FAKULTÄT FÜR !NFORMATIK

Faculty of Informatics

Automatische Vorhersage von Einkommen anhand von Jahresberichten

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Wirtschaftsinformatik

eingereicht von

Jure Zuljevic, Bsc
Matrikelnummer 01640007

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Allan Hanbury, Univ.Prof.Dr.

Wien, 13. Dezember 2018

Jure Zuljevic	Allan Hanbury

FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Automatically predicting revenue from management reports

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Business Informatics

by

Jure Zuljevic, Bsc
Registration Number 01640007

to the Faculty of Informatics

at the TU Wien

Advisor: Allan Hanbury, Univ.Prof.Dr.

Vienna, 13th December, 2018

Jure Zuljevic

Allan Hanbury

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.ac.at

Erklärung zur Verfassung der Arbeit

Jure Zuljevic, Bsc
Schelleingasse 36/513 1040 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 13. Dezember 2018

Jure Zuljevic

Danksagung

Zu Beginn möchte ich mich bei Univ.Prof.Dr. Allan Hanbury für die hervorragende Unterstützung bei meiner Arbeit bedanken. Insbesondere haben die ausführlichen und hilfreichen Kommentare zu meinen Entwürfen, die Qualität dieser Arbeit sehr gesteigert.

Mein besonderer Dank gilt Dr. Alessandro D'Alconzo (Siemens AG) durch den ich das Konzept der Arbeit entwickelt habe. Außerdem stand er mir für alle meine Fragen zur Verfügung. Ich danke auch Bernhard Bauer (Siemens AG) für seine Unterstützung bei der Erstellung der Diplomarbeit in Zusammenarbeit mit der Siemens AG sowie für die Interpretation der Ergebnisse.

Zum Schluss möchte ich meiner Familie und meiner Verlobten für die uneingeschränkte Unterstützung und Ermutigung während meiner Studienjahre und meiner Diplomarbeit danken. Diese Leistung wäre ohne sie nicht möglich gewesen. Vielen Dank.

Wien, Dezember 2018, Jure Zuljevic

Acknowledgements

At the beginning I would like to thank Univ.Prof.Dr. Allan Hanbury for the excellent support during my work on this thesis. In particular, for the detailed and helpful comments on my designs, which ensured the quality of the work and especially for the responsiveness during the past months.

My special thanks go to Dr. Alessandro D'Alconzo (Siemens AG), who developed the initial idea for this work with me and has always taken time to answer my questions. I would also like to thank Bernhard Bauer (Siemens AG), for his support on writing the thesis in cooperation with Siemens AG as well as for the interpretation of results.

Finally, I must express my profound gratitude to my family and to my fiancée for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Vienna, December 2018 Jure Zuljevic

Kurzfassung

Die Analyse der eigenen Umsatzzahlen der vergangenen Jahre kann für das Management eines Unternehmens große Einblicke bieten. Trotzdem hat dieser Ansatz ein Manko. Man kann sich zwar ein Bild über die eigenen Umsatzzahlen machen, jedoch fehlt der Überblick der restlichen Wettbewerber auf dem Markt.

Um zu wissen, wie sich andere Firmen auf dem Markt schlagen, benötigen Manager auch die Infomartionen über ihre Konkurrenz. Mit solchen Informationen können sie untersuchen auf welcher Position sie stehen und welche Gründe und Ursachen ihre (Miss)erfolge haben. Die Jahresberichte bieten eine relevante und zuverlässige Darstellung der Finanzinformationen. Jahresberichte werden durchschnittlich 6-7 Monate nach Ende des Kalenderjahres veröffentlicht. Da das veröffentlichte Dokument Zahlen für das letzte Jahr enthält, ist es eine gute Quelle für die Analyse der vergangenen Periode, jedoch ist der Nutzen dieser Untersuchung zur Vorhersage des laufenden Jahres nicht sehr relevant. Neben den Zahlen besteht der Geschäftsbericht auch aus dem Teil des Lageberichts. Für das besondere Beispiel der Umsatzprognose sind die textuellen Informationsquellen bisher nicht erschlossen.

Da der Standpunkt der Konkurrenzanalyse außerhalb des überwachten Unternehmens liegt, gibt es keine internen Informationen wie z.B. Umsatzzahlen oder Preisstrategien. Für die Prognose der jährlichen Umsatzwachstumsrate wird ein Durchschnitt der letzten drei Jahre genommen. Der Nachteil dieses Ansatzes ist, dass die Ausgabe des Modells sehr ungenau ist. Die anderen nützlichen Informationen wären Pläne, Strategie und Einstellung des Managements in Bezug auf den zukünftigen, wirtschaftlichen und politischen Status des Staates.

Diese Arbeit schlägt vor, die Stimmung aus den jährlichen Managementberichten zu bewerten. In den Berichten ist besonders der Abschnitt Äusblickïnteressant, in dem die Sicht der nahen Zukunft zum Ausdruck gebracht wird. Anhand der Techniken des Sentiment Mining werden wir die Korrelation der verschiedenen Meinungsmaßnahmen in den Outlook-Abschnitten analysieren und mit den entsprechenden historischen Umsatzzahlen vergleichen. Außerdem werden wir versuchen, die bisherige Basismethode zu verbessern. Die Ergebnisse der Experimente werden auf statistische Signifikanz geprüft und auf dieser Grundlage erschlossen. Diese Diplomarbeit wird in Kooperation mit der Siemens AG Österreich durchgeführt und soll am Beispiel ihrer Präsenz auf dem österreichischen

Markt der Industrieautomatisierung als Beispiel dienen. Alle Daten und Jahresberichte, die in dieser Arbeit verwendet werden, sind öffentlich verfügbar.

Abstract

Analysis of it's own revenue figures over the past years can provide great insight for the management of a company. Nevertheless, this approach has its shortcomings. Regardless of the information about its own revenue figures and growth or decline trends in the past, management of a company is lacking a broader view of the current competitors' activity and presence on the market.

To understand how other competitors are performing and what are the reasons for the growth or decline, managers need information for the whole market. With this information in hand, position and underlying reasons for the growth or decline can be deduced. The annual reports of companies provide relevant and faithful representation of the financial information. Annual reports are published on average 6-7 months after the end of the calendar year. Since the published documents provide figures for a last year it is a good source for the analysis of the past time. To forecast the revenues for the next year, financial figures published in those reports are not enough. However, in addition to the figures, the annual report contains the management report part. For the particular example of the revenue forecasting, the textual sources of information have remained untapped so far.

Since the point of view in competitor analysis is outside of the monitored company, one doesn't have internal information like sales figures or pricing strategy. For the forecast of annual revenue growth rate, the standard method is simple to average the revenue figures of the previous 3 years. The shortcoming of this approach is that output of the model is very inaccurate. The other useful information would be plans, strategy and management's attitude towards the future economic and political status of the country. The assumption is that relevant soft facts are taken into consideration when management reports are written. By mining the sentiment we should be able to partly grasp those soft facts.

This work proposes to mine the sentiment from the annual management reports. In the reports, the particularly interesting section is the Outlook section in which the view of the near future is expressed. Using the techniques of sentiment mining, we analyse correlation of the different opinion measures in the Outlook sections and compare it to the corresponding historical revenue figures. Correlation coefficients between revenues and sentiment scores are calculated. Results are interpreted by the domain experts and conclusions are made. This thesis is done in cooperation with Siemens AG Austria and

the use case of their presence on the Austrian market of industrial automation will serve as an example. All the data and annual reports that are used in this work are publicly available.

Contents

CHAPTER 1

Introduction

1.1 Problem definition

"Competitor analysis in marketing and strategic management is an assessment of the strengths and weaknesses of current and potential competitors" [Sim18]. This analysis provides both an offensive and defensive strategic context to identify opportunities and threats. Profiling combines all of the relevant sources of competitor analysis into one framework in the support of efficient and effective strategy formulation, implementation, monitoring and adjustment [FB07]. One part of the competitor analysis is the analysis of market share and its development.

Market share is the percentage of a specific market (in terms of revenue) held by a specific entity. "In a survey of nearly 200 senior marketing managers, 67% responded that they found the revenue dollar market share metric very useful, while 61% found unit market share very useful" [FBPR10]. This metric is used to give a general idea of the size of a company in relation to its market and its competitors. Forecasting revenues of the competitors could provide valuable information to the decision makers and executives. "Marketers need to be able to translate and incorporate sales targets into market share because this will demonstrate whether forecasts are to be attained by growing with the market or by capturing share from competitors. The latter will almost always be more difficult to achieve. Market share is closely monitored for signs of change in the competitive landscape, and it frequently drives strategic or tactical action." [FBPR10]

From the outside point of view of the monitored company, it is very hard to forecast revenue figures. Particularly, because revenues are calculated by multiplying the quantity of goods sold by the price of the goods.

$$TR = P * Q \tag{1.1}$$

where TR stands for the total revenue, P for the price of the goods, and Q for the quantity of goods sold. [Man14] From the outside point of view of the monitored company, one has no access to the pricing strategy nor to the sales forecasts. Therefore, a common way of evaluating and forecasting the revenues is by calculating the average annual growth rate. Average annual growth rate is considered a "hard fact". Things like economical situation in the country, political situation, inflation etc. are considered as soft facts. Since soft facts are hard to grasp in an objective manner, in this thesis sentiment of the outlook section in the management report will serve as a surrogate for the soft facts.

As a use case in this thesis, the focus will be on the Austrian market in the branch of industrial automation. This thesis is done in cooperation with Siemens AG Austria. According to the International Financial Reporting Standards, the following applies to Austria: All large companies must file in their management reports to the Register of Firms within nine months of the fiscal year end.[1] [2]

Information from those reports are generally not processed. According to work done by Petr Hajek, Vladimir Olej and Renata Myskova [HOM14], annual reports contain relevant information that can help forecast companies' financial performance. The sentiment (tone, opinion) is assessed by using several categorization schemas in order to explore various aspects of language used in the annual reports [HOM14].

The work done by Hajek et al. encompass the financial performance expressed in the terms of the Z-score bankruptcy model. Interestingly, there seems to be no work that deals directly with sentiment in the competitors' revenue forecasts.

1.2 Research question

This thesis answers the following question:

- To which extent is it possible to measure a company's attitude and opinion about the next fiscal year performance by utilising sentiment analysis?

In this context, the term denotes the analysis of documents provided by the companies in the branch of industrial automation with the objective to reveal information about their attitude and estimations regarding revenue in the next year or near future. A particularly interesting section of the management reports is the outlook section, which consists of an interpretation of the previous year and some predictions for the next fiscal year encoded in text. We had two hypothesises regarding the correlation between revenues and sentiment scores:

1. Negativity or uncertainty scores could indicate that the company is not confident and certain about the near future. This would be the early indicator of worsening the performance in terms of the market share.

[1] See https://www.corporate-governance.at/uploads/u/corpgov/files/code/corporate-governance-code-012015.pdf, accessed November 28th, 2018

[2] See https://www.jusline.at/gesetz/ugb/paragraf/243, accessed November 28th, 2018

2. High positivity could be indication of covering up problems inside the company and uncertainty in the environment. Especially interesting would be high positive score and high uncertain score.

The main concerns for the process and model used in this thesis are the size of the data set and clear correlation of the sentiment in the management reports and revenue development. In the work of Nopp and Hanbury(2015) [NH15], after the evaluation of the results, the conclusion was that results are meaningful only if figures are aggregated by the year. In other words, when considering single entities (banks), there was too much noise to predict the goal variable with statistical significance.

1.3 Methodological approach

To answer the research question following methodological approach is used.

First step is to gather the data. Dataset is created by downloading the annual reports of companies present in the Austrian market of industrial automation. For all 11 companies present on the market, annual reports were downloaded.

Since reports are printed on the paper and submitted to the authorities, those printed reports are than scanned and published on the internet through the certified third party companies. All the PDF files are encoded as a pictures. To overcome this problem optical character recognition tools are employed.

After gathering the data and transforming it into the text as described in previous section, the process of preprocessing the data begins. Preprocessing consists of the standard preprocessing tasks in natural language processing like tokenisation of the text files, case folding, removing stop words, stemming, etc.

When all the data is transformed and preprocessed we can create sentiment vectors. For weighting schema TF-IDF statistical measure. This enables us to associate correct weights to the terms based on their importance in the document and across the corpus of documents.

Dictionary based approach of sentiment analysis is context sensitive and therefore choice of correct dictionary is very important. In this thesis German business sentiment dictionary[3] is used. German business sentiment dictionary provides three categories of sentiment and therefore three different sentiment vectors were calculated for each document. Vectors are calculated by summing up TF-IDF values of the tokens which belong to the same sentiment category.

At the end the experiment is conducted. The Person correlation coefficients between revenue and different sentiment vectors are calculated. Also, graphical representations of the revenue and sentiment vectors are produced.

[3] See https://link.springer.com/article/10.1007%2Fs11573-018-0914-8, accessed December 8th, 2018

Results are presented to the domain experts. Based on their interpretation of the results conclusions are deducted.

1.4 State of the art

In the academic field, there is an ever growing amount of research regarding sentiment analysis. Furthermore, in the last years more and more users in financial domains use text mining and sentiment extraction to get insights from the information encoded in the text. Applied to different types of business communication such as earnings announcements, analyst reports, or IPO prospectuses, they have been used to extract relevant information for financial market participants [BPW18a].

In the work done by Petr Hajek, Vladimir Olej and Renata Myskova [HOM14], authors examined the role of annual reports' sentiment in forecasting financial performance. The sentiment is collected by employing different sentiment categories in order to explore various characteristics of language used in the annual reports of U.S. companies. Precisely, eleven categories of sentiment (ranging from negative and positive to active and common) are used as the inputs of the forecast models. They have used machine learning methods to forecast financial performance expressed in terms of the Z-score bankruptcy model. Z-score bankruptcy is a model that uses financial ratios to predict financial failure. It takes into consideration multiple financial ratios like current ration, return on capital employed, etc. [BCWH17] The results of the work done by Hajek et al.[HOM14] indicate that the sentiment information is an important forecasting determinant of financial performance and, thus, can be used to support decision-making process of corporate stakeholders. These results show that sentiment scores from the annual reports have predictive power to forecast financial performance like Z-score bankruptcy. The conclusion from the results is that a less optimistic and a more conservative tone was observed for the companies expecting worsening of their financial performance [HOM14].

The main difference is that in this thesis we are attempting to forecast revenue but not internal financial status of the company. As described in section 1.1, to calculate revenues, only valuable information is the sales forecast, pricing strategy and soft facts like managers attitude towards the future, economical status of the country etc. With regards to that, we have little to no use from financial ratios like current ration, return on capital employed, etc. used in work of Hajek et al.[HOM14] Also, business dictionaries available for sentiment mining in German language lacks the variety of the different sentiment categories.

One of the essential concepts in financial markets, volatility prediction, has recently been tackled using sentiment analysis methods. Hanbury et al.[RLB+17] investigated sentiment of annual disclosures of companies in stock markets to forecast volatility. "Our bag-of-words sentiment analysis approach benefited from state-of-the-art models in information retrieval which use word embeddings to extend the weight of the terms to the similar terms in the document."[RLB+17] The word embedding-based approach proposed by the Hanbury et al.[RLB+17] significantly outperforms state-of-the-art methods. The amount of the text in each report used in work by Hanbury et al. [RLB$^+$17] is much bigger then in use case of this master thesis. Therefore, the approach which combines state-of-the-art word embeddings to extend the weight of the terms to the similar terms in the document is not applicable in this use case.

Nopp and Hanbury [NH15] have investigated whether sentiment analysis is capable of measuring a bank's attitude and opinions towards risk by analysing text data. More than 500 CEO letters and outlook sections extracted from the bank annual reports were used. The finding were that at the level of individual banks, predictions are relatively inaccurate. In contrast, the analysis of aggregated figures revealed strong and significant correlations between uncertainty or negativity in textual disclosures and the quantitative risk indicator's future evolution. Work done by Nopp and Hanbury [NH15] inspired us to investigate relationship between revenue figures and sentiment scores from the annual reports of companies in the industrial automation present on Austrian market.

Bannier et al. [BPW17] analysed the market reaction to the sentiment of the CEO speech at the Annual General Meeting. They observed that sentiment from the transcripts of 338 CEO speeches of German corporates between 2008 and 2016 was significantly related to abnormal stock returns and trading volume around the AGM. They also used different dictionaries and German business sentiment dictionary appeared to be better suited to grasp the sentiment of German business documents compared to general dictionaries. [BPW17] In this thesis we will use the same German business sentiment dictionary[4] created by the Bannier et al.[BPW18a].

1.5 Structure of the Work

This master's thesis is divided into the chapters outlined below.

Chapter 2: Economic Background. In this chapter, the economic background of the competitor analysis is explained. Questions such as what revenue is and how revenue is calculated are answered. For the specific use case of this work, the industrial automation context will be explained.

Chapter 3: Technical Background The broad overview of the sentiment mining techniques and natural language processing are presented. Relevant challenges in the field are presented.

Chapter 4: Data Sources In this chapter, the data sources used in this work are presented. Ways to access the data and problems which occur during manipulation with the data are given.

Chapter 5: Methodology This section introduces which methodology is used in the work. It is stated how sentiment dictionary and term frequency help us mine opinion from the textual files.

[4] See https://www.uni-giessen.de/fbz/fb02/forschung/research-clusters/bsfa/textual_analysis,accessed December 8th, 2018

Chapter 6: Experiment and results The Pearson correlation is calculated between different sentiment scores and revenue figures. Analysis is done on both individual company level and on the level of the whole market. A visual representation of the correlation is presented as well. Interpretation of the results by the domain experts are presented.

Chapter 7: Conclusion. The last chapter consists of the conclusions of work. The limitations of the context are listed and explained. Further work possibilities are discussed.

CHAPTER 2

Economic background

In this chapter we are presenting economic background of the problem we are trying to tackle in this thesis. Explanation of the competitor analysis, terms like revenue and market share, why these measures are valuable information to the management of the company. At last, context of the industrial automation branch is described to understand the context and competitiveness of the market.

2.1 Competitor analysis

"Competitor analysis in marketing and strategic management is an assessment of the strengths and weaknesses of current and potential competitors" [Sim18]. This analysis provides both an offensive and defensive strategic context to identify opportunities and threats. Profiling combines all of the relevant sources of competitor analysis into one framework in support of strategy formulation, implementation, monitoring and adjustment [FB07]. One part of the competitor analysis is the analysis of market share and its development.

Market share is the percentage of a specific market (in terms of revenue) held by a specific entity. It is calculated by taking the company's sales over the period and dividing it by the total sales of the industry over the same period. This metric is used to give a general idea of the size of a company in relation to its market and its competitors. Forecasting revenues of the competitors could provide valuable information to the decision makers and executives. As we showed in section 1.1, market share measurement is valuable measurement. Furthermore, in the survey done by Farris et al.[FBPR10] 67% of the managers said that they found revenue dollar market share very useful. From this survey we can see that awareness of the competitors' activity in terms of revenue dollar market share is very useful to the managers. In this thesis we will also use the revenue market share measured in the Euros, since Euro is the domestic currency in Austria.

2.2 Revenue

In accounting, revenue is the income that a business acquires from its usual business activities, from the sale of goods and services to customers. Revenue is also referred to as sales or turnover [BCWH17]. Revenue from a single deal should reflect the amount agreed with the customer for the transfer of goods and services. It should be recognized when the business has satisfied its obligations towards the customer. This occurs when control of the services or goods is transferred to the customer. Where control is transferred over time, the revenue should be recognised over time. This is not the usual case in the industrial automation domain. Most of the time, revenue is recognised at a particular point in time when control is transferred. According to the act of transfer of control and not according to the act of making a payment, revenue is recognised unbounded from the cash transfer [AHM15]. This means that total sales revenue will often be different from the total cash received. According to that, in this work, we are focusing exclusively on the revenue figures, regardless of the cash payments done during the accounting period.

2.3 Industrial automation

"Automation is the technology by which a process or procedure is performed with minimum human assistance."[Gro10] Automation or automatic control is the use of various control systems for operating equipment such as machinery, processes in factories, steering and stabilization of aircraft and vehicles with minimal or reduced human intervention.

Automation refers to the process of replacing person's work with machines, usually through technical progress. Mechanization can be seen as a predecessor of an industrial automation. While the Mechanization of the work allows human beings to operate in an easier working conditions, the industrial automation reduces the need for human presence in performing certain activities. The industrial automation monitors the development of technology in production and shapes the implementation, management and other processes without direct human activity. Electronic engineering, mechanical engineering and computing are the main pillars of the industrial automation process. The goal is to create an efficient technological process [Gro10].

The industrial automation creates the ability to increase productivity and growth in production while reducing production costs and improving product quality. By implementing that, it allows to increase the efficiency of production control. Ultimately, automation results in greater productivity and reduced human workforce (and thus possible human errors) in production. On the other hand, it also helps to reduce total number of jobs, especially repetitive ones. The combination of industry automation, globalization and demographic change has undoubtedly a significant impact on the structure of the state economy.

In this industry branch we can see that effectiveness and feasibility of the automation is very important. This puts the pressure on the manufacturers of the machines due to the high competitiveness in the market.

Especially, problem with the competitiveness appears when we take into consideration interoperability and interchangeability of the machines produced by different manufactures. This is a double edge sword because if your machine can communicate and operate together with the machines of your competitors, than you can penetrate to the competitor's customers more easily. Also, this imposes a danger, because customers can swap your machines with the competitor's machines if better prices are offered. Due to this opportunities of wining customers and danger of losing customers, awareness of the competitors' impact on the market is very important.

CHAPTER 3

Technical background

As stated in the previous chapters, in highly competitive markets any information about competitors characteristics and activities is valuable. Since text from management annual reports will be used to answer research question, knowledge about text mining approaches will be needed to implement the analysis. This chapter covers the technical background knowledge by introducing and giving the overview of the natural language processing, sentiment analysis and Pearson correlation.

3.1 Natural language processing

"Natural language processing (NLP) is a subfield of computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data".

As a part of computer science and linguistics, the goal of natural language processing is to design mathematical models of language structures which enables automatic language processing. We can see it as a method to implement linguistic theoretical rules. Humans use language in everyday life by means of speaking, writing and listening. In an ideal situation, NLP would enable machines to understand and interact with humans by means of natural language. Talking virtual assistants and artificial intelligence that understands as a final goal has not been reached yet. Even if the ultimate goal has not been achieved yet, we are moving forward step-by-step.

"Natural language processing has a history nearly as old as that of computers and comprises a large body of work." [MNOMC11] Still, many attempts failed or never exceeded laboratory demonstrations. Compared to the technologies like operating systems, databases, and networking, natural language processing applications are still scarce. Even so, the number of commercial applications or significant prototypes which include language processing techniques are increasing.

Some of them are:

- Spelling and grammar checkers
- Text indexing and information retrieval from the Internet
- Speech dictation
- Voice control of domestic devices
- Conversational agents

"Unlike other computer programs, results of language processing techniques rarely hit a 100 % success rate. Speech recognition systems are a typical example. Their accuracy is assessed in statistical terms. Language processing techniques become mature and usable when they operate above a certain precision and at an acceptable cost." [MNOMC11]

3.2 Sentiment analysis

3.2.1 General introduction

Liu and Zhang[RR15] give a general definition: “Sentiment analysis or opinion mining is the computational study of people’s opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes” [RR15] Furthermore, Liu and Zhang[RR15] address sentiment analysis as one of the valuable techniques to gather information about what other people think. With the raising of the popularity and the amount of internet reviews and personal blogs, new challenges and opportunities arise. One of those challenges is how can a machine automatically grasp opinions of people encoded in textual form.

There are two common terms which denote, at least in essence, the same field: opinion mining and sentiment analysis. However, the common usage of the words opinion and sentiment differs. According to the Collins Dictionary, opinion is about “a belief not based on absolute certainty or positive knowledge but on what seems true, valid, or probable to one’s own mind [...]”[1], whereas sentiment is “a thought, opinion, judgment, or attitude, usually the result of careful consideration, but often colored with emotion[...]” [2], often based on careful consideration. Considering these differences, the term sentiment analysis seems to be more appropriate for this thesis.

[1] See `https://www.collinsdictionary.com/dictionary/english/opinion`, accessed November 28th, 2018.

[2] See `https://www.collinsdictionary.com/dictionary/english/sentiment`, accessed November 28th,2018

3.2.2 Challenges and Limitations

Sentiment analysis is a difficult task, especially because of the fact that sentiment in written statements is often very ambiguous. If two humans independently rate the sentiment of documents, they will not agree on every rating. For automated approaches, a proper selection of opinion words is crucial, but still not a guarantee for successful sentiment analyses. There are several known issues caused by the flexibility of natural language and other factors:

- Sentiment analysis is usually domain dependent. The same words and phrases can have completely different meanings—even in related contexts. Consider the sentiment for the word “easy” in a written review. This is an unambiguously positive statement in a digital camera review, e.g."this digital camera is easy to use." In a movie review, however, it would be a negative rather sentiment e.g. "the ending of this movie is easy to guess." Thus, a sentiment classifier trained in one domain usually cannot be applied to another domain directly [PL08] .
- In text with multiple entities, Entity-level sentiment analysis systems have to match text elements with the respective entities. However, “current accuracy in identifying the relevant text is far from satisfactory” [Fel13]. Simple example for this would be: "The quality of screen is great, however, the battery life is short." In this sentence screen and battery are two different entities and features like great quality or short life should be associated with only one entity. If the sentiment analysis is done on document level, this problem is not relevant.
- It is hard to detect sarcasm automatically, e.g. in “I work 40 hours a week to be this poor.” Approaches for automated sarcasm detection are still in their infancy.
- Using slang words, misspelling or incorrect punctuation, grammatically incorrect words may pose a problems in sentiment analysis. This is referred to as noisy text.

In this thesis, by using the German business sentiment dictionary created by Bannier et al. [BPW18a] will allow us to correctly interpret sentiment polarity of the words in the business domain. In this analysis we are using document level sentiment analysis and we are not concerned with different entities in the individual reports. Noisy text can occur because of running th optical character recognition on the annual report texts encoded as images can induce spelling mistakes or wrong punctuation. Partly this can be fixed by using the spell checker, as described in later chapters, to fix some of the spelling mistakes.

3.2.3 Dictionary-based approach

Dictionary-based approach belong to the class of unsupervised approaches since they work without class labels and employ words with a known polarity for sentiment extraction. A sentiment dictionary is incorporated for determining the degree of positivity or other features of the text [PL08]. It is evident that the dictionary's quality is crucial for this method. One of the examples of dictionaries is the Sentiment Orientation CALculator (SO-CAL) which uses "dictionaries of words annotated with their semantic orientation (polarity and strength), and incorporates intensification and negation." [TBT+11] Unlike the Sentiment Orientation CALculator, in this thesis, we will use a lexicon which provides only the polarity of the words and not the strength.

As the computer-aided text analyses have gained a lot of attention in recent years, especially in the business communication domain, a large amount of work done in this field employed dictionary-based methods which largely depends on the quality of the dictionary referred to. Since most of those dictionaries are compiled in English, number dictionaries for the other languages are lacking. [BPW18b]. "Our dictionary is based on the English dictionary by Loughran and McDonald [LM11], which is commonly used for examining finance- and accounting-specific texts."[BPW18b]

3.3 Term Weighting

It is obvious that in text documents, not all words are equally important.Therefore, just counting words would not lead towards meaningful results. A better approach would be to weigh different words according to a weighting schema. Chishom and Kolda [CK99] proposed a general term weighting formula covers three main elements of term weighting schemes, namely local weights, global weights, and document normalization:

$$w_{i},_{j} = L_{i},_{j}\, G_i N_j \tag{3.1}$$

In the equation 3.1:

- $w_{\mathrm{i,j}}$ is the weight of term i in document j
- $L_{\mathrm{i,j}}$ is local weight of term i in document j
- G_{i} is global weight of term i
- N_{j} is normalization factor which takes length of document j into consideration

In the paper of the Salton and Buckley [SB88] three essential components were analysed:

- Term Frequency (TF): frequently occurring terms in documents can indicate relevant items. Such local weights have been used for decades in content analysis systems. In its simplest form, TF just counts the occurrences of a term in a document [SB88]. Term which appears 4 times in a document is most likely not 4 times more important than another term which appears only once. Consequently, the impact of high frequency terms is reduced by applying the logarithm function. $tf_{\mathrm{i,j}}$ stands for term frequency of the term i in document j.

$$L_{i,j} = \begin{cases} 1 + log(tf_{i,j}), & \text{if } tf_{i,j} \geq 1 \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

- Inverse Document Frequency (IDF): the downside of term frequency is its poor performance when high frequency terms are omnipresent in the whole collection. Measuring inverse document frequency helps to assess a term's importance within the entire document corpus. It does so by favouring terms concentrated in a few documents. IDF is the most popular global term weight [SB88]. In the equation 3.3, N stands for number of documents and df_{i} denotes number of documents where term i occurs at least once.

$$G_i = log\left(\frac{N}{df_i}\right) \tag{3.3}$$

- Normalization: longer documents usually contain more distinct opinion words than shorter ones, but this should not automatically increase importance of the longer documents . Hence, an appropriate normalization factor should be used in the text analysis system. [SB88]. The rationale is to create weighted document vectors with equal lengths in order to make them better comparable. This operation is also referred to as cosine normalization and results in a vector with a length of one. The term weights used in this equation are the product of the local and the global weight as discussed before.

$$N_j = \frac{1}{\sqrt{\sum_{i=0}^{m}(G_i L_{i,j})^2}} \tag{3.4}$$

This approach is called TF-IDF technique. It is one of the most common techniques used in the information retrieval [MRS08].

3.4 Pearson correlation

In statistics, the Pearson correlation coefficient is a measure of the linear correlation. In essence, Pearson product correlation seeks a line of best fit through the data of two variables. Its value ranges between -1 and +1, where -1 is total negative linear correlation, 0 means no correlation, and +1 total positive linear correlation.

The Pearson correlation measures the strength of a linear relationship as well as direction of the relationship. This fact will provide quantitative information regarding the relationship between sentiment scores and revenues.

The Pearson correlation is defined as 3.5

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2(y_i - \overline{y})^2}} \tag{3.5}$$

In the equation 3.5:

- n is the sample size
- x_{i} and y_{i} are the individual sample points indexed with i
- $\bar{x}$ the sample mean for the values of x
- $\bar{y}$ the sample mean for the values of y

CHAPTER 4

Data sources

In this chapter process of creating data set is explained. First, the definition of annual report is given. Second, since we are using use case of companies in the industrial automation branch present on Austrian market, Austrian legislation on accounting and publishing annual reports is presented. It is important to understand timeliness, restrictions and availability of the reports for monitored companies.

4.1 Annual reports

"An annual report is a comprehensive report on a company's activities throughout the preceding year. Annual reports are intended to give shareholders and other interested people information about the company's activities and financial performance. Most jurisdictions require companies to prepare and disclose annual reports, and many require the annual report to be filed at the company's registry. " [Fri00]

Annual reports exists to give insights into the performance of the company to the different stakeholders. Stakeholders interested in the annual reports could be shareholders, employees, customers, suppliers, government, etc. To be able to trustworthy inform the stakeholders about the ongoing business in the company some general rules need to be followed. According to Bettner and Carcello [BCWH17] data in the annual reports need to be:

- Publicly accessible
- Contain verifiable evidence
- Great emphasis on objective
- Published annually

- Comparable among other companies

Since in this thesis a use case of the Siemens AG Austria will be used and data from the outlook sections will be used, also further conditions need to be satisfied:

- Written in German language
- Access to collection of records is granted via certain service providers [1]
- Contain a section with forward looking information

The outlook section is a part of the management report, which contains a textual summary of the company's results, business environment, market position, investments and internal development. In the outlook, companies write about the expected macroeconomic environment, management guidelines and priorities. Furthermore, the word choice of the managers strongly influence their company [AC06] which further supports our hypothesis that sentiment from the outlooks have influences on the performance of the company.

Information from those reports are generally not processed. According to work done by Petr Hajek, Vladimir Olej and Renata Myskova [HOM14], annual reports contain relevant information that can help forecast companies' financial performance.

4.2 Austrian legislation

Since use case used in this thesis focuses on the Austrian market to understand the context of the reports, investigation of the Austrian legislation regarding accounting and annual reports is required. Management reports shall describe the course of business, including the results of operations, and the situation of the company in such a way as to give a fair picture of the net assets, financial position, results of the operations, and the significant risks and uncertainties to which the company is exposed to. [2] In Austria, all small limited liability companies do not need to publish management report. [3] The limited liability company is considered small if it does not exceed at least 2 of the following conditions [4]:

- 5 million euros of the assets in total
- 10 million euros in sales in the twelve months preceding the balance sheet date
- Annual average number of employees is at least 50

[1] See https://www.usp.gv.at/Portal.Node/usp/public/content/laufender_betrieb/firmenbuch/firmenbuchabfrage/Seite.760006.htmlaccessed November 22th,2018

[2] See https://www.jusline.at/gesetz/ugb/paragraf/243, accessed November 17th, 2018

[3] See https://www.jusline.at/gesetz/ugb/paragraf/243, accessed November 17th, 2018

[4] See https://www.jusline.at/gesetz/ugb/paragraf/221, accessed November 17th, 2018

Furthermore, all big limited companies need to publish management reports. Large companies and concerns need to publish additional information with the management reports. [5]

4.3 Format and encoding

Unfortunately, the format of the data available is not convenient for the analysis conducted using computers. Namely, all the annual report provided by the credited third party providers[6] are scanned physical annual reports. Since documents are available online in the .pdf format, we expected that documents would be electronically created and uploaded in the original form. All the reports are encoded as pictures in .pdf documents. This fact makes it much more challenging to process and use them in text mining analysis. To be able to extract the texts, optical character recognition was employed. Even with the paid optical character recognition software[7] that uses German language, results were not 100% accurately transformed into text format. This was mainly because of the quality of the process of scanning the physical annual reports. With the bad alignment of reports and poor contrast effects, it was impossible for optical character recognition software to read the texts form the pictures.

Some of poor examples can be seen in Figures 4.1 and 4.2. In the first example, it is obvious that quality of the lighting and contrast is bad and even human can barely read it. In the other example, the quality of the scan is bad. Furthermore, bad alignment, and the wavy paper make it impossible for the optical character recognition software to recognize any meaningful words from the text in the picture.

4.4 Collection of Data

The annual reports for this work were obtained from the third party service Compass-Verlag GmbH [8]. In total, over 100 annual reports were downloaded. The data set contains annual reports of the 11 companies which operate in Austria in the industrial automation branch. Major companies present in the Austrian market of industrial automation are:

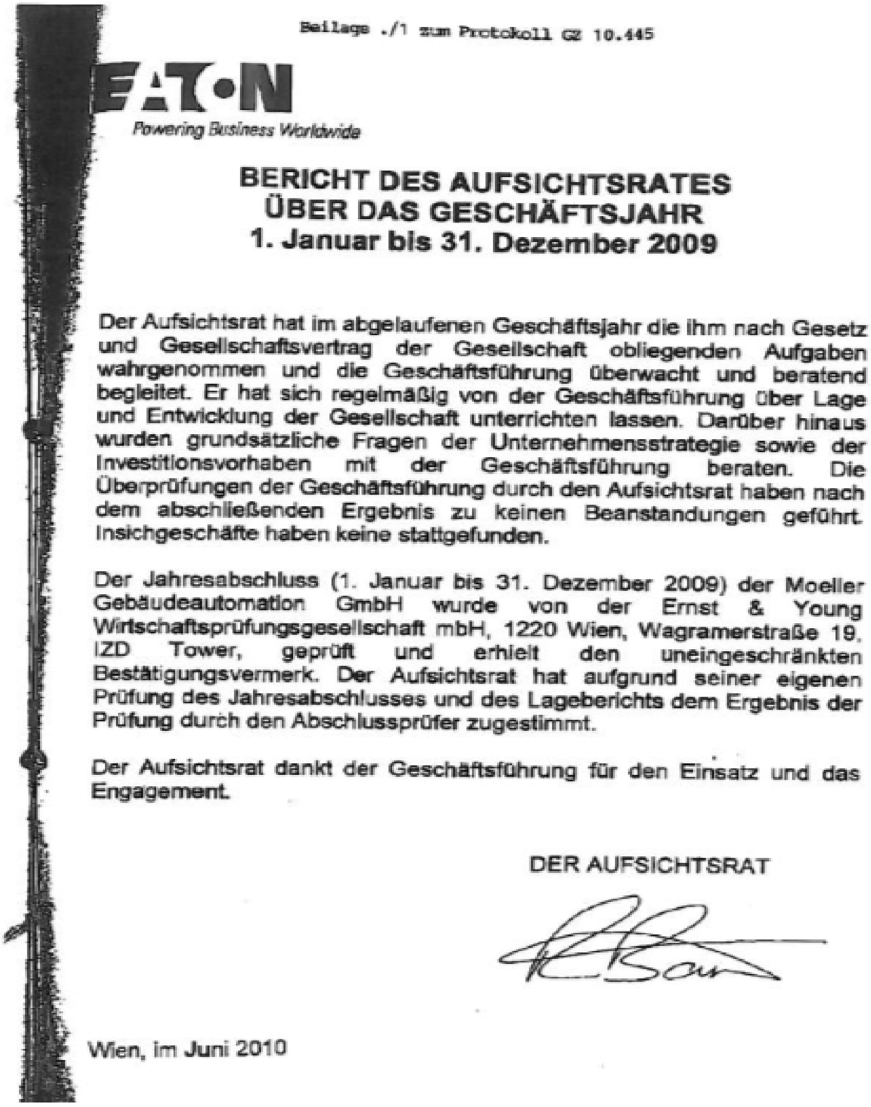
- ABB AG
- B&R Industrial Automation GmbH
- Danfoss Gesellschaft m.b.H.
- Eaton Industries Gmbh

[5] See `https://www.lindeverlag.at/buch/unternehmens-und-gesellschaftsrecht-6162/b/leseprobe/B02417-2.pdf`, accessed November 17th, 2018

[6] See `https://www.usp.gv.at/Portal.Node/usp/public/content/laufender_betrieb/firmenbuch/firmenbuchabfrage/Seite.760006.html`,accessed December 8th,2018

[7] See `https://convertio.co/ocr/german/`, accessed December 8th, 2018

[8] See `https://compass.at/en`, accessed November 25th,2018



Beilage ./1 zum Protokoll GZ 10.445

EATON

Powering Business Worldwide

BERICHT DES AUFSICHTSRATES
ÜBER DAS GESCHÄFTSJAHR
1. Januar bis 31. Dezember 2009

Der Aufsichtsrat hat im abgelaufenen Geschäftsjahr die ihm nach Gesetz und Gesellschaftsvertrag der Gesellschaft obliegenden Aufgaben wahrgenommen und die Geschäftsführung überwacht und beratend begleitet. Er hat sich regelmäßig von der Geschäftsführung über Lage und Entwicklung der Gesellschaft unterrichten lassen. Darüber hinaus wurden grundsätzliche Fragen der Unternehmensstrategie sowie der Investitionsvorhaben mit der Geschäftsführung beraten. Die Überprüfungen der Geschäftsführung durch den Aufsichtsrat haben nach dem abschließenden Ergebnis zu keinen Beanstandungen geführt. Insichgeschäfte haben keine stattgefunden.

Der Jahresabschluss (1. Januar bis 31. Dezember 2009) der Moeller Gebäudeautomation GmbH wurde von der Ernst & Young Wirtschaftsprüfungsgesellschaft mbH, 1220 Wien, Wagramerstraße 19, IZD Tower, geprüft und erhielt den uneingeschränkten Bestätigungsvermerk. Der Aufsichtsrat hat aufgrund seiner eigenen Prüfung des Jahresabschlusses und des Lageberichts dem Ergebnis der Prüfung durch den Abschlussprüfer zugestimmt.

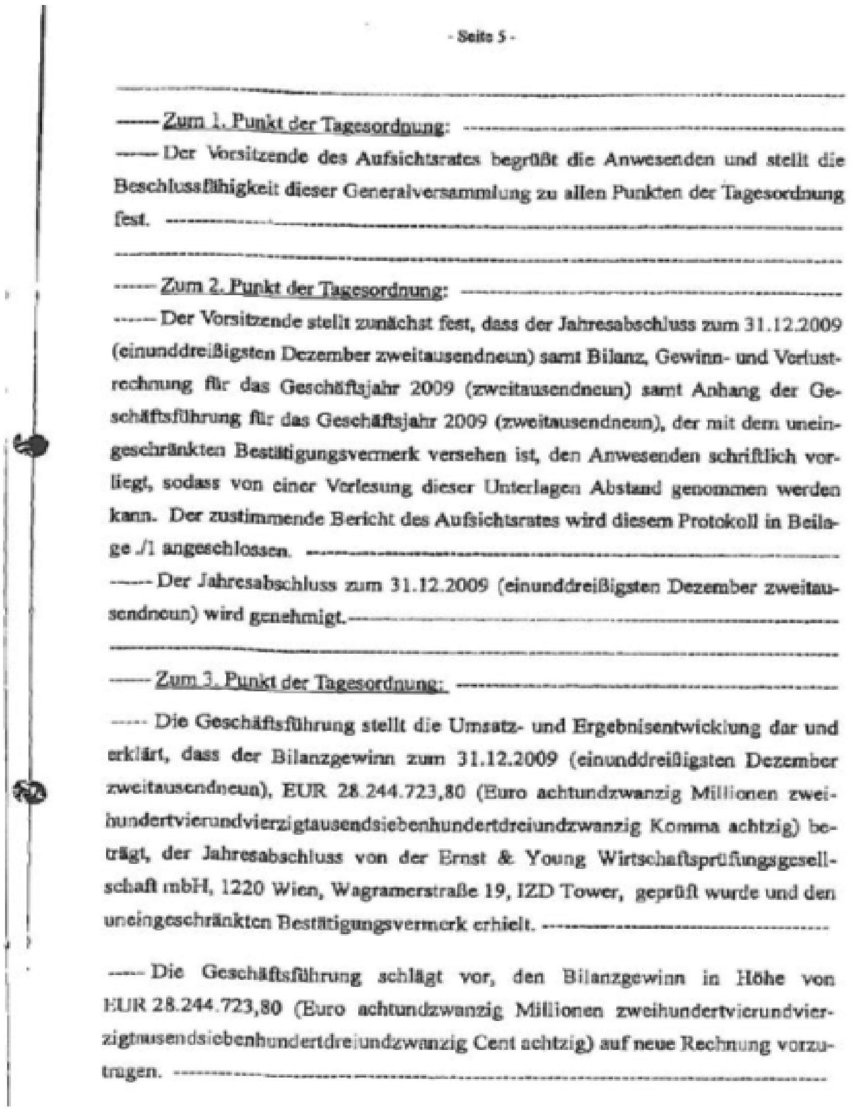
Der Aufsichtsrat dankt der Geschäftsführung für den Einsatz und das Engagement.

DER AUFSICHTSRAT

Wien, im Juni 2010

Figure 4.1: Example of the very poor quality and bad contrast between text and the background

- Endress+Hauser Gesellschaft mit beschränkter Haftung
- Keba AG
- Lenze Antriebstechnik GmbH
- Schneider Electric "Austria" Ges. m.b.H.
- SEW- Eurodrive Gesellschaft mit beschränkter Haftung
- Siemens Aktiengesellschaft Österreich



- Seite 5 -

------ Zum 1. Punkt der Tagesordnung: ------

------ Der Vorsitzende des Aufsichtsrates begrüßt die Anwesenden und stellt die Beschlussfähigkeit dieser Generalversammlung zu allen Punkten der Tagesordnung fest. ------

------ Zum 2. Punkt der Tagesordnung: ------

------ Der Vorsitzende stellt zunächst fest, dass der Jahresabschluss zum 31.12.2009 (einunddreißigsten Dezember zweitausendneun) samt Bilanz, Gewinn- und Verlustrechnung für das Geschäftsjahr 2009 (zweitausendneun) samt Anhang der Geschäftsführung für das Geschäftsjahr 2009 (zweitausendneun), der mit dem uneingeschränkten Bestätigungsvermerk versehen ist, den Anwesenden schriftlich vorliegt, sodass von einer Verlesung dieser Unterlagen Abstand genommen werden kann. Der zustimmende Bericht des Aufsichtsrates wird diesem Protokoll in Beilage ./1 angeschlossen. ------

------ Der Jahresabschluss zum 31.12.2009 (einunddreißigsten Dezember zweitausendneun) wird genehmigt. ------

------ Zum 3. Punkt der Tagesordnung: ------

------ Die Geschäftsführung stellt die Umsatz- und Ergebnisentwicklung dar und erklärt, dass der Bilanzgewinn zum 31.12.2009 (einunddreißigsten Dezember zweitausendneun), EUR 28.244.723,80 (Euro achtundzwanzig Millionen zweihundertvierundvierzigtausendsiebenhundertdreiundzwanzig Komma achtzig) beträgt, der Jahresabschluss von der Ernst & Young Wirtschaftsprüfungsgesellschaft mbH, 1220 Wien, Wagramerstraße 19, IZD Tower, geprüft wurde und den uneingeschränkten Bestätigungsvermerk erhielt. ------

------ Die Geschäftsführung schlägt vor, den Bilanzgewinn in Höhe von EUR 28.244.723,80 (Euro achtundzwanzig Millionen zweihundertvierundvierzigtausendsiebenhundertdreiundzwanzig Cent achtzig) auf neue Rechnung vorzutragen. ------

Figure 4.2: Example of the bad alignment and poor quality

- Sigmatek GmbH & Co KG

Minor companies are so small that they do not need to publish the management reports. Furthermore, their market share is less than 1% and therefore their influence on the course of development of the market is neglected. Some of the companies are present in markets outside of the Austria, primarily East and South Europe. the data set contains annual reports ranging from 2003 to 2017. Nevertheless, for some of the companies, annual reports are missing in the early years. This is due to their presence on the market and due to their size.

From the 4.3 it is clear that over the years, the length of the outlook sections, measured as the number of words after preprocessing task presented in section 5.2, is constantly increasing. This is a promising trend since in the past some of the companies wrote extremely short outlook sections which is changing now.

Annual reports contains financial information about the company. Documents like Profit and loss statement, Income statement and Cash flow statement are standard parts of the annual reports. Only information from this documents needed for the analysis in this thesis were total revenues. Total revenue figures are first entry in the Profit and loss statement and therefore easy to extract.

For the individual companies there were no issues expect the quality of the scanned documents. When grouping the variables by the year, another problem arises. Companies' fiscal years do not need to match the calendar years. This has an effect on grouping the revenues and sentiment scores in the terms of calendar years. For most of the companies, we associated its values to the calendar year that has the biggest overlapping with the fiscal year. In the majority of the cases this is an overlapping of 9 months, which is the best approximation available. One company that has its fiscal year form June of previous year to June in the present year would create bias in the data and therefore is marked as outlier and therefore excluded from the data set. Furthermore, for three more companies problems with data continuity occurred. Eaton Industries Gmbh has a discontinuity in existence of the annual reports. This company was previously part of the other company, and therefore annual reports and revenues of two different entities are incomparable. Second company with discontinuity is Sigmatek GmbH & Co KG. This company has switched its accounting practice regarding presenting the revenue in annual reports. In year 2012 they changed from publishing revenue to publishing cost of materials deduced from the revenue. This is allowed according to the Austrian accounting law[9]. Last company with discontinuity is Lenze Antriebstechnik GmbH. This company was founded in year 2010 and management reports are available from year 2011. All three companies were included into analysis on the individual company level. Since those three companies have discontinuity through the years, their data is not comparable and therefore both companies are excluded from the data set for the market basis level analysis.

[9] See https://www.jusline.at/gesetz/ugb/paragraf/279, accessed December 8th, 2018

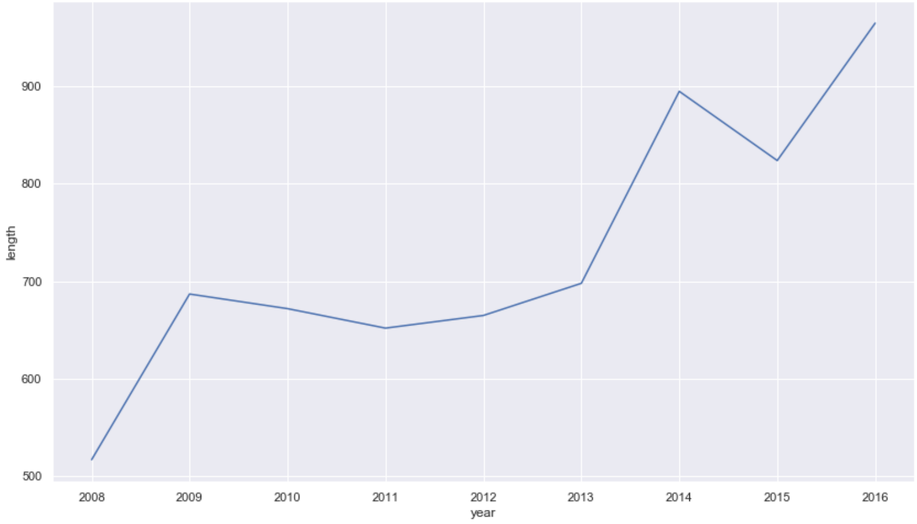


Figure 4.3: Development of the length of the outlook section summed up across the market measured as the number of words after preprocessing.

CHAPTER 5

Methodology

To answer the research question following methodology is used. Methodology comprises the following parts:

5.1 Data gathering and transforming

In the first step data is gathered. Dataset is created by downloading the annual reports of companies present in the Austrian market of the industrial automation. For all 11 companies annual reports were downloaded. Reports are publicly available and accessible online. Availability of the data starts since 2004. Even though, not all the companies started publishing the reports since 2004, but rather from the 2008. For 11 companies we have downloaded 106 annual reports.

Since reports are available online in PDF format, one would assume that they are stored as a text files in a PDF format. That was not the case. Reports are printed on the paper and submitted to the authorities. Those printed reports are than scanned and published on the internet through the certified third party companies. Inconveniently , all the PDF files are encoded as a pictures. To overcome this problem optical character recognition tools are employed.

For the optical character recognition primarily Adobe Acrobat was used. It is free to use software with very good optical character recognition features. For some specific pages, for which the Adobe optical character recognition software couldn't provide readable results, a paid service was employed. The paid optical character recognition service by Softo Ltd [1] provides almost perfect results. This service was used in approximately 5% cases.

[1] See `https://convertio.co/files`, accessed November 26th,2018

5.2 Preprocessing

After gathering the data and transforming it into the text as described in previous section, the process of preprocessing the data begins. It consists of the multiple operations.

1. First step is the tokenisation of the text. Tokenisation is the task of chopping up the text into pieces, called tokens. The NLTK tokeniser [2] for the German language was used.

2. The next step was using a case folding on all the tokens. Case folding is the task of "normalization" of text for the purposes of comparison. Since encodings define upper and lower case letters as different entities, all the characters need to be changed to one single convention. In this thesis we transformed all upper case letters to the corresponding lower case letters. This will enable us to match the same words with or without capital letters.

3. A method for the removing punctuation is applied. By removing the punctuation, tokens which consists of only punctuation characters will be removed. Also, punctuation like a dot at the end of the sentence, or commas in the middle of the sentence, which ended up connected to the word, will be removed.

4. In the next step, non alphabetical characters are removed to get rid of the numbers and special characters like currency symbols, percentage symbols, etc. This purification can be done, because numbers and special characters does not exist in the sentiment dictionaries.

5. Stop words from the sentiment dictionary are removed. Furthermore, stop words from the regular German language are removed by using the NLTK Universal Declaration of Human Rights corpora[3] for German language.

6. A process of stemming is employed. Stemming is the process of reducing inflected or derived words to their word stem(root). This task is important in text analysis because in natural language words are not used just in their root form. In this thesis the NLTK snowball stemmer[4] for the German language is used.

7. Since we have used optical character recognition, and its accuracy was not perfect, some post processing of the tokens is needed. To fix spelling mistakes induced by the process of transforming pictures into the text, pyspellchecker[5] spell checker is used. Initial check with the spell checker showed that around 79% of the tokens have corresponding entry in the spell checker. We need to bear in mind that those tokens contains also the names of the companies, physical persons etc. and spell

[2] See `https://www.nltk.org/api/nltk.tokenize.html`, accessed November 26th,2018

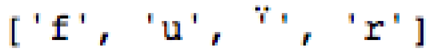
[3] See `https://www.nltk.org/book/ch02.html`, accessed December 7th, 2018

[4] See `https://www.nltk.org/api/nltk.stem.html`, accessed December 7th, 2018

[5] See `https://pypi.org/project/pyspellchecker/`, accessed November 26th,2018

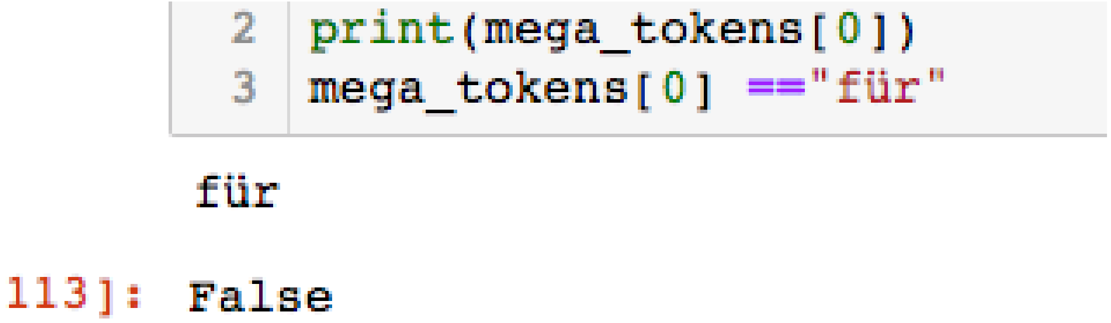
checker don't, so the accuracy in reality is higher. The spell checker successfully corrected some tokens and percentage of valid German words was increased to 87%.

8. An additional step in the preprocessing process is needed in this specific use case. During the transforming pictures into the text written in German language, wrong encoding of specific German letters occur. In this process utf-8[6] is used. Nevertheless, the software couldn't make distinction between regular symbols representing umlauts[7], and non umlaut letters in combination with the special characters that look like umlauts. On the picture 5.1, this phenomenon is showed when token is printed character by character. Furthermore on the picture 5.2 it is obvious that both encodings look the same when printed as strings, but equality between them is not valid.



```
['f', 'u', '̈', 'r']
```

Figure 5.1: String für encoded as regular letters and special character of double dots when printed character by character



```
print(mega_tokens[0])
mega_tokens[0] =="für"
```

```
für
```

```
113]: False
```

Figure 5.2: String für encoded as regular letters and special character of double dots compared for equality with the für encoded with the umlaut

Steps from 2 to 7 are also employed on the German business sentiment dictionary used in this thesis. This is done to increase the matching probability between the tokens and words form the German business sentiment dictionary.

5.3 Creation of the sentiment vectors

When all the tokens are correctly preprocessed and encoded, creation go the sentiment vectors can start.

[6] See `https://en.wikipedia.org/wiki/UTF-8`, accessed November 26th, 2018

[7] See `https://en.wikipedia.org/wiki/Umlaut_(linguistics)`, accessed November 26th, 2018

Not all words are equally important in a text file. Not all words are equally important across corpus of the text files. Not all text files are the same size and therefore have the same number of words in it. Due to this facts, associating the same weight to all words would lead to the wrong results. To overcome this , Term frequency - Inverse term frequency weighting schema is employed.

After calculating TF-IDF weights for each term, matching the terms with the sentiment dictionary terms takes a place.

The sentiment vectors are calculated for each document (annual report). Since sentiment analysis is context sensitive, in this thesis German business sentiment dictionary[8] is used[BPW18a].

Since sentiment dictionary provided three categories of sentiment (positivity, negativity, uncertainty), we have calculated three different vectors for each of the documents. Each token in the document is checked against the sentiment dictionary. If there is a match between the token and the word in sentiment dictionary, corresponding sentiment vector for the examined document is increased by the TF-IDF value of the token. When the process is being run for all the documents, sentiment vectors are created.

5.4 Experiment and results

Experiment consisted of the calculation of correlation coefficients, graphical representation and interpretation of the results by the domain experts.

Correlation coefficients are calculated by using the formula for Pearson correlation coefficient between two variables. Correlation coefficients between revenues and all three sentiment vectors are calculated. Next, graphical representation of the relationship between sentiment vectors and revenues are produced. Graphical representations are used to interpret the correlation coefficients. Both correlation coefficients and graphical representations were created for individual companies and for the whole market (grouped by years).

Both resulting correlation coefficients and graphs were presented to the domain experts. Results of the experiments were presented to the domain experts. Based on results and their interpretation, conclusions are made.

[8] See https://link.springer.com/article/10.1007%2Fs11573-018-0914-8, accessed December 7th, 2018

CHAPTER 6

Experiment and results

In this chapter experiment is explained and results are presented. Results for the correlation on individual basis and on market basis are presented. Graphical representation of revenues and different sentiment vectors are interpreted. Domain experts are interviewed and results are interpreted.

6.1 Experiment

The objective of this experiment is to show how the language of forward looking outlook sections, published by the companies in the industrial automation sector which are present in Austria, evolved within the last decade. The outcome of the lexicon-based approach consists of sentiment scores,calculated as presented in section 5.3, for each document representing the degrees of uncertainty, negativity, and positivity. We had two main hypotheses:

1. Negativity or uncertainty scores could indicate that the company is not confident and certain about the near future. This would be the early indicator of worsening performance in terms of the market share (percentage of an industry's total revenue).
2. High positivity could be an indication of covering up problems inside the company and uncertainty in the environment. Especially interesting would be a high positive score and high uncertainty score.

Evaluating these sentiment scores with classical performance measures would require the manual classification of a training set with labels like enlarging versus shrinking market share. Regarding the aims of this thesis, it is much more interesting to conduct a qualitative analysis. To put the figures in a historical context and to analyse how the financial crisis and other events were reflected in the opinions encoded in management reports. However, the sentiment scores are also compared to the revenue figures. For this purpose, correlation coefficients are calculated.

Since work done by the Nopp and Hanbury [NH15] showed that results of the sentiment analysis were meaningful only when aggregated, in this thesis we have performed experiment on both individual company level and whole market level. When grouping the variables by the year, another problem arises. This problem and solution of it is described in the section 4.4.

6.2 Evolution of the sentiment in the outlook section

As can be seen from Figure 4.3, over the years, the amount of the information in reports increases. Actually, over the period of the 9 years, it almost doubled. A possible explanation for this fact is that companies became more concerned about the future and discuss different scenarios. Another explanation would be that over the years, shareholders demand more information from the management of company. Nevertheless, if we take a look at the situation on the individual level, the length of the outlook sections even declines for some companies. This can be clearly seen in Figure 6.1.

As Figure 6.2 illustrates, all three different summed up scores increased significantly since the global financial crisis influenced the Austrian market in 2009[1]. In the outlooks from the years 2009 and 2010, sentiment scores almost doubled their values. This can be a clear representation of the awareness of the global financial crisis.

In Figure 6.2 it can be seen that in outlooks since 2009 the positivity scores doubled. This is one indicator for the validity of the hypothesis 2 which assumes that companies want to cover up some of the problems that they are aware of. This phenomenon needs to be interpreted with the regard to revenues, which will be done in the following section.

6.3 Results

Results from the experiment are split in two parts. Results on the individual basis and results on the market basis. Results on the individual basis didn't provided any meaningful results. Results on the market level showed moderate correlation between different sentiment scores and revenue figures.

6.3.1 Individual basis

For the experiments of the individual companies, results proved to contain too much noise as expected from the results obtained in the work done by Nopp and Hanbury [NH15]. Figure 6.3 represents relationship between uncertainty score and revenues through the

[1] See https://www.wifo.ac.at/jart/prj3/wifo/resources/person_dokument/person_dokument.jart?publikationsid=35604&mime_type=application/pdf, accessed December 8th, 2018

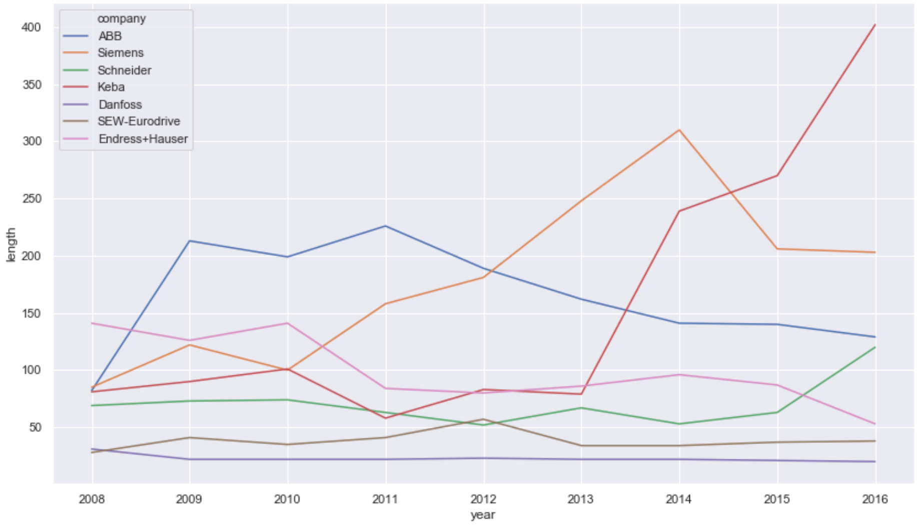


Figure 6.1: Development of the length of the Outlook section, after preprocessing task for the individual companies

years. In Figure 6.3 it is clear that from the year 2008 until the year 2010 there is obvious negative correlation between revenue figures and uncertainty sentiment score. From the year 2011 until 2016 there is no clear correlation between those two variables. It is easier to see this in Figure 6.4 that there is no clear correlation between those two variables. Furthermore, in Table 6.1 the Pearson correlation scores between sentiment vectors and revenue figures are presented. Small absolute values represent that there is no clear correlation between revenue and any of the sentiment scores. Siemens AG is used as an representative example. For other companies, when analysed on individual level, results are similar.

sentiment	revenue
positivity	-0.293880
negativity	0.189128
uncertainty	-0.026225

Table 6.1: Pearson correlation for Siemens AG

6.3.2 Whole market basis

Since on the level of an individual company there was too much noise to show any meaningful results, we grouped revenues and outlooks by the year. Since revenues of

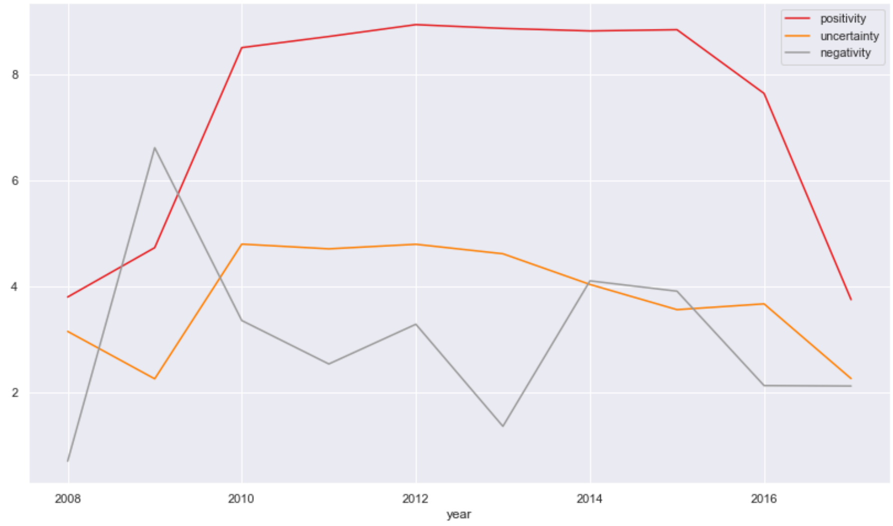


Figure 6.2: Development of the different sentiment scores, summed up across the companies, over the years

companies differ in big scale, revenues and sentiment scores are summed up. In this case, we are viewing sentiment scores and revenue figures for the Austrian industrial automation market. In Table 6.2 the Pearson correlation coefficients are much higher.

Uncertainty sentiment score has a moderate negative correlation coefficient towards revenues. This can be interpreted, in a logical sense, as if the managers feel less certain about the near future and the performance of the company, revenues in next year tend to be lower. This correlation supports our first hypothesis that uncertainty scores could indicate that the company is not confident and certain about the near future. This showed as a early indicator of worsening the performance in terms of the total revenues. So when the uncertainty across different companies is raising, revenues in the next year is expected to fall. Nevertheless, this information cannot tell us anything about the performance of the individual company but rather the size of the market. Therefore, this information is telling us more about the customers and their growth, more than growth of individual competitors. As Figure 6.5 shows, uncertainty is in the main part of the graph acting as a variable with a negative correlation towards the revenue. From 2013 to 2015, an inconsistency occurs. After 2015, the uncertainty sentiment score shows again an opposite behaviour to the slope of the revenue development.

In the interviews with the heads of the business units of Siemens AG Austria, some interesting facts appeared. In the years of the global financial crisis, Siemens and other competitors in the industrial automation scaled down production and laid off workers to be able to survive the recession. With the slowed down production and reduced personnel

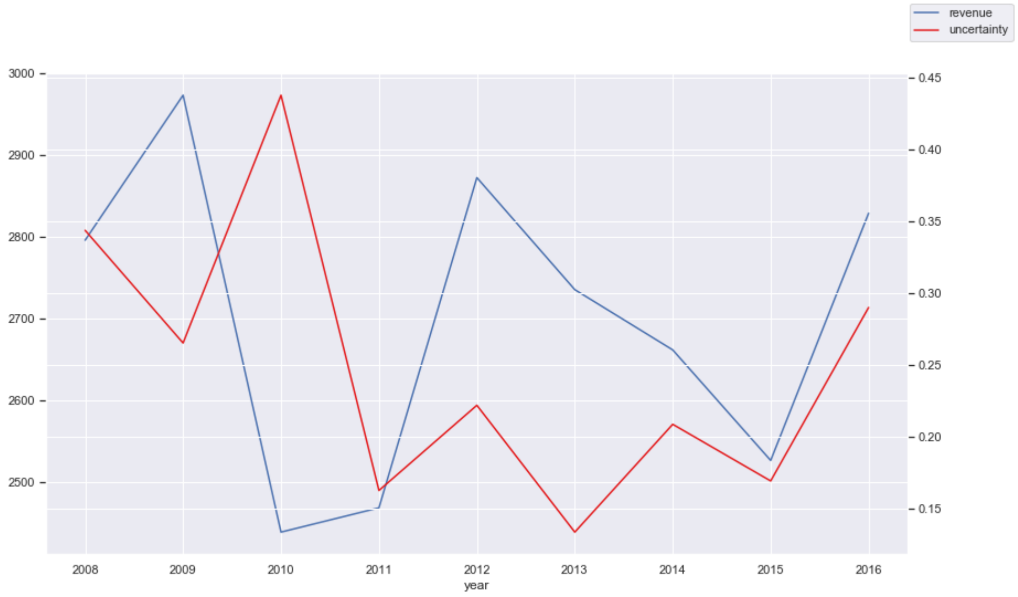


Figure 6.3: Development of the uncertainty over the years for Siemens AG

they were not capable to answer fast and huge market recovery. These insights explain a weird behaviour of the relationship between uncertainty and revenue figures, ranging from 2012 until 2015, as it can be seen in Figure6.5.

This was the result of their overreaction to the global financial crisis. Customers recovered faster than industrial automation companies expected. Huge jumps in terms of orders in years 2011 and 2012 threw the market into a big urge to produce and deliver goods as fast as possible. When companies felt huge demand they scaled up production immediately. Huge increase in orders was not caused only by the fact that customers recovered from the crisis, but also by the customer's desire to fill up the stocks and be on the safe side during the turbulent times. In Figure 6.5 it can bee seen that in year 2012 huge spike in terms of revenue pop up. Also, in following years revenues dropped which clearly represents overreaction of the management of the companies.

As interviewed domain experts stated "If companies hadn't overreacted, the whole process of the going through the crisis and recovering to the previous levels of sales and production would be much smoother." This also explains why we got significant decline of revenues in the years following the recovery of the financial crisis, even though nothing bad happened to the market during those years. In Figure 6.6 it is clear that positivity remained high during those years of high volatility in sales. Furthermore, uncertainty sentiment score in Figure 6.5 from year 2012, shows that uncertainty started dropping in the years following the recovery from the impact of global crisis.

From the year 2015 to the present day, revenues are rising constantly. Furthermore, the

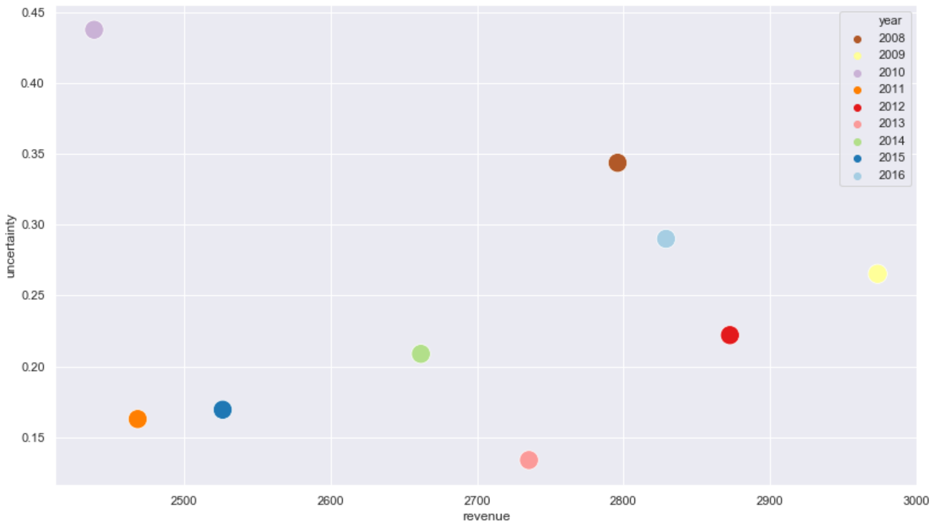


Figure 6.4: Scatter plot of the uncertainty and revenue for the Siemens AG

uncertainty score is in constant decrease as it can be seen in Figure 6.5. With the fall of the uncertainty and increase in the revenues, positivity also decreased to its "base level". Those are the clear indicators that the market is in the ascending course. A senior manager, who works in this industry branch, gives credits for this confidence to Industry 4.0. Drivers for this confidence under the term Industry 4.0 are digitalisation, Internet of things and cloud computing. With these technologies, even the smallest companies feel confident and certain about the future sales figures.

sentiment	revenue
positivity	-0.556096
negativity	-0.028260
uncertainty	-0.623176

Table 6.2: Pearson correlation for the whole market

Moreover, positivity sentiment score also has moderate negative Pearson correlation coefficient. This negative coefficient could be an support for the second hypothesis about covering up problems in the reports during the bad times. The relationship between positive sentiment score and revenue figures needs to be investigated further. For this purpose, the graph in Figure 6.6 could provide us insights. Figure 6.6 shows development of the positive sentiment score and development of the revenue through time. We can see that before the global financial crisis and after 2016, positive sentiment score maintains

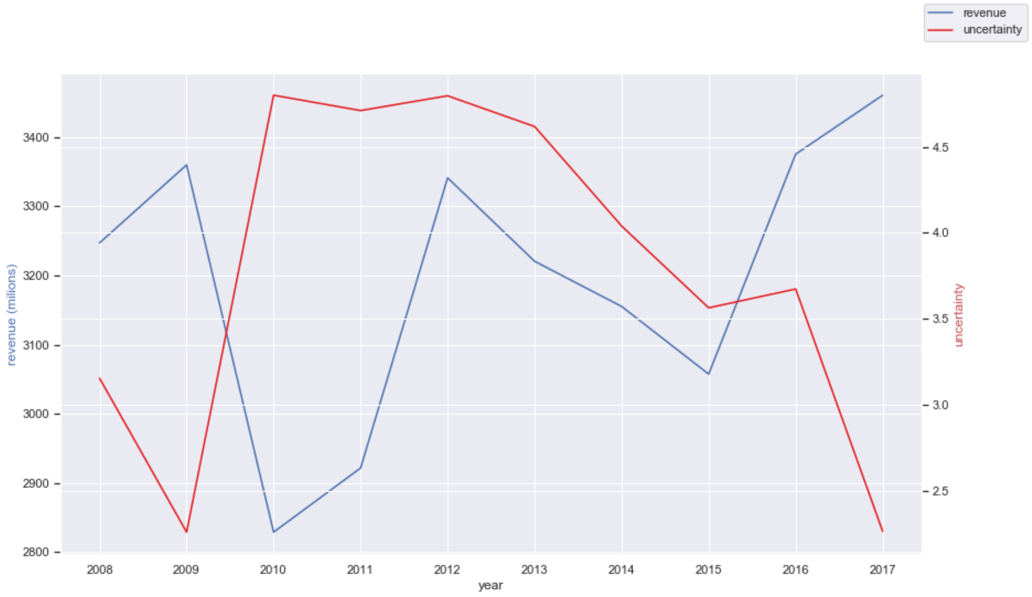


Figure 6.5: Development of the uncertainty score over the years for the whole market

lower levels. A big jump in the positive sentiment score occurs between 2009 and 2010, in the exact time when the revenue drops significantly. During the following years with the very volatile revenue figures, positivity sentiment score stays very high. This is a clear indication that authors of the outlooks use far more positive words during the uncertain times. the described phenomenon exactly confirms our second hypothesis that high positivity could be an indication of covering up problems inside the company and uncertainty in the environment. It is specially interesting that uncertainty and positivity have jumps in the same years. As long as there is high volatility in terms of the revenues, uncertain words and positive words are used much more frequently than usual.

For the relationship of negative sentiment and revenue figures, coefficient of correlation is very small as it can be seen in the Table6.2. Since there is no even slight correlation between negativity sentiment score and revenue figures, therefore negativity sentiment score seems like irrelevant for predicting the revenue figures.

6.4 Summary

Since on the individual basis there is no clear correlation between sentiment scores and revenue figures, creating any model to see prediction power would make no sense. From the example of Siemens in Table 6.1, it is clear that sentiment scores have no or very little predicting power.

Nevertheless, on the market basis, correlation between uncertainty, positivity and revenues

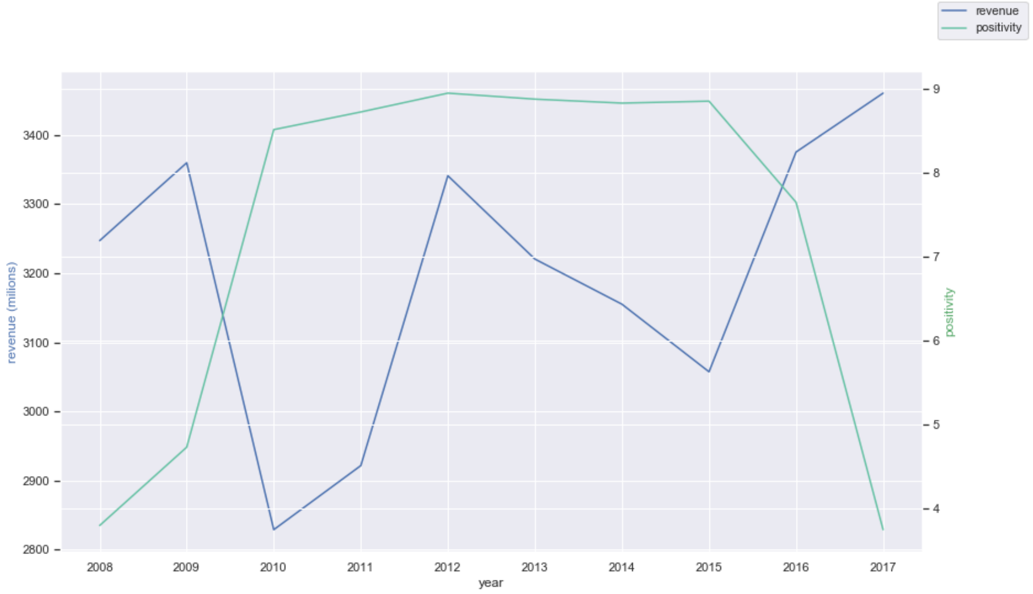


Figure 6.6: Development of the positivity score over the years for the whole market

have moderate coefficients. The Pearson correlation coefficient between negativity score and revenues is very low and therefore have little or no predicting power so it can be ignored. Unfortunately, when we group data by years we end up with only few data points (9). This is not enough to train and test a simple machine learning model. In this use case proof of the concept of predicting revenue figures from the sentiment scores could not be implemented. Because of inability to prove the concept, we can not demonstrate the feasibility of forecasting the revenues by using the sentiment scores.

CHAPTER 7

Conclusion

7.1 Summary and main findings

The management of a company is lacking a broader view of the competitors activity and presence on the market and we presented one of the approaches to particularly overcome this problem.

Relevant sources of the information for this analysis were annual management reports. In those reports forward looking sections provided us insights into the future development of the companies. We had two hypothesises about the text written in those sections of management reports.

1. Negativity or uncertainty scores could indicate that the company is not confident and certain about the near future. This would be the early indicator of worsening performance in terms of the market share (percentage of an industry's total revenue).
2. High positivity could be an indication of covering up problems inside the company and uncertainty in the environment. Especially interesting would be a high positive score and high uncertainty score.

After employing optical character recognition, preprocessing methods, coefficients between different sentiment scores and revenues where calculated. For the individual companies we have seen that there are no meaningful correlations between sentiment scores and revenues. The analysis of individual companies hasn't provide clear correlations between sentiment scores and revenues. This could be the because the data for individual companies contained too much noise.

Uncertainty and positivity scores showed moderate negative correlation to the revenue figures when the data is grouped by years. In other words, results for the whole market together showed meaningful results. To interpret those results, graphical representations of sentiment scores and revenues were created. Domain experts form Siemens AG were interviewed and based on answers form the interview conclusions were made. Our approach produced results which harmonise with the facts presented by the interviewed heads of business units. As described in chapter 6, interviewed domain experts provided insights into the past events, and reactions to those events, which influenced the evolution of the revenues.

Both of the hypothesis were supported by results of the experiment. First we showed that uncertainty has a moderate negative correlation to the revenue figures. Secondly, we have got the results which are supporting the hypothesis that companies, during the bad times, try to cover up problems by quite extensive use of positive words. Both of the hypotheses were also presented to the interviewed business unit heads which confirmed the intention of the managers who has a task of writing management reports.

Unfortunately, we were not able to produce a proof of concept in terms of the simple revenue forecasting model due to the lack of data points when data is grouped by years. The model would demonstrate the prediction power of the sentiment scores.

7.2 Limitations of the work

We have found out that use case of the Austrian market is not favourable for this kind of analysis. This is primarily due to the quality of the data representation through the official data sources. This fact introduced unnecessary mistakes in spelling and wrong encoding. Furthermore, the relatively short history of the annual reports available restrict us in terms of the amount of data for better analysis. Also, the small number of companies present in the market poses a restriction in the terms of amount of data.

Low frequency of the data puts a limit on the timeliness of the analysis and therefore time to react to the results of the analysis. If data frequency is higher, the responses of the managers to the competitors development would be faster. Due to this fact, annual reports are usually available after the 6 to 7 months from the end of fiscal year, which leaves one, who would use results of this analysis, with some time to react. Higher frequency like with quarterly reports would increase the impact of such analysis in terms of timely reactions to competitor's behaviour. Better timeliness and more frequent reports seems like main disadvantage of usage of this kind of analysis.

The inability to prove the concept in a quantitative way puts a question mark on the presented results. We are aware that coefficients between different sentiment scores and revenue figures could be caused by the underlying random distribution of the values. Because of this fact, it is important to bear in mind that results of the thesis need to be proven and there are many opportunities for future work.

7.3 Opportunities for future research

The experiments conducted in this thesis revealed interesting results. Unfortunately, we couldn't train a model to show the impact of sentiment scores on forecast of revenue figures. Nevertheless, there are approaches to improve methodology and make an expansion of the approach:

- It would be interesting to employ the same approach to different countries. Probably in different country contexts some of the restrictions we encountered here would not be present.

- Switching to a larger market. If the same approach was employed for the companies in much larger markets, where a larger number of competitors are present, we believe that the results would be more reliable and more certain.

- Scaling up to the global picture. If we try to grasp the worldwide market, new opportunities for research would arise. Primarily, most of the companies are present on the Stock Exchange markets. This obliges them to publish quarterly reports. By publishing quarterly reports, frequency of the data quadruplicates. This would allow us to get even more timely results which will be even more useful for the users of the analysis, due to the opportunity to act before it is too late. An additional advantage would be alignment of the fiscal year. No more problems with potential distort of the data when the data is grouped by years (quarters).

- Employing the same approach to different industry branches. There are no restrictions why this approach would not be effective in other branches of the industry.

- New relationships could be discovered with the availability of the new German business sentiment dictionaries with more sentiment categories, like English one used in the work of Hajek et al [HOM14].

List of Figures

List of Tables

Bibliography

[AC06] J Amernic and Russell Craig. CEO-speak: The language of corporate leadership. *CEO-Speak: The Language of Corporate Leadership*, pages 1–243, 02 2006.

[AHM15] Peter Atrill, David Harvey, and E. J. McLaney. *Accounting : an introduction.* Pearson Australia Melbourne, Vic, 6th edition. edition, 2015.

[BCWH17] M.S. Bettner, J.V. Carcello, J. Williams, and S. Haka. *Financial Accounting.* McGraw-Hill Education, 2017.

[BPW17] Christina E. Bannier, Thomas Pauls, and Andreas Walter. CEO-speeches and stock returns. Annual Conference 2017 (Vienna): Alternative Structures for Money and Banking 168192, Verein für Socialpolitik / German Economic Association, 2017.

[BPW18a] C. E. Bannier, T. Pauls, and A. Walter. *Dictionary of Finance and Investment Terms.* Springer Berlin Heidelberg, 2018.

[BPW18b] Christina E. Bannier, Thomas Pauls, and Andreas Walter. Content analysis of business communication: introducing a German dictionary. *Journal of Business Economics*, Aug 2018.

[CK99] Erica Chisholm and Tamara G. Kolda. New term weighting formulas for the vector space method in information retrieval. Technical Report ORNL-TM-13756, Oak Ridge National Laboratory, March 1999.

[FB07] C.S. Fleisher and B.E. Bensoussan. *Business and Competitive Analysis: Effective Application of New and Classic Methods.* Financial Times Press, 2007.

[FBPR10] Paul W. Farris, Neil T. Bendle, Phillip E. Pfeifer, and David J. Reibstein. *Marketing Metrics: The Definitive Guide to Measuring Marketing Performance.* Wharton School Publishing, 2nd edition, 2010.

[Fel13] Ronen Feldman. Techniques and applications for sentiment analysis. *Commun. ACM*, 56(4):82–89, April 2013.

[Fri00] J.P. Friedman. *Dictionary of Business Terms.* Barron's Business Dictionaries Series. Barron's Educational Series, 2000.

[Gro10] M.P. Groover. *Fundamentals of Modern Manufacturing: Materials, Processes, and Systems.* John Wiley & Sons, 2010.

[HOM14] Petr Hajek, Vladimir Olej, and Renata Myskova. Forecasting corporate financial performance using sentiment in annual reports for stakeholders' decision-making. *Technological and Economic Development of Economy, 01 December 2014, Vol.20(4)*, 2014.

[LM11] Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65, 2011.

[Man14] N.G. Mankiw. *Principles of Microeconomics.* Cengage Learning, 2014.

[MNOMC11] Prakash M Nadkarni, Lucila Ohno-Machado, and Wendy Chapman. Natural language processing: An introduction. *Journal of the American Medical Informatics Association : JAMIA*, 18:544–51, 09 2011.

[MRS08] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval.* Cambridge University Press, New York, NY, USA, 2008.

[NH15] Clemens Nopp and Allan Hanbury. Detecting risks in the banking system by sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 591–600, 01 2015.

[PL08] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January 2008.

[RLB$^+$17] Navid Rekabsaz, Mihai Lupu, Artem Baklanov, Allan Hanbury, Alexander Dür, and Linda Andersson. Volatility prediction using financial disclosures sentiments with word embedding-based IR models. *CoRR*, abs/1702.01978, 2017.

[RR15] Kumar Ravi and Vadlamani Ravi. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, 89:14 – 46, 2015.

[SB88] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5):513 – 523, 1988.

[Sim18] Dragos Simandan. Iterative lagged asymmetric responses in strategic management and long-range planning. *Time & Society*, page 0961463X1775265, jan 2018.

[TBT+11] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307, 2011.