Quo vadis, e-commerce?
Insights on and innovative approaches towards selected current challenges in the e-commerce context

Dissertation
zur Erlangung eines Doktors der Wirtschaftswissenschaft
der Rechts- und Wirtschaftswissenschaftlichen Fakultät
der Universität Bayreuth

vorgelegt
von
Theresa Maria Rausch
aus
Bayreuth

Dekan:	Herr Prof. Dr. Jörg Schlüchtermann
Erstberichterstatter:	Herr Prof. Dr. Daniel Baier
Zweitberichterstatter:	Herr Prof. Dr. Claas Christian Germelmann
Tag der mündlichen Prüfung:	22.04.2021

To my parents

Danksagung

Während meiner Promotion und der Zeit am Lehrstuhl für Marketing und Innovation haben mich viele Menschen begleitet, ohne die diese Zeit wahrscheinlich nicht möglich gewesen wäre. Ich möchte die Gelegenheit nutzen, um diesen Menschen zu danken und ihnen diese Seite zu widmen.

Zuerst möchte ich meinem Doktorvater, Herrn Prof. Dr. Baier, meine größte Dankbarkeit entgegenbringen. Er hat mich zur Promotion ermutigt und hat mich stets unterstützt. Er hatte immer ein offenes Ohr für mich und wertvolle fachliche Ratschläge parat. Sein Optimismus, seine Hingabe und die Fähigkeit, Menschen in seinem Umfeld anzutreiben und zu motivieren, waren sehr wertvoll für mich und meine Dissertation. Ohne ihn wäre diese Reise durch die wissenschaftliche und akademische Welt nie möglich gewesen – einen besseren Doktorvater hätte ich mir nicht wünschen können! Weiterhin möchte ich mich bei Herrn Prof. Dr. Germelmann bedanken, dass er die Rolle des Zweitgutachters übernommen hat und der mich während meines Bachelorstudiums dazu inspiriert hat, Marketing anhand der MuSe-Vertiefung im Masterstudium weiter zu verfolgen. Auch möchte ich mich bei PD Dr. Alexandra Rese bedanken, dass sie mir ebenfalls stets mit Rat und Tat beiseite stand.

Ebenfalls möchte ich mich bei meinen großartigen Kollegen bedanken: Benedikt Brand, Cristopher Kopplin und Karolina Ewers – wir hatten wirklich eine unvergessliche Zeit und ich bin extrem dankbar, dass ich solche Kollegen haben durfte. Egal ob wissenschaftlicher Diskurs, (künstlerisches) Brainstorming auf meinem Whiteboard oder anderweitige (gefühlt ewige) Gespräche, ohne euch wäre die Zeit am Lehrstuhl nicht annähernd so erinnerungswürdig gewesen – ich denke einen solchen Zusammenhalt im Team findet man nur selten. Danke für euer wertvolles Feedback und euren (vor allem zwischenmenschlichen) Rückhalt! Ich bin sehr stolz auf unsere vielen gemeinsamen Publikationen. Auch möchte ich mich bei Nicholas Derra und Tobias Albrecht für die tolle Zeit (auch während der langen Paper-Nächte) gemeinsam bedanken – dabei sind wirklich gelungene gemeinsame Publikationen entstanden. Bedanken möchte ich mich auch bei Jessica Kiesswetter für ihre tolle Organisation im Lehrstuhlalltag. Des Weiteren möchte ich mich bei allen Ko-Autoren bedanken, mit denen ich während meiner Promotionszeit publizieren durfte.

Zudem möchte ich mich bei meinen wunderbaren Eltern und meiner Oma bedanken – ihr habt mir sehr vieles im Leben ermöglicht und mich stets unterstützt. Ohne euch wäre die Promotion nicht möglich gewesen! Auch meine große Schwester Katharina war immer ein Vorbild für mich und ich konnte von ihrem selbstbewussten, starken Charakter einiges lernen. Erwähnen möchte ich auch meine restliche gesamte Familie – ich bin sehr dankbar, dass es euch gibt und ich einen solchen familiären Zusammenhalt erleben darf.

Zuletzt möchte ich mich bei allen meinen Freunden bedanken, die mich begleitet haben und mit denen ich wirklich alle meine Sorgen teilen konnte. Wir haben vieles gemeinsam erlebt und ich bin extrem dankbar, dass es euch gibt! Besonders hervorheben möchte ich meine langjährigen Freunde Maximilian Marquard, Laura Herath, Olga Havtyrko und Antonia Weber.

Abstract

E-commerce has seen a steady increase in usage since its establishment in the 1970s and 80s: By 2025, two-thirds of the world’s population (4,913.9M people) are expected to be e-commerce users. Throughout these decades, e-commerce businesses had to face a variety of different challenges, which, to some extent, determined their survival within their competitive environment. Within this thesis, two selected current phenomena are shed light on with which e-commerce businesses are struggling: A shift within society’s mindset towards environmental awareness and analytical approaches to manage the infinite pool of data about online consumer behavior. Since both research fields have an extremely granular spectrum of different facets, many sub-facets still lack a comprehensive investigation. The overall purpose of this research is thus twofold: (1) Gathering insights on consumers’ sustainable clothing consumption behavior and (2) proposing Artificial Intelligence-driven approaches for analytical problems in the e-commerce context.

More specifically, Part A focuses on consumers’ sustainable clothing consumption behavior as the textile industry causes an excessive environmental footprint considering valuable resources as ever inexhaustible and, simultaneously, yields the highest sales among all e-commerce segments. Research Paper No. 1 hence takes a macro-perspective on sustainable clothing consumption behavior by examining the determinants of consumers’ purchase intention for sustainable clothing and factors influencing the intention-behavior gap. Research Paper No. 2 and No. 3 take a deeper dive and provide micro-perspectives on the topic: the impact of specific sustainable clothing attributes on customer satisfaction is investigated (Research Paper No. 2). To complement these findings, the importance of specific sustainable clothing (and online shop) attributes is then compared to the importance of specific conventional clothing (and online shop respectively) attributes (Research Paper No. 3).

Within Part B of this thesis, Research Paper No. 4 and No. 5 focus on call center arrivals’ forecasting as call centers still constitute an essential customer touchpoint for e-commerce businesses: Reliable forecasts can enhance customer satisfaction with shortened waiting times and avoid overstaffing (and thus, unnecessary costs). Research Paper No. 4 therefore investigates the trade-off between accuracy and practicability of different machine learning models as these have been neglected by traditional forecasting literature. Research Paper No. 5 draws on these preceding findings and proposes a new dynamic harmonic regression model by incorporating the benefits of both approaches without (i.e., time series models) and with explanatory variables (i.e., machine learning and regression models). Research Paper No. 6 considers another prediction problem, which is particularly inherent to the online context of e-commerce, i.e., online shopping cart abandonment. It investigates the trade-off between accuracy and practicability of machine learning models for shopping cart abandonment prediction.

Overall, this thesis allows the reader to gather a better understanding of the underlying challenges by providing fruitful insights and proposes different approaches as a solution. Thereby, it makes several key contributions to extant literature and provides essential insights and implications for practitioners.

Table of contents

List of tables

List of figures

1 Introduction

Since the spread of the Internet's commercial usage, e-commerce sales are continuously increasing every year: sales are expected to rise from US$ 2,158,936M in 2020 to US$ 3,079,563M in 2025, with almost two-thirds of the world's population (4,913.9M people) being e-commerce users (Statista, 2021). Throughout past years, consumers' shopping behavior increasingly shifted from an offline to an online context and an enormous potential existed for the online purchase of goods and services.

From a consumer perspective, motives to shop (online) can be mainly differentiated into functional (i.e., utilitarian) and non-functional (i.e., hedonic) ones (Sheth, 1983). More specifically, functional motives aim to maximize the utility of the shopping experience (Babin et al., 1994), whereas non-functional motives relate to social and emotional needs (Babin et al., 1994; Childers et al., 2001). Established research considers online shoppers with functional motives to be concerned with purchasing products efficiently and in a timely manner to reduce their search costs (Alba et al., 1997; Childers et al., 2001) and, in contrast, non-functional shoppers' salient motives to conduct online purchases are escapism (Hirschman, 1983; Mathwick et al., 2001), fun as well as freedom (Wolfinbarger & Gilly, 2001), and entertainment (Childers et al., 2001; Mathwick et al., 2001).

While these consumers' core motives to shop online remained relatively stable over time, societal wants and interests are constantly shifting and thus, novel motives manifest whereas others vanish. One of the most controversially discussed topics within society throughout past years is climate change and its impact on environment. (Over-)exploitation of valuable resources to satisfy steadily increasing consumption levels cause an immense negative environmental impact with tons of CO_2 emissions. Therefore, modern consumers' awareness of their environmental footprint due to their overconsumption increased and led to an attitude shift (Paul & Rana, 2012; White et al., 2019). The shift in consumers' mindsets and the social change with respect to sustainability forces businesses to adapt their business models to thrive in the long term (Banerjee et al., 2003). Green consumers moved from a fringe to a mainstream issue for practitioners and, specifically, marketers, as their motives for shopping (online) change accordingly and substantially.

The negative environmental impact of rising consumption levels is particularly apparent in the textile industry with its fast-changing fashion trends and short life cycles: its global environmental impact is expected to amount 2,791M tons of CO_2 emissions and 118B cubic meters of water usage in 2030 (GFA & BCG, 2017). With the fashion industry yielding the highest sales within the e-commerce market among all segments (Statista, 2021), it seems inevitable to gather a comprehensive understanding of consumer behavior in the light of sustainable fashion e-commerce. As sustainability (especially in the context of clothing) particularly gained attention throughout recent years and is hence a rather novel challenge for marketers and e-commerce managers, research is still scarce and demanded (White et al., 2019). Part A therefore investigates sustainability as a challenge for e-commerce businesses and takes a macro- and a micro-perspective on consumers' sustainable clothing consumption behavior in the context of e-commerce.

Aside from the current shift towards sustainable consumption patterns, e-commerce businesses constantly had to face further challenges: a lack of trust of consumers (Gefen et al., 2003; Hoffman et al., 1999) and the shift towards mobile commerce (Ngai & Gunasekaran, 2007; Varshney & Vetter, 2002) are few examples that changed the e-commerce landscape since its emergence. While these aspects seem to be quite controllable in the meantime, exploding amounts of data about customers and their online behavior still constitute a major challenge for e-commerce businesses since several years. Artificial Intelligence and the era of big data triggered the *datafication* (Kelly & Noonan, 2017; Lycett, 2013) of consumer behavior and many e-commerce businesses cannot fully exhaust the multi-faceted potential of how to gather valuable information about their customers out of this data. This seems particularly critical considering that big data analytics ensures the e-commerce businesses' survival within their competitive environment, as they exhibit an estimated 5-6% higher productivity compared to their competitors when applying these methods (McAfee et al., 2012) and top-performing organizations were found to use analytics five times as often as lower performers (LaValle et al., 2011).

Artificial Intelligence and its sub-components are fundamentally changing how organizations make decisions and how organizations interact with external stakeholders (e.g., customers) (Kaplan & Haenlein, 2019). It can leverage opportunities to extract knowledge from highly granular, contextualized, and rich but complex data in a broad spectrum of high-impact areas within the organization (Chen et al., 2012). Thereby, the leading obstacles to adopting Artificial Intelligence-driven approaches are a lack of understanding regarding usage and improvement potential as well as a lack of skills (LaValle et al., 2011). Simultaneously, big data and Artificial Intelligence are considered a significant disruption within the academic ecosystem, causing a tsunami of scientific output to tackle these obstacles (Agarwal & Dhar, 2014). Although there is a myriad of publications, there are still many unexplored facets and application fields. Accordingly, Part B examines Artificial Intelligence as a challenge in the e-commerce context and, more specifically, approaches selected prediction problems. Figure 1 provides a summary of the structure and the included research papers.

The remainder is organized as follows: Section 2 comprises Part A, i.e., sustainability as a challenge in the e-commerce context. First, an overview of sustainability and sustainable fashion e-commerce is provided. Section 2.2 describes the content, research questions, and the context of Research Papers No. 1, 2, and 3. Part B (Section 3) investigates challenges regarding Artificial Intelligence in the e-commerce context. After introducing Artificial Intelligence and its applications fields in e-commerce, Section 3.2 elicits (similar to Section 2.2) the content, research questions, and context of Research Papers No. 4, 5, and 6. Section 4 draws an overall conclusion.

<table>
<tr><th colspan="5">Quo vadis, e-commerce?
Insights on and innovative approaches towards selected current challenges in the e-commerce context</th></tr>
<tr><th colspan="2">Part A: Sustainability in an e-commerce context</th><th colspan="3">Part B: Artificial Intelligence in an e-commerce context</th></tr>
<tr><td colspan="2">Macro-perspective
Research Paper No. 1: Bridge the gap: Consumers' purchase intention and behavior regarding sustainable clothing</td><td colspan="2" rowspan="2">Prediction problems 1 and 2:
Call center call arrivals' forecasting</td><td rowspan="2">Prediction problem 3:
Online shopping cart abandonment</td></tr>
<tr><td rowspan="2">Micro-perspective 1
Research Paper No. 2: The Drivers of Sustainable Apparel and Sportswear Consumption: A Segmented Kano Perspective</td><td rowspan="2">Micro-perspective 2
Research Paper No. 3: Does sustainability really matter to consumers? Assessing the importance of online shop and apparel product attributes</td></tr>
<tr><td>Research Paper No. 4: Call me maybe: Methods and practical implementation of artificial intelligence in call center arrivals' forecasting</td><td>Research Paper No. 5: Beyond the Beaten Paths of Forecasting Call Center Arrivals: On the Use of Dynamic Harmonic Regression with Predictor Variables</td><td>Research Paper No. 6: Predicting online shopping cart abandonment with machine learning</td></tr>
</table>

Figure 1: Overall structure.

2 Part A: Sustainability in an e-commerce context

2.1 Sustainability and sustainable fashion e-commerce

Although literature on sustainability mushroomed throughout the past few years as it became an extremely relevant topic within society, there is still no consensus on the definition of sustainability itself. Early definitions within literature in the 1970s and 1980s primarily focused on the ecological facet of sustainability and described concepts such as the long-term preservation of biological resources and maintenance of agricultural productivity (Brown et al., 1987; Conway, 1985) with a constant quality of environment and eco-systems as well as not exceeding the maximum carrying capacity of environment (i.e., the maximum population size that environment can support) (Odum, 1987). On a more abstract level, the United Nations considered development to be sustainable when it "meets the needs of the present without compromising the ability of future generations to meet their own needs" (United Nations, 1987, p. 37). Although the ecological perspective is by far the most considered throughout these decades, there were as well first discussions to incorporate an economic and social perspective into the definition of sustainability: Regarding the former, there was a controversy about the compatibility of (steady) economic growth and ecological sustainability, as economic growth was considered an inevitable consequence of increasing population size, the acquisitive nature of human beings, and technological progress (Brown et al., 1987). An economic definition of sustainability seemed rather elusive in the early 1980s (Brown et al., 1987) and was considered a fringe issue, dismissing it as an "alternative perspective[…] on sustainability" (Brown et al., 1987, p. 716). Social sustainability, in turn, was considered to aim at the survival and happiness of a maximum number of people and at fulfilling the minimum needs of the poorest members of society (Brown et al., 1987). From an ethical perspective, selfishness and competitiveness are replaced with values such as empathy and a sense of justice for all within a sustainable society (Milbrath, 1984).

The United Nations Conference on Environment and Development in 1992 created large-scale public awareness towards sustainability and can be considered the birth of the threefold definition of sustainability (United Nations, 1992): economic and social goals have to be treated equally to ecological goals for a sustainable development. All three facets are interdependent and have to be considered jointly. In an organizational context, this concept is nowadays more commonly referred to as the triple bottom line with its 3 P's (Elkington, 1997): The social facet (*people*) relates to employees, labor within the organization, and the community where an organization conducts its business. Organizations need to exhibit fair business practices and reciprocal behavior towards these stakeholder groups by making an effort to giving something back to society. The environmental facet (*planet*) requires organizations to benefit the environment, minimize their ecological footprint, or at least not to harm the environment. The *cradle to grave* principle is thereby an essential method to systematically assess the environmental impact of a product along its whole life cycle[1]. The economical facet (*profit*) captures the organization's

[1] This is closely related to the concept of life cycle assessment (ISO 2006).

impact on its economic local, national, and international environment by creating e.g. employment, wealth, and innovation. It shifts away from the traditional focus on financial profit a company makes and thus, accounts for the organization's societal and environmental impact as well.

Particularly in an e-commerce context, sustainability constitutes a severe issue along the whole life cycle of a product: aside from product-specific issues, working conditions and wages, amount and type of packaging, energy efficiency of the online shop and e-fulfillment center, e-fulfillment method, basket size, and return quota are – inter alia – critical aspects determining the sustainability of e-commerce (van Loon et al., 2015; Wiese et al., 2012). As the e-commerce market is continuously growing, yielding US$ 2,437,768M sales with 3,468.2M users worldwide in 2020 (Statista, 2021), it is inevitable to discuss issues as well as challenges surrounding sustainability within e-commerce. Especially green consumer behavior is frequently investigated within literature (see e.g., Chan, 2001; Kautish et al., 2019; Maichum et al., 2016; Paul et al., 2016; Taufique & Vaithianathan, 2018), as a shifted consumer behavior forces organizations, in turn, to adapt their internal structures.

As the fashion segment yields the highest revenues within the e-commerce market (US$ 665,629M in 2020 (Statista, 2021)) with most users worldwide (2,446.9M in 2020 (Statista, 2021)), it seems especially fruitful to gather more granular insights into consumer behavior within the fashion e-commerce segment. Since consumers demand contemporary, fast-changing styles due to their constantly varying preferences, rising consumption levels trigger a steadily growing market supply and thus, excessive usage of valuable natural resources (Achabou & Dekhili, 2013; Goworek et al., 2012). Consumers turned into a *throwaway society* perceiving the societal pressure to adopt the latest fashion trends leading to shortened life cycles of garments (i.e., fast fashion) and further, a loss of the garments' intrinsic value (Morgan & Birtwistle, 2009). Hence, consumers' mindsets in terms of clothing need to be shifted towards more sustainable ones and consumer behavior regarding sustainable clothing needs to be understood to minimize the negative environmental impact of the textile industry.

When investigating sustainable clothing consumption behavior, it becomes apparent that terms like organic (Hustvedt & Dickson, 2009), green (D'Souza et al., 2007), eco-conscious as well as eco-friendly (Hiller Connell, 2010; Laitala & Boks, 2012) are used synonymously as there is no industry standard uniformly defining sustainable clothing. Notwithstanding, literature mostly agrees that sustainable clothing consumption considers every phase along the garment's life cycle (Bianchi & Birtwistle, 2012; Lundblad & Davies, 2016). Thereby, sustainable clothing consumption behavior during the (pre-) purchase phase requires consumers to purchase garments made of environmentally friendly (e.g., recycled, upcycled, biodegradable) fibers, produced under environmentally friendly conditions (e.g., low pesticide, energy, and water usage), from second-hand shops or sharing economies, and – from a social perspective – garments manufactured under fair working conditions and with fair wages for the employees (Allwood et al., 2008; Armstrong et al., 2016; Bianchi & Birtwistle, 2012; Goworek et al., 2012; Joergens, 2006). Throughout the post-purchase phase, concepts concerning the garment's maintenance as well as care (e.g., laundering frequency, repairing), and discard (e.g., recycling, donation) to prolong the garment's life cycle are mostly discussed within literature (Armstrong et al., 2016; Goworek et al., 2018; Laitala & Boks, 2012; Morgan & Birtwistle, 2009).

2.2 A macro- and micro-perspective on sustainable fashion e-commerce

To gather a comprehensive understanding of sustainable clothing consumption behavior, a macro-perspective may be helpful to determine the antecedents of the purchase intention and behavior of sustainable clothing in a first step. Literature often draws on the Theory of Reasoned Action (TRA) (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975) (or its extension, the Theory of Planned Behavior (TPB) (Ajzen, 1985, 1988, 1991)) to investigate an individual's behavioral intention and the subsequent actual behavior across a variety of research areas, as behavioral intention was found to be the most immediate predictor of the respective actual behavior (Armitage & Conner, 2001; Bird, 1988; Locke & Latham, 2002). Within the TRA and the TPB, an individual's behavioral intention, in turn, is thereby assumed to be influenced by an intrinsic component (i.e., the individual's attitude A_B towards the behavior B) and an extrinsic component (i.e., the individual's social environment SN). With regard to attitude, behavioral beliefs b_i about the consequences (or outcome) of performing the behavior merge with evaluations e_i about the specific consequences ($\sum b_i e_i$) (Bagozzi, 1992):

$$A_B = \sum_{i=1}^{n} b_i e_i \tag{1}$$

Regarding subjective norm, normative beliefs b_j^* about whether the individual's significant others expect the individual to perform the behavior or not merge with the individual's motivation to comply m_j with these significant others ($\sum b_j^* m_j$) (Bagozzi, 1992):

$$SN = \sum_{j=1}^{k} b_j^* m_j \tag{2}$$

Actual behavior B as a proxy for behavioral intention BI can thus be defined as:

$$B \cong BI = \omega_1(A_B) + \omega_2(SN) \tag{3}$$

To further capture involitional behavior (Webb & Sheeran, 2006), the construct of perceived behavioral control was added within the TPB (Ajzen, 1991).

However, meta-analyses found the correlation of the intention-behavior relation to be only around 0.44 to 0.47 (Armitage & Conner, 2001) with vast proportions of variance in behavior remaining unexplained (Sheeran, 2002). Particularly in a green context (see e.g., Hughner et al., 2007; Kollmuss & Agyeman, 2002; Young et al., 2009), research found consumers to exhibit such an intention-behavior gap[2]: Albeit

[2] The intention-behavior gap is sometimes also referred to as the attitude-behavior gap (Kollmuss and Agyeman, 2002; White et al., 2019). A possible explanation for this may be that the intention-behavior relation within the TRA and TPB originally stems from preceding attitude-behavior models, which assume that an individual's attitude was the most accurate predictor for the subsequent behavior. However, research found inconsistencies in

consumers pretend a green attitude and intention, they struggle to translate this into environmentally friendly actions.

Exploratory research identified thus single potential barriers towards sustainable clothing consumption such as a lack of knowledge among consumers (Harris et al., 2016; Hiller Connell, 2010; Joergens, 2006), unstylish and unaesthetic appearance (Hiller Connell, 2010), high prices (Hustvedt & Dickson, 2009; Joergens, 2006), a lack of environmental concerns (Hustvedt & Dickson, 2009), or convenience as well as high search costs (Ellen, 1994). Nevertheless, research did not take a global perspective on sustainable clothing consumption by considering all crucial aspects identified by prior exploratory research and, more specifically, did not shed light on possible factors causing the intention-behavior discrepancy. Research Paper No. 1 takes a macro-perspective on sustainable clothing consumption by capturing these aspects with a structural equation model based on the TRA. It identifies key factors influencing the purchase intention for sustainable clothing and the intention-behavior relation, which might not only generate novel insights for e-commerce businesses, but further for brick-and-mortar stores. Research Paper No. 1 thus investigates the following research question:

RQ1: Which factors influence consumers' purchase intention for sustainable clothing as well as the intention-behavior gap?

After understanding the global determinants of sustainable clothing consumption behavior, it is inevitable to gather deeper insights into consumers' needs and wants. Gaining knowledge about consumers' attitude towards specific aspects concerning sustainable clothing (e.g. seals, materials, aesthetics) can then be used to ideally address these needs and wants. More specifically, Research Paper No. 2 takes a micro-perspective on sustainable clothing consumption behavior by inquiring consumers' importance of different sustainable clothing attributes as well as attributes of online shops distributing sustainable clothing. By using the Kano method (Kano et al., 1984), knowledge is gained about the effect of these attributes on consumers' satisfaction. Research Paper No. 2 intends to answer the subsequent research question:

RQ2: How does the existence (or absence) of specific sustainable clothing attributes influence customer satisfaction?

To complement these findings, Research Paper No. 3 contrasts the importance of sustainable clothing (and online shop) attributes with conventional clothing (and online shop respectively) attributes by using a best-worst scaling experiment (Finn & Louviere, 1992; Louviere & Woodworth, 1991). This allows deriving insights regarding the relative importance of sustainable clothing attributes and whether these are as important as conventional clothing attributes. Research Paper No. 3 further assesses consumers'

this relationship (LaPiere, 1934; Wicker, 1969) with attitude only weakly predicting actual behavior, and thus, intention was added to overcome this discrepancy (Fishbein and Ajzen, 1975).

willingness to pay for the investigated sustainable clothing attributes. The following research questions are examined within Research Paper No. 3:

RQ3: How important are specific sustainable clothing (and online shop) attributes to consumers compared to conventional clothing (and online shop respectively) attributes? What are consumers willing to pay for specific sustainable clothing attributes?

Table 1 sums up the publication status of the respective research papers. Figure 2 summarizes the research papers included in Part A.

Table 1: Publication status of research papers in Part A.

	Author(s) & Year	Title	Medium	Status
Research Paper No. 1	Rausch, T. M. & Kopplin, C. S. (2021)	Bridge the gap: Consumers' purchase intention and behavior regarding sustainable clothing	Journal of Cleaner Production, 278	Published
Research Paper No. 2	Baier, D., Rausch, T. M., & Wagner, T. F. (2020)	The Drivers of Sustainable Apparel and Sportswear Consumption: A Segmented Kano Perspective	Sustainability, 12 (7)	Published
Research Paper No. 3	Rausch, T. M., Baier, D., & Wening, S.	Does sustainability really matter to consumers? Assessing the importance of online shop and apparel product attributes	Journal of Retailing and Consumer Services	Under Review

Part A: Sustainability in an e-commerce context	
Macro-perspective: Research Paper No. 1: Bridge the gap: Consumers' purchase intention and behavior regarding sustainable clothing	
Content	Global and abstract perspective on the determinants of consumers' purchase intention for sustainable clothing and factors influencing the intention-behavior gap
Method	Structural equation model based on the Theory of Reasoned Action
Research question(s)	Which factors influence consumers' purchase intention for sustainable clothing as well as the intention-behavior gap?

Micro-perspective 1: Research Paper No. 2: The Drivers of Sustainable Apparel and Sportswear Consumption: A Segmented Kano Perspective		**Micro-perspective 2:** Research Paper No. 3: Does sustainability really matter to consumers? Assessing the importance of online shop and apparel product attributes	
Content	Granular perspective on the impact of specific sustainable clothing attributes on customer satisfaction	**Content**	Granular perspective on the importance of specific sustainable clothing (and online shop) attributes versus specific conventional clothing (and online shop respectively) attributes
Method	Kano method	**Method**	Best-worst scaling experiment
Research question(s)	How does the existence (or absence) of specific sustainable clothing attributes influence customer satisfaction?	**Research question(s)**	How important are specific sustainable clothing (and online shop) attributes to consumers compared to conventional clothing (and online shop respectively) attributes? What are consumers willing to pay for specific sustainable clothing attributes?

Figure 2: Summary of research papers in Part A.

2.2.1 Research Paper No. 1: Bridge the gap: Consumers' purchase intention and behavior regarding sustainable clothing

Authors: Rausch, T. M. & Kopplin, C. S. (2021)

Published in: Journal of Cleaner Production, 278, 1-15

DOI: https://doi.org/10.1016/j.jclepro.2020.123882

Abstract: With the textile industry satisfying steadily increasing consumption levels, excessive usage of valuable natural resources provokes a major environmental footprint: 118 billion cubic meters of water are expected to be utilized for global clothing production in 2030. Therefore, consumers' clothing consumption behavior needs to be shifted towards a more sustainable one. While green purchase behavior in general is well understood, research still lacks a comprehensive approach to explain consumers' purchase behavior of sustainable clothing. To provide a holistic framework which determines the main antecedents of purchase behavior of sustainable clothing and further, to shed light on the gap between purchase intention and subsequent purchase behavior of such clothes, we extended the Theory of Reasoned Action (TRA) approach with well-established constructs from green literature (i.e., perceived environmental knowledge and environmental concerns) and novel constructs derived from prior exploratory findings (i.e., greenwashing concerns, perceived economic risk, and perceived aesthetic risk). Four hundred sixty-four participants were inquired to assess these constructs in the context of sustainable clothing. Our findings indicate that attitude towards sustainable clothing has the highest impact on purchase intention. However, this relation is negatively influenced by consumers' greenwashing concerns. Moreover, we find evidence that consumers' perceived aesthetic risk negatively impacts the intention-behavior relation, whereas perceived economic risk has no significant effect on this relation.

Keywords: sustainable clothing consumption; intention-behavior gap; theory of reasoned action; purchase behavior; purchase intention; sustainability

1 Introduction

Steadily increasing consumption levels and consumer demand over the past decades led businesses to yield technological advances allowing for mass production and considering resources as ever inexhaustible (Csikszentmihalyi, 2000; Niinimäki & Hassi, 2011). Conventional business models primarily aim for profit maximization by satisfying growing demand disregarding the environmental facet of their actions. This phenomenon is particularly salient in the clothing industry, where manufacturing shifted to lower-cost countries with poor working conditions, price and quality of garments declined, and clothing's life cycle shortened to react to fast changing consumers' preferences and contemporary styles (Goworek et al., 2012). The demand for such fast fashion risen by the current 'throwaway society' and the subsequent growing market supply implies extreme obsolescence as well as a loss of intrinsic value of garments (Morgan & Birtwistle, 2009) and in turn, results in even more impulse purchasing and excessive waste of valuable resources (Achabou & Dekhili, 2013). The textile industry's environmental footprint negatively affects groundwater, air, and soil: its global environmental stress is expected to be around 2,791 million tons of CO_2 emissions, 118 billion cubic meters consumed water, and 148 million tons of textile waste in 2030 (GFA & BCG, 2017).

Due to increasing awareness of the clothing industry's resource intensity and its subsequent negative environmental impact, literature explored drivers and inhibitors of sustainable[3] clothing consumption. However, due to a lacking industry standard, sustainable clothing is not uniformly defined and terms like eco-conscious and eco-friendly (Hiller Connell, 2010; Laitala & Boks, 2012), ethical (Goworek et al., 2012; Joergens, 2006), green (D'Souza et al., 2007), and organic (Hustvedt & Dickson, 2009) are utilized interchangeably. Notwithstanding its different designations, there is consensus within literature on the conceptualization of sustainable clothing consumption behavior: it implies pro-environmental actions at every stage of the garment's life cycle from pre-purchase and purchase to post-purchase comprising its acquisition, storage, usage and care, maintenance, as well as discard (Bianchi & Birtwistle, 2012; Jacoby et al., 1977; Lundblad & Davies, 2016). Consequently, literature investigated how to minimize the negative environmental impact of the single stages. Thereby, sustainable behavior during the pre-purchase and purchase stages requires consumers to either purchase clothes made of environmentally preferable, recycled, upcycled, or biodegradable fibers manufactured under fair working conditions, or purchase garments from second-hand stores or sharing economies (Allwood et al., 2008; Armstrong et al., 2016; Goworek et al., 2012). Mostly, research focused on environmental issues occurring in the post-purchase stage by proposing strategies to prolong clothes' lifespans such as reusing (i.e., repairing, cleaning), recycling, and donation (Armstrong et al., 2016; Goworek et al., 2018; Laitala & Boks, 2012).

[3] The terms eco-conscious, environmentally/ecologically friendly, green, pro-environmental/ecological, and sustainable will be used interchangeably in this paper.

Albeit several concepts for sustainable clothing consumption have been proposed, most consumers still exhibit an intention-behavior gap regarding sustainable consumption, i.e. although they pretend a pro-environmental attitude and intention, they do not translate this into sustainable actions (Kollmuss & Agyeman, 2002; Young et al., 2009), particularly when it comes to the purchasing of sustainable clothes. Preliminary exploratory studies provide a number of aspects inhibiting green purchase behavior and its intention formation, respectively: interviews and focus group studies found limited knowledge (Harris et al., 2016; Hiller Connell, 2010; Joergens, 2006), the lack of environmental concerns (Hustvedt & Dickson, 2009), economic aspects (Hustvedt & Dickson, 2009; Joergens, 2006), unaesthetic appearance and fashion trend sensitivity (Hiller Connell, 2010; Lang et al., 2013), and high search costs (i.e., perceived time and effort) (Ellen, 1994) to be the main barriers for consumers to engage in sustainable consumption behavior.

Nevertheless, research still lacks a holistic framework investigating purchase intention as well as actual purchase behavior of sustainable clothing by integrating these preceding findings. Similarly to prior work investigating purchase behavior of sustainable products in general, we thus draw on the Theory of Reasoned Action (TRA) (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975) and extend it by employing well-known constructs from green literature as well as novel constructs derived from preceding exploratory findings. Thereby, we contribute to the body of knowledge by providing a thorough and comprehensive determination of established as well as unexplored, potential antecedents of consumer decision-making towards sustainable clothing consumption and further, by shedding light on the unexplored bivariate inconsistency between purchase intention and purchase behavior of sustainable clothes.

The remainder is structured as follows: The subsequent section reviews related work on sustainable clothing consumption and derives relevant constructs from prior findings as well as corresponding hypotheses. Section 3 describes the data collection, descriptive statistics, and items utilized in our questionnaire. Section 4 outlines the measurement and structural model evaluation. Section 5 discusses our contribution to the existing body of literature, managerial implications, enumerates limitations, and provides guidance for future research.

2 Related work and hypotheses

2.1 Purchase intention and purchase behavior

Across a variety of research fields such as entrepreneurial behavior (Kautonen et al., 2013; Kautonen et al., 2015; Shirokova et al., 2016), health-related behaviors (e.g., see Godin and Kok (1996) for a meta-analytic review), online purchase behavior (George, 2004; Pavlou & Fygenson, 2006), or ethical decisions (Shaw et al., 2000), behavioral intentions have been found to be immediate predictors of actual behaviors (Armitage & Conner, 2001; Bird, 1988; Locke & Latham, 2002). Thereby, scholars mostly exploited the insights of the TRA (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975) and its subsequent extension, the Theory of Planned Behavior (TPB) (Ajzen, 1985, 1988, 1991) to draw on the proposed intention-behavior relation and to investigate the antecedents of such behavioral intentions.

An essential impulse for the development of the TRA and the TPB, respectively, were preceding attitude-behavior models and more specifically, the identification of inconsistencies mentioned by – among others – LaPiere (1934) and Wicker (1969) indicating that an individual's attitude only weakly predicts actual behavior. This discrepancy provided a fruitful path for subsequent models in the late 1960s, combining these constructs with other factors to elucidate the attitude-behavior relation. Inter alia, the TRA (Fishbein & Ajzen, 1975) identified two additional constructs to overcome the bivariate inconsistency. First, a favorable attitude towards a specific behavior might not be translated into actual behavior due to a lacking social pressure from the individual's significant others or vice versa, the social pressure not to perform the behavior. Thus, in contrast to attitude capturing the personal influence on behavior, Fishbein and Ajzen (1975) suggested that measures of subjective norm capture the social influence on behavior. Second, attitude and subjective norm are assumed to affect behavior via a mediating cognitive link, i.e., the intention to perform the behavior. Behavioral intention captures motivational factors influencing the individual's behavior and reflects the amount of effort the individual is willing to exert (Ajzen, 1991). Thus, behavioral intention is considered to be the most immediate predictor of behavior with respect to the TRA and behavioral intention, in turn, is determined by attitude and subjective norm. Thereby, attitude is determined by behavioral beliefs (i.e., an individual's belief about the likelihood of the behavior's consequences) and subjective norm is determined by normative beliefs (i.e., an individual's belief about what relevant others think about the behavior).

The TRA was initially developed to predict volitional behavior, i.e., behavior over which the individual has control (Webb & Sheeran, 2006) or behavior which does not require skills, abilities, opportunities, or the cooperation of others (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975). However, this formulation was accused of creating a false dichotomy since most behavior is neither entirely volitional nor entirely involitional but ranges in between (Liska, 1984). Addressing this issue, Ajzen (Ajzen, 1985, 1988, 1991) added the concept of perceived behavioral control to the TRA yielding the TPB. Figure 1 depicts the TRA and the TPB.

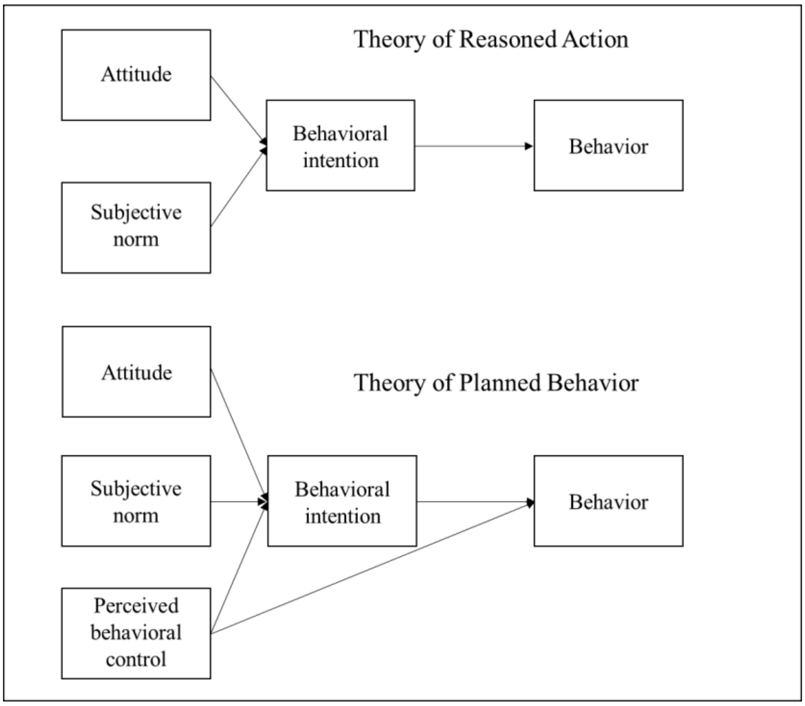

Figure 1: Theory of Reasoned Action and Theory of Planned Behavior.

Drawing on a sustainability context, both the TRA and the TPB were applied and further extended to investigate pro-environmental behaviors such as recycling (Cheung et al., 1999; Echegaray & Hansstein, 2017; Z. Wang et al., 2016), sustainable food consumption (Sparks & Shepherd, 1992; Vermeir & Verbeke, 2008), purchase behavior of energy-efficient products (Ha & Janda, 2012; Tan et al., 2017), purchase behavior of green cosmetic products (Hsu et al., 2017; Kim & Chung, 2011), or green purchase behavior in general (Chan, 2001; Kautish et al., 2019; Maichum et al., 2016; Paul et al., 2016; Taufique & Vaithianathan, 2018). Nevertheless, literature providing a comprehensive understanding of the determinants of consumers' purchase behavior for sustainable clothing is still scarce. We thus derive purchase intention, purchase behavior, attitude, and subjective norm from the TRA and the TPB as a basic framework for our model to investigate the phenomenon of sustainable clothing consumption:

H_1: Purchase intention for sustainable clothes has a positive impact on actual purchase behavior.

H_2: Attitude towards sustainable clothes has a positive impact on purchase intention.

H_3: Subjective norm has a positive impact on purchase intention for sustainable clothes.

2.2 Intention-behavior gap

Albeit intention is a good reference point to predict an individual's actual behavior, most people exhibit a substantial gap between their intentions and their subsequent behavior (Abraham et al., 1999; Bagozzi, 1992; Orbell & Sheeran, 1998). This intention-behavior gap was further identified in terms of sustainable consumption behavior, i.e. albeit consumers pretend to have pro-environmental intentions, they frequently struggle to translate them into green actions (Hughner et al., 2007; Kollmuss & Agyeman, 2002; Pickett-Baker & Ozaki, 2008; Young et al., 2009). Formally, meta-analyses of studies applying TRA and TPB found the intention-behavior correlation to be only 0.47 (185 studies) (Armitage & Conner, 2001) and 0.44 (28 studies) (Sheeran & Orbell, 1998) on average. Further, a meta-analysis of 10 meta-analyses indicated that intention accounted for only 28% of the variance in behavior on average (Sheeran, 2002), leaving substantial proportions of variance in behavior unexplained. Sheeran (2002) particularly identified – among others – properties of behavioral intentions and intention type to influence the degree of consistency between intentions and behavior.

Considering properties of behavioral intentions, prior research modeled different moderators intending to elucidate the intention-behavior discrepancy (see e.g., Sheeran (2002), Sheeran and Abraham (2003), or Webb and Sheeran (2006) for comprehensive reviews). It is assumed that people's intentions possess different dimensions or properties and thus, they might differ in the quality of their motivation or strength of their intention, respectively (Sheeran, 2002). Different properties affect the predictive ability of their intentions on actual behavior. For example, temporal stability of intentions (Sheeran & Orbell, 1998), past behavior (Kashima et al., 1993), self-schemas (Kendzierski & Whitaker, 1997), or anticipated regret (Sheeran & Orbell, 1999) are dimensions which might vary among individuals and thus affect predictive ability of their intentions, exhibiting a moderating effect on the intention-behavior relation.

Another line of research distinguished between different intention types occurring during different phases of the intention-behavior relation. Thereby, the lack of correspondence between behavioral patterns predicted by intentions and measures of actual behavior may be caused by two different groups: (1) intenders who do not transform their intention into subsequent action and (2) non-intenders who do take subsequent action (Abraham et al., 1999). The latter group requires exploring situational factors overcoming cognitive aversion to adopt new behaviors and thus, targeting intention formation. In contrast, the former group requires investigating cognitive changes other than those influencing intention formation (Abraham et al., 1999). Hence, it became common among social psychologists to distinguish between intention formation (or making a decision, respectively) and intention implementation (Ajzen, 1996; Beckmann & Kuhl, 1984; Kendzierski, 1990). Thereby, it was suggested that the intention-behavior relation encompasses four consecutive action phases (Gollwitzer, 1993): the (1) pre-decisional, (2) post-decisional but pre-actional, (3) actional, and (4) evaluative phases. Gollwitzer (Gollwitzer, 1990; 1993) detected obstacles preventing the successful realization of one's intentions to occur during the two pre-actional phases aligning with the mentioned distinction between intenders and non-intenders. Intentions associated with each of these two pre-actional phases can help to overcome these obstacles (Gollwitzer, 1990, 1993): The first pre-decisional phase involves deliberating wishes or desires and a consideration of desirability and feasibility of pursuing a goal. In case the wish is highly desirable and still feasible, the phase results in *goal intention* formation (or making the decision to perform a behavior respectively) (i.e., 'I intend to do X'). During the post-decisional but still pre-actional phase, an effective plan is formed specifying efforts to promote the initiation of relevant actions (i.e., 'I intend to do X in situation Y'). This plan is called *implementation* *intention* and commits the individual to a specific course of action underlying certain environmental conditions or situational factors (Gollwitzer, 1993). When these conditions are met, the performance of the intended behavior follows (and vice versa in case they are not met). Such situational factors or environmental conditions during the post-decisional (but still pre-actional) phase can thus strongly influence the intention-behavior relation and even inhibit the successful realization of an intended behavior.

With respect to the underlying sustainable clothing context, exploratory research identified several potential inhibitors to sustainable clothing consumption employing focus groups and interviews. First, sustainable apparel is frequently perceived as unfashionable or unstylish by consumers (Hiller Connell, 2010; Joergens, 2006) and does not match the perception of their lifestyle (Connolly & Prothero, 2003). They consider the appearance of sustainable fashion as unattractive and thus, it neither meets their wardrobe needs nor meets their aesthetic needs in contrast to conventional clothes. Harris et al. (2016) named the stigma and stereotypes associated with the design to be the key barriers to the mainstreaming of sustainable clothes. Second, consumers perceive the price of sustainable clothing (or sustainable products in general) as not comparable to conventional clothes (or conventional products, respectively) (Ali et al., 2011; Bray et al., 2011; Hustvedt & Dickson, 2009; Joergens, 2006; Young et al., 2009). Economic factors are found to have a strong influence on an individual's decisions and behavior (Kollmuss & Agyeman, 2002). Since only few technological advances have been made regarding the mass production of sustainable fashion, they often carry higher prices than conventional apparel, and thus are perceived as unaffordable to many consumers (Hiller Connell, 2010). Reflecting these insights regarding sustainable clothing in the light of prior intention-behavior findings, perceptions of aesthetic risk as well as economic risk might influence or even hinder the performance of an actual behavior after initial intention formation. That is, it is considerable that even though individuals initially form an intention towards sustainable clothing consumption, motivational quality differs among the individuals (Sheeran, 2002) and thus, high perceived aesthetic risk or economic risk might impact intention strength negatively during the post-decisional (or pre-actional respectively) phase. We thus hypothesize:

H_4: Perceived aesthetic risk negatively moderates the relationship between purchase intention and purchase behavior of sustainable clothes.

H_5: Perceived economic risk negatively moderates the relationship between purchase intention and purchase behavior of sustainable clothes.

2.3 Perceived environmental knowledge

Aside from the well-known constructs in the TRA and TPB, literature brought up several contextual factors which affect the purchase intention of individuals towards sustainable clothing embracing the traditional TRA and TPB approaches. Generally, behavioral literature reported a positive correlation between knowledge and actual behavior (Hoch & Deighton, 1989; Park et al., 1994). Reflecting these findings in a sustainability context, the measure of perceived environmental knowledge has been found to be an essential prerequisite of behavioral intention (or more specifically, purchase intention of sustainable products) (Chan, 2001; Kumar et al., 2017; Kwong & Balaji, 2016; Mostafa, 2006; P. Wang et al., 2014; Yadav & Pathak, 2016). Thereby, perceived environmental knowledge can be considered as an individual's "knowledge of facts, concepts, and relationships concerning the natural environment and its major ecosystems" (Fryxell & Lo, 2003). It is the state of individuals' knowledge about environment, the awareness of environmental issues, and the consciousness about consequences of human actions on the environment (do Paço & Reis, 2012; Kwong & Balaji, 2016). Within exploratory literature, consumers with greater environmental knowledge were found to be more likely to engage in eco-conscious clothing consumption (Harris et al., 2016; Hiller Connell, 2010). More specifically, consumers who are knowledgeable on environmental issues and impacts perceive a stronger responsibility towards environment and need for sustainable development (Fryxell & Lo, 2003) and further, are rather able to assess the environmental impact of conventional products. Thus, they may exhibit a higher purchase intention for sustainable products in order to meet their responsibilities.

Further, extant research substantiated the impact of perceived environmental knowledge as a cognitive component on green attitude formation (Jaiswal & Kant, 2018; Kumar et al., 2017; Maichum et al., 2016; Mostafa, 2007; Yadav & Pathak, 2016; Zhao et al., 2014). Knowledge enables consumers to differentiate the attributes and environmental impact of sustainable products from conventional products which in turn yields a positive, favorable attitude formation towards sustainable products (Kwong & Balaji, 2016; Pinto et al., 2011). Hence, we derive the following hypotheses:

H_6: Perceived environmental knowledge has a positive impact on purchase intention for sustainable clothes.

H_7: Perceived environmental knowledge has a positive impact on attitude towards sustainable clothes.

2.4 Environmental concern

Environmental concern (in some cases referred to as ecological affect) is an individual's extent of concern and emotional attachment towards environmental issues, environmental threats, and environmental protection, respectively (Chan, 2001; Crosby et al., 1981; Pinto et al., 2011). It is the individual's sense of responsibility and involvement regarding environmental protection (Dagher & Itani, 2014). Traditionally, environmental concern was considered to be a unidimensional construct ranging from unconcerned about the environment at the low end to concerned at the high end (Milfont & Duckitt, 2004). More sophisticated approaches assumed environmental concern to consist of concern for the self (egoistic), other people (altruistic), and the biosphere (biospheric) (Schultz, 2000). Notwithstanding the different conceptualizations of environmental concern, it established as a key construct within green behavioral literature: consistent empirical evidence has been found to support the relationship between environmental concern and purchase intention of sustainable products (Hartmann & Apaolaza-Ibáñez, 2012; Kwong & Balaji, 2016; Mostafa, 2006; Park & Lin, 2018; Prakash & Pathak, 2017) and actual purchase behavior (Lee et al., 2014).

Further, environmental concern focuses on an individual's affective evaluation of environmental issues (Newton et al., 2015). Since an individual's attitude comprises both cognitive as well as affective components to capture its knowledge and beliefs (Petty et al., 1991), prior research assumed environmental concerns to form an individual's attitude towards sustainable products aside from environmental knowledge (Chan, 2001; Jaiswal & Kant, 2018; Maichum et al., 2016; Mostafa, 2007; Yadav & Pathak, 2016). Consequently, we hypothesize that:

H_8: Environmental concern has a positive impact on purchase intention for sustainable clothes.

H_9: Environmental concern has a positive impact on attitude towards sustainable clothes.

2.5 Greenwashing concern

At its core, greenwashing is an organization's deceptive and misleading use of green marketing or green claims about the environmental impact of its products and practices in order to shape an overly positive public image and foster its reputation (Lyon & Maxwell, 2011; Lyon & Montgomery, 2015; Marquis et al., 2016). Greenwashers either choose to withhold negative information regarding their environmental impact or only partially disclose such information, and may even spread false positive information since they expect stakeholders to punish poor environmental performance (Lyon & Maxwell, 2011). Due to its increasing relevance in society, greenwashing has become a research hotspot in recent years (Bowen & Aragon-Correa, 2014; Seele & Gatti, 2017; Siano et al., 2017).

Research on the potential impact of an organization's greenwashing activities on consumers' green purchase intention and purchase behavior within the TRA and TPB frameworks is still sparse. Zhang et al. (2018) found consumers' greenwashing perception to negatively impact green purchase intention. Similarly, Kwong and Balaji (2016) found green skepticism to influence green purchase intention indirectly via environmental concern as well as environmental knowledge. This aligns with the findings of Mostafa (2006) who found consumers' skepticism towards environmental claims to be negatively related to green purchase intention.

We can thus assume a consumer's extent of suspicion towards an organization's intentional non-disclosure of negative environmental information or further, intentional disclosure of false positive environmental information about its products and practices, to affect the variables in the TRA and TPB framework. As stated in the preceding sections, the evaluative constructs attitude, subjective norm, perceived environmental knowledge as well as environmental concern are well-established immediate predictors of one's purchase intention towards sustainable products. Regarding an organization's environmental impact, consumers presume to be imperfectly informed due to non-transparent disclosure activities (Lyon & Maxwell, 2011). Thus, on the one hand, consumers are not fully aware of the true environmental impact of the considered product, and may have the suspicion that false positive claims are spread and negative environmental information is not disclosed. On the other hand, consumers cannot be completely sure whether and to which extent their greenwashing suspicions are legitimate. Due to this uncertainty regarding legitimation (in contrast to environmental concerns, for example), we assume a consumer's greenwashing concerns to influence the impact of attitude, subjective norm, perceived environmental knowledge, and environmental concern on purchase intention rather than having a direct effect on purchase intention. Therefore, we deduce the following hypotheses:

H_{10}: Greenwashing concern negatively moderates the relationship between perceived environmental knowledge and purchase intention for sustainable clothes.

H_{11}: Greenwashing concern negatively moderates the relationship between attitude towards sustainable clothes and purchase intention for sustainable clothes.

H_{12}: Greenwashing concern negatively moderates the relationship between environmental concern and purchase intention for sustainable clothes.

H_{13}: Greenwashing concern negatively moderates the relationship between subjective norm and purchase intention for sustainable clothes.

Table 1 summarizes the findings of extant literature on constructs and their relations derived for our study. Figure 2 displays the final research model.

Table 1: Extant (green) literature's findings on constructs and their relations.

Construct(s)	Description and relation(s)	Reference(s)
Attitude, Subjective norm, Purchase intention, Purchase behavior	Within the TRA and TPB, an individual's attitude and social influence on the individual are assumed to affect behavior via a mediating cognitive link, i.e., behavioral intention to perform the behavior	(Ajzen, 1985, 1988, 1991; Ajzen & Fishbein, 1980; Chan, 2001; Fishbein & Ajzen, 1975; Jaiswal & Kant, 2018; Kautish et al., 2019; Maichum et al., 2016; Paul et al., 2016; Taufique & Vaithianathan, 2018; Yadav & Pathak, 2016, 2017)
Perceived environmental knowledge	An individual's perceived environmental knowledge (awareness of environmental issues and consequences of human actions on environment) has been found to influence (1) purchase intention of sustainable products and (2) attitude towards sustainable products in prior studies	(Chan, 2001; Jaiswal & Kant, 2018; Kollmuss & Agyeman, 2002; Kumar et al., 2017; Kwong & Balaji, 2016; Maichum et al., 2016; Mostafa, 2006, 2007; P. Wang et al., 2014; Yadav & Pathak, 2016; Zhao et al., 2014)
Environmental concern	An individual's environmental concerns (sense of responsibility and involvement regarding environmental protection or issues) have been found to influence (1) purchase intention of sustainable products and (2) attitude towards sustainable products in prior studies	(Chan, 2001; Hartmann & Apaolaza-Ibáñez, 2012; Jaiswal & Kant, 2018; Kwong & Balaji, 2016; Maichum et al., 2016; Mostafa, 2006, 2007; Park & Lin, 2018; Paul et al., 2016; Prakash & Pathak, 2017; Yadav & Pathak, 2016)

Construct(s)	Description and relation(s)	Reference(s)
Greenwashing concern	An individual's extent of suspicion towards an organization's intentional non-disclosure of negative environmental information or intentional disclosure of false positive environmental information about its products and practices is assumed to affect the variables in the TRA and TPB framework. Due to imperfect information the individual can only be uncertain regarding the legitimation of its suspicions and thus, we assume greenwashing concerns to influence the relation between purchase intention and (1) attitude, (2) subjective norm, (3) perceived environmental knowledge, and (4) environmental concern	(Kwong & Balaji, 2016; Mostafa, 2006; Zhang et al., 2018)
Perceived economic risk, Perceived aesthetic risk	Albeit individuals pretend to have pro-environmental intentions, they frequently struggle to translate them into green actions. To elucidate the intention-behavior gap, we draw on exploratory literature's findings and assume perceived economic risk and perceived aesthetic risk to influence the purchase intention-purchase behavior relation	(Ali et al., 2011; Bray et al., 2011; Connolly & Prothero, 2003; Harris et al., 2016; Hiller Connell, 2010; Hughner et al., 2007; Hustvedt & Dickson, 2009; Joergens, 2006; Young et al., 2009)

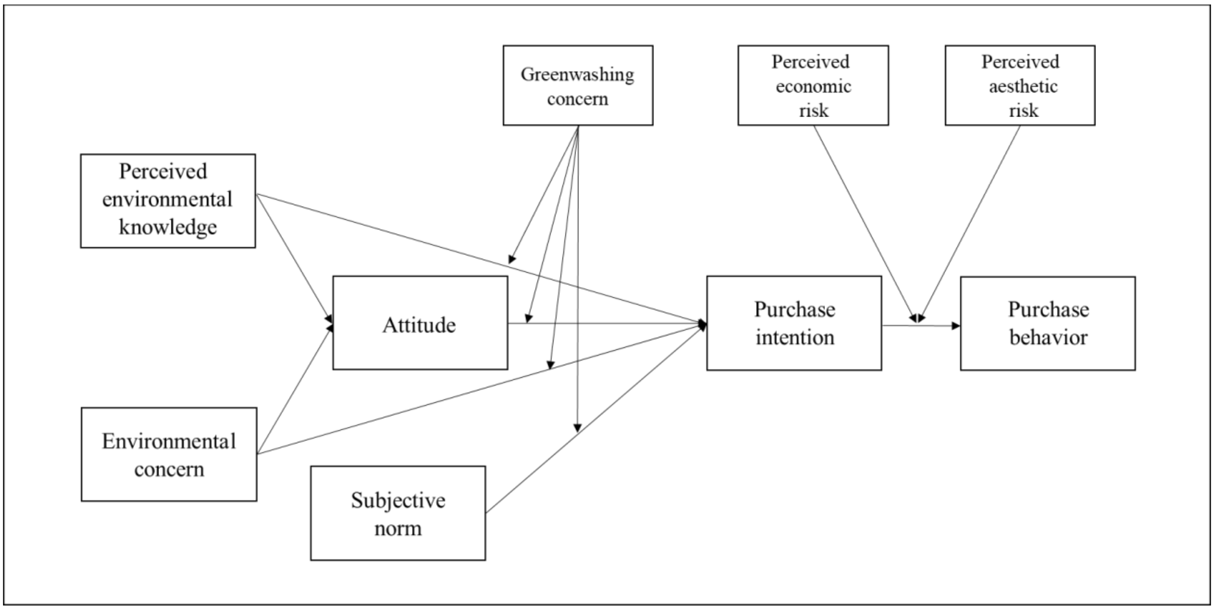

Figure 2: Research model.

3 Methodology

3.1 Data collection and descriptive statistics

To analyze the underlying constructs and their relations, an online questionnaire was developed using Qualtrics. Before conducting the main study, the questionnaire was pretested with 11 experienced participants to assess completeness, wording, clarity, structure, and appropriateness of the measurement items. After implementing minor modifications, the final questionnaire consisted of three major sections. We gained deeper insights into the participants' consumption behavior and perception of sustainability with four introductory questions: Participants were asked about their purchase frequency of clothing (items per month), their general perception of sustainability, their consumption frequency of sustainable products, and their general attitude towards sustainable products. In the main part, we first provided a scientific definition of sustainable clothing by describing the ‘from cradle to grave’ principle (i.e. the negative environmental impact of clothes has to be minimized throughout every consumption phase from acquisition through use and care to disposal in order to be deemed sustainable). Then, items measuring the constructs subjective norm (SN), attitude towards sustainable clothing (ATT), environmental concern (EC), perceived environmental knowledge (PEK), greenwashing concern (GC), perceived aesthetic risk (PAR), perceived economic risk (PER), purchase intention (PI), and actual purchase behavior (PB) were presented. The last part inquired participants’ demographics, i.e. gender, age, income, education level, employment status, and living conditions.

Data were gathered by spreading the self-administered questionnaire online across various social media channels and forums over the course of four weeks from March 26, 2020 to April 22, 2020 as we intended to target German online shoppers. The online context of our study can be deemed suitable with the international e-commerce market comprising 3,153.43m users worldwide in 2019 and more specifically, with the fashion segment yielding the highest revenue (i.e., 528,122.9m US dollar) among all market segments (Statista, 2020). A total of 553 responses was recorded. Eighty-nine (i.e. 16.09%) incomplete responses were excluded and thus, 464 responses were considered for further analysis.

Table 2 outlines the sample’s descriptive statistics und characteristics. Among the participants, 70.26% (n=326) were female. Age ranged from 15 to 77 with a mean of 30.49 years. Most participants were between 20 and 29 years old (n=274, 59.05%). Only 22.63% of the participants had a monthly income higher than 2001 Euros (n=105). Thus, our sample mainly comprises online shoppers with low or medium income. The majority of the participants was employed (n=208, 44.83%). Further, most participants’ highest education level was a high school diploma or below (n=347, 74.78%). Participants with a bachelor’s degree or above constituted a smaller proportion among the respondents (n=108, 23.27%).

Regarding their average purchase frequency of clothes, most participants indicated to buy one or two garments (n=215, 46.34%) or even less than one garment per month (n=168, 36.21%). Two hundred sixty-seven participants (57.54%) stated to purchase sustainable products occasionally, whereas only 86 participants (18.53%) indicated to buy sustainable products predominantly. However, most participants’ overall attitude towards sustainable products was positive (n=371, 79.95%).

Table 2: Descriptive statistics (n=464).

Demographics/Characteristics	Specifications	Counts	Proportion (in %)
Age	≤19 years	40	8.62
	20-29 years	274	59.05
	30-39 years	42	9.05
	40-49 years	43	9.27
	≥50 years	65	14.01

Gender	Female	326	70.26
	Male	137	29.53
	Diverse	1	0.21
Monthly income	≤1000 Euros	179	38.58
	1001-2000 Euros	136	29.31
	2001-3000 Euros	80	17.24
	≥3001 Euros	25	5.39
	No information provided	44	9.48
Education	High school or below	347	74.78
	Bachelor's degree	77	16.59
	Master's degree or above	31	6.68
	Other	9	1.94
Employment status	Student	160	34.48
	Self-Employed	8	1.72
	Employee	208	44.83
	Housewife/Househusband	13	2.80
	Unemployed	1	0.21
	Retiree	11	2.37
	Other	63	13.58
Purchase frequency of clothes per month	Less than one garment	168	36.21
	1-2 garments	215	46.34
	3-5 garments	71	15.30
	6-7 garments	6	1.29
	More than seven garments	4	0.86
Consumption frequency of sustainable products	Never	5	1.08
	Rarely	105	22.63
	Occasionally	267	57.54
	Mostly	86	18.53
	Always	1	0.21
Overall attitude towards sustainable products	Very negative	1	0.21
	Negative	1	0.21
	Neutral	91	19.61
	Positive	247	53.23
	Very positive	124	26.72

3.2 Measurement items

All constructs were measured using multiple items on a five-point Likert-type scale (1 = 'Strongly disagree' to 5 = 'Strongly agree'). The items contained an explicit key expression reflecting the specific construct. All items were derived from the literature and thus based on scales that have been previously validated. Since literature on sustainable clothing purchase behavior is still sparse, we drew on green purchase behavior literature for the established constructs and adapted the items to our context accordingly. Measures for greenwashing concern, perceived aesthetic risk, and perceived economic risk were based on previous exploratory findings or were derived by further development of related scales. Table 3 provides the items of each construct.

We assessed the participants' attitude towards sustainable clothes by adopting the measures of Park and Lin (2018) and further, by marginally adapting the scale from Chan (2001) stemming from Li (1997). Subjective norm was measured using the scale from Vermeir and Verbeke (2008). Items for perceived environmental knowledge were adapted from Ellen et al. (1997). Measures for environmental concern were formed by adapting scales from Lee (2008) and Dunlap et al. (2000). For the measurement of greenwashing concern, we generally based our items on the greenwashing perception or skepticism constructs of Chen and Chang (2013), Mohr et al. (1998), and Zhang et al. (2018), but we assume greenwashing to be an affective construct reflecting the consumer's suspicion of false environmental claims and simultaneously, consumer's uncertainty whether and to which extent his/her greenwashing suspicions are legitimate. This uncertainty in turn is being expressed in concerns. The measures for perceived aesthetic risk were operationalized from prior exploratory findings by Hiller Connell (2010) and Joergens (2006). Regarding the scale of perceived economic risk, we drew on Park and Lin (2018). The first endogenous variable, purchase intention towards sustainable clothing, was measured using four items derived from Park and Lin (2018) and Kumar et al. (2017). Measures for the second endogenous variable, purchase behavior towards sustainable clothing, were adopted from Lee (2008) and Schlegelmilch et al. (1996).

Table 3: Constructs, items, and references.

Construct	Item		Reference(s)
Attitude (ATT)	ATT1	Generally, I have a favorable attitude towards the sustainable version of clothes.	(Chan, 2001; Li, 1997; Park & Lin, 2018)
	ATT2	I am positive minded towards buying second hand clothes.	
	ATT3	I like the idea of buying sustainable clothes instead of conventional clothes to contribute to environmental protection.	
Subjective norm (SN)	SN1	My friends expect me to buy sustainable clothes.	(Vermeir & Verbeke, 2008)
	SN2	My family expects me to buy sustainable clothes.	
	SN3	People who are important to me expect me to buy sustainable clothes.	
Perceived environmental knowledge (PEK)	PEK1	I know how to behave sustainably.	(Ellen et al., 1997)
	PEK2	I know how I could lower the ecological harm with my behavior.	
	PEK3	I understand how I could reduce the negative environmental consequences of my behavior.	
	PEK4	I understand how to protect the environment in the long-term.	
Environmental concern (EC)	EC1	I am concerned about the environmental development.	(Dunlap et al., 2000; Lee, 2008)
	EC2	I am concerned about the long-term consequences of unsustainable behavior.	
	EC3	I often think about the potential negative development of the environmental situation.	
	EC4	I am concerned that humanity will cause a lasting damage towards the environment.	
Greenwashing concern (GC)	GC1	I am concerned that sustainable clothes are not produced of environmentally friendly materials.	(Chen & Chang, 2013; Mohr et al., 1998; Zhang et al., 2018)
	GC2	I am concerned that sustainable clothes are not manufactured under sustainable conditions.	
	GC3	I am concerned that the organization is only pretending its green image.	
Perceived aesthetic risk	PAR1	Sustainable clothing does not meet my aesthetic needs.	(Hiller Connell, 2010; Joergens, 2006)
	PAR2	Sustainable clothing does not match my clothing style.	

Construct	Item		Reference(s)
(PAR)	PAR3	Sustainable clothing does not meet my taste in clothing.	
Perceived economic risk (PER)	PER1	In my opinion, sustainable clothing is more expensive than conventional clothing.	(Park & Lin, 2018)
	PER2	I am worried about not getting my money's worth if I buy sustainable clothes instead of conventional clothes.	
	PER3	I think I would have to spend more for the sustainable version of a garment.	
Purchase intention (PI)	PI1	I consider purchasing sustainable clothes.	(Kumar et al., 2017; Park & Lin, 2018)
	PI2	I intend to buy sustainable clothes instead of conventional clothes in the future.	
	PI3	I might possibly buy sustainable clothes in the future.	
	PI4	I would consider to buy sustainable clothes if I happen to see them in a(n) (online) store.	
Purchase behavior (PB)	PB1	I choose to buy exclusively sustainable clothes.	(Lee, 2008; Schlegelmilch et al., 1996)
	PB2	I buy sustainable clothes instead of conventional clothes if the quality is comparable.	
	PB3	I purchase sustainable clothes even if they are more expensive than conventional clothes.	
	PB4	When buying clothes, I pay attention that they are sustainable.	

4 Results

4.1 Measurement model evaluation

Following the two-step analysis approach used in partial least squares structural equation modeling (PLS-SEM), model evaluation starts with the outer or measurement model. The algorithm is set to path weighting scheme, allowing 300 iterations at maximum and using a stop criterion of 10^{-7}. Results converged after two iterations. Outer loadings are checked employing a threshold of 0.708 (Hair et al., 2019), finding that all indicators survive. Construct reliability and validity are assessed drawing on composite reliability (CR) and average variance extracted (AVE), however, Cronbach's α is also provided due to the measure's high profile. All values exhibit satisfying values. Table 4 summarizes the results. The indicators' covariance matrix is provided in Table A.1 in the Appendix.

Table 4: Assessment of convergent validity and internal consistency reliability.

Latent variable	Indicators	Mean (SD)	Cronbach's α	CR	AVE
PAR	3	2.761 (1.012)	0.924	0.952	0.868
PER	3	3.465 (0.873)	0.777	0.862	0.675
ATT	3	3.752 (0.816)	0.757	0.860	0.674
EC	4	4.195 (0.613)	0.826	0.884	0.657
PEK	4	3.952 (0.588)	0.809	0.875	0.636
GC	3	3.392 (0.920)	0.865	0.917	0.786
PI	4	3.665 (0.777)	0.891	0.925	0.757
PB	4	2.689 (0.889)	0.854	0.901	0.696
SN	3	2.685 (0.969)	0.852	0.909	0.770

Note: ATT = attitude towards sustainable clothing, AVE = average variance extracted, CR = composite reliability, EC = environmental concern, GC = greenwashing concern, PAR = perceived aesthetic risk, PEK = perceived environmental knowledge, PER = perceived economic risk, PI = purchase intention for sustainable clothes, PB = purchase behavior, SD = standard deviation, SN = subjective norm.

Next, discriminant validity is checked. Cross-loadings, the Fornell-Larcker criterion (Fornell & Larcker, 1981), and heterotrait-monotrait ratio (HTMT, Henseler et al., 2015) are employed for analysis. Cross-loadings and Fornell-Larcker tabulation are provided in Tables A.2 and A.3 in the Appendix and HTMT results are displayed in Table 5. Considering HTMT, all pairings except for PI and ATT pass the conservative threshold of 0.85, while PI and ATT still meet the rather liberal value of 0.90 (Henseler et al., 2015). In order to derive 95 percent confidence intervals, a bootstrapping procedure drawing 10,000 samples is conducted. The critical value of 1 is excluded from all intervals, further corroborating discriminant validity. The bootstrapping run further corroborates that lower and upper limits for Cronbach's α and CR do not overshoot 0.70 and 0.95, respectively.

Table 5: Assessment of discriminant validity.

	PAR	PER	ATT	EC	PEK	GC	PI	PB	SN
PAR									
PER	0.257								
ATT	0.554	0.113							
EC	0.257	0.105	0.642						
PEK	0.133	0.067	0.412	0.352					
GC	0.044	0.096	0.162	0.169	0.089				
PI	0.534	0.099	0.875	0.573	0.409	0.171			
PB	0.491	0.202	0.696	0.405	0.376	0.114	0.727		
SN	0.075	0.060	0.281	0.311	0.179	0.090	0.228	0.375	

Note: ATT = attitude towards sustainable clothing, EC = environmental concern, GC = greenwashing concern, PAR = perceived aesthetic risk, PEK = perceived environmental knowledge, PER = perceived economic risk, PI = purchase intention for sustainable clothes, PB = purchase behavior, SN = subjective norm.

Assessment of the measurement model indicates absence of measurement problems. Construct reliability and validity and discriminant validity could be established.

4.2 Structural model evaluation

Moving on to evaluating the structural model, variance inflation factors (VIFs) are checked. Values are rather low, ranging from 1.066 to 1.444. Consequently, VIFs meet the conservative threshold of 3 for absence of collinearity issues (Hair Jr et al., 2016) and the threshold of 3.3 for common method bias (Kock, 2015). Next, R^2 values are checked, exhibiting 0.300 for ATT, 0.575 for PI, and 0.451 for PB (R^2 Adjusted: 0.297 for ATT, 0.567 for PI, and 0.445 for PB). Overall, in-sample predictive power can be considered moderate (Hair et al., 2019; Rigdon, 2012). As one of our main aims is to shed light on the intention-behavior gap and the moderating influences of PAR and PER, R^2 as measure for "explanatory modeling efforts" (Shmueli et al., 2016, p. 4555) is favored as quality criterion and preferred to Q^2. Due to completeness, however, a blindfolding procedure is used to derive Q^2 values for the endogenous constructs, yielding values of 0.195 for ATT, 0.425 for PI, and 0.306 for PB, respectively. These can be considered medium to large and indicate (pseudo) out-of-sample prediction ability (Hair et al., 2019). Having ensured that all measures work correctly, hypotheses are tested using a bootstrapping procedure with 10,000 subsamples. Point estimators as well as 95 percent confidence intervals are derived. Table 6 displays the results.

Table 6: Hypotheses testing.

Hypothesis				Path coefficients (effect size f^2)	Confidence intervals (bias-corrected, 95%)	T-statistics (p-value)
H1	PI	→	PB	0.594 (0.457)	[0.514, 0.669]	15.089 (< 0.001)
H2	ATT	→	PI	0.599 (0.565)	[0.526, 0.672]	16.129 (< 0.001)
H3	SN	→	PI	0.013 (< 0.001)	[-0.050, 0.076]	0.421 (0.674)
H4	PAR*	→	PI → PB	-0.107 (0.022)	[-0.171, -0.042]	3.270 (0.001)
H5	PER*	→	PI → PB	0.027 (0.001)	[-0.036, 0.089]	0.844 (0.399)
H6	PEK	→	PI	0.098 (0.019)	[0.030, 0.165]	2.814 (0.005)
H7	PEK	→	ATT	0.192 (0.048)	[0.094, 0.283]	3.965 (< 0.001)
H8	EC	→	PI	0.146 (0.034)	[0.067, 0.226]	3.576 (< 0.001)
H9	EC	→	ATT	0.459 (0.276)	[0.361, 0.550]	9.489 (< 0.001)
H10	GC*	→	PEK → PI	-0.049 (0.005)	[-0.118, 0.013]	1.449 (0.147)
H11	GC*	→	ATT → PI	-0.108 (0.019)	[-0.194, -0.016]	2.401 (0.016)
H12	GC*	→	EC → PI	0.088 (0.013)	[0.004, 0.173]	2.038 (0.042)
H13	GC*	→	SN → PI	0.013 (< 0.001)	[-0.052, 0.075]	0.394 (0.693)

Note: Asterisk (*) indicates moderating effect. ATT = attitude towards sustainable clothing, EC = environmental concern, GC = greenwashing concern, PAR = perceived aesthetic risk, PEK = perceived environmental knowledge, PER = perceived economic risk, PI = purchase intention for sustainable clothes, PB = purchase behavior, SN = subjective norm.

All hypotheses except for H3, H5, H10, H12, and H13 could be supported. PEK was found to positively impact both ATT and PI and EC showed the same influences. PI exhibits a moderate positive effect on PB. Moderators GC, PAR, and PER yield mixed results. PAR indeed does have a negative impact on the relation between PI and PB. However, this relation could not be supported for PER. GC's influence, which was hypothesized to moderate impacts on PI, could be confirmed for only one path, namely ATT to PI. While its negative influence on the ATT-PI relation appears reasonable, GC's positive effect on the EC-PI path, which was also detected, is counterintuitive. However, as the 95 percent confidence interval (i.e., [0.004, 0.173]) suggests, the lower interval boundary is very close to zero and therefore, the statistical significance may be a mathematical artifact. No evidence of GC affecting the relations of PEK and PI as well as SN and PI was found. Confidence intervals, which may also be interpreted as compatibility intervals spanning ranges particularly compatible with the data (Greenland, 2019), and f^2 values emphasize striking positive impacts of PI on PB and ATT on PI. ATT appears to be the major driver of PI, while EC has a higher influence on ATT compared to PEK. Assessing the hypothesized moderating effects, both confidence intervals and f^2 values indicate rather weak (Cohen, 1988), however statistically convincing impacts. Figure 3 summarizes the results from structural model evaluation.

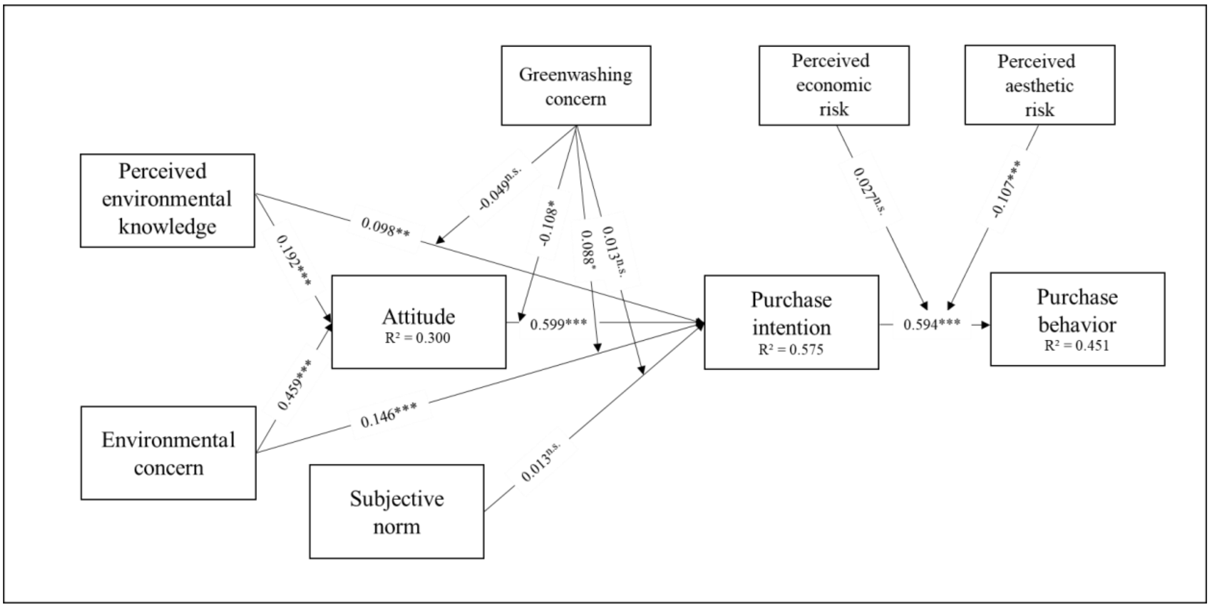

Figure 3: Structural model evaluation.

Note: $* = p < 0.05$, $** = p < 0.01$, $*** = p < 0.001$, n.s. = not significant.

5 Discussion

5.1 Theoretical contribution

While green purchase behavior is well elucidated by preceding literature (Chan, 2001; Jaiswal & Kant, 2018; Kautish et al., 2019; Kumar et al., 2017; Maichum et al., 2016; Paul et al., 2016; Taufique & Vaithianathan, 2018; Yadav & Pathak, 2016, 2017) with frameworks like TRA and TPB, research on sustainable clothing purchase behavior is sparse. To contribute to the existing body of literature by providing a holistic framework which determines the main antecedents of purchase intentions for sustainable clothing and further, by shedding light on the gap between purchase intention and subsequent purchase behavior of such clothes, we extended the TRA with well-established constructs from green literature (i.e. perceived environmental knowledge and environmental concerns) and novel constructs derived from prior exploratory findings (i.e. greenwashing concerns, perceived economic risk, and perceived aesthetic risk). Extant sustainable clothing literature drew on exploratory approaches (Harris et al., 2016; Hiller Connell, 2010; Joergens, 2006) or investigated purchase intention and purchase behavior separately (Park & Lin, 2018), whereas this study is one of the first in the sustainable clothing context intending to explain purchase intention, actual purchase behavior, and the intention-behavior gap with an extended TRA model.

Thereby, hypotheses derived from the TRA were corroborated in the context of sustainable clothing except for the relation between subjective norm and purchase intention. Our results thus mostly align with findings of preceding literature in the context of green purchase behavior in general (Chan, 2001; Jaiswal & Kant, 2018; Kumar et al., 2017; Yadav & Pathak, 2016). Also, the lack of evidence for the impact of subjective norm on purchase intention has already been found in green purchase behavior literature (Kumar et al., 2017; Park & Lin, 2018).

Greenwashing concerns indeed appear to influence consumer decisions on the intention-formation level as they were found to moderate the relation between attitude and purchase intention. In contrast, Zhang et al. (2018) modeled greenwashing perception as an immediate antecedent of green purchase intention and Kwong and Balaji (2016) found green skepticism to impact environmental knowledge and concerns. This study presumes greenwashing concerns to incorporate a consumer's suspicion about an organization's greenwashing activities but due to imperfect information the consumer can only be uncertain regarding the legitimation of his/her suspicions and thus, we assume (and partially confirmed) greenwashing concerns to impact the relation between purchase intention and its antecedents rather than having a direct effect on purchase intention.

We further yield new insights by elucidating that perceived aesthetic risk affects the relation between purchase intention and actual purchase behavior negatively, which represents a starting point to bridge the frequently identified gap (Kollmuss & Agyeman, 2002; Young et al., 2009) between the two variables. We show that sustainable clothing is apparently still associated with certain stereotypes implying an unfashionable perception among consumers which hinders the purchasing of such clothes despite successful initial intention formation. Moreover, we were not able to find evidence that perceived economic risk has an impact on the intention-behavior relation of sustainable clothing. Thus, we cannot confirm preceding exploratory findings from the early 2000s (Hiller Connell, 2010; Hustvedt & Dickson, 2009; Joergens, 2006) indicating that consumers perceive sustainable clothing as more unaffordable than conventional clothes.

5.2 Practical implications

Our findings regarding the determinants of consumers' purchase intention and purchase behavior of sustainable clothing provide several implications. Apart from clothing (online) retailers and manufacturers, several stakeholders (e.g., the government accomplishing its climate targets) might be interested in enhancing the purchase intention and further, purchase behavior of sustainable clothing. Particularly, findings regarding potential impacts of perceived aesthetic and economic risk on the intention-behavior relation and the influence of greenwashing concerns yield new and valuable insights.

Impacts of aesthetic and economic risks on the intention-behavior relation show the relevance of aesthetic worries over economic ones: Consumers' perceived economics risk towards sustainable clothing did not have an impact on the relationship between consumers' purchase intention and purchase behavior of sustainable clothes. As modern consumer environments provide a variety of data sources (e.g., platforms for exchange of experience such as social media, blogs, online reviews, and comparison websites), it may be rather easy for potential customers to collect and analyze information subjectively deemed necessary to make a decision. Hence, an individual's perceived risk of economic drawbacks may be attenuated in a way that once a purchase intention has been formed, it is translated into a behavior without regarding economic risk as a potential barrier.

From a managerial point of view, actions to mitigate the impact of consumers' greenwashing concerns and perceived aesthetic risk might yield more promising effects. The identified moderating effects of greenwashing concerns indicate that consumers' growing concerns about an organization withholding its negative environmental impact or even spreading false environmental claims significantly reduces consumers' intention to buy from that vendor. More specifically, greenwashing concerns influenced the impact of the participants' attitude towards sustainable clothing on their purchase intention for sustainable clothes. High transparency standards and established as well as renowned certificates may help to reduce imperfect information, i.e. consumers' uncertainty regarding the legitimation of their suspicion regarding the organization's disclosure activities. Moreover, clothing retailers may publish an annual sustainability report certified by independent auditors to verify the authenticity of their disclosed environmental claims. Government may impose strict penalties when false information is disclosed. Further, consumers that are able to retrace a product's fabrication process possess sufficient information for rational decision making and may decide which manufacturing step is the most important to them depending on their individual preferences. For example, while one consumer may emphasize ecological impacts of the manufacturing process itself such as water and energy use, working conditions, or CO_2 emissions, another might focus on the product's materials and their environmental impact during exploitation, manufacturing, and disposal such as pesticide use, materials' recyclability or biodegradability, and origin. Thus, clothing retailers are recommended to allow their customers to track the product's material origin and the manufacturing process.

Since we found consumers' perceived aesthetic risk to negatively influence the intention-behavior relation, several measures can be implemented to proactively avoid the subsequent potential gap between purchase intention and purchase behavior. Mitigating aesthetic risk also refers to transparency, yet in a slightly different way. Consumers need the possibility to get a true-to-life idea of the product before purchase, which may be carried out drawing on technological implementations such as 360 degree images, videos, and close-up images of details. Further, reviews from previous customers (which may be enhanced through means of videos and images) help getting an overview of the product in an everyday context from other consumers. Apparently, since consumers still perceive sustainable clothing as unfashionable not meeting their aesthetic needs, consumers' minds need to be shifted towards a more modern perception of such garments. Clothing retailers and manufacturers can overcome stereotypes and stigmata associated with sustainable clothes by cooperating e.g. with influencers or celebrities who promote and wear environmentally friendly apparel and thus, serve as a role model and nudge consumers towards the adoption of such clothing consumption behavior.

Further, since attitude towards sustainable clothes was found to have the largest impact on subsequent purchase intention, consumers' attitude needs to be shifted. Perceived environmental knowledge and particularly, environmental concerns were found to be essential cognitive and affective components forming consumers' attitudes. Hence, society's environmental knowledge and concerns need to be further enhanced with broad public campaigns to make consumers aware of environmental problems.

5.3 Limitations and future research

The study at hand was conducted examining clothing as example for sustainable products. Future investigations may assess the roles of greenwashing concerns and aesthetic as well as economic risks in other contexts to draw a generalized picture of the constructs' effects. Further research on economic risk may help to evaluate whether it does indeed not have a striking impact on the buying process or whether it is rather context-dependent. Particularly in online shopping scenarios, it is easy for potential customers to review a variety of alternative offers, to check for the best price over a myriad of vendors and distributors, and to incorporate public feedback into their decision-making. Further, the study was conducted in Germany and economic risks might be perceived as more severe in other countries. Moreover, as stated in earlier studies (Li, 1997), environmental concerns and environmental knowledge might vary by country. As we did not refer to a specific manufacturer or clothing company in our study, some constructs and particularly, greenwashing concerns may have appeared somewhat abstract to the respondents and thus, this may have influenced the results regarding greenwashing concerns.

As for all scientific studies, several methodological limitations need to be addressed. First, while the sample size is considerably large to draw statistical conclusions, it was collected by distributing the questionnaire across multiple social media channels. Hence, we cannot be sure whether the sample population is a representative instance of the target population that is interested in buying sustainable clothing. Constructs were measured using Likert-type scales for self-reporting. In the context of sustainability, which may be subject to social desirability and peer pressure, participants' evaluation of environmental knowledge and environmental concerns may be biased towards the high end.

6 Conclusion

With sustainability being an increasingly socially relevant issue, the textile industry, which causes a substantial environmental footprint, needs to experience a paradigm shift. Thereby, identifying consumers' motives for buying sustainable clothing constitutes a major challenge. Our study provides insights into the main antecedents of purchase behavior of sustainable clothing and further sheds light on the gap between purchase intention and subsequent purchase behavior. Therefore, we extended the TRA with well-established constructs from green literature (i.e., perceived environmental knowledge and environmental concerns) and novel constructs derived from prior exploratory findings (i.e., greenwashing concerns, perceived economic risk, and perceived aesthetic risk). Four hundred sixty-four participants evaluated these constructs in the context of sustainable clothing. Our findings show that attitude towards sustainable clothing has the highest impact on purchase intention and that consumers' greenwashing concerns negatively moderate this relation. We prove that consumers' perceived aesthetic risk negatively impacts the intention-behavior relation. Thus, a shift within consumers' mindset is needed to create a favorable attitude towards sustainable clothing and a stylish perception of sustainable clothes.

Appendix

Table A.1: Covariance matrix.

	PAR1	PAR2	PAR3	ATT1	ATT2	ATT3	PER1	PER2	PER3	EC1	EC2	EC3	EC4	GC1	GC2	GC3	PEK1	PEK2	PEK3	PEK4	PB1	PB2	PB3	PB4	PI1	PI2	PI3	PI4	SN1	SN2	SN3
PAR1	1.218																														
PAR2	0.952	1.162																													
PAR3	0.88	1	1.156																												
ATT1	-0.381	-0.4	-0.417	0.734																											
ATT2	-0.263	-0.337	-0.347	0.401	1.299																										
ATT3	-0.382	-0.401	-0.431	0.511	0.579	0.973																									
PER1	0.189	0.136	0.11	-0.016	-0.061	0.033	0.937																								
PER2	0.324	0.294	0.292	-0.147	-0.138	-0.111	0.505	1.353																							
PER3	0.162	0.146	0.121	-0.015	-0.036	0.027	0.694	0.555	1.054																						
EC1	-0.126	-0.094	-0.124	0.242	0.201	0.254	0.023	-0.074	0.051	0.489																					
EC2	-0.145	-0.123	-0.134	0.245	0.245	0.286	0.024	-0.077	0.047	0.308	0.533																				
EC3	-0.184	-0.165	-0.177	0.278	0.276	0.327	-0.016	-0.063	0.045	0.302	0.336	0.761																			
EC4	-0.147	-0.12	-0.139	0.217	0.199	0.233	0.019	-0.044	0.066	0.29	0.29	0.323	0.525																		
GC1	-0.003	-0.019	-0.061	0.085	0.091	0.08	0.035	0.203	0.082	0.066	0.058	0.135	0.104	1.071																	
GC2	-0.057	-0.023	-0.07	0.092	0.134	0.096	0.005	0.129	0.069	0.085	0.078	0.15	0.112	0.807	1.096																
GC3	-0.017	-0.019	-0.057	0.069	0.182	0.065	-0.014	0.058	-0.001	0.02	0.032	0.083	0.063	0.638	0.749	1.056															
PEK1	-0.037	-0.052	-0.053	0.161	0.145	0.139	-0.008	-0.059	-0.005	0.039	0.121	0.073	0.066	0.015	0.034	-0.011	0.601														
PEK2	-0.093	-0.073	-0.079	0.155	0.146	0.16	-0.002	-0.051	0.02	0.093	0.134	0.109	0.105	0.027	0.054	0.027	0.275	0.42													
PEK3	-0.064	-0.052	-0.044	0.15	0.145	0.155	0.017	-0.08	-0.007	0.084	0.142	0.119	0.085	0.008	0.049	0.038	0.252	0.268	0.539												
PEK4	-0.084	-0.09	-0.089	0.172	0.152	0.136	0.02	-0.067	0.013	0.093	0.149	0.151	0.082	0.083	0.081	0.06	0.268	0.285	0.319	0.632											
PB1	-0.335	-0.38	-0.392	0.299	0.35	0.389	-0.118	-0.159	-0.118	0.109	0.145	0.199	0.104	0.054	0.044	0.062	0.172	0.127	0.11	0.164	1.012										
PB2	-0.263	-0.384	-0.413	0.37	0.336	0.435	-0.127	-0.18	-0.119	0.157	0.206	0.212	0.124	0.109	0.15	0.091	0.181	0.156	0.165	0.22	0.581	1.199									
PB3	-0.421	-0.449	-0.44	0.386	0.365	0.476	-0.114	-0.294	-0.142	0.203	0.22	0.277	0.167	0.044	0.073	0.06	0.161	0.137	0.138	0.153	0.643	0.692	1.137								
PB4	-0.373	-0.4	-0.442	0.413	0.48	0.52	-0.038	-0.142	0.006	0.173	0.229	0.332	0.148	0.104	0.089	0.085	0.18	0.155	0.169	0.234	0.715	0.702	0.707	1.19							
PI1	-0.411	-0.403	-0.442	0.426	0.477	0.545	-0.042	-0.107	-0.017	0.224	0.265	0.302	0.216	0.128	0.149	0.091	0.171	0.187	0.175	0.181	0.388	0.462	0.486	0.516	0.789						
PI2	-0.366	-0.375	-0.409	0.459	0.45	0.564	-0.032	-0.084	0.011	0.214	0.244	0.323	0.241	0.152	0.141	0.116	0.181	0.187	0.177	0.164	0.428	0.488	0.505	0.536	0.689	0.87					
PI3	-0.396	-0.424	-0.46	0.455	0.484	0.531	-0.022	-0.153	0.015	0.196	0.243	0.281	0.179	0.099	0.074	0.069	0.185	0.169	0.159	0.166	0.426	0.499	0.498	0.546	0.608	0.642	0.884				
PI4	-0.272	-0.287	-0.325	0.31	0.302	0.409	-0.066	-0.123	-0.015	0.196	0.196	0.245	0.185	0.075	0.129	0.062	0.088	0.1	0.122	0.12	0.222	0.347	0.355	0.339	0.427	0.442	0.42	0.649			
SN1	-0.043	-0.086	-0.08	0.181	0.251	0.193	0.007	-0.021	0.041	0.106	0.186	0.251	0.089	0.104	0.041	0.059	0.103	0.065	0.037	0.113	0.21	0.331	0.262	0.299	0.155	0.193	0.191	0.101	1.115		
SN2	-0.026	-0.027	-0.072	0.14	0.089	0.129	-0.074	-0.09	-0.026	0.121	0.173	0.235	0.097	0.103	0.113	0.029	0.103	0.086	0.059	0.143	0.186	0.295	0.201	0.255	0.124	0.131	0.113	0.124	0.713	1.304	
SN3	-0.038	-0.097	-0.113	0.195	0.169	0.223	0.001	-0.089	0.034	0.13	0.182	0.218	0.101	0.093	0.054	0.026	0.079	0.07	0.044	0.124	0.249	0.373	0.296	0.346	0.152	0.189	0.161	0.147	0.776	0.905	1.224

Note: ATT = attitude towards sustainable clothing, EC = environmental concern, GC = greenwashing concern, PAR = perceived aesthetic risk, PEK = perceived environmental knowledge, PER = perceived economic risk, PI = purchase intention for sustainable clothes, PB = purchase behavior, SN = subjective norm.

Table A.2: Cross-loadings.

	PAR	ATT	PER	EC	GC	PEK	PB	PI	SN
PAR1	**0.898**	-0.401	0.246	-0.222	-0.026	-0.112	-0.357	-0.423	-0.035
PAR2	**0.956**	-0.447	0.217	-0.189	-0.021	-0.107	-0.422	-0.445	-0.072
PAR3	**0.940**	-0.472	0.201	-0.217	-0.064	-0.107	-0.442	-0.490	-0.087
ATT1	-0.462	**0.840**	-0.095	0.469	0.105	0.318	0.483	0.620	0.213
ATT2	-0.277	**0.737**	-0.087	0.330	0.125	0.221	0.379	0.485	0.164
ATT3	-0.407	**0.879**	-0.034	0.456	0.090	0.260	0.521	0.669	0.197
PER1	0.145	-0.011	**0.805**	0.023	0.012	0.011	-0.115	-0.053	-0.017
PER2	0.257	-0.141	**0.857**	-0.092	0.127	-0.094	-0.189	-0.128	-0.056
PER3	0.136	-0.006	**0.803**	0.083	0.058	0.011	-0.102	-0.002	0.022
EC1	-0.161	0.423	-0.018	**0.822**	0.094	0.194	0.261	0.382	0.176
EC2	-0.181	0.444	-0.020	**0.834**	0.086	0.321	0.309	0.418	0.258
EC3	-0.198	0.423	-0.026	**0.788**	0.157	0.221	0.332	0.425	0.279
EC4	-0.184	0.377	0.007	**0.798**	0.143	0.205	0.212	0.365	0.137
GC1	-0.029	0.102	0.130	0.140	**0.894**	0.053	0.084	0.142	0.100
GC2	-0.047	0.123	0.081	0.163	**0.930**	0.089	0.095	0.152	0.064
GC3	-0.031	0.114	0.023	0.075	**0.833**	0.048	0.081	0.106	0.040
PEK1	-0.061	0.241	-0.042	0.159	0.021	**0.751**	0.251	0.261	0.126
PEK2	-0.124	0.298	-0.030	0.280	0.063	**0.853**	0.249	0.321	0.115
PEK3	-0.071	0.256	-0.047	0.241	0.047	**0.799**	0.223	0.278	0.064
PEK4	-0.110	0.242	-0.027	0.245	0.104	**0.784**	0.272	0.256	0.163
PB1	-0.365	0.426	-0.152	0.225	0.056	0.240	**0.822**	0.470	0.225
PB2	-0.325	0.440	-0.152	0.262	0.118	0.278	**0.810**	0.528	0.319
PB3	-0.406	0.485	-0.210	0.332	0.060	0.234	**0.847**	0.557	0.251
PB4	-0.369	0.535	-0.074	0.328	0.092	0.284	**0.857**	0.573	0.290
PI1	-0.467	0.676	-0.081	0.463	0.153	0.346	0.588	**0.917**	0.171
PI2	-0.408	0.663	-0.052	0.445	0.160	0.328	0.591	**0.915**	0.195
PI3	-0.451	0.650	-0.082	0.391	0.094	0.309	0.590	**0.874**	0.177
PI4	-0.364	0.533	-0.109	0.416	0.123	0.229	0.443	**0.764**	0.159
SN1	-0.067	0.238	0.004	0.243	0.070	0.125	0.293	0.195	**0.876**
SN2	-0.037	0.136	-0.068	0.223	0.082	0.143	0.230	0.138	**0.846**
SN3	-0.076	0.225	-0.030	0.233	0.059	0.119	0.321	0.188	**0.909**

Note: Values corresponding to a construct's assigned indicators are highlighted in bold. ATT = attitude towards sustainable clothing, EC = environmental concern, GC = greenwashing concern, PAR = perceived aesthetic risk, PEK = perceived environmental knowledge, PER = perceived economic risk, PI = purchase intention for sustainable clothes, PB = purchase behavior, SN = subjective norm.

Table A.3: Fornell-Larcker evaluation.

	PAR	PER	ATT	EC	PEK	GC	PI	PB	SN
PAR	0.931								
PER	0.235	0.822							
ATT	-0.474	-0.085	0.821						
EC	-0.224	-0.019	0.516	0.811					
PEK	-0.116	-0.046	0.327	0.293	0.789				
GC	-0.041	0.093	0.127	0.148	0.073	0.887			
PI	-0.487	-0.091	0.728	0.492	0.352	0.153	0.870		
PB	-0.440	-0.176	0.567	0.347	0.311	0.098	0.640	0.834	
SN	-0.071	-0.031	0.235	0.266	0.145	0.079	0.202	0.326	0.877

Note: ATT = attitude towards sustainable clothing, EC = environmental concern, GC = greenwashing concern, PAR = perceived aesthetic risk, PEK = perceived environmental knowledge, PER = perceived economic risk, PI = purchase intention for sustainable clothes, PB = purchase behavior, SN = subjective norm.

References

Abraham, C., Sheeran, P., Norman, P., Conner, M., Vries, N. de, & Otten, W. (1999). When Good Intentions Are Not Enough: Modeling Postdecisional Cognitive Correlates of Condom Use. Jour-nal of Applied Social Psychology, 29(12), 2591–2612. https://doi.org/10.1111/j.1559-1816.1999.tb00127.x

Achabou, M. A., & Dekhili, S. (2013). Luxury and sustainable development: Is there a match? Jour-nal of Business Research, 66(10), 1896–1903. https://doi.org/10.1016/j.jbusres.2013.02.011

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. In J. Kuhl & J. Beck-mann (Eds.), Springer series in social psychology. Action control: From cognition to behavior (1st ed., pp. 11–39). Springer. https://doi.org/10.1007/978-3-642-69746-3_2

Ajzen, I. (1988). Attitudes, Personality and Behavior. Open University Press.

Ajzen, I. (1991). The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-t

Ajzen, I. (1996). The social psychology of decision making. In E. T. Higgins & A. W. Kruglanski (Eds.), Social psychology: Handbook of basic principles (pp. 297–325). Guilford Press.

Ajzen, I., & Fishbein, M. (1980). Understanding attitudes and predicting social behavior. Prentice-Hall. https://www.scienceopen.com/document?vid=c20c4174-d8dc-428d-b352-280b05eacdf7

Ali, A., Khan, A. A., Ahmed, I., & Shahzad, W. (2011). Determinants of Pakistani consumers' green purchase behavior: Some insights from a developing country. International Journal of Busi-ness and Social Science, 2(3), 217–226.

Allwood, J. M., Laursen, S. E., Russell, S. N., Rodríguez, C. M. de, & Bocken, N. (2008). An ap-proach to scenario analysis of the sustainability of an industrial sector applied to clothing and textiles in the UK. Journal of Cleaner Production, 16(12), 1234–1246. https://doi.org/10.1016/j.jclepro.2007.06.014

Armitage, C. J., & Conner, M. (2001). Efficacy of the Theory of Planned Behaviour: A meta-analytic review. The British Journal of Social Psychology, 40(Pt 4), 471–499. https://doi.org/10.1348/014466601164939

Armstrong, C. M., Niinimäki, K., Lang, C., & Kujala, S. (2016). A Use-Oriented Clothing Econo-my? Preliminary Affirmation for Sustainable Clothing Consumption Alternatives. Sustainable Development, 24(1), 18–31. https://doi.org/10.1002/sd.1602

Bagozzi, R. P. (1992). The Self-Regulation of Attitudes, Intentions, and Behavior. Social Psychology Quarterly, 55(2), 178. https://doi.org/10.2307/2786945

Beckmann, J., & Kuhl, J. (1984). Altering information to gain action control: Functional aspects of human information processing in decision making. Journal of Research in Personality, 18(2), 224–237. https://doi.org/10.1016/0092-6566(84)90031-x

Bianchi, C., & Birtwistle, G. (2012). Consumer clothing disposal behaviour: A comparative study. International Journal of Consumer Studies, 36(3), 335–341. https://doi.org/10.1111/j.1470-6431.2011.01011.x

Bird, B. (1988). Implementing Entrepreneurial Ideas: The Case for Intention. Academy of Manage-ment Review, 13(3), 442–453. https://doi.org/10.5465/amr.1988.4306970

Bowen, F., & Aragon-Correa, J. A. (2014). Greenwashing in Corporate Environmentalism Research and Practice. Organization & Environment, 27(2), 107–112. https://doi.org/10.1177/1086026614537078

Bray, J., Johns, N., & Kilburn, D. (2011). An Exploratory Study into the Factors Impeding Ethical Consumption. Journal of Business Ethics, 98(4), 597–608. https://doi.org/10.1007/s10551-010-0640-9

Chan, R. Y. K. (2001). Determinants of Chinese consumers' green purchase behavior. Psychology & Marketing, 18(4), 389–413. https://doi.org/10.1002/mar.1013

Chen, Y.-S., & Chang, C.-H. (2013). Greenwash and Green Trust: The Mediation Effects of Green Consumer Confusion and Green Perceived Risk. Journal of Business Ethics, 114(3), 489–500. https://doi.org/10.1007/s10551-012-1360-0

Cheung, S. F., Chan, D. K.-S., & Wong, Z. S.-Y. (1999). Reexamining the Theory of Planned Be-havior in Understanding Wastepaper Recycling. Environment and Behavior, 31(5), 587–612. https://doi.org/10.1177/00139169921972254

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associ-ates.

Connolly, J., & Prothero, A. (2003). Sustainable consumption: Consumption, consumers and the commodity discourse. Consumption Markets & Culture, 6(4), 275–291. https://doi.org/10.1080/1025386032000168311

Crosby, L. A., Gill, J. D., & Taylor, J. R. (1981). Consumer/voter Behavior in the Passage of the Michigan Container law. Journal of Marketing, 45(2), 19–32. https://doi.org/10.1177/002224298104500203

Csikszentmihalyi, M. (2000). The Costs and Benefits of Consuming. Journal of Consumer Research, 27(2), 267–272. https://doi.org/10.1086/314324

Dagher, G. K., & Itani, O. (2014). Factors influencing green purchasing behaviour: Empirical evi-dence from the Lebanese consumers. Journal of Consumer Behaviour, 13(3), 188–195. https://doi.org/10.1002/cb.1482 do Paço, A. M. F., & Reis, R. (2012). Factors Affecting Skepticism toward Green Advertising. Jour-nal of Advertising, 41(4), 147–155. https://doi.org/10.1080/00913367.2012.10672463

D'Souza, C., Taghian, M., Lamb, P., & Peretiatko, R. (2007). Green decisions: Demographics and consumer understanding of environmental labels. International Journal of Consumer Studies, 31(4), 371–376. https://doi.org/10.1111/j.1470-6431.2006.00567.x

Dunlap, R. E., van Liere, K. D., Mertig, A. G., & Jones, R. E. (2000). New Trends in Measuring Environmental Attitudes: Measuring Endorsement of the New Ecological Paradigm: A Revised NEP Scale. Journal of Social Issues, 56(3), 425–442. https://doi.org/10.1111/0022-4537.00176

Echegaray, F., & Hansstein, F. V. (2017). Assessing the intention-behavior gap in electronic waste recycling: the case of Brazil. Journal of Cleaner Production, 142, 180–190. https://doi.org/10.1016/j.jclepro.2016.05.064

Ellen, P. S. (1994). Do we know what we need to know? Objective and subjective knowledge effects on pro-ecological behaviors. Journal of Business Research, 30(1), 43–52. https://doi.org/10.1016/0148-2963(94)90067-1

Ellen, P. S., Eroğlu, D., & Webb, D. (1997). Consumer judgments in a changinginformation environ-ment: How consumers respond to ''green marketing claims" [Working Paper]. Georgia State Uni-versity.

Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention, and behavior: An introduction to theory and research. Addison-Wesley.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable varia-bles and measurement error. Journal of Marketing Research, 18(1), 39–50.

Fryxell, G. E., & Lo, C. W. (2003). The influence of environmental knowledge and values on mana-gerial behaviours on behalf of the environment: An empirical examination of managers in China. Journal of Business Ethics, 46(1), 45–69. https://link.springer.com/article/10.1023/a:1024773012398

George, J. F. (2004). The theory of planned behavior and Internet purchasing. Internet Research, 14(3), 198–212. https://doi.org/10.1108/10662240410542634

GFA, & BCG. (2017). Pulse of the fashion industry. https://globalfashionagenda.com/wp-content/uploads/2017/05/Pulse-of-the-Fashion-Industry_2017.pdf

Godin, G., & Kok, G. (1996). The theory of planned behavior: A review of its applications to health-related behaviors. American Journal of Health Promotion : AJHP, 11(2), 87–98. https://doi.org/10.4278/0890-1171-11.2.87

Gollwitzer, P. M. (1990). Action phases and mind-sets. In E. T. Higgins & R. M. Sorrentino (Eds.), Handbook of motivation and cognition: Foundations of social behavior (pp. 53–92). Guilford.

Gollwitzer, P. M. (1993). Goal Achievement: The Role of Intentions. European Review of Social Psychology, 4(1), 141–185. https://doi.org/10.1080/14792779343000059

Goworek, H., Fisher, T., Cooper, T., Woodward, S., & Hiller, A. (2012). The sustainable clothing market: An evaluation of potential strategies for UK retailers. International Journal of Retail & Distribution Management, 40(12), 935–955. https://doi.org/10.1108/09590551211274937

Goworek, H., Oxborrow, L., Claxton, S., McLaren, A., Cooper, T., & Hill, H. (2018). Managing sustainability in the fashion business: Challenges in product development for clothing longevity in the UK. Journal of Business Research, In Press. https://doi.org/10.1016/j.jbusres.2018.07.021

Greenland, S. (2019). Valid P -Values Behave Exactly as They Should: Some Misleading Criticisms of P -Values and Their Resolution With S -Values. The American Statistician, 73(sup1), 106–114. https://doi.org/10.1080/00031305.2018.1529625

Ha, H.-Y., & Janda, S. (2012). Predicting consumer intentions to purchase energy-efficient products. Journal of Consumer Marketing, 29(7), 461–469. https://doi.org/10.1108/07363761211274974

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. European Business Review, 31(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). A primer on partial least squares structural equation modeling (PLS-SEM). SAGE Publications.

Harris, F., Roby, H., & Dibb, S. (2016). Sustainable clothing: Challenges, barriers and interventions for encouraging more sustainable consumer behaviour. International Journal of Consumer Studies, 40(3), 309–318. https://doi.org/10.1111/ijcs.12257

Hartmann, P., & Apaolaza-Ibáñez, V. (2012). Consumer attitude and purchase intention toward green energy brands: The roles of psychological benefits and environmental concern. Journal of Business Research, 65(9), 1254–1263. https://doi.org/10.1016/j.jbusres.2011.11.001

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validi-ty in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115–135.

Hiller Connell, K. Y. (2010). Internal and external barriers to eco-conscious apparel acquisition. International Journal of Consumer Studies, 34(3), 279–286. https://doi.org/10.1111/j.1470-6431.2010.00865.x

Hoch, S. J., & Deighton, J. (1989). Managing What Consumers Learn from Experience. Journal of Marketing, 53(2), 1–20. https://doi.org/10.1177/002224298905300201

Hsu, C.-L., Chang, C.-Y., & Yansritakul, C. (2017). Exploring purchase intention of green skincare products using the theory of planned behavior: Testing the moderating effects of country of origin and price sensitivity. Journal of Retailing and Consumer Services, 34, 145–152. https://doi.org/10.1016/j.jretconser.2016.10.006

Hughner, R. S., McDonagh, P., Prothero, A., Shultz, C. J., & Stanton, J. (2007). Who are organic food consumers? A compilation and review of why people purchase organic food. Journal of Con-sumer Behaviour, 6(2-3), 94–110. https://doi.org/10.1002/cb.210

Hustvedt, G., & Dickson, M. A. (2009). Consumer likelihood of purchasing organic cotton apparel. Journal of Fashion Marketing and Management: An International Journal, 13(1), 49–65. https://doi.org/10.1108/13612020910939879

Jacoby, J., Berning, C. K., & Dietvorst, T. F. (1977). What about Disposition? Journal of Marketing, 41(2), 22–28. https://doi.org/10.1177/002224297704100212

Jaiswal, D., & Kant, R. (2018). Green purchasing behaviour: A conceptual framework and empirical investigation of Indian consumers. Journal of Retailing and Consumer Services, 41, 60–69. https://doi.org/10.1016/j.jretconser.2017.11.008

Joergens, C. (2006). Ethical fashion: Myth or future trend? Journal of Fashion Marketing and Management: An International Journal, 10(3), 360–371. https://doi.org/10.1108/13612020610679321

Kashima, Y., Gallois, C., & McCamish, M. (1993). The theory of reasoned action and cooperative behaviour: It takes two to use a condom. The British Journal of Social Psychology, 32 (Pt 3)(3), 227–239. https://doi.org/10.1111/j.2044-8309.1993.tb00997.x

Kautish, P., Paul, J., & Sharma, R. (2019). The moderating influence of environmental conscious-ness and recycling intentions on green purchase behavior. Journal of Cleaner Production, 228, 1425–1436. https://doi.org/10.1016/j.jclepro.2019.04.389

Kautonen, T., van Gelderen, M., & Fink, M. (2015). Robustness of the Theory of Planned Behavior in Predicting Entrepreneurial Intentions and Actions. Entrepreneurship Theory and Practice, 39(3), 655–674. https://doi.org/10.1111/etap.12056

Kautonen, T., van Gelderen, M., & Tornikoski, E. T. (2013). Predicting entrepreneurial behaviour: A test of the theory of planned behaviour. Applied Economics, 45(6), 697–707. https://doi.org/10.1080/00036846.2011.610750

Kendzierski, D. (1990). Decision Making Versus Decision Implementation: An Action Control Approach to Exercise Adoption and Adherence1. Journal of Applied Social Psychology, 20(1), 27–45. https://doi.org/10.1111/j.1559-1816.1990.tb00376.x

Kendzierski, D., & Whitaker, D. J. (1997). The Role of Self-Schema in Linking Intentions with Behavior. Personality and Social Psychology Bulletin, 23(2), 139–147. https://doi.org/10.1177/0146167297232003

Kim, H. Y., & Chung, J.-E. (2011). Consumer purchase intention for organic personal care products. Journal of Consumer Marketing, 28(1), 40–47. https://doi.org/10.1108/07363761111101930

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. International Journal of E-Collaboration (Ijec), 11(4), 1–10.

Kollmuss, A., & Agyeman, J. (2002). Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? Environmental Education Research, 8(3), 239–260. https://doi.org/10.1080/13504620220145401

Kumar, B., Manrai, A. K., & Manrai, L. A. (2017). Purchasing behaviour for environmentally sustainable products: A conceptual framework and empirical study. Journal of Retailing and Con-sumer Services, 34, 1–9. https://doi.org/10.1016/j.jretconser.2016.09.004

Kwong, G. S., & Balaji, M. S. (2016). Linking green skepticism to green purchase behavior. Journal of Cleaner Production, 131(10), 629–638. https://doi.org/10.1016/j.jclepro.2016.04.122

Laitala, K., & Boks, C. (2012). Sustainable clothing design: Use matters. Journal of Design Re-search, 10(1-2), 121–239.

Lang, C., Armstrong, C. M., & Brannon, L. A. (2013). Drivers of clothing disposal in the US: An exploration of the role of personal attributes and behaviours in frequent disposal. International Journal of Consumer Studies, 37(6), 706–714. https://doi.org/10.1111/ijcs.12060

LaPiere, R. (1934). Attitudes Versus Actions. Social Forces, 13, 230–237.

Lee, K. (2008). Opportunities for green marketing: Young consumers. Marketing Intelligence & Planning, 26(6), 573–586. https://doi.org/10.1108/02634500810902839

Lee, Y.-K., Kim, S., Kim, M.-S., & Choi, J.-G. (2014). Antecedents and interrelationships of three types of pro-environmental behavior. Journal of Business Research, 67(10), 2097–2105. https://doi.org/10.1016/j.jbusres.2014.04.018

Li, L.-Y. (1997). Effect of Collectivist Orientation and Ecological Attitude on Actual Environmental Commitment. Journal of International Consumer Marketing, 9(4), 31–53. https://doi.org/10.1300/J046v09n04_03

Liska, A. E. (1984). A Critical Examination of the Causal Structure of the Fishbein/Ajzen Attitude-Behavior Model. Social Psychology Quarterly, 47(1), 61–74. https://doi.org/10.2307/3033889

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American Psychologist, 57(9), 705–717. https://doi.org/10.1037/0003-066x.57.9.705

Lundblad, L., & Davies, I. A. (2016). The values and motivations behind sustainable fashion consumption. Journal of Consumer Behaviour, 15(2), 149–162. https://doi.org/10.1002/cb.1559

Lyon, T. P., & Maxwell, J. W. (2011). Greenwash: Corporate Environmental Disclosure under Threat of Audit. Journal of Economics & Management Strategy, 20(1), 3–41. https://doi.org/10.1111/j.1530-9134.2010.00282.x

Lyon, T. P., & Montgomery, A. W. (2015). The Means and End of Greenwash. Organization & Environment, 28(2), 223–249. https://doi.org/10.1177/1086026615575332

Maichum, K., Parichatnon, S., & Peng, K.-C. (2016). Application of the Extended Theory of Planned Behavior Model to Investigate Purchase Intention of Green Products among Thai Con-sumers. Sustainability, 8(1077), 1–20. https://doi.org/10.3390/su8101077

Marquis, C., Toffel, M. W., & Zhou, Y. (2016). Scrutiny, Norms, and Selective Disclosure: A Global Study of Greenwashing. Organization Science, 27(2), 483–504. https://doi.org/10.1287/orsc.2015.1039

Milfont, T. L., & Duckitt, J. (2004). The structure of environmental attitudes: A first- and second-order confirmatory factor analysis. Journal of Environmental Psychology, 24(3), 289–303. https://doi.org/10.1016/j.jenvp.2004.09.001

Mohr, L. A., Eroğlu, D., & Ellen, P. S. (1998). The Development and Testing of a Measure of Skepticism Toward Environmental Claims in Marketers' Communications. Journal of Consumer Af-fairs, 32(1), 30–55. https://doi.org/10.1111/j.1745-6606.1998.tb00399.x

Morgan, L. R., & Birtwistle, G. (2009). An investigation of young fashion consumers' disposal hab-its. International Journal of Consumer Studies, 33(2), 190–198. https://doi.org/10.1111/j.1470-6431.2009.00756.x

Mostafa, M. M. (2006). Antecedents of Egyptian Consumers' Green Purchase Intentions. Journal of International Consumer Marketing, 19(2), 97–126. https://doi.org/10.1300/J046v19n02_06

Mostafa, M. M. (2007). A hierarchical analysis of the green consciousness of the Egyptian consum-er. Psychology & Marketing, 24(5), 445–473. https://doi.org/10.1002/mar.20168

Newton, J. D., Tsarenko, Y., Ferraro, C., & Sands, S. (2015). Environmental concern and environmental purchase intentions: The mediating role of learning strategy. Journal of Business Re-search, 68(9), 1974–1981. https://doi.org/10.1016/j.jbusres.2015.01.007

Niinimäki, K., & Hassi, L. (2011). Emerging design strategies in sustainable production and consumption of textiles and clothing. Journal of Cleaner Production, 19(16), 1876–1883. https://doi.org/10.1016/j.jclepro.2011.04.020

Orbell, S., & Sheeran, P. (1998). 'inclined abstainers': A problem for predicting health-related behaviour. British Journal of Social Psychology, 37(2), 151–165. https://doi.org/10.1111/j.2044-8309.1998.tb01162.x

Park, C. W., Mothersbaugh, D. L., & Feick, L. (1994). Consumer Knowledge Assessment. Journal of Consumer Research, 21(1), 71–82. https://doi.org/10.1086/209383

Park, H. J., & Lin, L. M. (2018). Exploring attitude-behavior gap in sustainable consumption: Comparison of recycled and upcycled fashion products. Journal of Business Research, In Press. https://doi.org/10.1016/j.jbusres.2018.08.025

Paul, J., Modi, A., & Patel, J. (2016). Predicting green product consumption using theory of planned behavior and reasoned action. Journal of Retailing and Consumer Services, 29, 123–134. https://doi.org/10.1016/j.jretconser.2015.11.006

Pavlou, & Fygenson (2006). Understanding and Predicting Electronic Commerce Adoption: An Extension of the Theory of Planned Behavior. MIS Quarterly, 30(1), 115. https://doi.org/10.2307/25148720

Petty, R. E., Unnava, R. H., & Stratham, A. J. (1991). Theories of Attitude Change. In T. S. Robert-son & H. H. Kassarjian (Eds.), Handbook of consumer behavior (pp. 241–280). Prentice-Hall.

Pickett-Baker, J., & Ozaki, R. (2008). Pro-environmental products: Marketing influence on consum-er purchase decision. Journal of Consumer Marketing, 25(5), 281–293. https://doi.org/10.1108/07363760810890516

Pinto, D. C., Nique, W. M., Añaña, E. d. S., & Herter, M. M. (2011). Green consumer values: how do personal values influence environmentally responsible water consumption? International Jour-nal of Consumer Studies, 35(2), 122–131. https://doi.org/10.1111/j.1470-6431.2010.00962.x

Prakash, G., & Pathak, P. (2017). Intention to buy eco-friendly packaged products among young consumers of India: A study on developing nation. Journal of Cleaner Production, 141, 385–393. https://doi.org/10.1016/j.jclepro.2016.09.116

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. Long Range Planning, 45(5-6), 341–358.

Schlegelmilch, B. B., Bohlen, G. M., & Diamantopoulos, A. (1996). The link between green pur-chasing decisions and measures of environmental consciousness. European Journal of Marketing, 30(5), 35–55. https://doi.org/10.1108/03090569610118740

Schultz, P. W. (2000). New Environmental Theories: Empathizing With Nature: The Effects of Perspective Taking on Concern for Environmental Issues. Journal of Social Issues, 56(3), 391–406. https://doi.org/10.1111/0022-4537.00174

Seele, P., & Gatti, L. (2017). Greenwashing Revisited: In Search of a Typology and Accusation-Based Definition Incorporating Legitimacy Strategies. Business Strategy and the Environment, 26(2), 239–252. https://doi.org/10.1002/bse.1912

Shaw, D., Shiu, E., & Clarke, I. (2000). The Contribution of Ethical Obligation and Self-identity to the Theory of Planned Behaviour: An Exploration of Ethical Consumers. Journal of Marketing Management, 16(8), 879–894. https://doi.org/10.1362/026725700784683672

Sheeran, P. (2002). Intention-Behavior Relations: A Conceptual and Empirical Review. European Review of Social Psychology, 12(1), 1–36. https://doi.org/10.1080/14792772143000003

Sheeran, P., & Abraham, C. (2003). Mediator of moderators: Temporal stability of intention and the intention-behavior relation. Personality and Social Psychology Bulletin, 29(2), 205–215. https://doi.org/10.1177/0146167202239046

Sheeran, P., & Orbell, S. (1998). Do intentions predict condom use? Meta-analysis and examination of six moderator variables. The British Journal of Social Psychology, 37 (Pt 2)(2), 231–250. https://doi.org/10.1111/j.2044-8309.1998.tb01167.x

Sheeran, P., & Orbell, S. (1999). Augmenting the Theory of Planned Behavior: Roles for Anticipated Regret and Descriptive Norms. Journal of Applied Social Psychology, 29(10), 2107–2142. https://doi.org/10.1111/j.1559-1816.1999.tb02298.x

Shirokova, G., Osiyevskyy, O., & Bogatyreva, K. (2016). Exploring the intention–behavior link in student entrepreneurship: Moderating effects of individual and environmental characteristics. European Management Journal, 34(4), 386–399. https://doi.org/10.1016/j.emj.2015.12.007

Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. Journal of Business Research, 69(10), 4552–4564. https://doi.org/10.1016/j.jbusres.2016.03.049

Siano, A., Vollero, A., Conte, F., & Amabile, S. (2017). “More than words”: Expanding the taxono-my of greenwashing after the Volkswagen scandal. Journal of Business Research, 71, 27–37. https://doi.org/10.1016/j.jbusres.2016.11.002

Sparks, P., & Shepherd, R. (1992). Self-Identity and the Theory of Planned Behavior: Assesing the Role of Identification with “Green Consumerism”. Social Psychology Quarterly, 55(4), 388. https://doi.org/10.2307/2786955

Statista. (2020). eCommerce - worldwide. https://www.statista.com/outlook/243/100/ecommerce/worldwide

Tan, C.-S., Ooi, H.-Y., & Goh, Y.-N. (2017). A moral extension of the theory of planned behavior to predict consumers’ purchase intention for energy-efficient household appliances in Malaysia. Energy Policy, 107, 459–471. https://doi.org/10.1016/j.enpol.2017.05.027

Taufique, K. M., & Vaithianathan, S. (2018). A fresh look at understanding Green consumer behav-ior among young urban Indian consumers through the lens of Theory of Planned Behavior. Jour-nal of Cleaner Production, 183, 46–55. https://doi.org/10.1016/j.jclepro.2018.02.097

Vermeir, I., & Verbeke, W. (2008). Sustainable food consumption among young adults in Belgium: Theory of planned behaviour and the role of confidence and values. Ecological Economics, 64(3), 542–553. https://doi.org/10.1016/j.ecolecon.2007.03.007

Wang, P., Liu, Q., & Qi, Y. (2014). Factors influencing sustainable consumption behaviors: a survey of the rural residents in China. Journal of Cleaner Production, 63, 152–165. https://doi.org/10.1016/j.jclepro.2013.05.007

Wang, Z., Guo, D., & Wang, X. (2016). Determinants of residents' e-waste recycling behaviour in-tentions: Evidence from China. Journal of Cleaner Production, 137, 850–860. https://doi.org/10.1016/j.jclepro.2016.07.155

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychological Bulletin, 132(2), 249–268. https://doi.org/10.1037/0033-2909.132.2.249

Wicker, A. W. (1969). Attitudes versus Actions: The Relationship of Verbal and Overt Behavioral Responses to Attitude Objects. Journal of Social Issues, 25(4), 41–78. https://doi.org/10.1111/j.1540-4560.1969.tb00619.x

Yadav, R., & Pathak, G. S. (2016). Young consumers' intention towards buying green products in a developing nation: Extending the theory of planned behavior. Journal of Cleaner Production, 135, 732–739. https://doi.org/10.1016/j.jclepro.2016.06.120

Yadav, R., & Pathak, G. S. (2017). Determinants of Consumers' Green Purchase Behavior in a Developing Nation: Applying and Extending the Theory of Planned Behavior. Ecological Econom-ics, 134, 114–122. https://doi.org/10.1016/j.ecolecon.2016.12.019

Young, W., Hwang, K., McDonald, S., & Oates, C. J. (2009). Sustainable consumption: Green consumer behaviour when purchasing products. Sustainable Development, 18(1), 20-31. https://doi.org/10.1002/sd.394

Zhang, L., Li, D., Cao, C., & Huang, S. (2018). The influence of greenwashing perception on green purchasing intentions: The mediating role of green word-of-mouth and moderating role of green concern. Journal of Cleaner Production, 187, 740–750. https://doi.org/10.1016/j.jclepro.2018.03.201

Zhao, H.-H., Gao, Q., Wu, Y.-P., Wang, Y., & Zhu, X.-D. (2014). What affects green consumer behavior in China? A case study from Qingdao. Journal of Cleaner Production, 63, 143–151. https://doi.org/10.1016/j.jclepro.2013.05.021

2.2.2 Research Paper No. 2: The Drivers of Sustainable Apparel and Sportswear Consumption: A Segmented Kano Perspective

Authors: Baier, D., Rausch, T. M., & Wagner, T. F. (2020)

Published in: Sustainability, 12 (7), 1-21

DOI: https://doi.org/10.3390/su12072788

Abstract: The steady increase of sustainable consumer behavior leads companies to strengthen their efforts to become socially and ecologically more sustainable. Particularly in the clothing and footwear industry, more and more companies are aware of their need to fundamentally adapt the way they create value. Sustainability offerings are developed, e.g., usage of upcycled materials (e.g., ocean plastic), circular business models (e.g., decomposition of returned products into components for new ones), as well as adapted product ranges (e.g. smaller or with fewer fashion cycles). However, it is frequently unclear in advance, which offerings will increase (or decrease) satisfaction and consequently drive (or not drive) sustainable consumption. The application of a Segmented Kano perspective in an apparel and sportswear context is presented that helps to answer these questions: 17 potential offerings were assessed by a sample of 490 consumers. Our analysis demonstrates the usefulness of this methodology and that returning used products (to recycle them), discounts for buying sustainable products, sustainability level indicators, and biobased materials are highly attractive. However, the responsiveness varies across the derived consumer segments, from being decisive or attractive to indifferent or reverse. As assumed, gender and attitude towards sustainability are good predictors for segment membership.

Keywords: sustainable clothing consumption; Kano model; Segmented Kano perspective; sustainability offerings; circular economy; customer satisfaction

1 Introduction

Conventional business models are frequently linked to sales and profit maximization by satisfying consumers' needs and considering resources as ever inexhaustible (Csikszentmihalyi, 2000; Niinimäki & Hassi, 2011). Thereby, technological advances allowed consumption levels to increase steadily. This phenomenon is particularly apparent in the clothing and footwear industry: the manufacturing of these products – made using the same materials and equipment – shifted to lower-cost countries. The quality and price declined, and thus, the lifespans of products shortened (Goworek et al., 2012; Niinimäki & Hassi, 2011). Such so-called fast fashion, i.e., clothing and footwear in contemporary styles produced within reduced lead times to get products faster from concepts to consumers (Barnes & Lea-Greenwood, 2010), implies consequent obsolescence, impulse purchasing, and subsequently, an excessive usage of valuable natural resources (Achabou & Dekhili, 2013). Extremely fast fashion cycles cause an environmental burden by negatively impacting e.g. (ground-)water, soil, and air negatively (Niinimäki & Hassi, 2011): the clothing and footwear industry has a significant environmental footprint, polluting approximately 200 tons of water per ton of fabric (Nagurney & Yu, 2012), causing tons of CO_2 emissions (Goworek et al., 2018; Niinimäki & Hassi, 2011), and producing a growing amount of clothing and footwear waste (Niinimäki & Hassi, 2011).

The environmental consequences of the current 'throwaway' society and further, the increasing consciousness of its negative environmental impacts and its subsequent ethical issues led literature to investigate sustainable consumer behavior. For example, research focused on how to shift consumer behaviors to enhance sustainable consumption (White et al., 2019), how to encourage consumers to recycle (Sun & Trudel, 2017; Trudel et al., 2016; Wang et al., 2017; Winterich et al., 2019), motives and antecedents of consumers' sustainable purchase behavior (Eberhart & Naderer, 2017), or proposed a holistic customer-centric approach of mindful consumption (Sheth et al., 2011). Interestingly, a significant part of that research body has particularly focused on the consumption of garments since the clothing and footwear industry has a large-scale impact on the environment. In this context, consumers can reduce their negative environmental impact in every consumption phase from acquisition, use and care, to disposal (Laitala & Boks, 2012). Consumption levels can be decreased by prolonging lifespans by repairing or repurposing clothing, by using collaborative consumption concepts, or by establishing design strategies to extend fashion life cycles (Bianchi & Birtwistle, 2012; Goworek et al., 2018; Laitala, 2014; Maldini et al., 2019; Zamani et al., 2017).

Thus, antecedents and consequences of eco-conscious[4] consumer behavior in general are well understood and several concepts for sustainable clothing and footwear consumption have been proposed by existing literature. However, most consumers still exhibit an attitude-behavior gap regarding eco-conscious consumption, i.e., albeit they pretend pro-environmental attitudes and consciousness (Trudel

[4] The terms eco-conscious, eco-friendly, environmentally friendly, pro-environmental, and sustainable will be used interchangeably in this paper.

& Cotte, 2009), they frequently struggle to translate this into green actions and hence, do not behave sustainably (Auger & Devinney, 2007; Hughner et al., 2007; Young et al., 2010). Apparently, besides some drivers there are significant barriers inhibiting consumers to combine their clothing and footwear consumption habits with pro-environmental behavior, remedial offerings are needed. Thus, research still lacks a comprehensive understanding of how consumers assess sustainability in a clothing and footwear context and – more specifically – how different sustainability offerings are accepted by consumers.

In order to fill this research gap, we investigated these drivers, barriers, and remedies. Then, we concretized them in an apparel and sportswear context with respect to aspects like product range, labeling (i.e., 'traffic light' models as well as quality seals), processes (return and discount policies), and materials, and applied the Kano model (Kano et al., 1984) to a sample of typical apparel and sportswear consumers (n=490). They were asked to evaluate 17 concretized offerings as attractive, indifferent, one-dimensional, must-be, or reverse. Moreover, a Segmented Kano perspective (Baier et al., 2018; Baier & Rese, 2020; Rese, A., Schlee, T., & Baier, D., 2019) was developed to show that segment-specific differences in these evaluations exist and that segment membership can be related to background variables. Based on these analyses, recommendations regarding the prioritization of these offerings could be given. The resulting insights might help to overcome the attitude-behavior gap by assessing the segment-specific impact of surveyed offerings on sustainable apparel and sportswear consumption.

The remaining paper is structured as follows: In section 2, we review preceding literature on drivers, barriers, and remedies for sustainable clothing and footwear consumption. Then, in section 3, the methodology to categorize sustainability offerings, the use case selection in the apparel and sportswear industry and the conceptualization of the Kano questionnaire as well as the descriptive statistics of the customer sample are presented. Section 4 displays the derived results including the segment-specific findings regarding 17 sustainability offerings. Then, in section 5, the theoretical contribution, as well as limitations, and directions for future research are provided. The paper closes with conclusions in section 6.

2 Theoretical Background: Sustainable Clothing and Footwear Consumption

2.1 Drivers of Sustainable Clothing and Footwear Consumption

Sustainability in an organizational context is mostly referred to the triple bottom line concept. It accounts for social, environmental, and economic aspects (Elkington, 1997): The social facet pertains to fair business practices implying the well-being of corporate, labor, community, and region in which the organization operates. The environmental bottom line refers to environmental practices which benefit (or do not harm respectively) the planet and minimize the organization's environmental impact including, e.g., life cycle assessment of products. Further, the economic bottom line can be defined as the economic value created by the organization after the costs of all inputs are deducted. Today, a growing number of companies, also in the clothing and footwear industry, adopts and applies this concept as an accounting tool which allows to evaluate the company's performance in a broader perspective, not only with respect to the traditional return on investment by manufacturing and selling goods but further with respect to the impact on its customers' sustainable consumption.

However, due to a lacking industry standard, especially sustainable clothing and footwear consumption is not uniformly termed and defined (Joergens, 2006; Lundblad & Davies, 2016). Within literature, terms like eco(-conscious; -friendly) (Gam, 2011; Hiller Connell, 2010; Laitala & Boks, 2012; Niinimäki, 2010), ethical (Goworek et al., 2012; Joergens, 2006; McNeill & Moore, 2015; Shen et al., 2012), green (D'Souza et al., 2007; Nam et al., 2017), organic (Hustvedt & Dickson, 2009), and slow (Pookulangara & Shephard, 2013) are used frequently and interchangeably. Moreover, although mostly associated with eco-conscious logistics and manufacturing aspects, sustainable consumption extends well beyond the pre-purchase and purchase phase by additionally comprising e.g. cleansing or recycling of produced clothing and footwear (Goworek et al., 2012; Laitala et al., 2011). At its core, it is assumed that sustainable clothing and footwear consumption implies pro-environmental actions at every phase from pre-purchase, purchase, to post-purchase (Jacoby et al., 1977; Lundblad & Davies, 2016; Morgan & Birtwistle, 2009) comprising – inter alia – acquisition, storage, usage and care (e.g., laundering and cleaning respectively), maintenance (e.g., repairing), and discard (e.g., recycling, re-usage, or disposal) (Bianchi & Birtwistle, 2012; Hiller Connell, 2010).

2.1.1 Drivers in the Pre-Purchase and Purchase Phase

Regarding the environmental impact of these individual phases, opinions within literature are diverging. The steadily growing volume of clothing and footwear consumption, low employee wages, poor working conditions, and excessive pesticide use are some of the key issues during the early phases of the clothing life cycle (Bianchi & Birtwistle, 2010; Birtwistle & Moore, 2007; Joergens, 2006). Thus, sustainable behavior during this pre-purchase and purchase phases implies either purchasing products made of environmentally preferable, recycled, upcycled, or biobased fibers produced under fair conditions, purchasing from second-hand stores or collaborative platforms, or reducing the overall consumption level (Allwood et al., 2008; Armstrong et al., 2016; Goworek et al., 2012).

Further, design strategies have been proposed for manufacturers to prolong the clothing and footwear lifespan: Niinimäki and Hassi (Niinimäki & Hassi, 2011) developed different design and manufacturing strategies to decrease the environmental impact of the clothing and footwear industry by focusing on the customers' values and needs. Hirscher et al. (Hirscher et al., 2018) considered do-it-yourself, do-it-together, and participatory design strategies for value co-creation during the manufacturing process. Niinimäki (Niinimäki, 2010) stressed that manufacturers are hardly aware of their consumers' needs and wants regarding the aesthetics of eco-fashion and thus, such clothing and footwear often only appeals to a limited range of potential customers.

Kim and Kang (J. Kim et al., 2018) found social capital, i.e., an intangible force that unites society by transforming individuals into members of a community with shared assumptions and a sense of the common good, to have a strong impact on the purchase intention of sustainable fashion.

2.1.2 Drivers in the Post-Purchase Phase

Nevertheless, research focused mainly on environmental issues occurring in the post-purchase phase of the clothing and footwear life cycle. More specifically, the usage and care phase are assumed to cause a significant overall negative impact on the environment: particularly – in the washable clothing case – optimizations regarding the laundering process are suggested to decrease energy, water, and wax consumption (Allwood et al., 2008; Laitala et al., 2011; Laitala et al., 2012). E.g., Goworek et al. (Goworek et al., 2012) found that consumers were not willing to wash their clothing at a lower temperature in case this implies compromised cleanliness.

Another major strand of literature considers the discard phase to harm the environment the most and thus, aims at prevention of disposal since e.g. prolonging lifespans as well as reusing (including cleaning or repairing), reuse through organizations, and material recycling yield the highest energy and CO_2 savings: Laitala (Laitala, 2014) found the most common reasons for disposal to be wear and tear, poor fit, boredom, and a lack of storage space but suggested to deliver the apparel for reuse (e.g., donating it). Similar results were yielded by Lang et al. (Lang et al., 2013) who found that fashion trend sensitivity, shopping frequency, higher incomes, younger age, and being female are positively correlated to clothing disposal and hence, a behavioral shift is needed. Goworek et al. (Goworek et al., 2018) investigated life cycle assessment to generate clothing longevity via design, maintenance, and reuse to prevent early disposal. Morgan and Birtwistle (Morgan & Birtwistle, 2009) revealed that young female consumers are unaware of the need to recycle their clothing and thus, demand more information and clarification by e.g. media. This finding was replicated by Birtwistle and Moore (Birtwistle & Moore, 2007) indicating that clothing lost intrinsic value and hence, encouraging consumers to replace or dispose their apparel at an early phase during the clothings' life cycle. Goworek et al. (Goworek et al., 2012) found lacking consciousness among consumers regarding the facilities available to enable them to adopt more sustainable habits in terms of disposal and therefore, consumers' clarification is needed. Diddi et al. (Diddi et al., 2019) even demand for including education of repair skills in the high school curriculum to create a 'repair mindest' among young consumers to address the disposal culture. Further, Hu et al. (Hu et al., 2014) propose a closed-loop supply chain to adopt the circular use of clothing.

2.2 Barriers of Sustainable Clothing and Footwear Consumption and how to Avoid Them

Despite literature's suggestions regarding life cycle assessment to minimize the environmental impact of the clothing and footwear industry, consumers still exhibit an attitude-behavior gap (Kollmuss & Agyeman, 2002), i.e., although they pretend pro-environmental attitudes, they frequently do not behave sustainably. Drawing on Stern and Oskamp (Stern & Oskamp, 1987) and the subsequent work of Guagnano et al. (Guagnano et al., 1995), pro-environmental behaviors are the outcome of both internal (i.e., personal attitudes and values, beliefs, and knowledge) and external (i.e., macro-level forces outside of an individual's control like e.g. social institutions, economic forces, or physical structures) factors. Thereby, external conditions act as drivers of or barriers to certain behaviors (i.e., physical, financial, legal, or social sources). An individual's behaviors are consistent with her/his attitudes and values when external conditions are neutral. However, with external conditions making the resulting behavior e.g. difficult, inconvenient, time-consuming, or expensive, the behavior does not reflect one's attitudes or values (Guagnano et al., 1995; Stern & Oskamp, 1987). Consequently, previous literature identified internal and external barriers to sustainable behavior.

2.2.1 Internal Barriers and Remedies

Lacking knowledge and excessive amounts of complex information about sustainability were found to be internal inhibitors: Consumers with greater environmental knowledge were found more likely to engage in pro-environmental purchase behaviors in general (Goworek et al., 2018; McNeill & Moore, 2015; Meinhold & Malkus, 2005; Schahn & Holzer, 1990) and more specifically, in eco-conscious clothing and footwear consumption (Harris et al., 2016; Hiller Connell, 2010; H.-S. Kim & Damhorst, 1998). Consumers frequently require more information and better education of e.g. the materials used for production (Hill & Lee, 2012; Shaw et al., 2006; Shen et al., 2012) since they do not feel capable of making appropriate choices regarding eco-conscious clothing and footwear (Iwanow et al., 2005; Joergens, 2006). To overcome the barrier of limited consumers' knowledge, ecolabels or seals can be used to provide information for consumers and to indicate a product's environmental impact and sustainability level. Nevertheless, many different ecolabels are frequently associated with even more information and thus, might inhibit the purchasing process (D'Souza et al., 2007). Besides, Bly et al. (Bly et al., 2015) indicate that emotional associations of trust and authenticity rank as more sustainable than ecolabels.

Apart from limited knowledge, the lack of environmental concerns in consumers' attitudes (Hustvedt & Dickson, 2009; Shim, 1995), an overall negative attitude towards sustainable products (Hiller Connell, 2010; Hustvedt & Dickson, 2009; Song & Ko, 2017), and differing values (Blake, 2001) might be internal barriers to sustainable apparel consumption.

Another barrier to reduce consumption levels are perceived high search costs (i.e., perceived time and effort (Ellen, 1994)) of consumers associated with the maintenance of clothing and footwear by extending lifespans. E.g., using collaborative consumption principles or donating (Armstrong et al., 2016; Iran et al., 2019; Laitala, 2014; Retamal, 2019; Zamani et al., 2017), repairing or cleansing services (Goworek et al., 2012; Laitala & Boks, 2012), or even recycling (i.e., creating a circular economy) are environmentally friendly alternatives to disposal and potential drivers to enhance sustainable consumption.

Generally, several studies (Brough et al., 2016; Eisler & Eisler, 1994; Lee & Holden, 1999; Luchs & Mooradian, 2012) indicated that men are less likely than women to embrace environmentally friendly behaviors. This so-called gender gap is sometimes referred to personality differences between the sexes (Luchs & Mooradian, 2012). It is assumed that women tend to be more prosocial, altruistic, and empathetic (Lee & Holden, 1999), to take a future time perspective (Eisler & Eisler, 1994), or to bother more regarding health and safety (Davidson & Freudenburg, 1996). Brough et al. (Brough et al., 2016) found the alternative explanation that the gender gap might (partially) be due to the many men's assumption that 'greenness' and 'femininity' are cognitively linked and that some of them subsequently – in order to maintain their gender-identity – avoid sustainable behaviors. As a consequence, Brough et al. (Brough et al., 2016) proposed that the men's willingness to engage in sustainable behavior can be influenced by affirming their masculinity and further, by using masculine rather than conventional green branding.

2.2.2 External Barriers and Remedies

External barriers to pro-environmental clothing and footwear consumption arise particularly from potential consequences of sustainability on the manufacturers' product ranges: sustainability frequently implies smaller product ranges or fewer fashion cycles and collections respectively and hence, does not meet consumers' demand (Hiller Connell, 2010; Pookulangara & Shephard, 2013). Further, sustainable apparel is frequently perceived as unfashionable or unstylish respectively by consumers (Harris et al., 2016; Hiller Connell, 2010). Consumers consider the appearance of sustainable fashion as unattractive and thus, such clothing and footwear does not suit their wardrobe needs nor meet their aesthetic needs (Beard, 2008; Joergens, 2006). Price, quality (with respect to materials and craftmanship), and appearance are even more important criteria to many consumers than ethical aspects (Joergens, 2006). The lack of aesthetic appearance is stressed by the restriction of pro-environmental clothing to natural materials (Lundblad & Davies, 2016) and thus, few different styles (e.g., different colors) are available.

Aside from product range consequences, the price of sustainable products might constrain pro-environmental behavior (Hiller Connell, 2010; Hustvedt & Dickson, 2009; Joergens, 2006; Roberts, 1996): According to Joergens (2006), consumers have limited choice in eco-conscious clothing as they perceive the prices as not comparable to low-cost garments. Since there have been only few technological advances regarding the mass production of sustainable clothing and footwear, eco-conscious garments carry higher prices than conventional products and hence, they are unaffordable for many consumers (Hiller Connell, 2010). Collaborative consumption platforms (Armstrong et al., 2016; Geissinger et al., 2019; Zamani et al., 2017) or manufacturers discounting sustainable purchases (e.g., for returning used clothing and footwear to be recycled into components for new garments) might be suitable alternatives to support eco-conscious clothing consumption.

In the following, we discuss our research design including the Kano methodology and how this methodology was applied to measure the impact of the discussed drivers and remedies using the apparel and sportswear industry as a demonstration example.

3 Research Design: A Segmented Kano Perspective for the Apparel and Sportswear Industry

3.1 Kano Model and the Segmented Kano Perspective

The Kano model (Kano et al., 1984) is a commonly applied approach to investigate the relationship between the performance or the existence of product or service attributes (i.e., components, elements, features, technologies, in our case: sustainability offerings) and customer satisfaction. Moreover, it allows to predict customer satisfaction – and consequently behavioral change – when product or service attributes are varied. For instance, Ingaldi and Ulswicz (Ingaldi & Ulewicz, 2019) tested whether sustainability offerings (e.g., re-usable packaging, participation in ecological programs) drive customer satisfaction within an online shop for organic products in order to increase sustainable consumption. Moreover, Rese et al. (Rese, A., Schlee, T., & Baier, D., 2019) investigated which new services and technologies could convince consumers to visit physical fast fashion stores and try on clothes in order to reduce returns caused by ordering not fitting clothes.

The main idea of our Kano model application is to categorize investigated sustainability offerings with respect to their relationship between their existence and customer satisfaction using the well-known Kano categories (Matzler et al., 1996; Mikulić & Prebežac, 2011; Nilsson-Witell & Fundin, 2005):

1. Offerings categorized as **must-be (M)** are assumed to be taken for granted by the customer. Existence does not lead to customer satisfaction but, in contrast, absence leads to customer dissatisfaction.
2. Offerings categorized as **one-dimensional (O)** is proportional to customer satisfaction: existence leads to customer satisfaction, absence to dissatisfaction.
3. Offerings categorized as **attractive (A)** are assumed to be not expected by customers. Their existence leads to customer satisfaction and, in contrast, their absence does not lead to customer dissatisfaction. The popularity of these offerings is assumed to be rather short-term and thus, they disappear or turn into must-be offerings.
4. Offerings categorized as **indifferent (I)** are assumed not to affect customer satisfaction. Hence, neither their existence nor their absence impacts customer satisfaction or dissatisfaction respectively.
5. The presence of **reverse (R)** offerings leads to customer dissatisfaction and their absence leads to customer satisfaction.
6. If none of the above categories can be assumed or assessed, the offerings are categorized as **questionable (Q).**

Offerings can be classified into these categories based on a customer sample. For each offering a pair of questions is raised: the first question tests the customer’s reaction if the considered offering is present (functional question) whereas the second question tests the customer’s reaction if the considered attribute is absent (dysfunctional question) (Kano et al., 1984; Matzler et al., 1996; Mikulić & Prebežac, 2011). Ideally, the questions are formulated reflecting the ‘voice of the customer’, i.e., are written in a form that can be easily understood by the customer (Hauser & Clausing D., 1988). For both questions, the potential answers are ordinally scaled and range from (1) ‘I like it’, (2) ‘It must be that way’, (3) ‘I do not mind it’ to (4) ‘I can live with it’ and (5) ‘I do not like it’ (Berger, C., Blauth, R., Boger, D. et al., 1993; Matzler et al., 1996; Nilsson-Witell & Fundin, 2005). The answers reflect the extent to which customer satisfaction is generated if the offering is available and vice versa, the extent to which dissatisfaction is generated if the offering is absent. Combining both answers, offerings can subsequently be classified into one of the categories by using the Kano table (see Table 1).

The derived individual categorizations can be utilized further by aggregating them across all respondents using the customer satisfaction (CS+) and customer dissatisfaction (CS-) indices (Berger, C., Blauth, R., Boger, D. et al., 1993; Shahin et al., 2013; Shahin & Zairi, 2009):

$$\text{CS+}=\frac{\#A+\#O}{\#A+\#O+\#M+\#I} \tag{1}$$

$$\text{CS-}=-\frac{\#O+\#M}{\#A+\#O+\#M+\#I} \tag{2}$$

with #A, #I, #M, and #O being the categorization frequencies, i.e., number of respondents who classified the offering as attractive, indifferent, must-be, or one-dimensional respectively.

Table 1: Kano table: Categories derived from answers to the (dys-) functional questions.

		Dysfunctional question				
		(1) Like	(2) Must be	(3) Neutral	(4) Live with	(5) Dislike
Functional question	(1) Like	Q	A	A	A	O
	(2) Must be	R	I	I	I	M
	(3) Neutral	R	I	I	I	M
	(4) Live with	R	I	I	I	M
	(5) Dislike	R	R	R	R	Q

Note: A=Attractive; I=Indifferent; M=Must-Be; O=One-Dimensional; Q=Questionable; R=Reverse.

The indices reflect the proportion of respondents for whom the existence (absence respectively) of an offering attribute impacts customer satisfaction (customer dissatisfaction respectively). Additionally, CS- has a minus sign to emphasize the negative effects on customer satisfaction (for historical reasons). For each offering, the satisfaction index is within the range of [0, 1] and for customer dissatisfaction within [-1, 0]. A value close to 1 for CS+ indicates a high proportion of customers among whom satisfaction can be generated and a value close to -1 indicates a high proportion of respondents among whom dissatisfaction can be generated. The scale mean 0.5 for CS+ (or -0.5 for CS- respectively) indicates whether the majority of respondents can be positively (or negatively respectively) stimulated, yielding a two-dimensional grid with four quadrants:

$$\text{Attractive offerings, if } \begin{cases} 0.5 \leq \text{CS+} \leq 1 & \text{and} \\ 0 \geq \text{CS-} > -0.5 \end{cases}$$

$$\text{Indifferent offerings, if} \begin{cases} 0 \leq CS+ < 0.5 & \text{and} \\ 0 \geq CS- > -0.5 \end{cases}$$

$$\text{Must-be offerings, if} \begin{cases} 0 \leq CS+ < 0.5 & \text{and} \\ -0.5 \geq CS- \geq -1 \end{cases}$$

$$\text{One-dimensional offerings, if} \begin{cases} 0.5 \leq CS+ \leq 1 & \text{and} \\ -0.5 \geq CS- \geq -1 \end{cases}$$

The respondents classifying the offering as reverse (category R, frequency #R) or questionable (category Q, frequency #Q) are not reflected by the CS+ and CS- indices and the grid, since only respondents with 'strong' assessments are taken into consideration. Thus, aside from the satisfaction indices, we can determine the total strength for each offering, which indicates the proportion of attractive, one-dimensional, and must-be assessments of this offering among all assessments:

$$\text{Total Strength} = \frac{\#A + \#M + \#O}{\#A + \#I + \#M + \#O + \#Q + \#R} \tag{3}$$

Recently, an alternative to the above described aggregated analysis has been proposed and applied, the so-called Segmented Kano perspective. The respondents are clustered according to their assessments using two-mode metric cluster analysis with respect to their responses to the answers to the functional and dysfunctional questions (see Baier et al. (Baier et al., 2018) using e.g. double k-means algorithms) or using one-mode non-metric cluster analysis with respect to the derived categories (see Rese et al. (Rese, A., Schlee, T., & Baier, D., 2019) and Baier and Rese (Baier & Rese, 2020), using Chiu et al.'s (Chiu et al., 2001) well-known two-step algorithm or – with similar results – k-means after binary dummy coding of the categories). The number of clusters can be determined using the usual Bayesian Information Criteria with respect to the corresponding likelihood functions. Particularly in case highly innovative offerings are investigated, the Segmented Kano perspective is preferable since the usual categorizations as attractive or indifferent at the aggregated level are reduced and consumer segments can be identified which are highly receptive.

3.2 Apparel and Sportswear Industry Use Case Selection, Concretization of the Offerings, Questionnaire

During questionnaire development in order to measure the impact of the discussed offerings on sustainable clothing and footwear consumption, it became apparent that the offerings had to be concretized for a certain industry and product range to gain comparable valid insights.

For this purpose, the selection of the apparel and sportswear industry and the consumption of Adidas sneakers as a use case seemed to be a reasonable choice: Within the clothing and footwear industry, apparel and sportswear is a large but still rapidly growing market with a worldwide revenue of \$ 180.96 billion in 2019 and an estimated worldwide revenue of \$ 207.79 billion in 2025. Major players are Nike (\$ 34.88 billion in 2019), Adidas (\$ 23.64 billion), VF Corp. (\$ 13.29 billion), Puma (\$ 5.08 billion), Under Armour (\$ 4.86 billion). Innovations and sustainability are key factors for success in this market. The competition is high, and, moreover, market boundaries are disappearing and other clothing and footwear companies try to enter the attractive sneaker market. E-commerce is becoming increasingly important. Therefore, omnipresence and a good brand image are essential (see statistics and market overview in https://de.statista.com/themen/1626/sportartikel).

Moreover, sneaker consumption is closely related to clothing consumption, particularly when discussing successful sustainability offerings. Adidas – headquartered in Herzogenaurach, Bavaria (Germany) – recently become famous for selling more than 11 million eco-conscious Parley sneakers in 2019. Aside from these comfortable high-performance shoes whose prices vary from € 89 to € 179, the collection further comprises shirts and tights. The whole collection is all made of ocean plastic that has been processed to wool yarn. Currently, Adidas expands its eco-conscious efforts further and develops apparel and sportswear made of biobased materials (e.g., biodegradable biopolymers produced in large-scale bioreactors, the material called biosteel) and moreover, makes use of circular business models in which used and returned products are reutilized as components for new products (see Adidas reports at www.adidas.de or https://m.adidas.de/sustainability-parley-ocean-plastic).

Choosing Adidas sneakers as a case for our investigation and drawing on the extant literature on drivers, barriers, and remedies discussed in Section 2 as well as planned implementations in the apparel and sportswear industry (collected via press releases, newsletters, blogs in the internet), a preliminary list of essential sustainability aspects and offerings to consumers was developed by the authors. Then, this list was discussed with two experts responsible for the product range at a major global apparel and sportswear retailer, three experts responsible for product development, design, and marketing at two major global apparel and sportswear manufacturers, and one expert of a major global material science company. Additionally, two 90-minute workshops with a sample of 42 apparel and sportswear consumers (university students) served to finalize the list as shown in Table 2 with short descriptions of the sustainability offerings.

Table 2: Sustainable aspects and offerings for apparel and sportswear consumers together with references that discuss these offerings as helpful to increase sustainable clothing consumption.

Aspect	Offering	Detailed description	References
Range	Sustainable	Only sustainable products are offered.	(Ellen, 1994; Lundblad & Davies, 2016)
	Natural	Only products in natural colors are offered.	(Lundblad & Davies, 2016)
	Separate	A separate section with sustainable products is offered.	(Ellen, 1994; Lundblad & Davies, 2016)
	Small(er)	A small(er) range is offered (e.g. 1,000 instead of 3,000 sneakers).	(Hiller Connell, 2010)
	Few(er)	Few(er) fashion cycles are launched (e.g. a specific collection is sold for two monthles instead of a two weeks).	(Hiller Connell, 2010)

Aspect	Offering	Detailed description	References
Labeling	Traffic light	Products are classified with a traffic light to indicate the sustainability level (e.g. red: minimum, yellow: re/upcycled, green: biobased).	(D'Souza et al., 2007; Hiller Connell, 2010; H.-S. Kim & Damhorst, 1998)
	Removable	Products are tagged with a removable seal to indicate a high sustainability level.	(D'Souza et al., 2007; Hiller Connell, 2010; H.-S. Kim & Damhorst, 1998)
	Hidden	Products are tagged with a hidden seal to indicate a high sustainability level (e.g. on the sneaker sole).	(D'Souza et al., 2007; Hiller Connell, 2010; H.-S. Kim & Damhorst, 1998)
	Visible	Products are tagged with a visible seal to indicate a high sustainability level.	(D'Souza et al., 2007; Hiller Connell, 2010; H.-S. Kim & Damhorst, 1998)
	Certificate	Products are officially certified by an ecolabel to indicate a high sus-tainability level (e.g. Blauer Engel, Organic Textile, Fair trade).	(D'Souza et al., 2007; Hiller Connell, 2010; H.-S. Kim & Damhorst, 1998)
Processes	Return	A discount on the purchase of the next sustainable product is offered when products are returned to be recycled.	(Armstrong et al., 2016; Hiller Connell, 2010; Hustvedt & Dickson, 2009; Roberts, 1996)
	Discount	A discount on the purchase of the next sustainable product is offered when a sustainable product is bought.	(Armstrong et al., 2016; Hiller Connell, 2010; Hustvedt & Dickson, 2009; Roberts, 1996)
	Bonus card	Bonus points are collected when buying products. The more sustainable, the more points. Points can be redeemed when buying sustainnable products. Further, when reaching a minimum number of points, customers get early access to new sustainable products.	(Armstrong et al., 2016; Hiller Connell, 2010; Hustvedt & Dickson, 2009; Roberts, 1996)
Materials	Upcycled	Products are made of upcycled materials (e.g., ocean plastic).	(Allwood et al., 2008; Morgan & Birtwistle, 2009)
	Biobased	Products are made of biobased and -degradable materials (e.g., biosteel).	(Allwood et al., 2008; Morgan & Birtwistle, 2009)

Aspect	Offering	Detailed description	References
	Recycled	Products are made of components derived from returned products.	(Allwood et al., 2008; Morgan & Birtwistle, 2009)
	Cleansing	A repair and/or cleansing service is offered to prolong the product life cycle.	(Laitala & Boks, 2012; Niinimäki & Hassi, 2011)

Based on Table 2 our sustainability offerings were exemplified and illustrated for a questionnaire using – as discussed – Adidas sneakers for exemplification. Overall the questionnaire was constructed as follows: We first provided a general introduction to the study. Then, the respondents were asked few introductory questions regarding their consumption behavior in general (i.e., amount of sneakers bought per year, willingness to pay for conventional sneakers, purchase behavior regarding sneakers, importance of different criteria when buying sneakers, and sneaker brands bought in the last three years), her/his perception of sustainable consumption in general (i.e., perceived associations with sustainable consumption, perception of own knowledge regarding sustainability, and inhibitors of purchasing sustainable sneakers), and her/his willingness to pay for sustainable sneakers. The asked criteria when buying sneakers – as well as the list of offerings – were the results of the literature review combined with the expert interviews and consumer workshops. Thus, the criteria, which were identified as most important, were included into the questionnaire: price, quality (i.e. materials and craftmanship), appearance, brand, sustainability, comfort, and longevity.

The main part of the questionnaire was based on the Kano model. The respondents were introduced and made familiar with the specific technique and syntax of functional as well as dysfunctional questions of a Kano model being asked to declare satisfaction in case of presence and in case of absence of a sustainability offering. The 17 offerings of Table 2 were presented within a fictitious scenario comprising detailed explanations and exemplary presentations and illustrations. The offerings referred to the Adidas sneaker product range (even though it is clear that it could have also been referred to other apparel and sportswear ranges) and were not limited to purchases in the Adidas online shop nor in another online shop as well as an Adidas physical store or another apparel/sportswear physical store. For many offerings (e.g., colors, traffic lights, removable, hidden, or visible labeling, return or sustainability discount, upcycled, biobased, or recycled materials, and cleansing) the scenario was illustrated using Adidas sneakers as an example. The subsequent part of the survey contained control questions assessing the importance of the 17 sustainability offerings when buying sneakers using a 6-point Likert scale ranging from ‘I don’t mind it’ to ‘Very important’ (same scale as the importance questions in the introductory part). The last part included questions regarding demographic data (i.e., age, gender, employment status, place of residence, and monthly income) and boxes to leave comments regarding sustainable consumption behavior in general, the chosen case (Adidas) or the questionnaire design.

3.3 Descriptive Statistics of the Sample of Typical Apparel and Sportswear Consumers

The online questionnaire was spread in various social media channels and in several courses at a German university from December 17, 2019 to January 20, 2020. A total of 635 responses were recorded. 145 questionnaires (22.83%) were rejected since they had not been completed. This might be due to the length of the survey with an average response time of 18 minutes. Thus, a total of 490 filled out questionnaires was considered for further analysis. Table 3 outlines the sample’s descriptive statistics. The survey was more often completed by women (56.53% of responses). 54.49% of our participants were 15 to 24 years old. Participants, which were 35 to 64 years old, constituted a smaller proportion among the respondents (5.92%). This seems reasonable since these people are less likely to use social media and we distributed our questionnaire in university. The latter is reflected by the high proportion of students (77.35% of the respondents) participating in the study. The sample reflects to some extent the target segment for sneakers (younger than the average population) and the Adidas customers (younger and a higher percentage of university students and academics than the average population).

We gained further insights into the participants’ consumption behavior with our nine introductory questions. The first portion inquired their consumption behavior in general: 41.8% of the participants are buying 2 to 3 sneakers per year and 61.0% of the respondents indicated that their willingness to pay for conventional sneakers is around € 50 to € 99. 27.8% of the participants are buying sneakers when they perceive a need. 21.0% of the respondents get inspired during browsing in an online shop. Furthermore, respondents ranked the importance of different sneakers’ buying criteria on a 6-point Likert scale from ‘I do not mind’ (=1) to ‘Very important’ (=6). The most important buying criteria (in decreasing order) were appearance (mean 5.63; standard deviation (SD) 0.701), comfort (mean 5.16; SD 0.951), and quality (w.r.t. materials and craftmanship; mean 4.99; SD 0.942). The least important criteria were the sneakers’ sustainability (mean 3.14; SD 1.295), the brand (mean 3.95; SD 1.478), and price (mean 4.27; SD 1.214). The majority of the participants bought Adidas (360 responses, 73.9%) and Nike (325 responses, 66.7%) sneakers within the last three years.

Table 3: Descriptive statistics of the sample (n=490).

Demographics	Specification	Proportion
Age	**15-24**	54.49%
	25-34	39.59%
	35-64	5.92%
Gender	**Female**	56.53%
	Male	43.27%
Employment status	**Students**	77.35%
	Employed	14.49%
	Freelancer	2.65%
	Other	5.51%

The second portion was about the respondents' perception of sustainability in general. Participants mainly associate sustainable consumption with purchasing environmentally friendly products (388 responses, 79.2%), purchasing durable and repairable products (309 responses, 63.1%), and purchasing fair trade products (263 responses, 53.7%). Fewer respondents associated sustainable consumption with less purchasing (221 responses, 45.1%), few had no idea (n=15, 3.1%). Regarding the major inhibitors of sustainable consumption behavior, respondents referred to lacking information about sustainable products (247 responses, 50.4%), high prices (245 responses, 50.0%), few alternatives (n=206, 42.0%), and everyday routine/habits (202 responses, 41.2%). Additionally, the participants' willingness to pay for sustainable sneakers was inquired. On average, the respondents would spend € 116.53 (mean; SD € 15.32) for sustainable sneakers in case a comparable conventional sneaker would cost € 100. (Please note that prices for Parley sneakers range between € 89 and € 179). Hence, the majority exhibit a higher willingness to pay for eco-conscious sneakers than for conventional ones. This is comparable to the findings of Niinimäki (Niinimäki, 2010) who found the willingness to pay for sustainable sneakers to be approximately 10-14% higher than for ordinary sneakers.

4 Results: A Segmented Kano Perspective for the Apparel and Sportswear Industry

Table 4 reflects the overall assessment of the sustainability aspects and offerings based on the Kano model, indicating category frequencies, the total share TS as well as the customer satisfaction index CS+ and the customer dissatisfaction index CS-. The sustainability offerings are mostly categorized as attractive and indifferent. Particularly, the offerings discount for sustainable products, recycled materials, sustainable product range, discount for returned products, traffic lights, separate sustainability-section, and upcycled materials were categorized by more than half of the respondents as attractive. Implementing these offerings would increase overall customer satisfaction significantly. In contrast, natural colors, small(er) product range, and few(er) life cycles were categorized by more than half of the respondents as indifferent. Some offerings were categorized as one-dimensional by more than 20% of the respondents indicating that not only their presence would increase satisfaction but further their absence would decrease satisfaction: biobased materials, return discount for used products, traffic light indicating the sustainability level, cleansing and repairing service, a separate sustainability-section, and upcycled materials. Finally, few offerings even reduced consumer satisfaction in case they would be implemented: approximately 20% of the respondents categorized visible labeling and small(er) product range as reverse offerings.

Figure 1 illustrates the preceding findings. The offerings are positioned with respect to their CS+ and CS- values. The four quadrants visualize the respondents' majorities as discussed in section 3.1 with respect to the strong categories. Most offerings are categorized as attractive and few offerings are categorized as indifferent. Since attractive and indifferent categorizations of innovative offerings are frequently the case in Kano investigations, a Segment Kano perspective was developed to gain further insights as proposed in section 3.1: The individual categorizations were analyzed using the well-known two-step clustering procedure by Chiu et al. (Chiu et al., 2001) and the Bayesian information criterion for determining the number of clusters. A three-cluster-solution was found with cluster 1 termed 'Segment 1' (n=203, 41.4%), cluster 2 termed 'Segment 2' (n=142, 29.0%) and cluster 3 termed 'Segment 3' (n=145, 29.6%). Table 5 provides further insights into the categorizations at the segment level.

Table 4: Overall assessment of sustainability aspects and offerings.

	Offering	Overall category frequencies (n=490)								
		#A	**#I**	**#M**	**#O**	**#Q**	**#R**	**TS**	**CS+**	**CS-**
Range	**Sustainable**	**286**	125	15	49	1	14	64%	.7053	-.1347
	Natural	62	**305**	35	16	9	63	26%	.1866	-.1220
	Separate	**249**	100	21	114	1	5	58%	.7500	-.2789
	Small(er)	83	**257**	25	27	2	96	26%	.2806	-.1327
	Few(er)	104	**245**	30	82	2	27	33%	.4035	-.2430
Labeling	**Traffic light**	**263**	76	11	133	2	5	57%	.8199	-.2981
	Removable	**202**	170	27	73	2	16	50%	.5826	-.2119
	Hidden	**205**	**205**	10	46	2	22	45%	.5386	-.1202
	Visible	115	**241**	5	28	1	100	25%	.3676	-.0848
	Certificate	**159**	146	66	108	1	10	53%	.5574	-.3633
Processes	**Return**	**285**	46	9	146	2	2	61%	.8868	-.3189
	Discount	**313**	69	6	86	2	14	66%	.8418	-.1941
	Bonuscard	**207**	196	5	49	1	32	44%	.5602	-.1182
Materials	**Upcycled**	**245**	108	17	106	2	12	57%	.7374	-.2584
	Biobased	**228**	78	26	153	2	3	55%	.7856	-.3691
	Recycled	**283**	84	17	98	3	5	64%	.7905	-.2386
	Cleansing	**224**	125	21	116	2	2	53%	.6996	-.2819

Note: The most frequent category is marked in bold. A=attractive; I=indifferent; M=must-be; O=one-dimensional; Q=questionable; R=reverse; TS=total strength; CS+=customer satisfaction index; CS-=customer dissatisfaction index.

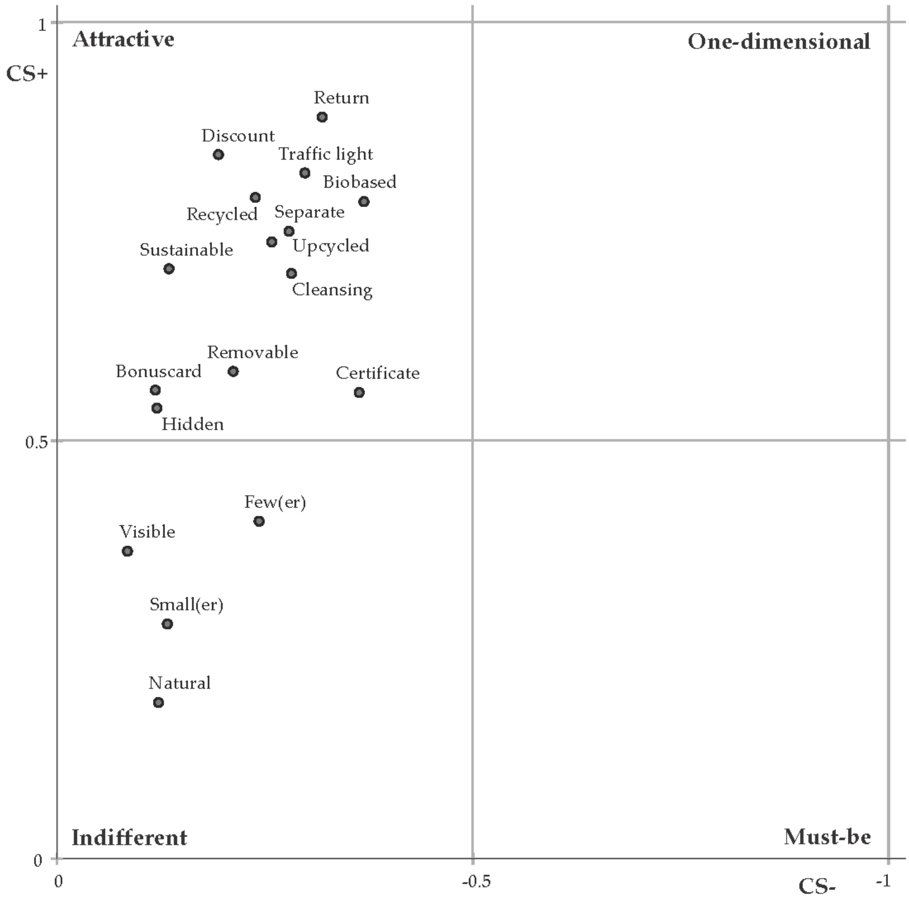

Figure 1: Depiction of the overall assessment of sustainability offerings (n=490).

Table 5: Segment-specific assessment of sustainability aspects and offerings.

	Offering	**Segment-specific category frequencies (Segment 1: n=203 / 2: n=142 / 3: n=145)**					
		#A	**#I**	**#M**	**#O**	**#Q**	**#R**
Range	**Sustainable*** **	**126** / 51 / **109**	17 / **78** / 30	11 / 2 / 2	48 / 1 / 0	0 / 1 / 0	1 / 9 / 4
	Natural* **	29 / 6 / 27	**120 / 109 / 76**	30 / 2 / 3	11 / 3 / 2	3 / 6 / 0	10 / 16 / 37
	Separate* **	87 / 49 / **113**	5 / **73** / 22	14 / 7 / 0	**97** / 8 / 9	0 / 1 / 0	0 / 4 / 1
	Small(er)* **	51 / 10 / 22	**95** / **87** / **75**	21 / 1 / 3	25 / 1 / 1	0 / 2 / 0	11 / 41 / 44
	Few(er)* **	47 / 16 / 41	**64** / **104** / **77**	26 / 3 / 1	62 / 9 / 11	1 / 1 / 0	3 / 3 / 1
Labeling	**Traffic light*** **	79 / **63** / **121**	6 / 60 / 10	6 / 3 / 2	**110** / 12 / 11	1 / 1 / 0	1 / 3 / 1
	Removable* **	**66** / 27 / **109**	57 / **90** / 23	18 / 4 / 5	58 / 8 / 7	0 / 2 / 0	4 / 11 / 1
	Hidden* **	**86** / 39 / **80**	71 / **85** / 49	7 / 3 / 0	35 / 3 / 8	0 / 2 / 0	4 / 10 / 8
	Visible* **	47 / 13 / **55**	**105** / **85** / 51	3 / 2 / 0	21 / 4 / 3	0 / 1 / 0	27 / 37 / 36
	Certificate* **	57 / 21 / **81**	27 / **89** / 30	33 / 19 / 14	**83** / 7 / 18	0 / 1 / 0	3 / 5 / 2
Processes	**Return*** **	**102** / **75** / **108**	14 / 28 / 4	5 / 4 / 0	81 / 32 / 33	0 / 2 / 0	1 / 1 / 0
	Sustainable* **	**109** / **77** / **127**	23 / 40 / 6	4 / 2 / 0	62 / 14 / 10	0 / 2 / 0	5 / 7 / 2
	Bonuscard* **	**87** / 25 / **95**	71 / **93** / 32	2 / 3 / 0	37 / 6 / 6	0 / 1 / 0	6 / 14 / 12
Materials	**Upcycled*** **	**86** / 52 / **107**	15 / **73** / 20	16 / 0 / 1	85 / 9 / 12	0 / 2 / 0	1 / 6 / 5
	Biobased* **	64 / 50 / **114**	5 / **70** / 3	15 / 5 / 6	**118** / 14 / 21	0 / 2 / 0	1 / 1 / 1
	Recycled* **	**102** / 62 / **119**	9 / **63** / 12	12 / 2 / 3	80 / 8 / 10	0 / 3 / 0	0 / 4 / 1

Offering	Segment-specific category frequencies (Segment 1: n=203 / 2: n=142 / 3: n=145)					
	#A	#I	#M	#O	#Q	#R
Cleansing* **	**78** / **65** / **81**	39 / 63 / 12	15 / 2 / 3	70 / 8 / 10	1 / 3 / 0	0 / 4 / 1

Note: The most frequent category per segment is marked in bold. A=attractive; I=indifferent; M=must-be; O=one-dimensional; Q=questionable; R=reverse; differences across segments are analyzed using the χ^2 test of independence with ***: p<0.01; **: p<0.05, *: p<0.1.

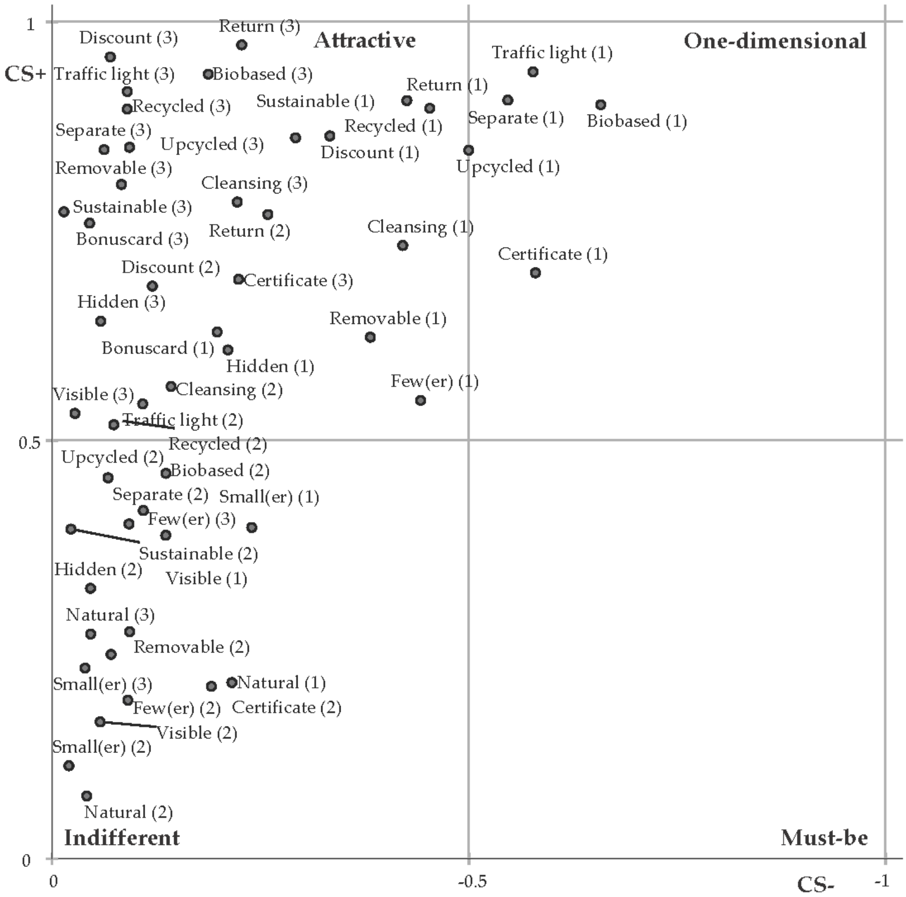

Figure 2: Depiction of the segment-specific assessment of sustainability offerings (Segment 1: n=203, Segment 2: n=142, Segment 3: n=145; segment numbers are in parentheses).

Apparently, the categorizations significantly vary across the segments and offerings. The majority of the respondents in Segment 1 categorizes biobased materials and a traffic light system indicating the sustainability level as one-dimensional offerings and further, a sustainable product range, a discount for sustainable products, and a discount for returned products as attractive offerings. In contrast to Segment 1, the majority of the Segment 3 rates almost every offering as attractive whereas a majority of Segment 2 rates almost every offering as indifferent. Figure 2 visualizes these differences with an illustration of the segment-specific assessments according to the CS+ and CS- values. Further, the differences between the three segments with respect to selected background variables can be seen in Table 6.

Table 6: Segment-specific descriptive statistics (segment 1: n=203, segment 2: n=142, segment 3: n=145).

Aspect	Specifications	Overall	Segment 1	Segment 2	Segment 3
Age	**15-24**	54.5%	56.6%	53.5%	52.5%

Aspect	Specifications	Overall	Segment 1	Segment 2	Segment 3
	25-34	39.6%	34.4%	45.1%	41.4%
	35-64	5.92%	9.0%	1.4%	6.1%
Gender* **	**Female**	56.5%	65.5%	46.5%	46.2%
	Male	43.3%	34.5%	52.8%	53.8%
Employment status	**Students**	77.4%	72.0%	83.8%	78.6%
	Employed	14.5%	16.3%	11.3%	15.2%
	Freelancer	2.7%	5.4%	0.0%	1.4%
	Other	5.5%	6.3%	5.9%	14.8%
Sneakers bought per year	**<1**	8.4%	11.8%	7.7%	4.1%
	1	34.1%	36.0%	31.7%	33.6%
	2-3	41.8%	38.4%	44.4%	44.1%
	>3	9.4%	13.8%	16.2%	18.2%
Sustainable consumption attitude in general	**I'm informed*** **	3.17 (0.911)	3.35 (0.833)	2.89 (0.939)	3.21 (0.927)
	Important for me* **	3.45 (0.954)	3.81 (0.876)	3.01 (0.956)	3.39 (0.859)
	I would pay more* **	3.69 (1.008)	4.07 (0.847)	3.15 (1.106)	3.68 (0.872)
	Labels are helpful* **	4.40 (0.908)	4.74 (0.559)	3.79 (1.166)	4.52 (0.698)
Sneakers buying motives	**I buy when needed**	3.48 (1.297)	3.48 (1.291)	3.57 (1.318)	3.37 (1.289)
	I buy when inspired	3.25 (1.321)	3.16 (1.316)	3.25 (1.274)	3.36 (1.372)
Importance when buying sneakers	**Price****	4.26 (1.214)	4.43 (1.130)	4.20 (1.216)	4.08 (1.299)
	Quality	4.99 (0.942)	5.06 (0.960)	4.91 (0.929)	4.97 (0.928)
	Appearance	5.63 (0.701)	5.59 (0.714)	5.58 (0.707)	5.73 (0.669)
	Brand* **	3.95 (1.478)	3.66 (1.538)	4.18 (1.462)	4.14 (1.342)
	Sustainability* **	3.14 (1.295)	3.46 (1.290)	2.80 (1.297)	3.02 (1.199)
	Comfort	5.16 (0.951)	5.26 (0.915)	5.06 (0.921)	5.12 (1.020)
	Longevity* **	4.35 (1.213)	4.46 (1.213)	4.07 (1.224)	4.47 (1.173)
Willingness to pay for sustainable sneakers compared to conventional ones for € 100.00* **		€ 116.53 (€15.32)	€ 120.61 (€ 14.52)	€ 110.66 (€ 16.30)	€ 116.58 (€ 13.57)

Note: Quality refers to materials and craftsmanship, differences across segments are analyzed with ***: p<0.01; **: p<0.05, *: p<0.1 using χ^2-tests of independence for the nominal aspects (age, gender, employment status, number of sneakers bought) and F-tests for the metric aspects (attitude and motives with scales 1='totally disagree' to 5='totally agree'; importance with scales 1='I do not mind' to 6='very important'; given are mean values (standard deviation)).

Segment 1 – the segment where most sustainability offerings are rated as attractive or even one-dimensional – has a significantly higher proportion of female respondents. For them, the importance of price, sustainability, and longevity is also significantly higher than for the average respondent and moreover, their willingness to pay for sustainable sneakers is significantly higher. However, the importance of brand is significantly lower in this segment. Segment 3 – the segment where most sustainability offerings are rated as attractive – has an above average importance of longevity and price than the average consumer, whereas for Segment 2 – the segment where almost every sustainability offering is rated as indifferent – the brand's importance is significantly higher than for the average respondent. However, it must be mentioned that the importance of sustainability across all segments is low (compared to the other criteria). In other words, the buying process of sneakers seems to be – across all segments – dominated by appearance, comfort, and quality (w.r.t. materials and craftsmanship).

Since the gender distribution differs significantly across the segments, additional χ^2-tests of independence were performed to see whether the categorizations of the offerings directly depend on the participants' gender. Across the 17 sustainability offerings, only one offering (recycled materials) exhibits significant dependency at the $p<0.01$ level and four offerings (traffic light, sustainable discount, upcycled materials, recycled materials) exhibit significant dependencies at the $p<0.05$ level. This further proves the usefulness of the Segmented Kano perspective: The three segments differ significantly across the categorizations of all attributes and the participants' gender differs significantly across the segments but the participants' gender cannot be used exclusively to derive the separable segments with respect to categorizations.

Overall, our results indicate that sustainability offerings are attractive and one-dimensional for many consumers, particularly for females, but – compared to purchase criteria like appearance, comfort, and quality (w.r.t. materials and craftmanship) – they are of inferior importance when purchasing sneakers. However, it is not clear, whether these findings can be extended to all sorts of apparel and sportswear.

5 Discussion

5.1 Theoretical Contribution

Our findings contribute to a deeper understanding of consumers' sustainable clothing and footwear consumption behavior particularly during the pre-purchase phase. Prior research focused either on sustainable consumer behavior in general or investigated sustainable clothing consumption behavior during the post-purchase phase to prevent clothing disposal. We fill a research gap by capturing the consumer's perspective regarding key sustainability aspects and offerings in terms of product range, labeling, processes, and materials. Hence, drivers (and potential inhibitors) of pro-environmental clothing consumption were determined which might help to overcome the consumers' attitude-behavior gap. Thereby, we make several theoretical contributions to extant literature:

First, our findings indicate that discounts for returned products, discounts for sustainable purchases, traffic lights indicating sustainability levels, and biobased materials are highly attractive to the participants. These results extend and align with extant literature (see e.g., (Hiller Connell, 2010; Hustvedt & Dickson, 2009; Joergens, 2006; Roberts, 1996)) proving that consumers' sustainable clothing and footwear consumption is frequently constrained due to high prices. Further, preceding research (see e.g., (Hill & Lee, 2012; Shaw et al., 2006; Shen et al., 2012)) found consumers to have limited knowledge regarding sustainability and thus, we identified labeling (e.g., traffic lights) indicating the sustainability level of products as suitable solution in order to enhance sustainability-related knowledge.

Approximately 20% of the respondents categorized visible labeling and a small(er) product range as reverse offerings leading to customer dissatisfaction. These findings are complementing the preceding work of e.g. Harris et al. (Harris et al., 2016), Hiller Connell (Hiller Connell, 2010), or Pookulangara and Shephard (Pookulangara & Shephard, 2013) who found that pro-environmental clothing frequently does not meet the aesthetic needs and wants of consumers. This is particularly reflected by our participants' aversion towards a smaller product range. Further, visible labeling might be perceived as unfashionable or disturbing by the consumers.

By applying the Segmented Kano perspective, we gained further insights into the participants' consumption behavior: particularly for female consumers[5], the importance of price, sustainability, and longevity of clothing is significantly higher than for the average respondent. This aligns with the gender gap findings of prior literature (Brough et al., 2016; Eisler & Eisler, 1994; Lee & Holden, 1999; Luchs & Mooradian, 2012) showing that women are more likely to engage in pro-environmental behavior than men. In our investigation, female respondents classified biobased materials and a traffic light system indicating the sustainability level as one-dimensional offerings, i.e., the customer satisfaction grows proportionally with an increasing degree of the offerings' implementation. Their categorizations are particularly emphasized by previous results of Morgan and Birtwistle (Morgan & Birtwistle, 2009) which found a lack of knowledge among female consumers (despite their pro-environmental mental attitudes) and thus, more information regarding the sustainability of products is needed e.g. by using a simple traffic light system.

Nevertheless, the overall importance of sustainability among consumers is still marginal compared to predominant purchasing criteria like appearance, comfort, and quality. This is strengthening the preceding results of Joergens (Joergens, 2006) who proved quality and appearance of clothing to be more important criteria to many consumers than ethical aspects.

5.2 Managerial Implications

This study provides several managerial implications. First of all, it needs to be highlighted that – in the apparel and sportswear industry – sustainability offerings still play a minor role compared to traditional buying arguments such as appearance, comfort, and quality (i.e. materials and craftmanship). Nevertheless, our study also shows that sustainability aspects and offerings can have a positive impact on customer satisfaction, ultimately leading to increased sales and brand value. In this research, we discussed and applied a new methodology which is able to test a variety of sustainability offerings related to the product range (e.g., purely sustainable product range), labeling (e.g., traffic light indicating sustainability level), sustainable processes (e.g., discount on future purchases for returned products in order to recycle them into components for new products), and used materials (e.g., products made from biobased materials). We could show that these offerings differ significantly in their impact on customer satisfaction. Also, we could show that female (vs. male) consumers are far more receptive for sustainable offerings. This has an implication both for sustainability-related innovations as well as their marketing activities. On the one hand, companies need to ensure that sustainability-related marketing activities fit to typical needs of female consumers. On the other hand, it could be promising to integrate female consumers in early phases of innovation processes aiming for sustainable product offerings as well as

[5] I.e., the segment with a significantly high proportion of female respondents.

the design of related marketing activities and business models (e.g., how to return products to enable recycling).

In terms of the analyzed sustainable offerings, our study provides guidance on two levels. First, since the utilization of upcycled and recycled materials for sustainable products were classified as attractive offerings by the majority of our participants, the apparel industry should increase their efforts to use such materials in production. For instance, Adidas Parley shoes, shirts, and tights are made of upcycled ocean plastics. Further efforts should be made regarding recyclable products, e.g., shoes which are recyclable from laces to sole in order to enable a fully circular business model. Second, sustainable products should be promoted broadly with different marketing techniques: E.g., a separate sustainability-section in online shops as well as stationary stores may attract the consumers' attention and minimizes search costs. Specific labeling of sustainable products such as a traffic light system and removable or hidden seals for eco-conscious products reduces search efforts, enhances clarity, and might lead consumers towards a sustainable consumption behavior during their purchasing decision. Discounts for returned products or for sustainable products in general might further boost sustainable sales.

In sum, companies should proactively develop strategies to combine sustainable offerings and commercial success following the triple bottom line accounting approach. Our research indicates that combining the social contributions, environmental contributions, and economic contributions requires, on the one hand, meaningful sustainable products and, on the other hand, creative measures to communicate these products to the consumers.

5.3 Limitations and Future Research

Our research is subject to several limitations which stimulate further research. First, our sample mainly consists of Germans and university students. This seems reasonable since this fits to some extent to the typical sneaker buyers and to Adidas customers (younger and more students than the population). However, the generalizability of our results to other clothing and footwear product ranges as well as other target segments is limited. Even though in Germany the percentage of university students among their age cohorts now is rather large, one suggestion for further research is to include in the sample, e.g., more people with another employment status as well as more even younger and more older repondents to gain more valid results.

Second, our investigation is geographically constrained to Germany and therefore, we did not consider cultural differences since consumers of other countries might assess these aspects differently. Future research could replicate our study in a differing cultural context.

Third, the high number of offerings classified as indifferent or attractive might be related to the newness of the offerings but also to the questionnaire's length (~ 18 minutes response time on average). Besides, an increasing completion time leads to a higher number of early terminations and further, fatigue effects might occur. Further research on this topic could reduce the number of investigated offerings.

Finally, whereas the Kano model and the Segmented Kano perspective are able to compare large numbers of sustainability offerings, a clear advantage of the approach, some methodological limitations must be mentioned: So, e.g., as discussed, it often is time-consuming for the respondents and often yields many indifferent and attractive categorizations when applied to innovative offerings. Here, it could be helpful in the next step to concentrate on fewer offerings and apply conjoint analysis and experimental (field) research in order to validate and enhance our findings.

6 Conclusion

To overcome the attitude-behavior gap in terms of sustainable behavior among consumers, we investigated how companies can find out which sustainability offerings drive (or inhibit) consumer satisfaction in the clothing and footwear industry. We proposed and applied the Kano model and the Segmented Kano perspective for this purpose in an apparel and sportswear context. Typical consumers (n=490) evaluated 17 sustainability offerings regarding product range, labeling, processes, and (re-) utilized materials. The analysis results show that there are several sustainability offerings that can promote sustainable consumption. However, on the other side, it could be shown that there are sustainability aspects and offerings that are not suitable for this and should therefore be avoided. We hope that our study motivates both researchers and practitioners to contribute to the diffusion of sustainable products not only in the apparel and sportswear industry but also in other areas. Mass adoption and use of sustainable products is still rare and this must change.

References

Achabou, M. A., & Dekhili, S. (2013). Luxury and sustainable development: Is there a match? Journal of Business Research, 66(10), 1896–1903. https://doi.org/10.1016/j.jbusres.2013.02.011

Allwood, J. M., Laursen, S. E., Russell, S. N., Rodríguez, C. M. de, & Bocken, N. (2008). An approach to scenario analysis of the sustainability of an industrial sector applied to clothing and textiles in the UK. Journal of Cleaner Production, 16(12), 1234–1246. https://doi.org/10.1016/j.jclepro.2007.06.014

Armstrong, C. M., Niinimäki, K., Lang, C., & Kujala, S. (2016). A Use-Oriented Clothing Economy? Preliminary Affirmation for Sustainable Clothing Consumption Alternatives. Sustainable Development, 24(1), 18–31. https://doi.org/10.1002/sd.1602

Auger, P., & Devinney, T. M. (2007). Do What Consumers Say Matter? The Misalignment of Preferences with Unconstrained Ethical Intentions. Journal of Business Ethics, 76(4), 361–383. https://doi.org/10.1007/s10551-006-9287-y

Baier, D., & Rese, A. (2020). How to Increase Multichannel Shopping Satisfaction? A Segmented Kano Perspective for the Selection of New Technologies. Journal of Retailing and Consumer Services, In Press.

Baier, D., Rese, A., & Röglinger, M. (2018). Conversational User Interfaces for Online Shops: A Segmented Kano Perspective. Proceedings on the 39th International Conference on Information Systems (ICIS).

Barnes, L., & Lea-Greenwood, G. (2010). Fast fashion in the retail store environment. International Journal of Retail & Distribution Management, 38(10), 760–772. https://doi.org/10.1108/09590551011076533

Beard, N. D. (2008). The Branding of Ethical Fashion and the Consumer: A Luxury Niche or Mass-market Reality? Fashion Theory, 12(4), 447–467. https://doi.org/10.2752/175174108X346931

Berger, C., Blauth, R., Boger, D., Bolster, C., Burchill, G., DuMouchel, W., Pouliot, F., Richter, R., Rubinoff, A., Shen, D., Timko, M., & Walden, D. (1993). Kano’s Methods for Understanding Dustomer-defined Quality. Center for Quality of Management Journal, 2(4), 3–36.

Bianchi, C., & Birtwistle, G. (2010). Sell, Give Away, Or Donate: An Exploratory Study of Fashion Clothing Disposal Behaviour in Two Countries. The International Review of Retail, Distribution and Consumer Research,, 20(3), 353–368.

Bianchi, C., & Birtwistle, G. (2012). Consumer Clothing Disposal Behaviour: A Comparative Study. International Journal of Consumer Studies, 36(3), 335–341. https://doi.org/10.1111/j.1470-6431.2011.01011.x

Birtwistle, G., & Moore, C. M. (2007). Fashion clothing – where does it all end up? International Journal of Retail & Distribution Management, 35(3), 210–216. https://doi.org/10.1108/09590550710735068

Blake, D. E. (2001). Contextual Effects on Environmental Attitudes and Behavior. Environment and Behavior, 33(5), 708–725.

Bly, S., Gwozdz, W., & Reisch, L. A. (2015). Exit from the high street: an exploratory study of sustainable fashion consumption pioneers. International Journal of Consumer Studies, 39(2), 125–135. https://doi.org/10.1111/ijcs.12159

Brough, A. R., Wilkie, J. E. B., Ma, J., Isaac, M. S., & Gal, D. (2016). Is Eco-Friendly Unmanly? The Green-Feminine Stereotype and Its Effect on Sustainable Consumption. Journal of Consumer Research, 43(4), 567–582. https://doi.org/10.1093/jcr/ucw044

Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001). A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 263–268.

Csikszentmihalyi, M. (2000). The Costs and Benefits of Consuming. Journal of Consumer Research, 27, 267–272.

Davidson, D. J., & Freudenburg, W. R. (1996). Gender and Environmental Risk Concerns. A Review and Analysis of Available Research. Environment and Behavior, 28(3), 302–339.

Diddi, S., Yan, R.-N., Bloodhart, B., Bajtelsmit, V., & McShane, K. (2019). Exploring young adult consumers' sustainable clothing consumption intention-behavior gap: A Behavioral Reasoning Theory perspective. Sustainable Production and Consumption, 18, 200–209. https://doi.org/10.1016/j.spc.2019.02.009

D'Souza, C., Taghian, M., Lamb, P., & Peretiatko, R. (2007). Green decisions: demographics and consumer understanding of environmental labels. International Journal of Consumer Studies, 31(4), 371–376. https://doi.org/10.1111/j.1470-6431.2006.00567.x

Eberhart, A. K., & Naderer, G. (2017). Quantitative and qualitative insights into consumers' sustainable purchasing behaviour: a segmentation approach based on motives and heuristic cues. Journal of Marketing Management, 33(13-14), 1149–1169. https://doi.org/10.1080/0267257X.2017.1371204

Eisler, A. D., & Eisler, H. (1994). Subjective Time Scaling: Influence of Age, Gender, and Type A and Type B Behavior. Chronobiologia, 21(3-4), 185–200.

Elkington, J. (1997). Cannibals with Forks: The Triple Bottom Line of 21st Century Business. Capstone.

Ellen, P. S. (1994). Do We Know What We Need to Know? Objective and Subjective Knowledge Effects on Pro-Ecological Behaviors. Journal of Business Research, 30, 43–52.

Gam, H. J. (2011). Are fashion-conscious consumers more likely to adopt eco-friendly clothing? Journal of Fashion Marketing and Management: An International Journal, 15(2), 178–193. https://doi.org/10.1108/13612021111132627

Geissinger, A., Laurell, C., Öberg, C., & Sandström, C. (2019). How sustainable is the sharing economy? On the sustainability connotations of sharing economy platforms. Journal of Cleaner Production, 206, 419–429. https://doi.org/10.1016/j.jclepro.2018.09.196

Goworek, H., Fisher, T., Cooper, T., Woodward, S., & Hiller, A. (2012). The sustainable clothing market: an evaluation of potential strategies for UK retailers. International Journal of Retail & Distribution Management, 40(12), 935–955. https://doi.org/10.1108/09590551211274937

Goworek, H., Oxborrow, L., Claxton, S., McLaren, A., Cooper, T., & Hill, H. (2018). Managing sustainability in the fashion business: Challenges in product development for clothing longevity in the UK. Journal of Business Research. Advance online publication. https://doi.org/10.1016/j.jbusres.2018.07.021

Guagnano, G. A., Stern, P. C., & Dietz, T. (1995). Influences on Attitude-Behavior Relationships. Environment and Behavior, 27(5), 699–718.

Harris, F., Roby, H., & Dibb, S. (2016). Sustainable clothing: challenges, barriers and interventions for encouraging more sustainable consumer behaviour. International Journal of Consumer Studies, 40(3), 309–318. https://doi.org/10.1111/ijcs.12257

Hauser, J. R., & Clausing D. (1988). The House of Quality. Harvard Business Review(May-June).

Hill, J., & Lee, H.-H. (2012). Young Generation Y consumers' perceptions of sustainability in the apparel industry. Journal of Fashion Marketing and Management: An International Journal, 16(4), 477–491. https://doi.org/10.1108/13612021211265863

Hiller Connell, K. Y. (2010). Internal and external barriers to eco-conscious apparel acquisition. International Journal of Consumer Studies, 34(3), 279–286. https://doi.org/10.1111/j.1470-6431.2010.00865.x

Hirscher, A.-L., Niinimäki, K., & Joyner Armstrong, C. M. (2018). Social manufacturing in the fashion sector: New value creation through alternative design strategies? Journal of Cleaner Production, 172, 4544–4554. https://doi.org/10.1016/j.jclepro.2017.11.020

Hu, Z.-H., Li, Q., Chen, X.-J., & Wang, Y.-F. (2014). Sustainable Rent-Based Closed-Loop Supply Chain for Fashion Products. Sustainability, 6(10), 7063–7088. https://doi.org/10.3390/su6107063

Hughner, R. S., McDonagh, P., Prothero, A., Shultz, C. J., & Stanton, J. (2007). Who are organic food consumers? A compilation and review of why people purchase organic food. Journal of Consumer Behaviour, 6(2-3), 94–110. https://doi.org/10.1002/cb.210

Hustvedt, G., & Dickson, M. A. (2009). Consumer likelihood of purchasing organic cotton apparel. Journal of Fashion Marketing and Management: An International Journal, 13(1), 49–65. https://doi.org/10.1108/13612020910939879

Ingaldi, M., & Ulewicz, R. (2019). How to Make E-Commerce More Successful by Use of Kano's Model to Assess Customer Satisfaction in Terms of Sustainable Development. Sustainability, 11(18), 4830. https://doi.org/10.3390/su11184830

Iran, S., Geiger, S. M., & Schrader, U. (2019). Collaborative fashion consumption – A cross-cultural study between Tehran and Berlin. Journal of Cleaner Production, 212, 313–323. https://doi.org/10.1016/j.jclepro.2018.11.163

Iwanow, H., McEachern, M. G., & Jeffrey, A. (2005). The influence of ethical trading policies on consumer apparel purchase decisions. International Journal of Retail & Distribution Management, 33(5), 371–387. https://doi.org/10.1108/09590550510596740

Jacoby, J., Berning, C. K., & Dietvorst, T. F. (1977). What about Disposition? Journal of Marketing, 41(2), 22–28.

Joergens, C. (2006). Ethical fashion: myth or future trend? Journal of Fashion Marketing and Management: An International Journal, 10(3), 360–371. https://doi.org/10.1108/13612020610679321

Kano, N., Seraku, N., Takahashi, F., & Tsuji S. (1984). Attractive quality and must-be quality. Journal of the Japanese Society for Quality Control, 14(2), 39–48.

Kim, H.-S., & Damhorst, M. L. (1998). Environmental Concern and Apparel Consumption. Clothing and Textiles Research Journal, 16(3), 126–133.

Kim, J., Kang, S., & Lee, K. H. (2018). How social capital impacts the purchase intention of sustainable fashion products. Journal of Business Research, 1–8. https://doi.org/10.1016/j.jbusres.2018.10.010

Kollmuss, A., & Agyeman, J. (2002). Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? Environmental Education Research, 8(3), 239–260. https://doi.org/10.1080/13504620220145401

Laitala, K. (2014). Consumers' clothing disposal behaviour - a synthesis of research results. International Journal of Consumer Studies, 38(5), 444–457. https://doi.org/10.1111/ijcs.12088

Laitala, K., & Boks, C. (2012). Sustainable clothing design: use matters. Journal of Design Research, 10(1/2), 121–139.

Laitala, K., Boks, C., & Klepp, I. G. (2011). Potential for environmental improvements in laundering. International Journal of Consumer Studies, 35(2), 254–264. https://doi.org/10.1111/j.1470-6431.2010.00968.x

Laitala, K., Klepp, I. G., & Boks, C. (2012). Changing laundry habits in Norway. International Journal of Consumer Studies, 36(2), 228–237. https://doi.org/10.1111/j.1470-6431.2011.01081.x

Lang, C., Armstrong, C. M., & Brannon, L. A. (2013). Drivers of clothing disposal in the US: An exploration of the role of personal attributes and behaviours in frequent disposal. International Journal of Consumer Studies, 37(6), 706–714. https://doi.org/10.1111/ijcs.12060

Lee, J. A., & Holden, S. J. S. (1999). Understanding the determinants of environmentally conscious behavior. Psychology & Marketing, 16(5), 373–392.

Luchs, M. G., & Mooradian, T. A. (2012). Sex, Personality, and Sustainable Consumer Behaviour: Elucidating the Gender Effect. Journal of Consumer Policy, 35(1), 127–144. https://doi.org/10.1007/s10603-011-9179-0

Lundblad, L., & Davies, I. A. (2016). The values and motivations behing sustainable fashion consumption. Journal of Consumer Behaviour, 15, 149–162. https://doi.org/10.1002/cb.1559

Maldini, I., Stappers, P. J., Gimeno-Martinez, J. C., & Daanen, H. A. (2019). Assessing the impact of design strategies on clothing lifetimes, usage and volumes: The case of product personalisation. Journal of Cleaner Production, 210, 1414–1424. https://doi.org/10.1016/j.jclepro.2018.11.056

Matzler, K., Hinterhuber, H. H., Bailom, F., & Sauerwein, E. (1996). How to delight your customers. Journal of Product & Brand Management, 5(2), 6–18.

McNeill, L., & Moore, R. (2015). Sustainable fashion consumption and the fast fashion conundrum: fashionable consumers and attitudes to sustainability in clothing choice. International Journal of Consumer Studies, 39(3), 212–222. https://doi.org/10.1111/ijcs.12169

Meinhold, J. L., & Malkus, A. J. (2005). Adolescent Environmental Behaviors. Environment and Behavior, 37(4), 511–532. https://doi.org/10.1177/0013916504269665

Mikulić, J., & Prebežac, D. (2011). Which products and services can be used to obtain a high level of customer satisfaction. Managing Service Quality: An International Journal, 21(1), 46–66.

Morgan, L. R., & Birtwistle, G. (2009). An investigation of young fashion consumers’ disposal habits. International Journal of Consumer Studies, 33(2), 190–198. https://doi.org/10.1111/j.1470-6431.2009.00756.x

Nagurney, A., & Yu, M. (2012). Sustainable fashion supply chain management under oligopolistic competition and brand differentiation. International Journal of Production Economics, 135(2), 532–540.

Nam, C., Dong, H., & Lee, Y.-A. (2017). Factors influencing consumers’ purchase intention of green sportswear. Fashion and Textiles, 4(1), 179. https://doi.org/10.1186/s40691-017-0091-3

Niinimäki, K. (2010). Eco-clothing, consumer identity and ideology. Sustainable Development, 18(3), 150–162. https://doi.org/10.1002/sd.455

Niinimäki, K., & Hassi, L. (2011). Emerging design strategies in sustainable production and consumption of textiles and clothing. Journal of Cleaner Production, 19, 1876–1883. https://doi.org/10.1016/j.jclepro.2011.04.020

Nilsson-Witell, L., & Fundin, A. (2005). Dynamics of service attributes: a test of Kano's theory of attractive quality. International Journal of Service Industry Management, 16(2), 152–168. https://doi.org/10.1108/09564230510592289

Pookulangara, S., & Shephard, A. (2013). Slow fashion movement: Understanding consumer perceptions—An exploratory study. Journal of Retailing and Consumer Services, 20(2), 200–206. https://doi.org/10.1016/j.jretconser.2012.12.002

Rese, A., Schlee, T., & Baier, D. (2019). The Need for Services and Technologies in Physical Fast Fashion Stores: Generation Y's Opinion. Journal of Marketing Management, 35(15-16), 1437–1459.

Retamal, M. (2019). Collaborative consumption practices in Southeast Asian cities: Prospects for growth and sustainability. Journal of Cleaner Production, 222, 143–152. https://doi.org/10.1016/j.jclepro.2019.02.267

Roberts, J. A. (1996). Green Consumers in the 1990s: Profile and Implications for Advertising. Journal of Business Research, 36, 217–231.

Schahn, J., & Holzer, E. (1990). Studies of Individual Environmental Concern. Environment and Behavior, 22(6), 767–786.

Shahin, A., Pourhamidi, M., Antony, J., & Hyun Park, S. (2013). Typology of Kano models: a critical review of literature and proposition of a revised model. International Journal of Quality & Reliability Management, 30(3), 341–358. https://doi.org/10.1108/02656711311299863

Shahin, A., & Zairi, M. (2009). Kano model: A dynamic approach for classifying and prioritising requirements of airline travellers with three case studies on international airlines. Total Quality Management & Business Excellence, 20(9), 1003–1028. https://doi.org/10.1080/14783360903181867

Shaw, D., Hogg, G., Wilson, E., Shiu, E., & Hassan, L. (2006). Fashion victim: the impact of fair trade concerns on clothing choice. Journal of Strategic Marketing, 14(4), 427–440.

Shen, B., Wang, Y., Lo, C. K., & Shum, M. (2012). The impact of ethical fashion on consumer purchase behavior. Journal of Fashion Marketing and Management: An International Journal, 16(2), 234–245. https://doi.org/10.1108/13612021211222842

Sheth, J. N., Sethia, N. K., & Srinivas, S. (2011). Mindful consumption: a customer-centric approach to sustainability. Journal of the Academy of Marketing Science, 39(1), 21–39. https://doi.org/10.1007/s11747-010-0216-3

Shim, S. (1995). Environmentalism and Consumers' Clothing Disposal Patterns: An Exploratory Study. Clothing and Textiles Research Journal, 13(1), 38–48.

Song, S., & Ko, E. (2017). Perceptions, attitudes, and behaviors toward sustainable fashion: Application of Q and Q-R methodologies. International Journal of Consumer Studies, 41(3), 264–273. https://doi.org/10.1111/ijcs.12335

Stern, P. C., & Oskamp, S. (1987). Managing scarce environmental resources. In D. Stokols & I. Altman (Eds.), Handbook of Environmental Psychology (pp. 1044–1088). Wiley.

Sun, M., & Trudel, R. (2017). The Effect of Recycling versus Trashing on Consumption: Theory and Experimental Evidence. Journal of Marketing Research, 54(2), 293–305. https://doi.org/10.1509/jmr.15.0574

Trudel, R., Argo, J. J., & Meng, M. D. (2016). The Recycled Self: Consumers' Disposal Decisions of Identity-Linked Products. Journal of Consumer Research, 43(2), 246–264. https://doi.org/10.1093/jcr/ucw014

Trudel, R., & Cotte, J. (2009). Does It Pay To Be Good? MIT Sloan Management Review, 50(2), 61–68.

Wang, T., Mukhopadhyay, A., & Patrick, V. M. (2017). Getting Consumers to Recycle NOW ! When and Why Cuteness Appeals Influence Prosocial and Sustainable Behavior. Journal of Public Policy & Marketing, 36(2), 269–283. https://doi.org/10.1509/jppm.16.089

White, K., Habib, R., & Hardisty, D. J. (2019). How to SHIFT Consumer Behaviors to be More Sustainable: A Literature Review and Guiding Framework. Journal of Marketing, 83(3), 22–49. https://doi.org/10.1177/0022242919825649

Winterich, K. P., Nenkov, G. Y., & Gonzales, G. E. (2019). Knowing What It Makes: How Product Transformation Salience Increases Recycling. Journal of Marketing, 83(4), 21–37. https://doi.org/10.1177/0022242919842167

Young, W., Hwang, K., McDonald, S., & Oates, C. J. (2010). Sustainable consumption: green consumer behaviour when purchasing products. Sustainable Development, 18, 20-31. https://doi.org/10.1002/sd.394

Zamani, B., Sandin, G., & Peters, G. M. (2017). Life cycle assessment of clothing libraries: can collaborative consumption reduce the environmental impact of fast fashion? Journal of Cleaner Production, 162, 1368–1375. https://doi.org/10.1016/j.jclepro.2017.06.128

2.2.3 Research Paper No. 3: Does sustainability really matter to consumers? Assessing the importance of online shop and apparel product attributes

Authors: Rausch, T. M., Baier, D., & Wening, S.

Submitted in: Journal of Retailing and Consumer Services

Abstract: Although there is a shift in consumers' consumption behavior towards more sustainable patterns across a variety of different contexts, sustainable apparel has still not become a mainstream trend despite the textile industry's excessive usage of valuable resources. Albeit extant research found different potential barriers elucidating why consumers hesitate to purchase such apparel, it remains unclear whether sustainability really matters to consumers in a clothing context and further, which aspects are of relevance during consumers' purchase decision. We thus conducted two studies with four best-worst scaling experiments in which 4,350 online shoppers assessed the importance of both conventional and sustainable apparel attributes, as well as sustainable apparel attributes only, and the willingness to pay for sustainable product attributes. We further inquired the importance of conventional as well as sustainable online shop attributes. Our findings indicate that conventional apparel attributes such as fit and comfort, price-performance ratio, and quality are of higher relevance to consumers than sustainable attributes. The most important sustainable apparel attributes are the garment's durability, fair wages and working conditions, as well as an environmentally friendly production process. Consumers also indicated to prefer the latter three attributes to a 20% discount. Moreover, consumers demand less as well as sustainable packaging, free returns, and discount campaigns. Our findings reveal a gender gap regarding green consumerism with female respondents assessing most sustainable attributes as more important than male respondents do.

Keywords: sustainable apparel; online shopping; willingness to pay; best-worst scaling; e-commerce

1 Introduction

Across a wide variety of industries, sustainability developed from a fringe to a mainstream issue for manufacturers, managers, marketers, and further stakeholders throughout the past years. The shift in consumers' mindsets and increasing awareness regarding their environmental impact triggered research to gather a better understanding of green purchase behavior (Kautish et al., 2019; Maichum et al., 2016; Paul et al., 2016; Taufique & Vaithianathan, 2018; Yadav & Pathak, 2016, 2017).

Particularly the textile industry can be considered a black sheep, and its environmental harm is frequently underestimated by consumers: it caused approximately 2.1 billion tons of greenhouse gas emissions in 2018 and thus, accounted for around 4% of the annual emissions globally, which is equivalent to the combined annual emissions of France, Germany, and United Kingdom (McKinsey & Company & Global Fashion Agenda, 2020). Nevertheless, sustainable apparel consumption is still marginal: Particularly consumers older than 35 years were found to have a neutral rather than a supportive attitude towards sustainable fashion (KPMG, 2019). Apparently, in the clothing context, a vast majority of consumers struggles to translate its overall green attitude into green actions (Kollmuss & Agyeman, 2002; Young et al., 2010). Albeit research found single aspects such as perceived higher prices (Hiller Connell, 2010; Joergens, 2006) and perceived aesthetic risk (Hiller Connell, 2010; Rausch & Kopplin, 2021) to deter consumers from buying sustainable clothing, literature still lacks a comprehensive understanding regarding the importance of sustainable apparel attributes and, further, their importance in comparison with conventional apparel attributes to successfully target the most crucial aspects. It remains unclear whether sustainability really matters to consumers in a clothing context and, more specifically, which aspects are of importance to consumers during their purchase decision. Moreover, as sustainable apparel is frequently perceived as more expensive than conventional apparel (Hiller Connell, 2010), research on consumers' willingness to pay for sustainable product attributes is needed as initial attempts mainly cover surcharges for materials (Ellis et al., 2012; Ha-Brookshire & Norum, 2011).

We thus contribute to literature by conducting two studies with four different best-worst scaling (BWS) experiments: We gathered data of 4,350 online shoppers assessing the importance of both conventional and sustainable apparel attributes as well as sustainable apparel attributes only and further, the willingness to pay for sustainable product attributes. As the textile industry yields the highest online sales across all e-businesses (Statista, 2020), we also asked the respondents to assess the importance of conventional as well as sustainable online shop attributes. To gain deeper insights, we further compare the results of male and female respondents.

The remainder of this paper is structured as follows: Section 2 reviews related literature and Section 3 outlines the theoretical basics of the BWS methodology. We then present our methodology, followed the conceptualization and results of our four BWS experiments. Lastly, we discuss our findings in the light of theoretical contribution, practical implications, and limitations.

2 Theoretical background

2.1 Apparel attributes affecting consumers' purchase decision

Sustainable clothing implies pro-environmental aspects throughout the whole lifecycle of a garment, from the pre-purchase, purchase, to the post-purchase phase (Jacoby et al., 1977; Lundblad & Davies, 2016; Morgan & Birtwistle, 2009). Albeit sustainable clothing is mostly associated with logistics or material usage and the respective environmental impact, there is no industry standard uniformly regulating the notion (Joergens, 2006; Lundblad & Davies, 2016). Within literature, the concept of sustainable clothing extends well beyond environmental aspects (e.g., usage of environmentally friendly materials) by further comprising social aspects (e.g., working conditions of employees) (Fulton & Lee, 2013; Goworek et al., 2012; Lundblad & Davies, 2016).

Throughout the early stages of the apparel lifecycle, employee wages, working conditions, amount of pesticide usage, material usage, and the country-of-origin are among the key aspects determining the product's sustainability (Goworek et al., 2012; Hustvedt & Dickson, 2009; Lundblad & Davies, 2016): Intuitively, sustainable clothing is frequently referred to recycled/upcycled or bio-based materials and locally manufactured products by most consumers (Allwood et al., 2008; Morgan & Birtwistle, 2009; Scherer et al., 2018), but good working conditions and fair wages (Fulton & Lee, 2013; Stöckigt et al., 2018) may also be an important sustainability facet. Thereby, increasing information complexity surrounding sustainable products and, as a result, higher search costs for consumers were frequently identified to be among the main barriers for sustainable clothing consumption (Ellen, 1994; Harris et al., 2016). To shift consumption behaviors towards more sustainable ones, prior literature thus recommended providing better information and education about materials used or manufacturing conditions, e.g., with eco or social labels, to consumers (D'Souza et al., 2007; Hiller Connell, 2010).

Considering the later stages of the garment's lifecycle, and specifically, the product's discard, it may be of importance to consumers during their purchase decision whether the material is recyclable or biodegradable to close the garment's lifecycle (Fulton & Lee, 2013). Further, companies offering take-back programs to recycle the garment were found to appear attractive to consumers (Baier et al., 2020).

Overall, female consumers were frequently found to be more likely to engage in sustainable behaviors than male consumers (Eisler & Eisler, 1994; Luchs & Mooradian, 2012) as women tend to be more altruistic as well as prosocial (Lee & Holden, 1999). Men associate greenness with femininity and thus, avoid sustainable behaviors to maintain their gender-identity (Brough et al., 2016).

Despite these initial exploratory findings on potentially essential sustainable apparel attributes, there is little knowledge on the importance of these different attributes with regard to consumers' clothing purchase behavior. Extant research primarily investigated the importance of conventional intrinsic as well as extrinsic attributes for consumers' purchase decision of apparel products: There is vast consensus within literature that aesthetic criteria with regard to the garment's physical appearance such as color, pattern, and design have the greatest impact on consumers' purchase decision (Abraham-Murali &

Littrell, 1995; Baier et al., 2020; Fiore & Damhorst, 1992; Zhang et al., 2002). Visual characteristics are assumed to fulfill implicit expectations such as fashionability, aesthetic appeal, and self-expression (Eckman et al., 1990). This seems particularly critical with regard to sustainable clothing consumption, as consumers frequently perceive sustainable apparel as unfashionable (Hiller Connell, 2010; Joergens, 2006), deterring consumers from purchasing despite initial purchase intention (Rausch & Kopplin, 2021). Sustainable clothing is mostly associated with specific stereotypes inhibiting mainstream consumption (Connolly & Prothero, 2003).

Aside from aesthetic criteria, the garment's quality and physical performance is another essential purchase criterion, i.e., attributes which are instrumental outcomes of the product's physical aspects (Abraham-Murali & Littrell, 1995; Zhang et al., 2002). Physical performance comprises – among others – aspects concerning the overall fit, comfort, the extent to which the garment can be maintained in a wearable condition during care, and workmanship (i.e., the level of excellence regarding construction as well as materials) (Eckman et al., 1990; Zhang et al., 2002).

Despite the striking importance of aesthetic and performance criteria, research found extrinsic criteria to exhibit a comparable impact on consumers' purchase decision: Price and perceived value for money are often among the most decisive purchase criteria for most consumers (Abraham-Murali & Littrell, 1995; Jegethesan et al., 2012; Zhang et al., 2002). Moreover, the garment's brand and, accordingly, brand familiarity are crucial to some consumers (Abraham-Murali & Littrell, 1995; Eckman et al., 1990; Jegethesan et al., 2012). Although it is of inferior importance than the garment's price, it still plays an important role during the purchase decision process (Abraham-Murali & Littrell, 1995; Zhang et al., 2002).

Research comparing the importance of both sustainable and conventional product attributes is sparse and mostly focused on materials as a sustainable product attribute: Although research indicated that environmentally friendly materials as a sustainable product attribute may influence consumers' purchase decision, conventional product attributes were found to be more important (Viciunaite & Alfnes, 2020).

2.2 Willingness to pay for sustainable clothing attributes

Consumers often exhibit an attitude-behavior gap regarding sustainable clothing, i.e., albeit they pretend environmentally friendly attitudes, they frequently struggle to translate this into green actions (Kollmuss & Agyeman, 2002; Young et al., 2010). One of the main barriers towards buying sustainable clothes is the perceived increased economic risk associated with such clothes: Preceding research found consumers to consider the price of sustainable apparel as higher compared to conventional clothes, and thus, they do not purchase such clothes despite their initial positive attitude (Hiller Connell, 2010; Hustvedt & Dickson, 2009; Joergens, 2006). Since sustainable clothes are hardly produced for the mass market, they often carry higher prices, and thus, are perceived as less affordable compared to conventional clothes (Hiller Connell, 2010). However, extant research determining consumers' willingness to pay for different sustainable clothing attributes and thus, trying to overcome the attitude-behavior gap is still sparse.

Mostly, research investigated consumers' willingness to pay for materials: (Ellis et al., 2012) found consumers to pay 25% more for t-shirts made of organic cotton, aligning with (Ha-Brookshire & Norum, 2011) findings that consumers are willing to pay a higher price for domestic-grown organic cotton shirts. Similarly, (Hustvedt & Bernard, 2008) found consumers to be willing to pay a premium for socks labeled as organic and produced locally. Similar findings were gathered in the food context, where research found a higher willingness to pay for local and organic food (de-Magistris & Gracia, 2016; Gil et al., 2000).

Aside from materials and country-of-origin, consumers were found to pay a premium in case the product had a labor-related labeling indicating social responsibility and fair trade (Hustvedt & Bernard, 2010). An increased willingness to pay for (e.g., fair trade or eco) labels was further found to be applicable in a food context (Delmas & Lessem, 2017; Paetz & Guhl, 2017; van Loo et al., 2015).

2.3 Online shop attributes affecting consumers' purchase decision

Compared to brick-and-mortar stores, e-commerce businesses are confronted with several issues inherent to the online context. As online transactions are frequently perceived as risky (particularly in terms of privacy as well as data security or product performance) and were associated with time as well as convenience loss (Forsythe & Shi, 2003; Miyazaki & Fernandez, 2001), it is crucial for e-businesses to assess the importance of different attributes regarding the online store itself to mitigate consumers' doubts.

Considering online store attributes which may be of importance to consumers during the purchase decision, research found particularly factors associated with shipping such as shipping fees, shipping speed, and return policy to be essential decision criteria (Bower & Maxham, 2012; Cao et al., 2018; Lewis, 2006; Ma, 2017; Oghazi et al., 2018; Stöckigt et al., 2018). E.g., (Smith & Brynjolfsson, 2001) found consumers to be approximately twice as sensitive to changes in shipping fees as they are to changes in the product's prices. Further, prior positive return experiences were found to enhance the customer's future buying behavior and thus, his or her lifetime value for the organization (Petersen & Kumar, 2009; Wood, 2001). Despite the striking importance of delivery speed and costs, the environmental impact of the shipping procedure can potentially influence consumers' purchase decision (Stöckigt et al., 2018).

Particularly in an e-commerce context, the availability and channels to contact the customer service seem to play an important role (Cao et al., 2018). It is well known that perceived service quality affects customer loyalty (Parasuraman et al., 1985; Zeithaml et al., 1996). This relation may be even more critical in an online context as consumers do not have the chance to interact face-to-face with agents like in brick-and-mortar stores and hence, may assess service quality as more intangible.

In a sustainability context, further online store attributes may be of relevance: E.g., consumers' demand for less packaging waste and/or sustainable packaging alternatives are steadily increasing and has been a fruitful path to investigate within literature throughout the past years (Nguyen et al., 2020; Prakash & Pathak, 2017; Schwepker & Cornwell, 1991). Thereby, purchase intention for eco-friendly packaging was particularly found to be determined by altruistic motives, i.e., consumers' environmental concerns (Prakash et al., 2019; Prakash & Pathak, 2017).

Generally, consumers value a large range of sustainable products as one of the main barriers towards sustainable clothing consumption is the lack of availability: Sustainability in a clothing context frequently implies fewer fashion cycles as well as collections, few different styles (e.g., in terms of color), and a smaller product range (Hiller Connell, 2010; Lundblad & Davies, 2016; Pookulangara & Shephard, 2013) as it is restricted to natural materials.

Overall, albeit the importance of single aspects has been indicated by extant research, literature still lacks a holistic comparison and evaluation of the importance of both conventional as well as sustainable store attributes.

3 Best-Worst Scaling

BWS is a rather new measurement approach for the subjective value (or importance, utility) of items (objects or attribute-levels for objects) that is based on the random utility framework by (Thurstone, 1927) and (McFadden, 1974). In this framework, it is assumed that individuals evaluate stimuli (objects, attribute-levels, or attribute-level-combinations) on a subjective utility scale and that these evaluations form their basis for choosing the 'best' stimulus (with maximum utility) among presented ones. These assumptions allow deriving unknown subjective values of items from observed choice frequencies among stimuli (Louviere et al., 2013). The individual evaluation might be superimposed by an additive random error, but when distributional assumptions are justified and individuals choose repeatedly, means and variances of these subjective values can be estimated (Louviere et al., 2015). On the basis of this framework, Jordan J. Louviere developed BWS in 1987 and applied the new approach together with Adam Finn for the first time (Finn & Louviere, 1992; Louviere & Woodworth, 1991).

BWS is an alternative to rating, ranking, and to other choice-based procedures that are common for measuring subjective values (Mühlbacher et al., 2013): For BWS, the basic idea is to repeatedly confront respondents with 'choice sets' of stimuli from which they have to select the 'best' and the 'worst' one (Cohen, 2003). In each of these choice tasks, the respondents are assumed to compare and evaluate all stimuli and select those two that they perceive to provide them the greatest and the least individual benefit. During this choice task, various cognitive processes have to take place: Respondents identify and evaluate all possible stimulus pairs. The utility differences between all pairs are calculated and the pair is selected with maximum difference (Flynn et al., 2007). This stochastic selection process can be expressed formally by the following so-called MaxDiff formulation:

$$\text{Prob}\left((j_1;j_2)\,|\,CS_{ik};i\right)=\text{Prob}\left((u_{ij_1}+\varepsilon_{ij_1})-(u_{ij_2}-\varepsilon_{ij_2})\geq\max_{j,j'\in CS_{ik},j\neq j'}\left((u_{ij}+\varepsilon_{ij})-(u_{ij'}+\varepsilon_{ij'})\right)\right)$$

$\text{Prob}\left((j_1;j_2)\,|\,CS_{ik};i\right)$ corresponds to the probability with which respondent i (i=1,…,I) assesses stimulus $j_1 \in CS_{ik}$ as the best and stimulus $j_2 \in CS_{ik}$ as the worst when the choice set CS_{ik} is presented (k=1,…,K). The expression $(u_{ij_1}+\varepsilon_{ij_1})-(u_{ij_2}-\varepsilon_{ij_2})$ represents the difference between the evaluations of stimulus j_1 and stimulus j_2 on the underlying scale. u_{ij} is the mean subjective value respondent i allocates to stimulus j and ε_{ij} is the additive random error. The MaxDiff expression $\max_{j,j'\in CS_{ik},j\neq j'}\left((u_{ij}+\varepsilon_{ij})-(u_{ij'}+\varepsilon_{ij'})\right)$ ultimately contains the greatest difference of all possible paired differences in CS_{ik} (cf. (Finn & Louviere, 1992). The pairwise comparisons within the choice sets force the respondent to make trade-off decisions between different alternatives (Lee et al., 2007). As a result, the respondent can neither accept all possible answers, evaluate them as good, reject them, nor classify them as bad. Nevertheless, the search for a pair with greatest difference in subjective values is closely related to the search for the 'best' stimulus with the highest and the 'worst' stimulus with the lowest subjective value in the set CS_{ik}. Ultimately, it is possible to estimate means and variances of the subjective values from repeated choice tasks (Cohen, 2003).

According to (Flynn & Marley, 2014) a distinction can be made between three cases of measurement, i.e., (1) object scaling (Case 1), (2) profile scaling (Case 2), and (3) multiprofile scaling (Case 3). Object scaling describes the basic archetype of BWS as discussed above. There is a (small) set of objects (say 6 to 30) and the presented choice sets are subsets of this set (Mühlbacher et al., 2013). The respondents have to select the best and the worst stimulus/object in each subset (Finn & Louviere, 1992). Further, if an i.i.d. exponential distribution for the additive random disturbance is assumed, we get a multinomial logit (MNL) formulation for the probability $\text{Prob}\left((j;.)\,|\,CS_{ik};i\right)$ that stimulus j is selected as the best (or most important) one by individual i in choice set CS_{ik}:

$$\text{Prob}\left((j;.)\,|\,CS_{ik};i\right)=\frac{\exp(u_{ij})}{\sum_{j'\in CS_{ik}}\exp(u_{ij'})}$$

Finn & Louviere, (1992) argue that the probability $\text{Prob}\left((.;j)\,|\,CS_{ik};i\right)$ that stimulus j is selected as the worst (or least important) one in choice set CS_{ik} can be formulated similarly:

$$\text{Prob}\left((.;j)\,|\,CS_{ik};i\right)=\frac{\exp(-u_{ij})}{\sum_{j'\in CS_{ik}}\exp(-u_{ij'})}$$

For object scaling with I individuals, J objects, K choice tasks and choice sets $CS_{ik} \subset \{1,...,J\}\ \forall\, i \in \{1,...,I\}, j \in \{1,...,J\}$, the observed selections of best and worst stimuli/objects across all choice sets and respondents can be used to derive a likelihood function across all observations and the model parameters (the unknown mean subjective values of the objects $u_{ij}\ \forall\, i \in \{1,...,I\}, j \in \{1,...,J\}$) are estimated using a Maximum Likelihood approach. The choice sets may vary across respondents, but (best possible) fulfillment of criteria like balance (objects appear with the same frequency aross all choice sets) and orthogonality (object pairs appear with the same frequency across all choice sets) support the efficient and uncorrelated estimation of the model parameters. It needs to be mentioned that the underlying MNL assumption of i.i.d. errors across the selections of best and of worst stimuli has been questioned by some authors since it may lead to biased estimates of subjective values (e.g., Dyachenko et al., 2013). However, Horne and Rayner (2013) showed in several analyses that this bias should not justify more advanced estimation procedures that additionally model ordering issues (as proposed, e.g., by Dyachenko et al., 2013)).

Instead, particularly Hierarchical Bayes (HB) with underlying MNL assumption is wide-spread for estimating the subjective values (Sawtooth Software, 2009a, 2013). The advantage of this approach is the sharing of observations across respondents by assuming a higher (aggregate) level model across all respondents besides the lower (individual) level model. This assumption allows reducing the number of necessary observations (choice sets) per respondent dramatically. Software packages like MaxDiff from Sawtooth Software (2009a) provide easy access to BWS and to HB estimation. When the same data is analyzed with Maximum Likelihood and HB, both approaches lead to very similar aggregate results (mean subjective values across all respondents), as Cheung et al. (2018) have demonstrated. According to this comparison, even the Average BW Score (Louviere & Flynn, 2010; Mühlbacher et al., 2016) is useful to derive valid proxies for the mean subjective values:

$$\text{Average BW Score}_j = \frac{n_j^{\text{Best}} - n_j^{\text{Worst}}}{n_j}$$

This Average BW Score for object j is calculated by relating the number of times j was selected as best stimulus (n_j^{Best}) minus the number of times object j was selected as worst stimulus (n_j^{Worst}) to the total number of times object j was presented (n_j). However, the advanced HB technique additionally allows deriving subjective values at the individual level – even with few observations per respondent – and thus, can be used to discuss segment-specific results by averaging them cross a priori defined groups of respondents or by applying clustering approaches. Before averaging or clustering, the estimated mean subjective values $u_{ij}\ \forall\, i \in \{1,...,I\}, j \in \{1,...,J\}$ have to be zero-centered so that $\sum_{j=1}^{J} u_{ij} = 0\ \forall\, i \in \{1,...,I\}$ holds. This transformation has no effect on the above individual choice probabilities. In some cases, the zero-centered mean subjective values are then further transformed into probabilities

$$p_{ij} = \frac{\exp(u_{ij})}{\exp(u_{ij}) + a - 1}$$

where a is the size of the choice sets (number of stimuli presented in a choice task). p_{ij} reflects the probability that object j is selected when presented together with $a - 1$ other stimuli j' with $u_{ij'} = 0$ (the average subjective value if zero-centered). In contrast to the subjective values u_{ij}, these ratio-scaled probabilities p_{ij} can be interpreted more easily (j has an x-times higher probability to be selected than j'). Further, these probabilities p_{ij} can additionally be normalized to Probability Scores so that they sum up to 100.

Profile Scaling – Case 2 – differs from object scaling in that stimuli are attribute-level-combinations (profiles). The respondents are presented one profile in a choice task and they have to select the best and worst attribute-level in the profile (Flynn & Marley, 2014). On the basis of this assessment, it is possible to determine the overall benefit by adding up the respective partworths for each attribute-level (Marley & Louviere, 2005). Finally, with multiprofile scaling, the respondents are presented more than one profile (attribute-level-combination) in a choice task and they have to select a best and a worst attribute-level-combination (Mühlbacher et al., 2013). In both cases, the subjective value of a stimulus/profile is assumed to be the sum of its attribute-level partworths. Again, using the above MNL formulation, Maximum Likelihood can be used for estimating means and variances of subjective values, but with the attribute-level partworths as model parameters (Sawtooth Software, 2009a, 2009b, 2013).

BWS has various advantages compared to conventional survey methods with rating scales. The crucial difference between BWS and traditional approaches is that BWS is an indirect approach where no direct evaluation of objects or attribute-levels is necessary, i.e., the risk of possible distortion of the results can be overcome (Auger et al., 2007). The following biases can thus be counteracted:

- Social desirability bias: Respondents tend to give answers that they think are correct and socially accepted (Fisher, 1993).
- Acquiescence bias: tendency of respondents to generally agree to questions regardless of their content (Schuman & Presser, 1981).
- Extreme response bias: the respondents' tendency to extreme responses (Culpepper & Zimmerman, 2006).

With BWS, these distortions can be avoided, and there is no distortion of mean values, which allows for valid comparisons (Cohen & Orme, 2004). In addition, significant differences in the response styles or in the use of rating scales between different countries have been empirically proven (Chen et al., 2016). Therefore, this method is frequently used in market research in a country comparison (Lee et al., 2007). Further advantages are ease of use for the respondents (Marley & Louviere, 2005) as well as pleasant implementations that allow a largely uncomplicated analysis with the help of standard software (Cohen, 2003). Finally, BWS leads to more precise results than standard rating scales (Cohen & Orme, 2004). This is the conclusion gathered by Chrzan and Golovashkina (2006), who examined customer satisfaction with restaurant services using various methods of comparative scaling. BWS turned out to be the model with the highest predictive validity. It is also best suited to differentiate between the importance of objects (Chrzan & Golovashkina, 2006). Further advantages that are particularly relevant for this work are the high level of forecast reliability, the achievement of reliable results, and the ability to identify a precise subjective value structure between the items (Mühlbacher et al., 2013). This is particularly important with regard to the determination of the heterogeneity of the individual evaluations with regard to the objects examined.

However, there are also weaknesses: Aside from an increased expenditure of time, there is a certain pressure to make a decision for the respondents. In addition, the respondents cannot express likes or dislikes of all items. However, this can be counteracted by adding a rejection option (Mühlbacher et al., 2013). Further disadvantages are the increased need for explanation of the method, the more complex data collection and analysis, and the availability of specific knowledge (Simon, 2010).

4 Methodology

We conducted our two studies in cooperation with BAUR, a large German online retailer for fashion and furniture: BAUR generated 361.8 million Euros online sales in 2018 (Statista, 2019), and thus, was among the top 10 fashion e-businesses in Germany. BAUR is primarily targeting women, which are between 40 and 55 years old.

To analyze the importance of the attributes, we developed two online questionnaires with Sawtooth Software's Lighthouse Studio 9.8.1. Before conducting the main study, the questionnaires were pre-tested with experienced participants and researchers to assess appropriateness, clarity, completeness, wording, and structure. Only minor amendments were made. The final questionnaires were structured identically and consisted of three major sections. To gather deeper insights into the participants' consumption behavior and perception of sustainable consumption, we developed five introductory questions. We inquired the respondents' general perception of sustainability (multiple selections with a maximum of three choices), their sustainable behavioral patterns in daily life (evaluation on a 6-point Likert type scale from '1 = Totally disagree' to '6 = Totally agree'), barriers towards making sustainable purchases (multiple selections), assessment of options to transparently communicate sustainability (evaluation on a 6-point Likert type scale from '1 = Very unimportant' to '6 = Very important'), and assessment of attributes indicating an organization's sustainability (multiple selections with a maximum of three choices). The main part comprised the respective BWS experiments (object scaling): Within Study 1, we inquired sustainable and conventional apparel attributes (Study 1a) as well as the willingness to pay for sustainable apparel attributes (Study 1b). Within Study 2, sustainable apparel attributes (Study

2a) and sustainable and conventional online shop attributes (Study 2b) were assessed. Lastly, we collected the respondents' demographics.

The two surveys were spread via the BAUR newsletter from July 10, 2020 to July 19, 2020. The surveys were mailed to an equal number of different BAUR customers (approximately 260,000 for each questionnaire), i.e., each customer could only participate in one survey. Among all participants, we gave away 15 vouchers á 15 Euros for the BAUR online shop.

The newsletter containing the first survey was opened by 51,578 customers. In total, 5,149 customers opened the first survey, and 2,244 responses were considered for further analysis. In line with the target customer segment of BAUR, our sample consisted of 1,770 (76.7%) females. Most respondents were between 50 and 54 years old (n=423, 18.3%) and 60 years or older (n=549, 23.8%). The majority of the respondents' household had a total net income between 1,501 and 2,000 Euros (n=287, 12.8%). Mostly, respondents associated sustainability with a decreased environmental impact (n=1,055, 47.0%) and durable as well as reparable products (n=1,054, 47.0%). Regarding their sustainable behavioral patterns in daily life, respondents mainly indicated to avoid plastic packaging when buying groceries (mean=4.95, standard deviation SD=1.27) and to purchase environmentally friendly cleaning products (mean=4.61, SD=1.45). The most frequently mentioned barriers towards sustainable consumption behavior were the lack of information (n=1,248, 55.6%) and high prices (n=1,139, 50.8%). To draw more attention towards sustainable products and to communicate sustainability transparently, respondents demand information about the product's environmental impact attached directly on the product (mean=4.70, SD=1.05) and discounts for sustainable products (mean=4.65, SD=1.10). Lastly, respondents indicated to assess an organization's sustainability level with regard to its manufacturing conditions (n=1,524, 67.9%) and use of pesticides as well as chemicals in its products (n=1,519, 67.7%). The descriptive statistics of Study 1 are depicted in Table A.1 in the Appendix.

The newsletter containing Study 2 was opened by 51,668 customers. In total, 5,065 customers opened the first survey, and 2,106 responses were considered for further analysis. The sample comprised 1,678 (79.7%) females. Mostly, respondents were between 50 and 54 years old (n=379, 18.0%) and 60 years or older (n=503, 23.9%). Most respondents' household had a total net income between 3,001 and 4000 Euros (n=241, 11.4%). Respondents mainly associated sustainability with a decreased environmental impact (n=1,030, 48.9%) and durable as well as reparable products (n=1,003, 47.6%). The majority of the respondents indicated to avoid plastic packaging when buying groceries (mean=4.84, SD=1.33) and to purchase environmentally friendly cleaning products (mean=4.56, SD=1.45) as common sustainable behavioral patterns in their daily life. Respondents stated the lack of information (n=1,107, 52.6%) and high prices (n=1,015, 48.2%) to be the main barriers towards sustainable consumption behavior. Further, respondents demand information about the product's environmental impact attached directly on the product (mean=4.62, SD=1.06) and discounts for sustainable products (mean=4.50, SD=1.12) to draw more attention towards sustainable products and to communicate sustainability transparently. Respondents assess an organization's sustainability level mainly with regard to its manufacturing conditions (n=1,532, 72.7%) and use of pesticides as well as chemicals in its products (n=1,385, 65.8%). Table A.2 outlines the descriptive statistics of Study 2.

5 Study 1a: Assessing sustainable and conventional apparel attributes

5.1 Conceptualization

As a first step, for both studies, we cooperated with several experts of BAUR to appropriately determine practically relevant attributes, which should then be assessed by the respondents. This procedure complements our literature findings from a practical perspective. Within the first BWS experiment of the first study's main part, we asked the respondents to evaluate both sustainable and conventional apparel attributes in the main part of the survey (see Table 1). Thus, there were ten items in the study and we displayed four items per choice set. We asked as many choice sets per respondent such that every item appeared at least three times per respondent. Respondents were asked to choose the most important and the least important attribute within each choice set while considering purchasing clothes.

Table 1: Sustainable and conventional apparel attributes used in Study 1a.

Attribute	Description	References
Brand	The garment's brand or reference to manufacturer	(Eckman et al., 1990; Jegethesan et al., 2012)
Design	The garment's appearance in terms of design and style	(Abraham-Murali & Littrell, 1995; Fiore & Damhorst, 1992)
Fit and comfort	How well the garment fits and conforms to the shape of the body	(Eckman et al., 1990; Zhang et al., 2002)
Number of customer reviews	The number of reviews written by other customers	(Park et al., 2007)
Price-performance ratio	Subjective evaluation about whether the price is appropriate for the perceived product performance (value for money)	(Abraham-Murali & Littrell, 1995; Zhang et al., 2002)
Quality	Level of excellence of the construction or material in the garment	(Abraham-Murali & Littrell, 1995; Zhang et al., 2002)
Quality of customer reviews	The quality of reviews written by other customers	(Park et al., 2007)
Bio-based materials[s]	Use of bio-based materials or fibres	(Baier et al., 2020; Fulton & Lee, 2013)
Fair wages and working conditions[s]	Manufactured under good working conditions; employees obtain appropriate wages	(Fulton & Lee, 2013; Goworek et al., 2012; Stöckigt et al., 2018)

Attribute	Description	References
Take-back program[s]	Garment can be returned to the retailer for disposal to prolong the life cycle; it will be recycled, re-used, or environmentally friendly disposed	(Baier et al., 2020)

Note: [s] = sustainable apparel attribute.

5.2 Results

For analysis, we applied HB estimation and ran 30,000 iterations (including 20,000 burn-in iterations). Complementary, we calculated average BW scores. Internal consistency is given, considering that the model's root likelihood (RLH) is 0.561, which is higher than the null model's RLH of 0.25.

Across the whole sample, fit and comfort was chosen the most important attribute, just before price-performance ratio. The quality and number of reviews are rather unimportant, as well as the garment's brand. The most important sustainable clothing attribute was social-facetted, i.e., fair wages and working conditions. Nevertheless, it was approximately only half as important as the garment's fit and comfort. Bio-based materials and the take-back program even had a negative average BW score. The results are reported in Table 2.

Overall, consumers apparently rather value intrinsic conventional apparel attributes than sustainable ones. Intrinsic attributes (i.e., fit and comfort, quality, and design) were found to be extremely important as well as the value consumers get for their money. Surprisingly, the conventional extrinsic attribute (brand) is of inferior relevance and only half as important as fair wages and manufacturing conditions (with respect to their probability scores).

Additionally, we performed a post-hoc Tukey HSD test (on the zero-centered utilities) to test the differences between the attributes. All attributes' utilities were found to be significantly different from each other, except for brand versus quality of customer reviews.

Table 2: Best-worst scaling results for Study 1a (n=2,244).

Attribute	Average BW score	Zero-centered utility	Probability score
Fit and comfort	0.621	3.314 (1.365)	22.631 (4.556)
Price-performance ratio	0.408	2.215 (1.458)	18.720 (6.746)
Quality	0.233	1.330 (1.544)	14.485 (7.109)
Design	0.114	0.729 (1.536)	11.765 (7.034)
Fair wages and working conditions[s]	0.087	0.426 (1.825)	10.515 (7.716)
Bio-based materials[s]	-0.132	-0.670 (1.850)	6.475 (6.475)

Attribute	Average BW score	Zero-centered utility	Probability score
Quality of customer reviews	-0.257	-1.502 (1.665)	4.074 (5.139)
Brand	-0.262	-1.460 (2.280)	5.231 (6.821)
Take-back program[s]	-0.324	-1.740 (2.025)	4.038 (5.644)
Number of customer reviews	-0.487	-2.642 (1.795)	2.066 (3.535)

Note: Standard deviations in parentheses, [s] = sustainable apparel attribute.

To gather further insights, we investigated the gender-specific attributes' importances (see Table 3). Apparently, male respondents are caring more about the garment's brand ($t = -6.997$, $p < 0.001$), and the opinion of other third parties in terms of number ($t = -8.429$, $p < 0.001$) and quality of customer reviews ($t = -5.193$, $p < 0.001$) compared to female respondents. Further, they value a take-back program more than female respondents ($t = -2.824$, $p = 0.005$). Apparently, male respondents generally care less about sustainable aspects, particularly when it comes to material usage and manufacturing conditions. In contrast, female respondents put more emphasis on fit and comfort when purchasing clothes ($t = 13.495$, $p < 0.001$), the garment's design and appearance ($t = 3.352$, $p = 0.001$), bio-based materials ($t = 6.008$, $p < 0.001$), and the social facet of sustainability, i.e., fair wages and working conditions ($t = 4.419$, $p = 0.003$).

Table 3: Gender-specific best-worst scaling results for Study 1a.

Attribute	Zero-centered utility		Probability score	
	Female (n=1,770)	Male (n=470)	Female (n=1,770)	Male (n=470)
Fit and comfort***	3.501 (1.341)	2.570 (1.288)	22.934 (4.209)	21.326 (5.375)
Price-performance ratio	2.212 (1.481)	2.186 (1.353)	18.447 (6.762)	19.655 (6.528)
Quality	1.360 (1.570)	1.230 (1.439)	14.426 (7.070)	14.728 (7.143)
Design**	0.788 (1.581)	0.521 (1.356)	11.950 (7.118)	11.104 (6.554)
Fair wages and working conditions[s] ***	0.511 (1.873)	0.094 (1.604)	10.785 (7.728)	9.583 (7.457)
Bio-based materials[s] ***	-0.539 (1.870)	-1.108 (1.634)	6.841 (6.617)	5.090 (5.495)
Quality of customer reviews***	-1.581 (1.687)	-1.134 (1.544)	3.852 (5.011)	5.015 (5.477)
Brand***	-1.652 (2.355)	-0.828 (1.904)	4.881 (6.699)	6.553 (6.985)
Take-back program[s] **	-1.797 (2.097)	-1.501 (1.710)	3.977 (5.634)	4.265 (5.541)

Attribute	Zero-centered utility		Probability score	
	Female (n=1,770)	**Male (n=470)**	**Female (n=1,770)**	**Male (n=470)**
Number of customer reviews***	-2.803 (1.825)	-2.030 (1.535)	1.907 (3.452)	2.680 (3.659)

Note: Standard deviations in parentheses. t-Test conducted with zero-centered utilities. [s] = sustainable apparel attribute, ***= $p < 0.001$, **= $p < 0.01$, *= $p < 0.05$.

6 Study 2a: Assessing sustainable apparel attributes

6.1 Conceptualization

Within the first BWS experiment of the second study's main part, we asked the respondents to evaluate only sustainable apparel attributes in the main part of the survey (see Table 4). We included nine items in the study and displayed four items per choice set. Every item was supposed to appear at least two times per respondent. This can be considered sufficient, as we expected our sample to be respectively large. For every choice set, respondents were asked to choose the most important and the least important attribute when purchasing clothes.

Table 4: Sustainable apparel attributes used in Study 2a.

Attribute	Description	References
Bio-based materials	Use of bio-based materials or fibres	(Baier et al., 2020; Fulton & Lee, 2013)
Country-of-manufacture	Country in which product was manufactured	(Fulton & Lee, 2013)
Durability	Product longevity in terms of robustness, reparability, timelessness	(Baier et al., 2020; Goworek et al., 2012)
Sustainability label	Product's sustainability is guaranteed with a label	(D'Souza et al., 2007; Hiller Connell, 2010)
Environmentally friendly production process	Production process saved valuable resources and caused few emissions	(Fulton & Lee, 2013)
Fair wages and working conditions	Manufactured under good working conditions; employees obtain appropriate wages	(Fulton & Lee, 2013; Goworek et al., 2012; Stöckigt et al., 2018)
Low-emission product	Product caused few emissions throughout the whole supply chain	(Fulton & Lee, 2013)
Recyclable materials	Use of recyclable materials or fibres	(Fulton & Lee, 2013)
Take-back program	Garment can be returned to the retailer for disposal to prolong the life cycle; it will be recycled, re-used, or environmentally friendly disposed	(Baier et al., 2020)

6.2 Results

Analogously to Study 1a, we applied hierarchical Bayes (HB) estimation and ran 30,000 iterations (including 20,000 burn-in iterations). Further, we calculated the average BW scores. The model's root likelihood (RLH) is 0.560, which is higher than the null model's RLH of 0.25, and thus, internal consistency is provided.

For our respondents, the garment's durability, fair wages and manufacturing conditions, a product causing few emissions, and an environmentally friendly production process are the most important sustainable apparel attributes. These attributes' probability scores are comparable, see Table 5. Durability can be considered not only a sustainable attribute reducing consumption levels of consumers and thus, decreasing overall apparel production, but further an economic attribute allowing consumers to wear their clothes longer with fewer replacements and hence, fewer investments. Respondents emphasize the social and environmental facets of sustainability, particularly those environmental aspects, which are obviously associated with CO_2 emissions. Material usage or the country-of-manufacture, which are only indirectly linked to CO_2 emissions (e.g., due to longer transportation routes), are considered only half as important. Moreover, programs for returning used clothes and labels indicating sustainable products are rather unimportant.

Post-hoc Tukey HSD test revealed that all attributes' zero-centered utilities are significantly different from each other, except for environmentally friendly production process versus low-emission product.

Table 5: Best-worst scaling results for Study 2a (n=2,106).

Attribute	Average BW score	Zero-centered utility score	Probability score
Durability	0.370	1.986 (2.275)	18.871 (10.174)
Fair wages and working conditions	0.359	1.787 (1.378)	18.800 (7.529)
Environmentally friendly production process	0.298	1.503 (0.952)	17.381 (5.240)
Low-emission product	0.267	1.598 (1.425)	17.552 (7.163)
Recyclable materials	0.010	0.101 (1.086)	9.183 (5.581)
Bio-based materials	-0.131	-0.831 (1.364)	5.550 (5.318)
Sustainability label	-0.291	-1.492 (1.662)	4.474 (5.779)
Country-of-manufacture	-0.406	-2.219 (2.158)	4.004 (6.598)
Take-back program	-0.426	-2.434 (2.440)	4.184 (7.169)

Note: Standard deviations in parentheses.

With regard to the respondents' gender, both groups consider the garment's durability as the most important attribute, and for male respondents it is even more decisive for their purchase decision ($t =$

-3.130, $p = 0.002$) (see Table 6). In contrast, the social facet of sustainability, i.e., fair wages and working conditions, is more important for female respondents than for male respondents ($t = 4.119$, $p < 0.001$). Furthermore, female consumers put more emphasis on environmental aspects such as an environmentally friendly production process ($t = 6.672$, $p < 0.001$), a product causing few emissions ($t = 7.924$ $p < 0.001$), and bio-based materials ($t = 3.243$, $p = 0.001$). Similar to the findings of Study 1a, male consumers tend to put more emphasis on the opinion of third parties as they consider sustainability labels as more important than female consumers ($t = -3.008$, $p = 0.003$), and they rather value take-back programs ($t = -3.240$, $p = 0.001$). Further, the country-of-manufacture is of higher importance to them compared to female respondents ($t = -3.013$, $p = 0.003$).

Table 6: Gender-specific best-worst scaling results for Study 2a.

Attribute	Zero-centered utility		Probability score	
	Female (n=1,678)	**Male (n=417)**	**Female (n=1,678)**	**Male (n=417)**
Durability**	1.892 (2.336)	2.277 (1.864)	18.257 (10.211)	21.562 (9.366)
Fair wages and working conditions***	1.835 (1.405)	1.529 (1.163)	18.808 (7.558)	18.535 (6.960)
Environmentally friendly production process***	1.584 (0.978)	1.231 (0.853)	17.652 (5.291)	16.636 (5.145)
Low-emission product***	1.724 (1.438)	1.117 (1.228)	17.977 (7.121)	15.819 (6.720)
Recyclable materials	0.085 (1.161)	0.155 (0.788)	9.102 (5.789)	9.590 (4.549)
Bio-based materials**	-0.782 (1.477)	-1.025 (0.795)	5.876 (5.730)	4.151 (3.038)
Sustainability label**	-1.542 (1.646)	-1.273 (1.583)	4.241 (5.529)	5.237 (6.326)
Country-of-manufacture**	-2.285 (2.219)	-1.931 (1.838)	3.927 (6.541)	4.237 (6.595)
Take-back program**	-2.512 (2.524)	-2.080 (2.020)	4.161 (7.300)	4.232 (6.479)

Note: Standard deviations in parentheses. t-Test conducted with zero-centered utilities. ***= $p < 0.001$, **= $p < 0.01$, *= $p < 0.05$.

7 Study 1b: Assessing willingness to pay for sustainable apparel attributes

7.1 Conceptualization

Within the second BWS experiment of the first study's main part, we determined how much consumers are willing to pay more for sustainable apparel attributes. In extension to the non-monetary attributes investigated in Study 2a, we added three monetary items – 10%, 15%, and 20% discount – to answer this question. In the highly competitive context of fashion online shopping, customers are used to monetary- and non-monetary product-related offerings. Discounts in this range are appropriate and can be used to comparatively measure the subjective value of the sustainable apparel attributes as alternative objects. Table 7 summarizes the list of items used in our BWS study.

Table 7: Sustainable apparel and price-related attributes used in Study 1b.

Attribute	Description	References
10% discount	A price reduction of 10% is offered	(Gaul, 1989; Gierl & Schwanenberg, 1997; Kaas, 1977)
15% discount	A price reduction of 10% is offered	(Gaul, 1989; Gierl & Schwanenberg, 1997; Kaas, 1977)
20% discount	A price reduction of 10% is offered	(Gaul, 1989; Gierl & Schwanenberg, 1997; Kaas, 1977)
Sustainability label	Product’s sustainability is guaranteed with a label	(D'Souza et al., 2007; Hiller Connell, 2010)
Bio-based materials	Use of bio-based materials or fibres	(Baier et al., 2020; Fulton & Lee, 2013)
Fair wages and working conditions	Manufactured under good working conditions; employees obtain appropriate wages	(Fulton & Lee, 2013; Goworek et al., 2012; Stöckigt et al., 2018)
Environmentally friendly production process	Production process saved valuable resources and caused few emissions	(Fulton & Lee, 2013)
Low-emission product	Product caused few emissions throughout the whole supply chain	(Fulton & Lee, 2013)
Recyclable materials	Use of recyclable materials or fibres	(Fulton & Lee, 2013)
Durability	Product longevity in terms of robustness, reparability, timelessness	(Baier et al., 2020; Goworek et al., 2012)
Country-of-manufacture	Country in which product was manufactured	(Fulton & Lee, 2013)

The usage of these discounts is in line with the approach that Kaas (1977) introduced to random utility theory and was later further elaborated by Gaul (1989) and Gierl and Schwanenberg (1997): The idea is to confront respondents with choice tasks that comprise monetary objects (e.g., certain amounts of money or discounts) as well as non-monetary objects (e.g., real or hypothetical products) and to use the collected choice frequencies to derive mean subjective values of the objects. Then, by comparing the subjective values of monetary and non-monetary objects, it is possible to calculate the monetary value of the non-monetary objects (via linear transformations). In their experiments with consumer goods, Kaas (1977), Gaul (1989), as well as Gierl and Schwanenberg (1997) demonstrated that this approach leads to valid monetary values of the non-monetary objects.

The chosen approach circumvents the traditional critique of choice-based conjoint analysis where the willingness to pay for attribute-levels is calculated by comparing partworths for attribute-levels with price coefficients or partworths for price-levels (G. Allenby et al., 2013; G. M. Allenby et al., 2014), since there assumptions of a base product (as a specified attribute-level-combination) and a competitive environment are made, which limits the general validity of the derived willingness to pay and results in a tendency to overestimate the effects of improved attribute-levels. In our BWS approach, we collect choices with respect to all monetary and non-monetary objects in Table 7 and therefore, do not assume additive models, competitors, or attribute-level-combinations as base products.

Based on the list of attributes (objects) in Table 7, a BWS experiment with eight choice tasks for each respondent, four objects in each choice task, and 300 versions of questionnaires were designed. Similar to the paired comparisons experiment in Kaas (1977), Gaul (1989), as well as Gierl and Schwanenberg (1997) three prohibitions were specified: A maximum of one discount attribute was allowed to appear in one choice tasks. Consequently, it was only possible to develop a slightly unbalanced and orthogonal choice task design. The monetary (discount) objects appeared less often (800 times in the $8 \cdot 4 \cdot 300 = 9.600$ choice sets) than the non-monetary objects (900 times) and pairs of monetary objects did not appear at all in the choice sets, whereas pairs of monetary and non-monetary objects appeared slightly more often than pairs of non-monetary objects.

7.2 Results

The choice frequencies were analyzed using HB for estimation. The mean RLH value of 0.615 across the 2,244 respondents (SD: 0.135) can be considered good compared to the random RLH value of 0.25.

The results are summarized in Table 8. First, it is obvious that the ranking of the zero-centered utilities and the probability scores are similar to the results in Table 5, implying quite valid results of Study 2 (without considering the monetary attributes): The attributes durability, low-emission product, environmentally friendly production process, as well as fair wages and working conditions are by far the highest-ranked attributes, whereas the attributes country-of-manufacture, bio-based materials, and sustainability label are the lowest-ranked attributes. Concerning the monetary attributes, it seems that the 10% discount is ranked rather low on average. The most important sustainable apparel attribute durability is equivalent to a 30.80% discount for the average respondent. However, it has to be mentioned that this discount was not offered directly to the respondents in the study.

Post-hoc Tukey HSD test showed that all attributes' zero-centered utilities were different from each other, except for environmentally friendly production process versus (1) low-emission product and (2) fair wages and working conditions; recyclable materials versus 20% discount; as well as bio-based materials versus 15% discount.

Table 8: Best-worst scaling results for Study 1b (n=2,244) and calculation of monetary values (discount equivalents) for non-monetary attributes.

Attribute	Zero-centered utility	Probability score	Discount equivalents
10% discount	-2.246 (3.563)	5.603 (8.647)	-
15% discount	-1.091 (3.953)	8.184 (9.844)	-
20% discount	0.033 (4.830)	11.089 (10.887)	-
Durability	2.493 (2.376)	16.038 (7.709)	30.80%
Low-emission product	1.457 (2.049)	12.755 (6.962)	26.25%
Environmentally friendly production process	1.250 (1.779)	11.889 (6.085)	25.34%
Fair wages and working conditions	1.132 (2.058)	11.280 (6.722)	24.83%
Recyclable materials	-0.209 (1.812)	6.664 (5.401)	18.94%
Sustainability label	-0.513 (2.464)	6.540 (6.510)	17.61%
Bio-based materials	-0.849 (2.103)	5.208 (5.166)	16.13%
Country-of-manufacture	-1.457 (2.670)	4.748 (6.109)	13.46%

Note: Standard deviations in parentheses.

We gathered similar results as in Study 2a (without monetary attributes) when comparing the subsamples of female and male respondents (see Table 9): Female respondents value every sustainable apparel attribute on average significantly higher than the male respondents (except for durability), whereas male respondents put more emphasis on the discounts and the garment's durability. The comparisons were made based on zero-centered utilities, using the Maximum Likelihood across the (sub)sample and assuming that identical subjective values lead to similar results. More specifically, the discount equivalents yielded for female respondents are higher for every attribute than for male respondents.

Table 9: Gender-specific best-worst scaling results for Study 1b and calculation of monetary values (discount equivalents) for non-monetary attributes.

Attribute	Zero-centered utility		Probability score		Discount equivalents	
	Female (n=1,770)	Male (n=470)	Female (n=1,770)	Male (n=470)	Female (n=1,770)	Male (n=470)
10% discount**	-2.364 (3.549)	-1.798 (3.593)	5.308 (8.521)	6.744 (9.051)	-	-
15% discount**	-1.209 (3.938)	-0.643 (3.985)	7.819 (9.593)	9.578 (10.318)	-	-
20% discount*	-0.081 (4.613)	0.465 (4.681)	10.801 (10.789)	12.166 (11.220)	-	-

Attribute	Zero-centered utility		Probability score		Discount equivalents	
	Female (n=1,770)	Male (n=470)	Female (n=1,770)	Male (n=470)	Female (n=1,770)	Male (n=470)
Durability**	2.423 (2.394)	2.747 (2.294)	15.771 (7.780)	16.997 (7.334)	30.97%	30.09%
Low-emission product***	1.533 (2.030)	1.166 (2.099)	13.038 (6.912)	11.658 (7.035)	27.07%	23.10%
Environmentally friendly production process***	1.323 (1.758)	0.972 (1.832)	12.172 (6.013)	10.818 (6.249)	26.15%	22.24%
Fair wages and working conditions***	1.209 (2.050)	0.846 (2.060)	11.502 (6.680)	10.461 (6.807)	25.65%	21.69%
Recyclable materials	-0.192 (1.819)	-0.278 (1.784)	6.711 (5.395)	6.482 (5.427)	19.51%	16.72%
Sustainability label	-0.490 (2.500)	-0.594 (2.332)	6.606 (6.524)	6.312 (6.477)	18.21%	15.32%
Bio-based materials***	-0.748 (2.089)	-1.236 (2.107)	5.437 (5.268)	4.343 (4.656)	17.08%	12.49%
Country-of-manufacture*	-1.406 (2.675)	-1.648 (2.654)	4.835 (6.123)	4.440 (6.071)	14.19%	10.67%

Note: Standard deviations in parentheses. t-Test conducted with zero-centered utilities. ***= $p < 0.001$, **= $p < 0.01$, *= $p < 0.05$.

8 Study 2b: Assessing sustainable and conventional store attributes

8.1 Conceptualization

Within the second BWS experiment of the second study's main part, we asked the respondents to assess sustainable and conventional online shop attributes (see Table 10). We included 14 items (i.e., five sustainable and nine conventional) in the study and displayed five items per choice set. Every item was supposed to appear at least three times per respondent. For every choice set, respondents were asked to choose the most important and the least important attribute of an online shop.

Table 10: Sustainable and conventional online shop attributes used in Study 2b.

Attribute	Description	References
Assurance seal	Online shop's security is verified by an assurance seal	(Kovar et al., 2000; Odom et al., 2002)
Availability of customer reviews	Online shop provides customer reviews for its products	(Chevalier & Mayzlin, 2006; Park et al., 2007)
Availability of customer service	Customer service is easily available and online shop has sufficient channels for contact	(Cao et al., 2018)

Broad product range	Online shop offers a broad variety of different products	
Data security	Customers' data are treated confidential and are not misued or passed to third parties	(Forsythe & Shi, 2003; Ingaldi & Ulewicz, 2019)
Discount campaigns	Online shop offers frequent discount campaigns	(Abraham-Murali & Littrell, 1995; Eckman et al., 1990)
Fast shipping	Orders are processed and shipped quickly	(Ma, 2017; Stöckigt et al., 2018)
Free returns	Returns are free of charge	(Bower & Maxham, 2012; Cao et al., 2018)
Free shipping	Shipping is free of charge	(Bower & Maxham, 2012; Lewis, 2006; Ma, 2017; Stöckigt et al., 2018)
Broad sustainable product range^{s}	Online shop offers a broad variety of different sustainable products	(Lundblad & Davies, 2016; Pookulangara & Shephard, 2013)
Climate-neutral shipping^{s}	Online shop offers climate-neutral shipping	(Stöckigt et al., 2018)
Less packaging^{s}	Online shop reduces packaging of its orders	(Nguyen et al., 2020; Schwepker & Cornwell, 1991)
Sustainable packaging^{s}	Packaging is environmentally friendly	(Fulton & Lee, 2013; Nguyen et al., 2020; Prakash & Pathak, 2017)
Sustainability seal^{s}	Online shop is certified with a sustainability seal	

Note: ^{s} = sustainable apparel attribute.

8.2 Results

We applied hierarchical Bayes (HB) estimation with 30,000 iterations (including 20,000 burn-in iterations). Additionally, we calculated the average BW scores. Internal consistency is given, considering the model's root likelihood (RLH) is 0.483, which is higher than the null model's RLH of 0.2. The results of Study 2b are reported in Table 11.

From an overall perspective, consumers demand less packaging of their orders, returns free of charge, and frequent discount campaigns. They also wish for more sustainable packaging, free shipping, and data security. Apparently, consumers mainly care about the 'how' (in terms of packaging) and 'how much' (in terms of fees) of their shipping. Nevertheless, consumers care less about climate-neutral shipping (as this frequently implies additional costs) and overall shipping speed. This indicates that our respondents are price-sensitive when it comes to shipping and return fees. They consider reducing packaging waste and using sustainable packaging as fruitful possibilities to reduce the environmental impact of online shopping. Further, consumers consider trust and sustainability seals of online shops as rather unimportant and do not mind other customers' reviews.

The post-hoc Tukey HSD test proved that – except for broad sustainable product range versus (1) data security, (2) free shipping, and (3) assurance seal; data security versus (1) free returns and (2) discount campaigns; availability of customer service versus (1) assurance seal, (2) broad product range, and (3) climate-neutral shipping; climate-neutral shipping versus (1) assurance seal and (2) broad product range; discount campaigns versus (1) free shipping and (2) sustainable packaging – all attributes' zero-centered utilities are different from each other.

Table 11: Best-worst scaling results for Study 2b (n=2,106).

Attribute	Average BW score	Zero-centered utility	Probability score
Less packaging[s]	0.287	1.835 (1.914)	11.793 (6.495)
Free returns	0.204	1.309 (2.082)	10.602 (7.080)
Discount campaigns	0.132	0.751 (2.469)	9.324 (7.687)
Sustainable packaging[s]	0.091	0.789 (1.739)	8.141 (5.802)
Data security	0.071	0.517 (2.198)	7.999 (6.704)
Free shipping	0.052	0.499 (2.326)	8.486 (7.246)
Broad sustainable product range[s]	0.051	0.401 (1.972)	7.304 (6.156)
Availability of customer service	0.006	-0.155 (1.860)	6.456 (5.934)
Assurance seal	0.005	0.019 (2.405)	7.160 (7.036)
Broad product range	0.0002	-0.145 (2.366)	7.009 (6.943)
Climate-neutral shipping[s]	-0.056	0.0865 (1.853)	6.117 (5.377)
Sustainability seal[s]	-0.114	-0.636 (2.524)	5.392 (6.139)
Availability of customer reviews	-0.322	-2.162 (1.967)	2.413 (3.983)
Fast shipping	-0.459	-3.110 (2.118)	1.804 (3.743)

Note: Standard deviations in parentheses, [s] = sustainable apparel attribute.

Comparing both female as well as male respondents (see Table 12), we found female respondents to put more emphasis on less ($t = 7.647, p < 0.001$) and sustainable ($t = 8.159, p < 0.001$) packaging, returns free of charge ($t = 5.091$, $p < 0.001$), data security ($t = 3.264$, $p = 0.001$), a broad sustainable product range to choose from ($t = 2.723, p = 0.007$), and climate-neutral shipping ($t = 4.967, p < 0.001$). Male respondents, in turn, are rather interested in the availability of customer service ($t = -6.745$, $p < 0.001$), a broad product range to choose from ($t = -3.588$, $p < 0.001$), the availability of customer reviews ($t = -7.630, p < 0.001$), and a fast shipping procedure ($t = -11.910, p < 0.001$).

Overall, all five sustainable online shop attributes were more important for female respondents than for male respondents (except for the sustainability seal). Further, female respondents seem to be more riskaverse regarding their personal data. Again, male consumers were found to emphasize more the opinion of third parties, i.e., other customers, and rather value customer touchpoints, i.e., the availability of customer service, to have the option of contacting in case of complaints and problems during the shopping process. They seem to be more price-sensitive regarding shipping fees than female consumers. In turn, male respondents care less about return fees.

Table 12: Gender-specific best-worst scaling results for Study 2b.

Attribute	Zero-centered utility		Probability score	
	Female (n=1,678)	Male (n=417)	Female (n=1,678)	Male (n=417)
Less packaging[s] ***	1.835 (1914)	1.038 (1.851)	12.193 (6.346)	10.238 (6.867)
Free returns***	1.309 (2.082)	0.730 (2.062)	10.785 (7.023)	9.715 (7.142)
Discount campaigns	0.751 (2.469)	0.714 (2.281)	9.248 (7.686)	9.506 (7.725)
Sustainable packaging[s] ***	0.789 (1.739)	0.029 (1.535)	8.561 (5.857)	6.372 (5.132)
Data security**	0.517 (2.198)	0.131 (2.007)	8.137 (6.644)	7.346 (6.674)
Free shipping	0.500 (2.326)	0.598 (2.185)	8.285 (7.191)	9.198 (7.328)
Broad sustainable product range[s] **	0.401 (1.972)	0.117 (1.604)	7.418 (6.239)	6.810 (5.671)
Availability of customer service***	-0.155 (1.860)	0.511 (1.538)	6.031 (5.789)	8.274 (6.090)
Assurance seal	0.019 (2.405)	-0.112 (2.205)	7.198 (7.065)	7.032 (6.927)
Broad product range***	-0.145 (2.366)	0.309 (2.076)	6.729 (6.866)	8.191 (7.170)
Climate-neutral shipping[s] ***	0.086 (1.853)	-0.409 (1.675)	6.443 (5.489)	5.184 (4.939)
Sustainability seal[s]	-0.636 (2.524)	-0.597 (1.977)	5.428 (6.238)	5.239 (5.663)
Availability of customer reviews***	-2.161 (1.967)	-1.351 (1.818)	2.142 (3.661)	3.405 (4.693)
Fast shipping***	-3.110 (2.118)	-1.708 (2.270)	1.401 (3.203)	3.489 (5.166)

Note: Standard deviations in parentheses. t-Test conducted with zero-centered utilities. [s] = sustainable apparel attribute, ***= $p < 0.001$, **= $p < 0.01$, *= $p < 0.05$.

9 Discussion

9.1 Theoretical contribution

While the importance of single conventional apparel attributes (Abraham-Murali & Littrell, 1995; Eckman et al., 1990) as well as online shop attributes (Stöckigt et al., 2018) for consumers has been discussed in several studies, research lacks comparing a comprehensive set of different conventional as well as sustainable apparel and online shop attributes: Extant research on comparing apparel attributes focused on materials as sustainable attributes only (Viciunaite & Alfnes, 2020). Further, literature on willingness to pay for sustainable apparel attributes is limited to materials (Ellis et al., 2012; Ha-Brookshire & Norum, 2011; Hustvedt & Bernard, 2008), country-of-manufacture (Hustvedt & Bernard, 2008), and labels (Hustvedt & Bernard, 2010). Thus, our contribution to literature is multifold:

First, we are the first study to compare the importance of an extensive set of both conventional as well as sustainable apparel attributes and, additionally, sustainable attributes only. Overall, we align with prior research, which compared the importance of sustainable materials and environmentally friendly production process with conventional attributes within a BWS experiment, finding that conventional apparel attributes are of higher importance to consumers (Viciunaite & Alfnes, 2020): We found fit and comfort, price-performance ratio, quality, and design to be the most important attributes when purchasing garments, which was previously indicated by research on conventional apparel attributes (Abraham-Murali & Littrell, 1995; Eckman et al., 1990). Concerning sustainable apparel attributes, consumers value durability and the social facet of sustainability, i.e., fair wages and working conditions. Materials were of inferior importance, contradicting the findings of Viciunaite and Alfnes (2020). With respect to the respondents' gender, we found female consumers to consider almost every sustainable attribute as more important than male consumers, particularly when it comes to the social facet of sustainability, i.e., fair wages and working conditions, as well as bio-based materials, and environmental aspects such as an environmentally friendly production process and a low-emission product. This aligns with prior research indicating that female consumers rather exhibit environmentally friendly behaviors and are more likely to engage in sustainable behaviors (Brough et al., 2016; Lee & Holden, 1999). Female consumers further value fit and comfort as well as design more than male consumers. In contrast, male respondents put more emphasis on the garment's durability and brand, the opinion of third parties (i.e., quality and number of customer reviews), signaling (i.e., sustainability label), and take-back programs.

Second, we extend previous research on willingness to pay for sustainable apparel attributes: we found durability as the most important sustainable attribute to be equivalent to a 30.80% discount. Regarding materials, we found bio-based as well as recyclable materials to be equivalent to a 16.13% and 18.94% discount, respectively. This is slightly lower compared to the results of Ellis et al. (2012), who found consumers to pay 25% more for organic cotton. Similarly to Hustvedt and Bernard (2010), who indicated that consumers are willing to pay a premium for labels indicating social responsibility and fair trade, we found fair wages and working conditions to yield among the highest discount equivalent. Female respondents were again found to value sustainable apparel attributes more, as almost all attributes' discount equivalents were higher for female than for male respondents.

Third, we provide first insights regarding the importance of both conventional and sustainable online shop attributes. Extant literature focusing on packaging waste and/or sustainable packaging in the e-commerce context flourished throughout the past years (Nguyen et al., 2020; Prakash & Pathak, 2017;

Schwepker & Cornwell, 1991), indicating its importance for consumers. Indeed, we found less as well as sustainable packaging to be among the four most important online shop attributes. Besides, we found returns and shipping free of charge to be emphasized by the respondents, which is in line with preceding findings (Bower & Maxham, 2012; Smith & Brynjolfsson, 2001; Stöckigt et al., 2018). Nevertheless, we found shipping speed and climate-neutral shipping to be less important, contradicting prior findings (Stöckigt et al., 2018). Female respondents were more interested in almost all sustainable online shop attributes (i.e., less and sustainable packaging, broad sustainable product range, climate-neutral shipping) compared to male respondents, highlighting the greenness of female consumers (Brough et al., 2016; Lee & Holden, 1999).

9.2 Practical implications

Our findings imply several practical implications for e-commerce managers, (sustainable) apparel retailers as well as manufacturers, and further stakeholders.

Overall, conventional apparel attributes are of striking importance to consumers, particularly fit and comfort, price-performance ratio, quality, and design. Nevertheless, in the context of sustainability, apparel retailers should keep the social facet of sustainability (i.e., fair wages and working conditions) in mind. To avoid image damages, manufacturers are recommended to guarantee fair working standards for their employees. Further, the garment's durability is decisive for the consumers' purchase: Consumers value robustness and reparability of garments and further, seek for timeless styles. Additionally, it seems of high importance to minimize the environmental impact of the production process and the product in general. Retailers and manufacturers are recommended to disclose this information for a transparent communication with their customers. Material usage, which primarily affects consumers' health, is of inferior importance. Apparently, consumers value altruistic (i.e., environmental and social) facets of sustainable clothing more than egoistic facets (i.e., aspects affecting health). Thus, retailers and e-commerce managers should highlight or even visualize the consumer's environmental and social impact when purchasing the respective sustainable garment rather than highlighting the consumer's personal benefit. To overcome the gender gap in green consumerism, e-commerce and retailer managers should consider addressing male consumers with male testimonials and ad campaigns targeting men.

E-commerce managers should further avoid unnecessary packaging and use biodegradable or recycled alternatives. Moreover, consumers were found to be sensitive in terms of return and shipping fees: While free returns are more important to women, men are rather interested in free shipping. Apparently, men's intention to return their order when purchasing apparel products is lower compared to women. The respondents' focus on monetary aspects is further reflected in the high importance of discount campaigns.

9.3 Limitations and future research

Notwithstanding our theoretical and practical implications, our research is subject to several limitations stimulating further studies. First, albeit our samples are appropriately reflecting the target population of the online retailer, younger as well as male consumers are rather underrepresented. Future research could replicate our studies with a younger sample and a higher number of male respondents.

Furthermore, we did not directly determine consumers' willingness to pay, but discount equivalents. We did not conduct the assessment based on a specific apparel product, and discount equivalents may depend on the product's base price.

Sustainable product attributes may have been underrepresented in Study 1a, and hence, future research could conduct a BWS experiment with a balanced number of conventional and sustainable apparel attributes.

Appendix: Descriptive Statistics

Table A.1: Descriptive statistics of Study 1 (n=2,244).

Demographics/Characteristics		Frequency	Proportion (in %)
Gender	Female	1,770	78.9
	Male	470	20.9
	Diverse	4	0.2
Age	≤29 years	109	4.9
	30-39 years	299	13.3
	40-49 years	501	22.3
	50-59 years	786	35.1
	≥60 years	549	24.5
Employment	Full-time employee	1,006	44.8
	Part-time employee	525	23.4
	Student	43	1.9
	Unemployed	52	2.3
	Retired	492	21.9
	Others	126	5.6
Household’s total net income	≤2,000 Euros	714	31.9
	2,001-3,000 Euros	425	18.9
	3,001-4,000 Euros	239	10.7
	4,001-5,000 Euros	110	4.9
	≥5,001 Euros	64	2.9
	No information provided	692	30.8
General perception of sustainability	Fair wages and working conditions	584	26.0
	No child labor	500	22.3
	Decreased environmental impact	1,055	47.0
	Buying local	328	14.6
	Avoid packaging waste	895	39.9
	Recycling	627	27.9
	Upcycling	157	7.0
	Durable and reparable products	1,054	47.0
	Reduce consumption level	289	12.9
	Animal welfare	408	18.2
	Reduce resource usage	740	33.0
Barriers towards sustainable consumption behavior	Lack of trust	515	23.0
	Lack of transparency	1,081	48.2
	High prices	1,139	50.8
	Lack of availability	637	28.4
	No clear marking as such	965	43.0
	Lack of information	1,248	55.6
	I always purchase sustainable products	72	3.2
	Other	67	3.0
Assessment of an organization’s sustainability	Labels	366	16.3
	Number of products with sustainability labels	327	14.6
	Take-back program	970	43.2
	Transparency	635	28.3
	Carbon footprint	270	12.0
	Use of regenerative energy	92	4.1
	Pesticide and chemicals usage	1,519	67.7

Demographics/Characteristics	Frequency	Proportion (in %)
Sustainable suppliers	556	24.8
Fair manufacturing conditions	1,524	67.9
Image of country-of-manufacturing	123	5.5
Level of social responsibility	330	14.7
Other	20	0.9

Table A.2: Descriptive statistics of Study 1 (n=2,106).

Demographics/Characteristics		Frequency	Proportion (in %)
Gender	Female	1,678	79.7
	Male	417	19.8
	Diverse	11	0.5
Age	≤29 years	119	5.7
	30-39 years	241	11.5
	40-49 years	488	23.2
	50-59 years	755	35.9
	≥60 years	503	23.9
Employment	Full-time employee	954	45.3
	Part-time employee	499	23.7
	Student	34	1.6
	Unemployed	52	2.5
	Retired	443	21.0
	Others	124	5.9
Household's total net income	≤2,000 Euros	615	29.2
	2,001-3,000 Euros	419	19.9
	3,001-4,000 Euros	241	11.4
	4,001-5,000 Euros	106	5.0
	≥5,001 Euros	62	2.9
	No information provided	663	31.5
General perception of sustainability	Fair wages and working conditions	623	29.6
	No child labor	475	22.6
	Decreased environmental impact	1,030	48.9
	Buying local	309	14.7
	Avoid packaging waste	772	36.7
	Recycling	576	27.4
	Upcycling	122	5.8
	Durable and reparable products	1,003	47.6
	Reduce consumption level	231	11.0
	Animal welfare	375	17.8
	Reduce resource usage	709	33.7
Barriers towards sustainable consumption behavior	Lack of trust	406	19.3
	Lack of transparency	919	43.6
	High prices	1,015	48.2
	Lack of availability	604	28.7
	No clear marking as such	863	41.0
	Lack of information	1,107	52.6
	I always purchase sustainable products	66	3.1
	Other	49	2.3
Assessment of an organization's sustainability	Labels	262	12.4
	Number of products with sustainability labels	194	9.2
	Take-back program	1,118	53.1
	Transparency	524	24.9
	Carbon footprint	249	11.8
	Use of regenerative energy	83	3.9
	Pesticide and chemicals usage	1,385	65.8
	Sustainable suppliers	493	23.4
	Fair manufacturing conditions	1,532	72.7

Demographics/Characteristics	Frequency	Proportion (in %)
Image of country-of-manufacturing	131	6.2
Level of social responsibility	330	15.7
Other	17	0.8

References

Abraham-Murali, L., & Littrell, M. A. (1995). Consumers' Conceptualization of Apparel Attributes. Clothing and Textiles Research Journal, 13(2), 65–74. https://doi.org/10.1177/0887302X9501300201

Allenby, G., Brazell, J., Howell, J., & Rossi, P. (2013). Using Conjoint Analysis to Determine the Market Value of Product Features. Sawtooth Software Conference Proceedings, Sawtooth Software Inc., 341–355.

Allenby, G. M., Brazell, J. D., Howell, J. R., & Rossi, P. E. (2014). Economic valuation of product features. Quantitative Marketing and Economics, 12(4), 421–456. https://doi.org/10.1007/s11129-014-9150-x

Allwood, J. M., Laursen, S. E., Russell, S. N., Rodríguez, C. M. de, & Bocken, N. (2008). An approach to scenario analysis of the sustainability of an industrial sector applied to clothing and textiles in the UK. Journal of Cleaner Production, 16(12), 1234–1246. https://doi.org/10.1016/j.jclepro.2007.06.014

Auger, P., Devinney, T. M., & Louviere, J. J. (2007). Using Best–Worst Scaling Methodology to Investigate Consumer Ethical Beliefs Across Countries. Journal of Business Ethics, 70(3), 299–326. https://doi.org/10.1007/s10551-006-9112-7

Baier, D., Rausch, T. M., & Wagner, T. F. (2020). The Drivers of Sustainable Apparel and Sportswear Consumption: A Segmented Kano Perspective. Sustainability, 12(7), 2788. https://doi.org/10.3390/su12072788

Bower, A. B., & Maxham, J. G. (2012). Return Shipping Policies of Online Retailers: Normative Assumptions and the Long-Term Consequences of Fee and Free Returns. Journal of Marketing, 76(5), 110–124. https://doi.org/10.1509/jm.10.0419

Brough, A. R., Wilkie, J. E. B., Ma, J., Isaac, M. S., & Gal, D. (2016). Is Eco-Friendly Unmanly? The Green-Feminine Stereotype and Its Effect on Sustainable Consumption. Journal of Consumer Research, 43(4), 567–582. https://doi.org/10.1093/jcr/ucw044

Cao, Y., Ajjan, H., & Hong, P. (2018). Post-purchase shipping and customer service experiences in online shopping and their impact on customer satisfaction: An empirical study with comparison. Asia Pacific Journal of Marketing and Logistics, 30(2), 400–416. https://doi.org/10.1108/APJML-04-2017-0071

Chen, C., Lee, S., & Stevenson, H. W. (2016). Response Style and Cross-Cultural Comparisons of Rating Scales Among East Asian and North American Students. Psychological Science, 6(3), 170–175. https://doi.org/10.1111/j.1467-9280.1995.tb00327.x (Psychological Science, 6(3), 170-175).

Cheung, K. L., Mayer, S., Simon, J., Vries, H. de, Evers, S. M., Kremer, I. E., & Hiligsmann, M. (2018). Comparison of statistical analysis methods for object case best–worst scaling. Journal of Medical Economics. Advance online publication. https://doi.org/10.1080/13696998.2018.1553781

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. Journal of Marketing Research, 43(3), 345–354. https://doi.org/10.1509/jmkr.43.3.345

Chrzan, K., & Golovashkina, N. (2006). An Empirical Test of Six Stated Importance Measures. International Journal of Market Research, 48(6), 717–740. https://doi.org/10.1177/147078530604800607

Cohen, S. (2003). Maximum difference scaling: improved measures of importance and preference for segmentation. Sawtooth Software Conference Proceedings, Sawtooth Software Inc., 61–74. http://www.sawtoothsoftware.com/download/techpap/maxdiff.pdf

Cohen, S., & Orme, B. (2004). What's your preference? Asking survey respondents about their preferences creates new scaling decisions. Marketing Research, 16, 32–37.

Connolly, J., & Prothero, A. (2003). Sustainable consumption: Consumption, consumers and the commodity discourse. Consumption Markets & Culture, 6(4), 275–291. https://doi.org/10.1080/1025386032000168311

Culpepper, R. A., & Zimmerman, R. A. (2006). Culture-based extreme response bias in surveys employing variable response items: An investigation of response tendency among Hispanic-Americans. Journal of International Business Research, 5(2), 75–84.

Delmas, M. A., & Lessem, N. (2017). Eco-Premium or Eco-Penalty? Eco-Labels and Quality in the Organic Wine Market. Business & Society, 56(2), 318–356. https://doi.org/10.1177/0007650315576119

de-Magistris, T., & Gracia, A. (2016). Consumers' willingness-to-pay for sustainable food products: The case of organically and locally grown almonds in Spain. Journal of Cleaner Production, 118, 97–104. https://doi.org/10.1016/j.jclepro.2016.01.050

D'Souza, C., Taghian, M., Lamb, P., & Peretiatko, R. (2007). Green decisions: demographics and consumer understanding of environmental labels. International Journal of Consumer Studies, 31(4), 371–376. https://doi.org/10.1111/j.1470-6431.2006.00567.x

Dyachenko, T., Naylor, R. W., & Allenby, G. (2013). The Ballad of Best and Worst. Sawtooth Software Conference Proceedings, Sawtooth Software Inc., 357–366.

Eckman, M., Damhorst, M. L., & Kadolph, S. J. (1990). Toward a Model of the In-Store Purchase Decision Process: Consumer Use of Criteria for Evaluating Women's Apparel. Clothing and Textiles Research Journal, 8(2), 13–22. https://doi.org/10.1177/0887302X9000800202

Eisler, A. D., & Eisler, H. (1994). Subjective Time Scaling: Influence of Age, Gender, and Type A and Type B Behavior. Chronobiologia, 21(3-4), 185–200.

Ellen, P. S. (1994). Do We Know What We Need to Know? Objective and Subjective Knowledge Effects on Pro-Ecological Behaviors. Journal of Business Research, 30, 43–52.

Ellis, J. L., McCracken, V. A., & Skuza, N. (2012). Insights into willingness to pay for organic cotton apparel. Journal of Fashion Marketing and Management: An International Journal, 16(3), 290–305. https://doi.org/10.1108/13612021211246053

Finn, A., & Louviere, J. J. (1992). Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. Journal of Public Policy & Marketing, 11(2), 12–25. https://doi.org/10.1177/074391569201100202

Fiore, A. M., & Damhorst, M. L. (1992). Intrinsic cues as predictors of perceived quality of apparel. Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, 5, 168–178.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. Journal of Consumer Research, 20(2), 303–315.

Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best–worst scaling: What it can do for health care research and how to do it. Journal of Health Economics, 26(1), 171–189. https://doi.org/10.1016/j.jhealeco.2006.04.002 (Journal of Health Economics, 26(1), 171-189).

Flynn, T. N., & Marley, A. (2014). Best-worst scaling: theory and methods. In S. Hess & A. Daly (Eds.), Handbook of Choice Modelling (pp. 178–202). Edward Elgar Publishing.

Forsythe, S. M., & Shi, B. (2003). Consumer patronage and risk perceptions in Internet shopping. Journal of Business Research, 56(11), 867–875. https://doi.org/10.1016/s0148-2963(01)00273-9

Fulton, K., & Lee, S.-E. (2013). Assessing sustainable initiatives of apparel retailers on the internet. Journal of Fashion Marketing and Management: An International Journal, 17(3), 353–366. https://doi.org/10.1108/JFMM-11-2012-0071

Gaul, W. (1989). Probabilistic Choice Behavior Models and Their Combination with Additional Tools Needed for Applications to Marketing. In K. C. Klauer, G. de Soete, & H. Feger (Eds.), Advances in psychology: Vol. 60. New Developments in Psychological Choice Modeling (1st ed., pp. 317–337). North-Holland. https://doi.org/10.1016/s0166-4115(08)60242-3

Gierl, H., & Schwanenberg, S. (1997). Die Schätzung der Preisresponsefunktion aus Paarvergleichsdaten. der markt, 36(1), 3–10. https://doi.org/10.1007/bf03032080

Gil, J. M., Gracia, A., & Sánchez, M. (2000). Market segmentation and willingness to pay for organic products in Spain. The International Food and Agribusiness Management Review, 3(2), 207–226. https://doi.org/10.1016/S1096-7508(01)00040-4

Goworek, H., Fisher, T., Cooper, T., Woodward, S., & Hiller, A. (2012). The sustainable clothing market: an evaluation of potential strategies for UK retailers. International Journal of Retail & Distribution Management, 40(12), 935–955. https://doi.org/10.1108/09590551211274937

Ha-Brookshire, J. E., & Norum, P. S. (2011). Willingness to pay for socially responsible products: case of cotton apparel. Journal of Consumer Marketing, 28(5), 344–353. https://doi.org/10.1108/07363761111149992

Harris, F., Roby, H., & Dibb, S. (2016). Sustainable clothing: Challenges, barriers and interventions for encouraging more sustainable consumer behaviour. International Journal of Consumer Studies, 40(3), 309–318. https://doi.org/10.1111/ijcs.12257

Hiller Connell, K. Y. (2010). Internal and external barriers to eco-conscious apparel acquisition. International Journal of Consumer Studies, 34(3), 279–286. https://doi.org/10.1111/j.1470-6431.2010.00865.x

Horne, J., & Rayner, B. (2013). Does the Analysis of MaxDiff Data Require Separate Scaling Factors? Sawtooth Software Conference Proceedings, Sawtooth Software Inc., 331–339.

Hustvedt, G., & Bernard, J. C. (2008). Consumer willingness to pay for sustainable apparel: The influence of labelling for fibre origin and production methods. International Journal of Consumer Studies, 32(5), 491–498. https://doi.org/10.1111/j.1470-6431.2008.00706.x

Hustvedt, G., & Bernard, J. C. (2010). Effects of social responsibility labelling and brand on willingness to pay for apparel. International Journal of Consumer Studies, 34(6), 619–626. https://doi.org/10.1111/j.1470-6431.2010.00870.x

Hustvedt, G., & Dickson, M. A. (2009). Consumer likelihood of purchasing organic cotton apparel. Journal of Fashion Marketing and Management: An International Journal, 13(1), 49–65. https://doi.org/10.1108/13612020910939879

Ingaldi, M., & Ulewicz, R. (2019). How to Make E-Commerce More Successful by Use of Kano's Model to Assess Customer Satisfaction in Terms of Sustainable Development. Sustainability, 11(18), 4830. https://doi.org/10.3390/su11184830

Jacoby, J., Berning, C. K., & Dietvorst, T. F. (1977). What about Disposition? Journal of Marketing, 41(2), 22–28.

Jegethesan, K., Sneddon, J. N., & Soutar, G. N. (2012). Young Australian consumers' preferences for fashion apparel attributes. Journal of Fashion Marketing and Management: An International Journal, 16(3), 275–289. https://doi.org/10.1108/13612021211246044

Joergens, C. (2006). Ethical fashion: Myth or future trend? Journal of Fashion Marketing and Management: An International Journal, 10(3), 360–371. https://doi.org/10.1108/13612020610679321

Kaas, K. P. (1977). Empirische Preisabsatzfunktionen bei Konsumgütern. Heidelberger betriebswirtschaftliche Studien. Springer. https://doi.org/10.1007/978-3-642-81160-9

Kautish, P., Paul, J., & Sharma, R. (2019). The moderating influence of environmental consciousness and recycling intentions on green purchase behavior. Journal of Cleaner Production, 228, 1425–1436. https://doi.org/10.1016/j.jclepro.2019.04.389

Kollmuss, A., & Agyeman, J. (2002). Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? Environmental Education Research, 8(3), 239–260. https://doi.org/10.1080/13504620220145401

Kovar, S. E., Burke, K. G., & Kovar, B. R. (2000). Consumer Responses to the CPA WEBTRUST™ Assurance. Journal of Information Systems, 14(1), 17–35. https://doi.org/10.2308/jis.2000.14.1.17

KPMG. (2019). Sustainable fashion: A survey on global perspectives. https://assets.kpmg/content/dam/kpmg/cn/pdf/en/2019/01/sustainable-fashion.pdf

Lee, J. A., & Holden, S. J. S. (1999). Understanding the determinants of environmentally conscious behavior. Psychology & Marketing, 16(5), 373–392.

Lee, J. A., Soutar, G. N., & Louviere, J. J [Jordan J.] (2007). Measuring values using best-worst scaling: The LOV example. Psychology & Marketing, 24(12), 1043–1058. https://doi.org/10.1002/mar.20197

Lewis, M. (2006). The effect of shipping fees on customer acquisition, customer retention, and purchase quantities. Journal of Retailing, 82(1), 13–23. https://doi.org/10.1016/J.JRETAI.2005.11.005 (Journal of Retailing).

Louviere, J. J., & Flynn, T. N. (2010). Using best-worst scaling choice experiments to measure public perceptions and preferences for healthcare reform in australia. The Patient, 3(4), 275–283. https://doi.org/10.2165/11539660-000000000-00000

Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). Best-worst scaling: Theory, methods and applications. Cambridge University Press. https://doi.org/10.1017/CBO9781107337855

Louviere, J. J., Lings, I., Islam, T., Gudergan, S., & Flynn, T. (2013). An introduction to the application of (case 1) best–worst scaling in marketing research. International Journal of Research in Marketing, 30(3), 292–303. https://doi.org/10.1016/j.ijresmar.2012.10.002

Louviere, J. J., & Woodworth, G. G. (1991). Best-worst scaling: A model for the largest difference judgments. Working Paper. https://scholar.google.de/citations?user=rp3xtscaaaaj&hl=de&oi=sra

Luchs, M. G., & Mooradian, T. A. (2012). Sex, Personality, and Sustainable Consumer Behaviour: Elucidating the Gender Effect. Journal of Consumer Policy, 35(1), 127–144. https://doi.org/10.1007/s10603-011-9179-0

Lundblad, L., & Davies, I. A. (2016). The values and motivations behing sustainable fashion consumption. Journal of Consumer Behaviour, 15, 149–162. https://doi.org/10.1002/cb.1559

Ma, S. (2017). Fast or free shipping options in online and Omni-channel retail? The mediating role of uncertainty on satisfaction and purchase intentions. The International Journal of Logistics Management, 28(4), 1099–1122. https://doi.org/10.1108/IJLM-05-2016-0130

Maichum, K., Parichatnon, S., & Peng, K.-C. (2016). Application of the Extended Theory of Planned Behavior Model to Investigate Purchase Intention of Green Products among Thai Consumers. Sustainability, 8(1077), 1–20. https://doi.org/10.3390/su8101077

Marley, A., & Louviere, J. J. (2005). Some probabilistic models of best, worst, and best–worst choices. Journal of Mathematical Psychology, 49(6), 464–480. https://doi.org/10.1016/j.jmp.2005.05.003

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), Frontiers in econometrics (pp. 105–142). Academic Press.

McKinsey & Company, & Global Fashion Agenda. (2020). Fashion on Climate: How the fashion industry can urgently act to reduce its greenhouse gas emissions. https://www.globalfashionagenda.com/publications-and-policy/fashion-on-climate/

Miyazaki, A. D., & Fernandez, A. N. A. (2001). Consumer Perceptions of Privacy and Security Risks for Online Shopping. Journal of Consumer Affairs, 35(1), 27–44. https://doi.org/10.1111/j.1745-6606.2001.tb00101.x

Morgan, L. R., & Birtwistle, G. (2009). An investigation of young fashion consumers' disposal habits. International Journal of Consumer Studies, 33(2), 190–198. https://doi.org/10.1111/j.1470-6431.2009.00756.x

Mühlbacher, A. C., Kaczynski, A., & Zweifel, P. (2013). Experimentelle Präferenzmessung im Gesundheitswesen mit Hilfe von Best-Worst Scaling (BWS). PharmacoEconomics German Research Articles, 11(2), 101–117. https://doi.org/10.1007/s40275-014-0012-7

Mühlbacher, A. C., Kaczynski, A., Zweifel, P., & Johnson, F. R. (2016). Experimental measurement of preferences in health and healthcare using best-worst scaling: An overview. Health Economics Review, 6(2), 1–14. https://doi.org/10.1186/s13561-015-0079-x

Nguyen, A. T., Parker, L., Brennan, L., & Lockrey, S. (2020). A consumer definition of eco-friendly packaging. Journal of Cleaner Production, 252, 119792. https://doi.org/10.1016/J.JCLEPRO.2019.119792 (Journal of Cleaner Production, 252, 119792).

Odom, M. D., Kumar, A., & Saunders, L. (2002). Web Assurance Seals: How and Why They Influence Consumers' Decisions. Journal of Information Systems, 16(2), 231–250. https://doi.org/10.2308/jis.2002.16.2.231

Oghazi, P., Karlsson, S., Hellström, D., & Hjort, K. (2018). Online purchase return policy leniency and purchase decision: Mediating role of consumer trust. Journal of Retailing and Consumer Services, 41, 190–200. https://doi.org/10.1016/j.jretconser.2017.12.007

Paetz, F., & Guhl, D. (2017). Understanding Differences in Segment-specific Willingness-to-pay for the Fair Trade Label. Marketing ZFP, 39(4), 37–46.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A Conceptual Model of Service Quality and Its Implications for Future Research. Journal of Marketing, 49(4), 41–50. https://doi.org/10.1177/002224298504900403

Park, D.-H., Lee, J., & Han, I. (2007). The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement. International Journal of Electronic Commerce, 11(4), 125–148. https://doi.org/10.2753/JEC1086-4415110405

Paul, J., Modi, A., & Patel, J. (2016). Predicting green product consumption using theory of planned behavior and reasoned action. Journal of Retailing and Consumer Services, 29, 123–134. https://doi.org/10.1016/j.jretconser.2015.11.006

Petersen, J. A., & Kumar, V. (2009). Are Product Returns a Necessary Evil? Antecedents and Consequences. Journal of Marketing, 73(3), 35–51. https://doi.org/10.1509/jmkg.73.3.035

Pookulangara, S., & Shephard, A. (2013). Slow fashion movement: Understanding consumer perceptions—An exploratory study. Journal of Retailing and Consumer Services, 20(2), 200–206. https://doi.org/10.1016/J.JRETCONSER.2012.12.002

Prakash, G., Choudhary, S., Kumar, A., Garza-Reyes, J. A., Khan, S. A. R., & Panda, T. K. (2019). Do altruistic and egoistic values influence consumers' attitudes and purchase intentions towards eco-friendly packaged products? An empirical investigation. Journal of Retailing and Consumer Services, 50, 163–169. https://doi.org/10.1016/j.jretconser.2019.05.011

Prakash, G., & Pathak, P. (2017). Intention to buy eco-friendly packaged products among young consumers of India: A study on developing nation. Journal of Cleaner Production, 141, 385–393. https://doi.org/10.1016/j.jclepro.2016.09.116

Rausch, T. M., & Kopplin, C. S. (2021). Bridge the gap: Consumers' purchase intention and behavior regarding sustainable clothing. Journal of Cleaner Production, 278, 1–15. https://doi.org/10.1016/j.jclepro.2020.123882

Sawtooth Software. (2009b). The CBC/HB System for Hierarchical Bayes Estimation Version 5.0 Technical Paper. https://www.sawtoothsoftware.com/support/technical- papers/hierarchical-bayes-estimation/cbc-hb-technical-paper-2009

Sawtooth Software. (2009a). MaxDiff Analysis: Simple Counting, Individual-Level Logit, and HB. https://sawtoothsoftware.com/resources/technical-papers/maxdiff-analysis-simple-counting-individual-level-logit-and-hb

Sawtooth Software. (2013). The MaxDiff System – Technical Paper. https://www.sawtoothsoftware.com/download/techpap/maxdifftech.pdf

Scherer, C., Emberger-Klein, A., & Menrad, K. (2018). Consumer preferences for outdoor sporting equipment made of bio-based plastics: Results of a choice-based-conjoint experiment in Germany. Journal of Cleaner Production(203), 1085–1094.

Schuman, H., & Presser, S. (1981). Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. Academic Press Inc.

Schwepker, C. H., & Cornwell, T. B. (1991). An Examination of Ecologically Concerned Consumers and Their Intention to Purchase Ecologically Packaged Products. Journal of Public Policy & Marketing, 10(2), 77–101. https://doi.org/10.1177/074391569101000205

Simon, A. (2010). Der Informationsbedarf von Patienten hinsichtlich der Krankenhausqualität. Gabler Verlag. https://link.springer.com/content/pdf/10.1007/978-3-8349-8709-9.pdf

Smith, M. D., & Brynjolfsson, E. (2001). Consumer Decision-making at an Internet Shopbot: Brand Still Matters. The Journal of Industrial Economics, 49(4), 541–558. https://doi.org/10.1111/1467-6451.00162

Statista. (2019). Die Top 10 umsatzstärksten Online-Shops im Fashionmarkt in Deutschland 2018. https://de.statista.com/prognosen/646166/top-online-shops-fashion-deutschland-ecommercedb

Statista. (2020). eCommerce - worldwide. https://www.statista.com/outlook/243/100/ecommerce/worldwide

Stöckigt, G., Schiebener, J., & Brand, M. (2018). Providing sustainability information in shopping situations contributes to sustainable decision making: An empirical study with choice-based conjoint analyses. Journal of Retailing and Consumer Services, 43, 188–199. https://doi.org/10.1016/J.JRETCONSER.2018.03.018

Taufique, K. M., & Vaithianathan, S. (2018). A fresh look at understanding Green consumer behavior among young urban Indian consumers through the lens of Theory of Planned Behavior. Journal of Cleaner Production, 183, 46–55. https://doi.org/10.1016/j.jclepro.2018.02.097

Thurstone, L. L. (1927). A law of comparative judgment. Psychological Review, 34(4), 273–286. https://doi.org/10.1037/h0070288 (Psychological Review, 34(4), 273-286).

van Loo, E. J., Caputo, V., Nayga, R. M., Seo, H.-S., Zhang, B., & Verbeke, W. (2015). Sustainability labels on coffee: Consumer preferences, willingness-to-pay and visual attention to attributes. Ecological Economics, 118, 215–225. https://doi.org/10.1016/j.ecolecon.2015.07.011

Viciunaite, V., & Alfnes, F. (2020). Informing sustainable business models with a consumer preference perspective. Journal of Cleaner Production, 242, 1–9. https://doi.org/10.1016/j.jclepro.2019.118417

Wood, S. L. (2001). Remote Purchase Environments: The Influence of Return Policy Leniency on Two-Stage Decision Processes. Journal of Marketing Research, 38(2), 157–169. https://doi.org/10.1509/jmkr.38.2.157.18847

Yadav, R., & Pathak, G. S. (2016). Young consumers' intention towards buying green products in a developing nation: Extending the theory of planned behavior. Journal of Cleaner Production, 135, 732–739. https://doi.org/10.1016/j.jclepro.2016.06.120

Yadav, R., & Pathak, G. S. (2017). Determinants of Consumers' Green Purchase Behavior in a Developing Nation: Applying and Extending the Theory of Planned Behavior. Ecological Economics, 134, 114–122. https://doi.org/10.1016/j.ecolecon.2016.12.019

Young, W., Hwang, K., McDonald, S., & Oates, C. J. (2010). Sustainable consumption: green consumer behaviour when purchasing products. Sustainable Development, 18(1), 20–31.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The Behavioral Consequences of Service Quality. Journal of Marketing, 60(2), 31–46. https://doi.org/10.1177/002224299606000203

Zhang, Z., Li, Y., Gong, C., & Wu, H. (2002). Casual wear product attributes. Journal of Fashion Marketing and Management: An International Journal, 6(1), 53–62. https://doi.org/10.1108/13612020210422464

3 Part B: Artifical Intelligence in an e-commerce context

3.1 Artificial Intelligence and applications in e-commerce

Artificial Intelligence first gathered attention in the 1950s with the Turing test (Turing, 1950): It determines a machine's ability to exhibit human-like intelligent behavior. The machine passes the test, if a human interrogator cannot reliably tell whether the responses in a natural language conversation come from a human being or a machine. However, the first work being recognized as Artificial Intelligence today was the modeling of artificial neurons copying the human nervous system in the 1940s (McCulloch & Pitts, 1943). Within literature, there is a myriad of definitions for Artificial Intelligence. Russell and Norvig (2020) intended to derive a classification scheme along two dimensions: definitions can rather focus on machines' behavior versus thought processes or focus on machines' human versus rational performance.

Table 2: Definitions of Artificial Intelligence.

	Human performance	**Rational performance**
Thought processes	**Thinking Humanly** AI are activities that humans associate with the human mind, e.g., decision-making or problem solving (Bellman, 1978)	**Thinking Rationally** AI is the "study of mental faculties through the use of computational models" (Charniak & McDermott, 1987, 1 ff.)
Human behavior	**Acting Humanly** AI is a study of "how to make computer do things at which, at the moment, people are better" (Rich & Knight, 1991, 1 ff.)	**Acting Rationally** "AI […] is concerned with intelligent behavior in artifacts" (Nilsson, 1998, p. 1)

Source: Own research based on Russell and Norvig (2020).

In the 1990s, the term Business Intelligence emerged and, in turn, in the 2000s, Business Analytics as a key analytical component within Business Intelligence established (Chen et al., 2012; Davenport, 2006). Artificial Intelligence and these sub-domains are considered the key enablers for digitalization (Brynjolfsson & McAfee, 2017). The subsequent automatization of processes as well as the rise of the Internet's commercial use (and the respective ability to track consumers' online behavior), resulted in the era of big data and an excessive *datafication* (Kelly & Noonan, 2017; Lycett, 2013) across almost every industry and the whole organizational environment. Big data is frequently described by four dimensions (*Volume, Variety, Velocity,* and *Veracity – the Four V's*) (Akter & Wamba, 2016; Chen et al., 2012): Volume refers to the magnitude of data, variety refers to the heterogeneity and noise in a dataset as well as the variety of sources where it originates from, velocity refers to the rate of data generation/delivery, and veracity refers to the authenticity of the data. Organizations swim in an expanding sea of data, which is too voluminous or too unstructured to analyze it with traditional approaches (Davenport et al., 2012).

Thus, more recently, the term (Big) Data Analytics was drawn on to describe analytical techniques to acquire intelligence from such extremely large and complex datasets (Chen et al., 2012; Gandomi & Haider, 2015) and can be viewed as a sub-process (aside from data management) of the whole process of *insight extraction* out of big data (Gandomi & Haider, 2015). Big data analytics was found to impact firm performance positively (Germann et al., 2014; Liu, 2014) and is considered a new key enabler of competitive advantage (Wamba et al., 2017). Big data analytics developed from a mere buzzword to a crucial game-changer across numerous industries such as healthcare (Liu, 2014; Raghupathi & Raghupathi, 2014), logistics as well as supply chain management (Govindan et al., 2018), security as well as public safety (Cardenas et al., 2013), and e-commerce.

Thereby, e-commerce businesses are among the earliest big data analytics adopters due to the constant perceived competitive pressure (Akter & Wamba, 2016). On an abstract level, big data analytics is – inter alia – utilized for decision-making and performance improvement purposes (see e.g., Constantiou & Kallinikos, 2015; George et al., 2014; Goes, 2014), market segmentation (see e.g., Davenport et al., 2012; Demirkan & Delen, 2013), or new product/market/business model innovations (see e.g., LaValle et al., 2011; McAfee et al., 2012) in an e-commerce context. Usually, transaction data, clickstream data, video, or voice data are utilized for these purposes (Akter & Wamba, 2016).

Hence, big data analytics is assumed to maximize business value from the big data explosion (Beath et al., 2012), i.e., it creates transactional, informational, and strategic benefits for e-commerce businesses (Akter & Wamba, 2016; Wixom et al., 2013). To create value, it makes use of a wide variety of different techniques; among the most important are text analytics (including information extraction, summarization, sentiment analysis, etc.), video and image analytics, social media analytics, and predictive analytics (Gandomi & Haider, 2015). These approaches can be used to personalize services or customize products in real-time (Koutsabasis et al., 2008), for dynamic pricing to set a competitive price (Davenport, 2006; Davenport et al., 2012), for customer engagement (Bijmolt et al., 2010), and customer service in general (Davenport, 2006; Ibrahim & Wang, 2019; Kiron et al., 2012; Lehrer et al., 2018).

3.2 Predictive analytics in the field of e-commerce

Predictive analytics is a generic term for a wide variety of different approaches to predict future outcomes based on current and historical data by revealing patterns and relationships within the data (Gandomi & Haider, 2015). Considering the latter aspects, predictive analytics approaches can be subdivided into two groups (Gandomi & Haider, 2015): While some approaches uncover historical patterns in the outcome variable(s) and extrapolate them to yield future values (e.g., time series models), others uncover interdependencies between the outcome (dependent) variable(s) and explanatory (independent) variable(s) and then generate predictions (e.g., regression or machine learning models). In an e-commerce context, predictive analytics approaches can enhance a company's success within its competitive environment (Chen et al., 2012; Fayyad et al., 1996).

Facing particularly the underlying online nature of e-commerce businesses, personal encounters with their customers along the customer journey are of uttermost importance. Thereby, despite the continuous digitalization of processes, call centers are still a critical touchpoint. Call center agents can thereby be considered customer service representatives delivering service to customers via telephone by making use of their firm's database (Aksin & Harker, 1999) and call centers still constitute the main or even only customer interface for information gathering, help desks, complaint resolution, or after-sales service (Dean, 2007; Gans et al., 2003).

Drawing on marketing literature, an organization's perceived service quality comprises outcome quality as well as interaction quality[6] (Brady & Cronin, 2001). Outcome quality, in turn, comprises – inter alia – customer's waiting times[7] (Brady & Cronin, 2001). Hence, when there is an insufficient number of call center agents available, customers may be placed on hold and thus, their waiting time increases. Additionally, individuals frequently tend to overestimate their waiting time (Hornik, 1984). This may contribute to a negative evaluation of outcome quality and further, the organization's perceived service quality[8] (Gans et al., 2003). This, in turn, decreases customer loyalty, i.e., the customer's attachment to an organization/service provider (Dean, 2007; Zeithaml et al., 1996). Simultaneously, call center managers should not only try to avoid understaffing, but further overstaffing as this causes unnecessary personnel costs. Hence, reliable and accurate call center call arrivals' forecasts with predictive analytics techniques are needed to optimally plan the call center staffing.

Within Research Paper No. 4, the predictive potential and practicability of machine learning approaches, which have been neglected by traditional forecasting literature, are investigated. Research Paper No. 4 thus not only draws on forecast accuracy as the only decision criterion, but further considers model complexity and computation time, which may be of high relevance in a practical setting. The study aims to initiate a paradigm shift away from traditional call center forecasting approaches towards Artificial Intelligence-driven methods. Research Paper No. 4 further provides a walk-through code example to implement such models in individual practical settings.

RQ4: Which model performs best with respect to forecast accuracy and practicability for call center call arrivals' forecasting?

Aside from proving the predictive power of machine learning approaches, Research Paper No. 4 reveals that models with explanatory variables are better able to capture special days (e.g., holidays), while models without explanatory variables are better able to capture ordinary weekdays. Research Paper No. 5 builds on these insights and proposes a new approach for forecasting call center call arrivals: it

[6] Brady and Cronin (2001) suggested three dimensions of service quality in total: interaction, outcome, and physical environmental quality. With respect to the call center context, physical environmental quality will be neglected.

[7] Aside from waiting times, outcome quality covers tangible evidence and valence. Interaction quality is reflected by the service consultant's attitude, behavior, and expertise (Brady & Cronin, 2001).

[8] For both outcome and interaction quality to contribute to an enhanced service quality, customers must perceive the received quality as reliable, responsive and empathetic (Brady & Cronin, 2001).

combines the advantages of approaches without explanatory variables (i.e., time series models) and with explanatory variables (i.e., machine learning as well as regression models). The proposed dynamic harmonic regression model with predictor variables thus captures both the dynamics of a time series and includes contextual information. The model is then compared to established forecasting models. Research Paper No. 5 further adds to the knowledge on relevant predictor variables in call center forecasting, particularly in a marketing context. Research Paper No. 5 thus intends to answer the following research questions:

RQ5: Can a dynamic harmonic regression model with predictor variables outperform established models? How do marketing-relevant predictor variables optimally enhance forecast accuracy?

Aside from personal encounters with their consumers, another major problem inherent to the online context of e-commerce businesses is consumers' non-purchase behavior. Research on conversion rates exploded throughout the past years (see e.g., Di Fatta et al., 2018; Gudigantala et al., 2016; McDowell et al., 2016; Moe & Fader, 2004b), as this implies untapped sales potential and can be considered a critical metric in the e-commerce context. Thereby, literature mostly investigated factors influencing consumers' purchase behavior (see e.g., Moe & Fader, 2004a, 2004b; Sismeiro & Bucklin, 2004; van den Poel & Buckinx, 2005) and non-purchase behavior with behavioral approaches (see e.g., Close & Kukar-Kinney, 2010; Kukar-Kinney & Close, 2010). In terms of non-purchase behavior, particularly online shopping cart abandonment constitutes a major challenge for e-commerce businesses. As the behavioral perspective is well understood by research (Close & Kukar-Kinney, 2010; Huang et al., 2018; Kukar-Kinney & Close, 2010), Research Paper No. 6 investigates this phenomenon with real clickstream data and compares different machine learning models to reliably predict online shopping cart abandonment. Thereby, to ensure practicability, not only prediction accuracy is considered as a decision criterion, but further model complexity and computation time. At its core, Research Paper No. 6 therefore intends to answer the following research question:

RQ6: Which model performs best with respect to accuracy and practicability for online shopping cart abandonment prediction?

Table 3 summarizes the publication status of all research papers included in Part B and Figure 3 summarizes content, method, and research question(s) of each research paper in Part B.

Table 3: Publication status of research papers in Part B.

	Author(s) & Year	Title	Medium	Status
Research Paper No. 4	Albrecht, T., Rausch, T. M. & Derra, N. D. (2021)	Call me maybe: Methods and practical implementation of artificial intelligence in call center arrivals' forecasting	Journal of Business Research, 123	Published

	Author(s) & Year	Title	Medium	Status
Research Paper No. 5	Rausch, T. M., Albrecht, T., & Baier, D.	Beyond the Beaten Paths of Forecasting Call Center Arrivals: On the Use of Dynamic Harmonic Regression with Predictor Variables	Journal of Business Economics	Under Review (First Revision)
Research Paper No. 6	Rausch, T. M., Derra, N. D., & Wolf, L. (2021)	Predicting online shopping cart abandonment with machine learning approaches	International Journal of Market Research (Forthcoming)	Published

Part B: Artificial Intelligence in an e-commerce context					
Prediction problems 1 and 2: Call center call arrivals' forecasting				**Prediction problem 3: Online shopping cart abandonment**	
Research Paper No. 4: Call me maybe: Methods and practical implementation of artificial intelligence in call center arrivals' forecasting		Research Paper No. 5: Beyond the Beaten Paths of Forecasting Call Center Arrivals: On the Use of Dynamic Harmonic Regression with Predictor Variables		Research Paper No. 6: Predicting online shopping cart abandonment with machine learning approaches	
Content	Investigates the trade-off between prediction accuracy and practicability (i.e., model complexity as well as computation time) of machine learning models and identifies the most accurate model for call center call arrivals' forecasting	**Content**	Novel dynamic harmonic regression model with predictor variables, which combines the strengths of models without explanatory variables (i.e., time series models) and with explanatory variables (i.e., machine learning and regression models), is proposed	**Content**	Investigates the trade-off between prediction accuracy and practicability (i.e., model complexity as well as computation time) of machine learning models and identifies the most suitable model for online shopping cart abandonment
Method	Machine learning models (compared to time series models)	**Method**	Own proposed dynamic harmonic regression model with predictor variables (compared to established time series and machine learning models)	**Method**	Machine learning models
Research question(s)	Which model performs best with respect to forecast accuracy and practicability for call center call arrivals' forecasting?	**Research question(s)**	Can a dynamic harmonic regression model with predictor variables outperform established models? How do marketing-relevant predictor variables optimally enhance forecast accuracy?	**Research question(s)**	Which model performs best with respect to accuracy and practicability for online shopping cart abandonment prediction?

Figure 3: Summary of research papers in Part B.

3.2.1 Research Paper No. 4: Call me maybe: Methods and practical implementation of artificial intelligence in call center arrivals' forecasting

Authors: Albrecht, T., Rausch, T. M., & Derra, N. D. (2021)

Published in: Journal of Business Research, 123, 267-278

DOI: https://doi.org/10.1016/j.jbusres.2020.09.033

Abstract: Machine learning (ML) techniques within the artificial intelligence (AI) paradigm are radically transforming organizational decision-making and businesses' interactions with external stakeholders. However, in time series forecasting for call center management, there is a substantial gap between the potential and actual use of AI-driven methods. This study investigates the capabilities of ML models for intra-daily call center arrivals' forecasting with respect to prediction accuracy and practicability. We analyze two datasets of an online retailer's customer support and complaints queue comprising half-hourly observations over 174.5 weeks. We compare practically relevant ML approaches and the most commonly used time series models via cross-validation with an expanding rolling window. Our findings indicate that the random forest (RF) algorithm yields the best prediction performances. Based on these results, a methodological walk-through example of a comprehensive model selection process based on cross-validation with an expanding rolling window is provided to encourage implementation in individual practical settings.

Keywords: artificial intelligence; machine learning; call center forecasting; predictive analytics

1 Introduction

Artificial Intelligence (AI) is considered the key enabler for the digitalization of a company in a broad spectrum of areas (Brynjolfsson & McAfee, 2017). Today, in the course of increasing availability of data, affordable as well as scalable processing power, and the development of advanced machine learning (ML) techniques, AI is about to radically transform how firms make decisions (Agrawal et al., 2019). It is expected to facilitate the internal decision-making processes of organizations by making it smarter, faster, and overall more efficient. To benefit from this potential competitive advantage, companies need to identify existing domain problems, find compatible AI solutions, and put an implementation concept into practice (Overgoor et al., 2019). This requires a thorough understanding of the task-specific capabilities and feasibility of AI methods like ML. So far, a lack of expertise in this area paired with a high level of perceived complexity is often preventing the implementation of ML solutions in practical settings (Tambe et al., 2019).

Particularly the interaction of companies with external stakeholders, such as customers, is about to be fundamentally transformed by AI (Kaplan & Haenlein, 2019). Fueled by an almost unlimited flow of information about their customers, service-oriented companies in particular, can capitalize on AI-driven decision support. Based on latent characteristics and previous customer behavior, ML techniques can predict future interactions (Wedel & Kannan, 2016). One of the most prevalent and dominant points of interaction between many organizations and their customers and therefore, critical for managing customer experience, are call centers or customer service centers (Whiting & Donthu, 2006). To constantly provide high service quality in the form of short waiting times at this touchpoint, a sufficient number of call center agents is needed (Atlason et al., 2008). Consequently, the process of predicting call arrival volumes and deciding on the required staffing level is a critical success factor in this area. In this connection, the capabilities of innovative ML techniques promise more flexible and precise predictions and thus, the possibility of enhanced organizational planning and better customer service.

Despite the encouraging prospects for service improvement and cost savings, a comprehensive understanding of the potential of ML models for creating additional value in call center forecasting is lacking. In order to gain more profound insights into the performance and practicability of such AI-driven models in this context, research comprising a methodological perspective with a focus on prediction accuracy as well as a practical angle on the selection and implementation of models is required.

This study proposes a two-step approach that, in the first step, provides a thorough understanding of the forecast accuracy of ML methods in call arrival forecasting and, in the second step, makes the underlying process of method comparison and selection feasible to decision-makers in practice. Specifically, we conduct an in-depth analysis of the forecast accuracy of viable ML models based on the call arrival data of a real German online retailer. Using two different datasets, i.e., the customer support and customer complaints queue of the corresponding call center, we perform a comprehensive method comparison opposing selected ML models to the three most commonly used time series models in this field. In the second step, we provide a methodological walk-through example for a valid model selection process based on cross-validation with an expanding rolling window. We illustrate the practical implementation of the process in a programming environment that is accessible to non-machine learning experts and practitioners using the random forest (RF) algorithm as the best-performing model for an in-depth example.

This paper therefore aims to present a starting point for shifting traditional call center forecasting towards a paradigm drawing on AI-driven methods. By systematically evaluating the predictive potential of ML models in comparison to commonly used methods, new sophisticated but yet applicable models for practical use are identified. In a business setting, following the explicated implementation in a reproducible programming environment is supposed to empower practitioners to develop insights on the use of ML for forecasting call center arrivals in individual data environments.

The remainder of this paper is structured as follows: In Section 2, we present the theoretical background of AI-driven methods in customer analytics and review the state of research in call center arrivals' forecasting, before adequate ML models for this field are introduced. Subsequently, in Section 3, we describe the methodology of our research. In Section 4, we present the results of our analysis for two different customer service channels and in Section 5, we discuss the theoretical contribution and the limitations of our study. We then illustrate the implementation of the best-performing RF model by giving a detailed code and walk-through example and demonstrate methodological as well as practical implications of the proposed approach. Finally, Section 7 presents a summary and concluding remarks.

2 Theoretical background

2.1 Artificial intelligence in customer analytics

For businesses, the strategic challenge of understanding and managing customer relationships is becoming increasingly important and demanding at the same time. While organizations today have easy access to enormous amounts of data about their customers, extracting relevant information to support prospective decision-making and thus, standing out from competitors in the long term has become a difficult hurdle to overcome for many of them (Kitchens et al., 2018). In the course of these changing market dynamics, businesses slowly realize the potential of AI in predictive analytics to enhance organizational decision-making by forecasting customer-related data and, therefore, effectively infer their future behavior (Huang & Rust, 2018). Predictive analytics techniques generally comprise statistical models and other empirical methods aimed at creating predictions as well as approaches for assessing the quality of those predictions in practice (Shmueli & Koppius, 2011). More recently, ML as a subset of AI has been added to the domains contributing effectively to business prediction problems as they provide a way to handle complex problems by forecasting future data based on more extensive sets of historical values (Chen et al., 2012). In literature, innovative ML approaches have been successfully applied to various customer analytics problems such as customer preferences analysis

(Yang & Allenby, 2003), customer retention (Donkers et al., 2003), and customer profitability management (Reinartz et al., 2005).

However, so far, the practical implementation of ML models in predictive customer analytics is limited (Wedel & Kannan, 2016). Drawing on the early distinction between forecasting methods and forecasting systems proposed by (Harrison & Stevens, 1976) may explain this slow adoption. While the former transforms input data into output information in a mere technical way, the latter in addition includes the people concerned with the forecast and the resulting actions. Based on that view, the evaluation and selection of a forecasting system explicitly go well beyond the accuracy of its prediction model and includes meaning and usability in practical implementation. In terms of this applicability, many ML approaches still exhibit shortcomings as they do not provide much insight into the influence and dynamics of the underlying factors that lead to the prediction results (Martens et al., 2011; Najafabadi et al., 2015). Due to this lack of comprehensibility and interpretability, many ML techniques are commonly considered as black box models (Doshi-Velez & Kim, 2017; Guidotti et al., 2019). Moreover, such models are frequently perceived as complex regarding the implementation. A high number of hyperparameters gives models the flexibility of adapting to a multitude of business problems but, at the same time, makes it complex for the user to build and optimize the ML algorithm. This especially applies to the broad class of artificial neural networks (Bergstra et al., 2011; Paliwal & Kumar, 2009). For the above reasons, other categories of ML approaches come into the focus for practical use.

Widely established methods like support vector machines and Bayesian approaches promise ease of use while maintaining good performance levels on data sets characterized by moderate complexity (Arora et al., 1998; Verbeke et al., 2011). Tree-based models, and in this field especially ensemble learning methods like RF and gradient boosting, gained popularity for their robustness and flexibility in modeling input–output relationships of various types and volumes of highly complex data (Fang et al., 2016; Lemmens & Croux, 2006). Research found them to provide high prediction accuracy as well as descriptive results in diverse customer analytics problems such as churn analysis (Burez & van den Poel, 2009) and credit risk management (Fantazzini & Figini, 2009). In addition, a small number of hyperparameters makes their construction, customization, and optimization more manageable and comprehensible (Breiman, 2001).

2.2 Call center arrivals' forecasting

In recent years, the role of call centers has fundamentally changed in many organizations and across all industries. While call centers previously only had an information function which did not exceed simple order processes, nowadays, more and more complex tasks and customer demands need to be fulfilled across multiple communication channels using modern digital technology (Aksin et al., 2007). However, instead of experiencing declining importance in the course of this transformational process, the opposite is the case. Call centers are increasingly transforming into customer interaction centers that form the basis for an efficient and value-oriented customer relationship management (Gans et al., 2003). They constitute an interface to the customer and provide complex services, while, at the same time, giving companies the opportunity of collecting large amounts of otherwise inaccessible customer data (Ibrahim et al., 2016). Subsequently, it is possible to anticipate customer needs and behavior through data analysis and forecasting techniques (Taylor, 2008). Based on those insights, internal processes and external expectations can be aligned to optimize business performance as well as customer experience.

One of the most important internal processes in call centers is the staffing of agents as customer service representatives who directly handle tasks such as order taking, complaint resolution, information, and help desk functions as well as after-sales and supplementary services (Dean, 2007; Koole & Pot, 2005). While overstaffing results in high personnel costs, understaffing can lead to extended waiting times for customers and consequently causes lower perceived service quality, decreasing customer satisfaction, and a lack of customer loyalty (Brady & Cronin, 2001). To determine the optimal staffing level, an accurate and robust prediction of call arrival volumes based on historical data is needed (Weinberg et al., 2007). Hence, the search for appropriate forecasting methods is the focus of scholars and practitioners alike. However, preceding literature so far mainly investigated traditional statistical models without taking into account the substantial changes coming along with the transforming role of call centers in organizations (Gans et al., 2003). Today, the increasing volume and variety of data through a multitude of channels as well as the necessity of real-time analysis and predictions call for more flexible and powerful methods.

Call center arrivals are count data limited to non-negative integers. Such discrete data are frequently estimated as Poisson arrival rates (see e.g., (Cezik & L'Ecuyer, 2008; Taylor, 2012). However, with arrival rates not being easily predictable, other researchers point out ascertained randomness of arrivals in real call centers (see e.g., (Aksin et al., 2007; Shen & Huang, 2008). Generally, call center arrivals data exhibit specific characteristics and challenges that affect the forecasting process. Firstly, an important feature of call arrival rates is their time dependence that typically manifests itself in intraday (or sub-daily), daily, weekly, monthly, or yearly seasonalities as repeating patterns in the arrival counts (Ibrahim et al., 2016). Secondly, the data are often high-dimensional and sensitive to contextual factors. Hence, additional information like holidays, promotional activities, and other special events may improve model predictions by indicating variations and outliers in the data (D. Barrow & Kourentzes, 2018). Thirdly, procedural characteristics are affecting the forecasting of incoming calls, such as (a) the specific call type (e.g., complaints, order taking, or after-sales service) associated with the forecast, (b) the length of forecast intervals, which may commonly range from monthly or weekly to daily or even sub-daily (i.e., hourly, half-hourly etc.) time spans, and (c) the period between the creation of the forecast and the first interval of the prediction, i.e., the lead time. Lead time is an organizational parameter resulting from staffing regulations and is assumed to strongly affect forecast accuracy as more recent data promise better predictions (Aksin et al., 2007; Rausch & Albrecht, 2020). Given these properties, the need for methods with high modeling flexibility, while being able to handle time dependencies and complex data structures, becomes evident.

With time dependence often being considered as one of the predominant features of the call arrival data, common forecasting techniques in research mostly originate from the field of time series analysis with call arrivals being a set of contiguous, dependent observations $y(t) = 0, 1, 2, ...$, each one being recorded sequentially at time t (G. E. Box et al., 2015). The most widely investigated and compared methods in literature include simple stationary time series models as well as the nonstationary seasonal autoregressive integrated moving average (ARIMA) model (G. E. P. Box & Jenkins, 1970), Holt Winters' exponential smoothing models (Holt, 2004; Winters, 1960), and random walk methods (Taylor, 2008). While ARIMA and exponential smoothing provide sophisticated complementary solutions to the general forecasting problem (Hyndman & Athanasopoulos, 2018) and constitute the most commonly used approaches in call center forecasting due to their high prediction accuracy (Andrews & Cunningham, 1995; D. Barrow & Kourentzes, 2018; Mabert, 1985; Taylor, 2012; Thompson & Tiao, 1971b), the random walk model is frequently utilized as a benchmark within literature due to its naïve forecasts and its informative value for model comparisons (Taylor, 2008). Besides, regression analysis in the form of generalized linear models (GLM), linear fixed-effects, random-effects, and mixed-effects models is implemented for call arrivals' forecasting (Avramidis et al., 2004; Ibrahim & L'Ecuyer, 2013; Nelder & Wedderburn, 1972).

In contrast, research on ML techniques in call center arrivals' forecasting is still in its infancy. (Ebadi Jalal et al., 2016) first indicate time-sensitive ML models to be eligible for forecasting call volumes in call centers. To improve short-term accuracy in call arrivals' forecasting, (D. K. Barrow, 2016) developed a hybrid method adjusting seasonal moving average predictions by means of nonlinear artificial neural networks and found it to outperform traditional time series models like ARIMA and Holt Winters'. Moreover, ML is shown to be capable of modeling complex outliers and thus, to improve call arrival prediction accuracy and to yield better results than ARIMA and an innovation state space model (ETS) (D. Barrow & Kourentzes, 2018). Recently, (Rausch & Albrecht, 2020) investigated RF algorithm as another ML method in their comparison of novel time series and regression models for call center arrivals forecasting. RF was found to yield higher prediction accuracy for nearly all of the considered lead time constellations. Despite first promising findings and the investigation of several approaches, current research lacks a comprehensive understanding of the full capabilities of ML in call center forecasting. To close this gap, an extensive assessment of the forecast accuracy of ML models in comparison to the most commonly used methods is still to be done. However, according to comparisons of common methods conducted on call center data, the selection of the best forecasting method can ultimately be highly dependent on the characteristics of the specific prediction problem (Andrews & Cunningham, 1995; Taylor, 2008). Therefore, a feasible process of model comparison and selection needs to be established to give methodological guidelines to practitioners and to match the set of researched forecasting methods with those considered in practice. Today, although a lot of progress has been made regarding the development of advanced methods, call arrivals' forecasting in real business environments is frequently still done based on experience or ordinary stochastic models with limited predictive capabilities (Ibrahim et al., 2016).

2.3 Machine learning approaches

Models from the field of ML are assumed to improve call center arrivals' forecasting and extend the range of feasible methods by providing additional robustness and accuracy to predictions. As the practicability of models play a central role in this field of application, non-parametric ML algorithms, that are comprehensible and comparatively easy to implement, such as tree-based models, k-nearest neighbor (KNN) algorithm, and support vector machines, come to the fore (Coussement & van den Poel, 2008; Li et al., 2010; Singh et al., 2017).

2.3.1 Bagging: Random forest

Tree-based methods are frequently utilized in business prediction problems since they yield desirable accuracies despite their ease of use (Breiman, 2001). In bagging, successive decision trees are grown independently from earlier trees, i.e., each tree is constructed using a bootstrap sample of the data (Breiman, 1996). A subclass of bagging methods are RFs, as proposed by (Breiman, 2001), which add an additional layer of randomness to bagging and change how the trees are constructed. Thereby, non-parametric the RF algorithm is one of the most widely used ML algorithms, supported by its robustness towards outliers and its moderate computation time compared with boosting and other bagging methods (Breiman, 2001).

The algorithm draws n_{tree} bootstrap samples from the training data and then grows an unpruned regression tree for each bootstrap sample by randomly sampling m_{try} of the predictors at each node and choosing the best split among them. More formally, the resulting RF is an ensemble of B trees $\{T_1(X), \dots, T_B(X)\}$, where $X = \{x_1, \dots, x_p\}$ is a p-dimensional vector of predictors associated with a dependent variable; the ensemble produces B outputs $\{\hat{y}_1 = T_1(X), \dots, \hat{y}_B = T_B(X)\}$, where $\hat{y}_b, b = 1, \dots, B$ is the prediction for the dependent variable by the bth tree (Svetnik et al., 2003). The outputs of all n_{tree} trees are aggregated to produce one final prediction $\hat{Y}$; for regression trees it is the average of the single tree predictions (Liaw & Wiener, 2002). I.e., the RF prediction is the unweighted average over the ensemble:

$$\hat{Y} = \frac{1}{B} \sum_{b=1}^{B} \hat{y}_B \left(T_B(X)\right)$$

To tune the hyperparameters, an estimate of the error rate based on training data can be obtained: at each bootstrap iteration, the data which is not in the bootstrap sample, i.e., the out-of-bag (OOB) data n, is predicted by using the tree grown with the bootstrap sample. Then the OOB predictions are aggregated and the error rate is calculated (Liaw & Wiener, 2002). In each bootstrap training set, about one-third of the sample is left out, i.e., is used for OOB predictions (Breiman, 2001).

2.3.2 Boosting: Gradient boosting machines

In contrast to bagging, boosting constructs successive weak learners (e.g., decision trees) to produce a final strong learner. Each sequentially added weak learner intends to correct the preceding learners (Schapire, 1990). Thereby, gradient boosting (machines) fits the new predictor or learner to the residual errors made by the preceding predictors or learners and uses gradient descent to identify the errors in the previous predictions, i.e., gradient boosting allows the optimization of an arbitrary differentiable loss function (Friedman, 2001, 2002). Formally, J_m are the number of leave and the tree partitions the input space into J_m joint regions $R_{1m}, \dots, R_{J_m m}$ and predicts a constant value in each region. γ_{im} is the multiplier chosen as an optimal value for each of the tree's regions to minimize the loss function L. Then the generic gradient tree boosting model can be defined as

$$\mathrm{F_m(x)}=\mathrm{F_{m\text{-}1}(x)} + \sum_{i=1}^{J_m} \gamma_{im}\, 1(x \in R_{im}), \quad \text{with } \gamma_{\mathrm{im}} = \arg\min_{\gamma} \sum_{x_i \in R_{im}} L(y_i, F_{m-1}(x_i) + \gamma).$$

Since gradient boosting frequently leads to overfitting, regularization techniques can be included to constrain the fitting procedure. E.g., dropout regularization – inspired by neural networks in a deep learning context – grows consecutive trees from the residual errors of a subset or sample of previous trees instead of using all previous trees (Rashmi & Gilad-Bachrach, 2015).

2.3.3 K-nearest neighbor

The KNN algorithm is frequently considered due to its simplicity in comparison with other ML approaches. The algorithm was first formalized by (Cover & Hart, 1967) for classification tasks: given an unlabeled instance, the algorithm finds a group of k most similar objects (or nearest neighbors respectively) given its features by computing the distance $\mathrm{d}(.,.)$ (e.g., Euclidean distance) between them and further, assigns a class label which matches the class of the majority of the k neighbors. This concept can easily be extended to regression tasks where the output is the average of the k nearest neighbors, i.e.,

$$\hat{\mathrm{Y}} = \frac{1}{\mathrm{K}} \sum_{\mathrm{i}=1}^{\mathrm{K}} \mathrm{y_i}$$

where $\mathrm{y_i}$ is the ith case of the nearest neighbors.

2.3.4 Support vector regression

Suppose we are given a space of input patterns $\{(x_1, y_1), \dots, (x_k, y_k)\} \subset \mathcal{X} \times \mathbb{R}$ with y_k being the output vectors and x_k are the input vectors. The basic support vector machine is a non-probabilistic binary linear classifier and it non-linearly maps input vectors into a higher dimensional feature space in which a linear decision surface, i.e., a separating hyperplane, is constructed (Cortes & Vapnik, 1995; Vapnik, 2000). Thus, its representation of the training data as points in the feature space is separated into categories by the hyperplane and predictions of new instances are classified into those categories. The main aim in ε-support vector regression (SVR) (Vapnik, 2006) is based on the same principles but with minor differences: the function $f(x)$ should have at most ε deviation from the actual targets for the training data and simultaneously, should be as flat as possible (Smola & Schölkopf, 2004). In the linear and most basic case, f is taking the form

$$f(x) = \langle \omega, x \rangle + b \quad \text{with } \omega \in \mathcal{X}, b \in \mathbb{R}$$

where $\langle \cdot,\cdot \rangle$ is the dot product in the space of input patterns $\mathcal{X}$. To ensure flatness, a small ω can be obtained by a convex optimization problem:

$$\text{minimize } \frac{1}{2}\|\omega\|^2$$

$$\text{subject to } \begin{cases} y_t - \langle \omega, x_i \rangle - b \leq \varepsilon \\ \langle \omega, x_i \rangle + b - y_i \leq \varepsilon \end{cases}$$

It assumes that function f approximates all pairs $\langle x_i, y_i \rangle$ with ε precision. Slack variables ξ_i, ξ_i^* can cope with such otherwise infeasible constraints of the optimization problem. Moreover, kernels can be used to make SV algorithms nonlinear by transforming the data into a higher dimensional feature space (Smola & Schölkopf, 2004).

3 Methodology

3.1 Preliminary data analysis

We analyze call center data of a leading German online retailer for fashion that were gathered and selected iteratively and in close exchange with the local data experts and department managers. Overall, the retailer's call center comprises four different queues: customer complaints, customer support, personal consultation service, and order taking. In this paper, we investigate two datasets describing the call arrival volume of the customer support and customer complaints queue. Both are open from 7 a.m. to 10 p.m. from Monday through Saturday. The half-hourly datasets comprise 31,410 observations or 174.5 weeks of data from January 2, 2016 to May 7, 2019. One day comprises 30 observations, one week consists of 180 observations, and one year comprises 9367.5 observations considering leap years. We exclude two weeks of data (or 360 observations) since these values are missing due to an internal system change for interval capturing.

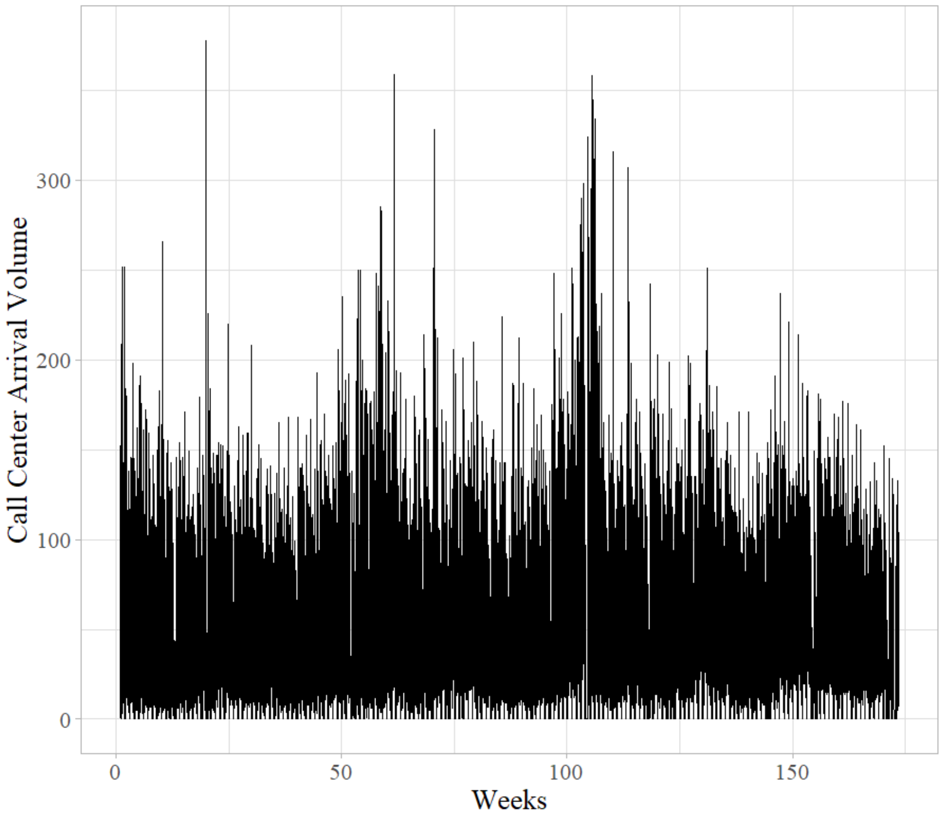

Figure 1: Overall call arrival volume of customer support queue.

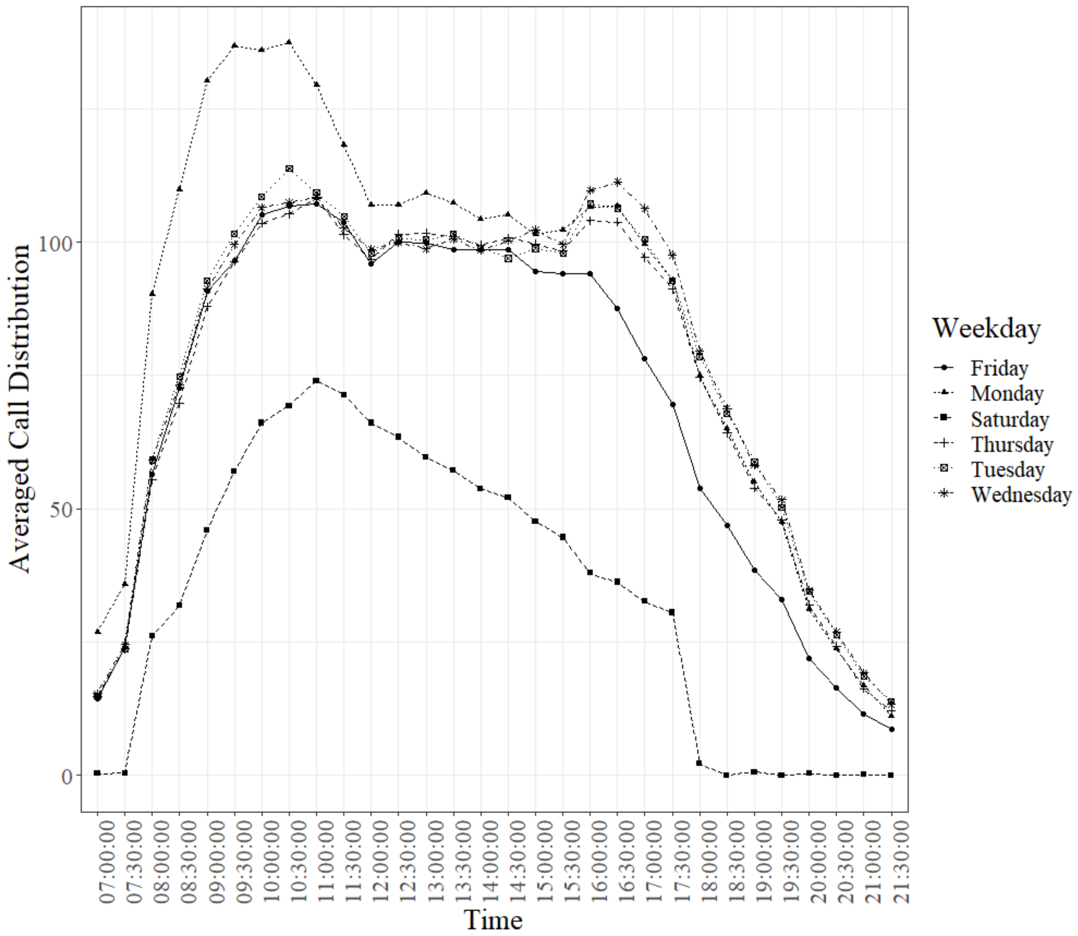

Figure 2: Averaged call distribution per day for customer support queue.

For the customer support queue, the maximum number of call arrivals per half hour is 378, and the data comprise 2218 zeros, i.e., intervals without call arrivals. The customer support data are overdispersed, exhibiting a mean of 70.9539 and a variance of 2181.6742. We conducted an Augmented Dickey Fuller (ADF) test to check whether the data have unit root and hence, are nonstationary: we cannot reject the null hypothesis of unit root in the data with a p-value of 0.9798 at lag order 9360 (value of test statistic -0.5469) and thus, assume that our data are nonstationary. Consequently, we have to apply time series decomposition to our time series models. Drawing on seasonal-trend decomposition based on Loess (STL) (Cleveland et al., 1990), the time series is detrended and deseasonalized resulting in a seasonal component $\hat{S}_t$ and a seasonally adjusted component $\hat{A}_t$, i.e., the data without a seasonal component. The latter can be forecasted with any non-seasonal forecasting method, whereas the seasonal component is forecasted by using the last period of the estimated component, i.e., a seasonal naïve method. Finally, inverting the decomposition's transformations yields the forecasts of the original time series (Brockwell et al., 2002).

Figure 1 depicts the arrival volume of the customer support queue during the 174.5 weeks of our data. Apparently, the call arrival volume remains more or less constant throughout the considered period. With respect to the averaged call distribution per day in Figure 2, Mondays are the busiest days with an extremely high peak in the morning hours. The remaining weekdays exhibit a relatively similar course with a peak in the morning and a second peak during the afternoon. In contrast, there are few call arrivals on Saturdays.

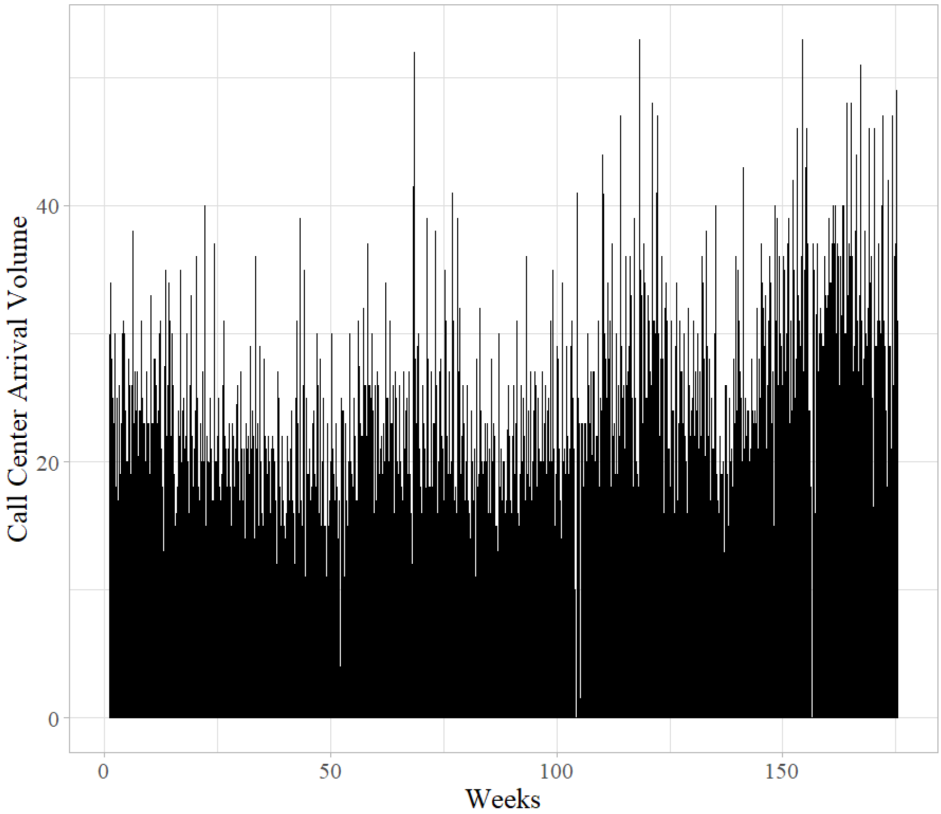

Figure 3: Overall call arrival volume of customer complaints queue.

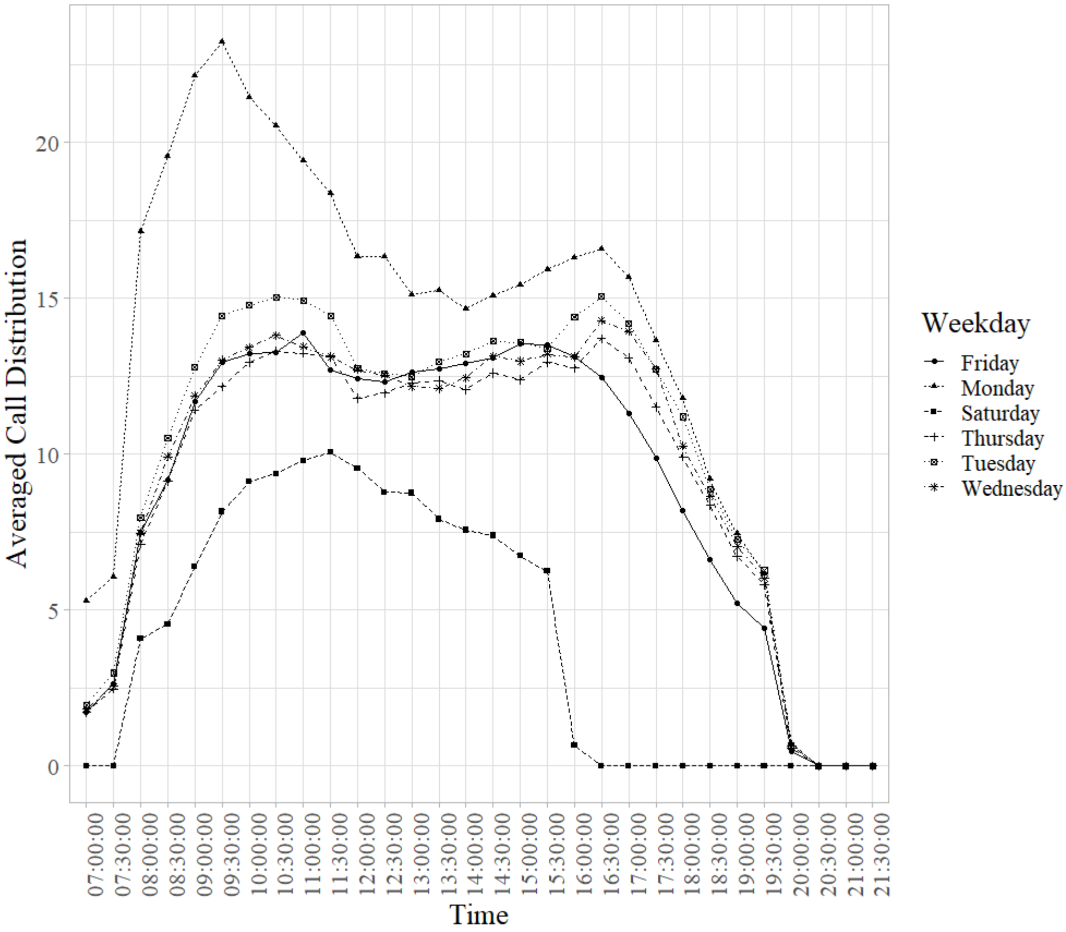

Figure 4: Averaged call distribution per day for customer complaints queue.

Regarding the customer complaints queue, the maximum number of call arrivals per interval is 53, and the dataset contains 6551 intervals without call arrivals. Since we cannot reject the null hypothesis of the ADF test with a p-value of 0.7905 at lag order 9360 (value of test statistic -1.5009) we assume our data to have unit root and, consequently, to be nonstationary. Accordingly, time series decomposition is applied to the time series model. Similar to the customer support queue, Figure 3 shows the overall arrival volume of the customer complaints queue: the call arrival volume remains relatively constant over time, but there is a slight increase towards the end of the dataset. Figure 4 reveals that the customer complaints' averaged call distribution per day is similar to the customer support queue on a lower level.

We model predictor variables (summarized in Table 1) to yield more accurate forecasts. Largely, our variables align with those of extant literature such as weekdays and billing periods (Aldor-Noiman et al., 2009) or holidays and catalog mailings (Andrews & Cunningham, 1995).

Table 1: Predictor variables.

Variable	Description
Time-of-the-day	Nominal variable capturing the time-of-the-day-effect; 30 half-hourly values ranging from 7 a.m. to 9:30 p.m.

Variable	Description
Day-of-the-week	Nominal variable capturing the day-of-the-week-effect; six values ranging from Monday to Saturday
Holiday	Nominal variable capturing the effect of German public holidays; 16 values for public holidays and ordinary weekdays
Day after holiday	Dummy variable capturing the effect of days after German public holidays, two values for days after holidays and ordinary weekdays
Outlier	Nominal variables capturing the effect of outliers; four values for extreme outliers as well as outliers (marked by the management), days on which the call center is closed, and ordinary weekdays
School holidays	Metric variable capturing the effect of German school holidays; specifying the number of German states having school holidays
Year	Nominal variable capturing the effect of busier seasons; eight values for semiannual sections from January 2016 to May 2019
CW0-3	Four dummy variables capturing the effect of catalog mailings on the first weekend, as well as the first, the second, and the third week after release
MMail1-2, MPost1-2, DMail1-2	Six dummy variables capturing the effect of reminders via e-mail (MMail) as well as via mail (MPost) and due date e-mails (DMail) on the day of delivery and the day after

3.2 Research design

To evaluate the predictive power of adequate ML approaches and to ensure the practical value of our study, we follow a two-step approach. It comprises the analysis of prediction performance in the form of a method comparison in line with extant forecasting research (see e.g., (Cao & Parry, 2009; Taylor, 2008) and, as proposed by (Buitinck et al., 2013), an in-depth walk-through example of the process of model comparison and selection to make the practical implementation accessible to decision-makers and non-experts.

In the first step, we conduct a model comparison of selected ML methods, presented in Section 2.3 (i.e., gradient boosting with dropout (GBD), gradient boosting with L1 and L2 regularization (GBR), KNN, RF, and SVR) with the three most commonly used time series models identified in Section 2.2 (i.e., ARIMA, ETS, and RW, for further formal information on these time series approaches readers are referred to the Appendix). The included methods summarized in Table 2 cover sophisticated ML and time series models as well as standard benchmark techniques. The model performance is evaluated based on the two datasets described in Section 3.1, and we include four different lead times in our experimental setup (three weeks, two weeks, one week, and no lead time from the forecast origin). This is done to validate our results as well as to assess the flexibility of the investigated models in an authentic forecasting situation that is comparable to real call center settings with specific organizational requirements like staffing regulations. We thereby aim to provide an extensive and robust assessment of the prediction accuracy of feasible ML models in call center arrivals' forecasting.

For model validation, we apply cross-validation with an expanding rolling window. Thereby, the initial model is fitted with its optimized hyperparameters using 118 weeks or 21,270 observations respectively from January 2, 2016 to April 7, 2018 as training data. We then predict one week or 180 observations respectively (i.e., forecast horizon $h = 180$). For the next iteration k, we roll the training data one week forward, re-optimize the model's hyperparameters or re-estimate the model respectively, and predict one week further. We repeat this step 52 times, i.e., for one year, and thus, $k = 52$. As stated earlier, we have to exclude two weeks of data from October 22, 2018 to November 4, 2018 and thus, we predict 9,000 observations. We evaluate the models' performance by comparing the predictions with the actual values, i.e., the test data, and hence, compute forecast accuracy.

As performance measures regarding forecast accuracy, we draw on the mean absolute error (MAE) and the root mean squared error (RMSE)

$$\text{MAE} = \frac{1}{T}\sum_{i=1}^{T}|Y_i - \hat{Y}_i| \qquad \text{RMSE} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(Y_i - \hat{Y}_i)^2}$$

where the test subset is given by Y_i, the predicted values are $\hat{Y}_i$, and T is the number of predicted values. Both error measures are frequently utilized by literature to determine accuracy (see e.g., (Aldor-Noiman et al., 2009; Ibrahim et al., 2016; Taylor, 2008; Weinberg et al., 2007) since they are easy to interpret and further, scale-dependent and therefore, suitable to compare forecasts on the same scale. Complementary, we report the computation time of both the benchmark time series models as well as the ML approaches to capture computational complexity and add practical value to the results.

Table 2: Models for comparison.

Model type	Model	Description
ML approaches	GBD	Algorithm builds an ensemble of weak tree learners, minimizes the model's loss by adding weak learners sequentially using a gradient descent like procedure, and randomly drops boosting tree members
	GBR	Algorithm builds an ensemble of weak linear base learners and utilizes L1 (Lasso Regression) as well as L2 (Ridge Regression) regularization
	KNN	Algorithm predicts an observation by averaging the values of the k nearest neighbors
	RF	Algorithm builds an ensemble of decision trees using a bootstrap sample of the data for each tree and averages the aggregated prediction of the trees

Model type	Model	Description
	SVR	Algorithm builds a separating hyperplane into the feature space of output and input vectors which should have at most ε deviation from the actual targets and should be as flat as possible
Time series models	STL + ARIMA	Time series is decomposed based on the Loess procedure and the seasonally adjusted component is forecasted based on the time series' lagged values and lagged errors
	STL+ETS	Time series is decomposed based on the Loess procedure and the seasonally adjusted component is forecasted based on previous level and error
	STL + RWDRIFT	Time series is decomposed based on the Loess procedure and the seasonally adjusted component is forecasted based on the time series' last observation and the average of changes between consecutive observations

Note: ARIMA = autoregressive integrated moving average, ETS = error, trend, seasonal (innovation state space model), GBD = gradient boosting with dropout, GBR = gradient boosting with regularization, KNN = k-nearest neighbor, RF = random forest, RWDRIFT = random walk with drift, STL = seasonal-trend decomposition based on Loess, SVR = support vector regression.

In the second step, in Section 6, we provide a methodological walk-through example for a valid model selection process based on cross-validation with an expanding rolling window. By illustrating different sequences of the implemented programming code used in the experimental design of the first step, we conduct the comparison and selection of the most suitable forecasting method comprehensible to organizational decision-makers and detach the study's value from specific characteristics of our datasets by making the implemented approach reproducible. Additionally, we aim to provide further evidence for the practical applicability of adequate ML algorithms in call center forecasting. Therefore, we do not only describe the generic programming of time series cross-validation with an expanding rolling window but further give detailed insights into the implementation of RF algorithm as the best-performing ML model in our preceding analysis. We also provide guidance on how to measure MAE, RMSE, and computation time in the process. For the methodological walk-through, we make use of the open-source statistical programming language R (Ihaka & Gentleman, 1996). Drawing on the combined results of both method evaluation and overall implementation process, we then derive practical implications for organizations.

4 Results

Drawing on the results for the customer support queue in Table 3 and Table 4, the RF algorithm outperforms the remaining approaches in every lead time constellation: with respect to both MAE and RMSE, the model yields the most accurate forecasts. The GBD, GBR, and SVR models yield comparable results, whereas the KNN approach was the most inaccurate forecasting method. Generally, every considered ML approach is superior to the benchmark time series models for all lead time constellations (except for the KNN method). Among the time series models, the ETS model is the best-performing approach. Overall, the models' performances worsen slightly with increasing lead time.

Regarding computation time, the RWDRIFT model was excelling with an estimation time of 39 seconds[9] for 52 iterations of the expanding rolling window. The remaining time series models yield comparable low computation times with 142.41 s for ETS and 1260.94 s for ARIMA. The AI-driven methods are computationally more intensive with 61,423.71 s estimation time for GBR, 93,861.33 s for KNN, 171,380.33 s for SVR, and 184,367.70 s for GBD. With 75,185.62 s for the estimation procedure of the rolling window, the RF algorithm provides an acceptable trade-off between accuracy and computation time: for the prediction of one iteration k (i.e., of the forecast horizon $h = 180$ observations), the model takes 24.1 min.

Table 3: MAE results for customer support arrivals' forecasts.

	Lead Time			
Model	**No lead time**	**One week**	**Two weeks**	**Three weeks**
GBD	13.4601	13.6603	13.9540	14.2203
GBR	12.9393	13.1488	13.3987	13.7386
KNN	18.2068	18.8704	19.2332	19.8064
RF	**11.7544**	**11.8129**	**12.0648**	**12.8134**
SVR	13.2325	13.2063	13.2256	13.6019
STL+ARIMA	14.5263	14.7407	15.5448	15.8520
STL+ETS	14.5152	14.5382	15.2424	15.7428
STL+RW	14.6651	14.6334	15.2941	15.7877

Note: The best accuracy results for each lead time are marked in bold. ARIMA = autoregressive integrated moving average, ETS = error, trend, seasonal (innovation state space model), GBD = gradient boosting with dropout, GBR = gradient boosting with regularization, KNN = k-nearest neighbor, RF = random forest, RWDRIFT = random walk with drift, STL = seasonal-trend decomposition based on Loess, SVR = support vector regression.

Table 4: RMSE results for customer support arrivals' forecasts.

	Lead Time			
Model	**No lead time**	**One week**	**Two weeks**	**Three weeks**
GBD	18.8706	19.0480	19.3644	19.9452
GBR	18.1216	18.3299	18.6043	19.3079
KNN	24.9867	25.9203	26.3528	27.6355
RF	**15.5678**	**16.6541**	**16.8929**	**18.4903**
SVR	18.3313	18.4081	18.3059	18.9199
STL+ARIMA	22.7009	23.1810	24.2187	25.0726

[9] With 40 GB RAM.

Model	Lead Time No lead time	One week	Two weeks	Three weeks
STL+ETS	22.9251	23.0876	23.9239	24.7768
STL+RW	23.0503	23.1555	23.9506	24.7793

Note: The best accuracy results for each lead time are marked in bold. ARIMA = autoregressive integrated moving average, ETS = error, trend, seasonal (innovation state space model), GBD = gradient boosting with dropout, GBR = gradient boosting with regularization, KNN = k-nearest neighbor, RF = random forest, RWDRIFT = random walk with drift, STL = seasonal-trend decomposition based on Loess, SVR = support vector regression.

To check the robustness of our results, we further consider the queue for customer complaints call arrivals. Since there are less call arrivals compared to the customer queue, the MAE and RMSE are generally lower. Similar to the previous findings, the RF yields the most accurate forecasts compared with the remaining approaches for all considered lead times except for the MAE result with two weeks lead time for which GBR is found to be superior (see Table 5 and Table 6). Aside from RF, GBR is outperforming the RWDRIFT model. The remaining models (i.e., GBD, KNN, and SVR) generate slightly more inaccurate forecasts. Moreover, with the lead time extending, the MAE and RMSE results worsen steadily in most cases.

Table 5: MAE results for customer complaints arrivals' forecasts.

Model	Lead Time No lead time	One week	Two weeks	Three weeks
GBD	3.7668	3.8067	3.8933	3.9694
GBR	3.6058	3.6962	**3.3783**	3.8362
KNN	4.5016	4.7366	4.8095	4.8350
RF	**3.3561**	**3.4348**	3.5629	**3.6746**
SVR	4.3283	4.2826	4.3224	4.2830
STL+ARIMA	3.7197	3.7639	3.8297	3.9073
STL+ETS	3.6990	3.7475	3.8199	3.9163
STL+RW	3.6589	3.7460	3.7968	3.9017

Note: The best accuracy results for each lead time are marked in bold. ARIMA = autoregressive integrated moving average, ETS = error, trend, seasonal (innovation state space model), GBD = gradient boosting with dropout, GBR = gradient boosting with regularization, KNN = k-nearest neighbor, RF = random forest, RWDRIFT = random walk with drift, STL = seasonal-trend decomposition based on Loess, SVR = support vector regression.

Table 6: RMSE results for customer complaints arrivals' forecasts.

Model	Lead Time No lead time	One week	Two weeks	Three weeks
GBD	5.3580	5.4212	5.5593	5.6714
GBR	5.2140	5.3527	5.4871	5.5734
KNN	6.4708	6.8279	6.9224	6.9549

Model	Lead Time			
	No lead time	**One week**	**Two weeks**	**Three weeks**
RF	**4.9422**	**5.0672**	**5.2338**	**5.3791**
SVR	5.9909	6.0502	6.1240	5.9487
STL+ARIMA	5.5152	5.5807	5.6783	5.8243
STL+ETS	5.4833	5.5559	5.6635	5.8210
STL+RW	5.3958	5.4754	5.5949	5.7647

Note: The best accuracy results for each lead time are marked in bold. ARIMA = autoregressive integrated moving average, ETS = error, trend, seasonal (innovation state space model), GBD = gradient boosting with dropout, GBR = gradient boosting with regularization, KNN = k-nearest neighbor, RF = random forest, RWDRIFT = random walk with drift, STL = seasonal-trend decomposition based on Loess, SVR = support vector regression.

To gain further insights regarding the models' performance, we plotted the last predicted week (i.e., 180 observations) for the customer support queue. Figure 5 depicts the time series models' predictions, whereas Figure 6 illustrates the machine learning models' predictions. On the first day of the week (i.e., Monday), the call center was closed, and consequently, this led to an exceptionally high arrival volume on the day after. Apparently, the time series models cannot capture such special days due to the lack of additional information, i.e., predictor variables indicating e.g. holidays and days after. The remaining ML models capture ordinary weekdays and further, such special days more accurately since they allow for the inclusion of explanatory variables for prediction. Consequently, the ML approaches exceed the time series models regarding predictive performance.

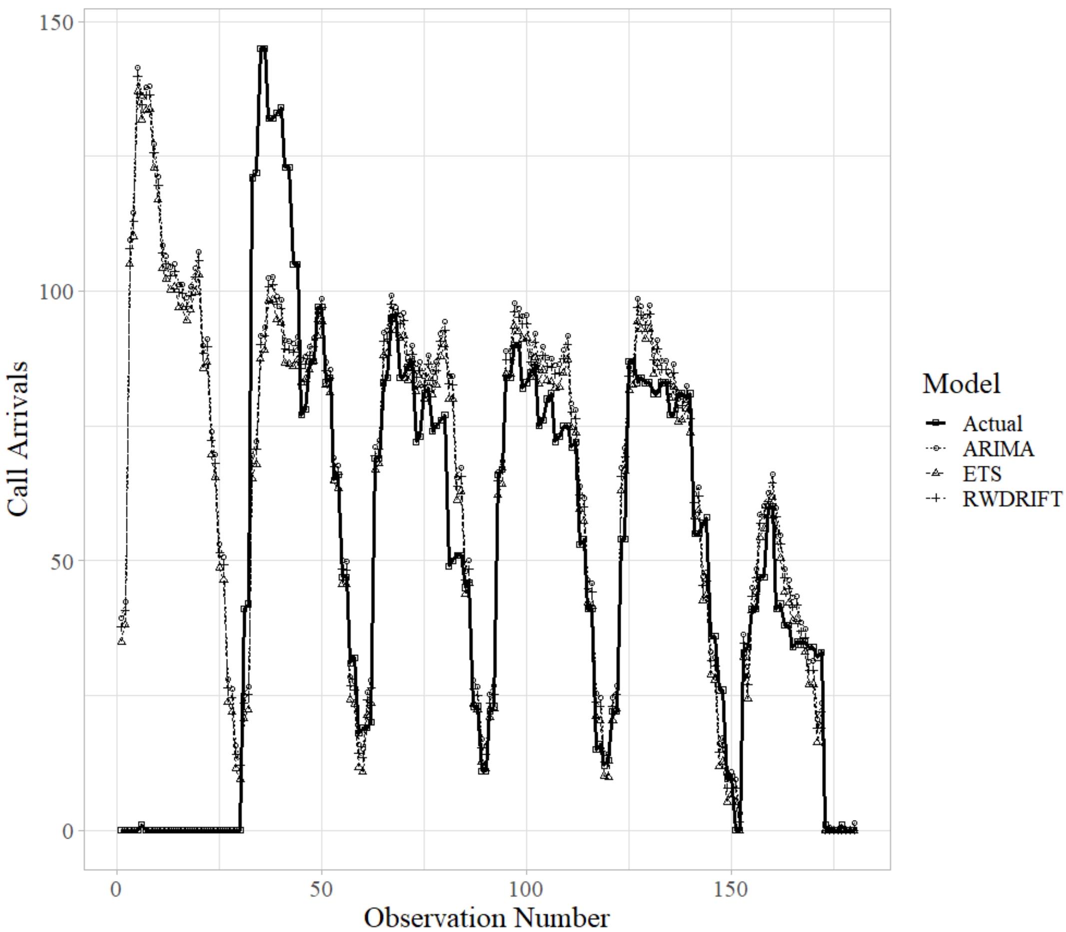

Figure 5: Last predicted week of the time series models.

Note: ARIMA = autoregressive integrated moving average, ETS = error, trend, seasonal (innovation state space model), RWDRIFT = random walk with drift.

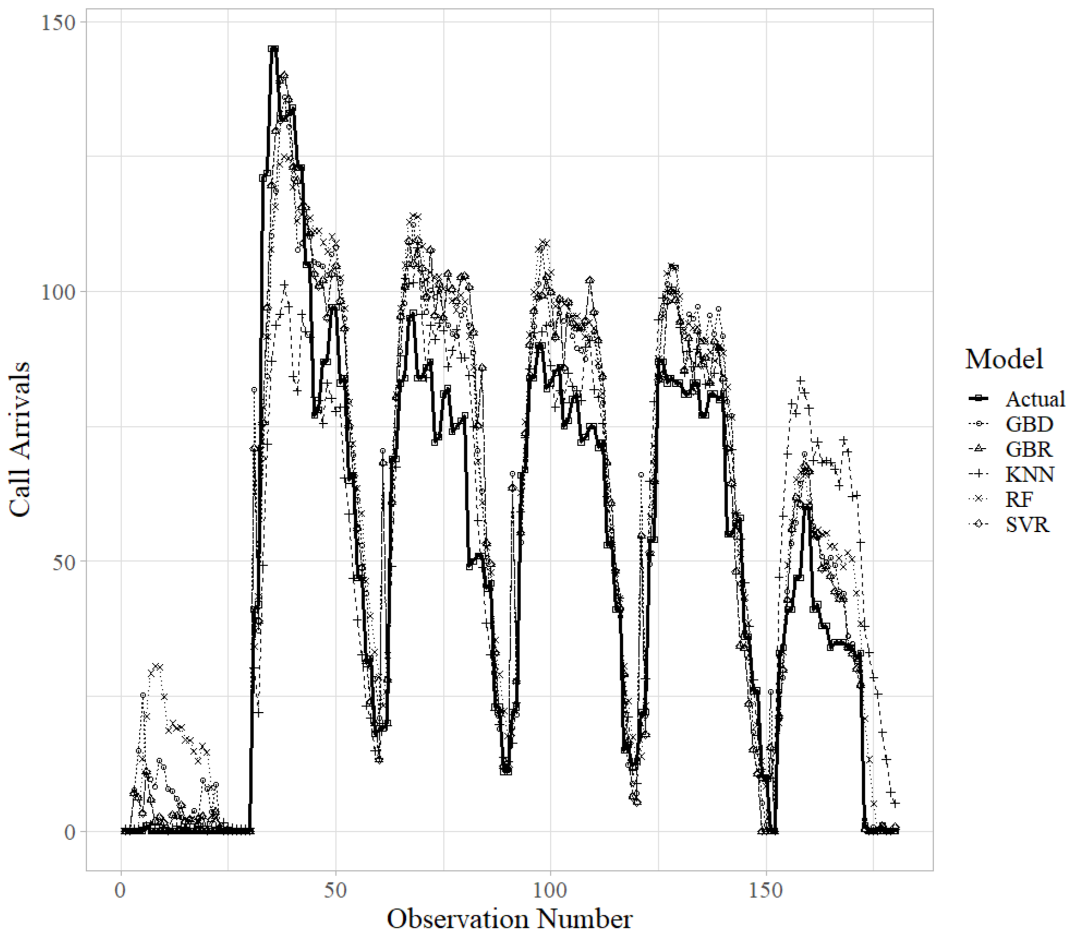

Figure 6: Last predicted week of the machine learning models.

Note: GBD = gradient boosting with dropout, GBR = gradient boosting with regularization, KNN = k-nearest neighbor, RF = random forest, SVR = support vector regression.

5 Discussion

The underlying investigation entails several theoretical implications and contributions made to literature. We present a starting point for shifting traditional call center forecasting literature towards a new paradigm drawing on AI-driven methods by providing a comprehensive understanding of the predictive potential of ML models. As traditional forecasting literature (Andrews & Cunningham, 1995; D. Barrow & Kourentzes, 2018; Mabert, 1985; Taylor, 2008, 2012; Thompson & Tiao, 1971a, 1971b) is predominantly characterized by the use of time series models, we intend to broaden this perspective: Across the two datasets examined, our investigated ML algorithms outperform benchmark models as well as more sophisticated time series models that prior studies most commonly focused on (e.g., ARIMA, exponential smoothing, etc.) in nearly all lead time constellations. Thereby, extending the research on call arrival forecasting techniques with ML approaches like GBR, GBD, KNN, RF, and SVR in this analysis leads to a wider range of methods to generate predictions that are more accurate. Our comprehensive model comparison underpins the preliminary findings of previous studies (D. K. Barrow, 2016; Ebadi Jalal et al., 2016; Rausch & Albrecht, 2020), which used single AI-driven methods like RF or neural networks, indicating that ML techniques are capable of improving the accuracy of call center arrivals' forecasts. Our results prove that tree-based methods and particularly the RF algorithm yield the highest potential for significantly improving forecast accuracy. This finding is replicated for a considerably lower level of call arrival volume in the customer complaints queue.

With regard to the models' practicability, which was neglected by extant literature so far, we are first to consider simultaneously different lead times (i.e., three weeks, two weeks, one week, and no lead time), the trade-off between complexity (i.e., estimation time and computation effort), and forecast accuracy in the model comparison. Extant call center forecasting literature focused mainly on forecast accuracy as a primary decision criterion or considered varying forecast horizons (Aldor-Noiman et al., 2009; D. K. Barrow, 2016; Taylor, 2012) while keeping lead times constant and neglecting complexity. Results prove the leading ML models, and especially RF, to be highly relevant for practical use as their forecast accuracy is less affected by lead time extension. Computation effort, on the other hand, is moderate, and implementation is feasible.

Additionally, we took a closer look at the main reasons for the superiority of ML models. Shedding light on the predictions of special days, such as days after holidays, indicates that ML methods excel in coping with anomalous values as predictor variables are included in the generated ex-post forecasts. Hence, one of the main aspects of ML approaches outperforming traditional time series models is assumed to be the ability to capture additional information on the predicted date or customer contact activities by businesses with the inclusion of predictor variables. Thereby, this characteristic of ML techniques makes them not only stand out in terms of forecast accuracy when it comes to outliers (D. Barrow & Kourentzes, 2018) but also positively affects the overall prediction performance over longer time periods. Nevertheless, albeit few suggestions regarding useful predictor variables have been made (e.g., catalog mailings and holidays (Andrews & Cunningham, 1995)or billing cycles (Aldor-Noiman et al., 2009)), research still lacks a comprehensive understanding on suitable predictor variables for call center arrivals' forecasting. We thus add to the existing body of literature by highlighting that variables such as the time of the day, day of the week, holidays, days after holidays, catalog mailings, and reminders provide valuable information for modeling ex-post forecasts.

The empirical results reported herein should be considered in the light of some limitations. The primary limitation to the generalization of these results accompanies one of the strengths of the study. Keeping in mind the required balance between prediction accuracy and model complexity, we focus on practical relevance in our model selection and neglect models like e.g. sophisticated types of artificial neural networks since such models are time-consuming in estimation, and thus, inadequate for practical use. We also refrain from developing and testing an own method. With an abundance of different ML methods and modifications in literature, we apply ready-to-use methods that are comparatively easy to implement and present a methodological extension to research in the form of a novel implementation focus. Second, the models' prediction performances are depending on the underlying data and, thus, are assumed to vary slightly for different datasets. Therefore, we validate the models' forecast accuracy on two datasets to prove the robustness of our results and further provide the methodological tutorial for testing the identified ML models' performance on other datasets. We do not distinguish between different forecast horizons like several other studies as we re-estimated our models for every week rolling forward from forecast origin, and thus, the forecast horizon constantly remains one week, i.e., 180 observations.

6 Practical implications: Methodological walk-through for call center arrivals' forecasting

Based on the results of the conducted model comparison, organizations are suspected of benefiting from including ML approaches in their process of evaluating and selecting the most suitable method for forecasting call center arrivals and therefore, to support their staffing decisions. To make the underlying process of method comparison and selection accessible to decision-makers in practice as well as to overcome its perceived high complexity and organizations' lack of expertise, we provide a methodological walk-through example based on cross-validation with an expanding rolling window. In doing so, we propose to view the question of method in call center forecasting as the overall issue of implementing a forecasting system that includes prediction accuracy as well as practicability for the user. By presenting a methodological tutorial, we aim to overcome the dependence of method comparisons on data characteristics and, at the same time, accelerate the adoption of ML techniques in this field. On these grounds, we provide a description of the generic cross-validation approach in the programming environment R as well as an in-depth example of RF algorithm as the best-performing model of our previous analysis.

Figure 7 illustrates a generic for-loop for the expanding rolling window that can be utilized to identify the most accurate model. Let n be the n^{th} observation (i.e., row) of the dataset, m be the m^{th} variable (i.e., column) of the dataset, and h be the forecast horizon.

```
results <- vector("numeric")
tic("Looptime")
for(i in 1:k){
  train_subset <- data[1:((n_train - h)+(i*h)),]
  test_subset <- data[((n_test - h)+(i*h)):(((n_test - h) - 1)+((i+1)*h)),]
  ## Insert Model here
}
toc(log = TRUE)
timelog <- tic.log(format = TRUE)
results <- pmax(results,0)

mae(data[n_test:(n_test + k * h), m_calls], results)
rmse(data[n_test:(n_test + k * h), m_calls], results)
```

Figure 7: R Code for rolling expanding window with generic for-loop.

Note: The bold variables have to be replaced depending on the specific dataset.

After analyzing and preprocessing the data as described in Section 3.1, we define an empty numeric vector, in which the results are stored during the loop. The for-loop itself can be iterated k times: let the forecast horizon h be e.g. one week and out-of-sample predictions with cross validation shall be generated for one year, then $k = 52$, i.e., 52 weeks. For each iteration $k = 1,2, \dots, K$ during the loop we define the training and test subset which roll forward for one unit of the forecast horizon h, i.e., $i * h$. Since $1 * h$ observations are added during the first iteration for syntax reasons, h observations are subtracted from the training and test subsets (n_{train} and n_{test} respectively) to yield the intended initial training and test subsets.

After the loop finishes, the looptime is reported with the *toc()* function since a model's computation time can potentially be a crucial aspect for decision makers. Further, in case some models might generate negative predictions we set the minimum value for predictions to zero with *pmax()*. The MAE and RMSE are both calculated by inserting the vector of actual values as the first argument and the vector of predicted values as the second argument.

To test a model's predictive ability, it can be integrated into the generic for-loop. Figure 7 demonstrates the R Code for the loop with the implemented RF. To achieve ease of use as well as to guarantee high model accuracy, we make use of R's *tuneRanger* package which automatically tunes the forest's hyperparameters (i.e., m_{try}, minimum node size, and sample fraction) by creating a regression task with *makeRegrTask()* (Probst et al., 2019).

```
results <- vector("numeric")
tic("Looptime")
for(i in 1:k){
  train_subset <- data[1:((n_train − h)+(i*h)),]
  test_subset <- data[((n_test − h)+(i*h)):(((n_test − h) − 1)+((i+1)*h)),]
  task <- makeRegrTask(data = train_subset, target = "m_calls")
  RF <- tuneRanger(task, num.trees = t, iters = j)
  pred_rf <- predict(RF$model, newdata = test_subset)
  pred_data <- pred_rf$data
  pred_response <- pred_data$response
  results <- append(results, pred_response)
  }
toc(log = TRUE)
timelog <- tic.log(format = TRUE)
results <- pmax(results,0)

mae(data[n_test:(n_test + k ∗ h), m_calls], results)
rmse(data[n_test:(n_test + k ∗ h), m_calls], results)
```

Figure 8: R Code for rolling expanding window with random forest.

Note: The bold variables have to be replaced depending on the specific dataset.

The package is favorable since it utilizes sequential model-based optimization (SMBO)[10] as a tuning strategy which is faster and moreover, better regarding its performance than standard tuning packages (Probst et al., 2019). It conducts a SMBO with 30 random points for the initial design (i.e., random points drawn from the hyperparameter space) and 70 iterative steps in the optimization procedure. Optionally, the number of iterations i can be inserted manually. m_{try} values are sampled from $[0, p]$ with p being the number of predictors. Sample size values are sampled from $[0.2 * n, 0.9 * n]$ with the number of observations n. Node size values are sampled with higher probability for smaller values by sampling x from $[0,1]$ and hence, transforming the value by $[(n * 0.2)^x]$. Further, out-of-bag predictions during the fitting procedure can be evaluated with several different error measures (mean squared error (MSE^{OOB}) as default for regression). The number of trees t can be inserted optionally: research found the model's performance peak to be reached during the construction of the first 100 trees (Probst et al., 2012).

Subsequent to the fitting procedure, the predictions based on the new and unknown test data are generated. By using the *append()* function, the predictions with length *h* are attached sequentially for

[10] For detailed information on the SMBO procedure, readers are referred to Probst et al. (2019).

k iterations. As described in Section 3.2, the MAE and RMSE results allow for a practically valid comparison of different models.

The methodological walk-through combined with the preceding results of the model comparison lead to several practical implications for businesses and organizational decision-makers. First, the hypothesized improvement of call center arrivals' forecasting accuracy was confirmed for the selection of feasible ML methods. Thereby, the range of applicable methods providing robust and accurate predictions in this field is extended to suitable ML algorithms. In comparison with commonly used forecasting techniques, ML models generate more precise forecasts in almost every case. That way, unnecessary costs caused by overstaffing as well as customer dissatisfaction originating from long waiting times due to understaffing can be avoided: In case the forecasts overestimate the actual customer support call arrival volume, decision-makers can save approximately 1.83[11] call center agents per day on average if RF (best-performing ML model) compared to ETS (best-performing time series model) is employed. Vice versa, in case the forecasts underestimate the actual call arrival volume, customers would need to wait approximately 0.41[12] minutes less on average if RF is implemented instead of ETS. Furthermore, the findings also indicate that decision-makers are recommended to minimize lead time in case it is possible in the scope of staffing regulations.

Overall, we exclusively investigated models standing out due to the favorable trade-off between accuracy and practicability, especially in terms of complexity regarding estimation time as well as computation effort. The comprehensibility and ease of implementation of tree-based models as best-performing methods is further verified by the applied example above. From a general perspective, organizations are encouraged to use the demonstrated process of cross-validation with an expanding rolling window not only to test and implement different approaches for call center arrivals' forecasting but also to adapt it for any forecasting task based on sequential data (e.g., e-mail arrivals, product sales, etc.). The implementation of this approach in an accessible programming environment further fills the need of practitioners for a task-specific guideline for the selection of AI-driven methods and helps to overcome the practicability issues identified in literature.

7 Conclusion

The process of forecasting call center arrival volumes in an increasingly complex data environment is predestinated to capitalize on AI-driven methods by improving internal decision-making. Accurate forecasts generated by ML algorithms are assumed to generate cost savings and service improvements through precise staffing. However, insights on and practical use of ML in call center arrivals' forecasting are limited.

Acting on the assumed potential of ML in this field as well as on the existing constraints regarding practicability in organizational use, this paper follows a two-step approach of model performance

[11] If the processing time is 10 min per call arrival and the working hours per call center agent are 8 h per day.
[12] If the processing time is 10 min per call arrival and there are 70.95 call arrivals per interval on average.

evaluation and practical implementation. The first step constitutes an extensive model comparison of selected feasible ML methods with common as well as sophisticated time series models using the call center arrival data of a large online retailer. In the second step, the implementation of the model evaluation and selection process based on cross-validation with an expanding rolling window is made accessible for practitioners by providing a methodological walk-through example.

The results of the method comparison confirm the hypothesized high potential of ML models for accuracy improvements based on two datasets and various lead times investigated. Tree-based methods and particularly RF algorithm yield the best prediction performances and therefore approve as preferable alternatives to commonly used methods. These findings are substantiated by the implementation example using RF as the best-performing model. By providing an efficient and reproducible way of assessing the case-specific value of ML methods in forecasting for organizations within a programming environment, the dependence of method comparison results on data characteristics as well as the lack of comprehensibility and methodological expertise in practical settings are mitigated or even eliminated.

This paper therefore presents a starting point for shifting traditional call center forecasting towards a new paradigm drawing on AI-driven methods by demonstrating the high predictive potential of ML models in comparison to commonly used methods. From a practical perspective, this study contributes to an improved understanding for businesses on how to deal with the increasingly complex task of forecasting call center arrivals caused by the datafication of customer relationships. Being aware of the general applicability of ML models to yield high forecast accuracy, organizations are now enabled to test ML techniques in individual practical settings by adapting the proposed implementation of a valid model selection process in time series forecasting. Improvements in prediction accuracy achieved by this approach can directly be capitalized on through optimized staffing. Future research is encouraged to extend the predictions to concrete staffing recommendations incorporating average service times. As a whole, this work suggests that taking the next step in call center arrivals' forecasting research towards advanced ML, such as deep neural networks and hybrid approaches, is likely to be beneficial. In this case, the evaluation of these methods beyond forecasting accuracy is recommended to ensure the practical value of future findings.

Appendix: Time series models

Overall, one strength of time series models is their ability to generate predictions only based on the time series' previous values without any other contextual information and thus, they are adequate models if information is scarce. The non-seasonal ARIMA (p, d, q) model (G. E. P. Box & Jenkins, 1970) assumes that a with d degrees differenced time series depends on its past values being periods apart and on a finite number q of prior forecast errors ε with p, d, and q being non-negative integers. Thus, it consists of an autoregressive process as well as moving average process

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

with y'_t being the differenced time series, ϕ_p being the parameter for autoregressive process, and θ_q being the parameter for moving average process. Since its development in the 1970s, the ARIMA model is among the most popular forecasting approaches across numerous application contexts, as it was found to perform well in the short-term (D. K. Barrow, 2016) and further, is suitable for a variety of data types with different characteristics as there are stationary as well as nonstationary ARIMA methods (Hyndman & Athanasopoulos, 2018).

While ARIMA models intend to capture autocorrelations in the data, exponential smoothing models draw on trend and seasonality in the data (Hyndman & Athanasopoulos, 2018). Holt-Winters' exponential smoothing model (Holt, 2004; Winters, 1960) was proposed in the late 1950s and weight the averages of the time series' previous observations. Thereby, the weights are decreasing exponentially the further the observations lie in the past. The component form of simple exponential smoothing can be defined as

$$\ell_t = (1 - \alpha)\ell_{t-1} + \alpha y_t$$

$$\hat{y}_{t+h|t} = \ell_t$$

with horizon $h = 1,2,..$, smoothing parameter $0 \leq \alpha \leq 1$ and series level (or smoothed value) ℓ_t at time t. If the exponential smoothing model further allows for additive or multiplicative errors, it evolves into an innovations state space model ETS(·,·,·) for (Error={Additive (A), Multiplicative (M)}, Trend={None (N), A}, Seasonal={N, A, M}):

$$\ell_t = \ell_{t-1} + \alpha \varepsilon_t$$

$$y_t = \ell_{t-1} + \varepsilon_t$$

where α is the smoothing parameter and ℓ_t is the series level (or smoothed value) at time t. Random walk models are frequently used for nonstationary data as random walks typically consist of long periods of apparent (upward or downward) trends and exhibit sudden changes in direction (Hyndman & Athanasopoulos, 2018). The forecasts from the random walk model are equal to the time series' last observation:

$$\hat{y} = y_{t-1} + \varepsilon_t$$

As an extension to the basic model, the drift parameter c is frequently added which is the average of changes between consecutive observations:

$$\hat{y} = c + y_{t-1} + \varepsilon_t$$

If c is positive, there is an increase in the average of changes between consecutive observations and thus, the prediction $\hat{y}$ will tend to drift upwards and vice versa for negative values of c.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2019). Exploring the Impact of Artificial Intelligence: Prediction versus Judgment. Information Economics and Policy(Working Paper 24626), 1–17. https://doi.org/10.3386/w24626

Aksin, Z., Armony, M., & Mehrotra, V. (2007). The Modern Call Center: A Multi-Disciplinary Perspective on Operations Management Research. Production and Operations Management, 16(6), 665–688. https://doi.org/10.1111/j.1937-5956.2007.tb00288.x

Aldor-Noiman, S., Feigin, P. D., & Mandelbaum, A. (2009). Workload forecasting for a call center: Methodology and a case study. The Annals of Applied Statistics, 3(4), 1403–1447. https://doi.org/10.1214/09-AOAS255

Andrews, B. H., & Cunningham, S. M. (1995). L. L. Bean Improves Call-Center Forecasting. Interfaces, 25(6), 1–13. https://doi.org/10.1287/inte.25.6.1

Arora, N., Allenby, G. M., & Ginter, J. L. (1998). A Hierarchical Bayes Model of Primary and Secondary Demand. Marketing Science, 17(1), 29–44. https://doi.org/10.1287/mksc.17.1.29

Atlason, J., Epelman, M. A., & Henderson, S. G. (2008). Optimizing Call Center Staffing Using Simulation and Analytic Center Cutting-Plane Methods. Management Science, 54(2), 295–309. https://doi.org/10.1287/mnsc.1070.0774

Avramidis, A. N., Deslauriers, A., & L'Ecuyer, P. (2004). Modeling Daily Arrivals to a Telephone Call Center. Management Science, 50(7), 896–908. https://doi.org/10.1287/mnsc.1040.0236

Barrow, D., & Kourentzes, N. (2018). The impact of special days in call arrivals forecasting: A neural network approach to modelling special days. European Journal of Operational Research, 264(3), 967–977. https://doi.org/10.1016/j.ejor.2016.07.015

Barrow, D. K. (2016). Forecasting intraday call arrivals using the seasonal moving average method. Journal of Business Research, 69(12), 6088–6096. https://doi.org/10.1016/j.jbusres.2016.06.016

Bergstra, J. S., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. Advances in Neural Information Processing Systems, 2546–2554.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control. John Wiley & Sons.

Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control. Holden-Day.

Brady, M. K., & Cronin, J. J. (2001). Some New Thoughts on Conceptualizing Perceived Service Quality: A Hierarchical Approach. Journal of Marketing, 65(3), 34–49. https://doi.org/10.1509/jmkg.65.3.34.18334

Breiman, L. (1996). Bagging Predictors. Machine Learning, 24(2), 123–140. https://doi.org/10.1007/bf00058655

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.

Brockwell, P. J., Davis, R. A., & Calder, M. V. (2002). Introduction to time series and forecasting. Springer.

Brynjolfsson, E., & McAfee, A. (2017). The Business of Artificial Intelligence. Harvard Business Review, 7, 1–20.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project.

Burez, J., & van den Poel, D. (2009). Handling class imbalance in customer churn prediction. Expert Systems with Applications, 36(3), 4626–4636. https://doi.org/10.1016/j.eswa.2008.05.027

Cao, Q., & Parry, M. E. (2009). Neural network earnings per share forecasting models: A comparison of backward propagation and the genetic algorithm. Decision Support Systems, 47(1), 32–41. https://doi.org/10.1016/j.dss.2008.12.011

Cezik, M. T., & L'Ecuyer, P. (2008). Staffing Multiskill Call Centers via Linear Programming and Simulation. Management Science, 54(2), 310–323. https://doi.org/10.1287/mnsc.1070.0824

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, 36(4), 1165–1188. https://doi.org/10.2307/41703503

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. Journal of Official Statistics, 6(1), 3–73.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273–297.

Coussement, K., & van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. Expert Systems with Applications, 34(1), 313–327. https://doi.org/10.1016/j.eswa.2006.09.038

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Dean, A. M. (2007). The Impact of the Customer Orientation of Call Center Employees on Customers' Affective Commitment and Loyalty. Journal of Service Research, 10(2), 161–173. https://doi.org/10.1177/1094670507309650

Donkers, B., Franses, P. H., & Verhoef, P. C. (2003). Selective Sampling for Binary Choice Models. Journal of Marketing Research, 40(4), 492–497. https://doi.org/10.1509/jmkr.40.4.492.19395

Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. http://arxiv.org/pdf/1702.08608v2

Ebadi Jalal, M., Hosseini, M., & Karlsson, S. (2016). Forecasting incoming call volumes in call centers with recurrent Neural Networks. Journal of Business Research, 69(11), 4811–4814. https://doi.org/10.1016/j.jbusres.2016.04.035

Fang, K., Jiang, Y., & Song, M. (2016). Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. Computers & Industrial Engineering, 101, 554–564. https://doi.org/10.1016/j.cie.2016.09.011

Fantazzini, D., & Figini, S. (2009). Random Survival Forests Models for SME Credit Risk Measurement. Methodology and Computing in Applied Probability, 11(1), 29–45. https://doi.org/10.1007/s11009-008-9078-2

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics(29), Article 5, 1189–1232.

Friedman, J. H. (2002). Stochastic gradient boosting. Computational Statistics & Data Analysis, 38(4), 367–378. https://doi.org/10.1016/s0167-9473(01)00065-2

Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone Call Centers: Tutorial, Review, and Research Prospects. Manufacturing & Service Operations Management, 5(2), 79–141. https://doi.org/10.1287/msom.5.2.79.16071

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. ACM Computing Surveys, 51(5), 1–42. https://doi.org/10.1145/3236009

Harrison, P. J., & Stevens, C. F. (1976). Bayesian Forecasting. Journal of the Royal Statistical Society: Series B (Methodological), 38(3), 205–228. https://doi.org/10.1111/j.2517-6161.1976.tb01586.x

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting, 20(1), 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015

Huang, M.-H., & Rust, R. T. (2018). Artificial Intelligence in Service. Journal of Service Research, 21(2), 155–172. https://doi.org/10.1177/1094670517752459

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.

Ibrahim, R., & L'Ecuyer, P. (2013). Forecasting Call Center Arrivals: Fixed-Effects, Mixed-Effects, and Bivariate Models. Manufacturing & Service Operations Management, 15(1), 72–85. https://doi.org/10.1287/msom.1120.0405

Ibrahim, R., Ye, H., L'Ecuyer, P., & Shen, H. (2016). Modeling and forecasting call center arrivals: A literature survey and a case study. International Journal of Forecasting, 32(3), 865–874. https://doi.org/10.1016/j.ijforecast.2015.11.012

Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics, 5(3), 299–314.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, 62(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. (2018). Advanced Customer Analytics: Strategic Value Through Integration of Relationship-Oriented Big Data. Journal of Management Information Systems, 35(2), 540–574. https://doi.org/10.1080/07421222.2018.1451957

Koole, G., & Pot, A. (2005). Approximate dynamic programming in multi-skill call centers. Proceedings of the Winter Simulation Conference 2005, 576–583.

Lemmens, A., & Croux, C. (2006). Bagging and Boosting Classification Trees to Predict Churn. Journal of Marketing Research, 43(2), 276–286. https://doi.org/10.1509/jmkr.43.2.276

Li, H., Sun, J., & Wu, J. (2010). Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. Expert Systems with Applications, 37(8), 5895–5904. https://www.sciencedirect.com/science/article/pii/s0957417410000552

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R News, 2(3), 18–22.

Mabert, V. A. (1985). Short interval forecasting of emergency phone call (911) work loads. Journal of Operations Management, 5(3), 259–271. https://doi.org/10.1016/0272-6963(85)90013-0

Martens, D., Vanthienen, J., Verbeke, W., & Baesens, B. (2011). Performance of classification models from a user perspective. Decision Support Systems, 51(4), 782–793. https://doi.org/10.1016/j.dss.2011.01.013

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. Journal of Big Data, 2(1), 886. https://doi.org/10.1186/s40537-014-0007-7

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. Royal Statistical Society: Series a (General), 135(3), 370–384.

Overgoor, G., Chica, M., Rand, W., & Weishampel, A. (2019). Letting the Computers Take Over: Using AI to Solve Marketing Problems. California Management Review, 61(4), 156–185. https://doi.org/10.1177/0008125619859318

Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. Expert Systems with Applications, 36(1), 2–17. https://doi.org/10.1016/j.eswa.2007.10.005

Probst, P., Wright, M., & Boulesteix, A.-L. (2019). Hyperparameters and Tuning Strategies for Random Forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(3), 281–300. https://doi.org/10.1002/widm.1301

Rashmi, K. V., & Gilad-Bachrach, R. (2015). DART: Dropouts meet Multiple Additive Regression Trees. AISTATS, 489–497. http://arxiv.org/pdf/1505.01866v1

Rausch, T. M., & Albrecht, T. (2020). The impact of lead time and model selection on the accuracy of call center arrivals' forecasts. Proceedings of the 28th European Conference on Information Systems (ECIS), 1–16.

Reinartz, W., Thomas, J. S., & Kumar, V. (2005). Balancing Acquisition and Retention Resources to Maximize Customer Profitability. Journal of Marketing, 69(1), 63–79. https://doi.org/10.1509/jmkg.69.1.63.55511

Schapire, R. E. (1990). The strength of weak learnability. Machine Learning (5), 197–227. https://doi.org/10.1109/sfcs.1989.63451

Shen, H., & Huang, J. Z. (2008). Interday Forecasting and Intraday Updating of Call Center Arrivals. Manufacturing & Service Operations Management, 10(3), 391–410. https://doi.org/10.1287/msom.1070.0179

Shmueli, G., & Koppius, O. (2011). Predictive Analytics in Information Systems Research. MIS Quarterly(35), Article 3, 553–572. https://doi.org/10.2139/ssrn.1606674

Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Kumar Roy, P. (2017). Predicting the "helpfulness" of online consumer reviews. Journal of Business Research, 70, 346–355. https://doi.org/10.1016/j.jbusres.2016.08.008

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199–222. https://doi.org/10.1023/b:stco.0000035301.49549.88

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. Journal of Chemical Information and Computer Sciences, 43(6), 1947–1958. https://doi.org/10.1021/ci034160g

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. California Management Review, 61(4), 15–42. https://doi.org/10.1177/0008125619867910

Taylor, J. W. (2008). A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center. Management Science, 54(2), 253–265. https://doi.org/10.1287/mnsc.1070.0786

Taylor, J. W. (2012). Density Forecasting of Intraday Call Center Arrivals Using Models Based on Exponential Smoothing. Management Science, 58(3), 534–549. https://doi.org/10.1287/mnsc.1110.1434

Thompson, H. E., & Tiao, G. C. (1971a). Analysis of Telephone Data: A Case Study of Forecasting Seasonal Time Series. The Bell Journal of Economics and Management Science, 2(2), 515. https://doi.org/10.2307/3003004

Thompson, H. E., & Tiao, G. C. (1971b). Analysis of Telephone Data: A Case Study of Forecasting Seasonal Time Series. The Bell Journal of Economics and Management Science, 2(2), 515. https://doi.org/10.2307/3003004

Vapnik, V. N. (2000). The Nature of Statistical Learning Theory. Springer. https://doi.org/10.1007/978-1-4757-3264-1

Vapnik, V. N. (2006). Estimation of Dependences Based on Empirical Data (Addendum 1). Springer. https://doi.org/10.1007/0-387-34239-7

Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Systems with Applications, 38(3), 2354–2364. https://doi.org/10.1016/j.eswa.2010.08.023

Wedel, M., & Kannan, P. K. (2016). Marketing Analytics for Data-Rich Environments. Journal of Marketing, 80(6), 97–121. https://doi.org/10.1509/jm.15.0413

Weinberg, J., Brown, L. D., & Stroud, J. R. (2007). Bayesian Forecasting of an Inhomogeneous Poisson Process With Applications to Call Center Data. Journal of the American Statistical Association, 102(480), 1185–1198. https://doi.org/10.1198/016214506000001455

Whiting, A., & Donthu, N. (2006). Managing Voice-to-Voice Encounters. Journal of Service Research, 8(3), 234–244. https://doi.org/10.1177/1094670505281703

Winters, P. R. (1960). Forecasting Sales by Exponentially Weighted Moving Averages. Management Science, 6(3), 324–342. https://doi.org/10.1287/mnsc.6.3.324

Yang, S., & Allenby, G. M. (2003). Modeling Interdependent Consumer Preferences. Journal of Marketing Research, 40(3), 282–294. https://doi.org/10.1509/jmkr.40.3.282.19240

3.2.2 Research Paper No. 5: Beyond the Beaten Paths of Forecasting Call Center Arrivals: On the Use of Dynamic Harmonic Regression with Predictor Variables

Authors: Rausch, T. M., Albrecht, T., & Baier, D.

Submitted in: Journal of Business Economics

Abstract: With call centers constituting a crucial customer touchpoint, short waiting times are proven to be an indispensable prerequisite of customer satisfaction. Therefore, high accuracy of call arrivals' forecasts is needed to avoid under- as well as overstaffing and to enhance a company's customer relationship. However, opinions within literature on the best-performing method for call center arrival forecasting are still diverging. Further, extant research barely shed light on the potential of combining the different benefits of both time series models and machine learning (ML) as well as regression models. Hence, we propose a new method for call arrival forecasting, which is able to capture the dynamics of a time series and to include contextual information in form of predictor variables. Thus, our extended dynamic harmonic regression (DHR) approach combines the strengths of the different methods. We test the predictive potential of our approach by forecasting the call and e-mail arrival series of a leading German online retailer comprising 174 weeks of data. We apply time series cross-validation with an expanding rolling window over 52 weeks and moreover, utilize established time series as well as ML models as benchmarks. Our proposed DHR model with predictor variables outperforms the remaining approaches with regard to forecast accuracy for every considered lead time and hence, provides empirical evidence for the potential of combining different types of method. We further contribute to extant knowledge on predictor variables by confirming that marketing-relevant variables – such as long periods with catalog releases or billing cycles – should be partitioned into shorter sequences to capture the varying effect on forecast accuracy over time. Additionally, we show that data on postal reminders particularly enhance call arrival prediction and, vice versa, data on e-mail reminders support e-mail arrival prediction.

Keywords: forecasting; call center arrivals; dynamic harmonic regression; time series analysis; machine learning; customer relationship

1 Introduction

In the retail industry, typical stages along the customer journey like order taking, after-sales service, and complaint resolution can easily be completed online (Gensler et al., 2012; Verhoef et al., 2015). Nevertheless, approximately 70 percent of customers still prefer to interact with a human counterpart for customer service requests (Sitel Group, 2018). For many organizations, call centers constitute the main or only point of human-to-human interaction with their customers. Thus, call centers are an essential communication channel for businesses and an important customer touchpoint across many industries (Aksin et al. 2007).

High perceived service quality at this customer interface contributes greatly to customer loyalty and is determined by short waiting times as well as the call experience and service outcome itself (Brady & Cronin Jr., 2001; Parasuraman et al., 1985; Zeithaml et al., 1996). Hence, to provide the correct number of call center agents as customer service representatives at the right time and to evaluate their required areas of expertise, call arrival volumes in different queues have to be predicted[13] reliably in advance. In this regard, preceding literature in the fields of operations management and forecasting so far focused on optimizing the opposite tendency of staffing costs and customer waiting times by enhancing forecast accuracy of predominant time series models (Dean, 2007; Gans et al., 2003). Methods like autoregressive integrated moving average (ARIMA) (Andrews & Cunningham, 1995; Barrow, 2016; Mabert, 1985; Thompson & Tiao, 1971), exponential smoothing (Taylor, 2003, 2008, 2012), or dynamic harmonic regression (DHR) (Young et al., 1999; Young, 1999) have traditionally been established as standard-setting approaches. However, such time series models generate predictions based on the time series' previous values without including any contextual data or other additional information available.

Meanwhile, new methods apart from time series models were developed and investigated to predict call center arrivals with high accuracy (Albrecht et al., 2021; Rausch & Albrecht, 2020). Regression models (Aldor-Noiman et al., 2009; Brown et al., 2005; Ibrahim et al., 2016; Ibrahim & L'Ecuyer, 2013) and machine learning (ML) approaches like random forest (RF) and neural networks (Barrow, 2016; Jalal et al., 2016) were found to yield accurate predictions by including meaningful predictor variables. Incorporating contextual data into call arrival forecasts not only positively affects prediction accuracy but also allows for a more customer-centric perspective in call center forecasting. By including information on customer motives and behavior, valuable marketing insights are gained. Thus, prior research recommended to model predictor variables (Taylor, 2008).

Hence, extant forecasting literature still disagrees on the best-performing model type for call arrival forecasting. Simultaneously, from a conceptual point of view, previous research did not explicitly investigate whether incorporating the benefits of both time series approaches as well as regression and ML models into one model yields an unrecognized predictive potential. We therefore contribute to literature by proposing a new method for call center arrival forecasting, which combines the strengths

[13] The terms predicting and forecasting will be utilized interchangeably in this paper.

of those model types – on the one hand by capturing the dynamics of the time series and on the other hand by including predictor variables. We hypothesize that this approach will lead to an increase in predictive performance and, simultaneously, will entail advantages for practical use. We extend the established Dynamic Harmonic Regression (DHR) model, which utilizes a sum of sinusoidal terms (i.e., Fourier terms) as predictors to handle periodic seasonality and an ARIMA error to capture short-term dynamics, by including predictor variables in the considered information space to generate predictions. We test the predictive potential of our approach with two different call and e-mail arrival series of a leading German online retailer comprising 174 weeks of data. We compare our proposed model to different established time series and ML approaches with time series cross-validation and an expanding rolling window. We further extend knowledge on suitable predictor variables in a marketing context, which has been neglected by prior research, as most datasets do not include such contextual information.

The remainder of this paper is structured as follows: Section 2 reviews related work on call center arrival forecasting approaches. In Section 3, we derive our DHR model including predictor variables. Section 4 and 5 analyze the customer support call and e-mail arrival series respectively by presenting the preliminaries of both datasets, the experimental design, and the analysis results. Section 6 discusses the results with regard to the practical implications as well as the theoretical contribution of the study before it reflects the limitations and provides guidance for future research. The paper concludes with a concise summary of its principal points in Section 7.

2 Related work

Call center call arrivals are count data and hence, discrete data restricted to non-negative integers. Therefore, a common model utilized for call arrivals' forecasts is a Poisson arrival process (Aksin et al., 2007; Cezik & L'Ecuyer, 2008; Gans et al., 2003; Taylor, 2012). However, one key feature of call center arrivals, which is not aligning with the preceding Poisson assumption, is their time dependence: call arrival counts exhibit obvious patterns (or seasonalities respectively) which are repeating itself sub-daily (e.g., half-hourly, hourly), daily, weekly, or yearly (Brown et al., 2005; Ibrahim et al., 2016). Thus, literature frequently draws on time series analysis to forecast call center arrivals assuming them to be a sequence of dependent, contiguous observations which are made continuously over a certain time interval (Brockwell & Davis, 2002).

Box and Jenkins (1970) developed a non-seasonal ARIMA (p,d,q) which assumes that – with d degrees differenced – the time series y_t' at time t is dependent on past values p periods apart (autoregressive part) and is related to a finite number q of preceding forecast errors ε (moving average part). The ARIMA model – or reduced models containing only sub-components respectively – are among the most common approaches to predict future call arrivals. The fields of application include e.g. a public telephone company (Thompson & Tiao, 1971), an emergency line (Mabert, 1985), banks in the US, UK, and Israel (Barrow, 2016), or a retailer for outdoor goods and apparel (Andrews & Cunningham, 1995). In the latter case, additional contextual covariates (i.e., catalog mailings and holidays) were modeled to enhance forecast accuracy (Andrews & Cunningham, 1995). Bianchi et al. (1998) forecasted call center arrivals for telemarketing centers and found ARIMA modeling to be superior to Holt-Winters' exponentially weighted moving average.

Exponential smoothing methods are predicting values by weighting averages of previous observations with exponentially decaying weights the further the observations lie in the past (Holt, 2004; Winters, 1960). Extensions of the Holt-Winters' approach comprised double seasonality (Taylor, 2003), a Poisson count data model with gamma-distributed stochastic arrival rate (Taylor, 2012), and robust exponential smoothing (Taylor, 2008). For relatively short forecast horizons (up to six days), Taylor (2008) found the Holt-Winters' extension to outperform established approaches such as seasonal autoregressive moving average (ARMA). A novel subclass of exponential smoothing models are innovation state space models that add an error term to exponential smoothing models yielding the label Error, Trend, Seasonal (ETS) $(\cdot,\cdot,\cdot)$ for each state space model (Hyndman et al., 2002). The single components can be defined as Error=$\{\text{Additive (A)}, \text{Multiplicative (M)}\}$, Trend=$\{\text{None (N)}, A\}$ and Seasonal=$\{N,A,M\}$. ETS models were found to be both superior (Hyndman et al., 2002) as well as inferior (Barrow & Kourentzes, 2018) to ARIMA in a call center forecasting context. De Livera et al. (2011) extended the basic ETS model which allows the seasonalities to slowly change over time by including Fourier terms for a trigonometric representation of seasonality and a Box-Cox transformation. The model's key features trigonometric seasonality, Box-Cox transformation, ARMA errors, trend and seasonal components yield the acronym TBATS.

The random walk (RW) method is an easy-to-implement time series approach frequently used as a benchmark model. Essentially, its forecasts equal the last actual value or observation respectively. By including the drift parameter c, the model additionally captures the average of changes between consecutive observations. Despite its simplicity, its performance is to some extent comparable – but not superior – to established methods in many experimental settings (Taylor, 2008).

Aside from time dependence, another key property of call arrivals is their overdispersion, i.e., the variance of an arrival count per time period usually exceeds its mean (Aldor-Noiman et al., 2009; Avramidis et al., 2004; Jongbloed & Koole, 2001), which is not consistent with the Poisson assumption. One option to deal with overdispersion is to assume the Poisson arrival process as doubly stochastic with random arrival rate (Whitt, 1999). Literature then drew on the root-unroot variance stabilizing data transformation for Poisson data assuming that with a large number of calls per interval, the square-root transformed counts are roughly normally distributed (Aldor-Noiman et al., 2009; Brown et al., 2005; Brown et al., 2010). Normality is then exploited to fit Gaussian linear models. Particularly, linear fixed-effects (Ibrahim & L'Ecuyer, 2013; Shen & Huang, 2008; Weinberg et al., 2007), random-effects, and mixed-effects models (Aldor-Noiman et al., 2009; Brown et al., 2005; Ibrahim et al., 2016; Ibrahim & L'Ecuyer, 2013) were utilized subsequently. Mostly, fixed effects comprise the effect of the day of the week and the time of the day and their interaction, whereas random effects capture the daily volume deviation of the fixed weekday effect. In call center forecasting, Aldor-Noiman (2009) further modeled billing cycles and delivery periods as additional covariates.

Recently, research started to investigate beyond time series and regression models. In the context of forecasting, approaches within the ML paradigm are characterized by detaching from distributional assumptions while, at the same time, promising immense predictive performance in a variety of application areas (Carbonneau et al., 2008). Rausch and Albrecht (2020) included RF as a powerful ML method in their comparison of novel time series and regression models for call center arrivals forecasting. RF was found to yield higher prediction accuracy for nearly all of the considered lead time constellations. Similar results were gathered in an extensive ML comparison study by Albrecht, Rausch, and Derra (2021). Besides, artificial neural networks such as multilayer perceptrons (Barrow, 2016) and recurrent neural networks (Jalal et al., 2016) attracted increasing attention. Barrow and Kourentzes (2018) found artificial neural networks to outperform traditional models for model complex outliers in call center arrival forecasting. However, call center forecasting research using ML approaches is still in its infancy.

While time series models generate predictions based on previous values in the time series but generally do not capture additional information, aforementioned ML and regression models allow for the inclusion of predictor variables. As these parameter values are typically available for both the past (i.e., the training data) and the future (i.e., the predictions and the test data), ex-post forecasts can be created. Such data specifying a forecasting period in the future at the time of the prediction may include, for instance, date-related information or data on scheduled customer contact activities. In contrast, time series models' forecasts are solely based on information available at the time of the prediction, i.e., the time series' historical values are used to generate ex-ante forecasts (Taylor, 2008). This difference regarding the forecasting method was previously found to significantly affect a models' ability to maintain stable prediction accuracy with varying lead time as the inclusion of predictor variables is assumed to make forecasting call center arrivals more robust and accurate (Rausch and Albrecht, 2020). On the other hand, ex-post forecasting models' lacking ability of capturing information of a time series' course and dynamics is supposed to prevent a significant further increase in model performance (Barrow 2016).

In case of ex-post forecasting models, including meaningful context factors in the form of predictor variables is critical as their informative value strongly affects prediction accuracy (Andrews & Cunningham, 1995). Previous literature identified an extensive list of possible variables that have been observed to affect arrival volumes. Data specifying date-related patterns such as variables indicating the time of day, the day of the week and holidays are widely used (Ibrahim et al., 2016; Ibrahim & L'Ecuyer, 2013; Shen & Huang, 2008; Weinberg et al., 2007). Additionally, information regarding customer contact activities on the part of the company like variables revealing upcoming billing cycles, delivery periods and catalog mailings has been investigated (Aldor-Noiman et al., 2009). However, their effect on call center arrival volumes has only been examined for a fixed point of prediction so that findings on the influence on customer call behavior are vague and do not allow for a thorough understanding of the relation. In this regard, capturing the temporal effect of influential factors over time to enable the estimation of short-time and medium-term effects as well as to assess their interrelation with data-related factors such as holidays and weekends is needed. This would lead to a better transferability of research into the complex and dynamic environment of practical call center arrival forecasting and, at the same time, provide valuable insights into the effect of customer contact activities on customer behavior.

3 Dynamic harmonic regression with predictor variables

The methods used in previous call center forecasting studies presented in Section 2 only allow for the inclusion of information from past observations of a time series or the incorporation of external data from predictor variables. Hence, either contextual information possibly stemming from predictor variables or information extracted from time series dynamics is lost. The latter applies to ordinary regression models of form

$$y_t = \beta_0 + \beta_1 x_t + \cdots + \beta_i x_{i,t} + N_t \tag{1}$$

with i predictor variables $x_{i,t}$ at time t and the time series' value y_t at time t. The error term is mostly assumed to be a set of zero-mean and normally distributed white noise random shocks a_t (Pankratz, 1991). Thus, if N_t in (1) has mean zero and is normally distributed white noise, then $N_t = a_t$. However, it is problematic estimating ordinary regression models of (1) with time series data (Pankratz, 1991). As stated earlier, regression models cannot capture previous dynamics of a time series: E.g., the error term might be autocorrelated, i.e., N_t is related to its previous values N_{t-1}, N_{t-2}, etc., i.e.,

$$N_t = \phi_1 N_{t-1} + a_t \tag{2}$$

with coefficient $\phi_1$and random shock component a_t. Alternatively,

$$N_t = a_t - \theta_1 a_{t-1} \tag{3}$$

with coefficient θ_1 and a_{t-1} being the random shock component of N_{t-1}. Thereby, equation (2) represents an autoregressive process whereas equation (3) displays a moving average process. Hence, combining both equations yields an ARIMA process. If we allow the error term N_t in (1) to contain autocorrelation of (2) and (3), we obtain a dynamic regression model

$$y_t = \beta_0 + \beta_1 x_t + \cdots + \beta_i x_{i,t} + \eta_t \tag{4}$$

with then η_t being the ARIMA process depicted in equations (2) and (3). The resulting regression model thus is able to capture previous dynamics of a time series.

In harmonic regression models, the observed time series is considered as being composed of a signal, i.e., consisting of a sum of sinusoidal terms (or Fourier terms respectively) (Bloomfield, 2000), so that any time series can be expressed as a combination of cosine (or sine) waves with differing periods. That is, the variation of a time series may be modeled as the sum of k different individual sinusoidal terms (harmonics) occurring at different frequencies of periodic variation ω (Bloomfield, 2000). Thus, a harmonic regression model can be defined as

$$y_t = \sum_{k=1}^{K} (\alpha_k \cos(\omega_k t) + \gamma_k \sin(\omega_k t)) + e_t \tag{5}$$

with white noise error e_t, coefficients α_k and γ_k, and $\omega_k, t = 1,2,.., N$ being the frequencies of periodic variation.

Combining both the autocorrelated error term of (4) and the Fourier terms of (5) yields a DHR model (Young et al., 1999; Young, 1999) using Fourier terms as predictors in combination with dynamic regression to handle periodic seasonality (Hyndman & Athanasopoulos, 2018)

$$y_t = \beta_0 + \sum_{k=1}^{K} (\alpha_k \cos(\omega_k t) + \gamma_k \sin(\omega_k t)) + \eta_t \tag{6}$$

with η_t being a non-seasonal ARIMA (p,d,q) process. The DHR as in (6) and its extensions have already been utilized by research to forecast sub-daily call arrivals (Taylor, 2008; Tych et al., 2002) since it allows for long seasonal periods compared to ARIMA and ETS models and short-term dynamics are handled by the ARIMA error (Hyndman & Athanasopoulos, 2018). Nevertheless, the DHR model in (6) does not include additional contextual information such as the effect of holidays, catalog mailings, or billing cycles. As mentioned in the previous section, prior research (Aldor-Noiman et al., 2009; Andrews & Cunningham, 1995) found such information to substantially enhance forecast accuracy and thus, recommended to include predictor variables (Taylor, 2008). Therefore, we extend the DHR model in (6) from extant call center literature by adding predictor variables aside from Fourier terms:

$$y_t = \beta_0 + \beta_1 x_t + \cdots + \beta_i x_{i,t} + \sum_{k=1}^{K} (\alpha_k \cos(\omega_k t) + \gamma_k \sin(\omega_k t)) + \eta_t \tag{7}$$

4 Analysis of customer support call arrival series

4.1 Preliminary data analysis

We gathered call center data from a leading German online retailer for fashion. The retailer's call center comprises arrival queues for each customer support, order taking, customer complaints, and consultation service. We investigate the customer support queue for e-mail arrivals as well as for call arrivals. The latter is open from 7 a.m. to 10 p.m. from Monday through Saturday. Since our proposed DHR model is computationally only capable of modeling one seasonality, we apply a common two-step temporal aggregation approach (Kourentzes et al., 2017): first, we aggregate our original high sampling frequency data (half-hourly data) at a pre-specified aggregation level with lower frequency (daily values) and thus, predict the daily arrival volume. Then, we disaggregate the daily predictions with respect to the averaged arrival distribution per weekday to re-yield the original high frequency, i.e., half-hourly predictions. The described temporal aggregation approach has gained substantial attention in recent methodological forecasting literature (Boylan & Babai, 2016; Kourentzes et al., 2014; Kourentzes & Petropoulos, 2016; Nikolopoulos et al., 2011) as it smooths the original time series, removes noise, improves forecast accuracy, and particularly simplifies the generation of forecasts (Kourentzes et al., 2017).

Hence, our aggregated daily data contain 1,045 observations from January 2, 2016 to May 4, 2019, i.e., 174 weeks of data. One week comprises six observations and one year contains 312.25 observations considering leap years. The maximum number of call arrivals per day are 5,300 arrivals and the mean of call arrivals per day is 2,105.56. Our data exhibits overdispersion with a variance of 675,914.17. We excluded two weeks of data (i.e., 12 observations) due to incorrect interval capturing.

To enable the use of time series models, the time series has to be stationary. We conduct an Augmented Dickey Fuller (ADF) test (Dickey & Fuller, 1979) to check for unit root in our data. With a p-value of 0.99 at lag order 312 (value of test statistic .1139), we cannot reject the null hypothesis of unit root in our data and thus, assume our data to be nonstationary.

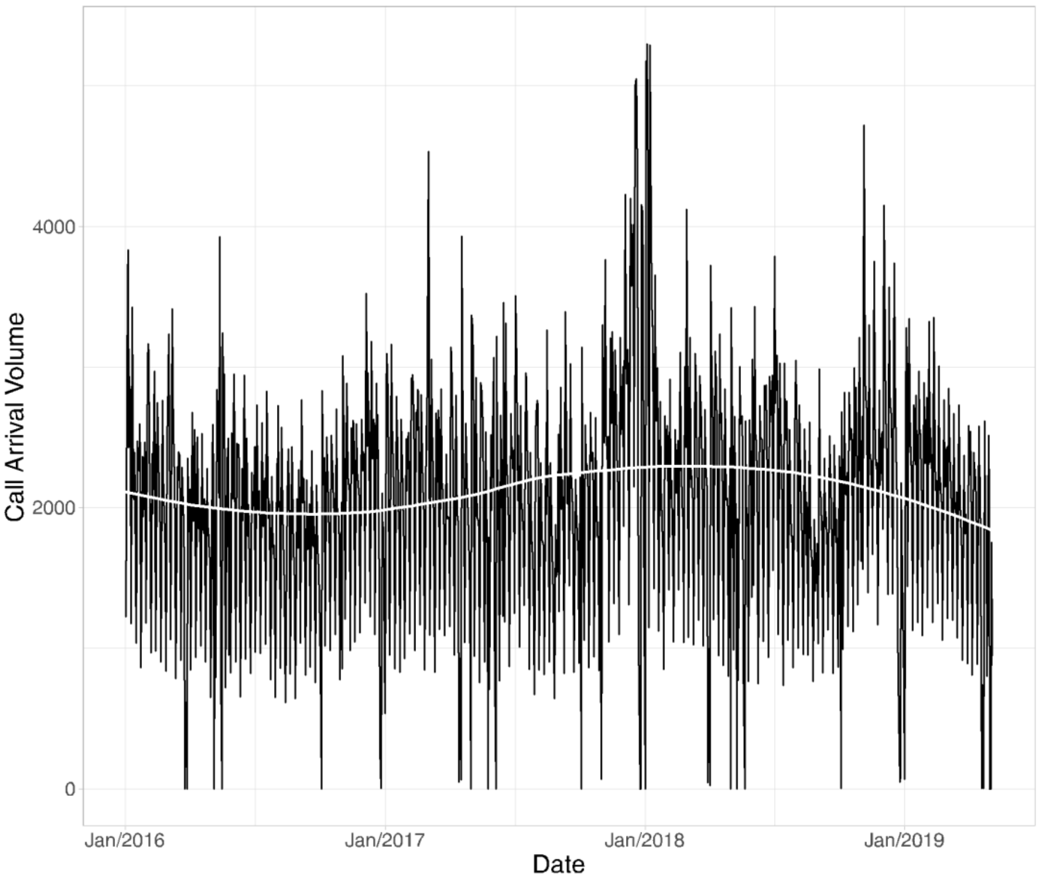

Figure 1: Overall call arrival volume of customer support queue.

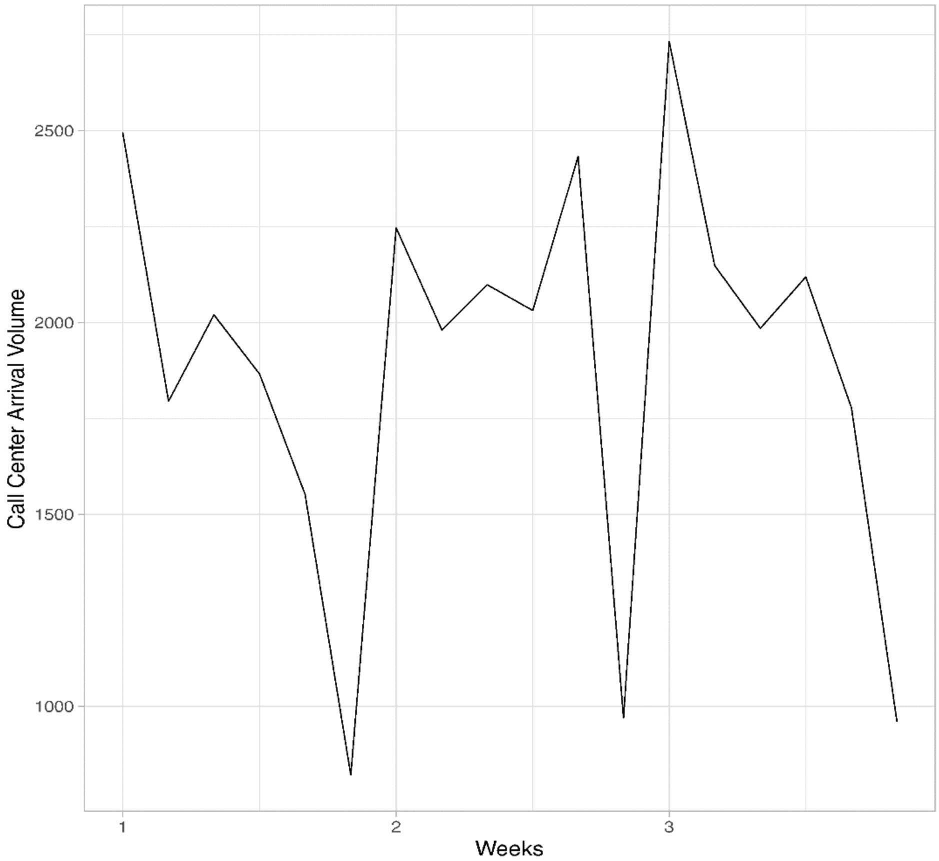

Figure 2: Call arrival volume of three consecutive weeks.

To determine appropriate frequencies for our DHR model and further, to make the time series stationary, we have to assess the degree of seasonality in our data. Figure 1 depicts the overall daily call arrival volume of the customer support queue and its smoothed trend. The volume remains relatively constant with a slight increase towards the beginning of the year 2018 and a decrease towards the end of the dataset. Figure 2 displays three consecutive weeks of data. The number of call arrivals peaks on Mondays, then decreases throughout the week, and exhibits a second peak during the course of the week. The arrival volume drops substantially on Saturdays. As this pattern repeats every week, we assume daily seasonality, i.e., $s = 6$. We do not assume yearly seasonality as the yearly pattern is rather weak or non-existent respectively, considering Figure 1. More formally, these findings further have been confirmed by the data's periodogram with a peak value at frequency $\omega = .1667$ and thus, the dominant period $T = \frac{1}{\omega}$ is 5.9999, i.e., it takes approximately six days to complete a full cycle. The periodogram does not indicate yearly seasonality.

We model different predictor variables for our data to improve forecast accuracy, summarized in Table 1. We utilize previous findings of literature regarding useful predictor variables (Aldor-Noiman et al., 2009; Andrews & Cunningham, 1995; Ibrahim et al., 2016) and include additional, potentially meaningful variables indicated by call center management or identified in preceding research (Rausch and Albrecht, 2020).

Table 1: Predictor variables.

Variable	Description	References
Day-of-the-week	Nominal variable: six values (Monday to Saturday) to capture the day-of-the-week-effect	(Aldor-Noiman et al., 2009; Ibrahim et al., 2016; Ibrahim & L'Ecuyer, 2013)
Holiday	Nominal variable: 16 values (15 public holidays and ordinary weekdays) to capture the effect of German public holidays	(Andrews & Cunningham, 1995)
Day-after-holiday	Dummy variable: two values (days after public holidays and ordinary weekdays) to capture the effect of days after public holidays	
Outlier	Nominal variable: four values (extreme outliers and outliers (marked as such by management: if an extraordinary high number of arrivals can be explained by e.g. a special offer), days on which the call center is closed, and ordinary weekdays) to capture the effect of outliers	
School holidays	Metric variable indicating the number of German states having school holidays to capture the effect of German school holidays	
Year	Nominal variable: eight values (semiannual sections from January 2016 to May 2019) to capture the effect of busier seasons and the long-term development of the arrivals' level	
CW0-3	Four dummy variables to capture the temporal effect of catalog mailings during the first weekend (CW0), the first week (CW1), the second week (CW2), and the third week (CW3) after release to capture the temporal effect of catalog mailings	(Aldor-Noiman et al., 2009; Andrews & Cunningham, 1995)
MMail1-2 MPost1-2 DMail1-2	Six dummy variables to capture the temporal effect of e-mail reminders (MMail) as well as postal reminders (MPost) and due date e-mails (DMail) on the day of delivery (MMail1, MPost1, and DMail1) and the day after (MMail2, MPost2, and DMail2)	(Aldor-Noiman et al., 2009)

4.2 Experimental design

To evaluate the performance of our proposed DHR model with predictor variables, we draw on several standard forecasting techniques of those presented in Section 2 for comparison listed in Table 2. Thereby, we utilize different time series models such as ARIMA, ETS, RW, TBATS, and further, the standard DHR approach without predictor variables as well as common high-performance ML methods such as RF and gradient boosting with L1 regularization (GBR) as benchmark approaches since they have been found to outperform other models by extant forecasting research.

Since our time series in nonstationary, we apply time series decomposition before generating predictions with the time series approaches[14]. The seasonal-trend decomposition based on Loess (STL) (Cleveland

[14] One of the anonymous reviewers pointed out that ETS models do not necessarily need time series decomposition prior to generating predictions. Nevertheless, although all ETS models are nonstationary, we also decompose the

et al., 1990) detrends and deseasonalizes the data yielding $y_t = \hat{S}_t + \hat{A}_t$ for additive decomposition ($y_t = \hat{S}_t * \hat{A}_t$ for multiplicative decomposition respectively) with $\hat{A}_t = \hat{T}_t + \hat{R}_t$ for additive decomposition ($\hat{A}_t = \hat{T}_t * \hat{R}_t$ for multiplicative decomposition respectively). Thereby, $\hat{S}_t$ is the seasonal component and $\hat{A}_t$ is the data without seasonality, i.e., the seasonally adjusted component. Both components are forecasted separately. The former is predicted by drawing on the last period of the estimated component, which equals a seasonal naïve method, whereas any non-seasonal forecasting approach can be utilized for the latter. The transformations of the decomposed time series are then inverted to yield the forecasts of the original time series $\{Y_t\}$ (Hyndman & Athanasopoulos, 2018).

To select appropriate predictor variables and filter out uninformative ones for our proposed DHR model, we draw on the forward variable selection procedure (Hyndman & Athanasopoulos, 2018). A prevailing approach to identify essential predictor variables is to drop those variables whose p-values are statistically insignificant (Aldor-Noiman et al., 2009). However, in a forecasting context, the p-value does not necessarily determine the variable's predictive performance regarding the out-of-sample predictions which are practically relevant. Hence, we begin with the null model comprising none of the variables and add each predictor variable at a time. The variable is maintained if it enhances forecast accuracy. This step is repeated until no further improvement of accuracy is yielded.

Table 2: Models for comparison.

	Model	Description
Time Series Models	**DHR**	Combination of a dynamic regression model and a harmonic regression model, i.e., a regression model with Fourier terms as predictors and an ARIMA error term (Young et al., 1999; Young, 1999).
	STL + (non-seasonal) ARIMA	Combination of a seasonal-trend decomposition of time series based on Loess and a non-seasonal ARIMA (p,d,q) model, i.e., predictions are generated based on prior values y_{t-p} of y_t and prior errors ε_{t-q} (Box & Jenkins, 1970).
	STL + ETS	Combination of a seasonal-trend decomposition of time series based on Loess and an exponential smoothing innovation state space model (exponential smoothing model with an error term), i.e., predictions are the exponentially weighted average of past observations (Hyndman et al., 2002).
	STL + RW with drift (RWDRIFT)	Combination of a seasonal-trend decomposition of time series based on Loess and a random walk model, i.e., predictions equal the last observation and the average of changes between consecutive observations.
	TBATS	Exponential smoothing innovation state space model (ETS model) with a Box-Cox transformation (stabilizes the time series' variance), Fourier terms (trigonometric expression of seasonality terms for complex seasonality as well as high frequency of seasonality; allow seasonality to change over time), and an ARMA (p,q) correction (De Livera et al., 2011).

time series before applying the ETS model as predictions are assumed to become more accurate by detrending and deseasonalizing the time series first compared to predictions based on the global series (Theodosiou 2011).

	Model	Description
ML Approaches	**GBR**	Ensemble of successive weak learners (i.e., models that achieve accuracy just above random guessing): Within boosting, weak learners are trained sequentially trying to correct its respective predecessor (Schapire et al., 1998). I.e., each learner is constructed using feedback from previously grown learners. More specifically, within gradient boosting, a subclass of boosting, weak learners are fitted to the residual errors made by preceding learners and gradient descent is used to identify the errors in previous predictions (Friedman, 2001, 2002).
	RF	Ensemble of successive decision trees: Withing bagging, each learner is grown independently from earlier learners. I.e., each tree is built using a bootstrap sample of the data (Breiman, 1996). More specifically, within random forests, a subclass of bagging, the algorithm draws n_{tree} bootstrap samples, grows an unpruned regression tree for each sample by randomly sampling m_{try} of the predictors at each node, and then chooses the best split among them. The outputs of the n_{tree} trees are aggregated and averaged to produce one final prediction (Breiman, 2001; Liaw & Wiener, 2002). An additional L1 regularization prevents the model from overfitting.

We evaluate model performance with time series cross-validation and an expanding rolling window. We use 118 weeks of data comprising 709 observations, i.e., the observations from January 2, 2016 to April 7, 2018, as our initial training data. We fit the models and predict one week or six observations respectively (i.e., forecast horizon $h = 6$). During the next iteration, we roll the training data one week forward, re-estimate our models, and predict one unit of our forecast horizon further. During each iteration $n = 1,2,..,N$, the ML models' hyperparameters are optimized by implementing 10-fold cross-validation with grid search. We further optimize the number of Fourier terms k of the DHR models by including a second loop within each iteration n: Since k can have a maximum value of $T/2$, the grid search for k is set within the range $[1; 3]$ and $k \in \mathbb{N}$. k is then optimized by fitting the model with the current training data and predicting the current out-of-sample data during each iteration $n = 1,2,..,N$. Forecast accuracy for the out-of-sample predictions is calculated for each k and subsequently, k is chosen with respect to the highest forecast accuracy results. Overall, the procedure is repeated 52 times, i.e., for one year, and hence, $N = 52$. As stated earlier, we exclude two weeks of data from October 22, 2018 to November 4, 2018 due to incorrect interval capturing and therefore, we predict 300 daily call arrival volumes.

As stated earlier, to yield relevance for practical use, we further disaggregate our daily predictions with respect to the averaged call arrival distribution for each weekday per half-hour interval to invert the daily predictions back to the series' original half-hour frequency, see Figure 3. We additionally include the averaged call distribution of holidays as they a divergent distribution compared to ordinary weekdays. Evidently, Mondays are the busiest days with a peak in the morning hours and a second smaller peak throughout the day. The remaining weekdays exhibit a similar course on a lower level. Saturdays register the fewest call arrivals on average throughout the day aside from holidays. Overall, each of the 300 predicted daily values is disaggregated into 30 half-hour intervals yielding a total of 9000 predictions.

To assess the models' performance, we compare the sub-daily forecasts with the out-of-sample or test data (i.e., the actual values) and compute forecast accuracy. We draw on mean absolute error (MAE) and root mean squared error (RMSE) as error measures:

$$\text{MAE} = \frac{1}{T}\sum_{i=1}^{T} |Y_i - \hat{Y}_i| \qquad \text{RMSE} = \sqrt{\frac{1}{T}\sum_{i=1}^{T} (Y_i - \hat{Y}_i)^2}$$

with test (or out-of-sample) data Y_i, forecasted values $\hat{Y}_i$, and the number of forecasted values T. Since they are scale-dependent measures, they are both appropriate to compare predictions on the same scale. Both MAE and RMSE are frequently used by research to determine their forecasts' accuracy (Aldor-Noiman et al., 2009; Barrow, 2016; Ibrahim et al., 2016; Taylor, 2008; Weinberg et al., 2007) as they can be calculated and interpreted easily (Hyndman & Athanasopoulos, 2018).

We further check the robustness of our comparison results by considering (1) no lead time, (2) one week lead time, (3) two weeks lead time, and (4) three weeks lead time. In this context, lead time refers to the period between the dataset's last actual observation and the first created forecast.

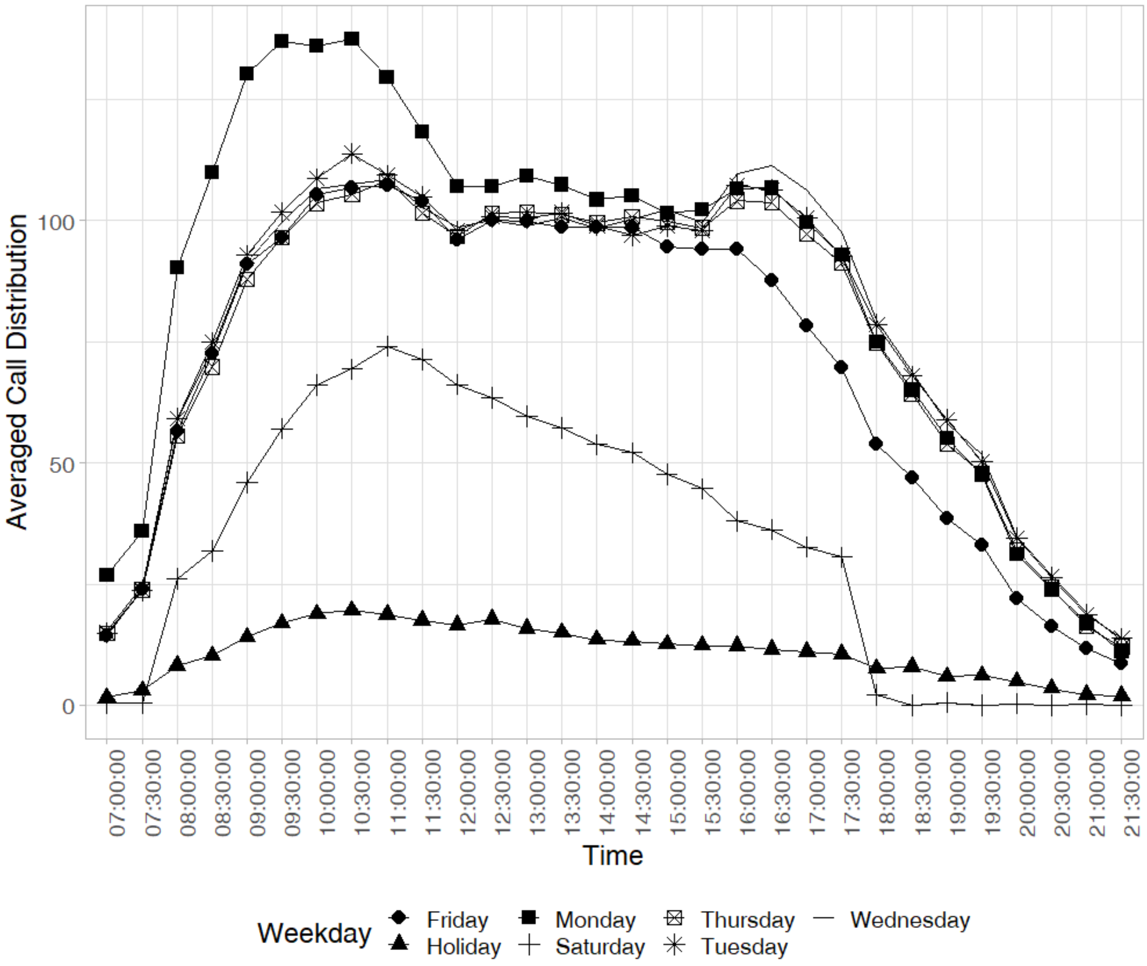

Figure 3: Averaged call arrival distribution per weekday.

4.3 Results

As stated earlier, we conduct forward variable selection for our proposed DHR model. Drawing on the results in Table 3, the variables year, outlier, school holidays, day-after-holiday, and CW2 do not enhance forecast accuracy regardless of the considered lead time constellation whereas the variables day-of-the-week, holiday, CW1, MMail2, MPost1, MPost2, and DMail2 improve accuracy for every lead time. The highest MAE improvement is yielded by the day-of-the-week and the holiday variable indicating their importance for the call center arrival forecasts.

Table 3: Forward variable selection MAE results for customer support call arrival forecasts.

	No lead time	**One week**	**Two weeks**	**Three weeks**
No predictor variables	18.6089	19.2526	19.7547	20.0912
Day-of-the-week	**13.5712**	**13.6240**	**13.8336**	**14.1911**
Holiday	**11.7666**	**11.8471**	**11.9814**	**11.8789**
Day-after-holiday	11.8434	11.9490	12.0462	12.0105
School holidays	11.8158	11.8982	12.0792	12.1299
Outlier	11.9839	12.0417	12.1656	12.1069
Year	12.1754	12.2806	12.4242	12.4873
CW0	**11.7389**	11.8578	**11.9731**	11.8851
CW1	**11.7127**	**11.8451**	**11.9690**	**11.8648**
CW2	11.7655	11.8510	11.9737	11.8873
CW3	11.7244	**11.8144**	11.9730	11.8720
MMail1	11.7489	11.8264	11.9795	**11.8644**
MMail2	**11.5700**	**11.7269**	**11.9324**	**11.7686**
MPost1	**11.3872**	**11.4660**	**11.6455**	**11.6049**
MPost2	**11.3369**	**11.3959**	**11.5747**	**11.4863**
DMail1	**11.3140**	**11.2669**	11.5805	**11.4648**
DMail2	**11.2336**	**11.2250**	**11.5571**	11.4688
All	11.5466	11.6480	11.9878	12.0333

Note: The bold values are an improvement to the respective preceding value and the corresponding variable is included in the final model.

Table 4 and 5 present the MAE and RMSE results of the investigated models. Our proposed DHR model with predictor variables outperforms the remaining models with respect to its MAE results for every considered lead time. Considering two weeks and three weeks of lead time, RF performs slightly better than our DHR model with regard to RMSE as evaluation metric. As the RMSE gives a higher weight to large errors, this indicates that our model made fewer large errors compared to RF. Nevertheless, as the discrepancy in RMSE between both models is only around 0.2, this seems rather negligible. This further indicates the superiority and importance of contextual information.

The decomposed ARIMA, ETS, and RW models yield the most inaccurate forecasts and among the time series models while the standard DHR model without predictor variables is the second best performing model. Comparing both DHR models, predictor variables apparently enhance forecast accuracy substantially. Further, since the ML models comprise predictor variables, their forecasts are comparable but slightly worse than those of our proposed model. Overall, ML models are superior to the time series models.

Evidently, ex-post forecasts (i.e., their predictor variables can be modeled for both past observations (the training data) as well as future observations (the out-of-sample data)) of our proposed model and the ML models are outperforming ex-ante forecasts (i.e., the models are only using information that is available at the time of generating the forecasts) of the time series models used.

Considering the lead times, the present results support previous findings in call center forecasting literature (Ibrahim et al., 2016; Rausch and Albrecht, 2020). Forecast accuracy declines steadily with increasing lead time for most of the models and the most accurate predictions for each model are yielded without any lead time.

Table 4: MAE results for customer support call arrival forecasts.

	No lead time	**One week**	**Two weeks**	**Three weeks**
DHR with predictor variables	**11.2336**	**11.2250**	**11.5571**	**11.4648**
DHR	14.3198	14.2243	14.3913	14.5242
STL+ARIMA	17.8254	17.3990	17.7603	18.2596
STL+ETS	17.8028	17.4140	17.8213	18.4370
STL+RWDRIFT	17.7657	17.5317	17.7757	18.2407
TBATS	16.3675	17.1001	17.2664	17.6892
GBR	13.4358	13.7363	13.7492	13.5698
RF	12.0249	12.0516	12.3048	12.3006

Note: The highest forecast accuracy for each lead time is marked in bold.

Table 5: RMSE results for customer support call arrival forecasts.

	No lead time	**One week**	**Two weeks**	**Three weeks**
DHR with predictor variables	**15.7521**	**16.6844**	17.0088	17.1805
DHR	22.4746	22.9245	23.1713	23.4540
STL+ARIMA	30.3195	29.0067	29.3980	30.2608
STL+ETS	30.1632	28.9110	29.4276	30.4809
STL+RWDRIFT	30.1719	28.9619	29.1190	29.8225
TBATS	24.8822	26.0952	26.3939	27.0433
GBR	18.6523	19.2414	19.0341	18.7706
RF	16.6278	16.7476	**16.9844**	**16.9797**

Note: The highest forecast accuracy for each lead time is marked in bold.

4.4 Robustness checks

To test the robustness of our results, we conducted further analyses. Although the temporal aggregation approach is assumed to remove noise and enhance forecast accuracy, it would not have been mandatory for the benchmark time series models, as these are computationally capable of modeling more than one seasonality[15]. Thus, we generated forecasts based on the original series (i.e., sub-daily data) without the temporal aggregation approach (see Table 6 and 7). Although forecast accuracy partially improves, our proposed DHR model with predictor variables is still superior in terms of MAE results. Nevertheless, without temporal aggregation, RF's RMSE values for forecasts without lead time and with one as well as two weeks lead time are slightly better than those of the DHR model.

Table 6: MAE results for customer support call arrival forecasts based on original sub-daily data without temporal aggregation.

	No lead time	**One week**	**Two weeks**	**Three weeks**
STL+ARIMA	14.5263	14.7404	15.5448	15.8520
STL+ETS	14.5263	14.7407	15.5448	15.8520
STL+RWDRIFT	14.6651	14.6334	15.2941	15.7877
TBATS	18.1162	17.9856	18.8661	18.9823
GBR	12.9393	13.1488	13.3987	13.7386
RF	11.7544	11.8129	12.0648	12.8134

Table 7: RMSE results for customer support call arrival forecasts based on original sub-daily data without temporal aggregation.

	No lead time	**One week**	**Two weeks**	**Three weeks**
STL+ARIMA	22.7009	23.1810	24.2187	25.0726
STL+ETS	22.9251	23.0876	23.9239	24.7768
STL+RWDRIFT	23.0503	23.1555	23.9506	24.7793
TBATS	30.8184	30.7156	31.6453	31.8839
GBR	18.1216	18.3299	18.6043	19.3079
RF	15.5678	16.6541	16.8929	18.4903

[15] Additionally, we generated forecasts based on the original sub-daily data with a double seasonal exponential smoothing model as suggested by one of the anonymous reviewers and found our models to be noticeably superior.

5 Analysis of customer support e-mail arrival series

5.1 Preliminary data analysis

To check the robustness of our results, we additionally investigate the e-mail arrivals of the customer support queue. The incoming e-mail data of this queue varies from the previous call analysis in the number of available observations per week, the level of average and maximum arrival count per interval, and the existence of trend in the arrival volume.

The aggregated daily data consist of 1,220 observations from January 2, 2016 to May 5, 2019, i.e., 174 weeks of data. Since e-mails arrive at any time throughout the day and on every weekday from Mondays to Sundays, one week consists of seven observations and one year comprises 365.25 observations considering leap years. The maximum number of e-mail arrivals is 3,240 per day. The customer support queue receives 1,670.15 e-mails on average each day and with a variance of 283,687.87 the data is overdispersed. With a p-value of 0.98 at lag order 365 (value of test statistic -.4533) of the ADF test, we cannot reject the null hypothesis of unit root in our data and thus, assume our data to be nonstationary.

Considering Figure 4, there is a steady upwards trend in the overall e-mail arrival volume. Figure 5 reveals that the e-mail arrival volume is high from Mondays to Fridays and drops noticeably on weekends. Since this pattern repeats every week, we assume weekly seasonality for the daily data, i.e., $s = 7$. Further, there is no considerable yearly seasonality. More formally, the data's periodogram exhibits a peak value at frequency $\omega = .1432$ and thus, the dominant period $T = \frac{1}{\omega}$ is 6.9832, i.e., it takes approximately seven days to complete a full cycle. The periodogram does not indicate a frequency which implies yearly seasonality.

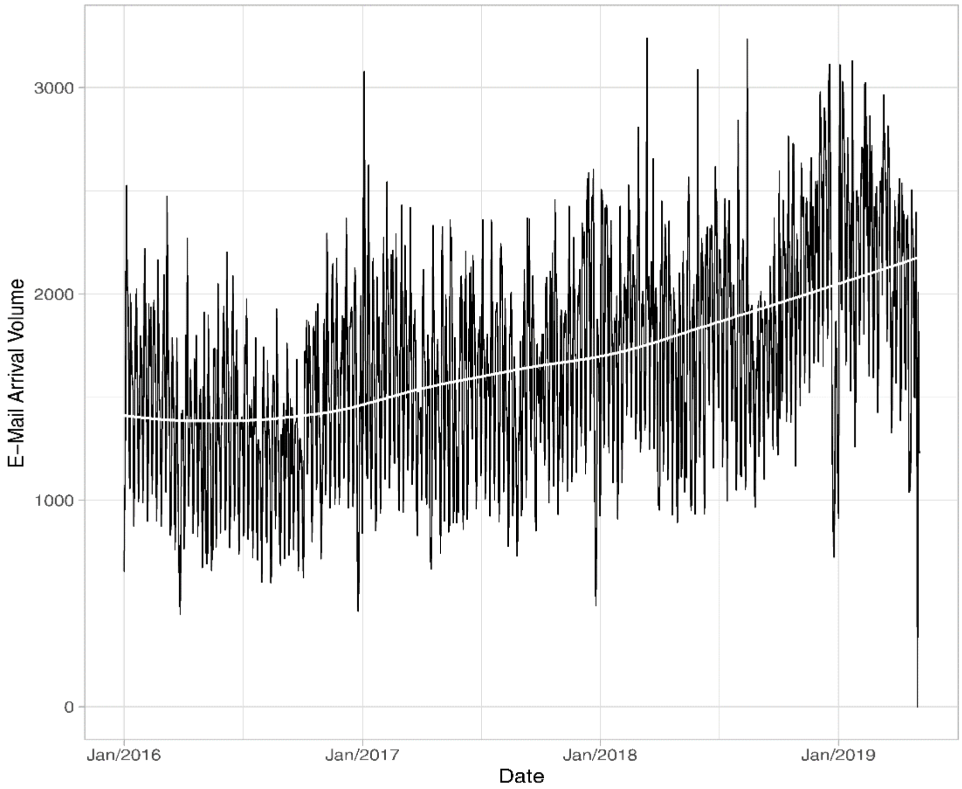

Figure 4: Overall e-mail arrival volume of customer support queue.

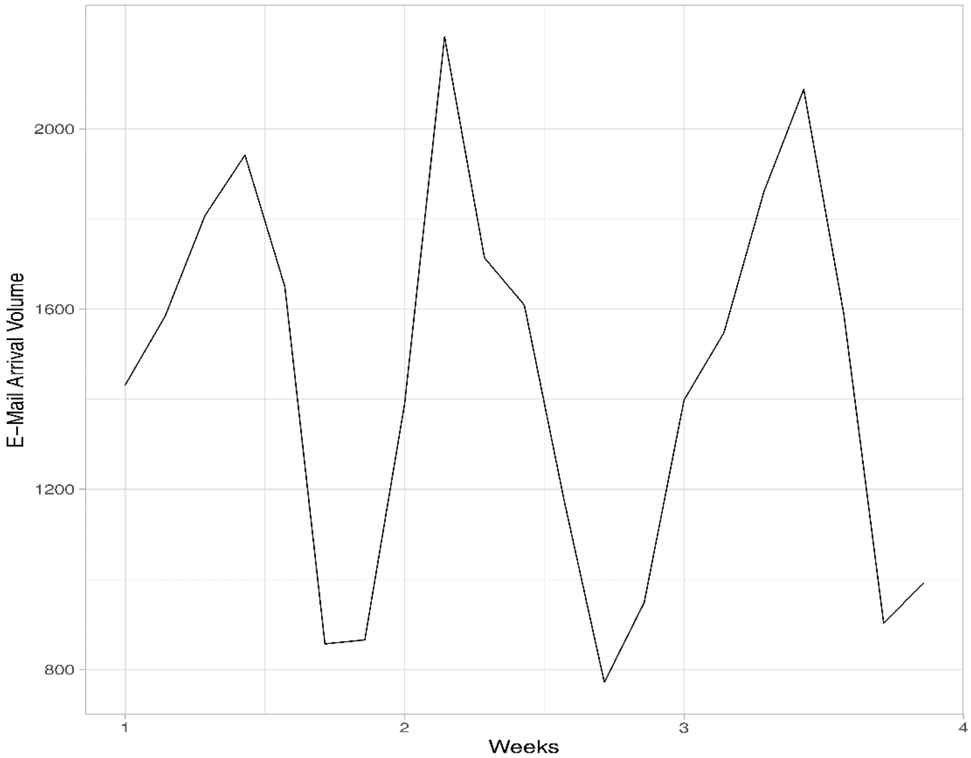

Figure 5: E-Mail arrival volume of three consecutive weeks.

We use the experimental design and predictor variables described in Section 4.2 and 4.1 respectively. Accordingly, we extend the day-of-the-week predictor variable as well as the forecast horizon h to seven days. To determine the averaged e-mail distribution per weekday, we draw on the original hourly dataset comprising 29,280 observations and average the cumulated e-mail arrivals for each interval per weekday (see Figure 6). Evidently, Mondays to Fridays exhibit a similar distribution with few e-mail arrivals during the night, a peak in the morning hours, and a second smaller peak throughout the day. The volume drops towards the end of the day. On Saturdays, Sundays, and holidays the e-mail arrival volume is relatively low but has a similar distribution.

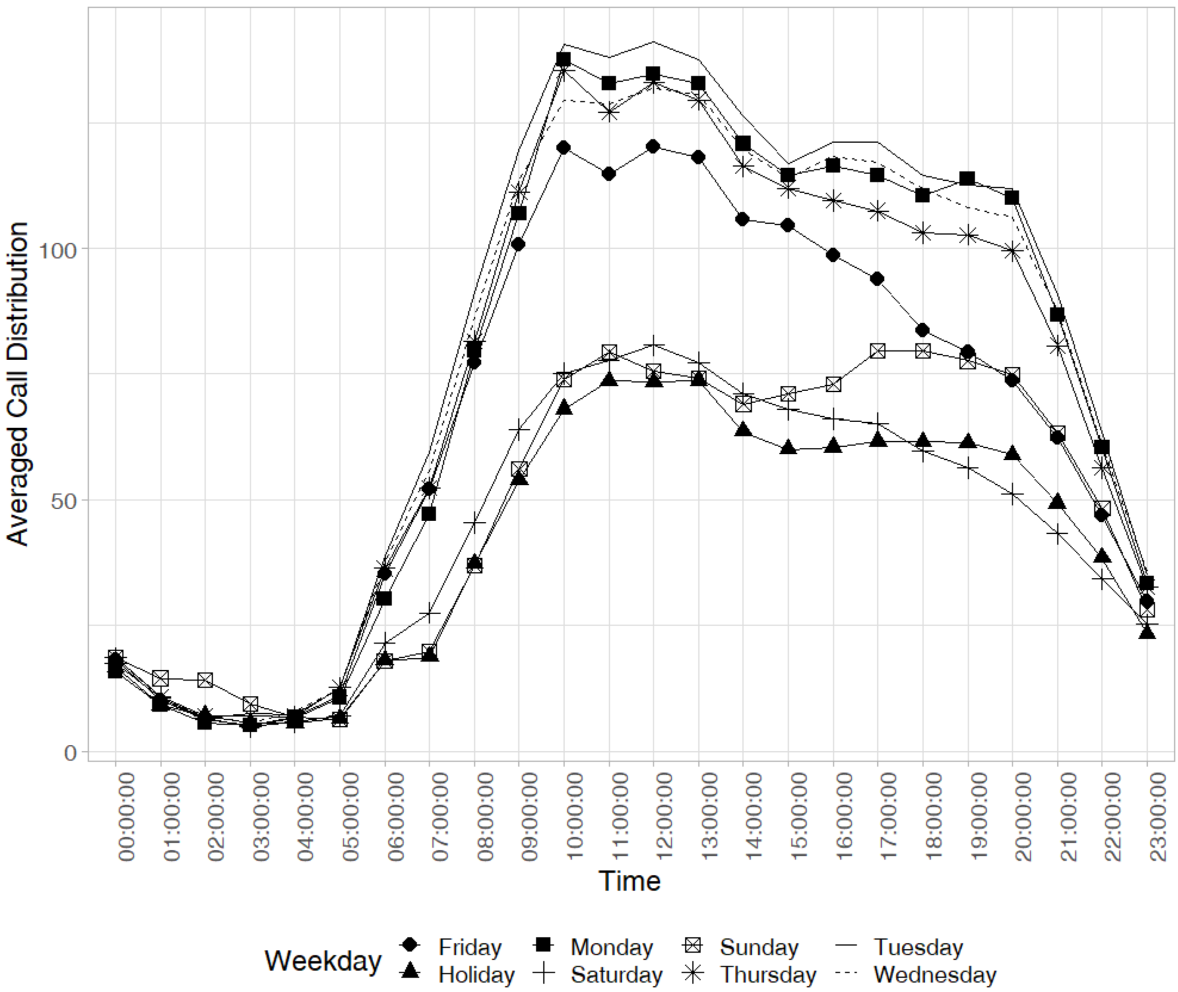

Figure 6: Averaged e-mail arrival distribution per weekday.

5.2 Results

Similarly to the customer support call arrival series, we conduct forward variable selection for our DHR model listed in Table 8. The variables day-after-holiday, school holidays, outlier, year, CW1, CW2, and DMail1 do not improve forecast accuracy regardless of the considered lead time constellation, whereas the variables day-of-the-week, holidays, CW0, and MMail1 have a positive impact on accuracy for every lead time. Analogously to the call arrival series, the day-of-the-week and holiday variables cause the highest MAE improvement and are assumed to be important elements contributing to the present prediction results.

Table 8: Forward variable selection MAE results for customer support e-mail arrival forecasts.

	No lead time	**One week**	**Two weeks**	**Three weeks**
No predictor variables	16.4636	17.3453	18.1969	18.5699
Day-of-the-week	**14.6233**	**14.7007**	**15.0112**	**15.1020**
Holiday	**14.2472**	**14.3957**	**14.5968**	**14.5782**
Day-after-holiday	14.2547	14.4004	14.5999	14.5856
School holidays	14.2520	14.4547	14.6330	14.6229
Outlier	14.2759	14.4185	14.5990	14.6019
Year	14.3672	14.5808	14.8253	14.8272
CW0	**14.2470**	**14.3674**	**14.5592**	**14.5604**
CW1	14.2881	14.3853	14.5636	14.5761
CW2	14.2483	14.3826	14.5679	14.5792
CW3	14.2476	14.3685	**14.5588**	14.5686
MMail1	**14.0404**	**14.2270**	**14.4386**	**14.4538**
MMail2	**14.0306**	**14.2055**	14.4413	14.4627
MPost1	14.0497	14.2740	14.4738	**14.4513**
MPost2	**14.0283**	14.2080	**14.4336**	14.4520
DMail1	14.0533	14.3239	14.5907	14.5382
DMail2	**13.9974**	**14.1827**	14.4355	**14.4230**
All	14.2046	14.4940	14.8711	14.9533

Note: The bold values are an improvement to the respective preceding value and the corresponding variable is included in the final model.

Table 9 and 10 summarize the MAE and RMSE results for the customer support e-mail arrival forecasts. Although the customer support queue receives less e-mails than calls on average, the MAE and RMSE obtained for the customer support e-mail arrival predictions are comparable to those of the customer support call arrival forecasts. Similarly to the call arrivals analysis, the proposed DHR model with predictor variables outperforms the remaining time series and ML approaches for every considered lead time constellation. Forecast accuracy mainly declines with higher lead times while the best performance for each model is yielded without lead time.

For the customer support e-mail arrivals, the performance gap between time series and ML models is not as evident as for the call arrivals. Under the conditions of a more apparent trend in the examined data, time series models operate in a comparable performance range as ML models. The DHR model without predictor variables yields the second-best forecasts indicating that the Fourier terms itself have a high predictive potential. The additional predictor variables in our model further enhance this predictive power. Among the ML models, the RF algorithm outperforms the gradient boosting approach with L1 regularization.

Table 9: MAE results for customer support e-mail arrival forecasts.

	No lead time	One week	Two weeks	Three weeks
DHR with predictor variables	**13.9974**	**14.1827**	**14.4336**	**14.4230**
DHR	14.9317	15.2102	15.4306	15.5281
STL+ARIMA	15.5983	15.6356	15.8708	15.9973
STL+ETS	15.7058	15.7584	15.9991	16.1562
STL+RWDRIFT	17.1299	17.6039	17.6647	17.9944
TBATS	15.6257	15.9940	16.2795	16.3678
GBR	16.6198	18.5997	17.3449	17.1729
RF	15.9811	16.6069	17.0371	17.1368

Note: The highest forecast accuracy for each lead time is marked in bold.

Table 10: RMSE results for customer support e-mail arrival forecasts.

	No lead time	One week	Two weeks	Three weeks
DHR with predictor variables	**25.8450**	**26.3258**	**26.7295**	**26.2177**
DHR	27.2817	27.7663	27.9761	28.8035
STL+ARIMA	28.3968	28.4400	28.7987	29.0450
STL+ETS	28.4915	28.5459	28.9247	29.2177
STL+RWDRIFT	30.2348	31.1480	31.4703	31.5576
TBATS	28.5707	29.1941	29.4854	29.5885
GBR	28.9195	32.0761	29.9215	29.6442
RF	28.1436	28.8698	29.3342	29.4563

Note: The highest forecast accuracy for each lead time is marked in bold.

5.3 Robustness checks

Similarly to the customer support call arrivals, we conducted additional analyses to test the results' robustness. We generated predictions based on the original sub-daily data without the temporal aggregation approach (see Table 11 and 12)[16]. Forecast accuracy does not improve for the time series

[16] Similar to the customer support call arrivals, we generated forecasts based on the original sub-daily data with a double seasonal exponential smoothing model as suggested by one of the anonymous reviewers and found it to be outperformed by all other benchmark models.

models, but for RF. Overall, our proposed DHR approach with predictor variables still outperforms the benchmark models in terms of both MAE as well as RMSE and for every lead time constellation.

Table 11: MAE results for customer support e-mail arrival forecasts based on original sub-daily data without temporal aggregation.

	No lead time	One week	Two weeks	Three weeks
STL+ARIMA	15.8875	17.0814	17.8252	17.8173
STL+ETS	17.1596	18.6449	19.1254	19.5232
STL+RWDRIFT	17.7435	19.1402	19.6219	20.1637
TBATS	19.8545	20.8313	21.7826	21.9521
GBR	18.5135	19.0513	17.5940	17.4820
RF	15.5333	16.1990	16.8110	16.9005

Table 12: RMSE results for customer support e-mail arrival forecasts based on original sub-daily data without temporal aggregation.

	No lead time	One week	Two weeks	Three weeks
STL+ARIMA	28.5696	30.0853	31.0859	30.7144
STL+ETS	29.4874	31.4311	32.1550	32.3495
STL+RWDRIFT	20.2880	31.9048	33.1407	33.4434
TBATS	32.0302	33.2508	34.2860	34.3550
GBR	32.4089	29.5703	30.2263	30.3050
RF	27.7846	28.7869	29.2917	29.3207

6 Discussion

The findings of the analysis using call as well as e-mail arrival data from the customer support queue of the call center demonstrate clear benefits of the use of our proposed model. In line with the hypothesis that combining the strengths of different forecasting model types will lead to an increase in prediction performance and, at the same time, entail advantages for the use in practice, the DHR model with predictor variables outperforms other approaches investigated. Thereby, we contribute not only to the existing body of literature in several ways but further provide practical implications for decision makers regarding methodological aspects on the one hand and meaningful contextual predictor variables on the other hand.

First, the results on both data sets show that our proposed DHR model with predictor variables yields better forecast accuracy than traditional time series models and ML approaches. Precisely, it outperforms established time series models used in previous research (Andrews & Cunningham, 1995; Bianchi et al., 1998; De Livera et al., 2011; Hyndman et al., 2002) such as ARIMA, ETS, TBATS, standard DHR, and RW as well as powerful ML approaches such as gradient boosting and RF for every considered lead time constellation. This is achieved by simultaneously capturing the dynamics of the time series and including additional predictor variables. Previous studies on call arrivals forecasting methods only focused on of these capabilities in the same model. Thus, the standard DHR model as applied by extant literature (Taylor, 2008; Tych et al., 2002) only relies on Fourier terms assuming that any time series can be expressed as a combination of cosine (or sine) waves with differing periods and on an ARIMA error term capturing short-term dynamics. At the same time, prior research suggests certain predictor variables to enhance forecast accuracy (Aldor-Noiman et al., 2009; Andrews & Cunningham, 1995). We therefore specifically contribute to call center forecasting literature by bringing methodological strings of research together and, in doing so, substantially increase the accuracy of call arrival forecasts. Additional robustness and generalizability are added to the presented results by replicating them for two different series with distinctions in trend, number of observations, as well as level of average arrival count. Reflecting our findings in a more conceptual and abstract manner, we thus contribute to literature by finding evidence that such hybrid models (combining both time series models as well as models with contextual information) unveil a high predictive potential.

Drawing on a broader perspective regarding data characteristics, our results additionally suggest that the magnitude of trend in the time series should be considered in model selection. For data exhibiting only a slight trend like the call arrival series, ML models are outperforming traditional time series models. However, the latter are more competitive and comparable to ML models if the data has a stronger trend like our e-mail arrival series. Further, from a more general perspective of model selection, we found ex-post forecasts of models with predictor variables, i.e., the ML models and our proposed DHR model with predictor variables, to be predominantly more accurate than ex-ante forecasts of models without predictor variables, i.e., time series models, aligning with prior findings (Rausch & Albrecht, 2020). This suggests that the general type of forecasting approach and its possibility of including contextual factors in the form of predictor variables particularly affects prediction accuracy in a practical call arrival forecast setting. Overall, preliminary call center forecasting literature recognized the predictive potential of ML approaches (Albrecht et al., 2021; Barrow, 2016; Jalal et al., 2016; Rausch & Albrecht, 2020) but is still in its infancy and thus, we substantiate the knowledge on the performance of ML models.

As call center managers strongly rely on the accuracy of call arrival predictions for staffing, the improvements achieved by our proposed model implicate high relevance for practice. To keep operating costs at a minimum by avoiding overstaffing and, at the same time, to maximize perceived service quality by shortening long waiting times caused by understaffing, the correct number of call center agents is crucial. Considering e.g. the customer support call arrivals' predictions without lead time of our proposed DHR model compared to ARIMA (most inaccurate model), call center managers would need approximately 4.12[17] call center agents on average less per day in case the model overestimates the arrival volume. Accordingly, on average customers would need to wait approximately 0.93[18] minutes less if the model underestimates the arrival volume. In comparison with the standard DHR model, the difference amounts to 1.93 agents less per day or respectively, an extension of customer waiting time of 0.43 minutes.

Additionally, results for both data sets fit with the theory that call center managers are recommended to minimize lead time in arrivals' forecasts, aligning with prior research (Ibrahim et al. 2016; Rausch and Albrecht 2020). For every model investigated, the best performance is yielded without lead time and generally forecast accuracy decreases steadily with longer lead times. However, in practice longer lead times are frequently mandatory due to personnel planning restrictions. Thus, to overcome this obstacle, managers might consider a two-stage forecasting process: first, producing an early forecast for the agents' scheduling with a pre-defined number of weeks in advance and then, adjusting this forecast right before the start of the predicted week. Our results indicate that the latter prediction with a shorter lead time is more accurate so that managers get more reliable information to incorporate immediate changes into the schedule.

Second, we increase existing knowledge on useful predictor variables. We determined the predictor variables' practical value by conducting a forward variable selection procedure. The results indicate that modeling the day of the week and holidays as predictor variables yields the highest improvement of forecast accuracy and thus, confirms prior research suggesting these influential factors (Aldor-Noiman et al., 2009; Andrews & Cunningham, 1995; Brown et al., 2005; Ibrahim et al., 2016; Ibrahim & L'Ecuyer, 2013). Moreover, the results illustrate that capturing the impact of catalog mailings during the first weekend (i.e., CW0) (and the first week after release (i.e., CW1) respectively) enhances prediction accuracy across all considered lead times for the customer support e-mail (and call arrival series respectively). This particularly indicates that variables including information on marketing actions such as mailings affect customer behavior in terms of e-mail and call volume directly after release. Further, it is shown that reminders via mail on the day of their delivery and the day after (variables MPost1 and MPost2) increased forecast accuracy for all lead times when included as predictor variables. In this context, the results suggest that postal reminders have a substantial effect on the call arrival volume. Vice versa, reminders via e-mail affect the e-mail arrival volume on the day of delivery (MMail1). These findings extend existing literature since the effect of periods with catalog mailings or billing cycles has not been investigated over time (Aldor-Noiman et al., 2009; Andrews & Cunningham, 1995). Thus, by partitioning billing and marketing mailing periods into sequential shorter periods, we are able to capture

[17] If the processing time is 10 minutes per call arrival and the working hours per call center agent are 8 hours per day.

[18] If the processing time is 10 minutes per call arrival and there are 70.95 call arrivals per interval on average.

their temporal effect regarding the first weekend and the first, second, and third week after a catalog release as well as the day of a reminder's delivery and the day after.

Consequently, the results first encourage practitioners to include the availability of explanatory data in their considerations when selecting forecasting models in a call center context. Then, when choosing to use ex-post forecasting models, not only date dependent predictor variables as commonly suggested by literature (such as weekday and holidays) but also factors related to enterprises' customer contact activities need to be considered. Thus, on the one hand forecast accuracy for staffing and ergo high service quality can be improved while on the other hand valuable insights on the effect of activities such as mailings on customer behavior can be gained.

The theoretical and practical implications notwithstanding, our research is subject to limitations that stimulate future research. Noticeably, the proposed method is only capable of modeling one seasonality which can limit its use for complex data with multiple seasonality like e.g. sub-daily (i.e., half-hourly or hourly) data. We overcome this constraint by applying a two-step temporal aggregation procedure to yield sub-daily forecasts. Although this aggregation-disaggregation approach is common practice in forecasting literature, it nevertheless poses an additional obstacle compared to direct sub-daily predictions. Also, we did not test for the optimal aggregation level for both series, i.e., whether to conduct temporal aggregation at a single level or at multiple levels. Besides, it is beyond the scope of this study to investigate different forecast horizons. To increase the reliability of results, in addition to considering different data sets and lead times, future research on this aspect of call center arrival forecasting is encouraged. Regarding model comparison, our study is limited to a selected range of commonly used models on the one hand and promising approaches from related fields on the other hand. In the authors' opinion, comparing the proposed DHR approach to more models, e.g., additional ML methods or mixed-effects models, as well as expanding its application to different businesses' call center data beyond online retail is a fruitful path for future research. In this connection, the high dependency of model performance on the availability and quality of explanatory data needs to be considered. Furthermore, future research can validate our findings on such hybrid models and confirm their superiority by combining the strengths of different model types.

7 Conclusion

Call centers constitute an important customer touchpoint for many businesses. To achieve a high level of customer service satisfaction through short waiting times and good customer support, providing the appropriate number of agents is a critical task for call center management. For this purpose, accurate and feasible forecasting methods to predict call center arrival volumes are needed.

Combining the strengths of different model types investigated by previous research, this study proposes a new method for call center arrivals' forecasting that is able to capture the dynamics of time series and, at the same time, include contextual information in the form of predictor variables. We hypothesize that this approach leads to an increase in prediction performance while also yielding advantages for practical use. The implemented forecasting method extends the established DHR model, which utilizes a sum of sinusoidal terms as predictors to handle periodic seasonality and an ARIMA error to capture short-term dynamics, by including predictor variables in the considered information space to generate predictions. To test the predictive potential of our approach, we analyzed two datasets comprising 174 weeks of data on the call and e-mail arrivals of the customer support queue of a leading German online retailer. We compare our method to traditional time series models (i.e., ARIMA, ETS, TBATS, and RW) as well as established ML approaches (i.e., RF and GBR). Further, we apply time series cross-validation and an expanding rolling window over 52 weeks to assess model performance.

Results show that our proposed DHR model with predictor variables outperforms traditional time series models and ML approaches with regard to forecast accuracy for both data sets and in all lead time constellations investigated. Reflecting this on a more abstract level, we find evidence that such hybrid models combining the benefits of both model types unleash a high predictive potential. Moreover, for the chosen ex-post forecasting method, the predictor variables' practical value can be determined by conducting a forward variable selection procedure. Beyond confirming date-related variables (such as weekday and holidays) as important influential factors for arrival volume, it is shown that catalog mailings and billing cycles exhibit a periodically enhancing effect on prediction accuracy when implemented as predictor variables over time. With the present study we contribute to existing literature by developing a new powerful method to be used in call center arrival forecasting as well as adding knowledge on the temporal effect of predictor variables on customer call and e-mail behavior in this context. We showed that data on e-mail reminders are particularly helpful to predict e-mail arrivals and vice versa, postal reminders are helpful to predict call arrivals. The model's ability to capture both time series information and predictor variables is well suited for the dynamic environment of practical call center arrival forecasting as it not only provides robust forecasts but also offers valuable insights into the effect of a company's customer contact activities. In this regard, future research on the use of other hybrid models in call center arrival forecasting is encouraged to broaden the spectrum of highly accurate methods with practical relevance.

References

Aksin, O. Z., Armony, M., & Mehrotra, V. (2007). The modern call center: A multi-disciplinary perspective on operations management research. Production and Operations Management, 16(6), 665–688.

Albrecht, T., Rausch, T. M., & Derra, N. D. (2021). Call me maybe: Methods and practical implementation of artificial intelligence in call center arrivals' forecasting. Journal of Business Research, 123, 267–278. https://doi.org/10.1016/j.jbusres.2020.09.033

Aldor-Noiman, S., Feigin, P. D., & Mandelbaum, A. (2009). Workload forecasting for a call center: Methodology and a case study. The Annals of Applied Statistics, 3(4), 1403–1447. https://doi.org/10.1214/09-AOAS255

Andrews, B. H., & Cunningham, S. M. (1995). L. L. Bean improves call-center forecasting. Interfaces, 25(6), 1–13.

Avramidis, A. N., Deslauriers, A., & L'Ecuyer, P. (2004). Modeling daily arrivals to a telephone call center. Management Science, 50(7), 896–908. https://doi.org/10.1287/mnsc.1040.0236

Barrow, D. K. (2016). Forecasting intraday call arrivals using the seasonal moving average method. Journal of Business Research, 69(12), 6088–6096. https://doi.org/10.1016/j.jbusres.2016.06.016

Barrow, D. K., & Kourentzes, N. (2018). The impact of special days in call arrivals forecasting: A neural network approach to modelling special days. European Journal of Operational Research, 264(3), 967–977. https://doi.org/10.1016/j.ejor.2016.07.015

Bianchi, L., Jarrett, J., & Choudary Hanumara, R. (1998). Improving forecasting for telemarketing centers by ARIMA modeling with intervention. International Journal of Forecasting, 14(4), 497–504. https://doi.org/10.1016/s0169-2070(98)00037-5

Bloomfield, P. (2000). Fourier analysis of time series: An introduction (2nd edition). John Wiley & Sons, Ltd.

Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: Forecasting and control. Holden-Day.

Boylan, J. E [John E.], & Babai, M. Z. (2016). On the performance of overlapping and non-overlapping temporal demand aggregation approaches. International Journal of Production Economics, 181, 136–144. https://doi.org/10.1016/j.ijpe.2016.04.003

Brady, M. K., & Cronin Jr., J. J. (2001). Customer orientation: Effects on customer service perceptions and outcome behaviors. Journal of Service Research, 3(3), 241–251. https://doi.org/10.1177/109467050133005

Breiman, L. (1996). Bagging predictors. Machine Learning, 24(123-140).

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32.

Brockwell, P. J., & Davis, R. A. (2002). Introduction to time series and forecasting (2nd ed.). Springer texts in statistics. Springer.

Brown, L., Cai, T., Zhang, R., Zhao, L., & Zhou, H. (2010). The root–unroot algorithm for density estimation as implemented via wavelet block thresholding. Probability Theory and Related Fields, 146(3), 401–433. https://doi.org/10.1007/s00440-008-0194-2

Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., & Zhao, L. (2005). Statistical analysis of a telephone call center. Journal of the American Statistical Association, 100(469), 36–50. https://doi.org/10.1198/016214504000001808

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. European Journal of Operational Research, 184(3), 1140–1154. https://doi.org/10.1016/j.ejor.2006.12.004

Cezik, M. T., & L'Ecuyer, P. (2008). Staffing multiskill call centers via linear programming and simulation. Management Science, 54(2), 310–323. https://doi.org/10.1287/mnsc.1070.0824

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on Loess. Journal of Official Statistics, 6(1), 3–73.

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association, 106(496), 1513–1527. https://doi.org/10.1198/jasa.2011.tm09771

Dean, A. M. (2007). The impact of the customer orientation of call center employees on customers' affective commitment and loyalty. Journal of Service Research, 10(2), 161–173. https://doi.org/10.1177/1094670507309650

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366a), 427–431. https://doi.org/10.1080/01621459.1979.10482531

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.

Friedman, J. H. (2002). Stochastic gradient boosting. Computational Statistics & Data Analysis, 38(4), 367–378. https://doi.org/10.1016/s0167-9473(01)00065-2

Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. Manufacturing & Service Operations Management, 5(2), 79–141. https://doi.org/10.1287/msom.5.2.79.16071

Gensler, S., Leeflang, P., & Skiera, B. (2012). Impact of online channel use on customer revenues and costs to serve: Considering product portfolios and self-selection. International Journal of Research in Marketing, 29(2), 192–201. https://doi.org/10.1016/j.ijresmar.2011.09.004

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting, 20(1), 5–10. https://www.sciencedirect.com/science/article/pii/s0169207003001134

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice. OTexts.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting, 18(3), 439–454. https://doi.org/10.1016/s0169-2070(01)00110-8

Ibrahim, R., & L'Ecuyer, P. (2013). Forecasting call center arrivals: Fixed-effects, mixed-effects, and bivariate models. Manufacturing & Service Operations Management, 15(1), 72–85.

Ibrahim, R., Ye, H., L'Ecuyer, P., & Shen, H. (2016). Modeling and forecasting call center arrivals: A literature survey and a case study. International Journal of Forecasting, 32(3), 865–874. https://doi.org/10.1016/j.ijforecast.2015.11.012

Jalal, M. E., Hosseini, M., & Karlsson, S. (2016). Forecasting incoming call volumes in call centers with recurrent Neural Networks. Journal of Business Research, 69(11), 4811–4814. https://doi.org/10.1016/j.jbusres.2016.04.035

Jongbloed, G., & Koole, G. (2001). Managing uncertainty in call centres using Poisson mixtures. Applied Stochastic Models in Business and Industry, 17(4), 307–318. https://doi.org/10.1002/asmb.444

Kourentzes, N., & Petropoulos, F. (2016). Forecasting with multivariate temporal aggregation: The case of promotional modelling. International Journal of Production Economics, 181, 145–153. https://doi.org/10.1016/j.ijpe.2015.09.011

Kourentzes, N., Petropoulos, F., & Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. International Journal of Forecasting, 30(2), 291–302. https://doi.org/10.1016/j.ijforecast.2013.09.006

Kourentzes, N., Rostami-Tabar, B., & Barrow, D. K. (2017). Demand forecasting by temporal aggregation: Using optimal or multiple aggregation levels? Journal of Business Research, 78, 1–9. https://doi.org/10.1016/j.jbusres.2017.04.016

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. R News, 2(3), 18–22.

Mabert, V. A. (1985). Short interval forecasting of emergency phone call (911) work loads. Journal of Operations Management, 5(3), 259–271. https://doi.org/10.1016/0272-6963(85)90013-0

Nikolopoulos, K., Syntetos, A. A., Boylan, J. E., Petropoulos, F., & Assimakopoulos, V. (2011). An aggregate–disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis. Journal of the Operational Research Society, 62(3), 544–554. https://doi.org/10.1057/jors.2010.32

Pankratz, A. E. (1991). Forecasting with dynamic regression models. John Wiley & Sons, Ltd.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual Model of service quality and its implications for future research. Journal of Marketing, 49(4), 41–50. https://doi.org/10.1177/002224298504900403

Rausch, T. M., & Albrecht, T. (2020). The impact of lead time and model selection on the accuracy of call center arrivals' forecasts. Proceedings of the 28th European Conference on Information Systems (ECIS), 1-16.

Schapire, R. E., Freund, Y., Bartlett, P., & Lee, W. S. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. The Annals of Statistics, 26(5), 1651–1686.

Shen, H., & Huang, J. Z. (2008). Forecasting time series of inhomogeneous Poisson processes with application to call center workforce management. The Annals of Applied Statistics, 2(2), 601–623. https://doi.org/10.1214/08-AOAS164

Sitel Group. (2018). 2018 CX Index: Brand Loyalty & Engagement, pp. 1–27.

Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. Journal of the Operational Research Society, 54(8), 799–805. https://doi.org/10.1057/palgrave.jors.2601589

Taylor, J. W. (2008). A comparison of univariate time series methods for forecasting intraday arrivals at a call center. Management Science, 54(2), 253–265.

Taylor, J. W. (2012). Density forecasting of intraday call center arrivals using models based on exponential smoothing. Management Science, 58(3), 534–549.

Theodosiou, M. (2011). Forecasting monthly and quarterly time series using STL decomposition. International Journal of Forecasting, 27(4), 1178–1195. https://doi.org/10.1016/j.ijforecast.2010.11.002

Thompson, H. E., & Tiao, G. C. (1971). Analysis of telephone data: A case study of forecasting seasonal time series. The Bell Journal of Economics and Management Science, 2(2), 515. https://doi.org/10.2307/3003004

Tych, W., Pedregal, D. J., Young, P. C., & Davies, J. (2002). An unobserved component model for multi-rate forecasting of telephone call demand: the design of a forecasting support system. International Journal of Forecasting, 18(4), 673–695. https://doi.org/10.1016/S0169-2070(02)00071-7

Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing. Journal of Retailing, 91(2), 174–181. https://doi.org/10.1016/j.jretai.2015.02.005

Weinberg, J., Brown, L. D., & Stroud, J. R. (2007). Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. Journal of the American Statistical Association, 102(480), 1185–1198.

Whitt, W. (1999). Dynamic staffing in a telephone call center aiming to immediately answer all calls. Operations Research Letters, 24(5), 205–212. https://doi.org/10.1016/s0167-6377(99)00022-x

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6(3), 324–342. https://doi.org/10.1287/mnsc.6.3.324

Young, P. C. (1999). Recursive and en-bloc approaches to signal extraction. Journal of Applied Statistics, 26(1), 103–128. https://doi.org/10.1080/02664769922692

Young, P. C., Pedregal, D. C., & Tych, W. (1999). Dynamic harmonic regression. Journal of Forecasting, 18(6), 369–394. https://doi.org/10.1002/(SICI)1099-131X(199911)18:6<369::AID-FOR748>3.0.CO;2-K

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The Behavioral Consequences of Service Quality. Journal of Marketing, 60(2), 31–46. https://doi.org/10.1177/002224299606000203

3.2.3 Research Paper No. 6: Predicting online shopping cart abandonment with machine learning approaches

Authors: Rausch, T. M., Derra, N. D., & Wolf, L.

Published in: International Journal of Market Research (Forthcoming)

DOI: https://doi.org/10.1177/1470785320972526

Abstract: Excessive online shopping cart abandonment rates constitute a major challenge for e-commerce companies and can inhibit their success within their competitive environment. Simultaneously, the emergence of the Internet's commercial usage results in steadily growing volumes of data about consumers' online behavior. Thus, data-driven methods are needed to extract valuable knowledge from such big data to automatically identify online shopping cart abandoners. Hence, this contribution analyzes clickstream data of a leading German online retailer comprising 821,048 observations to predict such abandoners by proposing different machine learning approaches. Thereby, we provide methodological insights to gather a comprehensive understanding of the practicability of classification methods in the context of online shopping cart abandonment prediction: our findings indicate that gradient boosting with regularization outperforms the remaining models yielding an F_1-Score of 0.8569 and an AUC value of 0.8182. Nevertheless, as gradient boosting tends to be computationally infeasible, a decision tree or boosted logistic regression may be suitable alternatives, balancing the trade-off between model complexity and prediction accuracy.

Keywords: e-commerce; shopping cart abandonment; prediction; classification; machine learning; supervised learning

1 Introduction

To strengthen a company's position within its competitive environment, marketers need to be able to precisely predict potential customers regarding their purchase and, further, non-purchase behavior. Considering this in the context of online shopping environment, customers frequently place items in their virtual shopping cart for reasons other than immediate purchase. This phenomenon is known as shopping cart abandonment and is particularly apparent in the context of e-commerce: it is the behavioral outcome of consumers placing item(s) in their online shopping cart without making a purchase by completing the checkout process during that online session (Huang et al., 2018; Kukar-Kinney & Close, 2010). Extant literature investigated the behavioral perspective of online shopping cart abandonment by identifying inhibitors to the purchase process: financial risks and concerns about delivery and return policies (Kukar-Kinney & Close, 2010) the usage of shopping carts as organization tools or for entertainment purposes (Kukar-Kinney & Close, 2010), and inhibitors at the checkout stage like perceived transaction inconvenience and privacy intrusion (Rajamma et al., 2009) are – inter alia – the main factors leading to online shopping cart abandonment.

With the spread of the Internet's commercial usage, the ability to track consumers' online activities allows companies to collect unbiased information about consumers' behavior. The detailed records of past usage behaviors comprised by log files and resulting clickstream data can be analyzed by marketers to gain valuable insights. In this context, clickstream data have frequently been modeled to derive implications for website design or advertising efforts (see, for example, (Chatterjee et al., 2003) and (Montgomery et al., 2004)) and further, to predict consumers' future behaviors, e.g. regarding purchase (see, for example, (Bucklin & Sismeiro, 2003) and (Moe & Fader, 2004a)).

Thus, the antecedents of online shopping cart abandonment are well understood by behavioral literature and clickstream data has been studied by methodological research to analyze consumers' behavior. The rise of the Internet and the era of big data resulted in an excessive 'datafication' (Kelly & Noonan, 2017; Lycett, 2013) of the organizational environment yielding the field of business intelligence comprising data analytics and predictive analytics approaches (Chen et al., 2012). However, despite the richness of clickstream data, prior shopping cart abandonment literature still lacks data-driven methods based on machine learning which make use of this information source to predict such abandoning customers. This might be due to the insufficient awareness of suitable intelligent approaches to extract knowledge from the steadily growing volumes of data (Fayyad et al., 1996).

To address this research gap, we utilize clickstream data of a leading German online retailer to train and subsequently compare different machine learning approaches for the prediction of online shopping cart abandonment (i.e., tree-based methods (more specifically, adaptive boosting, boosted logistic regression, decision tree, gradient boosting with regularization, gradient boosting, gradient boosting with dropout, random forest, and stochastic gradient boosting), k-nearest neighbor, naïve bayes, multi-layer perceptron with dropout, and a support vector machine with radial basis kernel). We successfully implement these machine learning methods for online shopping cart abandonment prediction and compare them with logistic regression as a standard non-machine learning benchmark model regarding their predictive performance.

Our paper makes several key contributions to the preceding literature. By combining the research fields of both shopping cart abandonment as well as clickstream data analysis with machine learning approaches, we particularly shed light on the practicability of machine learning methods in this application context, as this was neglected by prior research. Further, we provide insights into the characteristics of customers abandoning their shopping cart based on clickstream data that is unsusceptible to self-selection, relatively unobtrusive, and easy to gather. We extensively review literature on classification methods to identify shopping cart abandonments and present validation procedures as well as performance metrics for such methods. Our findings can be useful both for marketing intelligence research by extending the field of machine learning applications in marketing contexts through automatically predicting online shopping cart abandoners and for practitioners to actively prevent such abandonments by several real time reactions, e. g. providing real-time purchase incentives, and moreover, to gain insights into machine learning methods.

The remainder of this paper is organized as follows: the subsequent section describes the related work on online shopping cart abandonment and clickstream data. Further, Section 3 summarizes the background on machine learning approaches for classification. Section 4 outlines the methodology comprising a preliminary data analysis and the research design. In Section 5 and 6, we present the findings and discuss both theoretical and practical implications, limitations, as well as directions for future research. Finally, Section 7 draws a conclusion.

2 Related Work

2.1 Online Shopping Cart Abandonment

The online shopping cart abandonment phenomenon causes substantial losses of turnover for online retailers (Huang et al., 2018; Rajamma et al., 2009) resulting in a weakened position within their competitive environment. Therefore, extant marketing literature addressed this problem by drawing on a behavioral perspective to identify and understand essential determinants of online shopping cart abandonment: (Rajamma et al., 2009) focused on potential inhibitors at the checkout stage and found increased perceived transaction inconvenience (e.g., long registration forms) and high perceived risk (e.g., perceived security of information asked) to enhance online shopping cart abandonment. Partially, these findings seem to be applicable to new customers which are unfamiliar with the checkout process. Similarly, (Kukar-Kinney & Close, 2010) findings indicate that privacy intrusion and security concerns rather lead to the consumers' decision to buy the product from a stationary offline store. Further, they found the entertainment value of shopping carts, the use of shopping carts as an organization tool, the wait for sale, and the concerns about costs to be antecedents of shopping cart abandonment (Kukar-Kinney & Close, 2010). Their identified determinants were supported by (Kukar-Kinney & Close, 2010) proving that customers' tendencies to add items to the online shopping cart for reasons other than immediate purchase are – inter alia – due to organizational purposes. (Huang et al., 2018) focused on mobile shopping cart abandonment in their study. They found intrapersonal (i.e., conflicts regarding mobile shopping attributes and low self-efficacy regarding mobile shopping) and interpersonal (i.e., discrepancies from the other's attitudes to self-attitudes) conflicts to disturb consumers' emotions during mobile shopping, and in turn, implying shopping cart abandonment. Overall, their findings indicate that the utilized device for online shopping might impact purchase behavior as well. (Cho et al., 2006) proved that consumers' confusion by information overload, high value-consciousness, negative past experiences, intention to conduct price comparisons, and unreliable websites are likely to trigger online shopping cart abandonment[19].

2.2 Clickstream Data

Drawing on a more holistic perspective of online shopping behavior, further literature shifted away from explanatory behavioral approaches to data-driven methods predicting online purchase behavior in general. Typically, such predictions are based on clickstream data (see, e.g., (Moe & Fader, 2004a), (Sismeiro & Bucklin, 2004), or (van den Poel & Buckinx, 2005)). Clickstream data model the navigation path a customer takes through the online shop (Montgomery, 2001; Montgomery et al., 2004) and can be extracted from log files which register all requests and information transferred between the customer's computer and the company's commercial web server (Bucklin & Sismeiro, 2003).

Examples for using clickstream data to predict online shopping behavior are – inter alia - (Moe & Fader, 2004a) who proposed a conversion model predicting each customer's probability of making a purchase based on purchase and visit history. The same authors (Moe & Fader, 2004b) also developed a model for evolving visiting behavior and further, they examined the relationship between visiting frequency and purchasing propensity. They found consumers visiting an e-commerce site more frequently to have a greater propensity to buy (Moe & Fader, 2004b). (van den Poel & Buckinx, 2005) predicted purchase behavior and investigated the contribution of different variables: they proved (1) general clickstream variables (i.e., number of days since last visit, and speed of clickstream behavior during last visit), (2) more detailed clickstream variables (i.e., number of accessories (and personal pages and products respectively) viewed during last visit), (3) demographic variables (i.e., gender and the fact of supplying personal information), and (4) historical purchase behavior (i.e., number of days since last purchase and number of past purchases) to be meaningful predictors. (Montgomery et al., 2004) set up different models to predict purchase conversion probability by modeling path information.

Moreover, clickstream data was frequently utilized by research to predict not only purchase behavior but further similar outcome variables. For instance, (Bucklin & Sismeiro, 2003) investigated drivers

[19] Cho et al. (2006) defined online shopping cart abandonment rather as a hesitation reaction which implies that the customer actively drops items placed in his/her shopping cart. Thus, their definition differs slightly from the definition of Kukar-Kinney and Close (2010), which was used in this study for an understanding of shopping cart abandonment.

affecting the length of time spent viewing a website and the visitor's decision to continue browsing or to exit the website. (Sismeiro & Bucklin, 2004) decomposed the purchase process into sequences that must be completed for a purchase to take place (i.e., completion of product configuration, input of personal information, and order confirmation with provision of credit card data) and predicted the probability of completion for each task with covariates of browsing behavior, repeat visitation, use of decision aids, input effort, and information gathering.

3 Machine Learning Approaches for Classification

Overall, e-commerce as a research subject is suitable for the application of machine learning approaches as proposed by Kohavi & Provost (2001): online retailers can easily and inexpensively collect rich data with respect to the online behavior of customers (i.e., clickstream data) and, further, implement data mining and machine learning applications since political and social barriers are substantially lower than for traditional businesses. Consequently, typical problems for successfully applying machine learning (i.e., the need for a large volume of controlled and reliable data, data with sufficient descriptions, the ability to evaluate results, and to integrate applications successfully) are reduced by the characteristics of e-commerce environment (Kohavi & Provost, 2001).

Machine learning constitutes a new paradigm within data science research and emerged in the course of the artificial intelligence era, which, in turn, was first coined by Samuel (1959) describing it as "the programming of a digital computer to behave it in a way which, if done by human beings […], would be described as involving the process of learning". In this context, learning may be understood as the automatic search for more useful representations of data regarding a specific task (Chollet & Allaire, 2018). Machine learning algorithms and systems are consequently trained rather than explicitly programmed. During this process, these systems find statistical structure in given examples which are relevant to the task and derive rules for automating the task using guidance from a feedback signal (Cui et al., 2006). Thereby, classification algorithms are types of supervised learning approaches within machine learning which predict a qualitative response for an observation, i.e., they assign an observation to a category (James et al., 2013): Formally, let $\{y_k, x_k\}_{k=1}^{N}$ bet a training set, where $y_k \in \{0,1,2,\dots,K-1\}$ is the class membership and $x_k = \mathbb{R}^n$ is the vector of predictor values, then the task is to learn a function to predict the class label y_k from x_k. Thereby, $K = 2$ in case of binary classification and $K > 2$ in case of multi-class classification tasks.

Drawing on the online shopping cart abandonment problem, the prediction of purchasers and non-purchasers (i.e., customers abandoning their shopping cart) can be considered a binary classification task. Common machine learning approaches for binary classification include – inter alia – tree-based methods, support vector machines, naïve bayes, k-nearest neighbor, and neural networks. The approaches are explained in detail hereinafter.

3.1 Tree-Based Approaches

One of the most common machine learning approaches are tree-based methods which descend from single decision trees, as proposed by (L. Breiman et al., 1984). Basically, decision trees are flowchart-like structures that generate "if-else" rules and thereby allow for prediction of observation classes. Thereby, classification and regression tree models follow a recursive top-down approach in which binary trees aim to partition the predictor space with predictor variables $x_1, \dots, x_k$ into subsets in which the distribution of the dependent variable y is successively more homogeneous (Chipman et al., 1998).

Generally, single decision trees have the advantage of being easy to interpret and to understand (Moro et al., 2014). However, they frequently lead to overfitting, i.e., the model learns to identify specific characteristics of the training data which are irrelevant or even obstructive for the classification of unknown data (Friedman, 2001; Srivastava et al., 2014). This results in drawbacks of predictive performance and less expressiveness of the models. Ensemble learning methods that construct several individually trained decision trees and combine their results into a classifier outperforming the single predictions (Opitz & Maclin, 1999; Rokach, 2010) may offer a solution to this problem. In this context, two widely used methods of aggregating trees are boosting and bagging.

In boosting, a family of algorithms converts weak learners (i.e., models that achieve accuracy just above random guessing) to strong learners with a powerful predictive capacity. The idea is to train weak learners sequentially with each weak learner trying to correct its predecessor (Schapire et al., 1998). Thus, each decision tree is built using feedback from previously grown trees (James et al., 2013). Popular boosting algorithms include adaptive boosting "AdaBoost" (Freund & Schapire, 1997), boosted logistic regression "LogitBoost" (Friedman et al., 2000), gradient boosting machines "GB" (Friedman, 2001, 2002), and stochastic gradient boosting "SGB" (Friedman, 2002)[20]. For instance, AdaBoost as a basic boosting algorithm makes predicitions by combining the output of weak learners to a weighted sum and putting higher weights on incorrectly classified instances:

$$\hat{y} = \operatorname{sign}\left(\sum_{m=1}^{M} \alpha_m h_m(x)\right)$$

with the weak hypothesis h_m detected by the weak learner and its importance α_m.

In contrast to boosting, bagging (or bootstrap aggregating) grows successive trees independently from earlier trees, i.e., each tree is constructed using a bootstrap sample of the data and, hence, a majority vote is taken for prediction (Breiman, 1996). Random forests add an additional layer of randomness to bagging and change how the trees are constructed: in standard decision trees each node is split using the best split among all predictor variables whereas in random forests the nodes are split using the best

[20] The concepts of AdaBoost, LogitBoost, and gradient boosting are closely related as all approaches produce an ensemble of weak learners but – in contrast to AdaBoost and LogitBoost – gradient boosting models minimize the model's loss by adding weak learners sequentially using a procedure similar to gradient descent, i.e., it allows arbitrary differentiable loss functions to be used.

among a subset of predictors randomly chosen at that node (Leo Breiman, 2001; Liaw & Wiener, 2002). Due to the recursive structure of tree-based methods they often capture interaction effects between variables. However, since we focus on the performane of the models and not the importance of specific variables, we will not consider interaction effects further in our study.

Overall, tree-based methods have been found to outperform other established approaches across a variety of different classification tasks such as IP traffic flow classification (Williams et al., 2006), customer churn prediction (Vafeiadis et al., 2015), or – similar to our context – prediction of online purchase intention (Bogina et al., 2019; Boroujerdi et al., 2014; Zheng & Liu, 2018). They are particularly favorable since ensemble methods are able to reduce both bias and variance of the single learning algorithms: While individual models may get stuck in local minima, a weighted combination of several different local minima – produced by ensemble methods – are able to minimize the risk of choosing the wrong local minimum (Dietterich, 2002).

3.2 Support Vector Machines

Aside from tree-based methods, support vector machines are powerful tools for classification tasks (James et al., 2013). The basic support vector machine is solving pattern recognition problems by mapping data into a multidimensional input space and constructing an optimal hyperplane that separates the space into homogenous partitions[21] (Cortes & Vapnik, 1995; Vapnik, 1982). Predictions of new instances are then classified into those partitions. The support vector machine aims at constructing a classifier in the form of

$$\hat{y} = \operatorname{sign}\left[\sum_{i=1}^{N} \alpha_k y_k \psi(x, x_k) + b\right]$$

where α_k are positive real constants, b is a real constant, and $\psi(\cdot,\cdot)$ represents the hyperplane (e.g., $\psi(x, x_k) = x_k^T x$ in case of a linear support vector machine) (Suykens & Vandewalle, 1999). Aside from the linear case, Boser et al. (1992) proposed a non-linear classifier by applying the so-called kernel trick which allows the algorithm to fit the hyperplane in a transformed feature space.

We used a support vector machine with radial basis kernel for the comparison of machine learning models. However, support vector machines may become computationally infeasible on very large datasets like clickstream data (L'Heureux et al., 2017).

[21] A hyperplane is defined as a flat affine subspace of dimension $p - 1$ with p being the number of dimensions (i.e., the number of considered predictor variables) James et al. (2013). Basically, the 'hyperplane' is a line if the feature space is two-dimensional (i.e., two predictor variables) and a simple plane if the space is threedimensional (i.e., three predictor variables).

3.3 Naïve Bayes

The naïve bayes approach is a basic classifier based on applying the Bayes' theorem with the naïve assumption that the attributes are conditionally independent (Duda et al., 1973). The classifier assigns a new case to a class label $\hat{y} = C_k$ by deriving the maximum a posteriori probability:

$$\hat{y} = \underset{k\epsilon\{1,\ldots,K\}}{\arg\max}\, p(C_k)\prod_{i=1}^{n} p(x_i|\, C_k)$$

Naïve bayes as a generative classifier is frequently utilized for classification tasks due to its simplicity, efficiency, and efficacy (Muhammad & Yan, 2015).

3.4 K-Nearest Neighbor

Another basic approach, the k-nearest neighbor algorithm, classifies an observation by a majority vote of the observation's neighbors (Cover & Hart, 1967). The underlying assumption of the algorithm is that observations which lay closely together within the predictor space (i.e., neighbors) will have the same class label. Thus, the classifier weights the class of the nearest neighbors strikingly high in order to predict the class label of an unclassified sample (Cover & Hart, 1967). The class is thereby assigned by taking the majority vote of the k nearest neighbors, with k being the number of neighbors that are considered during the classification task. The nearest neighbors are determined with the help of arbitrary distance functions (e.g., Euclidian distance $d(.,.)$). For new observations (y, x) the nearest neighbor $(y_{(1)}, x_{(1)})$ within the training set is defined by

$$d\left(x, x_{(1)}\right) = \min_k(d(x, x_k))$$

and $\hat{y} = y_{(1)}$ – the class of the nearest neighbor – is selected as prediction for y. $x_{(j)}$ and $y_{(j)}$ describe the jth nearest neighbor of x and its class membership y.

K-nearest neighbor as a local learning approach may be suitable for online shopping cart abandonment prediction tasks since it is able to alleviate the challenge of imbalanced data (L'Heureux et al., 2017).

3.5 Artificial Neural Networks

Artificial neural networks are highly parallelized computer systems comprising process units (i.e., neurons) located on process layers with numerous weighted interconnections performing a learning process to create meaningful data representations (Jain et al., 1996). Regarding the concept of deep learning, artificial neural networks may use a number of hidden process layers (the depth of a network) between input and output layer containing non-linear operations in hierarchical architectures to learn characteristics and recognize patterns from given data (Bengio, 2009; Deng, 2011; Hinton et al., 2006). The concept of learning within deep learning (or artificial neural networks, respectively) describes a process of updating the network architecture and the weights of the neuron connections (Jain et al., 1996). To improve the performance, the optimizer is implementing a backpropagation algorithm to minimize the discrepancy between the actual and the target output vector (i.e., the loss score) by adjusting the weights (Rumelhart et al., 1986; Schmidhuber, 2015). To avoid overfitting, a regularization method called dropout can be integrated in the network which randomly sets a share of its output per layer to zero (Srivastava et al., 2014).

Concerning their connection structure (i.e., topology), neural network architectures can be distinguished between feedforward networks (e.g., multi-layer perceptrons (Deng, 2011; Q. Zhang et al., 2018) with neuron connections running to the output layer acyclically and recurrent networks (e.g., long short-term memories (Hochreiter & Schmidhuber, 1997) containing backward connections to build cyclic architectures (Jain et al., 1996; Schmidhuber, 2015). The most commonly used feedforward neural networks – multi-layer perceptrons – can be defined as

$$\hat{y} = \beta_0 + \sum_{h=1}^{H} \beta_h g\left(\gamma_{0i} + \sum_{i=1}^{I} \gamma_{hi} p_i\right)$$

where I denotes the number of inputs p_i, H is the number of hidden nodes in the network, the weights $\omega = (\beta, \gamma)$ with $\beta = [\beta_1, \ldots, \beta_H]$ and $\gamma = [\gamma_{11}, \ldots, \gamma_{HI}]$ are for the hidden and output layer respectively, $g(\cdot)$ is the transfer function (e.g., sigmoid logistic), and β_0 as well as γ_{0i} are the biases of each node (Zhang et al., 1998).

Multi-layer perceptrons were found to outperform other machine learning approaches for purchase intention prediction only after balancing the class distribution with oversampling (Sakar et al., 2019) since deep learning approaches are frequently sensitive to class imbalance (L'Heureux et al., 2017).

4 Methodology

4.1 Preprocessing and Preliminary Data Analysis

The purpose of this study is to predict shopping cart abandonment by making use of machine learning. The machine learning models explained in Section 3 are compared to find the best classifier for this task. The clickstream data were gathered from server log files of a leading German online retailer which primarily distributes fashion. The data were created by the online retailer through extracting the customers' chronological online shop activities out of sequential log files. Each log file observation comprised one action or activity (e.g., a click) of a certain customer such as adding a product to the cart or clicking on a product to view its details. Subsequently, each customer's activities during a session were assigned to summarizing variables. Hence, all activities of a customer were aggregated to one observation with different variables describing the session. Thereby, a session is a period of sustained web browsing or a sequence of the user's page viewings until the user exits the online shop (Montgomery et al., 2004). The data comprise 3,511,037 observations or sessions between February 1, 2019 and April 30, 2019, i.e., three months. Further, the data contain 18 explanatory variables for each observation or session listed in Table 1 many of which are consistent with van den Poel & Buckinx (2005) findings. We are only interested in visitors who made use of the virtual shopping cart during the session, i.e., who placed item(s) in their cart. In line with (Close & Kukar-Kinney, 2010), shopping cart usage is thus defined as necessary precondition for shopping cart abandonment. Thus, we filtered out customers which did not add any items to their shopping cart during the session, so-called just-browsing customers, and 821,048 observations (23,38%) remained. We modeled the dependent variable – shopping cart abandonment – as a dummy variable using the information about the customer's compiled and ordered shopping carts (variables BASKETS_BB and BASKETS) during the session:

$$Y = \begin{cases} 1 \text{ if number of compiled shopping carts} > 0 \text{ \& number of ordered shopping carts} = 0; \\ 0 \text{ if number of compiled shopping carts} > 0 \text{ \& number of ordered shopping carts} > 0. \end{cases}$$

Our data contain 520,653 (63.41%) observations of shopping cart abandonments (or non-purchasers respectively) and 300,395 (36.59%) observations of purchasers. Hence, the dataset is relatively balanced. We excluded the variable for the number of ordered shopping carts (BASKETS_BB) and the value of ordered shopping carts (VALUE_BB) further for prediction[22].

Table 1: Variables of Clickstream Data.

Variable	Index	Description
Shopping Cart Abandonment	SCA	Dependent dummy variable capturing customer's shopping cart abandonment $Y = \begin{cases} 1 \text{ if customer abandoned;} \\ 0 \text{ otherwise.} \end{cases}$
Number of Ordered Shopping Carts	BASKETS_BB	Metric predictor variable capturing the number of shopping carts ordered during the customer's session
Number of Compiled Shopping Carts	BASKETS	Metric predictor variable capturing the number of shopping carts compiled during the customer's session
Number of Logins	LOGS	Metric predictor variable capturing the number of logins during the customer's session
Number of Existing Customers' Logins to the Second Step of the Ordering Process	LOGS_CUST_STEP2	Metric predictor variable capturing the number of logins of existing customers to the second step of the purchasing process during the customer's session
Number of New Customers' Logins to the Second Step of the Ordering Process	LOGS_NEWCUST_STEP2	Metric predictor variable capturing the number of logins of new customers to the second step of the purchasing process during the customer's session
Number of Overall Page Viewings	PIS	Metric predictor variable capturing the number of overall page viewings during the customer's session
Number of Shopping Cart Page Viewings	PIS_AP	Metric predictor variable capturing the number of shopping carts page viewings during the customer's session
Number of Detailed Product Page Viewings	PIS_DV	Metric predictor variable capturing the number of detailed product page viewings during the customer's session
Number of Category Overview Page Viewings	PIS_PL	Metric predictor variable capturing the number of category overview page viewings (i.e., all products within a category) during the customer's session
Number of Department Page Viewings	PIS_SHOPS	Metric predictor variable capturing the number of department page viewings (i.e., all categories within a department) during the customer's session

[22] These variables are values referring to the customers' order and, thus, they would not be known ex-ante for prediction.

Variable	Index	Description
Number of Detailed Product Page Viewings Using Search Function	PIS_SDV	Metric predictor variable capturing the number of detailed product page viewings after using the search function during the customer's session
Number of Search Results Page Viewings	PIS_SR	Metric predictor variable capturing the number of overall search results page viewings during the customer's session
Number of Product Types in Shopping Cart	POSITIONS	Metric predictor variable capturing different product types in the shopping cart during the customer's session
Number of Items in Shopping Cart	QUANTITY	Metric predictor variable capturing the number of items in the shopping cart during the customer's session
Value of Ordered Shopping Carts	VALUE_BB	Metric predictor variable capturing the value of shopping carts ordered during the customer's session
New Customer	NEW_CUST	Predictor dummy variable capturing new customers $X_{16} = \begin{cases} 1 \text{ if new customer;} \\ 0 \text{ otherwise.} \end{cases}$
Accessing Online Shop via Desktop	WEB_CUST	Predictor dummy variable capturing customers that access the online shop via desktop $X_{17} = \begin{cases} 1 \text{ if accessing via desktop;} \\ 0 \text{ otherwise.} \end{cases}$
Accessing Online Shop via Mobile Phone	MOBILE_CUST	Predictor dummy variable capturing customers that access the online shop via mobile phone $X_{18} = \begin{cases} 1 \text{ if accessing via mobile phone;} \\ 0 \text{ otherwise.} \end{cases}$

Figure 1 illustrates the relationship between the page viewing and login variables by demonstrating the customer's clickstream in the online shop: the customer typically starts browsing departments (PIS_SHOPS), then selects a certain category within a department (PIS_PL), and further, chooses a certain product within a category (PIS_DV). Optionally, the customer uses the shop's search engine (PIS_SR) to look systematically for a specific product (PIS_SDV). To make a purchase, the customer can either directly sign in (LOGS) or check the items in the shopping cart (PIS_AP) first and then sign in and hence, proceed to the second step of the purchasing process (LOGS_CUST_STEP2 or LOGS_NEWCUST_STEP2). However, signing in to the second step of the purchasing process does not necessarily lead to a purchase of the customer.

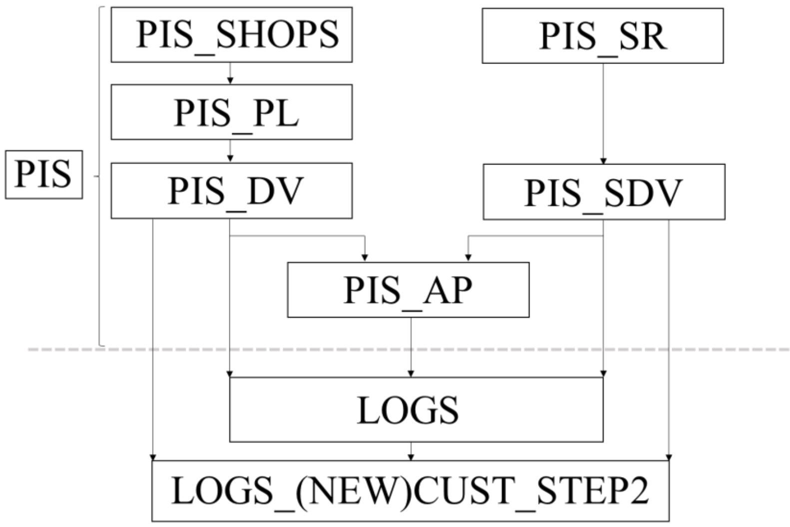

Figure 1: Main Clickstream of Customers in the Online Shop.

Note: LOGS = Number of Logins, LOGS_CUST_STEP2 = Number of Existing Customers' Logins to the Second Step of the Ordering Process, LOGS_NEWCUST_STEP2 = Number of New Customers' Logins to the Second Step of the Ordering Process, PIS = Number of Overall Page Viewings, PIS_AP = Number of Shopping Cart Page Viewings, PIS_DV = Number of Detailed Product Page Viewings, PIS_PL = Number of Category Overview Page Viewings, PIS_SDV = Number of Detailed Product Page Viewings Using Search Function, PIS_SHOPS = Number of Department Page Viewings, PIS_SR = Number of Search Results Page Viewings.

Nevertheless, with respect to the descriptive statistics in Table 2, we find that existing customers (or new customers respectively) which subsequently make a purchase sign in to the second step of the ordering process approximately 5.93 times (or 4.46 times respectively) more often than non-purchasers. Generally, purchasers sign in more often (1.03 logins on average) than non-purchasers (0.93 logins on average). This might indicate that the cause for shopping cart abandonment frequently occurs before the customer proceeds to the checkout stage.

Furthermore, the number of purchasers' overall page viewings is 2.09 times higher than of non-purchasers on average. Overall, customers abandoning their shopping cart browse less pages than purchasers – regardless of the pages' type. Particularly, the median reveals that there are significant differences regarding the number of page viewings between purchasers and abandoners: the median of abandoners' overall page viewings is 12, 1 for department viewings, and 0 for all other types of page viewings. In contrast, purchasers' median for overall page viewings is 35, 6 for department viewings, and e.g. 2 for shopping cart viewings.

On average, purchasers add more items and different product types (3.48 and 3.38 respectively) to their shopping cart than non-purchasers (2.95 and 2.88 respectively).

There is a larger absolute (48,839) and relative (9.38%) proportion of new customers among the observations of shopping cart abandonments than among those making a purchase (15,387 observations or 5.12% respectively). Moreover, there is a larger proportion of mobile shoppers among customers abandoning their shopping cart (45.85%) compared to the observations of purchasers (28.1%). The latter descriptive findings are consistent with the results of preceding (behavioral) research: e.g., as argued earlier, Huang et al. (2018) proved that online shopping cart abandonment occurs more frequently for customers using a mobile device due to high emotional ambivalence. Moe & Fader (2004a) found that – among new customers – online conversion rate is lower as purchasing thresholds and perceived risks are high for unexperienced visitors.

Table 2: Descriptive Statistics of Clickstream Data.

Variable	**Observations of Shopping Cart Abandonments (n=520,653)**					**Observations of Purchasers (n=300,395)**				
	Mean	**SD**	**Median**	**Min**	**Max**	**Mean**	**SD**	**Median**	**Min**	**Max**
BASKETS	0.99	0.11	1	0	2	1.09	0.40	1	0	49
LOGS	0.93	0.27	1	0	2	1.03	0.21	1	0	2
LOGS_CUST_ STEP2	0.06	0.23	0	0	1	0.32	0.47	0	0	1
LOGS_NEWCU ST_ STEP2	0.02	0.13	0	0	1	0.07	0.26	0	0	1
PIS	22.26	27.47	12	1	513	46.45	37.06	35	2	593
PIS_AP	1.05	2.17	0	0	71	3.06	3.42	2	0	57
PIS_DV	3.42	7.48	0	0	200	6.53	9.76	3	0	203
PIS_PL	3.99	11.37	0	0	279	8.75	17.00	1	0	315
PIS_SHOPS	7.68	17.55	1	0	405	15.87	25.14	6	0	396
PIS_SDV	1.40	3.92	0	0	142	3.13	5.46	1	0	127
PIS_SR	2.82	7.48	0	0	222	5.71	10.18	2	0	208
POSITIONS	2.88	3.31	2	1	66	3.38	3.31	2	1	111
QUANTITY	2.95	3.55	2	1	143	3.48	3.49	2	1	143
	Counts		**Proportion**			**Counts**		**Proportion**		
NEW_CUST	48,839		9.38%			15,387		5.12%		
WEB_CUST	214,455		41.29%			171,789		57.19%		
MOBILE_CUST	238,694		45.85%			84,401		28.1%		

Note: BASKETS = Number of Carts Compiled, LOGS = Number of Logins, LOGS_CUST_STEP2 = Number of Existing Customers' Logins to the Second Step of the Ordering Process, LOGS_NEWCUST_STEP2 = Number of New Customers' Logins to the Second Step of the Ordering Process, MOBILE_CUST = Customer Accessing via Mobile Phone**,** NEW_CUST = New Customer, PIS = Number of Overall Page Viewings, PIS_AP = Number of Shopping Cart Page Viewings, PIS_DV = Number of Detailed Product Page Viewings, PIS_PL = Number of Category Overview Page Viewings, PIS_SDV = Number of Detailed Product Page Viewings Using Search Function, PIS_SHOPS = Number of Department Page Viewings, PIS_SR = Number of Search Results Page Viewings, POSITIONS = Number of Product Types, QUANTITY = Number of Items, WEB_CUST = Customer Accessing via Desktop.

4.2 Experimental Setup

Since each machine learning approach and its subsequent refinements and modifications exhibit individual strengths and weaknesses in dependence of the underlying data and the requested task, it is highly recommended in the machine learning literature to compare and test different algorithms (Moro et al., 2014; Razi & Athappilly, 2005). Thus, we compared different models of those proposed in Section 3 to predict shopping cart abandonment for our data, listed in Table 3. Additionally, we included a standard logistic regression model in our comparison serving as a non-machine learning benchmark method.

Table 3: Machine Learning Approaches for Comparison.

Approach	Description
Adaptive Boosting (AdaBoost)	Ensemble of weak learners, algorithm puts higher weights on incorrectly classified instances
Boosted Logistic Regression (LogitBoost)	Algorithm applies logistic regression techniques to the AdaBoost method by minimizing the logistic loss
Decision Tree (DT)	Algorithm recursively partitions the predictor space into subsets in which the distribution of the dependent variable is successively more homogeneous
Gradient Boosting (Linear Base Learner) with L1 and L2 Regularization (GBReg)	Ensemble of weak learners (with linear base learners), algorithm applies L1 (Lasso Regression) and L2 (Ridge Regression) Regularization
Gradient Boosting (Tree Base Learner) (GBTree)	Ensemble of weak learners (with tree base learners), algorithm minimizes the model's loss by adding weak learners sequentially using a gradient descent like procedure
Gradient Boosting (Tree Base Learner) with Dropout (GBDropout)	See GBTree, but the algorithm randomly drops boosting tree members
k-Nearest Neighbor (KNN)	Algorithm classifies an observation by assigning it to the class most common among its k nearest neighbors
Multi-Layer Perceptron Network with Dropout (MLPDropout)	Feedforward Neural Network with dropout regularization technique
Naïve Bayes (NB)	Algorithm is based on the Bayes' theorem and classifies an observation by deriving the maximum a posteriori probability
Random Forest (RF)	Ensemble of decision trees, algorithm predicts new data by aggregating the predictions of the trees
Stochastic Gradient Boosting (SGB)	Algorithm fits base learner at each iteration on the subsample of the data – instead of the full – drawn at random without replacement
Support Vector Machine with Radial Basis Kernel (SVMRadial)	Support vector machine implementation with radial basis kernel

To estimate and, hence, validate the models, we randomly partitioned the data into a training and a test subset in a 67/33 ratio, i.e., 67% (or 550,098 observations respectively) of the data are used as training data and 33% (or 270,950 observations respectively) are used as test data.

We performed k-fold cross-validation with the training data to fit the models and optimized their hyperparameters respectively (Geisser, 1975; Stone, 1974): the sample, i.e., the training data, is randomly split into k equal sized subsamples $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_k$. Of the k subsamples, one single subsample is retained as validation data to test the fitted model subsequently and the remaining $k-1$ subsamples are used as training data to fit the model. This step is repeated k times with each of the k subsamples serving as validation data once. Drawing on machine learning literature, $k = 10$ is frequently utilized since it provides an adequate trade-off between method's variance and method's bias (i.e., trade-off between the estimated parameter's expected value and the estimated value) (Bradley, 1997; Leo Breiman, 1996; Kohavi, 1995; Tibshirani & Tibshirani, 2009; P. Zhang, 1993). Thus, we applied 10-fold cross-validation.

Further, to validate and evaluate our models' performance, we considered different performance metrics that indicate the models' predictive ability. In a binary decision problem, the classifier labels observations as either positive or negative. Consequently, the classification procedure yields four different outputs in a $2x2$ confusion matrix: the sample is either correctly classified as positive (true positive (TP)), correctly classified as negative (true negative (TN)), falsely classified as positive (false positive (FP) or Type II error), or falsely classified as negative (false negative (FN) or Type I error). Thereby, accuracy is one of the most commonly used measures for classification performance due to its simplicity (see e.g., Kohavi (1995)). It is the ratio between correctly classified samples to the total number of samples:

$$\text{Accuracy}=\frac{\text{TP+TN}}{\text{P+N}}$$

However, recent research shifted away from solely presenting accuracy results since accuracy assumes balanced class distribution and equal error costs (i.e., Type I errors are equivalent to Type II errors) which is rarely the case in real world applications (Davis & Goadrich, 2006; Provost & Fawcett, 1997). To address these problems, a receiver operating characteristics (ROC) curve and thus, the area under the ROC curve (AUC)[23] have been increasingly used by the machine learning community since they are insensitive to changes in class distributions and scale-invariant (Bradley, 1997; Fawcett, 2006). A ROC graph is a two-dimensional depiction of classification performance to measure different classifiers' performances and captures the trade-off between benefits (i.e., true positives) and costs (i.e., false positives) (Fawcett, 2006). It is created by plotting the true positive rate (TPR) (or sensitivity or recall respectively) against the false positive rate (FPR) (or $1-\text{specificity}$ respectively) (Bradley, 1997; Fawcett, 2006; Hand, 2009; Provost & Fawcett, 2001):

$$\text{TPR=Sensitivity=Recall}=\frac{\text{TP}}{\text{P}};\quad \text{FPR=1-Specificity}=\frac{\text{FP}}{\text{N}};\quad \text{Specificity}=\frac{\text{TN}}{\text{N}}$$

The classifier's AUC value is a portion of the area of the unit square and its value ranges from 0.0 to 1.0 (perfect classification). It should be higher than 0.5 which equals the AUC of an uninformative classifier (Bradley, 1997; Fawcett, 2006). An important statistical property of the AUC is that a classifier's AUC is equivalent to the probability that the classifier will rank a randomly chosen positive observation higher than a randomly chosen negative observation (Fawcett, 2006).

An alternate performance measure is the F_1-Score comprising both precision and recall:

$$\text{Precision}=\frac{\text{TP}}{\text{TP+FP}};\quad \text{F}_1=2\cdot\frac{\text{Precision}\cdot\text{Recall}}{\text{Precision+Recall}}$$

Ideally, the performance measure is chosen by properly reflecting the investigation's aims to avoid misleading conclusions. Since our data is relatively balanced it seems reasonable to consider accuracy as a basic performance metric. However, as we intend to convert customers abandoning their shopping

[23] In literature, the area under the ROC curve is frequently referred to as AUROC instead of AUC.

carts into purchasers our main aim is to correctly classify actual positives (i.e., observations of shopping cart abandonments) by minimizing the Type I error. Consequently, the higher the recall the less false negatives (i.e., shopping cart abandonments classified as purchasers) have been predicted. Besides, we intend to maximize the proportion of actual positives among the predicted positives by minimizing the Type II error, i.e., purchasing customers should not be classified as non-purchasers. Thus, the higher the precision the less false positives have been predicted. The F_1-Score considers the trade-off between recall and precision. Therefore, we determined the F_1-Score, recall, and precision as our main performance metrics for the test data. Additionally, to yield valid results, we considered the ROC curve or the AUC respectively as a performance metric since it is a common measure of separability capturing the trade-off between both TPR (or sensitivity or recall respectively, analogous to F_1-Score) and FPR (i.e., how many negative instances are falsely classified as positive among the negative instances). For the training data, the best classifier during hyperparameter optimization was automatically chosen based on the AUC values.

Although prediction accuracy (i.e., AUC, F_1-Score, and accuracy) is frequently the main decision criterion when comparing different machine learning models, the models' complexity in terms of computation time and computation effort (e.g., numbers of hyperparameters to be optimized) is of similar importance regarding the application in practice and should therefore be considered as well (Doshi-Velez & Kim, 2017; Guidotti et al., 2019; Tambe et al., 2019).

5 Findings

Drawing on the training results in Table 4, gradient boosting with regularization outperformed the remaining approaches with an AUC of 0.9008. The final gradient boosting model's fitted hyperparameters did not include the lasso regression technique (L1 regularization) but made use of the ridge regression technique (L2 regularization). The gradient boosting with tree base learners and random forest yielded comparable results (AUC of 0.8953 and 0.8954 respectively) whereas naïve bayes and boosted logistic regression realized the lowest AUC values (0.8218 and 0.8381 respectively).

Regarding estimation time, the benchmark logistic regression, decision tree, and boosted logistic regression performed the fastest 10-fold cross validation to optimize the hyperparameters (20.3, 225.07, and 380.0 seconds respectively). The support vector machine and adaptive boosting were the most time-consuming models to estimate (1,306,838.6 and 703,903.9 seconds respectively). Gradient boosting with regularization yielded a moderate estimation time (4,021.28 seconds) and thus, provides an appropriate trade-off between AUC and estimation time.

Table 4: Training Data Results.

Model	Fitted Parameters: Parameter	Fitted Parameters: Fitted Value	AUC	Estimation Time (Seconds)[24]
Logistic Regression			0.8003	20.3
AdaBoost	Number of Trees Method	50 Adaboost.M1	0.8698	703,903.9
LogitBoost	Number of Boosting Iterations	21	0.8381	380.0
DT	Complexity Parameter	0.0129	0.7988	225.07
GBReg	Number of Boosting Iterations L2 Regularization L1 Regularization Learning Rate	150 0.1 0 0.3	**0.9008**	4,021.28
GBTree	Number of Boosting Iterations Maximum Tree Depth Shrinkage Minimum Loss Reduction Subsample Ratio of Columns Minimum Sum of Instance Weight Subsample Percentage	150 3 0.4 0 0.8 1 1	0.8953	6,701.14
GBDropout	Number of Boosting Iterations Maximum Tree Depth Shrinkage Minimum Loss Reduction Subsample Ratio of Columns Minimum Sum of Instance Weight Subsample Percentage Fraction of Trees dropped Probability of Skipping Dropout	150 3 0.4 0 0.8 1 0.75 0.01 0.95	0.8952	49,794.27
KNN	Maximum Number of Neighbors Distance Kernel	30 2 Optimal	0.8828	127,773.4
MLPDropout	Number of Hidden Units Dropout Rate Batch Size Learning Rate Rho Learning Rate Decay Activation Function Epochs	768 0.35 64 0.000006 0.2 0 Sigmoid 30	0.8807	218,894.0
NB	Laplace Correction Distribution Type Bandwidth Adjustment	0 Kernel Density Estimation 0.3	0.8218	5,757.49
RF	Number of Randomly Selected Predictors	14	0.8954	171,587.7

[24] With 40 GB RAM.

Model	Fitted Parameters		AUC	Estimation Time (Seconds)[24]
	Parameter	Fitted Value		
Logistic Regression			0.8003	20.3
AdaBoost	Number of Trees Method	50 Adaboost.M1	0.8698	703,903.9
	Splitting Rule	Gini		
	Minimal Node Size	35		
SGB	Number of Boosting Iterations	150	0.8800	2,033.17
	Maximum Tree Depth	3		
	Shrinkage	0.1		
	Minimum Terminal Node Size	10		
SVMRadial	Sigma Cost	0.1818 0.5	0.8808	1,306,838.6

Note: The highest AUC value is marked in bold. AdaBoost = Adaptive Boosting, DT = Decision Tree, GBDropout = Gradient Boosting with Dropout, GBReg = Gradient Boosting with L1 and L2 Regularization, GBTree = Gradient Boosting with Tree Base Learners, KNN = k-Nearest Neighbor, LogitBoost = Boosted Logistic Regression, MLPDropout = Multi-Layer Perceptron Network with Dropout, NB = Naïve Bayes, RF = Random Forest, SGB = Stochastic Gradient Boosting, SVMRadial = Support Vector Machine with Radial Basis Kernel.

Since we are rather interested in the fitted models' performances on new and unknown data, the test data results in Table 5 exhibit a higher practical relevance than the preceding results: similarly to the training data results, the gradient boosting model with regularization was superior to the remaining models regarding the test data. It yielded the best AUC (0.8182) and accuracy (82.29%) results. In line with these findings, the F_1-Score (0.8569) proves that the model is the most suitable approach in our comparison to balance the trade-off between precision and recall. With respect to its confusion matrix in the Appendix, the gradient boosting model classified 28,209 abandonments falsely as purchasers (16.42% of all abandonments) and 19,767 purchasers as abandonments respectively (19.94% of all purchasers). This is further reflected by the model's precision (0.8790) and recall (0.8358), i.e., there is a high proportion of both correctly predicted abandonments among all correctly and falsely predicted abandonments (87.90%) and correctly predicted abandonments among all actual abandonments (83.58%).

Although naïve bayes realized an extremely high recall (0.9996), its precision (0.6351) is just slightly better than random guessing. This is due to its negligible Type I error (i.e., 68 abandonments classified as purchasers (0.0004% of all abandonments)) and its substantial Type II error (i.e., 98,677 purchasers classified as abandonments (99.52% of all purchasers)). Consequently, by focusing exclusively either on precision or recall, one could draw misleading conclusions regarding model selection. The F_1-Score of the naïve bayes model (0.7767) reveals that it constitutes a suboptimal choice.

Similarly, albeit the decision tree classified a high proportion of purchasers correctly and only 12,688 (i.e., 12.80% of all purchasers) wrong, it categorized 55,634 cart abandonments as purchasers (i.e., 32.38% of all abandonments). Thus, due to its high Type I error, its recall is extremely low (0.6762), but it realized the highest precision value of all models (0.9015).

Generally, our results indicate a substantial predictive ability of the most tree-based methods (i.e., gradient boosting with regularization (and linear base learners), gradient boosting (with tree base learners), gradient boosting with dropout (and tree base learners), and random forest) compared with the remaining machine learning approaches. The latter were outperformed by tree-based models with regard to all relevant performance metrics (AUC, accuracy, and F_1-Score).[25]

Logistic regression as a non-machine learning benchmark approach yielded the lowest F_1-Score but realized a higher AUC value than several other machine learning approaches like boosted logistic regression, k-nearest neighbor, multi-layer perceptron, naïve bayes, and support vector machine. Nevertheless, it did not perform better than the tree-based methods (except for adaptive boosting, decision tree, and stochastic gradient boosting) with regard to AUC.

Moreover, the k-nearest neighbor algorithm as a basic machine learning approach outperformed more sophisticated algorithms like the multi-layer perceptron, the stochastic gradient boosting, and adaptive boosting with respect to its AUC value (0.7962).

Table 5: Test Data Results.

Model	**Performance Metrics**				
	AUC	**Accuracy**	**Precision**	**Recall**	**F_1-Score**
Logistic Regression	0.8012	78.94%	0.6677	0.8454	0.7461
AdaBoost	0.7516	78.54%	0.8024	0.8777	0.8384
LogitBoost	0.7623	77.19%	0.8349	0.7981	0.8161
DT	0.7741	74.78%	**0.9015**	0.6762	0.7728
GBReg	**0.8182**	**82.29%**	0.8790	0.8358	**0.8569**
GBTree	0.8105	81.78%	0.8701	0.8377	0.8536
GBDropout	0.8123	81.84%	0.8731	0.8350	0.8536
KNN	0.7962	80.5%	0.8585	0.8290	0.8435
MLPDropout	0.7911	80.36%	0.8503	0.8378	0.8440
NB	0.5022	63.56%	0.6351	**0.9996**	0.7767
RF	0.8108	81.75%	0.8711	0.8359	0.8531
SGB	0.7902	80.08%	0.8521	0.8299	0.8409
SVMRadial	0.7956	81.23%	0.8479	0.8578	0.8528

Note: For each column, the highest value is marked in bold. AdaBoost = Adaptive Boosting, DT = Decision Tree, GBDropout = Gradient Boosting with Dropout, GBReg = Gradient Boosting with L1 and L2 Regularization, GBTree = Gradient Boosting with Tree Base Learners, KNN = k-Nearest Neighbor, LogitBoost = Boosted Logistic Regression, MLPDropout = Multi-Layer Perceptron Network with Dropout, NB = Naïve Bayes, RF = Random Forest, SGB = Stochastic Gradient Boosting, SVMRadial = Support Vector Machine with Radial Basis Kernel.

6 Discussion

Our findings contribute to a deeper understanding regarding the successful implementation of machine learning methods for predicting online shopping cart abandoners with a strong forecast performance in order to apply marketing techniques in real-time to convert them to purchasers. Thus, we discuss our findings' theoretical contribution and practical implications in this Section. We also discuss limitations and propose suggestions for future research.

[25] Tree-based approaches are typically not subject to multicollinearity (Climent et al. 2019). Thus, we did not remove any correlated variables during the training process.

6.1 Theoretical Contribution

Overall, we fill a research gap by identifying suitable machine learning approaches for online shopping cart abandonment prediction not only in terms of accuracy but, further, in terms of practicability. Thereby, we contribute to literature in several ways. First, we are able to characterize customers abandoning their shopping cart descriptively with our data. Preceding literature on shopping cart abandonment (e.g., Close & Kukar-Kinney (2010); Huang et al. (2018); Kukar-Kinney & Close (2010)) primarily shed light on behavioral aspects of the abandonment process with experimental designs. In contrast, our research deals with unbiased clickstream data comprising an exceptionally high number of observations. Our data indicate that there is a higher proportion of new customers and mobile shoppers among customers abandoning their shopping carts compared to purchasers whereas the latter add more items to their shopping cart and view an increased number of pages on average.

Second, we contribute to literature by proposing a broad range of machine learning models to compare their performance regarding online shopping cart abandonment prediction and, thus, to predict future customers abandoning their shopping carts in real-time. Prior literature either drew on a behavioral perspective to understand the antecedents of shopping cart abandonment or predicted – more generally – purchase behaviors with conservative approaches and less observations (see e.g., Huang et al. (2018); Kukar-Kinney & Close (2010); Sismeiro & Bucklin (2004)). For our data, the gradient boosting with regularization yielded the highest accuracy (82.29%). However, with respect to our main aim, to minimize the Type I error (i.e., abandoners falsely classified as purchasers) and the Type II error (i.e., purchasers falsely classified as abandoners), we focused on the F_1-Score capturing the trade-off between precision and recall. Consistent with the accuracy results, the gradient boosting with regularization outperformed the remaining models regarding the F_1-Score (0.8569). Additionally, it realized the highest AUC value (0.8182) compared to the other models.

Overall, we found tree-based methods to be superior to the remaining machine learning approaches and logistic regression as a benchmark non-machine learning approach aligning with prior research comparing machine learning approaches in different application fields like customer churn prediction or phishing detection (Abu-Nimeh et al., 2007; Caruana & Niculescu-Mizil, 2006; Vafeiadis et al., 2015) and – similar to our context – prediction of online purchase intention (Bogina et al., 2019; Boroujerdi et al., 2014; Zheng & Liu, 2018). Thus, we complement the literature on machine learning comparisons in a marketing context.

Moreover, despite the striking importance of prediction accuracy as a decision criterion for appropriate machine learning approaches, the models' practicability with respect to modeling complexity as an essential criterion is of particular importance (Doshi-Velez & Kim, 2017; Guidotti et al., 2019; Tambe et al., 2019) but, at the same time, is often neglected by current research. Thus, we considered the models' complexity in terms of computation time and computation effort (e.g., numbers of hyperparameters to optimize) to add to literature. Thereby, the decision tree approach and boosted logistic regression yielded only slightly worse AUC results compared to gradient boosting with regularization and, simultaneously, their complexity in terms of both computation effort and time was rather low. Hence, in case of online shopping cart abandonment prediction, a decision tree model and boosted logistic regression perform well in balancing the trade-off between accuracy and complexity. Further, as stated by prior literature, we found the support vector machine approach to be extremely computationally infeasible (L'Heureux et al., 2017) despite its acceptable prediction accuracy.

6.2 Practical Implications

Our research may help to gather a comprehensive understanding of machine learning approaches for prediction or classification, particularly with regard to online shopping cart abandonment prediction. More specifically, our research provides multifold practical implications for decision makers.

Since research about advanced machine learning approaches in marketing contexts is still in its infancy (e.g., Cheung et al. (2003) and Cui et al. (2006)) we reviewed relevant literature to provide an introduction to such models, its potential applications, as well as performance metrics, and common methods for validation: for machine learning models, k-fold cross-validation is a common method to optimize the models' hyperparameters. Decision makers should draw on either recall as a performance measure if their main aim is to correctly classify abandonments or precision if they intend to avoid falsely classified purchasers. The F_1-Score considers the trade-off between both. Besides, the AUC is a common measure of separability since it is insensitive to skewed class distributions. Overall, tree-based approaches and particularly boosting methods are superior to the remaining machine learning models regarding forecast accuracy within online shopping cart abandonment prediction. Random forest yields comparable results but is rather time-consuming to estimate (171,587.7 seconds estimation time). The support vector machine and adaptive boosting are computationally intensive with estimation times of 1,306,838.6 and 703,903.9 seconds respectively.

Aside from pointing out methodological aspects, we drew on an economical perspective to enhance an organization's turnover: with regard to our data, the mean value of purchasers' ordered shopping carts (VALUE_BB) is 271.73 euro and they add 3.479 items into their shopping cart on average and thus, we expect the online retailer's sales loss for each shopping cart abandonment to be around 230 euro with 2.945 items in their shopping cart on average. Therefore, we determined a suitable approach to correctly identify shopping cart abandonments as well as purchasers: our findings indicate that gradient boosting with regularization outperformed the remaining approaches. Organizations can implement this method to predict non-purchasers in real-time when a sufficient amount of information about the customer's activities during the session has been collected. Overall, we found particularly tree-based machine learning approaches such as random forest or gradient boosting to outperform traditional classification approaches such as logistic regression and decision tree, which are frequently utilized by practitioners.

Drawing on an overall practicability perspective, decision makers may take a slight loss in prediction accuracy into account if, instead, the model's complexity in terms of computation time and effort is substantially lower: in our application context, decision tree and boosted logistic regression yielded acceptable prediction results and their computation effort was substantially lower compared to gradient boosting methods.

6.3 Limitations and Future Research

Our research is subject to limitations which stimulate further research. First, the set of useful variables for prediction was limited. With respect to extant literature (see e.g., Bucklin & Sismeiro (2003), Moe & Fader (2004a), or van den Poel & Buckinx (2005)), we expect e.g. demographic variables, historical purchase behavior, or the time customers spend on the single pages to be informative variables. Further, we did not have information about the customers' identity and thus, could not determine whether there were recurring customers. However, this information could be of great interest for analyzing online behavior and predicting shopping cart abandonment. For instance, Huang et al. (2018) anticipated that some customers might use the mobile phone for initial purchase stages (i.e., browsing and collecting information) and then switch to the computer for completing the purchase. However, such customers are listed as two distinct sessions in the current data. Another missing information concerns the value of abandoned shopping carts. While there is a variable that indicates the value of ordered carts (i.e., VALUE_BB), the value of abandoned carts can only be estimated. In line with extant literature on shopping cart abandonment (e.g., Close & Kukar-Kinney (2010); Kukar-Kinney & Close (2010)), it can be assumed that the value of ordered items influences abandoning rates and, thus, could aid the (Kukar-Kinney & Close, 2010) prediction of such. Moreover, if detailed information about spent time and further, the chronological order of customers' actions in the online shop would be available, we could decompose the session into sequences or segments. Then, we could determine a critical point in the customer's session in which abandonment can be predicted reliably with the F_1-Score or the AUC exceeding a defined threshold (see e.g., Sismeiro & Bucklin (2004)). Hence, future research could replicate the present study with more detailed data, e.g. between-site clickstream data (i.e., panel data collected by media measurement company), that are typically more comprehensive and frequently used in clickstream analyses (see e.g., Moe & Fader (2004a)).

Second, we excluded just-browsing customers from our investigation. A possible direction for future research could be to conduct a multi-class classification by differentiating between purchasers, abandonments, and just-browsing customers, similar to the cluster analysis of Moe (2003).

Third, the models' performance strongly depends on the optimized hyperparameters which may be a time-consuming procedure for some of the models. Therefore, we considered only a limited range of possible hyperparameter values. Moreover, other values of k in cross-validation could lead to different results.

Lastly, a real-time implementation requires a certain amount of data to be collected before the model can make a reliable decision.

By implementing these models, companies may detect shopping cart abandoners in real-time and subsequently convert some of them into purchasers by making use of targeted marketing measures such as individual chat pop-ups, coupons or special discounts. For instance, Close & Kukar-Kinney (2010) suggest human-human interactions (i.e., live chats with employees or other online shoppers) to avoid shopping cart abandonment. These could pop-up on the website if the online user is predicted to abandon by the machine learning model. Therefore, future research is recommended to test whether pop-up messages and offers impact customers' online shopping behavior and can prevent online shopping cart abandonment.

7 Conclusion

Online shopping cart abandonment can inhibit corporate growth and hence, harm a company's success within its competitive environment. Simultaneously, the emergence of the Internet's commercial usage leads to the ability to track consumers' online activities and online behavior resulting in clickstream data.

Thus, to identify online shopping cart abandoners by extracting valuable knowledge from such clickstream data we proposed different machine learning approaches. We analyzed data of a German online retailer comprising 821,048 observations and fitted the models using 10-fold cross validation. Thereby, our paper contributes to extant literature by combining research fields of both online shopping cart abandonment and clickstream data with machine learning approaches.

Our data indicate that among customers abandoning their shopping carts there is a higher proportion of new customers and mobile shoppers compared to purchasers whereas the latter add more items to their shopping cart and have a higher number of page viewings on average. Moreover, our comparison results prove that gradient boosting with regularization is a suitable method to distinguish between abandonments and purchasers yielding an AUC of 0.8182, an F_1-Score of 0.8569, and an accuracy of 82.29%. Nevertheless, a decision tree or boosted logistic regression may be suitable alternatives yielding only slightly less accurate prediction results and being computationally more feasible.

Nevertheless, research on clickstream data combined with machine learning approaches is still in its infancy – particularly in a marketing context. Thereby, machine learning will be inevitable for e-commerce businesses to be successful in the long-term and the analysis provided in this paper shall stimulate further research on this topic.

Appendix: Confusion Matrices

Model	Prediction	Actual	
		0 (Purchaser)	1 (Abandonment)
Logistic Regression	0 (Purchaser)	83,817	41,722
	1 (Abandonment)	15,335	130,076
AdaBoost	0 (Purchaser)	62,009	21,005
	1 (Abandonment)	37,143	150,793
LogitBoost	0 (Purchaser)	72,036	34,692
	1 (Abandonment)	27,116	137,106
DT	0 (Purchaser)	86,464	55,634
	1 (Abandonment)	12,688	116,164
GBReg	0 (Purchaser)	79,385	28,209
	1 (Abandonment)	19,767	143,589
GBTree	0 (Purchaser)	77,662	27,875
	1 (Abandonment)	21,490	143,923
GBDropout	0 (Purchaser)	78,294	28,352
	1 (Abandonment)	20,858	143,446
KNN	0 (Purchaser)	75,687	29,383
	1 (Abandonment)	23,465	142,415
MLPDropout	0 (Purchaser)	73,803	27,869
	1 (Abandonment)	25,349	143,929
NB	0 (Purchaser)	475	68
	1 (Abandonment)	98,677	171,730
RF	0 (Purchaser)	77,903	28,197
	1 (Abandonment)	21,249	143,601
SGB	0 (Purchaser)	74,409	29,217
	1 (Abandonment)	24,743	142,581
SVMRadial	0 (Purchaser)	72,724	24,427
	1 (Abandonment)	26,428	147,371

Note: AdaBoost = Adaptive Boosting, DT = Decision Tree, GBDropout = Gradient Boosting with Dropout, GBReg = Gradient Boosting with L1 and L2 Regularization, GBTree = Gradient Boosting with Tree Base Learners, KNN = k-Nearest Neighbor, LogitBoost = Boosted Logistic Regression, MLPDropout = Multi-Layer Perceptron Network with Dropout, NB = Naïve Bayes, RF = Random Forest, SGB = Stochastic Gradient Boosting, SVMRadial = Support Vector Machine with Radial Basis Kernel.

References

Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007). A comparison of machine learning techniques for phishing detection. In Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit. ACM. https://doi.org/10.1145/1299015.1299021

Bengio, Y. (2009). Learning deep architectures for AI. Foundations and trends in machine learning: Volume 2, issue 1. Now Publishers Inc. http://lib.myilibrary.com/detail.asp?id=293219

Bogina, V., Mokryn, O., & Kuflik, T. (2019). Will this session end with a purchase? Inferring current purchase intent of anonymous visitors. Electronic Commerce Research and Applications, 34, 100836. https://doi.org/10.1016/j.elerap.2019.100836

Boroujerdi, E. G., Mehri, S., Sadeghi Garmaroudi, S., Pezeshki, M., Rashidi Mehrabadi, F., Malakouti, S., & Khadivi, S. (2014). A study on prediction of user's tendency toward purchases in websites based on behavior models. In 2014 6th Conference on Information and Knowledge Technology (IKT): 27 - 29 May 2014, Shahrood. IEEE. https://doi.org/10.1109/ikt.2014.7030334

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 144–152.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and regression trees. Chapman & Hall.

Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140.

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.

Bucklin, R. E., & Sismeiro, C. (2003). A Model of Web Site Browsing Behavior Estimated on Clickstream Data. Journal of Marketing Research, 40(3), 249–267.

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In W. Cohen (Ed.), Proceedings of the 23rd international conference on Machine learning. ACM. https://doi.org/10.1145/1143844.1143865

Chatterjee, P., Hoffman, D. L., & Novak, T. P. (2003). Modeling the Clickstream: Implications for Web-Based Advertising Efforts. Marketing Science, 22(4), 520–541. https://doi.org/10.1287/mksc.22.4.520.24906

Chen, H., Chiang, R. H. L., & Story, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, 36(4), 1165–1188.

Cheung, K.-W., Kwok, J. T., Law, M. H., & Tsui, K.-C. (2003). Mining customer product ratings for personalized marketing. Decision Support Systems, 35(2), 231–243. https://doi.org/10.1016/S0167-9236(02)00108-2

Chipman, H. A., George, E. I., & McCulloch, R. E. (1998). Bayesian CART Model Search. Journal of the American Statistical Association, 93(443), 935–948.

Cho, C.-H., Kang, J., & Cheon, H. J. (2006). Online Shopping Hesitation. CyberPsychology & Behavior, 9(3), 261–274.

Chollet, F., & Allaire, J. J. (2018). Deep Learning with R. Manning Publications.

Climent, F., Momparler, A., & Carmona, P. (2019). Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. Journal of Business Research, 101, 885–896. https://doi.org/10.1016/j.jbusres.2018.11.015

Close, A. G., & Kukar-Kinney, M. (2010). Beyond buying: Motivations behind consumers' online shopping cart use. Journal of Business Research, 63(9-10), 986–992. https://doi.org/10.1016/j.jbusres.2009.01.022

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Cui, G., Wong, M. L., & Lui, H.-K. (2006). Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming. Management Science, 52(4), 597–612. https://doi.org/10.1287/mnsc.1060.0514

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning, 233–240.

Deng, L. (2011). An Overview of Deep-Structured Learning for Information Processing. Proceedings of Asian-Pacific Signal & Information Processing Annual Summit and Conference (APSIPA-ASC), 1–14.

Dietterich, T. G. (2002). Ensemble Learning. In M. A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks (2nd ed., pp. 110–125). The MIT Press.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. ArXiv Preprint ArXiv:1702.08608.

Duda, R. O., Hart, P. E., & Stork, D. G. (1973). Pattern classification and scene analysis (3rd edition). Wiley.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. AI Magazine, 17(3), 37–54.

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, 55, 119–139.

Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 29(5), 1189–1232.

Friedman, J. (2002). Stochastic Gradient Boosting. Computational Statistics & Data Analysis, 38(4), 367–378.

Friedman, J., Hastie, T. H., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. The Annals of Statistics, 28(2), 337–407.

Geisser, S. (1975). The Predictive Sample Reuse Method with Applications. Journal of the American Statistical Association, 70(350), 320–328.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. ACM Computing Surveys, 51(5), 1–42. https://doi.org/10.1145/3236009

Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. Machine Learning, 77(1), 103–123. https://doi.org/10.1007/s10994-009-5119-5

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. Neural Computation, 18(7), 1527–1554.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Huang, G.-H., Korfiatis, N., & Chang, C.-T. (2018). Mobile shopping cart abandonment: The roles of conflicts, ambivalence, and hesitation. Journal of Business Research, 85, 165–174. https://doi.org/10.1016/j.jbusres.2017.12.008

Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial Neural Networks: A Tutorial. Computer, 29(3), 31–44.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning (Vol. 103). Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Kelly, S., & Noonan, C. (2017). The doing of datafication (and what this doing does): Practices of edification and the enactment of new forms of sociality in the Indian public health service. Journal of the Association for Information Systems, 18(2), 872–899.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. International Joint Conference on Artifical Intelligence (IJCAI), 1137–1145.

Kohavi, R., & Provost, F. (2001). Applications of Data Mining to Electronic Commerce. In Applications of Data Mining to Electronic Commerce (pp. 5–10). Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-1627-9_1

Kukar-Kinney, M., & Close, A. G. (2010). The determinants of consumers' online shopping cart abandonment. Journal of the Academy of Marketing Science, 38(2), 240–250. https://doi.org/10.1007/s11747-009-0141-5

L'Heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. M. (2017). Machine Learning With Big Data: Challenges and Approaches. IEEE Access, 5, 7776–7797. https://doi.org/10.1109/access.2017.2696365

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. R News, 2(3), 18–22.

Lycett, M. (2013). 'Datafication': making sense of (big) data in a complex world. European Journal of Information Systems, 22(4), 381–386. https://doi.org/10.1057/ejis.2013.10

Moe, W. W. (2003). Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream. Journal of Consumer Psychology, 13(1-2), 29–39. https://doi.org/10.1207/s15327663jcp13-1&2_03

Moe, W. W., & Fader, P. S. (2004a). Capturing evolving visit behavior in clickstream data. Journal of Interactive Marketing, 18(1), 5–19. https://doi.org/10.1002/dir.10074

Moe, W. W., & Fader, P. S. (2004b). Dynamic Conversion Behavior at E-Commerce Sites. Management Science, 50(3), 326–335. https://doi.org/10.1287/mnsc.1040.0153

Montgomery, A. L. (2001). Applying Quantitative Marketing Techniques to the Internet. Interfaces, 31(2), 90–108.

Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (2004). Modeling Online Browsing and Path Analysis Using Clickstream Data. Marketing Science, 23(4), 579–595. https://doi.org/10.1287/mksc.1040.0073

Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. Decision Support Systems, 62, 22–31. https://doi.org/10.1016/j.dss.2014.03.001

Muhammad, I., & Yan, Z. (2015). Supervised machine learning approaches: A survey. ICTACT Journal on Soft Computing, 05(03), 946–952. https://doi.org/10.21917/ijsc.2015.0133

Opitz, D., & Maclin, R. (1999). Popular Ensemble Methods: An Empirical Study. Journal of Artificial Intelligence Research, 11, 169–198. https://doi.org/10.1613/jair.614

Provost, F., & Fawcett, T. (1997). Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions. Proc of the 3rd International Conference on Knowledge Discovery and Data Mining, 43–48.

Provost, F., & Fawcett, T. (2001). Robust Classification for Imprecise Environments. Machine Learning, 42(3), 203–231.

Rajamma, R. K., Paswan, A. K., & Hossain, M. M. (2009). Why do shoppers abandon shopping cart? Perceived waiting time, risk, and transaction inconvenience. Journal of Product & Brand Management, 18(3), 188–197. https://doi.org/10.1108/10610420910957816

Razi, M. A., & Athappilly, K. (2005). A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. Expert Systems with Applications, 29(1), 65–74. https://doi.org/10.1016/j.eswa.2005.01.006

Rokach, L. (2010). Ensemble-based classifiers. Artificial Intelligence Review, 33(1-2), 1–39. https://doi.org/10.1007/s10462-009-9124-7

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536.

Sakar, C. O., Polat, S. O., Katircioglu, M., & Kastro, Y. (2019). Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks. Neural Computing and Applications, 31(10), 6893–6908. https://doi.org/10.1007/s00521-018-3523-0

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of Research and Development, 3(3), 210–229.

Schapire, R. E., Freund, Y., Bartlett, P., & Lee, W. S. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. The Annals of Statistics, 26(5), 1651–1686.

Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. Neural Networks, 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Sismeiro, C., & Bucklin, R. E. (2004). Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach. Journal of Marketing Research, 41(3), 306–323.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15, 1929–1958.

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 111–133.

Suykens, J. A. K., & Vandewalle, J. (1999). Least Squares Support Vector Machine Classifiers. Neural Processing Letters, 9(3), 293–300.

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. California Management Review, 61(4), 15–42. https://doi.org/10.1177/0008125619867910

Tibshirani, R. J., & Tibshirani, R. (2009). A Bias Correction for the Minimum Error Rate in Cross-Validation. The Annals of Applied Statistics, 3(2), 822–829.

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. (2015). A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory, 55, 1–9. https://doi.org/10.1016/j.simpat.2015.03.003

van den Poel, D., & Buckinx, W. (2005). Predicting online-purchasing behaviour. European Journal of Operational Research, 166(2), 557–575. https://doi.org/10.1016/j.ejor.2004.04.022

Vapnik, V. N. (1982). Estimation of Dependences Based on Empirical Data. Springer Publishing Company, Incorporated.

Williams, N., Zander, S., & Armitage, G. (2006). A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. ACM SIGCOMM Computer Communication Review, 36(5), 5–16.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14(1), 35–62.

Zhang, P. (1993). Model Selection Via Multifold Cross Validation. The Annals of Statistics, 21(1), 299–313.

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. Information Fusion, 42, 146–157. https://doi.org/10.1016/j.inffus.2017.10.006

Zheng, B., & Liu, B. (2018). A scalable purchase intention prediction system using extreme gradient boosting machines with browsing content entropy. In 2018 IEEE International Conference on Consumer Electronics (ICCE). IEEE. https://doi.org/10.1109/icce.2018.8326351

4 Conclusion

Since the spread of the Internet's commercial usage, e-commerce businesses have perceived constant pressure caused by different challenges. This thesis particularly shed light on two selected contemporary phenomenona with which e-commerce businesses are struggling: Society's increasing demand for sustainability as well as environmental awareness and Artificial Intelligence-driven approaches to manage the overwhelming flood of data about online consumer behavior. As both research fields are extremely granular and have an almost infinite number of different facets, many sub-areas are still underresearched. The overall aim of this research was therefore twofold: (1) Gathering a better understanding of consumers' sustainable clothing consumption behavior and (2) proposing solutional approaches for Artificial Intelligence-driven problems in the e-commerce context. As this thesis consists of the previously presented research papers, the following concluding remarks now intend to integrate the research papers' main findings into an overall summary instead of discussing the single contributions and limitations again.

When it comes to consumers' sustainable clothing consumption, consumers' purchase intention is determined by intrinsic (i.e., attitudinal) rather than extrinsic (i.e., social pressure from peers) motives. Further, design or aesthetics are extremely important to consumers and can deter them from buying despite an initial purchase intention (Research Paper No. 1). Aside from design, Research Paper No. 3 proved that such conventional apparel attributes (e.g., fit and comfort, quality, and price-performance ratio) are still more important to consumers than sustainable apparel attributes. The most important sustainable apparel attributes were the garment's durability and fair working conditions and employees' wages (social sustainability). Consumers' price sensitivity was further confirmed within Research Paper No. 2, as consumers highly value discounts. Overall, consumers still lack knowledge about sustainable clothing and demand information, education, as well as transparency (Research Paper No. 2). Consumers are concerned about false claims about a product's greenness and fear greenwashing (Research Paper No. 1). Furthermore, women were found to put more emphasis on sustainability than men (Reseach Papers No. 2 and 3).

With regard to the challenge of predictive analytics' usage in an e-commerce context, machine learning approaches (particularly random forest) were found to be powerful tools to reliably forecast future call arrivals (as opposed to traditional literature using time series models) with call centers still being a crucial customer touchpoint for e-commerce businesses. Moreover, models with explanatory variables were found to be better able to capture special days (e.g., holidays), while models without explanatory variables are better able to capture ordinary weekdays (Research Paper No. 4). Research Paper No. 5 built on these insights and proposed a new approach for forecasting call center call arrivals by incorporating the advantages of approaches without explanatory variables (i.e., time series models) and with explanatory variables (i.e., machine learning as well as regression models). The proposed model was found to outperform established benchmark models in different forecasting settings. Research Paper No. 6 investigated another prediction problem, which is specifically inherent to the online context: using real clickstream data, different machine learning models were compared with regard to prediction accuracy and practicability for the prediction of online shopping cart abandonment. Particularly gradient boosting or simple decision trees were found to be superior.

E-commerce businesses will most likely persist in the long term. Old challenges will be met, while novel challenges will pop up due to social change, progressive digitalization, or other unforeseen disruptions. Undoubtedly, e-commerce businesses cannot ignore the current and future challenges, otherwise they will be squeezed out of the market by their competitors. This thesis may serve as a guidance for e-commerce managers, bring some light into the e-commerce research darkness, and provide impulses for future research. I am sure there are many more challenges to come, while the current ones are still relatively unexplored.

Appendix: Additional publications

Table 4: Additional publications.

Author(s) & Year	Title	Medium	Status
Rausch, T. M. & Albrecht T. (2020)	The impact of lead time and model selection on the accuracy of call center arrivals' forecasts	28th European Conference on Information Systems (ECIS)	Published
Brand, B. M. & Rausch, T. M. (2021)	Examining sustainability surcharges for outdoor apparel using Adaptive Choice-Based Conjoint analysis	Journal of Cleaner Production, 289	Published
Rausch, T. M. & Brand, B. M.	Gotta buy 'em all? Online shopping cart abandonment among new and existing customers	International Journal of Electronic Business	Under Review (First Revision)
Rausch, T. M. & Kopplin, C. S.	Listen to your hearth: Consumers' purchase behavior of plant-based food substitutes	Psychology & Marketing	Under Review
Brand, B. M., Kopplin, C. S., & Rausch, T. M.	Cultural differences in processing online customer reviews: holistic versus analytic thinkers	Journal of Business Research	Under Review
Brand, B. M., Rausch, T. M., & Brandel, J.	The importance of sustainability aspects when purchasing online: comparing Generation X and Generation Z	European Journal of International Management	Under Review
Kopplin, C. S. & Rausch, T. M.	Above and beyond meat: the role of consumers' dietary behavior for the purchase of plant-based food substitutes	Review of Managerial Science	Under Review
Kopplin, C. S. & Rausch, T. M.	A Funnel Perspective on Technology Acceptance and Links to Preference	Information Systems Journal	Under Review

References

Achabou, M. A., & Dekhili, S. (2013). Luxury and sustainable development: Is there a match? Journal of Business Research, 66(10), 1896–1903. https://doi.org/10.1016/j.jbusres.2013.02.011

Agarwal, R., & Dhar, V. (2014). Editorial —Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. Information Systems Research, 25(3), 443–448. https://doi.org/10.1287/isre.2014.0546

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. In J. Kuhl & J. Beckmann (Eds.), Springer series in social psychology. Action control: From cognition to behavior (1st ed., pp. 11–39). Springer. https://doi.org/10.1007/978-3-642-69746-3_2

Ajzen, I. (1988). Attitudes, Personality and Behavior. Open University Press.

Ajzen, I. (1991). The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-t

Ajzen, I., & Fishbein, M. (1980). Understanding attitudes and predicting social behavior. Prentice-Hall. https://www.scienceopen.com/document?vid=c20c4174-d8dc-428d-b352-280b05eacdf7

Aksin, O. Z., & Harker, P. T. (1999). To Sell or Not to Sell. Journal of Service Research, 2(1), 19–33. https://doi.org/10.1177/109467059921003

Akter, S., & Wamba, S. F. (2016). Big data analytics in E-commerce: A systematic review and agenda for future research. Electronic Markets, 26(2), 173–194. https://doi.org/10.1007/s12525-016-0219-0

Alba, J., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A., & Wood, S. (1997). Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces. Journal of Marketing, 61(3), 38–53. https://doi.org/10.1177/002224299706100303

Allwood, J. M., Laursen, S. E., Russell, S. N., Rodríguez, C. M. de, & Bocken, N. (2008). An approach to scenario analysis of the sustainability of an industrial sector applied to clothing and textiles in the UK. Journal of Cleaner Production, 16(12), 1234–1246. https://doi.org/10.1016/j.jclepro.2007.06.014

Armitage, C. J., & Conner, M. (2001). Efficacy of the Theory of Planned Behaviour: A meta-analytic review. The British Journal of Social Psychology, 40(Pt 4), 471–499. https://doi.org/10.1348/014466601164939

Armstrong, C. M., Niinimäki, K., Lang, C., & Kujala, S. (2016). A Use-Oriented Clothing Economy? Preliminary Affirmation for Sustainable Clothing Consumption Alternatives. Sustainable Development, 24(1), 18–31. https://doi.org/10.1002/sd.1602

Babin, B. J., Darden, W. R., & Griffin, M. (1994). Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value. Journal of Consumer Research, 20(4), 644. https://doi.org/10.1086/209376

Bagozzi, R. P. (1992). The Self-Regulation of Attitudes, Intentions, and Behavior. Social Psychology Quarterly, 55(2), 178. https://doi.org/10.2307/2786945

Banerjee, S. B., Iyer, E. S., & Kashyap, R. K. (2003). Corporate Environmentalism: Antecedents and Influence of Industry Type. Journal of Marketing, 67(2), 106–122. https://doi.org/10.1509/jmkg.67.2.106.18604

Beath, C., Becerra-Fernandez, I., Ross, J., & Short, J. (2012). Finding value in the information explosion. MIT Sloan Management Review, 53(4), 18.

Bellman, R. (1978). An Introduction to artificial intelligence: Can computers think? Boyd & Fraser.

Bianchi, C., & Birtwistle, G. (2012). Consumer clothing disposal behaviour: A comparative study. International Journal of Consumer Studies, 36(3), 335–341. https://doi.org/10.1111/j.1470-6431.2011.01011.x

Bijmolt, T. H. A., Leeflang, P. S. H., Block, F., Eisenbeiss, M., Hardie, B. G. S., Lemmens, A., & Saffert, P. (2010). Analytics for Customer Engagement. Journal of Service Research, 13(3), 341–356. https://doi.org/10.1177/1094670510375603

Bird, B. (1988). Implementing Entrepreneurial Ideas: The Case for Intention. Academy of Management Review, 13(3), 442–453. https://doi.org/10.5465/amr.1988.4306970

Brady, M. K., & Cronin, J. J. (2001). Some New Thoughts on Conceptualizing Perceived Service Quality: A Hierarchical Approach. Journal of Marketing, 65(3), 34–49. https://doi.org/10.1509/jmkg.65.3.34.18334

Brown, B. J., Hanson, M. E., Liverman, D. M., & Merideth, R. W. (1987). Global sustainability: Toward definition. Environmental Management, 11(6), 713–719. https://doi.org/10.1007/BF01867238

Brynjolfsson, E., & McAfee, A. (2017). The business of artificial intelligence. Harvard Business Review, 7, 1–20.

Cardenas, A. A., Manadhata, P. K., & Rajan, S. P. (2013). Big Data Analytics for Security. IEEE Security & Privacy, 11(6), 74–76. https://doi.org/10.1109/msp.2013.138

Chan, R. Y. K. (2001). Determinants of Chinese consumers' green purchase behavior. Psychology & Marketing, 18(4), 389–413. https://doi.org/10.1002/mar.1013

Charniak, E., & McDermott, D. V. (1987). Introduction to artificial intelligence (Reprinted with corrections). Addison-Wesley Series in computer science. Addison-Wesley.

Chen, H., Chiang, R. H. L., & Story, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, 36(4), 1165–1188.

Childers, T. L., Carr, C. L., Peck, J., & Carson, S. (2001). Hedonic and utilitarian motivations for online retail shopping behavior. Journal of Retailing, 77(4), 511–535. https://doi.org/10.1016/S0022-4359(01)00056-2

Close, A. G., & Kukar-Kinney, M. (2010). Beyond buying: Motivations behind consumers' online shopping cart use. Journal of Business Research, 63(9-10), 986–992. https://doi.org/10.1016/j.jbusres.2009.01.022

Constantiou, I. D., & Kallinikos, J. (2015). New Games, New Rules: Big Data and the Changing Context of Strategy. Journal of Information Technology, 30(1), 44–57. https://doi.org/10.1057/jit.2014.17

Conway, G. R. (1985). Agroecosystem analysis. Agricultural Administration, 20(1), 31–55. https://doi.org/10.1016/0309-586X(85)90064-0

Davenport, T. H. (2006). Competing on analytics. Harvard Business Review, 84(1-10), 98. http://hosteddocs.ittoolbox.com/competinganalytics.pdf

Davenport, T. H., Barth, P., & Bean, R. (2012). How 'Big Data' is Different. MIT Sloan Management Review, 54(1), 22–24. https://www.hbs.edu/faculty/publication%20files/smr-how-big-data-is-different_782ad61f-8e5f-4b1e-b79f-83f33c903455.pdf

Dean, A. M. (2007). The Impact of the Customer Orientation of Call Center Employees on Customers' Affective Commitment and Loyalty. Journal of Service Research, 10(2), 161–173. https://doi.org/10.1177/1094670507309650

Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. Decision Support Systems, 55(1), 412–421. https://doi.org/10.1016/j.dss.2012.05.048

Di Fatta, D., Patton, D., & Viglia, G. (2018). The determinants of conversion rates in SME e-commerce websites. Journal of Retailing and Consumer Services, 41, 161–168. https://doi.org/10.1016/j.jretconser.2017.12.008

D'Souza, C., Taghian, M., Lamb, P., & Peretiatko, R. (2007). Green decisions: Demographics and consumer understanding of environmental labels. International Journal of Consumer Studies, 31(4), 371–376. https://doi.org/10.1111/j.1470-6431.2006.00567.x

Elkington, J. (1997). Cannibals with forks: The triple bottom line of 21st century business. Capstone.

Ellen, P. S. (1994). Do we know what we need to know? Objective and subjective knowledge effects on pro-ecological behaviors. Journal of Business Research, 30(1), 43–52. https://doi.org/10.1016/0148-2963(94)90067-1

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. AI Magazine, 17(3), 37. https://doi.org/10.1609/aimag.v17i3.1230

Finn, A., & Louviere, J. J. (1992). Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. Journal of Public Policy & Marketing, 11(2), 12–25. https://doi.org/10.1177/074391569201100202

Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention, and behavior: An introduction to theory and research. Addison-Wesley.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone Call Centers: Tutorial, Review, and Research Prospects. Manufacturing & Service Operations Management, 5(2), 79–141. https://doi.org/10.1287/msom.5.2.79.16071

Gefen, Karahanna, & Straub (2003). Trust and TAM in Online Shopping: An Integrated Model. MIS Quarterly, 27(1), 51. https://doi.org/10.2307/30036519

George, G., Haas, M. R., & Pentland, A. (2014). Big Data and Management. Academy of Management Journal, 57(2), 321–326. https://doi.org/10.5465/amj.2014.4002

Germann, F., Lilien, G. L., Fiedler, L., & Kraus, M. (2014). Do Retailers Benefit from Deploying Customer Analytics? Journal of Retailing, 90(4), 587–593. https://doi.org/10.1016/j.jretai.2014.08.002

GFA, & BCG. (2017). Pulse of the fashion industry. https://globalfashionagenda.com/wp-content/uploads/2017/05/Pulse-of-the-Fashion-Industry_2017.pdf

Goes, P. B. (2014). Editor's comments: big data and IS research. MIS Quarterly, 38(3), iii–viii. https://www.jstor.org/stable/26634980

Govindan, K., Cheng, T., Mishra, N., & Shukla, N. (2018). Big data analytics and application for logistics and supply chain management. Transportation Research Part E: Logistics and Transportation Review, 114, 343–349. https://doi.org/10.1016/j.tre.2018.03.011

Goworek, H., Fisher, T., Cooper, T., Woodward, S., & Hiller, A. (2012). The sustainable clothing market: An evaluation of potential strategies for UK retailers. International Journal of Retail & Distribution Management, 40(12), 935–955. https://doi.org/10.1108/09590551211274937

Goworek, H., Oxborrow, L., Claxton, S., McLaren, A., Cooper, T., & Hill, H. (2018). Managing sustainability in the fashion business: Challenges in product development for clothing longevity in the UK. Journal of Business Research, In Press. https://doi.org/10.1016/j.jbusres.2018.07.021

Gudigantala, N., Bicen, P., & Eom, M. (2016). An examination of antecedents of conversion rates of e-commerce retailers. Management Research Review, 39(1), 82–114. https://doi.org/10.1108/MRR-05-2014-0112

Harris, F., Roby, H., & Dibb, S. (2016). Sustainable clothing: Challenges, barriers and interventions for encouraging more sustainable consumer behaviour. International Journal of Consumer Studies, 40(3), 309–318. https://doi.org/10.1111/ijcs.12257

Hiller Connell, K. Y. (2010). Internal and external barriers to eco-conscious apparel acquisition. International Journal of Consumer Studies, 34(3), 279–286. https://doi.org/10.1111/j.1470-6431.2010.00865.x

Hirschman, E. C. (1983). Predictors of Self-Projection, Fantasy Fulfillment, and Escapism. The Journal of Social Psychology, 120(1), 63–76. https://doi.org/10.1080/00224545.1983.9712011

Hoffman, D. L., Novak, T. P., & Peralta, M. (1999). Building consumer trust online. Communications of the ACM, 42(4), 80–85. https://doi.org/10.1145/299157.299175

Hornik, J. (1984). Subjective vs. Objective Time Measures: A Note on the Perception of Time in Consumer Behavior. Journal of Consumer Research, 11(1), 615. https://doi.org/10.1086/208998

Huang, G.-H., Korfiatis, N., & Chang, C.-T. (2018). Mobile shopping cart abandonment: The roles of conflicts, ambivalence, and hesitation. Journal of Business Research, 85, 165–174. https://doi.org/10.1016/j.jbusres.2017.12.008

Hughner, R. S., McDonagh, P., Prothero, A., Shultz, C. J., & Stanton, J. (2007). Who are organic food consumers? A compilation and review of why people purchase organic food. Journal of Consumer Behaviour, 6(2-3), 94–110. https://doi.org/10.1002/cb.210

Hustvedt, G., & Dickson, M. A. (2009). Consumer likelihood of purchasing organic cotton apparel. Journal of Fashion Marketing and Management: An International Journal, 13(1), 49–65. https://doi.org/10.1108/13612020910939879

Ibrahim, N. F., & Wang, X. (2019). A text analytics approach for online retailing service improvement: Evidence from Twitter. Decision Support Systems, 121, 37–50. https://doi.org/10.1016/j.dss.2019.03.002

ISO. (2006). ISO 14040:2006 Environmental management — Life cycle assessment — Principles and framework. ISO. https://undocs.org/en/A/CONF.151/26/Rev.1(vol.I)

Joergens, C. (2006). Ethical fashion: Myth or future trend? Journal of Fashion Marketing and Management: An International Journal, 10(3), 360–371. https://doi.org/10.1108/13612020610679321

Kano, N., Seraku, N., Takahashi, F., & Tsuji S. (1984). Attractive quality and must-be quality. Journal of the Japanese Society for Quality Control, 14(2), 39–48.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, 62(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kautish, P., Paul, J., & Sharma, R. (2019). The moderating influence of environmental consciousness and recycling intentions on green purchase behavior. Journal of Cleaner Production, 228, 1425–1436. https://doi.org/10.1016/j.jclepro.2019.04.389

Kelly, S., & Noonan, C. (2017). The doing of datafication (and what this doing does): Practices of edification and the enactment of new forms of sociality in the Indian public health service. Journal of the Association for Information Systems, 18(2), 872–899.

Kiron, D., Prentice, P. K., & Ferguson, R. B. (2012). Innovating with analytics. MIT Sloan Management Review, 54(1), 47.

Kollmuss, A., & Agyeman, J. (2002). Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? Environmental Education Research, 8(3), 239–260. https://doi.org/10.1080/13504620220145401

Koutsabasis, P., Stavrakis, M., Viorres, N., Darzentas, J. S., Spyrou, T., & Darzentas, J. (2008). A descriptive reference framework for the personalisation of e-business applications. Electronic Commerce Research, 8(3), 173–192. https://doi.org/10.1007/s10660-008-9021-1

Kukar-Kinney, M., & Close, A. G. (2010). The determinants of consumers' online shopping cart abandonment. Journal of the Academy of Marketing Science, 38(2), 240–250. https://doi.org/10.1007/s11747-009-0141-5

Laitala, K., & Boks, C. (2012). Sustainable clothing design: Use matters. Journal of Design Research, 10(1-2), 121–239.

LaPiere, R. (1934). Attitudes Versus Actions. Social Forces, 13, 230–237.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. MIT Sloan Management Review, 52(2), 21–32.

Lehrer, C., Wieneke, A., vom Brocke, J., Jung, R., & Seidel, S. (2018). How Big Data Analytics Enables Service Innovation: Materiality, Affordance, and the Individualization of Service. Journal of Management Information Systems, 35(2), 424–460. https://doi.org/10.1080/07421222.2018.1451953

Liu, Y. (2014). Big data and predictive business analytics. The Journal of Business Forecasting, 33(7), 40.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American Psychologist, 57(9), 705–717. https://doi.org/10.1037/0003-066x.57.9.705

Louviere, J. J., & Woodworth, G. G. (1991). Best-worst scaling: A model for the largest difference judgments. Working Paper. https://scholar.google.de/citations?user=rp3xtscaaaaj&hl=de&oi=sra

Lundblad, L., & Davies, I. A. (2016). The values and motivations behind sustainable fashion consumption. Journal of Consumer Behaviour, 15(2), 149–162. https://doi.org/10.1002/cb.1559

Lycett, M. (2013). ‘Datafication’: making sense of (big) data in a complex world. European Journal of Information Systems, 22(4), 381–386. https://doi.org/10.1057/ejis.2013.10

Maichum, K., Parichatnon, S., & Peng, K.-C. (2016). Application of the Extended Theory of Planned Behavior Model to Investigate Purchase Intention of Green Products among Thai Consumers. Sustainability, 8(1077), 1–20. https://doi.org/10.3390/su8101077

Mathwick, C., Malhotra, N., & Rigdon, E. (2001). Experiential value: conceptualization, measurement and application in the catalog and Internet shopping environment. Journal of Retailing, 77(1), 39–56. https://www.sciencedirect.com/science/article/pii/s0022435900000452

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. Harvard Business Review, 90(10), 60–68.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5(4), 115–133. https://doi.org/10.1007/BF02478259

McDowell, W. C., Wilson, R. C., & Kile, C. O. (2016). An examination of retail website design and conversion rate. Journal of Business Research, 69(11), 4837–4842. https://doi.org/10.1016/j.jbusres.2016.04.040

Milbrath, L. (1984). A proposed value structure for a sustainable society. The Environmentalist, 4(2), 113–124. https://doi.org/10.1016/S0251-1088(84)80007-8

Moe, W. W., & Fader, P. S. (2004a). Capturing evolving visit behavior in clickstream data. Journal of Interactive Marketing, 18(1), 5–19. https://doi.org/10.1002/dir.10074

Moe, W. W., & Fader, P. S. (2004b). Dynamic Conversion Behavior at E-Commerce Sites. Management Science, 50(3), 326–335. https://doi.org/10.1287/mnsc.1040.0153

Morgan, L. R., & Birtwistle, G. (2009). An investigation of young fashion consumers’ disposal habits. International Journal of Consumer Studies, 33(2), 190–198. https://doi.org/10.1111/j.1470-6431.2009.00756.x

Ngai, E., & Gunasekaran, A. (2007). A review for mobile commerce research and applications. Decision Support Systems, 43(1), 3–15. https://doi.org/10.1016/j.dss.2005.05.003

Nilsson, N. J. (1998). Artificial intelligence: A new synthesis. The Morgan Kaufmann Series in Artificial Intelligence. Kaufmann. http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10595529

Odum, E. P. (1987). Basic ecology. Saunders College Publishing.

Paul, J., Modi, A., & Patel, J. (2016). Predicting green product consumption using theory of planned behavior and reasoned action. Journal of Retailing and Consumer Services, 29, 123–134. https://doi.org/10.1016/j.jretconser.2015.11.006

Paul, J., & Rana, J. (2012). Consumer behavior and purchase intention for organic food. Journal of Consumer Marketing, 29(6), 412–422. https://doi.org/10.1108/07363761211259223

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. Health Information Science and Systems, 2(1), 3. https://doi.org/10.1186/2047-2501-2-3

Rich, E., & Knight, K. (1991). Artificial intelligence (2nd edition). McGraw-Hill.

Russell, S. J., Norvig, P., Davis, E., & Edwards, D. (2020). Artificial intelligence: A modern approach (4th edition). Always learning. Pearson.

Sheeran, P. (2002). Intention-Behavior Relations: A Conceptual and Empirical Review. European Review of Social Psychology, 12(1), 1–36. https://doi.org/10.1080/14792772143000003

Sheth, J. N. (1983). An Integrative Theory of Patronage Preference and Behavior. In W. R. Darden & R. F. Lusch (Eds.), Patronage behavior and retail management (pp. 9–28). North-Holland.

Sismeiro, C., & Bucklin, R. E. (2004). Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach. Journal of Marketing Research, 41(3), 306–323. https://doi.org/10.1509/jmkr.41.3.306.35985

Statista. (2021). eCommerce worldwide. https://www.statista.com/outlook/243/100/ecommerce/worldwide

Taufique, K. M., & Vaithianathan, S. (2018). A fresh look at understanding Green consumer behavior among young urban Indian consumers through the lens of Theory of Planned Behavior. Journal of Cleaner Production, 183, 46–55. https://doi.org/10.1016/j.jclepro.2018.02.097

Turing, A. (1950). Computing machinery and intelligence. Mind(236), 433–460. https://www.abelard.org/turpap/turpap.php

United Nations. (1987). Report of the World Commission on Environment and Development: Our Common Future. https://netzwerk-n.org/wp-content/uploads/2017/04/0_Brundtland_Report-1987-Our_Common_Future.pdf

United Nations. (1992). Report of the United Nations Conference on Environment and Development. https://undocs.org/en/A/CONF.151/26/Rev.1(vol.I)

van den Poel, D., & Buckinx, W. (2005). Predicting online-purchasing behaviour. European Journal of Operational Research, 166(2), 557–575. https://doi.org/10.1016/j.ejor.2004.04.022

van Loon, P., Deketele, L., Dewaele, J., McKinnon, A., & Rutherford, C. (2015). A comparative analysis of carbon emissions from online retailing of fast moving consumer goods. Journal of Cleaner Production, 106, 478–486. https://doi.org/10.1016/j.jclepro.2014.06.060

Varshney, U., & Vetter, R. (2002). Mobile Commerce: Framework, Applications and Networking Support. Mobile Networks and Applications, 7(3), 185–198. https://doi.org/10.1023/A:1014570512129

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. Journal of Business Research, 70, 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychological Bulletin, 132(2), 249–268. https://doi.org/10.1037/0033-2909.132.2.249

White, K., Habib, R., & Hardisty, D. J. (2019). How to SHIFT Consumer Behaviors to be More Sustainable: A Literature Review and Guiding Framework. Journal of Marketing, 83(3), 22–49. https://doi.org/10.1177/0022242919825649

Wicker, A. W. (1969). Attitudes versus Actions: The Relationship of Verbal and Overt Behavioral Responses to Attitude Objects. Journal of Social Issues, 25(4), 41–78. https://doi.org/10.1111/j.1540-4560.1969.tb00619.x

Wiese, A., Toporowski, W., & Zielke, S. (2012). Transport-related CO2 effects of online and brick-and-mortar shopping: A comparison and sensitivity analysis of clothing retailing. Transportation Research Part D: Transport and Environment, 17(6), 473–477. https://doi.org/10.1016/j.trd.2012.05.007

Wixom, B. H., Yen, B., & Relich, M. (2013). Maximizing value from business analytics. MIS Quarterly Executive, 12(2).

Wolfinbarger, M., & Gilly, M. C. (2001). Shopping Online for Freedom, Control, and Fun. California Management Review, 43(2), 34–55. https://doi.org/10.2307/41166074

Young, W., Hwang, K., McDonald, S., & Oates, C. J. (2009). Sustainable consumption: Green consumer behaviour when purchasing products. Sustainable Development, 18(1), 20-31. https://doi.org/10.1002/sd.394

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The Behavioral Consequences of Service Quality. Journal of Marketing, 60(2), 31–46. https://doi.org/10.1177/002224299606000203