Advancements in Sentiment Analysis and Dictionary Building for German Financial Texts

Dissertation

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

vorgelegt
von
Matthias Pöferlein
aus
Nürnberg

Dekan/in:	Prof. Dr. André Meyer
Erstberichterstatter:	Prof. Dr. Klaus Schäfer
Zweitberichterstatter:	Prof. Dr. Rolf Uwe Fülbier StB
Tag der mündlichen Prüfung:	16.08.2024

Acknowledgements

During the process of writing a dissertation, you do not just learn a lot about the relevant subject, but also about yourself. The need to keep motivated despite numerous setbacks, such as unfeasible research projects or rejected submissions, was an immense challenge. Without the support and encouragement of special people in my life, this project would never have been completed.

First and foremost, I would like to thank Prof. Dr. Klaus Schäfer, for his support during this process, not only with his expert advice and in finding a topic. Without his support in difficult personal times during my doctorate, which went far beyond what I would have taken for granted, this thesis would never have been completed. I would also like to thank Prof. Dr. Rolf Uwe Fülbier StB for his willingness to be available as a second reviewer.

Many of my colleagues at the Chair of Finance and Banking at the University of Bayreuth were an important part of the motivation and professional exchange I always needed: Daniel Hertrich, Julian Kettl, Dorian Marshall and Maximilian Zurek supported me and helped me not to lose the fun of doing a doctorate. The two good souls of the chair, Iris Gödde and Christine Landmann, who always had an open ear and good advice for me, are of course also responsible for the positive atmosphere and support in everyday life.

Special thanks go to the many research assistants at our chair, without whom a thesis of this scope would not have been possible. I would especially like to mention Kilian Trost and Mike-Roland Westermayer, who helped me more than once with their in-depth Python knowledge and commitment.

The biggest gratitude, however, goes to my family, especially my wife Alexandra. Without her and her support, I would have given up on my dissertation project several times. The amount of free time I have been given, the evenings and weekends we have missed together and the countless times I have been persuaded not to give up cannot be outweighed by my gratitude. Words cannot describe how grateful I am to you.

An important motivator in my decision to pursue a doctorate was that I wanted to make my parents proud at the time, and today, I also want to make my children proud. Certainly, the changes in my family situation during my doctorate, with all the ups and downs we had to overcome together, extended the time I spent on my doctorate. But I don't regret a single minute that I spent with my children rather than on my laptop.

Abstract

Textual analysis is an increasingly important field in accounting and finance research, primarily focusing on the analysis of English texts. This thesis discusses several important adaptions and extensions as well as use cases for textual analysis of German-speaking finance-related texts.

The first two parts of this thesis are based on the first finance-related dictionary available for the German language, contributed by Bannier et al. (2019b). The initial part proposes several reforms and extensions to the original word list and tests the suitability of the most common measurements for sentiment. We are able to show that the adapted dictionary in combination with a relative measurement of sentiment, is able to calculate more significant relations between the sentiment of a speech by a CEO at the Annual General Meeting and subsequent abnormal stock returns.

The second part applies further improvements to obtain more significant results in calculating the sentiment of German-speaking annual reports to forecast future return on assets and future return on equity. Furthermore, we successfully tested different adaptions of negations in order to further optimize the results obtained.

In the third and final part, we propose an alternative approach to the usage of surveys in order to answer the question whether and to what extend German savings and cooperative banks use artificial intelligence. Therefore, we introduce a combined methodology from the approaches Word2vec and bag-of-words, to obtain an individual word list. This approach allows us to obtain a detailed overview of the application of artificial intelligence in the analyzed banking groups.

Contents

List of Figures

List of Tables

1. Introduction

The well-known quote “the pen is mightier than the sword” (Bulwer-Lytton, 1839, p. 39) can be seen figuratively as the guiding idea of this thesis since all three parts are based on the fundamental idea that it is possible to extract deeper information from words.

The accounting and finance research field of textual analysis, which has become increasingly important in recent years, focuses primarily on the sentiment of texts and their financial implications, for example on stock prices or financial ratios (Bannier et al., 2019b, pp. 82f.; Kearney & Liu, 2014, pp. 182-184). A very striking definition of sentiment is given by Algaba et al. (2020, p. 514). They define that “sentiment is the disposition of an entity toward an entity, expressed via a certain medium.” The stated disposition can be conveyed quantitatively through numbers although it is primarily expressed qualitatively, using text, audio, or visual media (Algaba et al., 2020, p. 514).

The two primary techniques for converting qualitative sentiment information into measurable sentiment variables are the dictionary-based method, often known as bag-of-words, and machine learning (Chakraborty & Bhattacharjee, 2020, p. 762; Kearney & Liu, 2014, pp. 174f.). In the dictionary-based approach predefined sets of words are used by an algorithm to categorize the words of a text into various sentiment groups such as positive or negative. From the categorized word counts, several sentiment metrics can be derived (F. Li, 2010, p. 146; Loughran & McDonald, 2015, p. 1). One of the most crucial success factors is the appropriate choice of a word list. A distinction is made between general and domain specific word lists, where specific word lists for the finance domain have proven to be superior (Algaba et al., 2020, p. 523; Chakraborty & Bhattacharjee, 2020, p. 764; Loughran & McDonald, 2011, p. 36, 2015, pp. 2f.).

On the other hand, the machine learning technique trains on a selection of typically labeled linguistic data to obtain models. These models subsequently classify and predict the sentiment present within texts (Algaba et al., 2020, p. 525; Kearney & Liu, 2014, pp. 175f.; Rice & Zorn, 2019, p. 1). In machine learning, a distinction can be made between supervised and unsupervised learning. Supervised machine learning requires a labeled data set, for example sentiment values, to build the specific model. Whereas an unsupervised approach is building its own categories or representations and therefore is looking for hidden structures in the data (Algaba et al., 2020, p. 525; Loughran & McDonald, 2020b, p. 364). However, machine learning models are often black-boxes and are therefore hard to explain or replicate (Algaba et al., 2020, p. 523; Krause et al., 2016, p. 5686).

Various aspects must be considered when comparing and deciding on a specific methodology. A comprehensible issue could be the drifting apart of the underlying hypothesis, where machine learning techniques might identify relevant words for future stock returns, that are not capturing sentiment. A possible example could be the identification of words for firm attributes that relate to positive or negative stock returns (Loughran & McDonald, 2020b, p. 365). Another important issue for future research is the possibility to replicate results. Due to the nature of constructing machine learning algorithms, studies based on those techniques are not easy to replicate (Luo & Zhou, 2020, p. 107).

There is currently no consensus in the academic literature regarding the preferred methodology. Due to the possible creation of words that are pseudodummy variables for identifying a particular firm or industry, when using machine learning techniques Loughran and McDonald (2020b, p. 365) are in favor of the dictionary-based method. Frankel et al. (2022) on the other hand, find that machine learning offers a significant improvement in explanatory power, when capturing the sentiment at 10-K filing and conference-call dates. Nevertheless, according to Chakraborty and Bhattacharjee (2020, p. 772) and Luo and Zhou (2020, p. 108), the dictionary-based approach is the most commonly used methodology. Therefore, this thesis focuses on the dictionary-based approach, as it is the most commonly used and most replicable.

English is the predominant language in research communication and therefore can be considered as the almost exclusive language of science (Drubin & Kellogg, 2012, p. 1399; Garfield & Welljams-Dorof, 1990, p. 10). Due to this fact it is reasonable that most of the literature on the analysis of financial texts is based on English language dictionaries (Bannier et al., 2018, pp. 39-42; Kearney & Liu, 2014, pp. 182-184). Under these conditions, the dictionary created by Loughran and McDonald (2011) has become the established standard in the field of financial textual analysis (Kearney & Liu, 2014, p. 175; Loughran & McDonald, 2016, pp. 1204-1206).

The unavailability of a finance-specific German dictionary restricted research in the German-speaking world primarily to general dictionaries such as SentiWS (Remus et al., 2010) and LIWC (Meier et al., 2018; Wolf et al., 2008). Therefore, resulting research contributions were only available to a limited extent (Ammann & Schaub, 2016; Dorfleitner et al., 2016; Fritz & Tows, 2018). Addressing this noticeable gap, Bannier et al. (2019b) unveiled a German domain-specific dictionary, based on the lists established by Loughran and McDonald (2011).

Since its introduction, this dictionary has proven valuable in various studies (Bannier et al., 2017, 2019a; Röder & Walter, 2019; Tillmann & Walter, 2018, 2019).

The first part of this thesis uses the described first finance-related dictionary available for the German language and improves the existing word list by proposing several reforms and extensions. Additionally, the most commonly used measurements of sentiment are compared to evaluate the one being most appropriate for measuring the tone of textual documents in finance. We show that those applications lead to more significant relations between the sentiment of a speech by a CEO at the Annual General Meeting and subsequent abnormal stock returns.

Based on these findings we conduct further improvements in the second part, to provide more significant results. In addition, the so far not applied negations were tested in various forms. We are able to show that those implementations lead to more significant results in forecasting future return on assets and future return on equity, by using the adapted calculation of the sentiment on German-speaking annual reports.

In the third part, we show the variety of possible applications of the dictionary-based approach, by complementarily creating an own dictionary as an alternative to traditional surveys on the topic of artificial intelligence.

Currently artificial intelligence is an emerging topic in the European and German banking market. The European Central Bank conducted a survey among 105 large banks in the summer of 2022 to assess the status of their digital transformation. Even though classified as having a lower business relevance, 60% of institutions surveyed, stated that they use artificial intelligence (European Central Bank, 2023a, 2023c, p. 8). In addition, Prof. Dr. Joachim Wuermeling, member of the Executive Board of the Deutsche Bundesbank, sees incredible opportunities in the upcoming developments in artificial intelligence for the financial sector (Wuermeling, 2023).

Savings and cooperative banks account for a large portion of German financial institutions (Deutsche Bundesbank, 2023, pp. 6, 12). Whether and to what extent artificial intelligence is currently used in these two banking groups represents a current gap in financial research. One possibility to answer this question is the usage of questionnaires or interviews. However, this approach brings with it the additional challenge of being costly and hard to replicate (Algaba et al., 2020, p. 519). Therefore, given the assumption that German savings and cooperative banks communicate publicly about the successful implementation or use of new technologies, such as artificial intelligence, we use a dictionary-based approach to identify banks using artificial intelligence. Due to the absence of a sufficient word list, we used multiple iterations of the neural network Word2vec to create a customized list of relevant words.

Considering the results of all three parts the contribution of this thesis to the literature of textual analysis of German-speaking financial texts is an extension and optimization of the first available German finance-related dictionary. Therefore, future research in analyzing the sentiment of German-speaking texts in finance can be conducted more accurately and more thoroughly. Additionally, this thesis provides a domain-specific dictionary for analyzing the usage of artificial intelligence by German banks. The proposed method for creating domain-specific dictionaries, as well as the approach to use external communication instead of surveys can be applied to solve other research questions.

2. Sentiment Analysis of German Texts in Finance: Improving and Testing the BPW Dictionary

[1]

Abstract

Using the dictionary-based approach to measure the sentiment of finance-related texts is primarily focused on English-speaking content. This is due to the need for domain-specific dictionaries and the primary availability of those in English. Through the contribution of Bannier et al. (2019b), the first finance-related dictionary is available for the German language. Because of the novelty of this dictionary, this paper proposes several reforms and extensions of the original word lists. Additionally, we tested multiple measurements of sentiment. We show that using the edited and extended dictionary to calculate a relative measurement of sentiment, central assumptions regarding textual analysis can be fulfilled and more significant relations between the sentiment of a speech by a CEO at the Annual General Meeting and subsequent abnormal stock returns can be calculated.

[1] This chapter has been published as: *Sentiment Analysis of German Texts in Finance: Improving and Testing the BPW Dictionary* in *Journal of Banking and Financial Economics*, 16 (2), pp. 5–24. To standardize the three components of this thesis, the term "I" has been replaced by "we".

2.1 Introduction

In recent years, textual analysis has become an important part of accounting and finance research. This is due to the fact that the availability and quantity of digitally available texts are constantly increasing. Additionally, the information encoded in those texts in the form of sentiment can be obtained in an easier and more targeted way through recent developments in the field of textual analysis (Bannier et al., 2019b, pp. 82f.; Gentzkow et al., 2019, p. 535; Loughran & McDonald, 2015, p. 1).

Algaba et al. (2020, p. 514) define sentiment "[…] as the disposition of an entity toward an entity, expressed via a certain medium. […] This disposition can be conveyed numerically but is primarily expressed qualitatively through text, audio, and visual media." The two most common methods for transforming qualitative sentiment data into quantitative sentiment variables are the dictionary-based approach (also referred to as bag-of-words) and machine learning (Kearney & Liu, 2014, pp. 174f.). The dictionary-based approach is a rule-based approach that uses an algorithm to classify a text's words or phrases into different categories based on predefined rules or categories like dictionaries[2] (F. Li, 2010, p. 146). More specifically, the dictionary assigns words into different categories like positive or negative. Using the total count of positive, negative, and all words, several measurements of sentiment can be calculated (Loughran & McDonald, 2015, p. 1). The machine learning or statistical approach relies on statistical techniques to classify the content of documents (Kearney & Liu, 2014, p. 175; F. Li, 2010, p. 146).

When using the dictionary-based approach, the chosen dictionary has a specific importance (Bannier et al., 2019b, p. 80; Loughran & McDonald, 2015, p. 1). As described in the following section, the newly developed word list provided by Bannier et al. (2019b) (BPW Dictionary) gives researchers the possibility to analyze German-speaking texts in finance in a more targeted way.

Due to the novelty of this BPW Dictionary, we propose several reforms and extensions with the objective of improving its performance. Therefore, the main hypothesis of this paper is that the edited version of the BPW (BPW_N) can improve results compared to its original (BPW_O). So far, the BPW Dictionary has been used primarily to analyze the market reaction to the sentiment of CEO speeches held at the Annual General Meeting (AGM) of German stock

[2] As stated in Loughran and McDonald (2015, p. 10), the terms dictionary and word list are used interchangeably.

companies (Bannier et al., 2017, 2019a). Therefore, this paper also uses comparable speeches for testing the possible improvements.

As stated in the following course of this paper, there are several different possibilities to measure the sentiment of textual documents in a dictionary-based approach. Given the fact that this is the first German domain-specific dictionary for the field of finance, the additional research question is which sentiment measure is the most appropriate for measuring the tone of textual documents in the field of finance using a German domain-specific dictionary. This topic is especially relevant, given the previous use of exclusively four different measurements of sentiment using the BPW Dictionary (Bannier et al., 2017, p. 11, 2019a, p. 10; Röder & Walter, 2019, p. 396; Tillmann & Walter, 2018, pp. 9, 21, 2019, pp. 69f.).

The contribution of this paper to the literature on textual analysis of German texts is the extension and reform of the only existing German finance-related dictionary and testing the performance of the original against the new dictionary. Additionally, the suitability of the primarily used measures of sentiment in a business context is analyzed. This should make it possible for researchers to measure the sentiment of German texts in finance more accurately and more thoroughly.

The paper proceeds as follows. In the following section, we will give a short review of the relevant literature regarding textual analysis with a particular focus on analyzing financial texts. The data and the parsing procedure applied to it, as well as the used dictionaries form the third section. The used measurements of sentiment and the empirical approach to obtain the results given in section five are presented in the fourth section. Section six concludes.

2.2 Literature review

The extensive field of textual analysis in finance is ideally pictured in the surveys of Kearney and Liu (2014) and the online appendix of Bannier et al. (2019b). Other important surveys giving additional information and areas of caution regarding textual analysis in finance are Algaba et al. (2020) and Loughran and McDonald (2016).

One of the first steps in measuring the tone of a text is selecting a dictionary or word list (Loughran & McDonald, 2015, p. 1). According to Loughran and McDonald (2016, p. 1200), four different word lists have been primarily used by researchers classifying English finance-related texts. These are the two general dictionaries – General Inquirer (Stone et al., 1966) and DICTION (Hart, 2000) – and the two word lists generated for finance-related texts: Henry (Henry, 2006, 2008) and Loughran and McDonald (Loughran & McDonald, 2011).

In the contributions of Henry (2006, 2008) and Loughran and McDonald (2011), the usage of general word lists for different forms of textual content like news, earnings press releases or annual reports was widely criticized in favor of domain-specific word lists, because of the high possibility of misclassification (Algaba et al., 2020, pp. 523-525; Lewis & Young, 2019, pp.598f.; Mengelkamp et al., 2016, p. 7; Price et al., 2012, p. 1006). Loughran and McDonald (2011, p. 49) analyzed that 73.8% of negative words in the general dictionary General Inquirer do not have a negative meaning in a business context.

Despite the fact that the Henry word lists have been used for different purposes like conference calls (Davis et al., 2015, pp. 641, 647; Price et al., 2012, pp. 996f.) or news (Jandl et al., 2014, pp. 4, 7), the lists provided by Loughran and McDonald have become predominant (Kearney & Liu, 2014, p. 175) in the field of finance. They have been used in the classification of many different kinds of written financial content like news (Garcia, 2013, pp. 1272, 1274; Gurun & Butler, 2012, pp. 562, 566), conference calls (Mayew & Venkatachalam, 2012, pp. 2, 20) and annual reports (Ahmed & Elshandidy, 2016, p. 179; Jegadeesh & Wu, 2013, pp. 713, 715).

Due to the absence of a German domain-specific dictionary for the field of finance, research was limited to different versions of general dictionaries like LIWC (Meier et al., 2018; Wolf et al., 2008) or SentiWS (Remus et al., 2010), resulting in little research (Ammann & Schaub, 2016; Dorfleitner et al., 2016; Fritz & Tows, 2018). The first public available business-related dictionary for the German language was introduced by Bannier et al. (2019b). The introduced word lists are based on the predominant lists by Loughran and McDonald (Bannier et al., 2019b, p. 79) and have already been successfully used (Bannier et al., 2017, 2019a; Röder & Walter, 2019; Tillmann & Walter, 2018, 2019).

As stated in Bannier et al. (2019a, p. 2), the contributions of Bannier et al. (2017, 2019a) are the primary studies analyzing the information content of CEO speeches delivered at the Annual General Meeting. Thus, this paper is also an essential complementary contribution to the information content of CEO speeches.

2.3 Data

2.3.1 Data Source

We collected the transcripts of the CEO speeches from the companies' homepages, since there is no database for German CEO speeches delivered at the AGM. We screened the web pages of all companies listed in the DAX, MDAX, SDAX or TECDAX between 2008 and 2019 for transcripts of CEO speeches delivered at the AGM. Since not all companies publish transcripts on their homepage, we could find 976 speeches of 139 companies for the initial sample. We had to remove 53 speeches that were not delivered by the CEO. All available additional information, such as annotations, audio and video material provided by the company or other providers, was evaluated to confirm that the speeches were initially delivered in German. Therefore we had to exclude another 50 speeches. Additionally, 49 transcripts contained speeches of several speakers and required filtering of the relevant parts. Due to a delisting, we had to delete one additional speech. The final sample consists of 872 speeches from 125 companies. Comparing the contributions of Bannier et al. (2017, p. 10) (338 speeches) and Bannier et al. (2019a, p. 7) (457 speeches), this is the most comprehensive collection of German CEO speeches so far. An overview of the sample creation is given in Table 2.1. We obtained all other variables from Thomson Reuters Datastream.

Table 2.1
Sample Creation

Source/Filter	Sample Size	Removed Observations
CEO speeches found on the companies' homepages	976	
Speeches not held by the CEO	923	53
Speeches held initially in English	873	50
Speeches where no CAR or CAV could be calculated	872	1
Final Sample	872	

2.3.2 Used Dictionaries

The mutated vowels "ä", "ö" and "ü" in the German language can alternatively be written as "ae", "oe" and "ue". To get the updated form of the BPW_O (BPW_N), the first step is to add the alternative spelling of words with mutated vowels because the BPW_O does not include those. As a part of the parsing procedure, we deleted hyphens. Therefore, stop words written with hyphens had to be included without hyphens. Overall, we deleted 21 words that also appear on the positive and negative list of the BPW_O from the stop word list. In total, 144 stop words occurred twice and had to be deleted, because 110 surnames match company or given names (e.g. "kummer"). After extending for mutated vowels and hyphens, another 34 words occurred twice. Finally, we added 244 additional stop words through a translation of the generic list provided by Loughran & McDonald (2020a) (LMD stop words). A summary of the conducted steps and the resulting alteration of the number of words on the different lists is given in Table 2.2.

Table 2.2
Updating of the BPW

	Positive	Negative	Stop words
BPW_O total words	2,223	10,147	3,682
Adding mutated vowels	+ 626	+ 2,514	+ 218
Including words without hyphens			+ 153
Delete doubles (positive/negative)			- 21
Delete doubles			- 144
Adding additional LMD stop words			+ 244
BPW_N total words	2,849	12,661	4,132

Due to the update of the BPW_O, this paper examines the suitability of two different dictionaries.

2.3.3 Parsing

Given expressed criticism regarding unspecified parsing rules and the related difficulty to replicate existing studies (Loughran & McDonald, 2015, p. 2), We give a detailed overview of performed text manipulation.

In the first step, the collected PDF files were transferred into TXT files using UTF-8 encoding (Bannier et al., 2017, p. 10, 2019a, p. 9; Meier et al., 2018, p. 29). In order to automatically process the speeches, they need to be parsed. Due to the unique and unsystematic character of the collected texts, manual corrections need to be conducted before using an automated parser. Those include the removal of headlines, disclaimers, legal notices, and additional information (e.g. the positioning of slides).

The subsequent automated parser was programmed using Python. First of all, we replaced typographic ligatures (Bannier et al., 2017, p. 10, 2019a, p. 9) and hyphens (Loughran & McDonald, 2011, internet appendix) and converted all words to lowercase (Fritz & Tows, 2018, p. 61; Picault & Renault, 2017, p. 139). Additionally, we removed special characters (Allee & Deangelis, 2015, p. 247; Mengelkamp et al., 2016, p. 4), numbers (Boudt & Thewissen, 2019, p. 84; Schmeling & Wagner, 2016, p. 8), punctuation (Gentzkow et al., 2019, p. 538; Loughran et al., 2009, p. 41), and multiple whitespaces (González et al., 2019, p. 433; Schmeling & Wagner, 2016, p. 8). Finally, we removed words with fewer than three characters (Bannier et al., 2017, p. 10, 2019a, pp. 9f.; Loughran et al., 2009, p. 42). Depending on the used dictionary

(BPW_O or BPW_N), we deleted the predefined individual stop words. Stop words are very common words but have relatively little meaning or rarely contribute information on their own, despite being essential to the grammatical structure of a sentence (Bannier et al., 2017, p. 10; Gentzkow et al., 2019, p. 538).

Furthermore, we included an important automated alteration[3] of the words “betrug” and “sorgen” prior to the automated parser. When written in lowercase, the words were changed to “betrugnoneg” and “sorgennoneg.” This is because of the very frequent occurrence of those words in the analyzed texts (betrug: 812, sorgen: 344) and the characteristics of the German language. When written with a first capital letter, both words are nouns, where the word “Betrug” means “fraud” and the word “Sorgen” means “sorrow,” which are both negative words in a business context and due to that are justifiably on the list of negative words. But when written entirely in lowercase, both words are verbs. In this case, the word “betrug” means “amounted” and “sorgen” means “care,” which does not have a negative connotation. Without this automated alteration, the exclusive use of lowercase words would lead to a wrong and exaggerated number of negative words.

2.4 Methodology

2.4.1 Measurement of Sentiment

Using Python, we counted the occurrence of positive (p) and negative (n) words from each of the two dictionaries as well as the total number of words (w) for each document. By using those three numbers, a variety of measurements of sentiment can be calculated. Even though the notations differ in several contributions, this paper focuses on the most widely used measurements to evaluate which sentiment measure is the most appropriate for the tone of textual documents in finance.

First of all, we calculated a simple share of negative and positive words as in Loughran and McDonald (2011, p. 46), Ferguson et al. (2015, p. 7), and Ammann and Schaub (2016, p. 2):

$$N=\frac{n}{w} \tag{2.1}$$

$$P=\frac{p}{w} \tag{2.2}$$

[3] Note that this automated alteration was only implemented when using the updated form of the dictionary provided by Bannier et al. (2019b) (BPW_N).

Other studies, as stated below, use the relation of positive and negative words rather than their individual fractions. However, there are different approaches to measure this relation. In this paper, we used the three most prominent relative measurements of sentiment.

Following the approach of Davis et al. (2015, p. 646), Loughran and McDonald (2015, p. 4), and Picault and Renault (2017, p. 141), we measured the sentiment of a text as the number of positive words minus the number of negative words divided by the total number of words:

$$Tone=\frac{p\text{-}n}{w} \tag{2.3}$$

Other contributions switch the numerator while retaining the notation "*Tone*" (Franke, 2018, p. 9; Y. H. Kim & Meschke, 2014, p. 33). To prevent misinterpretations, this paper uses the term *ITone* for inverted tone.

$$ITone=\frac{n\text{-}p}{w} \tag{2.4}$$

In contrast to *Tone* and *ITone*, the variable *NTone* used by Henry (2008, p. 386), Price et al. (2012, p. 998), and Henry and Leone (2016, p. 159) only focuses on the number of positive and negative words and is not altered by the length of the analyzed text. It therefore gives the NetTone:

$$NTone=\frac{p\text{-}n}{p+n} \tag{2.5}$$

Also, a fourth relative variable *NToneSQ* as in Henry (2008, p. 393) is estimated, by squaring the variable *NTone*.

Given this variety of six different measurements of sentiment, this paper adds the two measurements *InvTone* and *NToneSQ* to the four already tested calculations, when using the BPW_O (Bannier et al., 2017, p. 11, 2019a, p. 10; Röder & Walter, 2019, p. 396; Tillmann & Walter, 2018, pp. 9, 21, 2019, pp. 69f.).

In this paper, following Apel and Blix Grimaldi (2012, p. 9), Davis et al. (2015, p. 653), and Bannier et al. (2017, p. 15), all words found are weighted equally. This approach makes it possible for other researchers to replicate and further develop the results of this contribution, due to the independence of the weighting scheme from the texts used. This approach and the superiority of equal weighting is also supported by Henry and Leone (2016, p. 166).

2.4.2 Empirical Approach

By using linear regressions, we conduct one of the most common approaches for analyzing the impact of sentiment on stock prices (Kearney & Liu, 2014, p. 177). Therefore, we performed several linear regressions for ten different dependent variables in the following form:

$$Dep_j=\alpha_0+\alpha_1 Sentiment_j+\sum_{k=1}^{K}\propto_k Control_{kj}+\varepsilon_j \tag{2.6}$$

Dep represents two different forms of variables to measure the effect of speech sentiment on stock prices and trading.

To obtain the effect on stock prices, we calculated cumulative abnormal returns (*CAR*). The abnormal returns are calculated by the market adjusted model using the value weighted market index CDAX. Following Henry (2006, p. 5, 2008, p. 385), Loughran and McDonald (2011, p. 41), Henry and Leone (2016, p. 159), and Bannier et al. (2017, p. 12, 2019a, p. 8), the CARs are calculated through cumulating the abnormal returns (*AR*) over a predefined event period (event window) with length *T*. We obtained the individual *AR*s by subtracting the returns (*R*) of the analyzed stock (*j*) from the return of the CDAX for a given day (*t*):

$$AR_{j,t}=R_{j,t}-R_{CDAX.t} \tag{2.7}$$

$$CAR_{j,T}=\sum_{t=0}^{T}AR_{j,t} \tag{2.8}$$

Based on Loughran and McDonald (2011, p. 41), Boudt and Thewissen (2019, p. 95), and Bannier et al. (2019a, p. 9), this paper solely uses event windows beginning on the day of the AGM ($t=0$), to only measure the effect of the CEO speeches. Therefore, the five different trading day event windows [0,1], [0,3], [0,5], [0,15], and [0,30] were used following contributions examining similar texts like CEO letters or CEO conference calls (Bannier et al., 2019a, p. 9; Boudt & Thewissen, 2019, p. 95; Doran et al., 2012, p. 412; Loughran & McDonald, 2011, p. 41; Mayew & Venkatachalam, 2012, p. 20).

Additionally, we performed all regressions with cumulative abnormal trading volumes (*CAV*) for the five different event windows. We calculated the different CAVs according to Bannier et al. (2017, p. 47, 2019a, p. 38) and Price et al. (2012, p. 1000) as:

$$AV_{j,t}=\frac{VOL_{j,t}}{\overline{VOL_{j,t}}}-1 \tag{2.9}$$

$$CAV_{j,T}=\sum_{t=0}^{T} AV_{j,t} \tag{2.10}$$

Here $VOL_{j,t}$ is the trading volume for firm *j* at day *t*, and $\overline{VOL_{j,t}}$ is the mean volume for firm *j* from trading day $t=-252$ to $t=-2$. Due to different estimation windows in the primary studies of Bannier et al. (2017, p. 47, 2019a, p. 38), we selected a combined period of time in accordance with Price et al. (2012, p. 1000).

We used the six above mentioned measurements of sentiment separately for each of the ten different dependent variables *Dep*.

The comprehensive set of control variables *Control* consist of eleven different variables (*K*), which include the firm size (*SIZE*), the market to book value (*M2B*), leverage (*LEV*), volatility (*VOLA*), volume (*VOL*), number of words (*COUNT*), individual words (*IND*), return on assets (*ROA*), the earnings surprise (*EPS_SP*), and the dividend surprise (*DIV_SPP* and *DIV_SPN*) (Bannier et al., 2017, p. 47, 2019a, pp. 38f.; Doran et al., 2012, p. 426; Loughran & McDonald, 2011, p. 63). The calculation of the individual control variables can be found in the appendix (section 2.7,Table 2.11).

We used the variables *SIZE*, *VOL*, and *COUNT* in a logarithmic form. When using *CAV*, the variable *VOL* is excluded from the regression. Additionally, we used year fixed effects.

2.5 Results

2.5.1 Summary Statistics

We report summary statistics for the analyzed sample of 872 CEO speeches in the following three tables.

Table 2.3 provides descriptive statistics for all calculated CARs and CAVs. While we could calculate CARs for all different event windows, the calculation of CAVs is only partially possible based on the availability of data. As stated in Bannier et al. (2017, p. 16), the means of all CARs are economically small, indicating no market reaction due to the AGM. In comparison, CAVs are in the mean higher than 1, indicating an abnormal trading volume caused by the AGM.

Table 2.3
Descriptive Statistics for CARs and CAVs

Statistic	N	Mean	St. Dev.	Min	Max	Pctl. (25)	Pctl. (75)
CAR01	872	0.001	0.027	-0.184	0.104	-0.013	0.015
CAR03	872	-0.0002	0.031	-0.285	0.116	-0.017	0.018
CAR05	872	-0.002	0.037	-0.171	0.138	-0.021	0.018
CAR015	872	-0.004	0.059	-0.271	0.229	-0.035	0.033
CAR030	872	-0.005	0.087	-0.459	0.321	-0.057	0.046
CAV01	849	2.790	2.192	0.041	32.141	1.654	3.195
CAV03	841	4.825	3.076	0.054	37.987	3.130	5.645
CAV05	839	6.787	3.705	0.087	41.084	4.604	7.927
CAV015	827	16.498	7.859	0.595	82.829	12.060	19.007
CAV030	817	30.614	12.434	0.931	124.574	23.843	35.132

Source: Author’s calculation based on data from Thomson Reuters Datastream.

Because of the extension of the stop word list, the mean words counted are 22.7% lower for BPW_N, as given in Table 2.4. In addition to the change of sentiment measures, the reduction of words also improves calculation times of algorithms for measuring textual sentiment. The deletion of positive words from the stop words list leads to an increase in the number of positive words. In contrast, the mean number of negative words decreases due to the treatment of the words “betrug” and “sorgen.” The combination of those changes leads to an increase in all six sentiment measures on average. The mean number of positive and negative words combined with positive means for the measurements *Tone*, *NTone*, and *NToneSQ* show that the speeches delivered by the CEOs are on average positive. This positivity of speeches is slightly higher for the BPW_N dictionary. As stated in Doran et al. (2012, p. 414) for earnings conference calls using the Henry word list, it is not surprising that the general sentiment is positive, reflecting the effort of CEOs to present their information as positive as possible. This positive wording is also reflected in the characteristics of values of *NTone*, which by construction is bounded between -1 and 1. While the minimum value is -0.455 and thus relatively far from the highest possible minimum, the maximum value of 0.941 for BPW_O and 0.943 for BPW_N shows that in the most positive speeches hardly any negative words were used. This finding is additionally confirmed by the positivity of the 25% quartile and by the minimum number of one negative and eleven positive words.

Table 2.4
Descriptive Statistics for Sentiment Variables

Statistic	N	Mean	St. Dev.	Min	Max	Pctl. (25)	Pctl. (75)
COUNT_BPW_O	872	2,411.709	834.021	759	5,625	1,817.5	2,909
IND_NUM_BPW_O	872	1,153.603	334.053	433	2,402	920.8	1,331.5
IND_BPW_O	872	0.490	0.046	0.368	0.642	0.457	0.519
P_NUM_BPW_O	872	90.142	32.124	11	206	65	112
N_NUM_BPW_O	872	38.556	25.082	1	152	21	49
N_BPW_O	872	0.015	0.007	0.001	0.046	0.010	0.019
P_BPW_O	872	0.038	0.009	0.010	0.068	0.032	0.044
Tone_BPW_O	872	0.023	0.013	-0.029	0.062	0.014	0.032
NTone_BPW_O	872	0.428	0.241	-0.455	0.941	0.283	0.606
ITone_BPW_O	872	-0.023	0.013	-0.062	0.029	-0.032	-0.014
NToneSQ_BPW_O	872	0.241	0.188	0.000	0.886	0.083	0.367
COUNT_BPW_N	872	1,864.443	646.324	589	4,431	1,405	2,247.2
IND_NUM_BPW_N	872	1,098.989	326.592	399	2,323	873	1,277
IND_BPW_N	872	0.602	0.052	0.456	0.777	0.566	0.634
P_NUM_BPW_N	872	92.905	32.992	11	212	68	116
N_NUM_BPW_N	872	37.361	24.830	1	149	20	48
N_BPW_N	872	0.019	0.010	0.001	0.062	0.012	0.024
P_BPW_N	872	0.051	0.011	0.015	0.095	0.043	0.058
Tone_BPW_N	872	0.031	0.017	-0.039	0.090	0.020	0.043
NTone_BPW_N	872	0.454	0.238	-0.455	0.943	0.304	0.630
ITone_BPW_N	872	-0.031	0.017	-0.090	0.039	-0.043	-0.020
NToneSQ_BPW_N	872	0.263	0.195	0.000	0.889	0.095	0.396

We conducted a dependent-samples t-test to compare the alteration of positive and negative words found. There was a significant difference in the number of positive words found concerning the use of the BPW_O ($M=90.142, SD=32.124$) and BPW_N ($M=92.905, SD=32.992$), $t(871)=-22.939, p<.001$. This also applies to the number of negative words found when using the BPW_O ($M=38.556, SD=25.082$) and the BPW_N ($M=37.361, SD=24.830$), $t(871)=18.471, p<.001$.

Table 2.5 gives the descriptive statistics for the additional control variables used in the regression. In accordance with Bannier et al. (2017, p. 17), the number of observations in which the dividend per share is unchanged compared to the previous year is 31.1%. In 51.4% of the observations the dividend per share increased, and in 17.5% decreased.

Table 2.5
Descriptive Statistics for Control Variables

Statistic	N	Mean	St. Dev.	Min	Max	Pctl. (25)	Pctl. (75)
SIZE	870	9,883.827	16,996.830	30.200	104,226.900	845.245	10,287.470
M2B	869	2.208	2.267	-17.640	19.070	1.160	2.930
LEV	865	0.637	0.209	0.094	1.811	0.519	0.753
VOLA	872	0.020	0.010	0.002	0.130	0.014	0.024
VOL	852	2,108.435	4,949.786	0.100	47,270.600	67.925	1,518.850
ROA	865	0.037	0.065	-0.483	0.679	0.007	0.063
EPS_SP	848	1.685	16.275	-140.625	196.193	-1.607	2.625
DIV_SPP	872	0.514	0.500	0.000	1.000	0.000	1.000
DIV_SPN	872	0.175	0.381	0.000	1.000	0.000	0.000

Source: Author’s calculation based on data from Thomson Reuters Datastream.
Note: The definitions of all variables are given in the appendix (Section 2.7, Table 2.11).

Overall, editing stop words leads to a word reduction of 22.7% (477,216 words), as stated in Table 2.6. Deleting the 21 words from the stop word list that are also on the positive and negative list leads to 3.1% (2,409) more positive words found, with only eight more individual words. Although there are three more individual negative words, the number of negative words found decreases by 3.1% (1,042). This is because of the correction for “betrug” and “sorgen” described in the parsing process.

Table 2.6
Total Number of Words

	BPW_O	BPW_N
	All words	
Number of words	2,103,010	1,625,794
Individual words	100,151	99,970
	Positive words	
Number of words	78,604	81,013
Individual words	1,123	1,131
	Negative words	
Number of words	33,621	32,579
Individual words	2,180	2,183

Table 2.7 displays the number and cumulative fraction of the ten most frequent positive words in all speeches after correcting for stop words. The only difference is the deletion of the word “große” from the stop word list of the dictionary BPW_N.

Table 2.7
Ten most Frequent Positive Words

BPW_O			BPW_N		
Word	Number	cumulative %	Word	Number	cumulative %
erfolgreich	2,143	2.73%	erfolgreich	2,143	2.65%
erfolg	2,015	5.29%	erfolg	2,015	5.13%
erreicht	1,624	7.36%	erreicht	1,624	7.14%
erreichen	1,566	9.35%	erreichen	1,566	9.07%
großen	1,546	11.31%	großen	1,546	10.98%
besser	1,515	13.24%	besser	1,515	12.85%
positiv	1,157	14.71%	große	1,209	14.34%
stärker	1,089	16.10%	positiv	1,157	15,77%
positive	1,040	17.42%	stärker	1,089	17.11%
stärken	1,035	18.74%	positive	1,040	18.40%

As Table 2.8 illustrates, the adjustment in the parsing process for the words "betrug" and "sorgen" leads to an extensive decrease of those words, to the extent to which they do not appear in the ten most frequent negative words.

Table 2.8
Ten most Frequent Negative Words

BPW_O			BPW_N		
Word	Number	cumulative %	Word	Number	cumulative %
herausforderungen	1,019	3.03%	herausforderungen	1,019	3.13%
betrug	876	5.64%	krise	845	5.72%
krise	845	8.15%	schwierigen	792	8.15%
schwierigen	792	10.51%	rückgang	728	10.39%
rückgang	728	12.67%	gegen	650	12.38%
gegen	650	14.60%	minus	483	13.86%
minus	483	16.04%	verfügung	476	15.33%
verfügung	476	17.46%	wider	415	16.60%
wider	415	18.69%	leider	356	17.69%
sorgen	398	19.87%	finanzkrise	330	18.71%

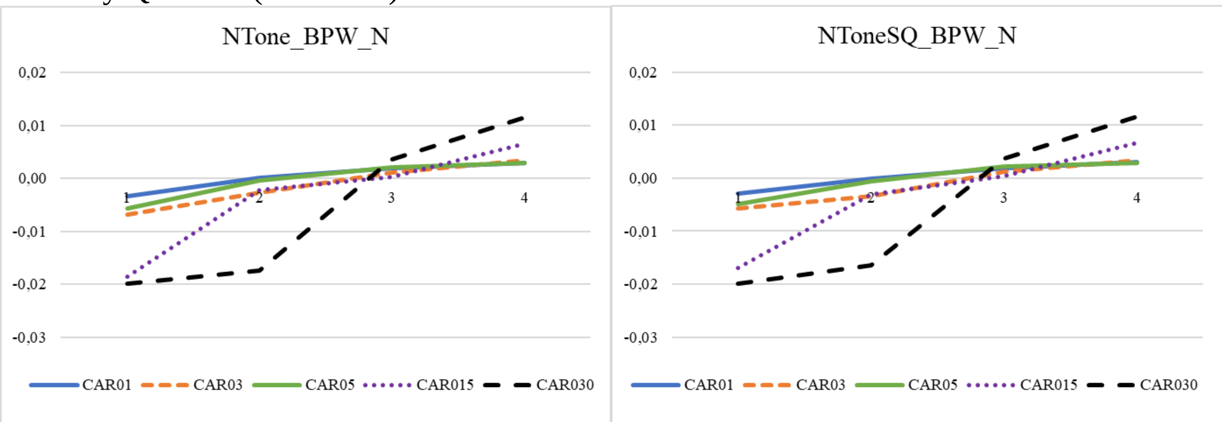
An English translation of all words listed in Table 2.7 and Table 2.8 is given in the appendix (section 2.7, Table 2.12). Note that an important distinction of German words through small and capital letters is not possible due to the nature of the parsing procedure and structure of the dictionaries. Because of their impact, we only considered this distinction for the words "betrug" and "sorgen."

Of the 2,223 (BPW_N: 2,849) positive words available, we only found 1,123 (BPW_N: 1,131) words. A comparably small fraction of those words found is able to account for 18.74% (BPW_N: 18.40%). The same applies to the more extensive list of 10,147 (BPW_N: 12,661) negative words. Of this list, we only found 2,180 (BPW_N: 2,183) words in the speeches, with ten words accounting for 19.87% (BPW_N: 18.71%) of all negative words found. These results clearly indicate that the correct words are more important than the mere extent of the used list.

2.5.2 Sentiment Measurements

Following Loughran and McDonald (2011, pp. 50f.), the assumption that the sentiment of certain texts is relevant leads in the case of CEO speeches to the assumption that speeches with a more positive measurement of sentiment lead to higher abnormal returns and higher abnormal trading volumes. By dividing all texts into quartiles based on the different sentiment measures[4] and analyzing the median CARs and CAVs, a visual examination can be conducted. Figure 2.1 gives the only two measurements that meet the stated assumptions. Using the sentiment measures *NTone* and *NToneSQ*, it is possible to have ascending quartile medians for all five event windows.

Figure 2.1
CARs by Quartiles (Sufficient)

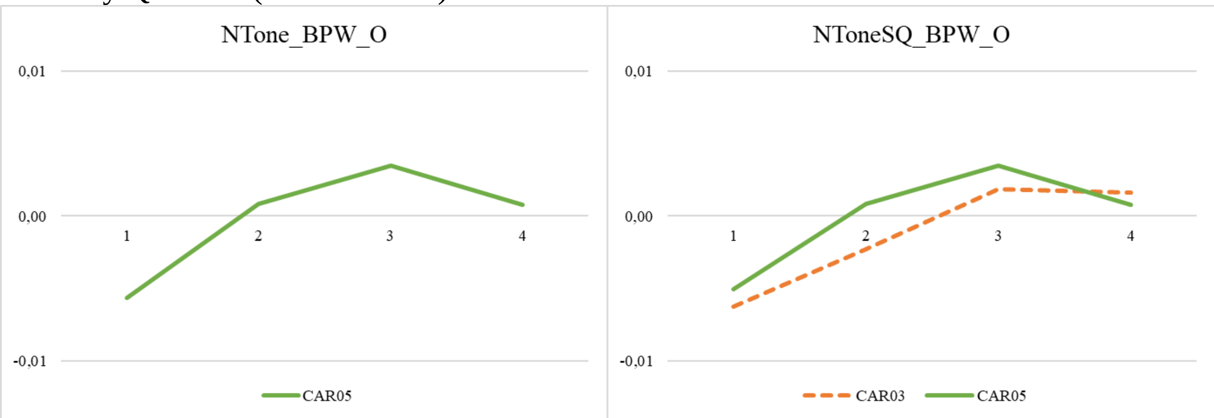


The equivalent measures for the BPW_O cannot provide comparable sufficient results for all analyzed event windows. The affected windows and the not sufficient results for the

[4] Note that only the share of negative words (*N*) was sorted in the descending order. All other sentiment measures are sorted in the ascending order.

associated quartiles are given in Figure 2.2. Here the window CAR [0,5] does not meet the assumptions for the sentiment measurement *NTone*. The same applies to the two windows CAR [0,3] and CAR [0,5] for *NToneSQ*. Other measurements of sentiment using the BPW_O or BPW_N do not meet this assumption either and therefore are not discussed further.

Figure 2.2
CARs by Quartiles (not Sufficient)



With regard to the visual examination of the CAVs for different sentiment measures, no measurement meets the above stated assumptions. Therefore, we excluded those figures.

Another essential assumption independent of certain event windows is the separation of above and below average abnormal returns through the use of sentiment measures as precisely as possible. Therefore, following Bannier et al. (2019a, pp. 17f., 37) and Price et al. (2012, pp. 1001f.), Figure 2.3 gives the average cumulative abnormal returns for up to 30 days following the AGM, divided by the above and below median sentiment measures *NTone* and *NToneSQ*. Additionally, the average CARs for all days are given[5].

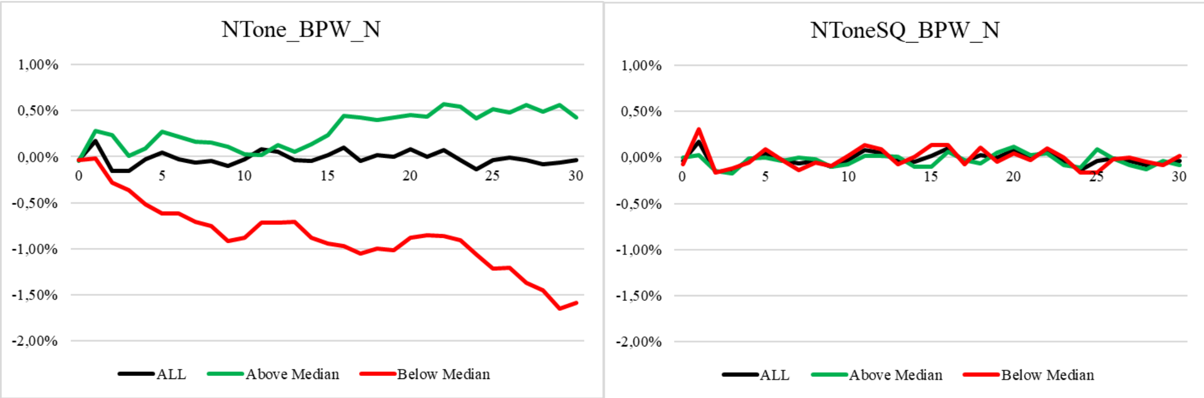
The accumulation of abnormal returns in Figure 2.3 for up to 30 days following the AGM shows that the average CARs are close to zero. By dividing the different observations into above and below median *NTone*, it is possible to separate positive and negative CARs. This is in accordance with the results of Bannier et al. (2019a, pp. 17f., 37). This separation can only be conducted using *NTone*. The same analysis using *NToneSQ* allows no distinction of positive and negative CARs using above and below median *NToneSQ*.

It therefore can be stated as an interim result that only the usage of the reformed and extended BPW_N dictionary with *NTone* as a sentiment measure is able to meet one of the

[5] Due to the results stated in Figure 2.1 and Figure 2.2, only the results for *NTone* and *NToneSQ* calculated using the BPW_N are given.

central assumptions stated in the pioneer paper by Loughran and McDonald (2011, pp. 50f.) and the additional assumption of distinction.

Figure 2.3
CARs over time



2.5.3 Significance of Results

Based on the preceding results, this section examines the relation between *NTone* and CARs for different event windows in a multivariate context using the control variables that we described above. Table 2.9 reports the regression results for *NTone* using the BPW_N and the five different event windows for CARs.

The results show a high statistical significance of the coefficient of the sentiment measurement *NTone* that we calculated using the BPW_N and the five different CARs as dependent variables. Thus, more positive speeches of CEOs can be associated with higher abnormal returns. An increase in *NTone* by the interquartile change of 0.326 leads to a minor increase of 0.42% in CAR [0,1], but a major increase of 1.53% in CAR [0,30]. This role as a key factor in the market reaction to AGMs becomes more interesting, when other variables, based on the performance or the dividend policy are considered. The ROA negatively relates to all five event windows and is only significant for the first three windows. The dividend surprise can only partially account for the significance of the longer event windows. We could verify only a significant association with individual event windows for the analyzed control variables. None of the variables are able to explain all windows.

Regarding the significant relation of *NTone* as a relative measurement of sentiment and short- and long-term event windows, the results are consistent with Price et al. (2012, pp. 1004f.) and Bannier et al. (2017, p. 37, 2019a, p. 34).

Table 2.9
Regression of NTone_BPW_N and CARs

	Dependent variable:				
	CAR01	CAR03	CAR05	CAR015	CAR030
	(1)	(2)	(3)	(4)	(5)
NTone_BPW_N	0.014***	0.018***	0.018***	0.035***	0.064***
	(0.005)	(0.006)	(0.007)	(0.011)	(0.017)
LN_COUNT_BPW_N	0.009**	0.004	0.011**	0.011	0.014
	(0.004)	(0.005)	(0.005)	(0.008)	(0.012)
IND_BPW_N	0.071***	0.045	0.041	0.042	0.056
	(0.027)	(0.032)	(0.036)	(0.055)	(0.082)
LN_SIZE	0.001	0.001	0.002	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
M2B	-0.0002	-0.0003	0.0004	0.001	0.001
	(0.0004)	(0.0005)	(0.001)	(0.001)	(0.002)
LEV	-0.002	-0.006	-0.007	-0.003	-0.007
	(0.005)	(0.006)	(0.007)	(0.010)	(0.016)
VOLA	0.028	-0.091	-0.144	-0.679**	-1.162**
	(0.201)	(0.247)	(0.210)	(0.326)	(0.499)
LN_VOL	-0.001	-0.001*	-0.001	-0.001	0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
ROA	-0.044**	-0.080***	-0.084***	-0.035	-0.054
	(0.021)	(0.024)	(0.025)	(0.039)	(0.063)
EPS_SP	0.00002	0.00001	0.00002	-0.0001	0.0004
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
DIV_SPP	-0.0002	0.003	0.007**	0.007	0.018***
	(0.002)	(0.002)	(0.003)	(0.005)	(0.007)
DIV_SPN	-0.003	-0.002	0.002	-0.003	-0.022**
	(0.003)	(0.004)	(0.004)	(0.007)	(0.010)
Constant	-0.119**	-0.064	-0.115*	-0.128	-0.152
	(0.046)	(0.052)	(0.062)	(0.097)	(0.141)
Observations	829	829	829	829	829
Year Fixed Effects	YES	YES	YES	YES	YES
R2	0.032	0.050	0.053	0.073	0.121
Adjusted R2	0.004	0.022	0.026	0.046	0.095
Residual Std. Error (df = 805)	0.026	0.031	0.036	0.057	0.082
F Statistic (df = 23; 805)	1.149	1.826**	1.977***	2.747***	4.800***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by * p < 0.1; ** p < 0.05; *** p < 0.01.

Despite the insufficient fulfillment of the assumption that speeches with a more positive measurement of sentiment lead to higher abnormal returns for *NTone* using the BPW_O, Table 2.10 shows that the positive relation between this measurement and the different CARs is almost as significant as the usage of BPW_N.

Table 2.10
Regression of NTone_BPW_O and CARs

	Dependent variable:				
	CAR01	CAR03	CAR05	CAR015	CAR030
	(1)	(2)	(3)	(4)	(5)
NTone_BPW_O	0.012**	0.016***	0.017**	0.034***	0.062***
	(0.005)	(0.006)	(0.007)	(0.011)	(0.017)
LN_COUNT_BPW_O	0.008**	0.003	0.010*	0.009	0.011
	(0.004)	(0.005)	(0.005)	(0.009)	(0.012)
IND_BPW_O	0.075**	0.036	0.037	0.022	0.032
	(0.031)	(0.037)	(0.042)	(0.064)	(0.097)
LN_SIZE	0.001	0.001	0.002	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
M2B	-0.0001	-0.0003	0.0004	0.001	0.001
	(0.0004)	(0.0005)	(0.001)	(0.001)	(0.002)
LEV	-0.002	-0.006	-0.007	-0.003	-0.007
	(0.005)	(0.006)	(0.007)	(0.010)	(0.016)
VOLA	0.022	-0.098	-0.149	-0.688**	-1.172**
	(0.202)	(0.248)	(0.210)	(0.324)	(0.496)
LN_VOL	-0.001	-0.001*	-0.001	-0.001	0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
ROA	-0.044**	-0.080***	-0.083***	-0.035	-0.054
	(0.021)	(0.024)	(0.025)	(0.039)	(0.063)
EPS_SP	0.00002	0.00002	0.00002	-0.0001	0.0004
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0003)
DIV_SPP	-0.00001	0.003	0.007**	0.008	0.018***
	(0.002)	(0.002)	(0.003)	(0.005)	(0.007)
DIV_SPN	-0.003	-0.002	0.002	-0.003	-0.022**
	(0.003)	(0.004)	(0.004)	(0.007)	(0.010)
Constant	-0.110**	-0.043	-0.104	-0.096	-0.111
	(0.048)	(0.054)	(0.065)	(0.098)	(0.145)
Observations	829	829	829	829	829
Year Fixed Effects	YES	YES	YES	YES	YES
R2	0.029	0.047	0.052	0.072	0.120
Adjusted R2	0.001	0.020	0.025	0.046	0.095
Residual Std. Error (df = 805)	0.026	0.031	0.036	0.057	0.082
F Statistic (df = 23; 805)	1.042	1.736**	1.937***	2.721***	4.790***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by * p < 0.1; ** p < 0.05; *** p < 0.01.

Only for the event windows CAR [0,1] and CAR [0,5], the coefficient is significant at a 5% level. Due to the smaller interquartile change of 0.323, a change in *NTone* by this change leads to a 0.39% higher CAR [0,1] and a 1.45% higher CAR [0,30]. Interestingly, these results show higher significance than Bannier et al. (2019a, p. 34), where maximum significance at the 5% level was achieved (CAR [0,30]: 10%).

Based on the already stated results for the necessary assumptions of the cumulative abnormal trading volumes under 2.5.2, we will not discuss those regressions further.

2.6 Conclusions

This paper focuses on textual analysis as an important part of accounting and finance research using the dictionary-based approach with the first available finance-related dictionary for the German language (BPW_O). Due to the novelty of this dictionary, the aim of this paper is to propose several reforms and extensions (BPW_N) to improve its performance and to find the most appropriate measurement of sentiment.

Based on the visual examination of the two central assumptions that speeches with a more positive measurement of sentiment lead to higher abnormal returns and that it is possible to separate above and below average abnormal returns through the use of sentiment measures, the use of the measurement *NTone* calculated using the BPW_N should be preferred. Additionally, we were able to supplement the significance of these results by several regressions. Here the use of *NTone*, calculated by using the BPW_N, could provide highly statistically significant results for all five analyzed event windows. Thus, more positive speeches of CEOs can be associated with higher abnormal returns following the Annual General Meeting. Based on the event window, an increase in *NTone* by the interquartile change of 0.326 leads to an increase in cumulative abnormal returns ranging from 0.42% (CAR [0,1]) to 1.53% (CAR [0,30]).

Using the most comprehensive collection of German CEO speeches so far, this paper is able to give two contributions to the literature on textual analysis of German texts. Through implementing reforms and extensions, we improved the results of the original BPW_O and confirmed the stated hypothesis. Additionally, the combination of the BPW_N and the relative measurement of sentiment *NTone* has proven to be the most suitable one for measuring business texts and therefore answers the additional research question.

Due to the results of the proposed adjustments on the newly developed BPW_O, additional improvements should be considered and tested. Moreover, this new version of the BPW (BPW_N) should be compared to old and new versions of general German dictionaries.

As there is a wide range of publicly available textual data, the BPW_N should be used to analyze other types of corporate disclosures.

2.7 Appendix

Table 2.11
Description of Variables

Variable	Description
SIZE	Firm Size: Daily market value
M2B	Market to Book Value: Ratio of the market value of the ordinary (common) equity to the balance sheet value of the ordinary (common) equity
LEV	Leverage: Ratio of the total liabilities to the total assets
VOLA	Volatility: Standard deviation of the daily returns for the ninety trading-day window ending ten days prior to the AGM
VOL	Volume: Number of shares traded on the day of the AGM
COUNT	Total number of Words. Due to different stop word lists calculated individually for BPW_O and BPW_N
IND_NUM	Number of individual words. Due to different stop word lists calculated individually for BPW_O and BPW_N.
IND	Individual Words: *IND_NUM* divided by *COUNT*
ROA	Return on Assets: Net income divided by total assets
EPS_SP	Earnings Surprise: Calculated according to Bannier et al., 2017: The difference between the last reported earnings per share at time *t* minus the latest reported earnings per share in the year prior to date *t*, divided by the stock price one year before *t* times 100 $EPS_{SP}=\frac{EPS_t-EPS_{t-1}}{Price_{t-1}}\cdot 100$
DIV_SPP	Dividend Surprise Positive: Calculated according to Bannier et al., 2017: *DIV_SPP* equals one if the dividend per share is increased compared to the previous year, zero otherwise
DIV_SPN	Dividend Surprise Negative: Calculated according to Bannier et al., 2017: *DIV_SPN* equals one if the dividend per share is decreased compared to the previous year, zero otherwise
P_NUM	Number of positive words
N_NUM	Number of negative words

Table 2.12
Translation of ten most Frequent Words

positive words		negative words	
German	English	German	English
besser	better	betrug	fraud, amounted
erfolg	success	finanzkrise	financial crisis
erfolgreich	successful	gegen	against
erreichen	achieve	herausforderungen	challenges
erreicht	achieved	krise	crisis
große	large	leider	unfortunately
großen	large	minus	minus
positiv	positive	rückgang	decline
positive	positive	schwierigen	difficult
stärken	strenghten	sorgen	sorrow, care
stärker	stronger	verfügung	decree
		wider	against

Note that the listed translations represent only one of several possibilities. Due to the nature of the parsing procedure and structure of the dictionaries, an important distinction of German words through small and capital letters is not possible.

3. Using Negations in Analyzing German Texts in Finance

[6]

Abstract

Domain-specific dictionaries have prevailed, when conducting the dictionary-based approach to measure the sentiment of textual data in finance. Through the contributions of Bannier et al. (2019b) and Pöferlein (2021), two versions of a dictionary suitable for analyzing German finance-related texts are available (BPW dictionary). This paper conducts and tests further improvements of the given word lists by calculating the sentiment of German-speaking annual reports to forecast future return on assets and future return on equity. This corrected and expanded version provides more significant results. Despite the broad usage of negations, this type of improvement in combination with the BPW dictionary has not yet been tested when conducting the dictionary-based approach. Therefore, this paper additionally tests different negation lists to show that implementing negations can improve results.

[6] This chapter has been published as: *Using Negations in Analyzing German Texts in Finance* in *Credit and Capital Markets*, Vol. (2024), Online First, pp. 1–36

3.1 Introduction

Public companies use annual reports as a tool of external communication with investors. Investors use these reports as a basis for their investment decisions. In addition to business figures, these reports contain a large amount of text, which is purely qualitative information. By using methods of textual analysis, the quantitative information encoded in these texts can be obtained and further processed. Therefore, obtaining annual reports' sentiment to prove correlations with financial ratios or share prices, represents an established field in accounting and finance research (Chakraborty & Bhattacharjee, 2020, p. 767; T. Kang et al., 2018, pp. 370f.; Kearney & Liu, 2014, p. 173; Loughran & McDonald, 2011, p. 35). We focus our paper on the two variables future return on assets ($FROA$) and future return on equity ($FROE$) one year ahead, which are frequently used as an independent performance measure in relevant studies (Daniel et al., 2004, pp. 568–570; King et al., 2004, p. 191; Koelbl, 2020, p. 194; Myšková & Hájek, 2020, p. 1428; Vojinović et al., 2020, p. 136).

Algaba et al. (2020, p. 514) define that “sentiment is the disposition of an entity toward an entity, expressed via a certain medium.” The specified disposition can be conveyed quantitatively through numbers, although it is primarily expressed qualitatively, using text, audio, or visual media (Algaba et al., 2020, p. 514). This sentiment provides a measure of the degree of positivity or negativity and can potentially offer an additional perspective in the process of stock price formation. As a result, it can help address key questions in the field of behavioral finance (Kearney & Liu, 2014, p. 172).

The two most common textual analysis methods for obtaining sentiment from qualitative data are the dictionary-based approach (or bag-of-words) and machine learning (Chakraborty & Bhattacharjee, 2020, p. 762; Kearney & Liu, 2014, pp. 174f.). Using a mapping algorithm, the dictionary-based approach utilizes predefined word lists to assign words into positive, negative, or other sentiment categories like uncertainty. By counting these classified words, several measurements of sentiment can be calculated (F. Li, 2010, p. 146; Loughran & McDonald, 2015, p. 1; Rice & Zorn, 2019, p. 2). The machine learning approach uses a subset of linguistic labeled texts to train complex models. These models are then used to predict the sentiment of a given set of texts (Rice & Zorn, 2019, p. 1; Shapiro et al., 2022, p. 224). Contributions like Frankel et al. (2022, p. 5514) and Mishev et al. (2020, p. 131677) show that to measure the sentiment of financial text, machine learning approaches can be superior. However, this advantage has the additional disadvantage that machine learning approaches are often a black-box and are therefore almost unreplicable and difficult to explain (Algaba et al., 2020, p. 523; Krause et al., 2016, p. 5686). To prevent these challenges and provide a replicable approach for future research, this paper focuses on the dictionary-based approach.

When using the dictionary-based approach, domain-specific dictionaries have proven to be superior and prevailed in analyzing financial texts (Y. Kang et al., 2020, p. 149; Kearney & Liu, 2014, p. 177; Loughran & McDonald, 2015, p. 1; Luo & Zhou, 2020, p. 107; Shapiro et al., 2022, pp. 223, 227). The newly developed finance word lists by Bannier et al. (2019b) (BPW_O) have been improved by Pöferlein (2021) (BPW_N). Due to the novelty of those dictionaries, the first hypothesis of this paper is that further correcting and expanding the BPW_N dictionary, to get an expanded BPW_E dictionary, improves the results of forecasting future ROAs and ROEs from the sentiment of annual reports. One possible improvement that has not yet been tested in the context of the BPW word lists is the implementation of negations. Due to their potential high impact and widespread usage (Bochkay et al., 2020, p. 43; Borochin et al., 2018, p. 80; Loughran & McDonald, 2011, p. 44; Shapiro et al., 2022, 228f.), the additional hypothesis of this paper is that accounting for negations additionally improves results.

The contribution of this paper to the literature on analyzing German-speaking financial texts is the further extension and optimization of the edited version of the BPW dictionary. Additionally, this paper is the first contribution using different negations combined with the two versions of the BPW dictionary. Therefore, future research in analyzing the sentiment of German-speaking texts in finance can be conducted more precisely.

This paper proceeds as follows. In the second part, we provide a short review of the relevant literature on textual analysis, focusing on analyzing financial texts with and without using negations. The third section presents the data and the applied parsing procedure, in addition to the usage and creation of the dictionaries. The fourth section highlights the empirical approach used to obtain the results presented in section five. Lastly, the sixth section concludes.

3.2 Literature Review

Several contributions like Chakraborty and Bhattacharjee (2020), Kearney and Liu (2014), and Luo and Zhou (2020) provide an excellent overview of the extensive field of textual analysis in finance. Moreover, certain overview papers provide additional information about specific areas of caution (Algaba et al., 2020; Loughran & McDonald, 2016) and ideas for future research (Kaya et al., 2020). Due to the above-mentioned reasons this paper and therefore the following literature review focuses on the dictionary-based approach.

One of the first steps in measuring the sentiment of a text is selecting a dictionary or word list (Loughran & McDonald, 2015, p. 1). According to Loughran and McDonald (2016, p. 1200), four different word lists have been primarily used by researchers in classifying English finance-related texts. These can be divided into two general dictionaries, namely “General Inquirer” (Stone et al., 1966) and “DICTION” (Hart, 2000), and two word lists generated for finance-related texts by Henry (2006, 2008) and Loughran and McDonald (2011).

Through the contributions of Henry (2006, 2008) and Loughran and McDonald (2011), the usage of general word lists for different forms of finance-related textual content like news (Tetlock, 2007; Tetlock et al., 2008), earnings press releases (Davis et al., 2012; Davis & Tama-Sweet, 2012) or annual reports (Feldman et al., 2008; Yuthas et al., 2002) was widely criticized in favor of domain-specific word lists (Algaba et al., 2020, p. 523; Chakraborty & Bhattacharjee, 2020, p. 764; Lewis & Young, 2019, pp. 598f.; Loughran & McDonald, 2015, pp. 2f.; Mishev et al., 2020, p. 131677; Price et al., 2012, p. 1006).

In the field of finance, the word lists provided by Loughran and McDonald are primarily used (Kearney & Liu, 2014, p. 175; Loughran & McDonald, 2016, pp. 1204–1206), for different kinds of finance-related textual data. These lists were used to analyze news (Ferguson et al., 2015; Hillert et al., 2018), conference calls (Da Tonin & Scherer, 2022; Druz et al., 2020), and annual reports (Berns et al., 2022; T. Kang et al., 2018).

The above-mentioned domain-specific problems regarding the German language were also present. Research was primarily limited to general dictionaries like SentiWS (Remus et al., 2010) and LIWC (Meier et al., 2018; Wolf et al., 2008). In order to rectify this problem, Bannier et al. (2019b) introduced a German domain-specific dictionary in the field of finance. After the usage of the original word lists in different contributions (Bannier et al., 2017, 2019a; Röder & Walter, 2019; Tillmann & Walter, 2018, 2019), a reformed and extended version was introduced by Pöferlein (2021).

An essential element in the approach introduced by Loughran and McDonald (2011) is the use of negations. They account for simple negations for their list of positive words using the six negations “no, not, none, neither, never, nobody” occurring within three words preceding a positive word (Loughran & McDonald, 2011, p. 44). In accordance with the work of Loughran and McDonald, negations are widely used in the textual analysis of business texts. These are either used in the form proposed by Loughran and McDonald (Huang et al., 2014, p. 1089; Renault, 2017, p. 29), as an extended version of the six negations (Borochin et al., 2018, p. 80;

Brau et al., 2016, p. 5; Correa et al., 2021, p. 94) or in other forms (Jandl et al., 2014, p. 5; Jegadeesh & Wu, 2013, p. 716).

Despite having the contribution by Loughran and McDonald (2011) as a theoretical foundation (Bannier et al., 2019b, p. 79), Bannier et al. (2017, 2019a, 2019b) and other authors using the BPW have not yet accounted for negations in their papers (Pöferlein, 2021; Röder & Walter, 2019; Tillmann & Walter, 2018, 2019).

3.3 Data

3.3.1 Data Source

We get the initial sample of relevant companies and all the financial variables from the Amadeus database provided by Bureau van Dijk. Hereby we focus on stock-listed companies from three German-speaking countries, Austria, Germany, and Switzerland. Additionally, we only select companies with available reports for at least one year between 2010 and 2020. From the initial sample of 893 companies, 740 companies published at least one annual report on their web page. We were able to find and manually download 6,275 annual reports[7]. Table 3.1 provides an overview of the Amadeus search strategy, and the following sample creation. We obtained all other variables from Amadeus.

Table 3.1
Sample Creation

Source / Filter	Sample Size
Active companies in Amadeus	3,105,008
Country: Austria, Germany, Switzerland	480,282
Stock listed companies	10,738
At least one available annual report in the years 2010 to 2020	893
Company with annual report available on Homepage	740
Final sample of annual reports	6,275

3.3.2 Used Dictionaries

We use the BPW_N dictionary proposed by Pöferlein (2021) to analyze the annual reports. These word lists also build the foundation for constructing the BPW_E word lists. Additionally, we use the original word lists by Bannier et al. (2019b) (BPW_O) to compare results.

[7] 620 annual reports have a different fiscal year. Due to available data in Amadeus those reports weren't removed from the sample.

To get the extended version of the BPW_N (BPW_E) we manually check all word lists and delete words with a different or ambiguous meaning (e.g. “prolongiert” (English: prolonged) on the negative word list). During the review of all three relevant lists, we deleted 22 words on the positive list, 141 words on the negative list and 259 words on the stop words list.

In order to find missing words in all three word lists, we use the German news corpus 2020 from Universität Leipzig (2022) to check every word for missing basic forms and variations. Additionally, we account for synonyms, their basic forms, and variations. We manually check all words found for their plausibility regarding the different word lists. Out of the 35,254 basic forms found, we add 1,911 positive, 3,157 negative, and 779 stop words. Through the 17,630 synonyms found, we are able to add another 746 positive, 2,389 negative, and 85 stop words. Finally, we add an alternative spelling of mutated vowels according to Pöferlein (2021, p. 8). A summary of the conducted steps and the resulting alteration of the three word lists is presented in Table 3.2.

Table 3.2
Updating the BPW_N Dictionaries

	Positive	Negative	Stop words
BPW_O total words	2,223	10,147	3,682
BPW_N total words	2,849	12,661	4,132
Delete words with a different meaning	- 22	- 141	- 259
Adding basic forms	+ 1,911	+ 3,157	+ 779
Adding synonyms	+ 746	+ 2.389	+ 85
Adding mutated vowels	+ 692	+ 1,336	+ 84
BPW_E total words	6,176	19,402	4,821

We use four different lists of negations. Firstly, we obtain the two German lists of the Linguistic Inquiry and Word Count LIWC2001 and LIWC2015 in their original form (Meier et al., 2018; Wolf et al., 2008), containing 13 and 39 negations. Additionally, we generate two own lists based on the six negations given by Loughran and McDonald (2011, p. 44)[8]. Furthermore, we account for the criticism of Picault and Renault (2017, p. 139) by adding the word “lower”, resulting in seven negations[9]. To obtain the German version of these two lists, we screen 30 corresponding annual statements of the DAX companies in 2017 for the negations

[8] Negation list LMD.
[9] Negation list PR.

given by Loughran and McDonald (2011) and Picault and Renault (2017) and their matching German translations. This approach is based on Bannier et al. (2019b, pp. 94, 102) where they evaluated their dictionary by using corresponding German and English quarterly and annual reports from DAX and MDAX companies. Overall, we find 8,063 translations of the Loughran and McDonald negations, resulting in 25 individual negations. Due to the additional word "lower", 9,201 translations can be found for the Picault and Renault negations, resulting in 316 individual negations (including mutated vowels).

We apply the above-described approach of obtaining the extended version of the three word lists to the four negation lists resulting in 26 LIWC2001, 49 LIWC2015, 28 LMD, and 916 PR negations. Altogether we manually check 2,525 basic forms and 1,151 synonyms for their plausibility. Finally, we add the above used alternative spelling of mutated vowels. Table 3.3 summarizes all steps and the resulting alterations.

Table 3.3
Creating and Updating Negations

	LIWC 2001	LIWC 2015	LMD	PR
Basic form / Translation (BPW_N)	13	39	25	316
Delete words with a different meaning				- 6
Adding basic forms	+ 12	+ 8	+ 1	+ 397
Adding synonyms	+ 1	+ 2	+ 2	+ 84
Adding mutated vowels				+ 125
BPW_E total words	26	49	28	916

3.3.3 Parsing

Based on the criticism of Loughran and McDonald (2015, p. 2), we follow Pöferlein (2021, p. 9) in giving a detailed overview of performed text manipulation. Owing to this approach, difficulties in replicating this study due to unspecified parsing rules are avoided.

First and foremost, we convert the manually collected PDFs to UTF-8 encoded TXT files (Bannier et al., 2017, p. 10, 2019a, p. 9; Y. Kang et al., 2020, p. 157; Meier et al., 2018, p. 29). We conduct the following parsing procedure in accordance with Pöferlein (2021, p. 9) using an automated parser programmed in Python. We replace typographic ligatures (Bannier et al., 2017, p. 10, 2019a, p. 9), hyphens (Loughran & McDonald, 2011, internet appendix), and convert all words to lowercase (Pengnate et al., 2020, p. 193; Picault & Renault, 2017, p. 139; Tillmann & Walter, 2018, p. 8). Furthermore, we remove irrelevant content in the form of special characters (Allee & Deangelis, 2015, p. 247; Fritz & Tows, 2018, p. 61), numbers (Ferris et al., 2013, p. 998; Gentzkow et al., 2019, p. 536), punctuation (Iqbal & Riaz, 2022, p. 2702; Picault & Renault, 2017, p. 139), and multiple whitespaces (González et al., 2019, p. 433; Schmeling & Wagner, 2016, p. 8). Eventually, we follow Bannier et al. (2017, p. 10, 2019a, pp. 9f.) and delete all words with less than three characters. Depending on the dictionary, we use the associated stop word list (BPW_O, BWP_N or BPW_E).

Following Pöferlein (2021, p. 9), we include an automated alteration of the words "betrug" and "sorgen" prior to the parsing procedure when using the BPW_N word lists. Additionally, when using the BPW_E, we add the word "bremse" from the BPW_N word list and the two words "stahl" and "sucht" from the BPW_E dictionary to the automated alteration. When written in lowercase the words "betrug", "sorgen" and "sucht" are changed to "betrugnoneg", "sorgennoneg" and "suchtnoneg". Additionally, the words "bremse" and "stahl" are changed to "bremsenoneg" and "stahlnoneg" when written with a first capital letter. These alterations are due to the change in meaning of certain words when written with a first capital or lowercase letter. Due to peculiarities of the German language, in addition to the approach of Pöferlein (2021, p. 9), occurrences of the word "betrug" at the beginning of a sentence are changed to "betrugnoneg". Table 3.4 displays an overview of these different meanings. Due to this pre parsing procedure, we are able to additionally reduce the stated exaggeration of negative words in Pöferlein (2021, p. 9).

Table 3.4
Differences Between Capital and Lowercase Letters

Words with a first capital letter	Translation	Words with a first lowercase letter	Translation
Betrug	fraud	**betrug**	amounted
Bremse	brake	bremse	slow down
Sorgen	sorrow	**sorgen**	care
Stahl	steel	stahl	steal
Sucht	addiction	**sucht**	search

Note: German words altered using the suffix "noneg" are bold.

3.4 Methodology

3.4.1 Measurement of Sentiment and Implementation of Negations

We use Python to count the occurrence of positive (p) and negative (n) words from each of the three dictionaries. We use the relative measurement of NetTone ($NTone$), which is the most common measurement regarding the BPW-Dictionary (Bannier et al., 2017, p. 11, 2019a, p. 10; Tillmann & Walter, 2018, p. 9) and has proven to be superior to other measurements (Pöferlein, 2021, p. 20). This measurement solely focuses on the number of positive and negative words and is not altered by the length of analyzed documents:

$$NTone = \frac{p - n}{p + n} \tag{3.1}$$

In the existing literature, negations are considered in two different ways. In order to provide a fully comprehensive analysis of the influence of negations, this paper uses both approaches. We follow Druz et al. (2020, pp. 51f.), Loughran and McDonald (2011, p. 44), and Shapiro et al. (2022, p. 228) in counting words as negated if there is a negation among the three preceding words. In handling negated words, we use two different approaches. In accordance with Bushman et al. (2016, p. 783) and Druz et al. (2020, p. 51), negated words are not counted. Measurements using this approach are marked with the suffix "_ig" (for ignore). Additionally, the more common approach of handling negations is term shifting (Algaba et al., 2020, p. 524; Bochkay et al., 2020, p. 43; Jandl et al., 2014, p. 5; Taboada et al., 2011, p. 276). Here the negated word is counted as a word from the opposite dictionary. Measurements using this approach are marked with the suffix "_ts" (for term shifting). Depending on the respective dictionaries, the corresponding negation lists are used.

Following Bannier et al. (2017, p. 15), Davis et al. (2015, p. 653), and Pöferlein (2021, p. 10), all words found are weighted equally. Due to this, other researchers can replicate and further develop the results of this paper. Henry and Leone (2016, p. 166) also support this approach and the superiority of equal weighting.

3.4.2 Empirical Approach

The most common approach for measuring the impact of sentiment on future profitability using a bag-of-words model is linear regression (Bannier et al., 2019a, p. 13; Boudt & Thewissen, 2019, p. 103; Henry et al., 2021, p. 8; Patelli & Pedrini, 2014, p. 29). Therefore, we apply the following linear regression model using two different dependent variables:

$$Dep_j = \alpha_0 + \alpha_1 NTone_j + \sum_{k=1}^{K} \alpha_k Control_{kj} + \varepsilon_j \tag{3.2}$$

Dep represents the two different variables, future return on assets (*FROA*) and future return on equity (*FROE*) one year ahead. Both variables are used frequently as an independent performance measure (Daniel et al., 2004, pp. 568–570; King et al., 2004, p. 191; Koelbl, 2020, p. 194), even though ROA is considered to be more accurate and less influenced by accounting (Myšková & Hájek, 2020, p. 1428; Vojinović et al., 2020, p. 136).

We use five different control variables (*Control*) as well as year and industry fixed effects based on relevant research findings (Alshorman & Shanahan, 2022, p. 132; Aly et al., 2018, p. 66; Boudt & Thewissen, 2019, pp. 103, 110; Davis & Tama-Sweet, 2012, pp. 814f., 827; González et al., 2019, p. 442; T. Kang et al., 2018, p. 375). These include the age of the company (*AGE*), a dummy variable to identify loss firms (*LOSS*), the leverage (*LEV*), the current return on assets (*ROA*) and the current return on equity (*ROE*). When using *FROA* as a dependent variable *ROE* is excluded from the regression. The same applies for using *FROE* and *ROA*. The calculation of all variables can be found in the appendix (section 3.7, Table 3.16).

3.5 Results

According to Loughran and McDonald (2011, p. 39), we exclude annual reports with less than 2,000 words from the sample. Additionally, we eliminate reports with less than 200 individual words to remove corrupted data. Due to different stop word lists connected with the particular dictionaries, the numbers of excluded reports and, therefore, the numbers of analyzed annual reports vary. A possible alternative of considering the following analyses on a uniform data sample is not carried out, as this contradicts the general basic logic of using different dictionaries.

3.5.1 Summary Statistics

The following three tables report the summary statistics for all three dictionaries used. Table 3.5 provides descriptive statistics for all variables used to analyze the original dictionary by Bannier et al. (2019b) (BPW_O). It can be observed that the future and present return variables have a high standard deviation, with values ranging from highly negative to highly positive.

Table 3.5
Descriptive Statistics for BPW_O Variables (N = 4,168)

Statistic	Mean	St. Dev.	Min	Max	Pctl. (25)	Pctl. (75)
FROA	4.303	11.967	-93.678	100.000	1.367	8.676
FROE	6.592	43.348	-783.269	372.161	3.029	18.954
NTone	-0.075	0.183	-0.750	0.703	-0.195	0.035
AGE	49.641	45.224	0.000	555.000	16.000	92.000
LOSS	0.178	0.383	0.000	1.000	0.000	0.000
LEV	1.648	3.015	0.000	111.411	0.581	1.881
ROA	4.534	11.764	-91.969	90.525	1.596	8.856
ROE	9.089	40.780	-783.269	924.023	3.639	19.430

As shown in Table 3.6, the mean *NTone* using BPW_N slightly increases, while the standard deviation and minimum values remain the same. Additionally, the maximum value slightly decreases. The additional usage of negations leads to higher values of *NTone*, where using a combination of PR negations and term shifting creates a positive mean.

Table 3.6
Descriptive Statistics for BPW_N Variables (N = 4,112)

Statistic	Mean	St. Dev.	Min	Max	Pctl. (25)	Pctl. (75)
FROA	4.309	11.928	-93.678	100.000	1.383	8.684
FROE	6.659	43.503	-783.269	372.161	3.064	19.035
NTone	-0.051	0.183	-0.750	0.696	-0.172	0.062
NTone_LIWC01_ig	-0.043	0.185	-0.745	0.711	-0.165	0.070
NTone_LIWC15_ig	-0.032	0.186	-0.733	0.730	-0.156	0.081
NTone_LMD_ig	-0.039	0.184	-0.739	0.702	-0.161	0.074
NTone_PR_ig	-0.024	0.188	-0.733	0.723	-0.146	0.092
NTone_LIWC01_ts	-0.031	0.177	-0.708	0.708	-0.147	0.075
NTone_LIWC15_ts	-0.008	0.172	-0.630	0.719	-0.122	0.096
NTone_LMD_ts	-0.024	0.178	-0.679	0.698	-0.143	0.085
NTone_PR_ts	0.010	0.165	-0.630	0.673	-0.093	0.112
AGE	49.803	45.309	0.000	555.000	16.000	92.750
LOSS	0.177	0.382	0.000	1.000	0.000	0.000
LEV	1.663	3.032	0.000	111.411	0.593	1.898
ROA	4.545	11.737	-91.969	90.525	1.606	8.874
ROE	9.163	40.985	-783.269	924.023	3.738	19.498

Table 3.7
Descriptive Statistics for BPW_E Variables (N = 4,116)

Statistic	Mean	St. Dev.	Min	Max	Pctl. (25)	Pctl. (75)
FROA	4.310	11.922	-93.678	100.000	1.384	8.680
FROE	6.660	43.482	-783.269	372.161	3.068	19.015
NTone	0.139	0.141	-0.558	0.740	0.046	0.227
NTone_LIWC01_ig	0.148	0.143	-0.553	0.764	0.053	0.239
NTone_LIWC15_ig	0.149	0.144	-0.553	0.764	0.054	0.240
NTone_LMD_ig	0.146	0.142	-0.553	0.756	0.053	0.234
NTone_PR_ig	0.154	0.146	-0.554	0.781	0.058	0.247
NTone_LIWC01_ts	0.147	0.137	-0.484	0.767	0.055	0.231
NTone_LIWC15_ts	0.148	0.138	-0.484	0.767	0.056	0.232
NTone_LMD_ts	0.148	0.140	-0.537	0.758	0.056	0.235
NTone_PR_ts	0.147	0.133	-0.463	0.731	0.060	0.227
AGE	49.826	45.322	0.000	555.000	16.000	93.000
LOSS	0.177	0.382	0.000	1.000	0.000	0.000
LEV	1.662	3.031	0.000	111.411	0.592	1.898
ROA	4.546	11.732	-91.969	90.525	1.606	8.869
ROE	9.162	40.965	-783.269	924.023	3.738	19.491

Table 3.7 shows that further extending the three word lists leads to an increase in $NTone$, resulting in a positive mean. In contrast to using the BPW_N dictionary, the usage of negations also leads to positive means.

To compare the alteration of $NTone$ when using BPW_N and BPW_E, we conduct a dependent-samples t-test. There is a significant difference between $NTone$, when using BPW_N ($Mean = -0.078,\ St.\ Dev. = 0.210$) or BPW_E ($Mean = 0.123,\ St.\ Dev. = 0.156$), $t(6247) = -161.77,\ p < .001$.[10]

As highlighted in Table 3.8, regarding all 6,275 analyzed reports, the editing of stop words leads to an alteration of total and individual words found. Interestingly in contrast to the BPW_N, individual words using BPW_E decrease, while the total number of words increase. Expanding the positive and negative word lists of the BPW_N leads to an immense increase in total and individual words.

[10] For conducting the t-test, all 6,275 data points are used.

Table 3.8
Total Number of Words

	BPW_O	BPW_N	BPW_E
	All words		
Number of words	156,966,254	127,408,125	129,692,675
Individual words	1,143,403	1,143,083	1,142,806
	Positive words		
Number of words	2,169,243	2,219,778	5,709,076
Individual words	1,702	1,718	4,075
	Negative words		
Number of words	2,488,910	2,436,004	4,323,617
Individual words	5,013	5,028	8,341

After correcting for dictionary-specific stop word lists, Table 3.9 displays the cumulative fraction of the ten most frequently used positive words. Despite having minor differences in fractions, the positive words used in BPW_O and BPW_N are identical. In contrast, the ten most frequently used words of the BPW_E are entirely different. This shows the high impact of the above-described extension.

Table 3.9
Ten most Frequent Positive Words

BPW_O		BPW_N		BPW_E	
word	cum %	word	cum %	word	cum %
ertrag	2.06 %	ertrag	2.01 %	erträge	2.31 %
erreicht	3.79 %	erreicht	3.70 %	chancen	4.08 %
erfolg	5.50 %	erfolg	5.38 %	zusammen	5.63 %
zusammenarbeit	7.04 %	zusammenarbeit	6.88 %	wachstum	7.12 %
erfolgreich	8.56 %	erfolgreich	8.37 %	wert	8.59 %
erreichen	10.05 %	erreichen	9.82 %	führen	9.77 %
positiven	11.49 %	positiven	11.23 %	vermögens	10.85 %
positiv	12.92 %	positiv	12.62 %	bedeutung	11.74 %
positive	14.34 %	positive	14.01 %	sicherheit	12.62 %
möglichkeit	15.76 %	möglichkeit	15.40 %	aktiven	13.46 %

Note: We obtained frequencies from the complete sample of 6,275 annual reports.

Table 3.10
Ten most Frequent Negative Words

BPW_O		BPW_N		BPW_E	
word	cum %	word	cum %	word	cum %
gegen	3.72 %	gegen	3.80 %	nicht	17.10 %
verpflichtungen	7.26 %	verpflichtungen	7.42 %	risiken	23.11 %
verluste	10.20 %	verluste	10.42 %	risiko	25.43 %
betrug	12.56 %	wertminderungen	12.51 %	gegen	27.57 %
wertminderungen	14.61 %	verfügung	14.47 %	verpflichtungen	29.61 %
verfügung	16.53 %	wertminderung	16.31 %	verluste	31.30 %
wertminderung	18.33 %	wertberichtigungen	17.99 %	wertminderungen	32.48 %
wertberichtigungen	19.97 %	ermittlung	19.65 %	verfügung	33.58 %
ermittlung	21.60 %	rückgang	21.29 %	wertminderung	34.62 %
rückgang	23.20 %	verpflichtung	22.90 %	wertberichtigungen	35.56 %

Note: We obtained the frequencies from the complete sample of 6,275 annual reports.

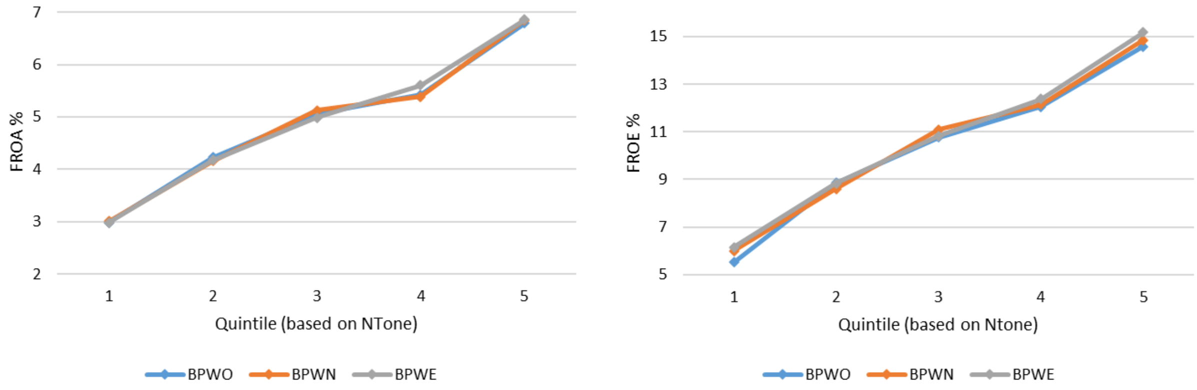
Considering the most frequent negative words in Table 3.10, the main difference between BPW_O and BPW_N is the above-described correction of the word “betrug”, accounting for 2.36 % of all negative words. Due to the extension of the word list, the results for BPW_E show three new words accounting for 25.43 % of all negative words and therefore have a higher fraction than the ten most frequent words on the other lists. Due to their meaning, some words appear both on the lists of negative words and on the corresponding lists of negations. This is particularly clear in the case of the word "nicht", which is the most frequently used negative word in the BPW_E dictionary. All duplications were checked and, in our view, represent both negations and words to be counted as negative.

These findings are consistent with Shapiro et al. (2022, pp. 227f.), stating that apart from domain specificity, the size of the word list is important. A translation of the words used in Table 3.9 and 3.10 can be found in the appendix (section 3.7, Table 3.25).

To test the suitability of the three word lists, we apply the assumption of Loughran and McDonald (2011, pp. 50f.) that the value of sentiment has a direct impact on the particular dependent variable in Figure 3.1. Moreover, higher values in sentiment should lead to higher values in the dependent variables. All three word lists show different and ascending values for *FROA* and *FROE* in all quintiles. Therefore, the necessary assumptions can be considered as given for all three dictionaries.

Figure 3.1

Dependent Variables by Quintile

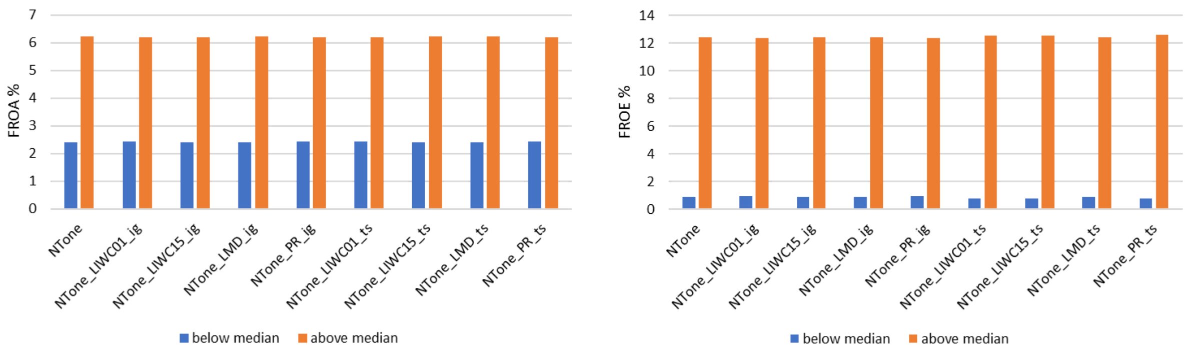


Additionally, we conduct Kruskal-Wallis Tests for all six measurements shown in Figure 3.1. The tests show a statistically significant difference between the quintiles of each measurement. Detailed test statistics can be found in the appendix (section 3.7, Table 3.17).

In addition, we create two groups with above and below median *NTone*, to compare the average *FROA* and *FROE*. For every pair given in Figure 3.2, we perform an independent-samples t-test. All pairs are significantly different from one another. In addition to the given results for the BPW_E in Figure 3.2, we conduct the same tests for below and above measurements for BPW_O and BPW_N. These additional tests show that all pairs for all three word lists are significantly different. The results for all t-tests can be found in the appendix (section 3.7, Tables 3.18 to 3.20).

Figure 3.2

FROA and FROE Grouped by Above and Below Median Sentiment (BPW_E)



3.5.2 Significance of Results

Table 3.11 presents the results for the relation between the two dependent variables (future ROA and future ROE) and *NTone* for all three used dictionaries in a multivariate context, as described in section 3.4.2.

Table 3.11

Regression of NTone and the Three Dictionaries (BPW_O, BPW_N, BPW_E)

	Dependent variable:					
	FROA (BPW_O) (1)	FROA (BPW_N) (2)	FROA (BPW_E) (3)	FROE (BPW_O) (4)	FROE (BPW_N) (5)	FROE (BPW_E) (6)
NTone	1.346	1.517	2.231*	10.397**	10.474**	15.987***
	(0.947)	(0.937)	(1.215)	(4.054)	(4.092)	(5.266)
AGE	-0.0002	-0.001	-0.001	-0.012	-0.013	-0.013
	(0.003)	(0.003)	(0.003)	(0.015)	(0.016)	(0.016)
LOSS	-1.427*	-1.377*	-1.356*	-20.234***	-20.266***	-20.041***
	(0.754)	(0.757)	(0.754)	(3.849)	(3.888)	(3.868)
LEV	0.076	0.076	0.079	-0.005	-0.003	0.018
	(0.057)	(0.057)	(0.058)	(0.947)	(0.946)	(0.949)
ROA	0.586***	0.590***	0.590***			
	(0.043)	(0.043)	(0.043)			
ROE				0.273***	0.272***	0.270***
				(0.071)	(0.071)	(0.072)
Constant	1.547	1.380	0.945	12.613	12.276	9.146
	(3.680)	(3.682)	(3.686)	(11.940)	(11.988)	(11.879)
Observations	4,168	4,112	4,116	4,168	4,112	4,116
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Industry Fixed Effects	YES	YES	YES	YES	YES	YES
R2	0.405	0.411	0.411	0.174	0.174	0.174
Adjusted R2	0.396	0.402	0.402	0.162	0.161	0.162
Residual Std. Error	9.299 (df = 4106)	9.222 (df = 4050)	9.216 (df = 4054)	39.692 (df = 4106)	39.846 (df = 4050)	39.811 (df = 4054)
F Statistic	45.819*** (df = 61; 4106)	46.360*** (df = 61; 4050)	46.425*** (df = 61; 4054)	14.164*** (df = 61; 4106)	13.939*** (df = 61; 4050)	14.015*** (df = 61; 4054)

Significance levels are based on robust standard errors (given in parentheses) and are indicated by * p < 0.1; ** p < 0.05; *** p < 0.01.

The displayed results show a significant relationship between the dependent variables and *NTone* using the extended BPW dictionary (estimation (3) and (6)). Based on those findings, we can confirm the first hypothesis that further correcting and expanding the BPW dictionary improves its ability to forecast future ROAs and ROEs. This shows that the *NTone* of annual reports seems to contain relevant information for future ROAs and ROEs. An increase in *NTone* by the interquartile change of 0.181 for the BPW_E word lists leads to an increase of 40.38% in *FROA* and 289.36 % in *FROE*. Similar relationships were also found while using the dictionaries Henry (2006, 2008) and Ruscheinsky et al. (2018) on English-speaking annual reports (Henry et al., 2021, p. 20; Koelbl, 2020, p. 196). When analyzing conference calls Druz et al. (2020, p. 54) stated that managers could possibly reveal information about future earnings through their usage of sentiment. Although this is a possible reason, we are unable to confirm such a relationship based on the given data.

Additionally, there is a highly significant relationship between the two dependent variables and the current parameters of those variables (*ROA* and *ROE*). The binary variable *LOSS* also shows a significant impact on *FROA* and *FROE*. These results are consistent with Davis and Tama-Sweet (2012, p. 828), Davis et al. (2012, p. 857), and Henry et al. (2021, pp. 20f.).

Table 3.12 and Table 3.13 display the results for using the four different negation lists separated for *FROA* and *FROE*, when using the BPW_E dictionary. The usage of the two LIWC negation lists and the PR negation list improves the significance of results for *FROA* when using the approach of term shifting negated words. The already highly significant results for *FROE* kept their level of significance when using negations. Therefore, we can confirm the second hypothesis that using negations further improves results. The other significant relationships regarding *ROA*, *ROE* and *LOSS* remain unchanged.

Table 3.12
Regression of NTone and FROA for BPW_E (term Shift Negated Words)

	Dependent variable:				
	FROA	FROA	FROA	FROA	FROA
	(7)	(8)	(9)	(10)	(11)
NTone	2.231*				
	(1.215)				
NTone_LIWC01_ts		2.739**			
		(1.255)			
NTone_LIWC15_ts			2.636**		
			(1.254)		
NTone_LMD_ts				2.294*	
				(1.235)	
NTone_PR_ts					2.619**
					(1.318)
AGE	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
LOSS	-1.356*	-1.324*	-1.330*	-1.350*	-1.343*
	(0.754)	(0.752)	(0.752)	(0.753)	(0.754)
LEV	0.079	0.080	0.080	0.079	0.080
	(0.058)	(0.059)	(0.059)	(0.058)	(0.059)
ROA	0.590***	0.589***	0.589***	0.590***	0.589***
	(0.043)	(0.044)	(0.043)	(0.043)	(0.044)
Constant	0.945	0.891	0.905	0.933	0.919
	(3.686)	(3.688)	(3.690)	(3.687)	(3.669)
Observations	4,116	4,116	4,116	4,116	4,116
Year Fixed Effects	YES	YES	YES	YES	YES
Industry Fixed Effects	YES	YES	YES	YES	YES
R2	0.411	0.412	0.411	0.411	0.411
Adjusted R2	0.402	0.403	0.403	0.402	0.403
Residual Std. Error (df = 4054)	9.216	9.214	9.215	9.216	9.215
F Statistic (df = 61; 4054)	46.425***	46.476***	46.463***	46.430***	46.451***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by * p < 0.1; ** p < 0.05; *** p < 0.01.

Table 3.13

Regression of NTone and FROE for BPW_E (term Shift Negated Words)

	Dependent variable:				
	FROE	FROE	FROE	FROE	FROE
	(12)	(13)	(14)	(15)	(16)
NTone	15.987***				
	(5.266)				
NTone_LIWC01_ts		17.400***			
		(5.302)			
NTone_LIWC15_ts			17.053***		
			(5.297)		
NTone_LMD_ts				15.719***	
				(5.212)	
NTone_PR_ts					17.286***
					(5.375)
AGE	-0.013	-0.013	-0.013	-0.013	-0.013
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
LOSS	-20.041***	-19.915***	-19.941***	-20.045***	-20.011***
	(3.868)	(3.864)	(3.868)	(3.872)	(3.860)
LEV	0.018	0.025	0.025	0.020	0.028
	(0.949)	(0.949)	(0.949)	(0.949)	(0.950)
ROE	0.270***	0.270***	0.270***	0.270***	0.270***
	(0.072)	(0.071)	(0.071)	(0.072)	(0.071)
Constant	9.146	8.999	9.053	9.146	9.109
	(11.879)	(11.883)	(11.904)	(11.894)	(11.773)
Observations	4,116	4,116	4,116	4,116	4,116
Year Fixed Effects	YES	YES	YES	YES	YES
Industry Fixed Effects	YES	YES	YES	YES	YES
R2	0.174	0.174	0.174	0.174	0.174
Adjusted R2	0.162	0.162	0.162	0.162	0.162
Residual Std. Error (df = 4054)	39.811	39.803	39.806	39.813	39.808
F Statistic (df = 61; 4054)	14.015***	14.045***	14.035***	14.005***	14.025***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by * p < 0.1; ** p < 0.05; *** p < 0.01.

We also performed all regression using BPW_E and negations with the approach of ignoring the negated words. Only the usage of the LIWC 2001 negations was able to improve results. Due to the minor level of improvement compared to term shifting, the results are given in the appendix (section 3.7, Table 3.21 and 3.22) and are not discussed further.

Table 3.14
Regression of NTone and FROA for BPW_N (term Shift Negated Words)

	Dependent variable:				
	FROA (17)	FROA (18)	FROA (19)	FROA (20)	FROA (21)
NTone	1.517				
	(0.937)				
NTone_LIWC01_ts		1.977**			
		(0.965)			
NTone_LIWC15_ts			2.028**		
			(0.996)		
NTone_LMD_ts				1.635*	
				(0.970)	
NTone_PR_ts					2.168**
					(1.051)
AGE	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
LOSS	-1.377*	-1.345*	-1.340*	-1.369*	-1.347*
	(0.757)	(0.756)	(0.755)	(0.757)	(0.756)
LEV	0.076	0.076	0.077	0.077	0.077
	(0.057)	(0.057)	(0.058)	(0.058)	(0.058)
ROA	0.590***	0.590***	0.590***	0.590***	0.590***
	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Constant	1.380	1.410	1.381	1.367	1.321
	(3.682)	(3.673)	(3.677)	(3.684)	(3.674)
Observations	4,112	4,112	4,112	4,112	4,112
Year Fixed Effects	YES	YES	YES	YES	YES
Industry Fixed Effects	YES	YES	YES	YES	YES
R2	0.411	0.411	0.411	0.411	0.411
Adjusted R2	0.402	0.403	0.403	0.402	0.403
Residual Std. Error (df = 4050)	9.222	9.219	9.219	9.221	9.219
F Statistic (df = 61; 4050)	46.360***	46.413***	46.413***	46.369***	46.421***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 3.15
Regression of NTone and FROE for BPW_N (term Shift Negated Words)

	Dependent variable:				
	FROE	FROE	FROE	FROE	FROE
	(22)	(23)	(24)	(25)	(26)
NTone	10.474**				
	(4.092)				
NTone_LIWC01_ts		11.706***			
		(4.169)			
NTone_LIWC15_ts			11.697***		
			(4.241)		
NTone_LMD_ts				10.715***	
				(4.156)	
NTone_PR_ts					11.961***
					(4.255)
AGE	-0.013	-0.013	-0.013	-0.013	-0.013
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
LOSS	-20.266***	-20.149***	-20.154***	-20.247***	-20.218***
	(3.888)	(3.883)	(3.889)	(3.891)	(3.876)
LEV	-0.003	-0.0001	0.005	0.002	0.005
	(0.946)	(0.945)	(0.946)	(0.947)	(0.946)
ROE	0.272***	0.271***	0.271***	0.272***	0.271***
	(0.071)	(0.071)	(0.071)	(0.071)	(0.071)
Constant	12.276	12.258	12.057	12.124	11.681
	(11.988)	(11.914)	(11.932)	(11.999)	(11.905)
Observations	4,112	4,112	4,112	4,112	4,112
Year Fixed Effects	YES	YES	YES	YES	YES
Industry Fixed Effects	YES	YES	YES	YES	YES
R2	0.174	0.174	0.174	0.173	0.174
Adjusted R2	0.161	0.161	0.161	0.161	0.161
Residual Std. Error (df = 4050)	39.846	39.839	39.841	39.846	39.843
F Statistic (df = 61; 4050)	13.939***	13.966***	13.957***	13.937***	13.950***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Additional proof of the importance of implementing negations is given in Table 3.14 and Table 3.15. When using BPW_N, negations and the approach of term shifting, the levels of significance are equal to the usage of the superior BPW_E word lists. These results underline the importance of implementing negations. Based on the results visible, the word list PR, developed specifically for the financial context, should be used. These results are consistent with Shapiro et al. (2022, pp. 227f.) and their findings, which claim that using negations improves the prediction of human sentiment ratings.

The improvement shown above also partially applies when using the approach of ignoring negated words. The relevant tables are given in the appendix (section 3.7, Tables 3.23 and 3.24).

3.6 Conclusion

This paper uses the dictionary-based approach to compute the sentiment of German-speaking annual reports. Due to the novelty of the used dictionary, the aim of this paper is to improve the given BPW_O and BPW_N word lists by further correction and expansion. Additionally, we test the use of different negations to further improve the results.

The expansion of the BPW_N word lists leads to an immense increase in total words found (positive: 157 %, negative: 77 %). Additionally, the ten most frequent positive and negative words found underwent an enormous change. This leads to a significant change in NTone calculated by using BPW_E. Despite the fundamental alteration, we successfully test basic assumptions visually and statistically. By using the new and extended BPW_E, we are able to improve regression results compared to the two previous versions and therefore confirm the first hypothesis. Additionally, we can show that negations should be implemented because they are able to improve results. A deterioration of results caused by the usage of negations could not be observed and should therefore be implemented in the form of term shifted PR negations.

Furthermore, by successfully improving the second version of the BPW dictionary and testing the implementation of negations, this paper contributes immensely to the existing literature on analyzing German corporate disclosures.

Due to this successful improvement of the BPW dictionary, further research on finance related texts should be conducted by using the BPW_E. Based on the novelty of this dictionary, other types of corporate disclosure should be analyzed, and a comparison to general German dictionaries should be conducted.

3.7 Appendix

Table 3.16
Description of Variables

Variable	Description
AGE	Age of the Company: Difference between the year of observation and the date of incorporation
FROA	Future Return on Assets: Return on Assets (ROA) one year ahead
FROE	Future Return on Equity: Return on Equity (ROE) one year ahead
LEV	Leverage: Sum of non-current liabilities and current liabilities, divided by shareholders funds
LOSS	LOSS equals one if the Profit and Loss before tax is negative, zero otherwise
NTone	Net Tone: Difference between the number of positive and negative words, divided by the sum of positive and negative words
ROA	Current Return on Assets: Profit and Loss before tax divided by total assets times 100
ROE	Current Return on Equity: Profit and Loss before tax divided by shareholders funds times 100

Table 3.17
Kruskal-Wallis test Statistics

	FROA			FROE		
	BPW_O	BPW_N	BPW_E	BPW_O	BPW_N	BPW_E
Kruskal-Wallis-H	207.201	210.450	249.461	242.842	240.057	256.486
df	4	4	4	4	4	4
Asymp. Sig.	<.001	<.001	<.001	<.001	<.001	<.001

Table 3.18
Independent Samples t-test for Below and Above Median Sentiment (BPW_O)

Dependent variable	Sentiment measure	Statistics	df	p	Mean below	Mean above
FROA	NTone	-8.377	4072	<.001	2.763	5.844
FROE	NTone	-7.975	3187	<.001	1.277	11.910

Table 3.19

Independent Samples t-test for Below and Above Median Sentiment (BPW_N)

Dependent variable	Sentiment measure	Statistics	df	p	Mean below	Mean above
FROA	NTone	-9.516	3992	<.001	2.558	6.060
	NTone_LIWC01_ig	-10.890	3879	<.001	2.312	6.307
	NTone_LIWC15_ig	-10.300	3939	<.001	2.417	6.201
	NTone_LMD_ig	-9.506	4009	<.001	2.560	6.059
	NTone_PR_ig	-10.080	3929	<.001	2.456	6.162
	NTone_LIWC01_ts	-11.120	3878	<.001	2.271	6.348
	NTone_LIWC15_ts	-10.800	3914	<.001	2.328	6.291
	NTone_LMD_ts	-9.363	3996	<.001	2.586	6.033
	NTone_PR_ts	-9.846	3931	<.001	2.499	6.120
FROE	NTone	-8.312	3140	<.001	1.066	12.250
	NTone_LIWC01_ig	-9.025	3024	<.001	0.595	12.720
	NTone_LIWC15_ig	-8.760	3029	<.001	0.770	12.550
	NTone_LMD_ig	-8.230	3153	<.001	1.120	12.200
	NTone_PR_ig	-8.669	3026	<.001	0.830	12.490
	NTone_LIWC01_ts	-9.123	3020	<.001	0.530	12.790
	NTone_LIWC15_ts	-8.996	3019	<.001	0.614	12.700
	NTone_LMD_ts	-8.341	3080	<.001	1.047	12.270
	NTone_PR_ts	-8.400	3015	<.001	1.008	12.310

Table 3.20
Independent Samples t-test for Below and Above Median Sentiment (BPW_E)

Dependent variable	Sentiment measure	Statistics	df	p	Mean below	Mean above
FROA	NTone	-10.350	3875	<.001	2.411	6.209
	NTone_LIWC01_ig	-10.240	3877	<.001	2.430	6.190
	NTone_LIWC15_ig	-10.350	3874	<.001	2.411	6.208
	NTone_LMD_ig	-10.410	3874	<.001	2.400	6.220
	NTone_PR_ig	-10.240	3873	<.001	2.430	6.190
	NTone_LIWC01_ts	-10.300	3852	<.001	2.419	6.201
	NTone_LIWC15_ts	-10.380	3853	<.001	2.406	6.213
	NTone_LMD_ts	-10.390	3884	<.001	2.403	6.217
	NTone_PR_ts	-10.230	3872	<.001	2.433	6.187
FROE	NTone	-8.607	3105	<.001	0.878	12.440
	NTone_LIWC01_ig	-8.489	3102	<.001	0.956	12.360
	NTone_LIWC15_ig	-8.554	3100	<.001	0.912	12.410
	NTone_LMD_ig	-8.581	3104	<.001	0.895	12.420
	NTone_PR_ig	-8.501	3099	<.001	0.947	12.370
	NTone_LIWC01_ts	-8.728	3016	<.001	0.798	12.520
	NTone_LIWC15_ts	-8.787	3016	<.001	0.759	12.560
	NTone_LMD_ts	-8.553	3103	<.001	0.913	12.410
	NTone_PR_ts	-8.798	3008	<.001	0.752	12.570

Table 3.21

Regression of NTone and FROA for BPW_E (Ignore Negated Words)

	Dependent variable:				
	FROA	FROA	FROA	FROA	FROA
	(27)	(28)	(29)	(30)	(31)
NTone	2.231*				
	(1.215)				
NTone_LIWC01_ig		2.392**			
		(1.212)			
NTone_LIWC15_ig			2.322*		
			(1.208)		
NTone_LMD_ig				2.228*	
				(1.214)	
NTone_PR_ig					2.216*
					(1.202)
AGE	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
LOSS	-1.356*	-1.340*	-1.344*	-1.353*	-1.351*
	(0.754)	(0.753)	(0.753)	(0.754)	(0.754)
LEV	0.079	0.079	0.079	0.079	0.079
	(0.058)	(0.058)	(0.058)	(0.058)	(0.058)
ROA	0.590***	0.589***	0.589***	0.590***	0.589***
	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Constant	0.945	0.917	0.925	0.939	0.928
	(3.686)	(3.687)	(3.688)	(3.686)	(3.679)
Observations	4,116	4,116	4,116	4,116	4,116
Year Fixed Effects	YES	YES	YES	YES	YES
Industry Fixed Effects	YES	YES	YES	YES	YES
R2	0.411	0.411	0.411	0.411	0.411
Adjusted R2	0.402	0.403	0.402	0.402	0.402
Residual Std. Error (df = 4054)	9.216	9.215	9.216	9.216	9.216
F Statistic (df = 61; 4054)	46.425***	46.447***	46.439***	46.427***	46.431***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by

* p < 0.1; ** p < 0.05; *** p < 0.01.

Table 3.22
Regression of NTone and FROE for BPW_E (Ignore Negated Words)

	Dependent variable:				
	FROE	FROE	FROE	FROE	FROE
	(32)	(33)	(34)	(35)	(36)
NTone	15.987***				
	(5.266)				
NTone_LIWC01_ig		16.281***			
		(5.184)			
NTone_LIWC15_ig			16.014***		
			(5.173)		
NTone_LMD_ig				15.707***	
				(5.205)	
NTone_PR_ig					15.547***
					(5.048)
AGE	-0.013	-0.013	-0.013	-0.013	-0.013
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
LOSS	-20.041***	-19.967***	-19.984***	-20.037***	-20.020***
	(3.868)	(3.866)	(3.868)	(3.869)	(3.864)
LEV	0.018	0.022	0.022	0.019	0.023
	(0.949)	(0.949)	(0.949)	(0.949)	(0.949)
ROE	0.270***	0.270***	0.270***	0.270***	0.270***
	(0.072)	(0.072)	(0.072)	(0.072)	(0.072)
Constant	9.146	9.038	9.074	9.133	9.065
	(11.879)	(11.879)	(11.889)	(11.885)	(11.827)
Observations	4,116	4,116	4,116	4,116	4,116
Year Fixed Effects	YES	YES	YES	YES	YES
Industry Fixed Effects	YES	YES	YES	YES	YES
R2	0.174	0.174	0.174	0.174	0.174
Adjusted R2	0.162	0.162	0.162	0.162	0.162
Residual Std. Error (df = 4054)	39.811	39.807	39.808	39.812	39.810
F Statistic (df = 61; 4054)	14.015***	14.032***	14.025***	14.012***	14.019***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by

* p < 0.1; ** p < 0.05; *** p < 0.01.

Table 3.23
Regression of NTone and FROA for BPW_N (Ignore Negated Words)

	Dependent variable:				
	FROA	FROA	FROA	FROA	FROA
	(37)	(38)	(39)	(40)	(41)
NTone	1.517				
	(0.937)				
NTone_LIWC01_ig		1.648*			
		(0.929)			
NTone_LIWC15_ig			1.624*		
			(0.934)		
NTone_LMD_ig				1.527	
				(0.942)	
NTone_PR_ig					1.638*
					(0.931)
AGE	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
LOSS	-1.377*	-1.363*	-1.363*	-1.375*	-1.365*
	(0.757)	(0.757)	(0.757)	(0.757)	(0.757)
LEV	0.076	0.076	0.077	0.076	0.077
	(0.057)	(0.057)	(0.058)	(0.058)	(0.058)
ROA	0.590***	0.590***	0.590***	0.590***	0.590***
	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Constant	1.380	1.394	1.380	1.373	1.368
	(3.682)	(3.677)	(3.680)	(3.683)	(3.677)
Observations	4,112	4,112	4,112	4,112	4,112
Year Fixed Effects	YES	YES	YES	YES	YES
Industry Fixed Effects	YES	YES	YES	YES	YES
R2	0.411	0.411	0.411	0.411	0.411
Adjusted R2	0.402	0.402	0.402	0.402	0.402
Residual Std. Error (df = 4050)	9.222	9.221	9.221	9.222	9.221
F Statistic (df = 61; 4050)	46.360***	46.379***	46.377***	46.362***	46.381***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by

* p < 0.1; ** p < 0.05; *** p < 0.01.

Table 3.24
Regression of NTone and FROE for BPW_N (Ignore Negated Words)

	Dependent variable:				
	FROE	FROE	FROE	FROE	FROE
	(42)	(43)	(44)	(45)	(46)
NTone	10.474**				
	(4.092)				
NTone_LIWC01_ig		10.706***			
		(4.032)			
NTone_LIWC15_ig			10.595***		
			(4.066)		
NTone_LMD_ig				10.446**	
				(4.115)	
NTone_PR_ig					10.383***
					(3.956)
AGE	-0.013	-0.013	-0.013	-0.013	-0.013
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
LOSS	-20.266***	-20.207***	-20.205***	-20.253***	-20.231***
	(3.888)	(3.886)	(3.888)	(3.888)	(3.883)
LEV	-0.003	-0.001	0.001	-0.001	0.001
	(0.946)	(0.946)	(0.946)	(0.947)	(0.946)
ROE	0.272***	0.271***	0.271***	0.272***	0.271***
	(0.071)	(0.071)	(0.071)	(0.071)	(0.071)
Constant	12.276	12.280	12.202	12.210	12.087
	(11.988)	(11.951)	(11.961)	(11.995)	(11.995)
Observations	4,112	4,112	4,112	4,112	4,112
Year Fixed Effects	YES	YES	YES	YES	YES
Industry Fixed Effects	YES	YES	YES	YES	YES
R2	0.174	0.174	0.174	0.174	0.174
Adjusted R2	0.161	0.161	0.161	0.161	0.161
Residual Std. Error (df = 4050)	39.846	39.843	39.843	39.846	39.844
F Statistic (df = 61; 4050)	13.939***	13.951***	13.949***	13.940***	13.945***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by

* p < 0.1; ** p < 0.05; *** p < 0.01.

Table 3.25
Translation of ten most Frequent Words (all Three Dictionaries)

Positive words		Negative words	
German	English	German	English
chancen	chances	betrug	fraud, amounted
erfolg	success	ermittlung	investigation
erfolgreich	successful	gegen	against
erreichen	achieve	nicht	not
erreicht	achieved	risiken	risks
ertrag	return, revenue	risiko	risk
erträge	returns, revenues	rückgang	decline
führen	lead	verfügung	decree
positiven	positive	verluste	losses
vermögens	assets	verpflichtung	obligation
wachstum	growth	verpflichtungen	obligations
wert	value	wertberichtigungen	value adjustments
zusammen	together	wertminderung	impairment
zusammenarbeit	cooperation	wertminderungen	impairments

4. Building a Domain-Specific Dictionary for Artificial Intelligence Using Word2vec: A Contextual Approach to Keyword Extraction and Search for German Savings and Cooperative Banks

[11]

Abstract

The use and implementation of artificial intelligence (AI) represent an increasingly important area of interest. The application of AI by the two largest groups in the German banking market, based on the number of institutions, can generally be analyzed in a structured manner using surveys. In this paper, we introduce a combined methodology from the approaches Word2vec and bag-of-words, to create an own AI dictionary. This approach allows us to present a comprehensive analysis of the application of AI by savings and cooperative banks without being confronted with survey issues such as non-response bias. Additionally, we are able to provide a dictionary for further research.

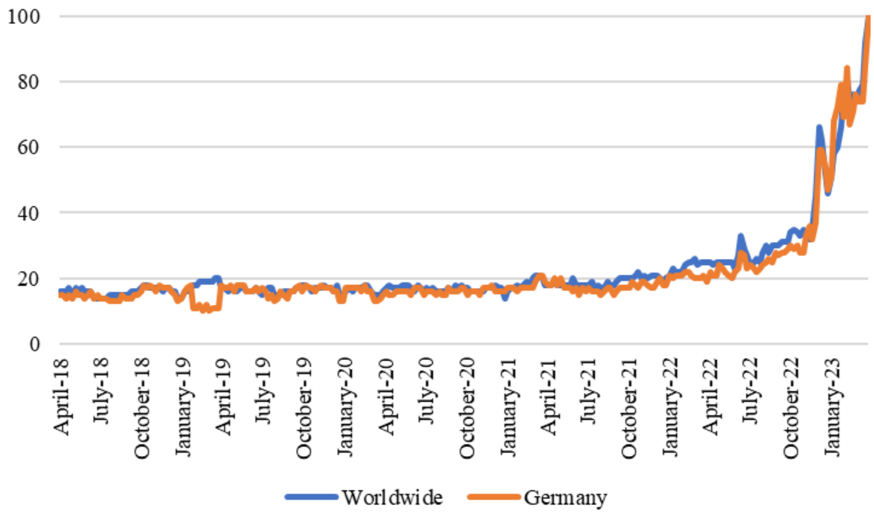
[11] The author gratefully acknowledges financial support by DZ BANK Stiftung.

4.1 Introduction

There are several definitions of artificial intelligence (AI), focusing on different aspects. AI can be defined as "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan & Haenlein, 2019, p. 17) or "as a set of theories and techniques used to create machines capable of simulating intelligence. AI is a general term that involves the use of a computer to model intelligent behavior with minimal human intervention" (Wamba-Taguimdje et al., 2020, p. 1894). We combine these aspects of learning and intelligence and define AI as "a system's capability of simulating intelligent behavior, by interpreting external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation". The topic of artificial intelligence has become increasingly important in business and research. Google Trends shows a worldwide increase in searches for AI of 300% between the end of March 2022 and March 2023 (Germany 426%), as shown in Figure 4.1. In published articles of the Handelsblatt, an increase of 148%, related to articles about AI, could be observed in March 2023 compared to the previous year (Handelsblatt, 2023). There is also an increasing trend in finance research on the topic of AI (Goodell et al., 2021, p. 4) which covers multiple areas, such as marketing (Mauro et al., 2022; Schiessl et al., 2022), human resources (Kaushal et al., 2021), operations management (Grover et al., 2022) and strategic management (Keding, 2021).

Figure 4.1

Google Trends for Artificial Intelligence



Source: Google (2023).

Being a disruptive and transformative technology, AI will change multiple sectors and society (Davenport et al., 2018, p. 116). Although being a highly regulated domain, AI is an important topic in the financial industry (Weber et al., 2023, p. 1). Where several topics like data mining to predict bankruptcy, optimize risk models, credit and loan analysis, bank efficiency, and customer relations were covered in research (Doumpos et al., 2023, p. 4; Fares et al., 2022, pp. 3–5). Artificial intelligence is even taking hold of banking supervision. The European Central Bank has created its own hub to teach supervisors more about current technologies and help them use advanced analytics and develop new tools (European Central Bank, 2023b).

Surprisingly, the current and future application of AI regarding financial institutions represents a major research gap. For this reason, it is partly necessary to fall back on publications from consulting companies. Nevertheless, it is clear that AI will play a significant role in the future of banking, affecting all areas of banks (Milana & Ashta, 2021, p. 204; OliverWyman, 2022, p. 11). Fields of significant future developments are customer service and experience (Deloitte, 2021, p. 2; Dietzmann et al., 2023, pp. 17f.; Evident Insights, 2023, p. 5), crime detection and prevention (Deutsche Bank, 2023; Evident Insights, 2023, p. 5), credit underwriting and risk assessment (Deloitte, 2021, p. 3; Evident Insights, 2023, p. 5).

The German banking sector consists of different groups of financial institutions (Deutsche Bundesbank, 2023, pp. 203ff.). According to the Deutsche Bundesbank, savings and cooperative banks, with 362 and 733 institutions respectively, represent a large proportion of the total of 1,386 reporting institutions in Germany (Deutsche Bundesbank, 2023, pp. 6, 12). Therefore, it would be of utmost interest to investigate the current use of AI in these two banking groups. The resulting research question is whether and to what extent savings and cooperative banks use artificial intelligence.

One possibility to answer this question is the usage of questionnaires or interviews. When using surveys the accuracy and the significance of the results are influenced by various challenges, like random sampling error, occurring when a sample is not representative of the population, or respondent error, when respondents to surveys are answering questions falsely or not at all (Sreejesh et al., 2014, pp. 73–77). Especially the problem of not answering surveys and incentives to cope with this non-response error in organizational science is a widely analyzed field in research (Anseel et al., 2010; Baruch & Holtom, 2008; Rogelberg & Stanton, 2007) where response rates for managers or top executives are on average between 36.2% and 47.1% (Anseel et al., 2010, p. 341; Baruch & Holtom, 2008, p. 1148). The alternative of conducting interviews with corporate representatives is a viable but very time-consuming approach due to the large number of savings and cooperative banks. Although surveys are a viable approach to gain important insights, we use an alternative approach to avoid the stated downsides.

Based on the assumption that German savings and cooperative banks communicate publicly about the successful implementation or use of new technologies, such as artificial intelligence, this paper takes a bag-of-words approach (BoW) to identify these banks. The bag-of-words or dictionary-based approach is a rule-based algorithm, that classifies words or phrases of a given text into different categories based on predefined rules or categories like dictionaries or word lists (F. Li, 2010, p. 146). To identify AI-using banks, we need as comprehensive a word list of AI terms as possible. Due to the specificity of this required word list, we will not be able to use existing lists. For this reason, we use the neural network Word2vec to create a customized list of AI-relevant words. Therefore, this paper hypothesizes that it is possible to identify AI-using savings and cooperative banks through their external communication. Analogous to the challenges of the survey approach, it must be noted that success is only achievable under the premise of current available external communication on the topic of AI. If savings and cooperative banks surveyed do not communicate externally about AI, it would not be possible to obtain results using the selected approach. One possibility of non-existent communication may be due to internal processes that should not be disclosed externally, such as fraud detection.

The contribution of this paper to the literature on textual analysis of German texts is the provision of a domain-specific dictionary for analyzing the usage of AI by German banks. Additionally, the proposed method of using Word2vec for creating domain-specific dictionaries based on multiple iterations, as well as the approach to use external communication instead of surveys can be applied to solve other research questions.

This paper proceeds as follows. In the second part, we provide a short review of the relevant literature on textual analysis, focusing on the creation of dictionaries, as well as the usage of Word2vec for creating those. The third section presents the data and the different parsing procedures applied. The methodology to obtain the dictionary and the subsequent and customized usage of the dictionary approach are given in passage four. The fifth part presents the obtained results. Lastly, the sixth section concludes.

4.2 Literature Review

Most contributions to textual analysis focus primarily on the calculation of sentiment (Gupta et al., 2020, p. 3). There are several summarizing papers like Algaba et al. (2020), Chakraborty and Bhattacharjee (2020), Kearney and Liu (2014) and Loughran and McDonald (2015, 2016), addressing this research topic. Nevertheless, the findings therein, in conjunction with the discussion of the need for domain specificity of word lists can be adapted for the analysis of savings and cooperative banks regarding the use of AI.

Based on this necessary domain specificity, presented in the literature shown, as well as the absence of an existing dictionary, the creation of a separate word list is necessary to analyze the external communication of savings and cooperative banks, regarding the use of AI. General information about the creation of dictionaries using machine learning is given in Loughran and McDonald (2020b, p. 364). Gentzkow et al. (2019) provide a comprehensive review of the most prominent machine learning methods. These methods encompass a wide range of techniques used to create dictionaries. On the one hand, there are simple approaches, such as manually selecting relevant words and adding synonyms (S.-M. Kim & Hovy, 2004, p. 2; Wilson et al., 2005, p. 349) or merging several already existing lists (Jegadeesh & Wu, 2013, p. 713). Other approaches are based on categorized data by classifying datasets in advance in a certain way, for example in terms of sentiment or stock price returns (Bos & Frasincar, 2022, pp. 445f.; Liu & Alsaadi, 2020, pp. 354f.; Yekrangi & Abdolvand, 2021, pp. 134f.). Since neither existing lists nor classified data can be used to generate the list of AI-relevant words, the methodology of this paper is partly based on the contributions of Das et al. (2022, pp. 3f.), Du et al. (2021, p. 8), S. Li et al. (2021, p. 6), Tsai and Wang (2014, p. 1454) and Xue et al. (2021, p. 101608:5). They search semantically similar words to extend existing seed words based on word embeddings using the Word2vec approach introduced by Mikolov et al. (2013).

Word embeddings are representations of words as vectors of real numbers, where similar vectors are grouped in a vector space, meaning that those words are related to the same concepts and contexts (Bos & Frasincar, 2022, p. 448; Das et al., 2022, p. 3; Du et al., 2021, p. 8; Mahmoudi et al., 2021, p. 10). Primarily used algorithms are FastText, GloVe, and Word2vec (Bos & Frasincar, 2022, p. 448; Das et al., 2022, p. 3; Naili et al., 2017, p. 340).

4.3 Data

4.3.1 Data Source

Three comprehensive data corpora form the data basis of this paper. The two data corpora of external communication of savings and cooperative banks to be analyzed are compiled by a web crawler programmed in Python (analysis corpora). Here, all contents from the homepages of active savings and cooperative banks as of the end of 2021 (Bundesverband der Deutschen Volksbanken und Raiffeisenbanken, 2022; Deutscher Sparkassen- und Giroverband, 2022) are scanned systematically, in the period from 08/01/2022 to 08/12/2022, and the individual homepages and PDFs are downloaded and stored. Due to the lack of information on homepages, we determined the URLs for all banks manually. As given in Table 4.1 1,156 homepages were scanned to download 2,217,269 HTML and PDF files.

Table 4.1
Data Corpora of External Communication

	Combined	Savings Banks	Cooperative Banks
Overall	2,217,269	877,036	1,340,233
HTML	2,075,033	820,629	1,254,404
PDF	142,236	56,407	85,829
Number of Banks	1,156	370	786

Note: We base our analysis on the number of institutions provided by Bundesverband der Deutschen Volksbanken und Raiffeisenbanken and Deutscher Sparkassen- und Giroverband, which differ from the numbers reported by Deutsche Bundesbank.

The third data corpus forms the basis of the dictionary of AI-relevant terms to be created and is used to train the neuronal net (training corpus), which is described in 4.4.1. For this purpose, textbooks and papers dealing with AI were collected through a systematic search. Additionally, the databases wiso and Nexis were used to collect newspaper articles regarding AI, published between January 2017 and June 2022. Table 4.2 shows the type and the number of collected files.

Table 4.2
Data Corpora of Training Corpus

	Number
Textbooks	302
Papers	1,440
News	67,690

4.3.2 Parsing

We use a conversion tool programmed in Python to get the textual data from the homepages and PDF files. For the homepages we use the library BeautifulSoup (Renault, 2017, p. 27; Vytautas et al., 2018, p. 58) to delete any HTML code and save the text in UTF-8 encoded TXT files (Y. Kang et al., 2020, p. 157; Meier et al., 2018, p. 29). We convert PDF files into TXT files using the package PDFMiner.

To obtain a dataset of plain text for all three corpora, we follow approaches used in the field of sentiment analysis, by using an automated parser programmed in Python. For all files we replace typographic ligatures (Bannier et al., 2017, p. 10, 2019a, p. 9) and hyphens (Loughran & McDonald, 2011, internet appendix). Additionally, we remove numbers (Koelbl, 2020, p. 186; Picault & Renault, 2017, p. 139), special characters (Allee & Deangelis, 2015, p. 247; Gentzkow et al., 2019, p. 536), and multiple whitespaces (González et al., 2019, p. 433; Schmeling & Wagner, 2016, p. 8). Based on a random check of all files, we additionally delete multiple full stops, URLs, and single-letter words. Due to different spellings in the given data, we replaced all mutated vowels.

For further processing and due to the peculiarities of the Python libraries used, we take complementary steps for the two analysis corpora. We convert all words to lowercase (Pengnate et al., 2020, p. 193; Picault & Renault, 2017, p. 139), remove punctuation (Algaba et al., 2020, p. 529; Iqbal & Riaz, 2022, p. 2702) and convert bigrams (see 4.4.2).

4.4 Methodology

4.4.1 Word2vec

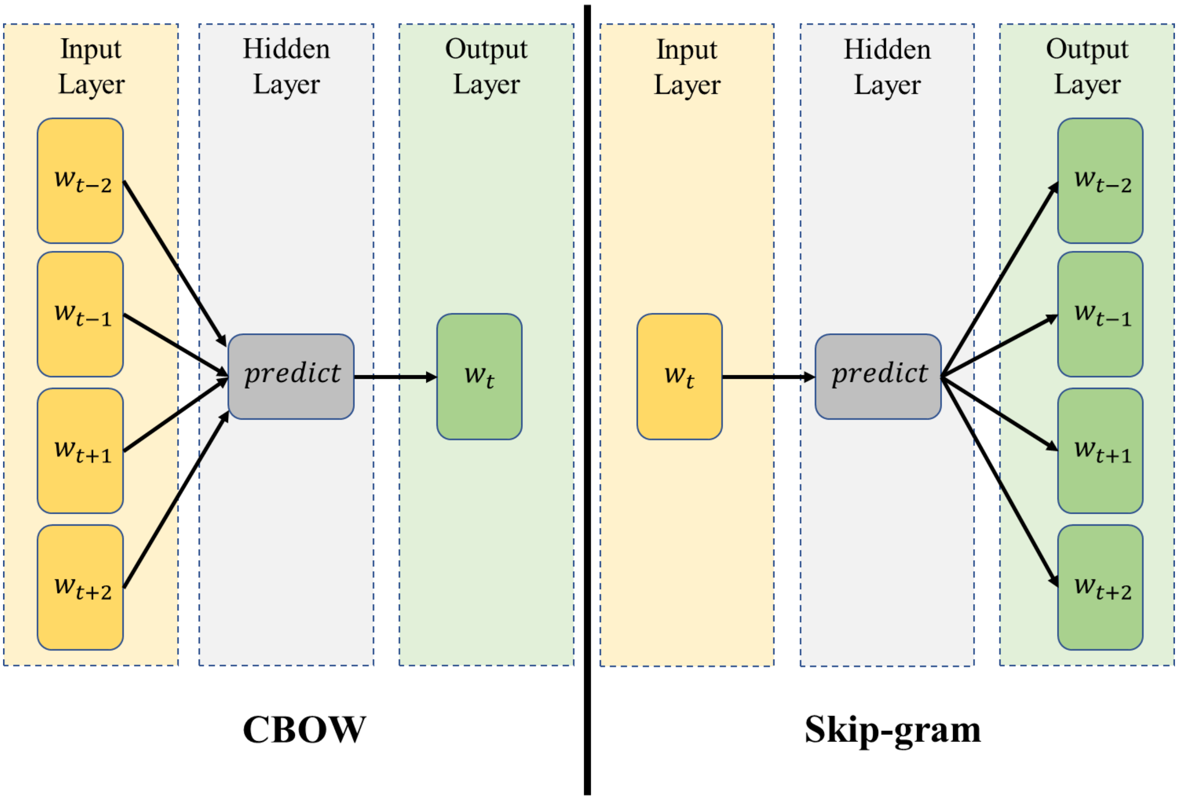
For creating our dictionary, we apply the Word2vec approach presented in 2013 by Mikolov et al., by using the Python library Gensim (Chen & Sokolova, 2021, p. 414:3; Y. Kang et al., 2020, p. 160).

The basic idea is that a word is most accurately described by its context (Y. Kang et al., 2020, p. 160; Naili et al., 2017, p. 341). Word2vec can identify words that appear in a similar context (Xue et al., 2021, p. 101608:2). Word2vec generates word embeddings, which are multi-dimensional numerical vectors for individual words using neural networks (NN), these embeddings are based on how many times specific words appear as the context of a word (Chen & Sokolova, 2021, p. 414:4; Sagnika et al., 2020, p. 32391; Xue et al., 2021, p. 101608:5). Those vectors allow to capture important semantic relationships by using simple algebraic operations (Kale et al., 2023, p. 3; Mikolov et al., 2013, p. 2; Yang et al., 2018, p. 184). Mikolov et al. (2013, p. 2) observed that the vector ”King” – vector ”Man” + vector ”Women” results in a vector that is closest, in terms of cosine similarity, to the vector of the word ”Queen” (Kale et al., 2023, p. 3). In summary, words with a similar context are closer to each other than words that do not have this context. (Bos & Frasincar, 2022, p. 448)

For the training of those word embeddings the two different unsupervised approaches Continuous bag-of-words (CBOW) and Skip-gram (SG) model presented in Figure 4.2 are available (Mishev et al., 2020, p. 131665). CBOW predicts a target word (w_t) based on its context (w_{-2} to w_{+2}), by using three layers of a NN. The context words are used in the input layer of the NN to be predicted and projected into the output layer by the hidden layer. Skip-gram is the opposite of CBOW, by predicting the context (w_{-2} to w_{+2}) of a given target word (w_t) (Chen & Sokolova, 2021, p. 414:4; Naili et al., 2017, p. 342). The output of both models are word vectors, which enable both approaches to learn vector representation of words in a vector space (Asudani et al., 2023, p. 12).

Figure 4.2

CBOW and Skip-gram Model Architecture



Source: adapted from Mikolov et al. (2013, p. 5), Torregrossa et al. (2021, p. 88), and Kumar (2021, p. 49).

When using Word2vec, we are able to set several parameters to train the model. We need to define the vector length and the window length (Kale et al., 2023, p. 2; Lauren et al., 2018, p. 630), where vector length is the number of dimensions and window length determines the distance to the left and right of the target word (Lauren et al., 2018, pp. 626, 630).

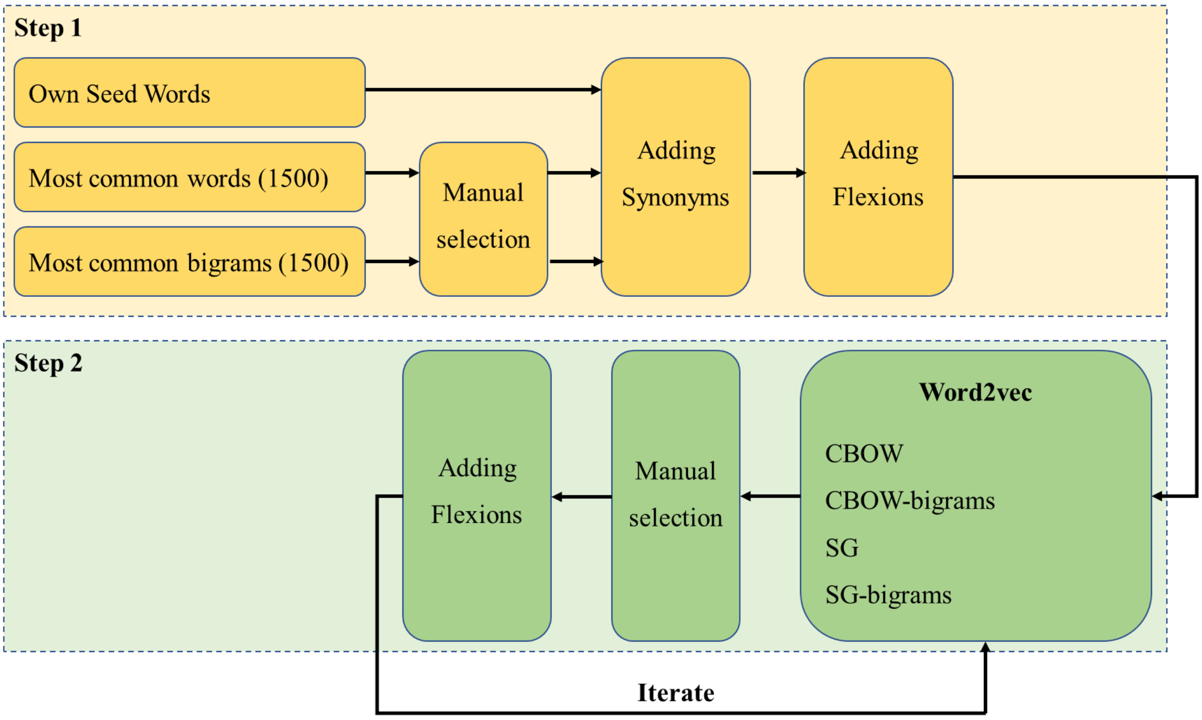
Most studies using Word2vec do not specify the used parameters (Du et al., 2021; Tsai & Wang, 2014; Xue et al., 2021), therefore we base our setting on the studies of S. Li et al. (2021, p. 6) and Theil et al. (2020, p. 6:7). Both use a vector length of 100 but differ in window size (S. Li et al., 2021, p. 6; Theil et al., 2020, p. 6:7). Due to the findings of Levy and Goldberg (2014, p. 304) that a larger window length captures more broad topical content, we choose the approach of S. Li et al. (2021, p. 6) and use a windows size of 10.

Naili et al. (2017, p. 349) found that CBOW is more efficient in the case of frequent words, while SG is more efficient in the case of infrequent words. Due to this differentiation, we use both approaches with the above-described parameters to obtain the most comprehensive results possible.

We base our methodology visualized in Figure 4.3 on the contributions of Das et al. (2022, pp. 3-5), Du et al. (2021, p. 8), S. Li et al. (2021, p. 6), Tsai and Wang (2014, p. 1454) and Xue et al. (2021, p. 5).

Figure 4.3

Methodology for Building AI Dictionary



They use models for word embeddings to expand given lists of relevant words (seed words). Seed words are equivalent to target words in the CBOW and SG models described above. The goal of all five contributions is to find similar context words based on defined seed words. Therefore, different seed words like existing dictionaries (S. Li et al., 2021, p. 6; Tsai & Wang, 2014, p. 1454) or self-defined relevant words are used (Das et al., 2022, pp. 2f.; Du et al., 2021, pp. 8f.). Since we cannot use existing dictionaries to create our word list on AI-relevant words, we follow the approach of selecting our own seed words. Here, we follow the work of Xue et al. (2021, p. 101608:5), who use three seed words to find out whether companies are pursuing a Big Data strategy. We use three sources of seed words. Through reading relevant literature, we are able to identify 36 relevant unigrams and bigrams (unigram refers to single words and bigram refers to two-word pairs). We add relevant synonyms and flexions of those words[12] to get 86 seed words. Additionally, we filter the 1,500 most common occurrences for unigrams and bigrams and manually select all relevant words. After adding synonyms and flexions we end with a final sample of 475 seed words.

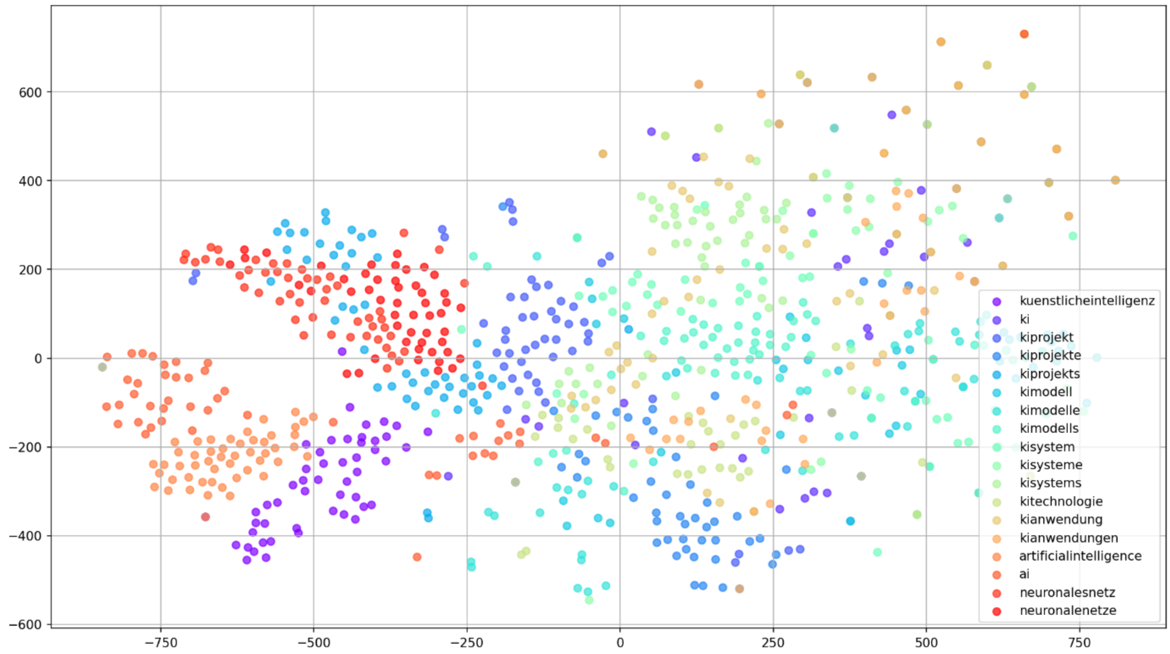
Due to our initial seed words consisting of unigrams and bigrams we train two separate models each for unigrams and bigrams. We use those seed words in our four trained Word2vec models to obtain the most similar words, based on cosine similarity, for each model. The number of similar words in our underlying literature ranges from five (S. Li et al., 2021, p. 102673:6; Tsai & Wang, 2014, p. 1454) to 1,000 (Das et al., 2022, p. 4). Given the number of seed words, as well as the iterations described below, we set the number of similar words to 50.

Figure 4.4 shows a graphical examination of this approach in the context of a T-SNE two-dimensional projection of the Word2vec vectors following Gastaldi (2021, pp. 156f.) and Tang et al. (2016), for the 50 most similar words of selected seed words using the unigram CBOW model. Based on the clusters formed, it becomes apparent that certain similar words are closer together than others. This different degree of similarity makes it necessary to validate the words found and argues against an uncontrolled inclusion of these, as performed by S. Li et al. (2021, p. 102673:6) and Tsai and Wang (2014, p. 1454).

[12] In the further course the term "word" refers to unigrams and bigrams.

Figure 4.4

CBOW Unigram T-SNE Projection



Based on the works of Du et al. (2021, p. 10) and Xue et al. (2021, p. 101608:5), in accordance with the observations stated above, we manually check all words found for suitability for our dictionary. This approach is also consistent with the acknowledgments of Das et al. (2022, p. 3), who emphasize the need for manual post-processing in their found words. In addition, we again add flexions. To account for alternative spellings or changes in words due to parsing, we also add all bigrams in a united form as unigrams. Through this approach, we are able to obtain 2,585 new words.

We use these 2,585 words in a subsequent iteration using the presented method. Except for the contribution by Du et al. (2021, p. 10), all five underlying papers use only a single iteration (Das et al., 2022, pp. 3-5; S. Li et al., 2021, p. 6; Tsai and Wang, 2014, p. 1454; Xue et al., 2021, p. 5). We extend this approach, which uses three iterations, and repeat the procedure until no more relevant words are found. Finally, all of the relevant words are checked again, and any irrelevant words are deleted. An overview of all 12 iterations and the performed adjustments can be found in Table 4.3.

Table 4.3

Overview Building AI Dictionary (12 Iterations)

Seed Words	Own seed words	86	475
	Most common words	13	
	Most common bigrams	376	
Words found	CBOW	48,426	196,450
	Skip Gram	28,741	
	CBOW bigrams	75,839	
	Skip Gram bigrams	43,444	
Words added	CBOW	927	10,445
	Skip Gram	669	
	CBOW bigrams	1,204	
	Skip Gram bigrams	1,061	
	Flexions	3,074	
	Unite Words	3,510	
Words deleted	Delete Bigrams	2,098	4,190
	Delete united Bigrams	2,092	
Words added	Add parts of Bigrams	205	205
			6,935

4.4.2 Bag-of-Words

To classify the two data corpora of external communication of savings and cooperative banks, we use an adapted bag-of-words approach, which we refer to as contextual bag-of-words approach (ConBoW) in the following. Normally BoW classifies words or phrases of a text into different categories, for example positive or negative (F. Li, 2010, p. 146; Loughran & McDonald, 2015, p. 1). Based on random checks of the matches of the created AI dictionary, we find that a large number of matches have ambiguous meanings with respect to AI, as they are used both in the application of AI and in misclassifications. Therefore, we created a Python program that gives us the context, in the form of 10 words before and after each match. Based on this analysis, all matches were manually classified into the categories shown and defined in Tables 4.4 and 4.5. The individual categories and subcategories were continuously developed based on the ongoing classification and adapted to the respective matches. This approach contradicts a classical bag-of-words approach since it disregards the order of the words to be analyzed (Algaba et al., 2020, p. 520; Loughran & McDonald, 2016, p. 1199). However, since there are approaches that take into account the use of negations and thus the order of the words (Loughran & McDonald, 2011, pp. 40, 44; Shapiro et al., 2022, pp. 225, 228), the approach we have chosen can still be considered a bag-of-words approach.

Table 4.4
Main Categories for Classification

Categories	Definition
Application	The bank uses AI or sells or provides products that apply or incorporate AI.
Article	The bank informs about AI topics.
Link	Link which uses AI terms and redirects to AI-relevant pages (Application or Article).
Misclassification	Terminology is not used in an AI context.
Incorrect data	Context does not consist of words of the German language.

Appropriate subcategories are created exclusively for the "Application" category, since the aim of this work is to find out whether savings and cooperative banks use AI.

Table 4.5
Subcategories for Classification

Categories	Definition
Automated decision	Use of AI to make automated decisions
Banking app	Offering a banking app with AI features
Banking software	Offering banking software with AI features
Chatbot	Using a chatbot based on AI
External app	Offering a third-party app with AI features
External service	Offering external services using AI
Robo-advisor	Offering a robo-advisor based on AI
Robot	Using humanoid robots
Robotic Process Automation	Using Robotic Process Automation based on AI
Voice assistant	Offering a voice assistant based on AI
Other	Other applications of AI

In order to provide an overview of the basic application or potential application of AI in savings and cooperative banks, all active applications are first classified in the listed subcategories. However, to show exactly which banks are using artificial intelligence themselves, a more detailed classification will be made later.

4.5 Results

4.5.1 Summary Statistics

Overall, 1,008,772 out of 2,217,269 files contain AI-relevant terms from our word list (AI files). This equals a fraction of 45.5%. A summary of the analysis additionally subdivided into savings and cooperative banks, can be found in Table 4.6. It should be noted here that this is an analysis prior to the actual classification. A classification into the categories presented above has not yet been made and will be presented in Table 4.7.

Table 4.6
Summary AI Files

	Combined		Savings Banks		Cooperative Banks	
	All files	AI files	All files	AI files	All files	AI files
Overall	2,217,269	1,008,772	877,036	794,419	1,340,233	214,353
HTML	2,075,033	971,918	820,629	785,458	1,254,404	186,460
PDF	142,236	36,854	56,407	8,961	85,829	27,893
Banks	1,156	1,143	370	370	786	773

Notably, the use of AI terms is proportionately much lower among cooperative banks than among savings banks. However, the high fraction at savings banks is primarily characterized by the references in HTML files. In contrast, cooperative banks have a higher proportion of relevant PDF files. Since the number of HTML files at both savings and cooperative banks far exceeds the number of PDF files, this results in the widely differing proportions of files containing AI-relevant terms.

The number of documents containing AI terms categorized as "application" is much lower, as shown in Table 4.7.

Table 4.7
Documents Classified as Application

	Combined		Savings Banks		Cooperative Banks	
	AI files	Application	AI files	Application	AI files	Application
Overall	1,008,772	115,570	794,419	90,924	214,353	24,646
HTML	971,918	112,628	785,458	89,885	186,460	22,743
PDF	36,854	2,942	8,961	1,039	27,893	1,903
Banks	1,143	1,127	370	370	773	757

The majority of documents containing AI terms that can be classified as "application" and therefore considered as AI potentially being used, are HTML files. Despite the smaller number of banks, savings banks have a much larger share of relevant documents (78.7%). Interestingly the fraction of documents of the initial classification by the dictionary that are manually classified as "application" is almost the same for cooperative banks (11.5%) as for savings banks (11.4%). There are more cooperative banks using AI-relevant terms than banks using terms classified as "application". In contrast, all savings banks use both AI-relevant terms and values classified as "application".

Due to partially several matches per document, the number of matches exceeds the number of documents. An overview can be taken from Table 4.8.

Table 4.8
Summary Matches AI / Matches Application

	Combined		Savings Banks		Cooperative Banks	
	AI files	Application	AI files	Application	AI files	Application
Overall	3,055,550	211,634	2,391,577	165,261	663,973	46,373
HTML	2,881,296	204,418	2,362,917	160,334	518,379	44,084
PDF	174,254	7,216	28,660	4,927	145,594	2,289

Only a small fraction of all matches can be classified as "application" (savings banks 6.9%, cooperative banks 7.0%). This is also reflected in the number of matches per file. The average number of matches for all files containing AI words is 3.0 for savings banks and 3.1 for cooperative banks. When classifying all files, additionally to the immense reduction of files shown in Table 4.7, the average number of words per file decreases to 1.8 for savings banks and 1.9 for cooperative banks.

The analysis presented above could lead to the conclusion that savings banks use AI to a greater extent than cooperative banks. However, this statement needs to be evaluated in more detail, based on subcategories, and is examined in section 4.5.2. Based on these initial results, it can be stated that due to the low percentage of documents and matches in the manual classification of the category "application", a revision of the constructed dictionary will be necessary.

4.5.2 Application of AI

Table 4.9 shows the composition of the manual classification of all matches that are assessed as "application" into the subcategories shown in Table 4.5.

The number of relevant matches differs significantly, with banking apps and chatbots having by far the most matches. However, not all subcategories that represent an application can be considered a true application in their own regard. Banking apps are used by all savings and cooperative banks[13], but are not developed by them, instead they are provided centrally. Therefore, the features based on AI, like photo money transfer, cannot be directly attributed. The same applies to banking software, external apps, external services, and partially for voice

[13] 34 cooperative banks use banking apps but do not describe them, using dictionary words and are therefore not listed in Table 4.9.

assistants. As a result, only the remaining six categories and, in some cases, voice assistants will be analyzed in more detail in the subsequent course of the paper.

Table 4.9
Subcategories of “Applications”

Categories	Combined		Savings Banks		Cooperative Banks	
	Matches	Banks	Matches	Banks	Matches	Banks
Automated decision	1,014	366	814	322	200	44
Banking app	102,299	1,122	61,216	370	41,083	752
Banking software	11,202	378	11,190	370	12	8
Chatbot	94,418	261	90,759	223	3,659	39
External app	175	53	175	53	0	0
External service	1,223	31	89	23	1,134	8
Robo-advisor	412	3	409	2	3	1
Robot	20	6	18	4	2	2
Robotic Process Automation	152	20	53	4	99	16
Voice assistant	204	19	192	15	12	4
Other	515	35	346	11	169	24

The subcategory of automated decisions represents a gray area. 366 banks state that they use automated individual case decisions in various areas, but only Sparda-Bank München eG and TeamBank AG Nürnberg can be assumed to use AI with certainty. For all other banks, AI use is possible, but not verifiable based on the matches found. While Sparda-Bank München eG directly states that AI is used in the context of the digital customer file, TeamBank AG Nürnberg only states the general use of AI in automated decisions.

One of the most frequently used categories is chatbots, which are used by 261 banks. It is striking that the number of savings banks using chatbots is much higher than that of cooperative banks, despite the smaller number of institutions. Chatbots are systems that can communicate with human users using natural language. Through the use of AI and the associated improvements, as made widely available by ChatGPT in late 2022, it is possible to replace human chat service agents (Adam et al., 2021, p. 427; Teubner et al., 2023, p. 95). All savings banks use a chatbot named "Linda" that is, according to the Kreissparkasse Ravensberg, based on artificial intelligence. The cooperative banks on the other hand use 32 differently named chatbots. Of the 38 cooperative banks using chatbots, seven banks state that their bots are based on AI. Six savings and cooperative banks had already discontinued the use of their chatbots at the time of the analysis.

"Robo-advisors are digital interfaces that guide investors through an entirely automated process of investment advisory from assessing financial goals, evaluating consumers' risk profile, and ultimately managing the entire portfolio" (Hildebrand & Bergner, 2021, p. 659). They can base those tasks of giving financial advice on artificial intelligence. When based on AI, robo-advisors exceed many human abilities in the process of giving financial advice (Flavián et al., 2022, pp. 295f., 300). AI-based robo-advisors represent a niche product, as they are only used by two savings banks and one cooperative bank. Sparkasse Duisburg and Bremen use Smavesto and GLS Gemeinschaftsbank eG uses GLS onlineinvest. Other robo-advisor applications are available, but they are not based on artificial intelligence.

Humanoid robots in the form of Pepper are used by all six banks in the subcategory robots. Pepper is a small social humanoid robot with voice recognition abilities, being able to recognize faces and human emotions (Søraa et al., 2021, p. 207). However, the use of Pepper cannot be classified as a real use of AI, as it is primarily used for show purposes or marketing.

Robotic Process Automation (RPA) is a collective term for tools that mimic human employee behavior by operating on the user interface of a computer system. The system itself remains unchanged and therefore RPA enables the rapid automatization of simple and repetitive tasks (Herm et al., 2021, p. 289; van der Aalst et al., 2018, p. 269). Without the application of artificial intelligence, RPA comes to a point where the development of a program becomes inefficient due to complex variants and rules (Herm et al., 2021, p. 290). Similar to the considerations around automated decisions, in most cases it is not clear to what extent the 20 banks using RPA do so with or without AI. Only PSD Bank Nürnberg eG and VR Bank Südpfalz eG mention the use of artificial intelligence in the context of RPA.

Most banks use external solutions for voice assistants such as Alexa Skills or Google Homeconnect, which, analogous to the previous explanations, are not classified as their own application of AI. Only six savings banks use a voice computer as part of their telephone banking. Based on the statements made by the banks, it can be assumed that AI is used in the context of telephone banking.

The applications of the 35 banks classified as "Other" relate primarily to building services issues, such as heating or power supply. Sparkasse Bonn uses AI terms to describe its digital experience world. Exceptions are Sparkasse Bamberg and Volksbank RheinAhrEifel eG, which say they use AI but do not define this in more detail. Among the cooperative banks, DZ Bank stands out, stating that, in addition to AI initiatives, it intends to further expand AI and use it to manage large data volumes.

Therefore, it can be stated that the use of AI at savings and cooperative banks is primarily limited to chatbots. Other areas of application, such as automated decisions, robo-advisors, Robotic Process Automation, or own voice assistants tend to be the exception. Table 4.10 summarizes the results presented above.

Table 4.10
Subcategories Containing own "Applications"

Categories	Combined		Savings Banks		Cooperative Banks	
	Matches	Banks	Matches	Banks	Matches	Banks
Automated decision	4	2	-	-	4	2
Chatbot	94,218	256	90,728	223	3,490	33
Robo-advisor	112	3	109	2	3	1
Robotic Process Automation	6	2	-	-	6	2
Voice assistant	19	6	19	6	-	-
Other	25	4	16	2	9	2

4.5.3 AI-Words used

By manually classifying the matches into the category "application", a total of 100 of the 6,935 words in the dictionary were used. The subsequent further classification regarding the own application of AI, as shown in Table 4.10, leads to a word list of 36 relevant words. It should be emphasized that all words were found within the first three iterations of our Word2vec approach. Therefore, a reduction of iterations to three, following Du et al. (2021, p. 10), is recommended. All occurrences of relevant words were manually reviewed again. This identified 11 words that were responsible for the majority of misclassifications. However, the additional removal of these words has a minimal impact on the final number of banks, since documents containing these words were sometimes classified as "application" twice. Exclusively Sparkasse Vest Recklinghausen with its telephone banking under the category "voice assistant" is excluded when deleting these 11 words. For this reason, those 11 words were removed from the final list. This leaves 25 relevant words for the analysis of the application of AI. Although the words were not found in the entire analysis corpus, a total of 11 inflections are added to these 25 identified words to enable the use of the dictionary in future analyses. This finalized dictionary will be referred to as the "Final AI Dictionary" in the further course of this paper. All changes due to the use of the Final AI Dictionary are highlighted in bold and underlined in Table 4.11.

Table 4.11
Subcategories Containing own "Applications" – Final AI Dictionary

Categories	Combined		Savings Banks		Cooperative Banks	
	Matches	Banks	Matches	Banks	Matches	Banks
Automated decision	4	2	-	-	4	2
Chatbot	**89,044**	256	**85,560**	223	**3,484**	33
Robo-advisor	112	3	109	2	3	1
Robotic Process Automation	6	2	-	-	6	2
Voice assistant	**18**	**5**	**18**	**5**	-	-
Other	**22**	4	16	2	**6**	2

Due to the use of the Final AI Dictionary, the matches to be classified have been reduced from 3,055,550 to 204,848. Despite this reduction of 93.3%, manual classification is still necessary, although to a much lesser extent. Table 4.12 gives an overview of the final composition of the dictionary.

Table 4.12
Final AI Dictionary

Terms		Added Flexions
artificialintelligence	roboter	chatterbot
chatbot	roboterassistent	finanzchatbot
chatbots	robotergestuetzter	kuenstlicherintelligenzen
chatterbots	servicebot	roboterassistenten
finanzchatbots	sprachassistent	robotergestuetzte
intelligente	sprachassistenten	robotergestuetzten
intelligentesoftware	spracherkennung	servicebots
ki	sprachroboter	spracherkennungen
kuenstlicheintelligenz	virtuelleassistenten	sprachrobotern
kuenstlichenintelligenz	virtuellenassistenten	virtuellerassistenten
kuenstlichenintelligenzen	virtuellerassistent	voicebot
kuenstlicherintelligenz	voicebots	
roboadvisor		

Due to the composition of words in the dictionary, an additional merging of words such as "artificial intelligence" to "artificialintelligence" should be done as part of the parsing, described in 4.3.2. By optimizing the dictionary, we are able to successfully accomplish the word list revision required by 4.5.1 and reduce the percentage of documents found that were not manually classified as own applications from 96.9% to 56.5%.

4.6 Conclusion

This paper uses a combination of Word2Vec and bag-of-words to create a dictionary for analyzing the external communication of savings and cooperative banks. In order to answer the question of the extent to which savings and cooperative banks use artificial intelligence, we use this dictionary as an alternative to surveys.

Based on four trained Word2vec models and a sample of 475 seed words, we are able to create a comprehensive dictionary of 6,935 elements. We use this dictionary in a contextual bag-of-words approach to manually classify 3,055,550 matches into the respective categories. This initial classification of the application of AI is further analyzed and reviewed to get a final sample of 89,206 matches for the own application of AI by savings and cooperative banks. By analyzing the final classification, we are successfully able to concentrate our initial dictionary on 36 AI-relevant terms and therefore reduce the matches to be classified by 93.3% without worsening the classification results.

Based on our approach, we can demonstrate a very moderate usage of AI at savings and cooperative banks. This application is primarily limited to the use of chatbots belonging to the field of customer service and experience, which belongs to significant future developments. Other fields of application, such as automated decisions, robo-advisors, RPA, or voice assistants represent niche applications of individual banks. Other future significant topics like crime detection and prevention or credit underwriting and risk assessment could not be found.

By using our combined approach of Word2vec and contextual bag-of-words, this paper is able to contribute a structured approach as an alternative to surveys, without being subject to challenges, such as the non-response bias. Additionally, the developed final AI Dictionary can be used for analyzing other German banking groups like private banks or public sector banks. Furthermore, a frequent classification of savings and cooperative banks based on the presented methodology is of interest to analyze the change in the use of artificial intelligence over time.

5. Conclusions

This thesis focuses on text analysis as an important part of accounting and finance research. Significant reforms, extensions and clarifications are proposed to further improve the analysis of German-speaking financial texts, based on the first available finance-related dictionary for the German language by Bannier et al. (2019b). In addition, the importance of text analysis in the form of the dictionary approach is emphasized as an alternative to surveys.

Using the most comprehensive collection of German CEO speeches we are able to contribute two very important insights in chapter 2, by analyzing major assumptions of sentiment analysis using the six most common measurements of sentiment in financial texts. Exclusively a combination of the modified dictionary BWP_N and the relative measurement *NTone* is able to fulfill the central assumptions, that speeches with a more positive measurement of sentiment lead to higher abnormal returns and that it is possible to separate above and below average abnormal returns through the use of those sentiment measures. These findings are reinforced by highly statistically significant results for several regressions.

The findings of the previous chapter are expanded on in chapter 3, by further increasing the scope of the dictionary by 11,179 words through adding basic forms and synonyms for existing words of the BPW_N. Additionally, we found 442 words that did not correspond to the assigned categories due to their meaning. By computing the sentiment of German-speaking annual reports using this further expanded and corrected BPW_E dictionary, we are able to successfully test the stated central assumptions and improve regression results. In a complimentary approach, we analyze the employment of negations by using four different lists of negations, as well as two different approaches of implementation. We are able to show that the usage of negations can improve, but not deteriorate results and should therefore be implemented in the form of term shifted PR negations.

Therefore, chapters 2 and 3 can give three important contributions to the field of textual analysis in the financial domain through verifying the superiority of the characteristics of the three parameters choice of a word list, measurement of sentiment and usage of negations.

The fourth chapter demonstrates the broad range of applications for text analysis by examining the external communication of savings and cooperative banks as an alternative for surveys to determine the extent to which these banks use artificial intelligence. We train four different neural networks based on Word2vec on a sample of 475 seed words in 12 iterations to create a comprehensive dictionary of 6,935 elements. Through further manual classification based on the respective references in the text corpus, we are able to reduce the relevant words for the identification of the usage of artificial intelligence to 36 relevant terms. Based on our approach, we can demonstrate a very moderate usage of AI at savings and cooperative banks, primarily limited to the use of chatbots. The demonstrated combined approach of Word2Vec and contextual bag-of-words can contribute as a structured alternative to surveys, which can be adapted to the respective research purpose to be investigated.

Based on the findings of this thesis further research on finance related texts should be conducted by calculating the sentiment measurement *NTone* using the BPW_E in combination with term shifted PR negations. Due to the novelty of the proposed approach, the methodology should be applied to other financial publications such as conference calls or earnings press releases. In addition, the approach presented in the fourth chapter should be applied to other aspects, such as the importance of sustainability for German financial institutions.

References

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31 (2), pp. 427–445.

Ahmed, Y., & Elshandidy, T. (2016). The effect of bidder conservatism on M&A decisions: Text-based evidence from US 10-K filings. *International Review of Financial Analysis* (46), pp. 176–190.

Algaba, A., Ardia, D., Bluteau, K., Borms, S., & Boudt, K. (2020). Econometrics meets sentiment: An overview of methodology and applications. *Journal of Economic Surveys*, 34 (3), pp. 512–547.

Allee, K. D., & Deangelis, M. D. (2015). The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion. *Journal of Accounting Research*, 53 (2), pp. 241–274.

Alshorman, S. A. A., & Shanahan, M. (2022). The voice of profit: exploring the tone of Australian CEO's letters to shareholders after the global financial crisis. *Corporate Communications: An International Journal*, 27 (1), pp. 127–147.

Aly, D., El-Halaby, S., & Hussainey, K. (2018). Tone disclosure and financial performance: evidence from Egypt. *Accounting Research Journal*, 31 (1), pp. 63–74.

Ammann, M., & Schaub, N. (2016). *Social Interaction and Investing: Evidence from an Online Social Trading Network*. Working Paper, retrieved on 11.07.2018 from https://www.rsm.nl/fileadmin/home/Department_of_Finance__VG5_/PAM2016/Final_Papers/Nic_Schaub.pdf

Anseel, F., Lievens, F., Schollaert, E., & Choragwicka, B. (2010). Response Rates in Organizational Science, 1995–2008: A Meta-analytic Review and Guidelines for Survey Researchers. *Journal of Business and Psychology*, 25 (3), pp. 335–349.

Apel, M., & Blix Grimaldi, M. (2012). *The Information Content of Central Bank Minutes* (Sveriges Riksbank Working Paper Series No. 261). Sveriges Riksbank, Stockholm, retrieved on 13.02.2020 from http://archive.riksbank.se/Documents/Rapporter/Working_papers/2012/rap_wp261_120426.pdf

Asudani, D. S., Nagwani, N. K., & Singh, P. (2023). Impact of word embedding models on text analytics in deep learning environment: A review. *Artificial Intelligence Review*, pp. 1–81.

Bannier, C. E., Pauls, T., & Walter, A. (2017). *CEO-Speeches and Stock Returns*. Working Paper, retrieved on 15.08.2019 from https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3051151_code1882913.pdf?abstractid=3051151&mirid=1

Bannier, C. E., Pauls, T., & Walter, A. (2018). *Content analysis of business specific text documents: Introducing a German dictionary*. Working Paper, retrieved on 29.10.2023 from https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3105884_code2184042.pdf?abstractid=2961820&mirid=1

Bannier, C. E., Pauls, T., & Walter, A. (2019a). *The Annual General Meeting revisited: The role of the CEO speech*. Working Paper, retrieved on 11.12.2021 from https://ssrn.com/abstract=2869785

Bannier, C. E., Pauls, T., & Walter, A. (2019b). Content analysis of business specific text documents: Introducing a German dictionary. *Journal of Business Economics,* 89 (1), pp. 79–123.

Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, 61 (8), pp. 1139–1160.

Berns, J., Bick, P., Flugum, R., & Houston, R. (2022). Do changes in MD&A section tone predict investment behavior? *Financial Review*, 57 (1), pp. 129–153.

Bochkay, K., Hales, J., & Chava, S. (2020). Hyperbole or Reality? Investor Response to Extreme Language in Earnings Conference Calls. *The Accounting Review*, 95 (2), pp. 31–60.

Borochin, P. A., Cicon, J. E., DeLisle, R. J., & Price, S. M. (2018). The effects of conference call tones on market perceptions of value uncertainty. *Journal of Financial Markets*, 40, pp. 75–91.

Bos, T., & Frasincar, F. (2022). Automatically Building Financial Sentiment Lexicons While Accounting for Negation. *Cognitive Computation*, 14, pp. 442–460.

Boudt, K., & Thewissen, J. (2019). Jockeying for Position in CEO Letters: Impression Management and Sentiment Analytics. *Financial Management*, 48 (1).

Brau, J. C., Cicon, J., & McQueen, G. (2016). Soft Strategic Information and IPO Underpricing. *Journal of Behavioral Finance*, 17 (1), pp. 1–17.

Bulwer-Lytton, E. (1839). *Richelieu, or, The conspiracy: a play, in five acts ; to which are added, Historical odes on The last days of Elizabeth, Cromwell's dream, the death of Nelson* (2nd edition), Saunders and Otley.

Bundesverband der Deutschen Volksbanken und Raiffeisenbanken. (2022). *Alle Genossenschaftsbanken per Ende 2021*, retrieved on 31.02.2022 from https://www.bvr.de/Presse/Zahlen_Daten_Fakten

Bushman, R. M., Hendricks, B. E., & Williams, C. D. (2016). Bank Competition: Measurement, Decision-Making, and Risk-Taking. *Journal of Accounting Research*, 54 (3), pp. 777–826.

Chakraborty, B., & Bhattacharjee, T. (2020). A review on textual analysis of corporate disclosure according to the evolution of different automated methods. *Journal of Financial Reporting and Accounting*, 18 (4), pp. 757–777.

Chen, Q., & Sokolova, M. (2021). Specialists, Scientists, and Sentiments: Word2vec and Doc2Vec in Analysis of Scientific and Medical Texts. *SN Computer Science,* 2 (5), pp. 414.

Correa, R., Garud, K., Londono, J. M., & Mislang, N. (2021). Sentiment in Central Banks' Financial Stability Reports. *Review of Finance*, 25 (1), pp. 85–120.

Da Tonin, J. M. F., & Scherer, L. M. (2022). Market reaction to the tones of earnings conference calls. *Journal of Business Management*, 62 (1), Article e2020-0301, pp. 1–18.

Daniel, F., Lohrke, F. T., Fornaciari, C. J., & Turner, R. (2004). Slack resources and firm performance: a meta-analysis. *Journal of Business Research*, 57 (6), pp. 565–574.

Das, S. R., Donini, M., Zafar, M. B., He, J [John], & Kenthapadi, K. (2022). FinLex: An effective use of word embeddings for financial lexicon generation. *The Journal of Finance and Data Science*, 8, pp. 1–11.

Davenport, T. H., Ronanki, R., & others (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96 (1), pp. 108–116.

Davis, A. K., Ge, W., Matsumoto, D., & Zhang, J. L. (2015). The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies*, 20 (2), pp. 639–673.

Davis, A. K., Piger, J. M., & Sedor, L. M. (2012). Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language. *Contemporary Accounting Research*, 29 (3), pp. 845–868.

Davis, A. K., & Tama-Sweet, I. (2012). Managers' Use of Language Across Alternative Disclosure Outlets: Earnings Press Releases versus MD&A. *Contemporary Accounting Research*, 29 (3), pp. 804–837.

Deloitte. (2021). *Artificial Intelligence: Transforming the future of banking*, retrieved on 14.10.2023 from https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/financial-services/Banking/lu-ai-transforming-future-of-banking.pdf

Deutsche Bank. (2023). *How AI is changing banking*, retrieved on 14.10.2023 from https://www.db.com/what-next/digital-disruption/better-than-humans/how-artificial-intelligence-is-changing-banking/index?language_id=1

Deutsche Bundesbank. (2023). *Bankenstatistiken: Aktualisierte Ausgabe*, retrieved on 06.04.2023 from https://www.bundesbank.de/resource/blob/803956/49838dfbcd0420b42b494447e17f40aa/mL/0-bankenstatistiken-data.pdf

Deutscher Sparkassen- und Giroverband. (2022). *Rangliste der Sparkassen 2021*, retrieved on 12.04.2022 from https://www.dsgv.de/sparkassen-finanzgruppe/organisation/sparkassen.html

Dietzmann, C., Jaeggi, T., & Alt, R. (2023). Implications of AI-based robo-advisory for private banking investment advisory. *Journal of Electronic Business & Digital Economics*, 2 (1), pp. 3–23.

Doran, J. S., Peterson, D. R., & Price, M. S. (2012). Earnings Conference Call Content and Stock Price: The Case of REITs. *Journal of Real Estate Finance and Economics*, 45 (2), pp. 402–434.

Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., & Kammler, J. (2016). Description-text related soft information in peer-to-peer lending: Evidence from two leading European platforms. *Journal of Banking & Finance* (64), pp. 169–187.

Doumpos, M., Zopounidis, C., Gounopoulos, D., Platanakis, E., & Zhang, W. (2023). Operational research and artificial intelligence methods in banking. *European Journal of Operational Research*, 306 (1), pp. 1–16.

Drubin, D. G., & Kellogg, D. R. (2012). English as the universal language of science: Opportunities and challenges. *Molecular Biology of the Cell*, 23 (8), pp. 1399.

Druz, M., Petzev, I., Wagner, A. F., & Zeckhauser, R. J. (2020). When Managers Change Their Tone, Analysts and Investors Change Their Tune. *Financial Analysts Journal*, 76 (2), pp. 47–69.

Du, Z., Huang, A. G., Wermers, R., & Wu, W. (2021). Language and Domain Specificity: A Chinese Financial Sentiment Dictionary. *Review of Finance*, pp. 1–47.

European Central Bank. (2023a). *Banks' digital transformation: where do we stand?* European Central Bank, retrieved on 18.01.2024 from https://www.bankingsupervision.europa.eu

/press/publications/newsletter/2023/html/ssm.nl230215_2.en.html#/search/artificial%20intelligence/1

European Central Bank. (2023b). *Bringing artificial intelligence to banking supervision*, retrieved on 14.10.2023 from https://www.bankingsupervision.europa.eu/press/publications/newsletter/2019/html/ssm.nl191113_4.en.html

European Central Bank. (2023c). *Take-aways from the horizontal assessment of the survey on digital transformation and the use of fintech*, retrieved on 18.01.2024 from https://www.bankingsupervision.europa.eu/ecb/pub/pdf/Takeaways_horizontal_assessment~de65261ad0.en.pdf

Evident Insights. (2023). *The Evident AI Index Key Findings report*, retrieved on 27.08.2023 from https://evidentinsights.com/ai-index/

Fares, O. H., Butt, I., & Lee, S. H. M. (2022). Utilization of artificial intelligence in the banking sector: a systematic literature review. *Journal of Financial Services Marketing.* Advance online publication.

Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. (2008). *The Incremental Information Content of Tone Change in Management Discussion and Analysis* (NYU Working Paper No. JOSHUA LIVNAT-09), retrieved on 11.07.2018 from https://ssrn.com/abstract=1301338

Ferguson, N. J., Philip, D., Lam, H. Y. T., & Guo, J. M. (2015). Media Content and Stock Returns: The Predictive Power of Press. *Multinational Finance Journal*, 19 (1), pp. 1–31.

Ferris, S. P., Hao, Q., & Liao, M.-Y. (2013). The Effect of Issuer Conservatism on IPO Pricing and Performance. *Review of Finance*, 17 (3), pp. 993–1027.

Flavián, C., Pérez-Rueda, A., Belanche, D., & Casaló, L. V. (2022). Intention to use analytical artificial intelligence (AI) in services – the effect of technology readiness and awareness. *Journal of Service Management*, 33 (2), pp. 293–320.

Franke, B. (2018). *Qualitative Information and Loan Terms: A Textual Analysis*. Working Paper, retrieved on 15.09.2019 from https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3209201_code1660824.pdf?abstractid=3152458&mirid=1

Frankel, R., Jennings, J., & Lee, J. (2022). Disclosure Sentiment: Machine Learning vs. Dictionary Methods. *Management Science*, 68 (7), pp. :5514-55.

Fritz, D., & Tows, E. (2018). Text Mining and Reporting Quality in German Banks: A Cooccurrence and Sentiment Analysis. *Universal Journal of Accounting and Finance*, 6 (2), pp. 54–81.

Garcia, D. (2013). Sentiment during Recessions. *The Journal of Finance*, 68 (3), pp. 1267–1300.

Garfield, E., & Welljams-Dorof, A. (1990). Language Use in International Research: A Citation Analysis. *The ANNALS of the American Academy of Political and Social Science*, 511 (1), pp. 10–24.

Gastaldi, J. L. (2021). Why Can Computers Understand Natural Language? *Philosophy & Technology*, 34 (1), pp. 149–214.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, 57 (3), pp. 535–574.

González, M., Guzmán, A., Téllez, D. F., & Trujillo, M. A. (2019). What you say and how you say it: Information disclosure in Latin American firms. *Journal of Business Research*, 127 (3), pp. 427–443.

Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance*, 32, pp. 100577.

Google. (2023). *Google Trends*, retrieved on 05.04.2023 from https://trends.google.de/trends/explore?date=all&q=%2Fm%2F0mkz&hl=de

Grover, P., Kar, A. K., & Dwivedi, Y. K. (2022). Understanding artificial intelligence adoption in operations management: insights from the review of academic literature and social media discussions. *Annals of Operations Research*, 308 (1-2), pp. 177–213.

Gupta, A., Dengre, V., Kheruwala, H. A., & Shah, M. (2020). Comprehensive review of text-mining applications in finance. *Financial Innovation*, 6 (39).

Gurun, U. G., & Butler, A. W. (2012). Don’t Believe the Hype: Local Media Slant, Local Advertising, and Firm Value. *The Journal of Finance*, 67 (2), pp. 561–598.

Handelsblatt. (2023). *wiso database*, retrieved on 05.04.2023 from www.wiso-net.de

Hart, R. P. (2000). *DICTION 5.0*, retrieved on 09.06.2020 from https://rhetorica.net/diction.htm

Henry, E. (2006). Market Reaction to Verbal Components of Earnings Press Releases: Event Study Using a Predictive Algorithm. *Journal of Emerging Technologies in Accounting*, 3, pp. 1–19.

Henry, E. (2008). Are Investors Influenced By How Earnings Press Releases Are Written? *Journal of Business Communication*, 45 (4), pp. 363–407.

Henry, E., & Leone, A. J. (2016). Measuring Qualitative Information in Capital Markets Research: Comparison of Alternative Methodologies to Measure Disclosure Tone. *The Accounting Review*, 91 (1), pp. 153–178.

Henry, E., Thewissen, J., & Torsin, W. (2021). International Earnings Announcements: Tone, Forward-looking Statements, and Informativeness. *European Accounting Review*, pp. 1–35.

Herm, L.-V., Janiesch, C., Reijers, H. A., & Seubert, F. (2021). From Symbolic RPA to Intelligent RPA: Challenges for Developing and Operating Intelligent Software Robots. In A. Polyvyanyy, M. T. Wynn, A. van Looy, & M. Reichert (Eds.), *Lecture Notes in Computer Science. Business Process Management, Vol.* 12875, pp. 289–305, Springer International Publishing.

Hildebrand, C., & Bergner, A. (2021). Conversational robo advisors as surrogates of trust: onboarding experience, firm perception, and consumer financial decision making. *Journal of the Academy of Marketing Science,* 49 (4), pp. 659–676.

Hillert, A., Jacobs, H., & Müller, S. (2018). Journalist disagreement. *Journal of Financial Markets*, 41, pp. 57–76.

Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone Management. *The Accounting Review*, 89 (3), pp. 1083–1113.

Iqbal, J., & Riaz, K. (2022). Predicting future financial performance of banks from management’s tone in the textual disclosures. *Quality & Quantity* (56), pp. 2691–2721.

Jandl, J.-O., Feuerriegel, S., & Neumann, D. (2014). *Long- and Short-Term Impact of News Messages on House Prices: A Comparative Study of Spain and the United States* (Thirty Fifth International Conference on Information Systems), Auckland, retrieved on 15.09.2019 from https://aisel.aisnet.org/icis2014/proceedings/DecisionAnalytics/17/

Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. *Journal of Financial Economics*, 110 (3), pp. 712–729.

Kale, A. S., Pandya, V., Di Troia, F., & Stamp, M. (2023). Malware classification with Word2Vec, HMM2Vec, BERT, and ELMo. *Journal of Computer Virology and Hacking Techniques*, 19 (1), pp. 1–16.

Kang, T., Park, D.-H., & Han, I. (2018). Beyond the numbers: The effect of 10-K tone on firms' performance predictions using text analytics. *Telematics and Informatics*, 35 (2), pp. 370–381.

Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics*, 7 (2), pp. 139–172.

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), pp. 15–25.

Kaushal, N., Kaurav, R. P. S., Sivathanu, B., & Kaushik, N. (2021). Artificial intelligence and HRM: identifying future research Agenda using systematic literature review and bibliometric analysis. *Management Review Quarterly*. Advance online publication.

Kaya, D., Maier, C., & Böhmer, T. (2020). Empirische Kapitalmarktforschung zu Conference Calls: Eine Literaturanalyse. *Schmalenbachs Zeitschrift Für Betriebswirtschaftliche Forschung*, 72, pp. 183–212.

Kearney, C., & Liu, S. (2014). Textual Sentiment in Finance: A Survey of Methods and Models. *International Review of Financial Analysis* (33), pp. 171–185.

Keding, C. (2021). Understanding the interplay of artificial intelligence and strategic management: four decades of research in review. *Management Review Quarterly*, 71 (1), pp. 91–134.

Kim, S.-M., & Hovy, E. (2004). Determining the sentiment of opinions. In *COLING '04: Proceedings of the 20th international conference on Computational Linguistics - COLING '04*, 1367-es, Association for Computational Linguistics.

Kim, Y. H., & Meschke, F. (2014). *CEO Interviews on CNBC*. Working Paper, retrieved on 12.02.2020 from http://dx.doi.org/10.2139/ssrn.1745085

King, D. R., Dalton, D. R., Daily, C. M., & Covin, J. G. (2004). Meta-analyses of post-acquisition performance: indications of unidentified moderators. *Strategic Management Journal*, 25 (2), pp. 187–200.

Koelbl, M. (2020). Is the MD&A of US REITs informative? A textual sentiment study. *Journal of Property Investment & Finance*, 38 (3), pp. 181–201.

Krause, J., Perer, A., & Ng, K. (2016). Interacting with Predictions. In J. Kaye, A. Druin, C. Lampe, D. Morris, & J. P. Hourcade (Eds.), *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5686–5697, ACM.

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, 28 (1), pp. 40–80.

Lauren, P., Qu, G., Yang, J., Watta, P., Huang, G.-B., & Lendasse, A. (2018). Generating Word Embeddings from an Extreme Learning Machine for Sentiment Analysis and Sequence Labeling Tasks. *Cognitive Computation*, 10 (4), pp. 625–638.

Levy, O., & Goldberg, Y. (2014). Dependency-Based Word Embeddings. In K. Toutanova & H. Wu (Eds.), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 302–308, Association for Computational Linguistics.

Lewis, C., & Young, S. (2019). Fad or future? Automated analysis of financial text and its implications for corporate reporting. *Accounting and Business Research*, 49 (5), pp. 587–615.

Li, F. (2010). Textual Analysis of Corporate Disclosures: A Survey of the Literature. *Journal of Accounting Literature* (29), pp. 143–165.

Li, S., Shi, W., Wang, J., & Zhou, H. (2021). A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: a Case Study in Financial Distress Prediction. *Information Processing & Management*, 58 (5).

Liu, Y., & Alsaadi, F. E. (2020). A Novel Way to Build Stock Market Sentiment Lexicon. In J. He, P. S. Yu, Y. Shi, X. Li, Z. Xie, G. Huang, J. Cao, & F. Xiao (Eds.), *Communications in Computer and Information Science. Data Science*, *Vol.* 1179, pp. 350–361, Springer Singapore.

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66 (1), pp. 35–65.

Loughran, T., & McDonald, B. (2015). The Use of Word Lists in Textual Analysis. *Journal of Behavioral Finance*, 16 (1), pp. 1–11.

Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research* (54), pp. 1187–1230.

Loughran, T., & McDonald, B. (2020a). *Stop Words*, retrieved on 21.01.2021 from https://drive.google.com/file/d/0B4niqV00F3mseWZrUk1YMGxpVzQ/view?usp=sharing

Loughran, T., & McDonald, B. (2020b). Textual Analysis in Finance. *Annual Review of Financial Economics*, 12 (1), pp. 357–375.

Loughran, T., McDonald, B., & Yun, H. (2009). A Wolf in Sheep's Clothing: The Use of Ethics-Related Terms in 10-K Reports. *Journal of Business Ethics*, 89 (1), pp. 39–49.

Luo, Y., & Zhou, L. (2020). Textual tone in corporate financial disclosures: a survey of the literature. *International Journal of Disclosure and Governance*, 17, pp. 101–110.

Mahmoudi, N., Olech, Ł. P., & Docherty, P. (2021). A comprehensive study of domain-specific emoji meanings in sentiment classification. *Computational Management Science.* Advance online publication.

Mauro, A. de, Sestino, A., & Bacconi, A. (2022). Machine learning and artificial intelligence use in marketing: a general taxonomy. *Italian Journal of Marketing*, 2022 (4), pp. 439–457.

Mayew, W. J., & Venkatachalam, M. (2012). The Power of Voice: Managerial Affective States and Future Firm Performance. *The Journal of Finance*, 67 (1), pp. 1–43.

Meier, T., Boyd, R. L., Pennebaker, J. W., Mehl, M. R., Martin, M., Wolf, M., & Horn, A. B. (2018). *"LIWC auf Deutsch": The Development, Psychometrics, and Introduction of DE-LIWC2015*, retrieved on 08.03.2019 from https://osf.io/tfqzc/

Mengelkamp, A., Wolf, S., & Schumann, M. (2016). *Data Driven Creation of Sentiment Dictionaries for Corporate Credit Risk Analysis* (Proceedings of the 22nd Americas Conference on Information Systems (AMCIS)), retrieved on 10.07.2018 from https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1058&context=amcis2016

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*, retrieved on 23.02.2022 from http://arxiv.org/pdf/1301.3781v3

Milana, C., & Ashta, A. (2021). Artificial intelligence techniques in finance and financial markets: A survey of the literature. *Strategic Change*, 30 (3), pp. 189–209.

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access*, 8, pp. 131662–131682.

Myšková, R., & Hájek, P. (2020). Mining risk-related sentiment in Corporate annual reports and its effect on financial performance. *Technological and Economic Development of Economy*, 26 (6), pp. 1422–1443.

Naili, M., Chaibi, A. H., & Ben Ghezala, H. H. (2017). Comparative study of word embedding methods in topic segmentation. *Procedia Computer Science*, 112, pp. 340–349.

OliverWyman. (2022). *The AI Revolution in Banking*, retrieved on 14.10.2023 from https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2022/sept/ai-revolution-in-banking-report.pdf

Patelli, L., & Pedrini, M. (2014). Is the Optimism in CEO’s Letters to Shareholders Sincere? Impression Management Versus Communicative Action During the Economic Crisis. *Journal of Business Ethics*, 124, pp. 19–34.

Pengnate, S., Lehmberg, D. G., & Tangpong, C. (2020). Top management's communication in economic crisis and the firm's subsequent performance: sentiment analysis approach. *Corporate Communications: An International Journal*, 25 (2), pp. 187–205.

Picault, M., & Renault, T. (2017). Words are not all created equal: A new measure of ECB communication. *Journal of International Money and Finance* (79), pp. 136–156.

Pöferlein, M. (2021). Sentiment Analysis of German Texts in Finance: Improving and Testing the BPW Dictionary. *Journal of Banking and Financial Economics*, 16 (2), pp. 5–24.

Price, M. S., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, 36 (4), pp. 992–1011.

Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS - a Publicly Available German-language Resource for Sentiment Analysis. *Proceedings of the 7th International Language Ressources and Evaluation (LREC'10)*, pp. 1168–1171.

Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, 84, pp. 25–40.

Rice, D. R., & Zorn, C. (2019). Corpus-based dictionaries for sentiment analysis of specialized vocabularies. *Political Science Research and Methods*, 67, pp. 1–16.

Röder, F., & Walter, A. (2019). What drives Investment Flows into Social Trading Portfolios? *The Journal of Financial Research*, 42 (2), pp. 383–411.

Rogelberg, S. G., & Stanton, J. M. (2007). Introduction: Understanding and Dealing With Organizational Survey Nonresponse. *Organizational Research Methods*, 10 (2), pp. 195–209.

Ruscheinsky, J. R., Lang, M., & Schäfers, W. (2018). Real estate media sentiment through textual analysis. *Journal of Property Investment & Finance*, 36 (5), pp. 410–428.

Sagnika, S., Mishra, B. S. P., & Meher, S. K. (2020). Improved method of word embedding for efficient analysis of human sentiments. *Multimedia Tools and Applications*, 79 (43-44), pp. 32389–32413.

Schiessl, D., Dias, H. B. A., & Korelo, J. C. (2022). Artificial intelligence in marketing: a network analysis and future agenda. *Journal of Marketing Analytics*, 10 (3), pp. 207–218.

Schmeling, M., & Wagner, C. (2016). *Does Central Bank Tone Move Asset Prices?* (The 77th Annual Meeting of American Finance Association. AFA 2017), retrieved on 29.06.2018 from https://research.cbs.dk/en/publications/does-central-bank-tone-move-asset-prices(c6401864-a921-401c-90db-57d42d6b5022).html

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, 228 (2), pp. 221–243.

Søraa, R. A., Nyvoll, P. S., Grønvik, K. B., & Serrano, J. A. (2021). Children's perceptions of social robots: a study of the robots Pepper, AV1 and Tessa at Norwegian research fairs. *AI & SOCIETY*, 36 (1), pp. 205–216.

Sreejesh, S., Mohapatra, S., & Anusree, M. R. (2014). *Business Research Methods*, Springer International Publishing.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The General Inquirer: A Computer Approach to Content Analysis*, The M.I.T. Press.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37 (2), pp. 267–307.

Tang, J., Liu, J., Zhang, M., & Mei, Q. (2016). Visualizing Large-scale and High-dimensional Data. In J. Bourdeau, J. A. Hendler, R. N. Nkambou, I. Horrocks, & B. Y. Zhao (Eds.), *Proceedings of the 25th International Conference on World Wide Web*, pp. 287–297, International World Wide Web Conferences Steering Committee.

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62 (3), pp. 1139–1168.

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance*, 63 (3), pp. 1437–1467.

Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W., & Hinz, O. (2023). Welcome to the Era of ChatGPT et al. *Business & Information Systems Engineering*, 65 (2), pp. 95–101.

Theil, C. K., Štajner, S., & Stuckenschmidt, H. (2020). Explaining Financial Uncertainty through Specialized Word Embeddings. *ACM/IMS Transactions on Data Science*, 1 (1), pp. 1–19.

Tillmann, P., & Walter, A. (2018). *ECB vs Bundesbank: Diverging Tones and policy Effectiveness* (Joint Discussion Paper Series in Economics No. 20-2018), retrieved on 13.02.2020 from https://www.uni-marburg.de/fb02/makro/forschung/magkspapers/paper_2018/20-2018_tillmann.pdf

Tillmann, P., & Walter, A. (2019). The effect of diverging communication: The case of the ECB and the Bundesbank. *Economics Letters*, 176 (C), pp. 68–74.

Torregrossa, F., Allesiardo, R., Claveau, V., Kooli, N., & Gravier, G. (2021). A survey on training and evaluation of word embeddings. *International Journal of Data Science and Analytics*, 11 (2), pp. 85–103.

Tsai, M.-F., & Wang, C.-J. (2014). Financial Keyword Expansion via Continuous Word Vector Representations. In Q. C. R. I. Alessandro Moschitti, G. Bo Pang, & U. o. A. Walter Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1453–1458, Association for Computational Linguistics.

Universität Leipzig. (2022). *Deutscher Wortschatz: Korpus News 2020*, retrieved on 15.04.2022 from https://corpora.uni-leipzig.de/de?corpusId=deu_news_2020

van der Aalst, W. M. P., Bichler, M., & Heinzl, A. (2018). Robotic Process Automation. *Business & Information Systems Engineering*, 60 (4), pp. 269–272.

Vojinović, Ž., Milutinović, S., & Leković, B. (2020). Micro-specific Profitability Factors of the Serbian Insurance Industry: A Panel Data Estimation. *Ekonomie a Management*, 23 (1), pp. 135–155.

Vytautas, K., Degrande, N., & De Weerdt, J. (2018). Using sentiment analysis to predict interday Bitcoin price movements. *The Journal of Risk Finance*, 19 (1), pp. 56–75.

Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects. *Business Process Management Journal*, 26 (7), pp. 1893–1924.

Weber, P., Carl, K. V., & Hinz, O. (2023). Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature. *Management Review Quarterly*. Advance online publication.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (pp. 347–354), retrieved on from https://aclanthology.org/H05-1044/

Wolf, M., Horn, A. B., Mehl, M. R., Haug, S., Pennebaker, J. W., & Kordy, H. (2008). Computergestützte quantitative Textanalyse: Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry and Word Count. *Diagnostica*, 54 (2), pp. 85–98.

Wuermeling, J. (2023). *Big Data - Die Verheißung unstrukturierter Daten für das Finanzwesen: Rede auf der Veranstaltung „BaFinTech 2023" der Bundesanstalt für Finanzdienstleistungsaufsicht.* Deutsche Bundesbank, retrieved on 18.01.2024 from https://www.bundesbank.de/de/presse/reden/big-data-die-verheissung-unstrukturierter-daten-fuer-das-finanzwesen-915904

Xue, F., Li, X [Xiaoyu], Zhang, T., & Hu, N. (2021). Stock market reactions to the COVID-19 pandemic: The moderating role of corporate big data strategies based on Word2Vec. *Pacific-Basin Finance Journal*, 68.

Yang, X., Macdonald, C., & Ounis, I. (2018). Using word embeddings in Twitter election classification. *Information Retrieval Journal*, 21 (2-3), pp. 183–207.

Yekrangi, M., & Abdolvand, N. (2021). Financial markets sentiment analysis: developing a specialized Lexicon. *Journal of Intelligent Information Systems*, 57, pp. 127–146.

Yuthas, K., Rogers, R., & Dillard, J. F. (2002). Communicative Action and Corporate Annual Reports. *Journal of Business Ethics*, 41 (1-2), pp. 141–157.