M. KANMAZ

METU 2018

THE EFFECT OF FINANCIAL NEWS ON BIST STOCK PRICES: A MACHINE LEARNING APPROACH

MEDET KANMAZ

AUGUST 2018

THE EFFECT OF FINANCIAL NEWS ON BIST STOCK PRICES:
A MACHINE LEARNING APPROACH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MEDET KANMAZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF ECONOMICS

AUGUST 2018

Approval of the Graduate School of Social Sciences

Prof. Dr. Tülin Gençöz
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Meltem Dayıoğlu Tayfur
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Serkan Küçükşenel
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Seza Danışoğlu	(METU, BA)	
Assoc. Prof. Dr. Serkan Küçükşenel	(METU, ECON)	
Assist. Prof. Dr. Nuh Aygün Dalkıran	(Bilkent Uni., ECON)	

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Medet Kanmaz

Signature :

ABSTRACT

THE EFFECT OF FINANCIAL NEWS ON BIST STOCK PRICES:
A MACHINE LEARNING APPROACH

Kanmaz, Medet
M.S. Department of Economics
Supervisor: Assoc. Prof. Dr. Serkan Küçükşenel

August 2018, 106 pages

This thesis examines the relationship between the price data of companies in different sectors in the Borsa Istanbul (BIST) stock exchange and the verbal data revealed in the financial news related to these companies. In this work, sentiment analysis, natural language processing and the effect of financial news on individual stock performances are studied with a simple and novel method. Sentiment analysis is created by automatically labelling the news for companies publicly traded in BIST as positive or negative on the basis of the daily performance of stocks with different methods in machine learning. These algorithms determine the polarity in financial news with an accuracy of around 70%. As a result of this study, it was seen that positive or negative news had a positive / negative effect on the related stock prices. Whether the outcome of this algorithm provides incentives to make a profit in the market or not is also questioned. On the other hand, it is shown that it is hard to gain profit from this public information unless there is insider information.

Keywords: Sentiment Analysis, Financial Analysis, SVM, Naïve Bayes, Automated Labeling

ÖZ

FİNANSAL HABERLERİN BİST HİSSE SENEDİ FİYATLARINA ETKİSİ: MAKİNE ÖĞRENMESİ YAKLAŞIMI

Kanmaz, Medet
Yüksek Lisans, İktisat Bölümü
Tez Yöneticisi: Doç. Dr. Serkan Küçükşenel

Ağustos 2018, 106 sayfa

Bu tez, Borsa İstanbul'da (BIST) bulunan farklı sektörlerdeki şirketlere ait fiyat verileri ile bu şirketlerle ilintili olan finansal haberlerde kullanılan ifadeler arasındaki ilişkiyi incelemektedir. Bu çalışmada, duygu analizi, doğal dil işleme ve finansal haberlerin bireysel hisse senedi performansları üzerindeki etkisi basit ve yeni bir yöntemle incelenmiştir. Duygu analizi, pay piyasasında halka açık bir şekilde işlem gören şirketlerin hisse senedi fiyatlarının günlük performansına göre, bu firmalara ait finansal piyasalarda yer alan haberlerin pozitif veya negatif olacak şekilde otomatik etiketlenmesi suretiyle makine öğrenmesi yöntemi kullanılarak yapılmaktadır. Oluşturulan algoritmalar, finansal haberlerdeki kutupluluğu %70 civarında bir doğruluk oranıyla belirlemektedir. Bu çalışmanın sonucunda, olumlu veya olumsuz haberlerin, ilgili hisse senedi fiyatları üzerinde olumlu/ olumsuz etkisinin bulunduğu görülmüştür. Çalışmada ayrıca, algoritma çıkarımının hisse senedi piyasasında kar elde etmek için teşvik sağlayıp sağlamadığı sorgulanmış, içeriden öğrenilen bilgiler dışında kamuya açık bilgilerden kâr elde etmenin zor olduğu görülmüştür.

Anahtar Kelimeler: Duygu Analizi, Finansal Analiz, SVM, Naïve Bayes, Otomatik Etiketleme

To My Parents, My Wife and My Son. I love you all dearly.

ACKNOWLEDGMENTS

I owe many thanks to my supervisor Assoc. Prof Dr. Serkan Küçükşenel for supporting and encouraging me in choosing the research topic without any constraints and for his guidance and criticism to improve the content of this thesis. Also, without his good sense of humor, this research would be harder and stressful.

I owe a debt of gratitude to Assoc. Prof. Dr. Seza Danışoğlu. During the data collection process, her guidance is invaluable.

I am really thankful to the examining committee members Assist. Prof. Dr. Nuh Aygün Dalkıran and Assoc. Prof. Dr. Seza Danışoğlu for their time, interest, evaluations, assistances in eliminating the errors and giving fresh ideas about future work.

My family has been the sources of my motivation for doing my best in my life and my academic research as well. I acknowledge my wife, Yasemin, for her great patience and continuous supports and my little energizer bunny, Ahmet Mert, for his playful smile and never-ending questions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BIST	Borsa Istanbul
CAPM	Capital Asset Pricing Model
DT	Decision Tree
EMH	Efficient Market Hypothesis
GI	General Inquirer Dictionary
HML	High Minus Low
KAP	Public Disclosure Platform
L&M	Loughran and McDonald
ME	Maximum Entropy
ML	Machine Learning
MM	Market Model
NB	Naïve Bayes
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
RID	Regressive Imagery Dictionary
S&P	Standard and Poor's
SMB	Small Minus Big
SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

In our daily life, data has become one of the most important resources. Contrary to nature of resources, even comes in abundance. With the help of powerful machine learning algorithms, it can be turned into intelligence. There has probably never been a better time to step into the machine learning field and learn how to use this knowledge in different areas. This is because qualitative analysis, more specifically textual analysis, has been interested by many researchers. Think about your Gmail box. When was the last time you encounter spam mails in your email box? As we all remember, early mailboxes were full of spams, it took quite a while to delete these mails. However, identifying markers and features in an email and making spam detection becomes as easy as ABC by means of powerful algorithms (spam filters). The important step taken here was, undoubtedly, the textual analysis.

As for accounting and finance side, news media, social media, conference calls, financial statements, public disclosures and even chatrooms[1] are the main sources of information for making investment decisions. Hence, any tool that can help to tease out sentiment from this text data would be interesting.

The hidden information held in a text is called sentiment. This study aims to extract sentiment (positive or negative) from financial news for firms automatically. The text data below is an example for positive news story for the firm NETAS.

[1] Tirunillai and Tellis (2012) find that one can beat the S&P 500 Index by 7.9 percentage points over the four-year period by making investment decision on the stocks (hypothetical portfolio) mostly talked in chat rooms. Bordino et al. (2012) find that trading volume and query volume (query in search engines) of NASDAQ-100 stocks are highly correlated.

Probil Bilgi won the Fatih Project Tender, the shares flew. Probil Bilgi, a subsidiary of Netaş Telekom, offered the best bid with a total of 249.94 million TL to Fatih Project 2. Phase Local Area Network Installation. After the announcement Netaş shares went up to 9.79 Turkish Lira.[2]

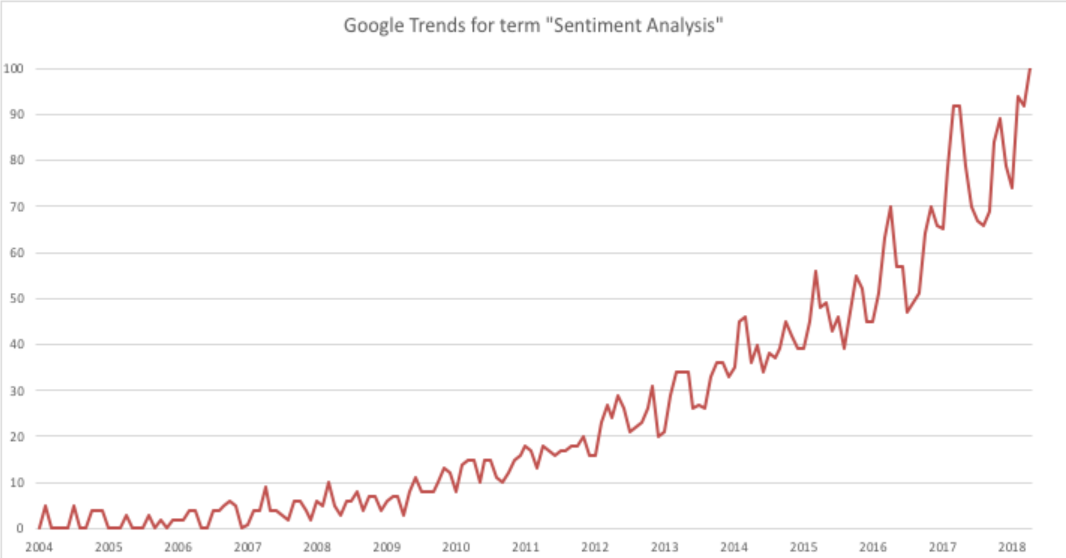


Figure 1- Google Trends

Figure 1 displays search interest for the term "sentiment analysis" in a period from 2004 to 2018. A value of 100 means that the term has the highest popularity and a value of 0 means that there is not enough data for the term "sentiment analysis". It can be seen that there has been growing interest since 2004 and it has reached the highest popularity in May 2018.

Even tough "Sentiment Analysis" is in its peak period in these days, sadly to say that there is limited work in Turkish. The main reason for this problem is that sentiment dictionaries, word lists, labeled and tagged datasets which are available mostly in English provide ample fodder for studies concentrating broadly in this language. Therefore, the aim of this thesis is to propose a novel approach towards data collection

[2] The original news story is: "Fatih Projesi ihalesini kazandı, hisseleri uçtu. Netaş Telekom'un bağlı ortaklığı Probil Bilgi, Fatih Projesi 2. Faz Yerel Alan Ağ Kurulum ihalesine toplam 249,94 milyon TL ile en iyi teklifi verdi. Açıklama sonrası Netaş hisseleri 9.79 liraya kadar çıktı."

and labeling on Turkish data; perform experiments on this dataset and highlight areas of future research.

The structure of the remainder of the thesis is as follows. Section 1.1 points out the motivation for the thesis. Section 1.2 reviews recent literature on sentiment analysis and Section 1.3 shows main findings and remarks of this thesis. Chapter 2 summarizes the structure of the data and its description. In Chapter 3, textual processing methodology is covered. Chapter 4 reviews different market models for identifying the effect of financial news on BIST stock prices and examines whether there is trade incentive or not. Chapter 5 concludes the thesis and provides recommendations for future studies.

1.1. MOTIVATION

Understanding the nature of human intelligence is a challenging topic. Technical innovation in every sector depends on human intelligence from individual companies to government institutions. For this reason, any approach that mimics the feature of human mind and simulates the way of thinking / problem solving skills always fascinates me. For me, it was the turning point when the world chess champion, Garry Kasparov agreed to play a match against Deep Blue, the IBM supercomputer. After the match[3], February 1996, he said that:

> "I had played a lot of computers but had never experienced anything like this. I could feel, I could smell a new kind of intelligence across the table. It was a wonderful and extremely human move."

[3] Interested readers can refer to Kasparov, G.'s article: "The day that I sensed a new kind of intelligence."

Similarly, in May 2017 DeepMind's AlphaGo[4], the first computer program which defeated the world Go champion Ke Jie, took the stage. After the match, the grandmaster Ke Jie said that:

> When I first saw it, I thought it was almost an impossible move for human players to come up with, since it is obviously a later step. But afterward, I realized it was really an astonishing move. Compared to last year, AlphaGo's understanding of Go has progressed so much. Last year it played in a human-like way, but this time, it's almost like the God of Go.

However, Ke Ji was wrong when he called AlphaGo as "God of Go". A few months later, in October 2017, the newest version of DeepMind's AlphaGo Zero, which was created without human data, surpassed AlphaGo with a score 100-0.

Discovering the success of artificial intelligence in different areas was the starting point of my journey to enjoy doing research on concepts of machine intelligence. For this reason, my main motivation for writing this thesis is to show that the tools and techniques commonly used in the field of computer science can also be applied to the field of economics. Even though my research, fieldworks and findings are restricted to the domain of financial news and market data, the main contribution of this thesis is to give insights about future work in economics engaging with computer science. I hope this thesis gives a sense of how scientists working in the field of economics should benefit from developments in computer science.

1.2. RELATED WORKS / LITERATURE REVIEW

There are many studies concentrating on textual analysis and classification. These studies date back to the late 1990s. Learning from the data itself and dictionary-based learning methods are two different approach in sentiment analysis.

[4] Silver et al. (2017) trained their initial version of deep neural network by "supervised learning from human grandmasters' games and reinforcement learning from games of self-play".

Learning from the data itself consists of many different techniques. Argamon-Engelson, Koppel and Avneri (1998), define the features from the frequencies of words and parts-of-speech triples. Then, they use machine learning techniques to classify the stories from New York Time News, NY Times Editorial, NY Daily news and Newsweek. Turney (2002) is the other sentiment-based classification paper focuses on non-financial domain. The author classifies restaurant and automobile reviews as positive (thumbs up) or negative (thumbs down). The adjectives and adverbs in a review is used for prediction of semantic orientation of the review.

In this approach, data should be labeled either manually or automatically. Dave, Lawrence and Pennock (2003) classify product reviews as positive or negative using machine learning algorithm with manually labelled data. Pang, Lee and Vaithyanathan (2002) aim to determine whether a review is positive or negative. They used manually labeled movie reviews as the training data. They propose that standard machine learning techniques definitively outperform human-produced baselines. Koppel and Shtrimberg (2006) introduced automated labelling according to returns in a given stock. They used Support Vector Machines (SVM), decision trees and Naive Bayes learners to detect sentiment from news stories. Their linear SVM model yield 70.3% accuracy. Similarly, Généreux, Poibeau and Koppel (2008) have hypothesized that "the market reaction to a news is a good indicator for labelling financial news, and that a machine learning algorithm can be trained on those news to build models detecting price movement effectively". They propose a pool of features to detect price changes via machine learning algorithms. Their model has 69% of accuracy.

As far as dictionary-based learning methods is considered, Henry (2006), Henry (2008), Li (2006) are the pivotal papers. This approach mainly focuses on word counts. There is no need to pre-label the data. Each dictionary has different features (word lists) on specific range. The negative/positive count score for a given text, for example, stands for the outcome (label) of the news story. Many researchers use the list of words or dictionaries for sentiment analysis. Li (2006), for example, uses annual reports to detect risk sentiments by using specific word list for risks and uncertainty.

General Inquirer Dictionary (GI), a work of Smith et al. (1967), has already been widely used by many researchers.[5] Tetlock (2007), Engelberg (2008) uses this dictionary for the content analysis method. Tetlock (2007) argues that high value of media pessimism leads downward pressure on prices and media pessimism forecasts increase in market volatility. Sentiment analysis based on GI gives important results in finance literature. Engelberg (2008) explains that both quantitative and qualitative information in firms' earning announcement have an impact on future returns. Tetlock, Saar-Tsechansky and Macskassy (2008) use GI dictionary for counting negative words in news articles. They find that stock market prices reflect the negative information with one-day delay; negative words in stories about firm fundamentals affect earnings and returns more than negative words in any other stories.

As dictionary-based categorization focuses on counting positive and negative words, Boudoukh, Feldman, Kogan and Richardson (2013) argue that "employing a more sophisticated textual analysis methodology further improves the results". For this reason, they introduce one other category for sentiment analysis. They propose a new categorization for news as relevant and irrelevant as "identify the news, and more accurately evaluate its tone, there is considerably more evidence of a strong relationship between stock price changes and information" (p. 4). If the news stories are identified and contain related information about company's fundamentals, they are accepted as relevant. We take this hypothesis and gather related news stories for the news data set explained in Chapter 2.

There are some studies that compare the cross performances of different dictionaries. In their paper, Heston and Sinha (2016) compare the return predictability of different

[5] In Chapter 3, we discuss GI and other dictionaries in more detail.

dictionaries and effects of daily / weekly news stories on stock returns. The authors find that "stocks with positive (negative) news over one day have subsequent predictably high (low) returns for 1 to 2 days that are largely reversed. But in contrast to the published literature, aggregating news over one week produces a dramatic increase in predictability" (p. 4).

Similarly, there are some other studies comparing the performances of both dictionary-based method and self-learners. Azar (2009) uses machine learning algorithms to detect sentiments and categorizes news in two classes: positives and negatives. He uses financial texts and returns as a source of the models. The author uses both GI dictionary and self-learners for labelling the news stories. He shows that SVM model has comparable performance, even to human raters.

The other example for cross comparison of these two methods is shown in Huang, Zang and Zheng (2013). They classify the opinions in analyst reports for S&P 500 firms during the 1995–2008 period. The authors find that Naive Bayes learners have higher accuracy than dictionary-based approaches and negative sentences in reports have higher importance in terms of informative about companies' earnings growth rate performance.

As for literature on sentiment analysis of Turkish texts, the number of studies is limited. Since these sentiment dictionaries are mostly in English, there is limited work on dictionary-based method is Turkish literature. Most of the studies based on prelabelled set and self-learner methods. Whereas, Dehkharghani, Saygin, Yanikoglu and Oflazer (2016) propose the first Turkish polarity resource, SentiTurkNet. They present three polarity scores for each level (positive, negative and neutral) to each synsets in the Turkish WordNet (contains almost 15,000 synsets). This is an important source so that it can be used as dictionary-based approach for sentiment analysis in Turkish texts.

Due to its rich morphology, natural language processing is challenging in Turkish language. Stemming (taking the roots of a word) is the main handicap. Kaya, Fidan and Toroslu (2012) compare the performance of well-known supervised machine learning algorithms (Naive Bayes (NB), Maximum Entropy (ME), Support Vector Machine (SVM) and the character-based N-Gram Language Model) for sentiment analysis of Turkish political columns. They use the stemming tool, Zemberek[6]. Their models reach accuracies of 65% to 77%. In his thesis, Eroğul (2009) applies SVM to prelabelled data sets from movie reviews. He uses stemming tool for examining the effects of this. He finds that using root of words causes 1% decrease in performance of classifiers in terms of F-measure. Boynukalin (2012) presents emotion analysis of Turkish text. She generates a new data set from two different sources (translation of ISEAR dataset and Turkish fairy tales). She proposes different approaches and compares the accuracy results. Her models reach accuracies of 76% to 81%.

There are many sources of information related to the firms. Social media texts such as tweets are the other source for sentiment analysis. Limited number of characters in this text data give an incentive to focus on detecting sentiment using both dictionary-based method and self-learners. In their study, Türkmenoğlu and Tantuğ (2014) compare two different sentiment analysis frameworks for Turkish social media (twitter and movie reviews). They use natural language processing tool to build lexicons for the comparison. They find that "accuracy of movie dataset is better than accuracy of twitter dataset in both lexicon based and ML based sentiment analysis methods."

On finance side, news from Public Disclosure Platform (KAP) is an important source for firms. Eyuboglu and Bulut (2015) examine the effect of financial news from Public Disclosure Platform on the BIST-30 companies' stock prices. Instead of sentiment analysis of news, they use the category titles as benchmarks for their analysis. One of the latest work on sentiment analysis is the paper of Eliacik and Erdogan (2018). In

[6] Zemberek is the stemming tool for Turkish language. All the code and APIs can be acquired from https://github.com/ahmetaa/zemberek-nlp. Accessed May 9, 2018.

the said paper the authors take into account the social network information with PageRank based algorithm that detects influential users on that topic (considering users and influencers) and calculate the community's (Bloomberg HT anchor and CNBCE anchor and their community) sentiment polarity. They also made a correlation analysis between the polarity results and the BIST-100 performance.
As for accuracy scores, there is a wide range of acceptance level. The main threshold for the accuracy score is the surpass of the occurrence rate of any class in a given set. For simplicity, if the negative news consists 55% percent of the all data set, then any score above 55% could be accepted. Each method gives different scores. Catal and Nangir (2017) provide a more recent survey about sentiment classification based on multiple classifiers. Their "majority voting" classification method gives better accuracy scores than individual classifiers on three different domains (book, movie, shopping). Also, their Table 3 provides a bibliography of most sentiment-related papers in Turkish language published prior to 2017.

In economics, sentiment analysis is used for different areas as well. In their working paper, Shapiro, Sudhof and Wilson (2018) propose macroeconomic perspective for news sentiment. They develop an index from news sentiment "to measure the economic sentiment embodied in the news media" (p. 2). The authors find that "the news sentiment indexes correlate strongly with contemporaneous key business cycle indicators and aid in forecasting economic variables, particularly for inflation and the federal funds rate. Innovations to news sentiment, orthogonal to business cycle fundamentals, or what can be referred to as "news sentiment shocks," affect future economic activity" (p. 4). This thesis takes the working hypothesis of Généreux et al. (2008) and implements this into Turkish Stock Exchange Market and financial news in Turkish. In this work, we prefer self-learner method for our sentiment analysis. The outcome of the learners is also questioned, whether it provides incentives to make a profit in the market or not.

1.3. MAIN CONTRIBUTION

As a result of our novel approach, we achieved to get valuable information from unlabeled data. We made a dictionary with effective words for financial news. Each word has also labels as negative or positive. One valuable thing is that the entire process is managed automatically. Thus, the automatic labelling approach that we built in this work gives some incentives about generating a large corpus easily.

The other important thing added to this process is the collection of market sense as a determiner of the polarity. As a result of this, it has been gathered more reliable judgement.

We have found that our classifiers can learn to detect sentiment in the news stories with moderate success. As we compared several learners, we got better results from Naive Bayes classifiers about 70% of accuracy and 80% score for F-Score. We have room for optimism that more sophisticated feature set might improve this score.

As for market concerns, we have found that both negative and positive news has an effect on its related stock prices. We showed that news dummies are statistically significant in market models. It is very likely, though, that labelling news stories would be an important input as far as it is implemented in a very short period of time after the stories published. For this reason, we question whether there is trade incentive or not within this context. We showed that it is hard to gain profit from this information unless there is insider information.

CHAPTER 2

DATA COLLECTION AND DESCRIPTIONS

In this chapter, we briefly introduce the data sets used for both sentiment analysis and calculation for abnormal returns in this work. These are the corpus of financial news; daily stock performance of publicly traded firms /market performance of BIST and two-year government bond as a proxy for risk free rate of Turkey. Section 2.1 highlights the financial news data and Section 2.2 covers the descriptions of price data used in analysis.

2.1. NEWS DATA

Our news data source[7] consists of over 100 different sources including newspapers, monthly magazines, internet news, bulletins and public disclosure platform news. We gathered about 75000 news stories about publicly traded companies from BIST for the years 1996-2018. This corpus set also includes some important elements: date of publication, name of the company, headline and summary of the body.

From this point, we make some assumptions about news items and take some steps towards initial data set. The first one is about ad-containing news. In general, articles in the magazines have some ad-containing news stories. Using this data may always result in getting positive sentiment. Also, because of the fact that the news in the monthly magazines is written before the date of publication, it may not correspond to the change in the price of the stock on the publishing day. For this reason, articles from monthly and weekly magazines are disregarded.

[7] Financial Information News Network (FINNET)

Since our aim is to analyze the effect of news stories to the stock return, which are observed at daily basis, we merge the news stories released on a day into one body. Another thing that we take step is that if the news story published on weekend or holidays, the reference date is being changed to the next working day. The same approach has been taken for the news stories published after the closing time of the market. So, we use the reference day for the news stories instead of publishing date for those news items. For instance, the following news items are the sample news for the company Bimeks for different dates.

> "Dünya devine Türkiye'de büyük şok! Bimeks'ten yapılan açıklamada ise eylülde satın aldıkları Elektro World'ün toplamda 18 franchise mağazası bulunduğu belirtilerek bunlardan 9'unun tabelalarını değiştirerek Bimeks'e geçtiği vurgulandı." (Published on 12 February 2014)

> "Bir süredir mali açıdan zor günler yaşayan ve birçok mağazasını kapatmak zorunda kalan Bimeks, kurtuluş için alacaklılarına iki alternatifli bir plan sundu. Buna göre ya borçlu firmalar alacakları ölçüsünde şirkete ortak olacak ya da alacaklarını bir şirkete devredip o şirket vasıtasıyla hissedar olacaklar." (Published on 5 February 2018.)

The news data has no label for the items. The important step is to categorize each news item. To label each news as positive or negative, one can read each of them and mark its sentiment. This could be achievable but it is both time consuming and inevitably subjective. Furthermore, the assessment of sentiment within given text may arise significant disagreement (Wiebe et al. (2001)). Due to these reasons, only the news for publicly traded companies is selected among the initial data set, and in order to be objective, efficient and effective in labelling the news, the rates of return are used as an indicator for the polarity.

After that, we get the data set of 38808 news item. Table 1 provides an overview of this data set. There are 38808 news stories from 347 different publicly traded companies. Maximum number of story for a company is 1687 whereas there are some companies who have just one story. Average number of stories per company is about 110.

Table 1- Descriptive statistics of news data for firms

Mean	110
Median	36
Mode	1
Range	1686
Minimum	1
Maximum	1687
Sum	38088
Count	347

Figure 2 shows the companies with the most news item in our data set. There are 5 firms that have over 1000 news stories. These are Turkish Airlines (THYAO) with 1687 news item; Turkcell (TCELL) with 1428 news item, Is Bank (ISBTR) with 1141 news item, Garanti Bank (GARAN) with 1087 news item and Koc Holding (KCHOL) with 1037 news item.

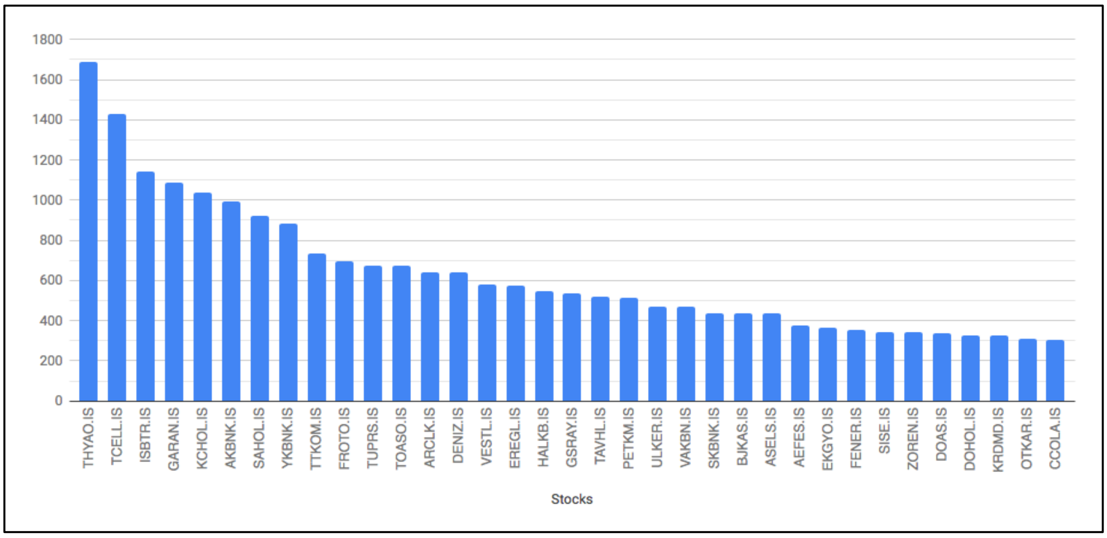


Figure 2- Number of news stories

The last limitation for the news stories is about the length of the news items. For this reason, we take only the news items which has the summary body more than 18 words.

We set this restriction in order to get more detail news about the firm and to construct more informative feature set in our machine learning algorithm.

After all restrictions, we end up a data set consists of 29302 news items. On average, the news items have 110 words. The maximum word counts that a news item has is 223. Table 2 provides an overview of this data set. The final news data set contains about 946 thousand of words for our analysis.

Table 2- Descriptive statistics of news data for the corpus set

Mean	32
Median	28
Mode	24
Minimum	18
Maximum	223
Sum	946471
Count	29302

2.2. ECONOMIC DATA

The other source that we use in this work is the performance data of stocks. For this purpose, we bought the daily stock price data from Borsa Istanbul Data Store. The data consists of opening and closing prices of all the stocks in BIST for the fiscal years between 1996 and 2018 (up to February). For the daily performance of BIST 100 data, we get historical data from investing.com Inc.

As described in previous section, the news data has no label. In order to be labelled initially, we match each story with its own stock. For each event date (reference day for news stories), we take the difference between the closing price of the related stock on the reference day (t) and the closing price of the preceding day (t-1) divided by the closing price of the stock (t-1). For example, if the news published on April 20, we take the difference of the closing prices on April 20 and April 19; divided by the closing price on April 19. This is the return for the stock on day (t).

For the period January 1996-November 2015, trades are executed in two trading sessions. So, the price data consists of two different sessions for each day. Session 1 runs in the morning and Session 2 runs in the afternoon. The biggest effect in absolute value is taken as an indicator of the news story on that day. The same procedure is applied to preceding day (t-1) as well. We add the return for day (t-1) as another benchmark for our labelling approach. In other words, we take the biggest result for the sessions on day (t) and compare to the outcome of day (t-1). Then, we reach the final decision about the sign (categorization) of the news item.

For the period November 2015[8]-February 2018, instead of sessions, we cover the intraday price changes for the related stocks (closing price - opening price of the stock on day t). We take into account of both the difference in closing prices of the stock and intraday performance of the stock. The biggest effect in absolute value was taken as an indicator of the news story. In short, we compared the absolute values of the rate of returns on day (t) and (t-1) and make a final decision about the polarity for the news story. For instance, the following item about the company, Is Bank Inc, appeared on 24 March 2015:

> ADNAN BALİ'DEN SERT SÖZLER! İş Bankası Genel Müdürü, banka hakkında son dönemde çıkan iddialara cevap verdi. "Kimse üzerimizden cephe genişletmeye çalışmasın." İş Bankası Genel Müdürü Adnan Bali, "Kimse kendi hedefi üzerinden ya da kendi hedefi uğruna gayri şeffaf medyalarda İş Bankası üzerinden, İş Bankası gibi bu ülkenin göz bebeği kurumları yıpratmayı denemesin." dedi.

At the closing on 24 March 2015, the price of the stock reached 1340 TRY, whereas at the closing on 23 March 2015 the price was 1330.5. Therefore, there is a positive price change of 0.7%[9] for the day 24 March 2015. On the other hand, for the preceding

---

[8] This is the date of Borsa Istanbul Management made revision in BIST.

[9] (1340-1330.5) / 1330.5 =0.007

day (23 March 2015), there is a significant price decrease compared to the closing price of day 20 March 2015. Returns on 23 March yields negative 10%[10] return. So, the news item is classified as being negative. The same reasoning applied to all news items while creating the training data set.

In this work, 2-year government bond of Turkey is taken as the risk-free rate for the market models created in Chapter 4. The daily historical data for risk-free rate is taken for the time period of 8/1/2006-2/28/2018.

Our analysis mainly consists of two parts. In Chapter 3, we initially discuss the methodology and then we build machine learning algorithms that distinguish negative and positive news among news stories. We use these algorithms to test the accuracy for the testing data set and build a new variable (news item dummy variable). Then, in Chapter 4, we build market models and show the effect of both positive and negative sentiments over stock prices. Additionally, we use the algorithm to earn profit from trade in different scenarios.

---

[10] (1330.5-1478.5) / 1478.5 = - 0.100

CHAPTER 3

TEXTUAL PROCESSING METHODOLOGY

Machine Learning is the application of computation and set of algorithms which helps to solve complex problems like voice recognition, disease detection, face and iris recognition.[11] Machine learning is also widely used in natural language processing in a broad context. While it could be as easy as counting word frequencies at one extreme, on the other extreme, it could be as complex as human understanding and expressions. More specifically, Natural Language Processing (NLP), the computer manipulation of any language, is the field of machine learning focuses on this issue.

In this chapter, we initially cover some problems regarding the data and how to overcome the problems in machine learning. Next, we discuss the labelling and feature selection methodology. In Section 3.5, initially, we describe how the news data is labelled automatically based on the assumptions held in previous chapter. Then we train and test our model and show how the scores vary with different models. Section 3.6 covers the results of the textual processing of the financial news data and provides a summary of the chapter.

In this work, we use Python Programming Language (Raschka (2014)); Scikit-Learn library (Pedregosa et al. (2012)) and NLTK library (Bird et al. (2009)) and related modules for the computation and for the selection of all the models.

[11] Interested readers can be referred to Keynote: Google I/O, 2018.

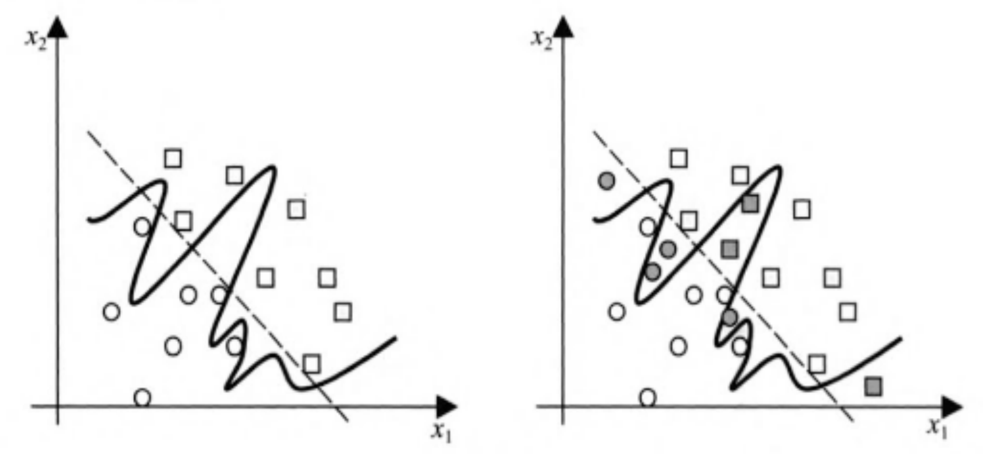
3.1. SOME PROBLEMS IN MACHINE LEARNING

"With four parameters I can fit an elephant
and with five I can make him wiggle his trunk."
John von Neumann

Although machine learning algorithms are strong in mathematical computation on different topics, one can get wrong results with these methods as well. The issues may arise from within data itself. There are some common undesirable possibilities in machine learning. These are underfitting and overfitting.

Underfitting occurs when a model doesn't include enough knowledge to precisely model real-life situations. Underfitting takes place if the model or algorithm reveals low variance with high bias. In general, underfitting is a problem about the sample size. When the sample size is too small or the model has low number of feature set, it is hard to come up with true results. So, the simplest solution for this problem is to increase the sample size. On the other hand, too specific models have other problem: overfitting. According to the Oxford Living dictionary overfitting is "*the production of an analysis which corresponds too closely or exactly to a particular set of data and may therefore fail to fit additional data or predict future observations reliably."*

In particular, if the model has more feature sets/ parameters, it represents exactly the data. For this reason, the model gives wrong results when it encounters with a new data. Likewise, overfitting takes place if the model or algorithm reveals high variance with low bias. The figures below show the overfitting and underfitting problem. In figure 3 (a) shows perfect separation of the training data by both low order linear line and high order non-linear curve. It is clear that both model has achieved to classify the item. Both model has the same training score and clearly separates the circles from squares.



(a)Training Data (b)Test Data

Figure 3- The overfitted algorithm with linearly separable observations[12]

On the other hand, the non-linear model sharply separates the items and overfits the training set and will probably performs worse on new/unseen examples. In Figure 3 (b), the filled circles and squares are the new/unseen examples in the new data. So, we can say that all the filled items are wrongly classified by the non-linear (overfitted) model. On the other hand, low order linear line performs better on this new (test) data. From this point of view, Burnham and Anderson (2002) state the following about underfitting and overfitting problem in modelling.

> ... an underfitted model would ignore some important replicable (i.e., conceptually replicable in most other samples) structure in the data and thus fail to identify effects that were actually supported by the data. In this case, bias in the parameter estimators is often substantial, and the sampling variance is underestimated, both factors resulting in poor confidence interval coverage. Underfitted models tend to miss important treatment effects in experimental settings. Overfitted models, as judged against a best approximating model, are often free of bias in the parameter estimators, but have estimated (and actual) sampling variances that are needlessly large (the precision of the estimators is poor, relative to what could have been accomplished with a more parsimonious model) (p. 32).

[12] Adapted from Kecman (2005). (p. 8)

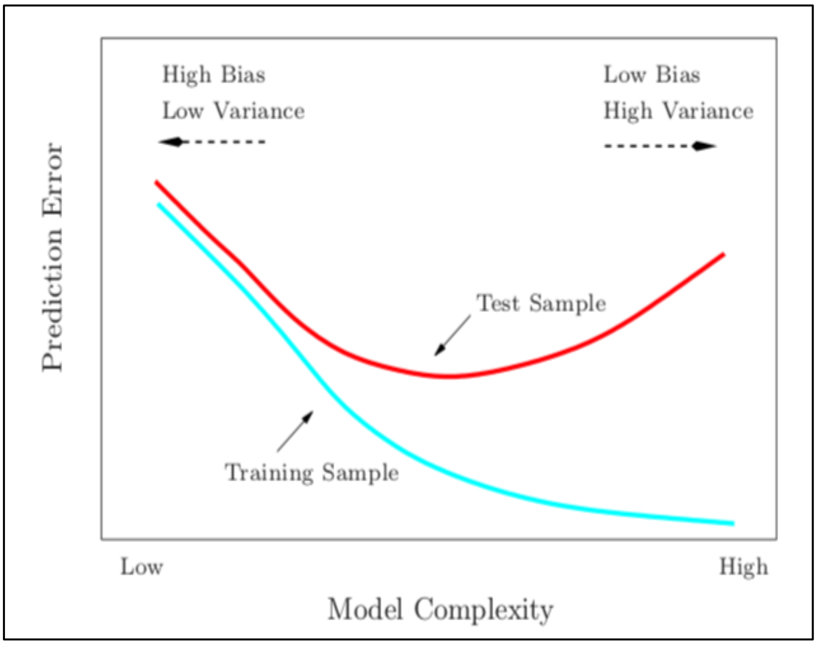


Figure 4- Test and Training Error as a function of model complexity[13]

Figure 4 shows the prediction error and model complexity relationship. The blue line shows the training score (error) and the red line shows the testing error. The training error has a tendency to decrease as the model complexity increases. The reason for this is that all the details (features) of training data can be captured by high number of estimators in complex models. Excessive fitting, on the other hand, reflects too closely to the training data set and will not result in acceptable outcomes (i.e., have large test error) when it encounters a new data set. The results will get high variance compared to the underfitted model. In contrast, if the model complexity is low, it will underfit and may have low variance-high bias, again resulting in poor outcomes. In low complex models, both training and testing error is high. As the model complexity (features) increases, testing error decreases at one point and reaches its minimum value with an intermediate level of complexity.

[13] Adapted from Hastie et al. (2009) (p. 220).

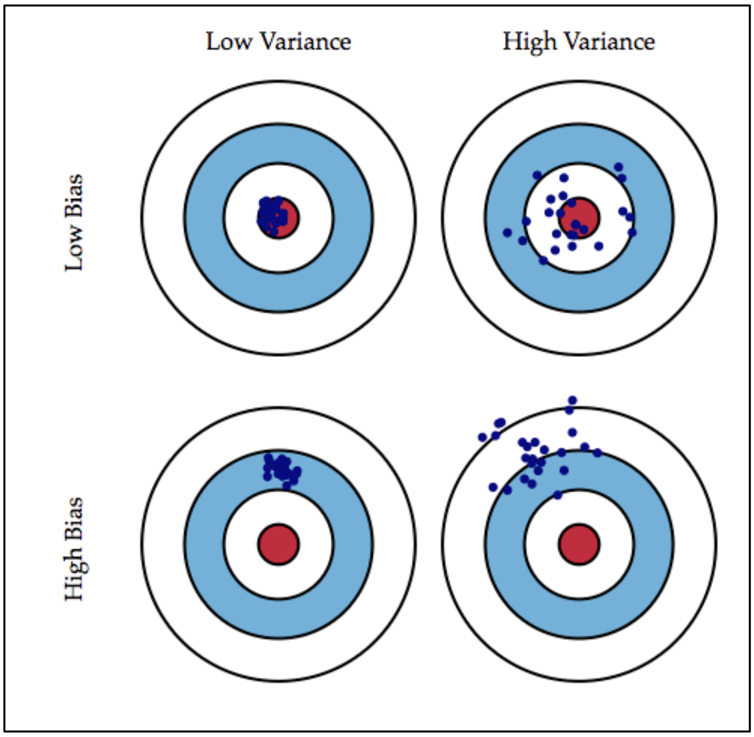


Figure 5- Bias and variance illustration[14]

Bias refers to prediction error of a model, while variance is the amount that the output of the model will change as the training data altered. The graphical illustration of bias and variance is shown in Figure 5. It is assumed that red areas are the true values of the population and blue points are the estimates of the model. As shown, the model bias increases as the estimation of the model gives higher prediction errors. Given model complexity, high variance occurs under the assumption that increases as the training data set changes.

The model bias can only be reduced by including richer features to the model and enhancing the variables in the model. Hastie et al. (2009). So, a moderate complex model with a rich feature set gives better accuracy results. The other solution for these underfitting and overfitting problems is randomly splitting dataset into training and

[14] Retrieved from http://scott.fortmann-roe.com/docs/BiasVariance.html (Accessed July 1, 2018) Fortmann-Roe, Scott. "Understanding the Bias-Variance Tradeoff".

test set and taking the cross-validation scores. For this purpose, we can create K training and test sets ("fold"). For each k in K, we apply k-1 folds to train data and the remaining subset (the last fold) as test data. After running all classifiers, we can measure the performance of the test set for each round. Then, we can choose the best learner based on this cross-validated performance. In the below figure, yellow parts are the validation (test) set and blue ones are the training sets. After each round, we can take average score for the performance of classifiers.

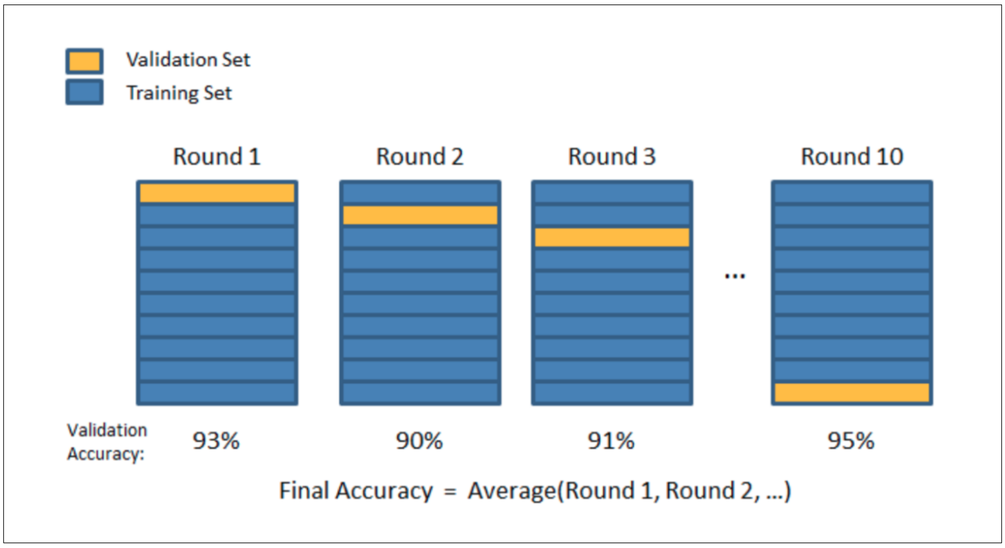


Figure 6- Visual Representation of Cross Validation K-Fold[15]

3.2. LABELLING THE NEWS DATA

The text data should be labelled in order to be used as an input in the supervised machine learning analysis. As mentioned before, one can read the columns and gives his opinion about the article whether it is positive or negative. There are some platforms like Amazon's MTurk, Figure-Eight and Upwork[16] for especially hiring

[15] Nelson J. (2018). Retrieved from https://medium.com/@josephofiowa , Accessed June 9, 2018.

[16] We used Upwork platform for cross-checking the human performance versus the machine learning classifiers we built with the data set used in Chapter 4.4.

people for labelling the data. This could be costly, inefficient and the results depends on freelancer's subjective opinion as well. Therefore, we prefer an efficient and objective technique for labelling the data.

In the following subsections, we discuss the most common content analysis methods in textual processing literature. These are the dictionary-based approach and automated labelling (learning from the data itself) method. To bridge qualitative and quantitative analysis, fully automatic approaches are widely used.

3.2.1. The Dictionary Based Approach

In the literature, some common list of words/ dictionaries, have been used to classify textual documents as positive or negative. Words in these dictionaries are pre-classified. After calculating the overall score for each class scaled by the total number of words in the corpus, a text corpus matched with these predefined classes is assumed as the label (e.g. positive, negative, neutral, uncertain) of the document. In short, these dictionaries provide a comparative measurement for each sentiment. After counting all the words associated with the predefined feature set (list of words), the label is assigned to the document. For instance, after calculating the overall score of words in a given text document, the lower percentage of negative words is assumed as a more optimistic sentiment, so the label could be positive; whereas the higher proportion of negative words shows pessimistic tone so, the label could be negative.

In the dictionary-based approach, the most important step is the selection of the right dictionary for the textual analysis since each dictionary has an orientation for specific purposes and refined by different domains. For example, the dictionary based on movie reviews underperforms in finance literature (Azar (2009)). Similarly, dictionaries for sociology and psychology literature may give wrong results in economics in terms of sentiment. "Tax", "cost", "capital", "board", "liability" or "foreign" for example, are negative words in one dictionary but not at all in finance literature. Similarly, "litigation", "restated", "misstatement" or "unanticipated" are classified negative words in finance but not negative in psychology dictionaries. Thus, deciding on right dictionary plays an important role in textual analysis. According to Loughran and McDonald (2015), some common dictionaries are inappropriate for determining the tone of the disclosures:

> We argue that Diction is inappropriate for gauging the tone of financial disclosures. About 83% of the Diction optimistic words and 70% of the Diction pessimistic words appearing in a large 10-K sample are likely misclassified. Frequently occurring Diction optimistic words like respect, security, power, and authority will not be considered positive by readers of business documents. Similarly, over 45% of the Diction pessimistic 10-K word-counts are not and no. (p. 20)

Some well-known dictionaries especially used in textual processing in accounting and finance literature are listed below.

1. The Henry (2008) Word List

In literature, the initial work for creating finance word list belongs to Henry (2008). She scanned earning press disclosures and created her dictionary based on these corpora. Although the list consists of low number of words (has just 85 negative words in the list) compared to the other sources, it gives better result in determining the tone of earning conference calls. Price et al. (2012) and Davis et al. (2015) assessed the Henry (2008) word list in their papers and showed the performance of the list.

2. General Inquirer Dictionary

General Inquirer Dictionary, often called Harvard GI Dictionary, dated back to Smith et al. (1967), The General Inquirer: A Computer Approach to Content Analysis paper. The latest version of the dictionary consists of 1045 positive words and 1160 negative words with 182 tag categories from four sources. These are:

(1) “Harvard IV-4” dictionary,

(2) “Lasswell” dictionary,

(3) “Several categories” constructed based on the social cognition work of Semin and Fiedler,

(4) “marker" categories for disambiguation[17].

Most of the works in literature use this dictionary because of the fact that complete list of the dictionary is free and readily available for use.[18] Tetlock (2007), Tetlock et al. (2008) and Kothari et al. (2009) for example, use the Harvard IV-4 as a tool for determining the polarity.

3. Diction

Diction is a text-analysis program which was originally created by Roderick P. Hart. The latest version consists of 10,000-word corpus, 686 positive and 920 negative words with 35 different tag categories and has user created dictionaries. Similar to the Harvard IV-4, diction is also widely used in finance world. On the other hand, the source of data is not freely distributed but can be purchased.[19] Rogers et al. (2011) and Davis et al. (2012) use Diction for determining the tone of verbal messages in their papers.

4. Loughran-McDonald Word List

The other well-known word list is the Loughran and McDonald (2011). They created six different word lists which are based on business background. These lists consist of 354 optimistic and 2,329 pessimistic words with eight different tags: “Negative”, “Positive”, “Uncertainty”, “Litigious”, “Constraining”, “Superfluous”, “Interesting” and “Modal”. Tetlock (2007), Garcia (2013) and Liu and McConnel (2013) use

---

[17] Disambiguation part consist of words like race which has several meanings such as a contest; moving rapidly; group of people of common descent and as an idiom “rat race”.

[18] See the webpage http://www.wjh.harvard.edu/~inquirer/homecat.htm, for further information about the descriptions of Inquirer Categories and Use of Inquirer Dictionaries.

[19] See the webpage for more details. http://www.dictionsoftware.com/ , Accessed June 1, 2018.

Loughran and McDonald (L&M) dictionaries to uncover the tone of the text documents. L&M word lists are freely available for use.[20]

5. Regressive Imaginary Dictionary

According to the Kovach Technical Services Company, "the latest version of the English Regressive Imagery Dictionary (RID) is composed of about 3200 words and roots assigned to 29 categories of primary process cognition, 7 categories of secondary process cognition, and 7 categories of emotions." These categories were derived from the work of Martindale (1975) and Martindale (1990). The data set can be purchased through the website[21].

6. WordNet

WordNet Miller (1995)[22], is a web-based dictionary developed by Princeton University. WordNet is available to researchers and commercial users free of charge. Thus, there are some other sources that use these lists to establish sentiment-based dictionaries such as SentiWordNet[23], "an automatically generated lexical resource that assigns to each synsets of WordNet a triplet of positivity, negativity and objectivity scores" (p. 1).

7. SentiTurk, Zemberek, Metu-Sabancı Treebank

Most of the sentiment dictionaries and word lists based on Indo-European languages. Due to its rich morphology, natural language processing is challenging in Turkish language. That's why the number of dictionaries is limited.

---

[20] https://sraf.nd.edu/textual-analysis/resources/, Accessed June 1, 2018.

[21] http://www.kovcomp.co.uk/wordstat/RID.html, Accessed June 1, 2018.

[22] The database can be gathered from https://wordnet.princeton.edu. Accessed July 11, 2018.

[23] The resource is described in details in this webpage: http://sentiwordnet.isti.cnr.it. Accessed July 11,2018.

Sentiturk, polarity lexicon data for Turkish language, created by Dehkharghani et al. (2016) through direct translation (mapping) from SentiWordNet. Sentiturk has three categories of emotions: positivity, negativity, and objectivity (neutrality) levels.

Metu-Sabancı Treebank, created by Oflazer et al. (2002), is another database for NLP. It has over 2 million words from various books, newspapers and magazines. Similar to the Metu-Sabancı Treebank, Zemberek, created by Akın and Akın (2007), is open source database for Turkish NLP. Both platforms provide valuable outputs for the users.

3.2.2. Creating Automated Word List

Dictionary based judgement is reliable since there is no human decision making. As for Turkish language, there is limited source for economics and finance literature. Thus, in our study we introduce a novel method of automatically labelling for the textual analysis. Since performance of a stock is an indicator for future news (Akbas et al. (2013)) and "the market reaction to a news is a good indicator for labelling financial news, and that a machine learning algorithm can be trained on those news to build models detecting price movement effectively" (Généreux et al. (2008)), we can measure the polarity of the news issued in the media based on the performance of the individual stock returns, instead of counting words.

As discussed in Chapter 2.2, we have discussed how we label "the news data" based on individual performance of the stocks. We also argue the techniques and features in sections 3.4. Feature Selection and 3.5. Training and Testing.

3.3. CLASSIFICATION METHODS

In machine learning, classifiers are the algorithms or mathematical functions that divide the input data into groups, classes or categories. In this thesis, we use three main classifiers for determining the sentiment of the news. These machine learning algorithms are: Naïve Bayes Classification, Decision Tree Classification and Support Vector Machine Classification. In this section, we introduce these machine learning techniques by following the calculations and formulations as presented in Pedrosa et al. (2011).

3.3.1. Naïve Bayes

Naive Bayes is one of the foremost efficient and effective algorithms in supervised machine learning. Basically, Naïve Bayes (NB) classifier assumes that each feature (i.e. each word of the news item) is generated independently of its position. In his paper Zhang (2004) showed that:

> …the dependence distribution; i.e., how the local dependence of a node distributes in each class, evenly or unevenly, and how the local dependencies of all nodes work together, consistently (supporting a certain classification) or inconsistently (cancelling each other out), plays a crucial role. Therefore, no matter how strong the dependences among attributes are, naive Bayes can still be optimal if the dependences distribute evenly in classes, or if the dependences cancel each other out. (p.1)

Naïve Bayes assumes that any attribute is independent from its position and there is no dependencies among attributes. We can explain the naïve assumption of NB classification using the following example. Let's assume that we have five attributes and E is a tuple of attribute values (x_1, x_2, x_3, x_4, x_5) where x_i is the value of attribute X_i. The attribute vector consists of binary values (0, 1) depends on the availability of attributes. For example, if attributes are E= ("increase", "decrease", "tax", "price", "ratios") and the news story is "Tax increase affected the price of cars." Then, the text represents as the vector (1,0,1,1,0). If the story is "Central Bank of Turkey has an option to decrease the required reserve ratios", then the story is represented as the vector (0,1,0,0,1). NB classification holds the assumption that the attribute "increase", i.e, is independent from those others ("decrease", "tax", "price", "ratios").

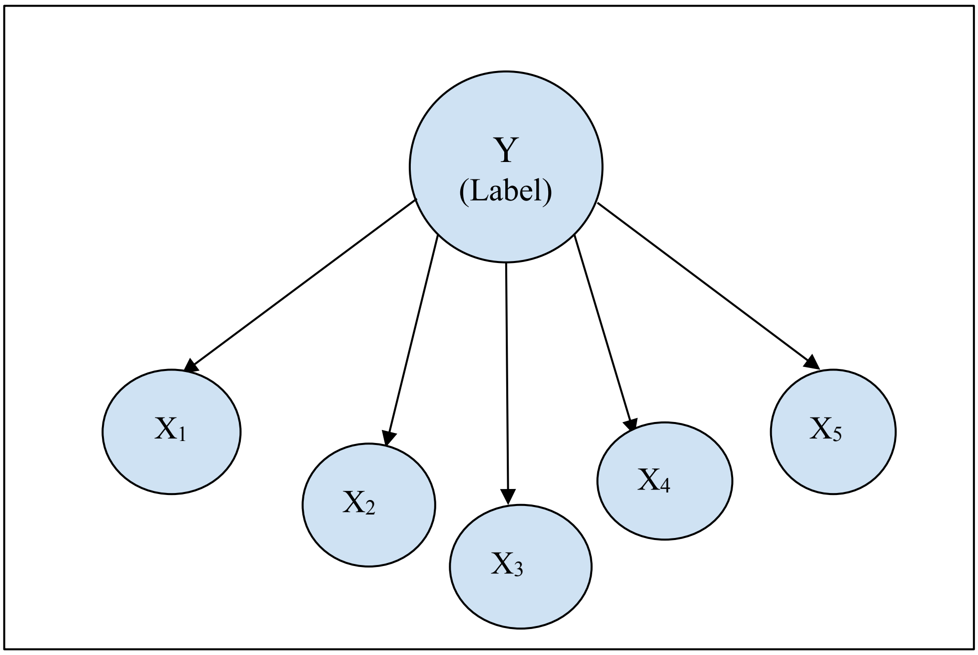


Figure 7- Naive Bayes Classifier

As shown in Figure 6, Y is a classification variable (i.e. it is 1 for positives and 0 for negatives) and y is the value of Y. Each of the attributes (X_i) is independent. The conditional probability distribution of the model is represented by this notation:

$P(X_1 = x_1, \ldots X_5 = x_5 | Y = 0)$

$P(X_1 = x_1, \ldots X_5 = x_5 | Y = 1)$

The mathematical expression of NB is as follows.

Given a class variable y and an attribute vector of a document ($x_1, x_2, x_3, , , x_n$),

$$P(\, y |\, x_1, \ldots., x_n \,) = \frac{P(y)\, P(x_1, \ldots., x_n |\, y\,)}{P(x_1, \ldots., x_n)}$$

The predicted class ($\hat{y}$) is estimated by the highest probability of occurring the given attribute vector. However, even in this 5-attribute simple feature set, there could be 32 possible feature vectors and 2 the binary classifications variable. This means that we have to learn these 64 parameters[24] with the training data. On the other hand, we have to increase the number of features to get better results from the model. It is clear that an increase in the number of features set makes it harder to learn with a limited number of observations. So, with its "independence" assumption of features, NB simplifies the problem. That is, the probability of each feature occurring in a document is independent of the occurrence of other features in a document.

$$P(X_i = x_i \mid Y = y)$$

We can explain this with our initial example. Let's assume that we have 5 binary features and a binary class variable Y. A binary feature has 2 possible instances (0 or 1) and corresponds 2 possible parameters for the class Y (0 or 1). Thus, instead of 64 parameters, this assumption relies on 4 . 5 = 20 parameters. This assumption is very useful when we have more features. If we have 100 features in the feature set, then we have to cover 400 instances in our training data. On the other hand, it is impossible to come up with these numbers of examples.

Given the "naive" independence assumption, then:

$$P(x_i \mid y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i \mid y)$$

For all *i* the algorithm simplified as:

$$P(y \mid x_1, \dots., x_n) = \frac{P(y) \prod_{i=1}^{n} P(x_i \mid y)}{P(x_1, \dots., x_n)}$$

Since $P(x_1, x_2, x_3, , , x_n)$ is constant given the input, then we drive the NB classifier as follows:

$$P(y \mid x_1, \dots., x_n) \propto P(y) \prod_{i=1}^{n} P(x_i \mid y) \Rightarrow$$

[24] The number of possible instances is calculated by $2^6 = 64$.

$$\hat{y} = \underset{y}{arg\,max}\, P(y) \prod_{i=1}^{n} P(x_i |\, y\,)$$

3.3.2. Gaussian Naive Bayes

NB Classifiers differ mainly by the "naive" independence assumption: the distribution of $P(x_i|\,y\,)$. Gaussian NB algorithm proposes the likelihood of the features as:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

where σ_y and μ_y are the parameters estimated by maximum likelihood.

3.3.3. Multinomial Naive Bayes

In Multinomial NB Classifiers, the distribution is represented by;

$$\theta_y = (\theta_{y1}, \ldots, \theta_{yn})$$

for each class of y, where n is the total number of features (words). θ_{yi} represents the naïve assumption for this particular NB classifier so that the probability $P(x_i|y)$ of feature *i* appearing in a sample document belonging to class y. θ_y is estimated by this smoothing function:

$$\hat{\theta}_{yi} = \frac{N_{yi} + \alpha}{N_y + \alpha n}$$

where α is the smoothing parameter;

$$N_{yi} = \sum\nolimits_{x \in T} x_i$$

is the occurrence number of feature *i* in the sample of class *y* in the training set (T);

$$N_y = \sum_{i=1}^{|T|} N_{yi}$$

is the total count of all the features for class y in the training set.

3.3.4. Bernoulli Naive Bayes

As for Bernoulli NB, the distribution function becomes:

$$P(x_i \mid y) = P(i \mid y)x_i + (1 - P(i \mid y))(1 - x_i)$$

The main difference between Multinomial NB and Bernoulli NB is that in Bernoulli NB there is a penalty term where non-occurrence of a feature i and it focuses on occurrences of a feature, not the counting. For this reason, Bernoulli NB might give better results on short text documents. We discuss the estimation results of these NB classifiers for the news data in Section 3.5.

3.3.5. Decision Trees

Decision Trees (DTs) are also common classifiers in machine learning methods. The aim of DTs is to establish a model that predicts the target value by if-then-else decision patterns. In simple, DT is a flowchart where shows labels for feature series. DT starts with root node and continues with decision nodes and leaf nodes. In this sequence, there could be a decision stump as well. Decision stump is the node that shows the label based on single feature. It decides whether the feature is in the text or not. Main advantage of DTs is it is easy to visualize the decision pattern. On the other hand, it is not time efficient if the feature set has high number of items.

The real example taken from the data in this study is shown in Figure 8. This DT classifier starts with the root node X_5 which is the 6th element of the feature set.[25] It is the word of "görmek" which means "to see" in English. If a text document has the word, then it goes to the left node, else continues with the right node. The feature X[77] represents the word "giant" in English. If the text document doesn't contain the word "to see" and has the word "giant", then it is classified as Positive; if not, it goes the next leaf. The feature in this leaf is "share". If the text has the word "share" then it will be categorized as Positive, else Negative.

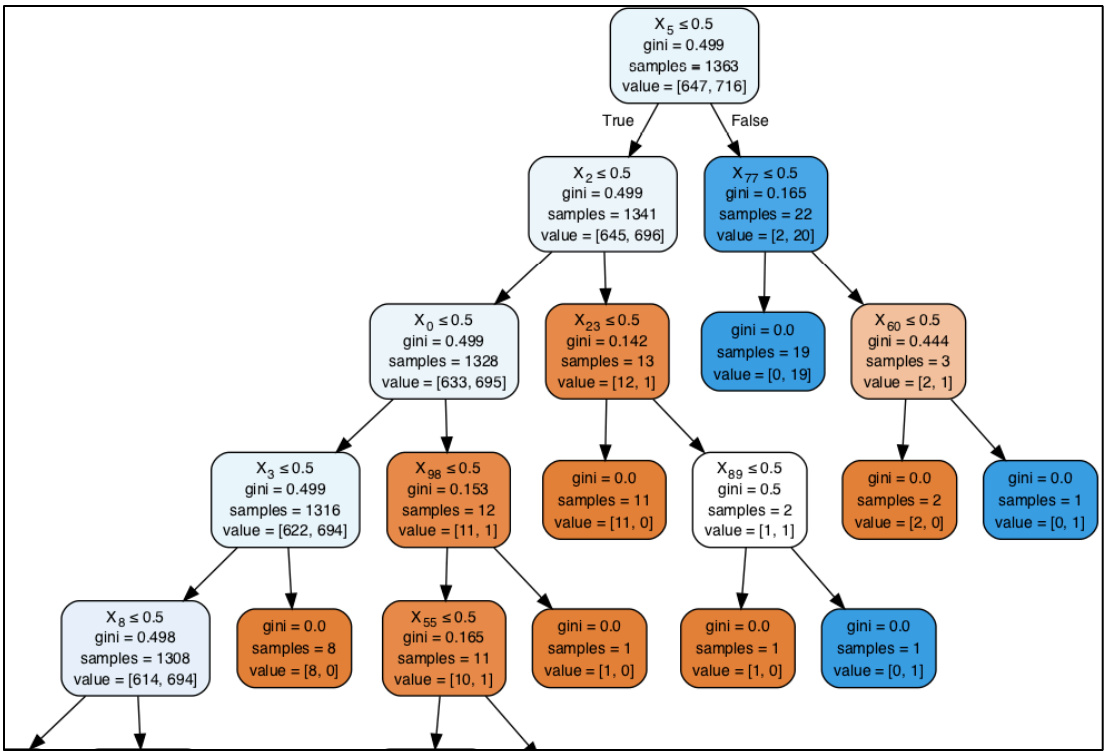


Figure 8- Visual representation of Decision Tree

3.3.6. Support Vector Machine

Support Vector Machine (SVM) is a supervised learning method used for classification and outlier detection by generating input-output mapping algorithms

[25] The first element in the feature set is represented by X[0].

from a given labelled training set. SVMs use hyperplanes that has the largest distance to the nearest points of classes (in our example labels, e.g. positives or negatives). SVM uses the term “kernel” for the separation line. If the separation line is linear, then it is called SVM with linear kernel. Dedicated readers to this topic can be referred to Kecman (2005).

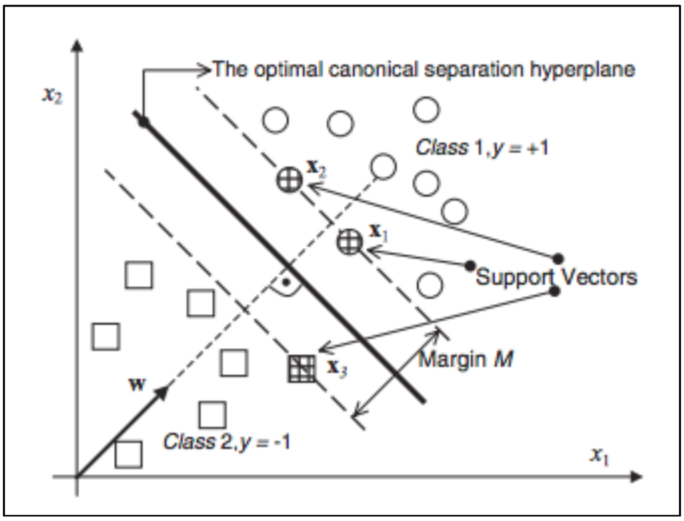


Figure 9- Support Vector Machine Hyperplane[26]

Figure 8 shows the decision boundary for two classes. The hyperplane in this figure is the black line, which has the largest margin to the nearest points; support vectors. Circles (x_1 , x_2) are support vectors for class 1 (positive sentiments); x_3 on the other hand, is the support vector from class 2 (negative sentiments).

SVM is applicable to more classes as well. As shown in the figure below, SVM is applied for three different classes. Each decision boundary has shown as different colors in the Figure. This is an example of how classes are separated by SVMs with different kernels.

[26] Retrieved from Kecman (2005) (p.15)

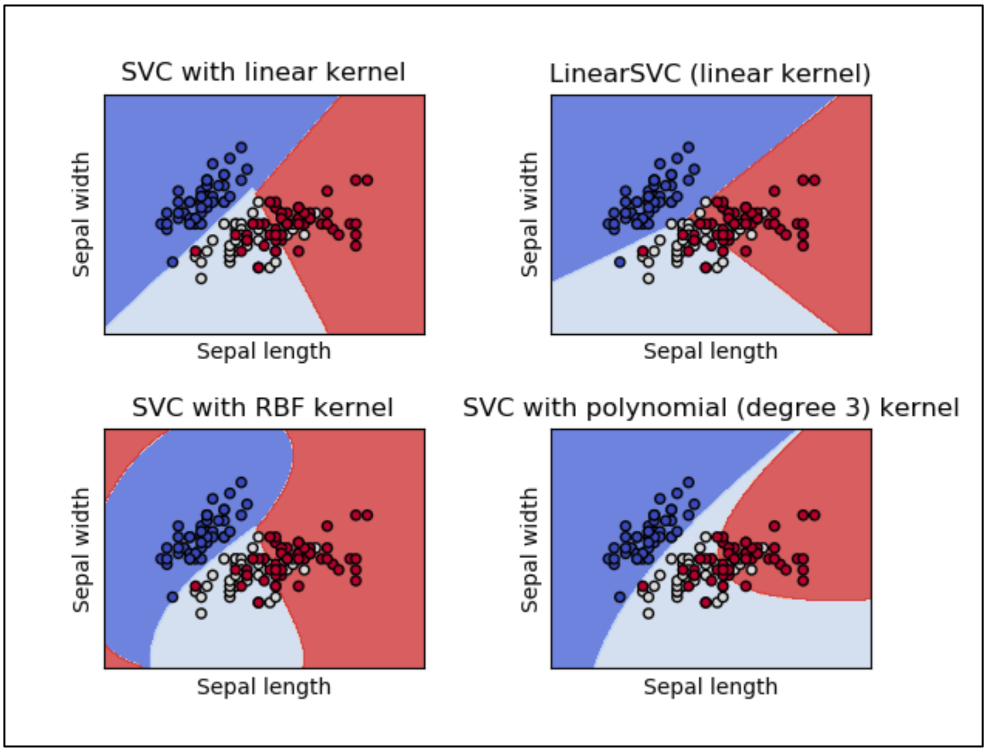


Figure 10- SVM with different kernels[27]

3.4. FEATURE SELECTION

3.4.1. Natural Language Processing[28]

Natural Language Processing (NLP) is a computer-based technology deals with everyday communication by humans and sources such as newspapers, articles, emails, handwritten documents and so on. NLP has the key importance in scientific, economic and social research. From spam detection to chat bots, from machine translation to sentiment analysis, NLP applications are widely used for dealing with human utterances.

[27] Reprinted from Scikit-Learn Tutorial Page., retrieved from http://scikit-learn.org/stable/modules/svm.html, Accessed May 11, 2018.

[28] Natural Language Processing is also called as "Computational Linguistics".

There are many NLP tools for text analysis. These are the useful tools help us to do some work before starting analysis. In order to “tokenize”, “lemmatize”, “spell check” we need NLP tools. They also provide some other facilities. In this study, I use the Natural Language Toolkit (NLTK) for Python programming language. Zemberek and Metu-Sabancı Treebank are the other valuable tools for NLP.

3.4.2. Tokenization

The first step for text processing is making sensible chunks. In an article, for instance, paragraphs, sentences, words and punctuations are the tokenized outputs. Until the raw text data converted into the list of words, this process is called tokenizing.

If the news story is:

> “*EXPECTATIONS: EXPECTS REVENUE GROWTH TO BE AROUND 20 PERCENT IN LIRA. EXPECTS EBITDA MARGIN TO BE AROUND 10 PERCENT FOR 2018. EXPECTS LONG TERM EBITDA MARGIN TO BE AROUND 11 PERCENT.*”

then, the tokenized version of this news story would be:

> ['EXPECTATIONS', ':', 'EXPECTS', 'REVENUE', 'GROWTH', 'TO', 'BE', 'AROUND', '20', 'PERCENT', 'IN', 'LIRA', '.', 'EXPECTS', 'EBITDA', 'MARGIN', 'TO', 'BE', 'AROUND', '10', 'PERCENT', 'FOR', '2018', '.', 'EXPECTS', 'LONG', 'TERM', 'EBITDA', 'MARGIN', 'TO', 'BE', 'AROUND', '11', 'PERCENT', '.']

After tokenization, the next step for text analysis is the normalizing of the text data. This is the step for text data to lowercase so that the difference between “For” and “for” is ignored.

3.4.3. Lemmatization

In text analysis, lemmatizing is the algorithmic process of removing the suffixes. As a result of this process, the distinction between “women” and “woman” is ignored, for instance. After lemmatization, “better” turns to “good”, its lemma. Since Turkish language has complex structure, in this study we do not use this process.

3.4.4. Feature Set

Choosing the right, informative, representative and independent features is an important step in effective machine learning. For this reason, after the normalization process, the source text should be tokenized. From this list of tokens, some of the features should be ignored. These could be the ending words, punctuations or transition words.

For our analysis, initially, we construct our feature set by all the words that appear at least 50 times in the corpus. This process results in 2100 different features. We use this feature set in all of the learning models to vectorize the text. Table 3 provides the top 50 frequented words appeared in the corpus.

Table 3- The top 50 frequent words in the corpus

Word	Word	Frequency	Word	Word	Frequency
milyon	million	14292	büyük	big	2464
yüzde	percent	9483	aldı	took	2241
bir	one	6718	trilyon	trillion	2206
için	for	6481	şirket	company	2117
milyar	billion	6305	grubu	group	2096
yeni	new	5137	yönetim	governance	2070
ilk	first	4849	geçen	last	2027
yıl	year	4719	ciro	endorsement	1843
genel	general	4475	daha	more	1812
bin	thousand	4403	satış	sales	1757
en	most	4149	enerji	energy	1702
türk	turkish	4059	yapı	structure	1609
kredi	credit	4020	hava	air	1571
holding	holding	3636	üretim	production	1516
dolar	dollar	3548	hedefliyor	aims	1499
yatırım	investment	3426	etti	done	1495
türkiye	turkey	3366	ytl	ytl	1463
net	net	3094	anadolu	anatolia	1458
lira	lira	2975	kadar	until	1440
müdür	director	2958	satın	buy	1418
olarak	as	2905	halka	public	1406
tl	try	2902	şirketi	company	1388
iş	business	2654	son	last	1386
kurulu	board	2521	toplam	total	1385
dolarlık	dollar	2511	aylık	monthly	1373

3.5. TRAINING AND TESTING THE DATA

As described in Chapter 2, initial constraints on the news data source ends up with 29302 news items for the years 1996-2018. Firstly, in this section, we label the data as positive or negative based on individual performance of the stock. In order to be objective in labelling, we take the stock performances as a benchmark. Also, we take some other measures against the stories. That is, a news item of a company is labelled as positive when the associated stock has at least 9.3% increase and has better performance than the overall stock market. Likewise, if a stock has worse performance than the market and its price declines 7.6% or more, then the story is labelled as negative. These high thresholds are chosen for strengthened the likelihood that the price changes in the stock would have changed as a result of news stories, not the consequences of market dynamics or random fluctuation. We use lower threshold for negative news in order to get close number of stories compared to positive ones. Additionally, we only consider the stories which have 18 words or higher. Hence, these limitations result in 47% negative news out of 1,436 news stories.

Then, we create "Training" and "Testing" set from this labelled data. The training set that we create consists of randomly selected 80% of 1,436 news stories. We train our learning models by this training data set. Then, for each learning models, we get the estimation accuracy scores from testing data set and reach conclusion about compared performances. Furthermore, we split 1436 stories into another two parts with regard to the publishing dates. This time, we set the "training" data by news stories published between the years of 1996 and 2015 and the "testing" data consists of stories published afterwards. This time-based selection of training and testing data set helps to examine the performance of the learners for the recent news data. Table 4 shows the associated number of news item for both training and test set under each of the selection processes.

Table 4- Training and Testing set

	Training Data	Test Data
Random Selection	1148	288
Time Based Selection	1363	73

During the training phase, pairs of feature set and labels are added into different machine learning algorithms in order to generate prediction models for each learner. The figure below illustrates this process. In Figure 11-a, the "input" symbolizes the news items created for the training set. Based on what our feature set, the "feature extractor" constructs vector representation of features in a single story. Then, both feature vectors and predefined labels are fed into the machine learning algorithm. Consequently, this algorithm produces a classifier model. In 11-"(b) Prediction" schema, the "input" represents the news stories in the testing set. The same feature extractor turns the text into feature vector. In this phase, since the stories are accepted as unseen, there is no label for them. Only feature vectors from testing set are then fed into the classifier model which creates predicted labels for each of the story.

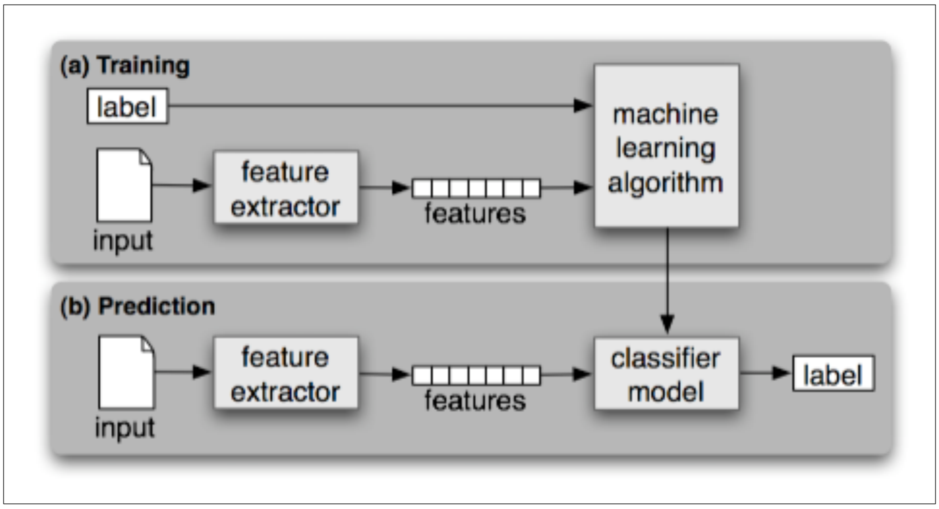


Figure 11- Machine Learning Algorithm Train and Test Data[29]

[29] Reprinted from Bird and Loper (2009) (p. 222).

In the next section, we focus on the performance of Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Decision Tree and Support Vector Machine classifiers. Then, we discuss how good or bad is these results of each classifier.

3.6. RESULTS AND CHAPTER SUMMARY

The simplest but important method to evaluate the performance of a learner is the accuracy, which measures the percentage of correctly labelled inputs in the test set. The other important metrics generally used in classification and sentiment analysis are Precision, Recall and F-Measure. While "precision" is the fraction of relevant items among the retrieved items, "recall" is the fraction of relevant items that have been retrieved over the total amount of relevant items. F-Measure takes into consideration both of "precision and "recall" by calculating the harmonic mean of them. In order to calculate these metrics, the number of items are taken into account. These are True positives, True negatives, False positives (Type I errors), False negatives (Type II errors).

True Positive (TP) **:** Correctly labelled as positive
True Negative (TN) **:** Correctly labelled as negative
False Positive (FP) **:** Incorrectly labelled as positive
False Negative (FN) **:** Incorrectly labelled as negative

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\boldsymbol{Precision} = \frac{TP}{TP + FP}$$

$$\boldsymbol{Recall} = \frac{TP}{TP + FN}$$

$$\boldsymbol{F\ Measure} = 2 * (Precision * Recall)/(Precision\ + Recall)$$

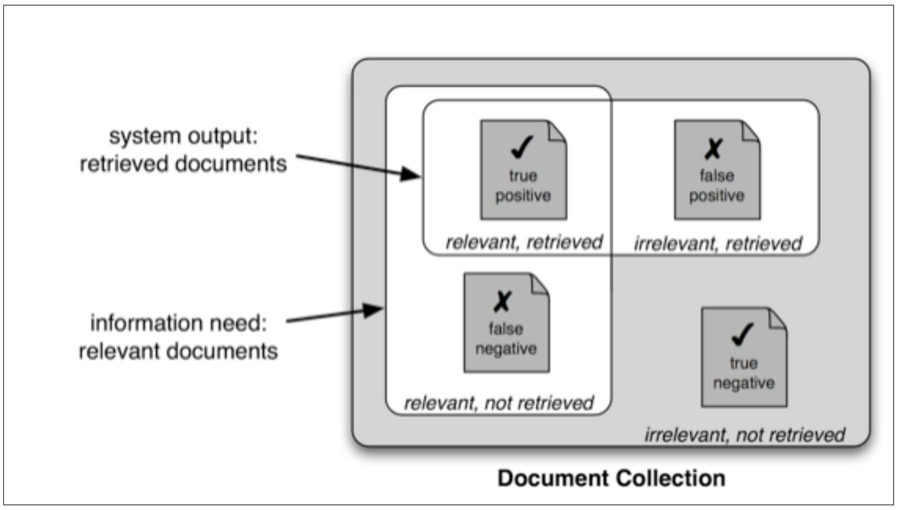


Figure 12- Precision, Recall and Confusion Matrix[30]

The figure above illustrates the confusion matrix. The horizontal white rectangle represents the precision area and the vertical white one is for recall area in this confusion matrix. The diagonal entries indicate accuracy. These are the labels correctly predicted by the models.

As we discussed in Section 3.5, we split the data into 2 different sets. The first one is based on random selection (80% of the data for training and 20% for testing); the second one is time-based selection. We initially select the “random selection” training data set to train and test the models. Then, we continue with “time-based selection” training data set to run the same performance tests. The scores are calculated by using cross validation method. In Figures 13-16, learning curves of Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Decision Tree and Support Vector Machine classifiers are shown. In these figures, the red line is the training score line and the green line represents the cross-validation score line.

Figure 13, under page, provides that as the sample size increases the learning curve decreases about 10%. This is basically because of the fact that with low number of

[30] Amended from Bird and Loper (2009) (p. 240).

samples the model overfits to the data, as the sample size increases it becomes harder for the model to fit the data - so results in a lower training score, but the model works better with new training examples. The same pattern of learning curves is seen, in Figure 14, for Bernoulli Naïve Bayes classifier.

On the other hand, Figure 15 and Figure 16 reflects another form for training curves. That is, since the training score line in these learners fits the data well, there is no or little change on the accuracy rate as the sample size increases. As for cross-validation scores, both Decision Tree classifier and Support Vector Machine classifier obtain similar results to Naïve Bayes classifiers.

As can be seen from the graphs in Figure 13-16, all the models have similar accuracy scores. How good or bad these results are an important question at this point. As the population of the labelled news data has 53% of positive in the training data, an algorithm that always estimates "positive" results in 53% accuracy rate. So, any classifier performs greater accuracy than this critical point, is accepted as a good classifier. It is clear that all the models outperform this threshold.

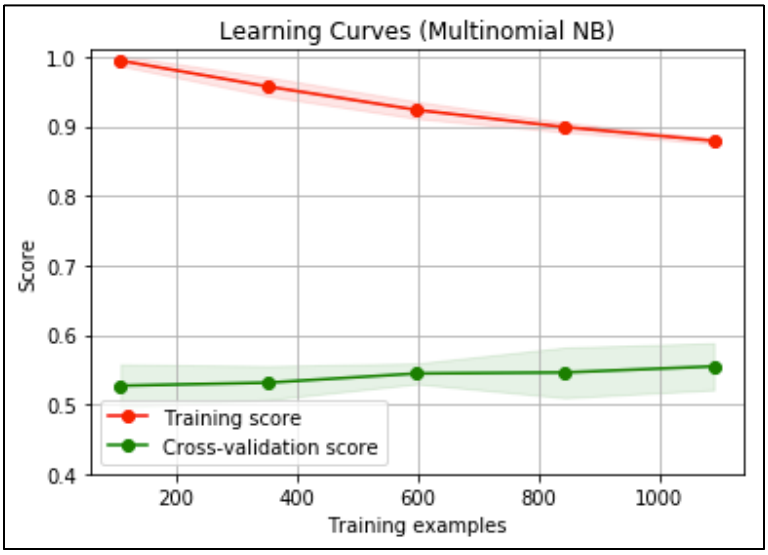


Figure 13- Learning curves of Multinomial Naïve Bayes Classifier

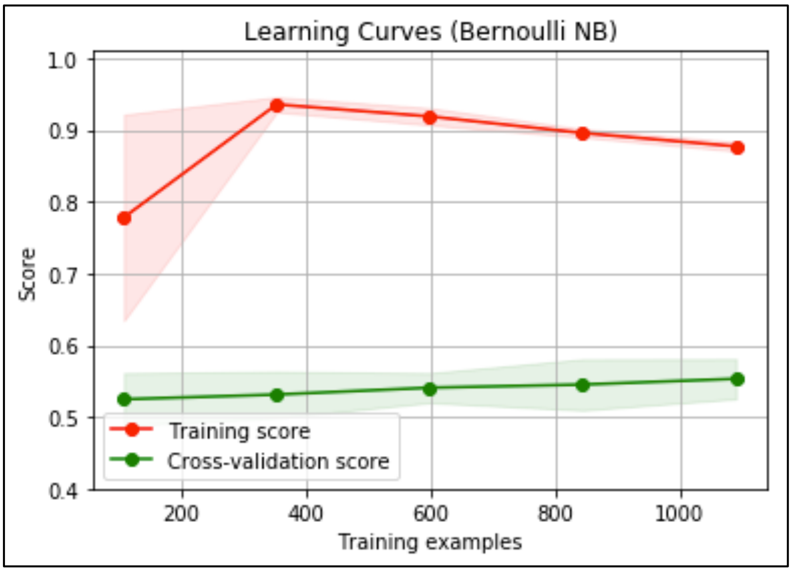


Figure 14- Learning curves of Bernoulli Naïve Bayes Classifier

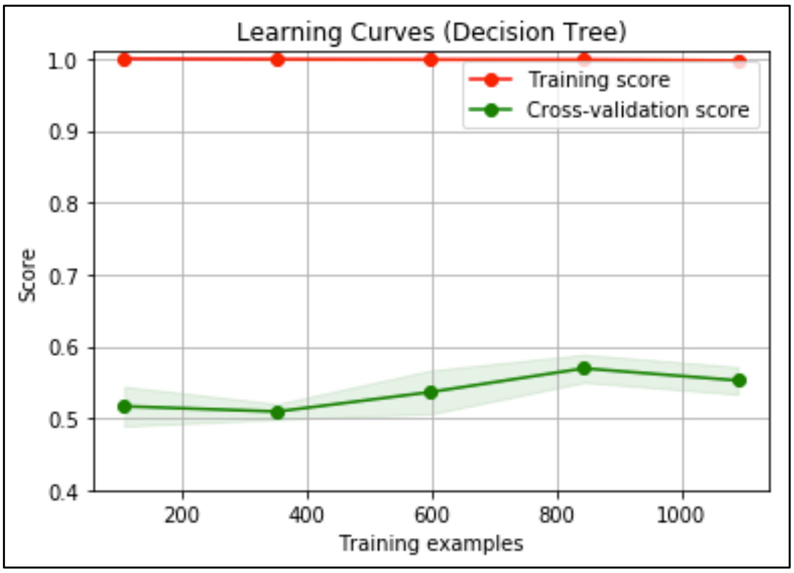


Figure 15- Learning curves of Decision Tree Classifier

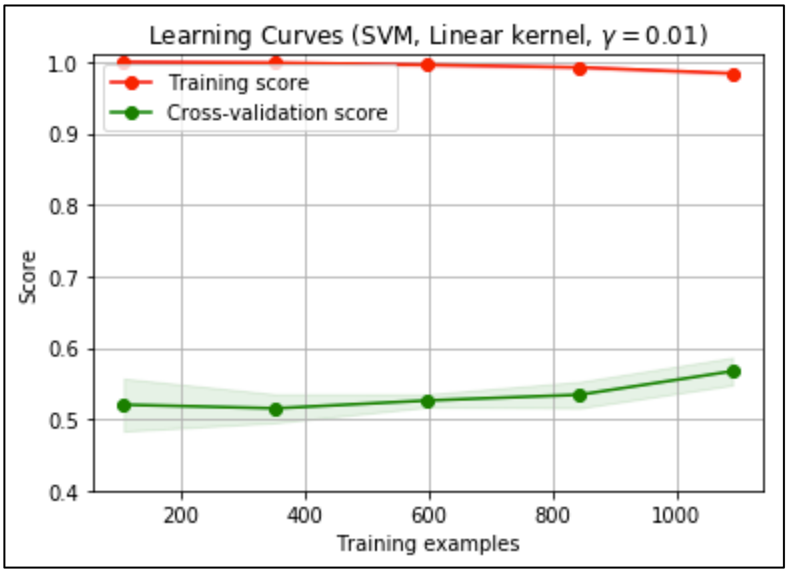


Figure 16- Learning curves of Support Vector Machine Classifier

Figure 17-20 show the learning curves of learners based on "time-based" training data set. It is clear that the learning curves of both selection method follow similar pattern. As the number of training data in time-based selection is higher than those in random selection one, the accuracy rate has increased slightly.

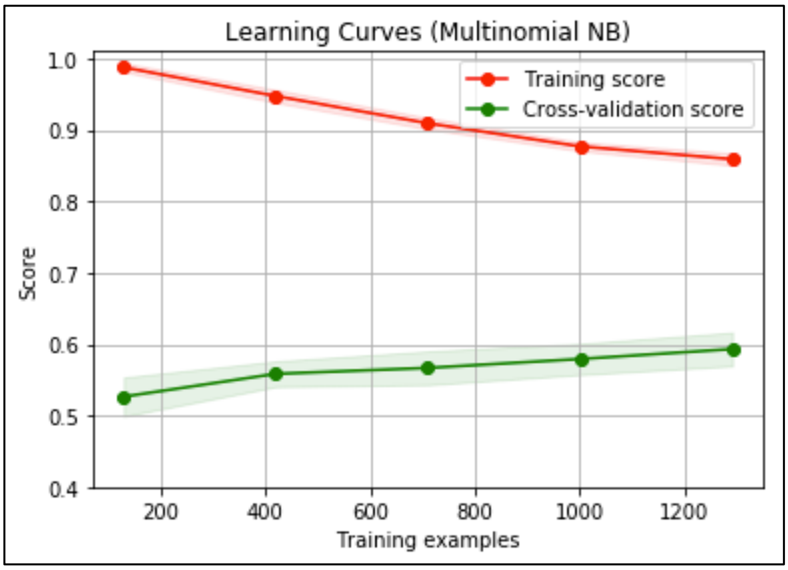


Figure 17- Learning curves of Multinomial NB Classifier (Time Based)

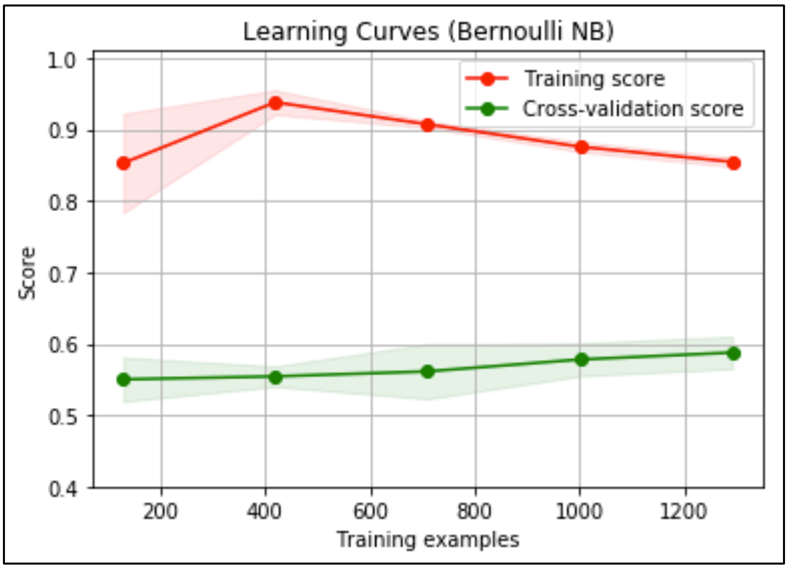


Figure 18- Learning curves of Bernoulli NB Classifier (Time Based)

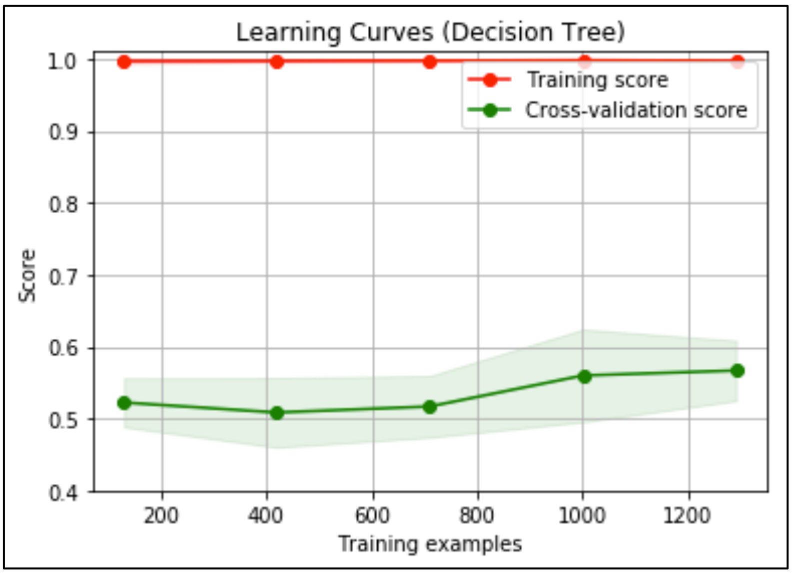


Figure 19- Learning curves of Decision Tree Classifier (Time-Based)

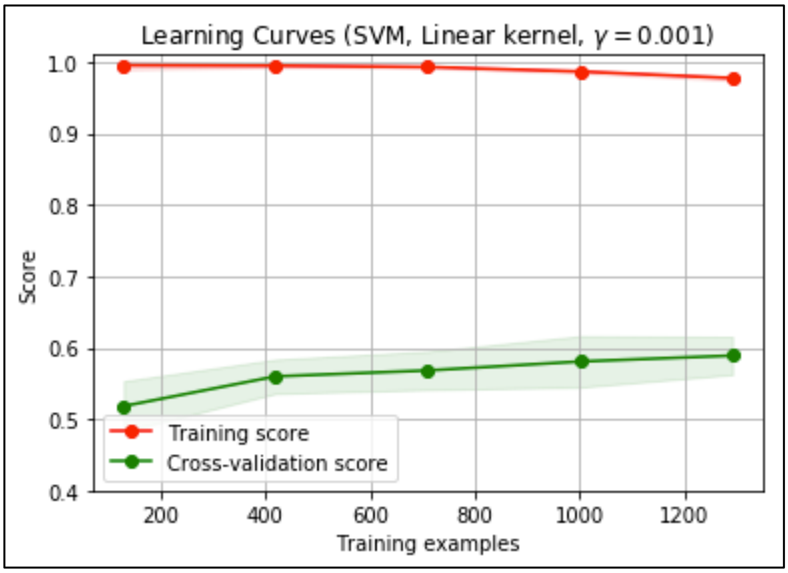


Figure 20- Learning curves of Support Vector Machine Classifier (Time Based)

It is clear that, in Figure 13-20, cross validation score fluctuates as new data comes in. This is because the models (over)fits the data but they do not generalize well. Even though we have high training accuracy, the model might not work well with new data. Our models seem to have high variance due to the complex structure (high number of features) for the small amount of data. The gap between the two curves shows this issue, too. The solution is simple. Either getting more data or using a simpler model. For this reason, first I increase the training set of random selection method from 80% to 90% with decreasing the number of features[31] from 2100 to 100. A new feature set is chosen by the highest information gain by the Naïve Bayes learner. Consequently, the new feature set has the most 100 informative words.

Table 5 shows the most informative features in our initial training data set. The feature “sürüyor” (“to last” in English), for instance, has the maximum effect in the corpus. This word is seen 8.1-fold more in negative news items than positive ones. Likewise, “sanayi” (“industry” in English) is an important feature for positive news item which is used in 6.6-fold more compared to negative stories.

[31] Recall that we construct our feature set by all the words that appear at least 50 times in the corpus. This process results in 2100 different features.

Table 5- The most Informative 15 Features (entire corpus)

Feature	English Meaning	Sign	Effect
sürüyor	last	Negatives	8.1
sanayi	industry	Positives	6.6
vurdu	hit	Negatives	6.5
geriledi	declined	Negatives	6.5
çıkan	outgoing	Negatives	6.5
görmeye	seeing	Positives	6.1
teknoloji	technology	Positives	6.1
düştü	fell	Negatives	5.8
çıkardı	released	Positives	5.4
artırım	issue	Negatives	5.0
bedelsiz	bonus	Negatives	5.0
karşı	against	Negatives	5.0
ağırlıklı	weighted	Negatives	5.0
az	little	Negatives	5.0
ihale	bid	Positives	4.9

These are the markers for detecting polarity of the news. It is clear that most of them are negative markers. A closer analysis of these results indicates interesting insights. The markers of positive features have limited with compared to the negative ones; thus, positive sentiments are characterized by the non-appearance of negative markers as well. This finding is similar to Koppel and Shtrimberg (2006) and Baumeister et al. (2001).

Table 6- Training and Testing set

	Training Data	Test Data
Random Selection	1292	144
Time Based Selection	1363	73

As demonstrated in Table 6, the number of news item in the training data set increases to 1292. After making changes towards models' complexity, the overfitting problem dissolves and accuracy scores increased significantly. As can be seen from the figure below, since the training and cross- validation score lines get close to each other, the models reveal low variance now. The graphs below show the new learning curves of each models. It is clear that Multinomial Naive Bayes classifier has relatively yields better scores and reaches 70% of accuracy rate in some validation chunks (light green area).

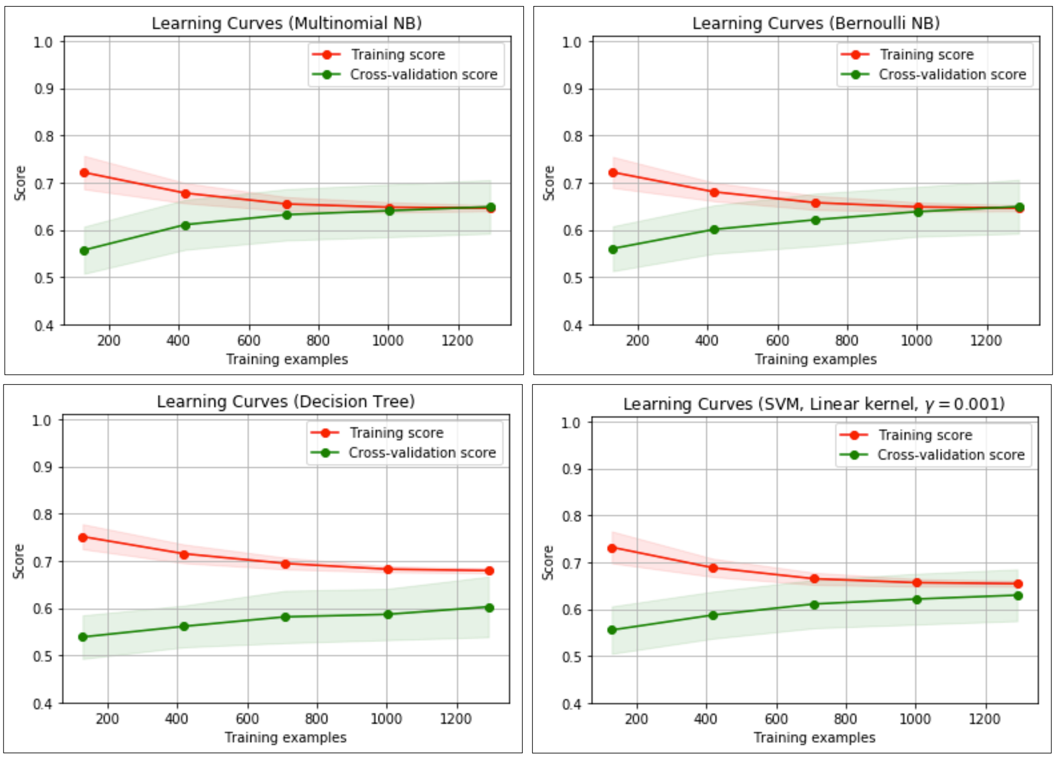


Figure 21- Learning curves of multiple classifiers (Random Based Selection)

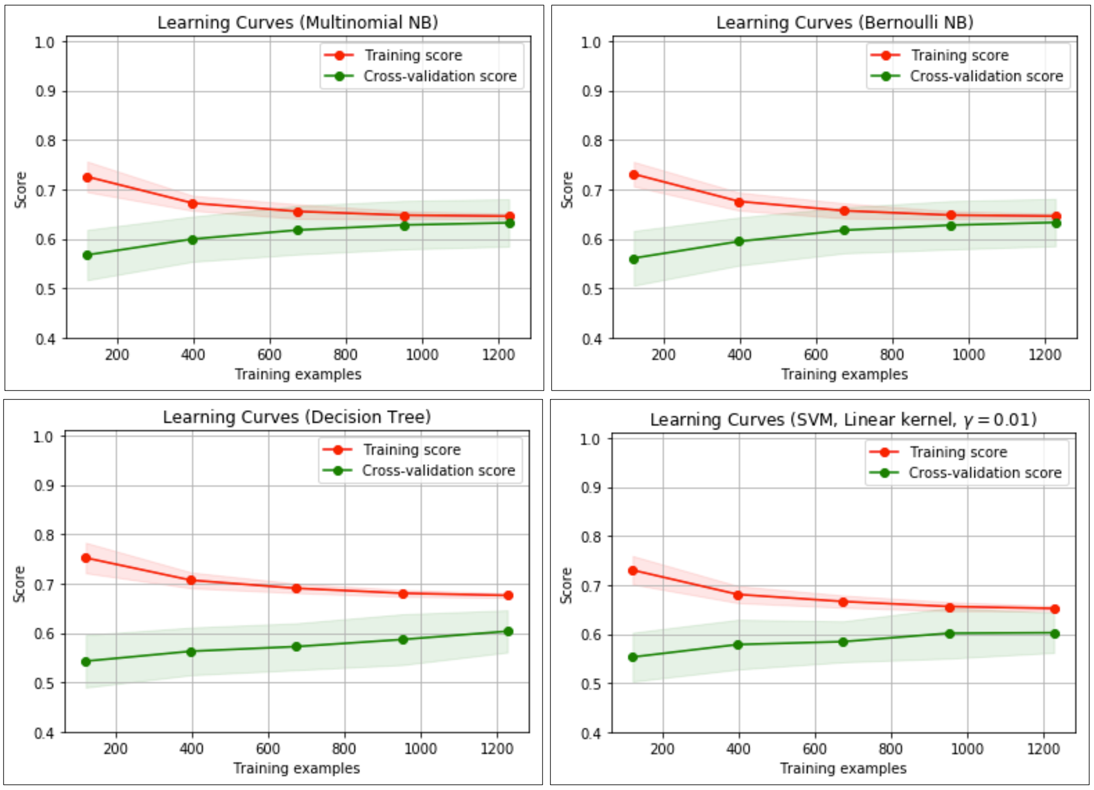
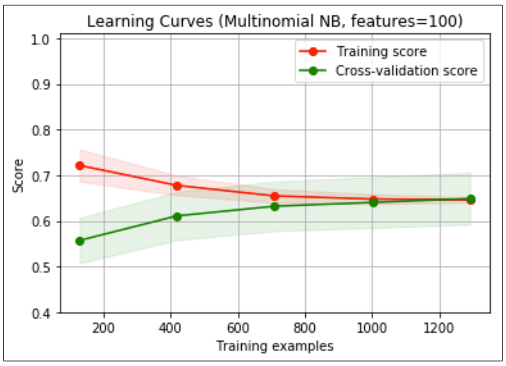


Figure 22- Learning curves of multiple classifiers (Time Based Selection)

If we increase the features, by adding another 50 most informative words, keeping the training and testing set constant, what happens to the learning curves? Do they still keep the same yield level or the accuracy rates improve/decline? Figure 23 confirms that that there is no barely noticeable improve in accuracy rates.

In general, it is expected that as the new data comes in to the training set, a model yields better results. For this reason, we can make adjustment to the restrictions and measure the cross performances. That is, if we modify the “word count” restriction (i.e. it is 18) or change the constraints for labelling (i.e. stock performance lower bounds (-7.6% ; 9.3%) while keeping the population of negatives on same level (i.e. it is 47%) and the testing set constant, we can get different accuracy results as well.



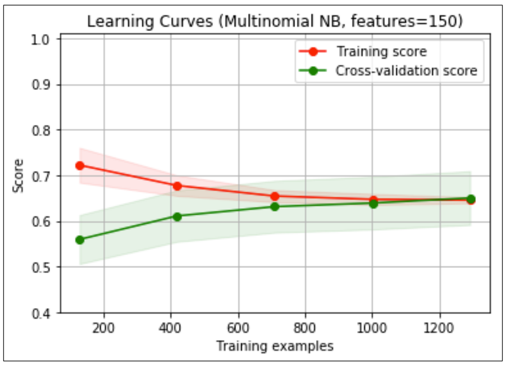


Figure 23- Comparison of Multinomial Naïve Bayes Classifiers

A summary of result in Table7 shows how the accuracy score of Multinomial NB and SVM classifiers change with respect to the certain restrictions. In this analysis, testing set is held constant to 2016-2018 corpus. As can be seen from the table, Multinomial NB yielded accuracy of 71.43% when we study the news item has greater than 18 words, and the related stock has the lower bounds of -9.10% and 10.30%. Other learners, including Decision Tree and Bernoulli NB, yield relatively lower scores compared to Multinomial NB classifier.

Table 7- Accuracy scores from different restrictions

Negative Population	News Data	Training Data	Testing Data	Word Count	Negative Bound	Positive Bound	Multinomial NB	SVM (Linear)
48.18%	909	818	91	18	-9.10%	10.30%	**71.43%**	53.57%
47.61%	899	809	90	18	-9.20%	10.30%	70.91%	67.27%
48.58%	883	794	89	19	-9.10%	10.30%	70.91%	56.36%
48.19%	855	769	86	20	-9.10%	10.30%	70.37%	**68.52%**
47.89%	831	747	84	21	-9.10%	10.30%	70.37%	68.52%
47.02%	889	800	89	18	-9.30%	10.30%	70.37%	66.67%
48.38%	833	749	84	18	-9.40%	10.40%	70.37%	66.67%
48.00%	873	785	88	19	-9.20%	10.30%	70.37%	55.56%
47.43%	818	736	82	18	-9.50%	10.40%	69.81%	67.92%
48.02%	756	680	76	18	-9.70%	10.50%	69.81%	67.92%

Table 8- Score metrics for Naive Bayes Classifier

Accuracy	71.43%	**Precision**	73.30%
Recall	89.10%	**F-Measure**	80.40%

Table 9- The Most Informative 15 Features, (1996-2015 corpus)

Feature	Feature (English)	Sign	Effect
işbirliği	cooperation	Positive	6.1
kaybetti	lost	Negative	5.9
görmeye	seeing	Positive	5.6
temettü	dividends	Negative	5.2
borcunu	the debt	Positive	4.8
teknoloji	technology	Positive	4.8
çağrı	call	Positive	4.8
düşüş	decline	Negative	4.5
bedelsiz	bonus	Negative	4.5
çimento	cement	Positive	4.4
düşüşle	declining	Negative	4.4
bilanço	balance	Positive	4.2
çıkardı	released	Positive	4.1
oran	rate	Negative	3.8
tavan	ceiling	Positive	3.7

Table 8 describes the related metrics of the accuracy score for Multinomial NB classifier. The recall for positive stories is high (89.1%) but precision is lower (73.3%); thus, the misclassified news stories are mostly those negative news which has no distinct word in the feature set. Also, if we question how the overall performance of the MNB classifier is, the answered is F-Measure, which is 80.4%. It is interesting to note that the most informative features of the new corpus set, given in Table 9, has different percentages with compared to the former one. Positive markers have reached 60% of the top informative features, which was formerly 33%. This finding corresponds with the low precision score we have.

In conclusion, as a result of our novel approach, we achieve to get valuable information from unlabeled data. We made a list of effective words for finance domain. These words are also labelled as negative or positive. The entire process is managed automatically. The automatic labelling approach that we built in this work gives some incentives about generating a large corpus easily. The other important thing that we added this process is the collection of market sense as a determiner of the polarity. As a result of this, we gathered more reliable judgement.

We have found that our classifiers can learn to detect sentiment in the news stories with a moderate success. Compared to several learners, we got better results from Naïve Bayes classifiers: about 70% of accuracy and 80% score for F-Score. The accuracy scores can be improved by several ways. The simplest one is to increase the news data set. This can be achieved by getting data from other external sources in finance domain. Also, financial text data can be obtained through social media and forums. The second way is about the labelling process. The automatic labelling approach that we built in this work takes the market sense as a determiner of the polarity. With the help of human reviewers, it could be gathered more consistent labels.

In the next chapter we begin a new discussion about the market dynamics and financial news. We describe three related models: Single Factor Market Model, Capital Asset Pricing Model, and Single Factor Market Model (with panel data). Since we get good Recall and F-measure scores, we could question to develop a strategy which could weight these learners to make investments based on news stories. Is there any opportunity to get profit from this algorithm in the stock market? This question is also answered in Chapter 4.

CHAPTER 4

THE EFFECT OF FINANCIAL NEWS ON BIST STOCK PRICES

In this chapter, we begin by examining the related literature and highlight the efficient market hypothesis (EMH) and forms of EMH. Then, we explain the data structure and methodology that are used in the research. As a result of machine learning algorithms that we established in the previous chapter, we can convert the news item into mathematical form and estimate its polarity. More specifically, we apply the news item as an input to the market analysis. In Section 4.3, we run different market models by examining how the stock return data is affected by the news. We also discuss whether the coefficients of dummies of negative and positive news are statistically significant and theoretically consistent in explaining the changes in expected return of individual stocks or not. In the last section, it is questioned whether the algorithm that categorizes the news as positive or negative provides the trade incentive.

4.1. THEORETICAL BACKGROUND

The Efficient Markets Hypothesis (EMH) states that the stock prices always adjust immediately to any change in the market and it is not possible to earn higher return permanently, since the prices are always in equilibrium after a quick response. Based on empirical evidence, Ehrhardt and Brigham (2010) emphasize that *the strong form of EMH* does not hold. The main statement of this hypothesis is earning abnormal returns in stock market is impossible due to the direct reflection of public or private information to prices. *The weak form of EMH* states that the past information of stock prices is rapidly reflected to current stock prices is claimed to be highly efficient in the market while *the semi-strong form of EMH* with the notion of current market prices reflect all the publicly available information is asserted to be reasonably efficient.

Market Model (MM) and Capital Asset Pricing Model (CAPM) clearly demonstrate how the individual stock return is affected by the market return. MM is a regression of expected market return on expected return of an individual stock, the beta coefficient of market return demonstrates how much the individual stock return changes when expected market return changes by one unit. In comparison to MM, CAPM, estimates the effect of excess return on market return $[R_m - R_f]$ on excess return on stock of a firm $[R_i - R_f]$ with the inclusion of risk free return. CAPM developed by Sharpe (1964) and Lintner (1965) and its building blocks are constructed by Markowitz's model of portfolio choice (1959). CAPM is a widely preferred simple method for asset pricing in terms of describing the relation with risk and expected return.

<u>MM:</u>

$$R_i = c + \beta_{i,m} R_m + \varepsilon$$

<u>CAPM:</u>

$$R_i = R_f + \beta_{i,m} [R_m - R_f] + \varepsilon$$

$$[R_i - R_f] = c + \beta_{i,m} [R_m - R_f] + \varepsilon$$

$$(c \text{ should be zero for CAPM to hold})$$

CAPM is criticized to be insufficient and weak in supporting the empirical findings due to the several simplifying assumptions. Fama and French (1992) constructed a model with the cross-section regression approach to explain expected stock returns with the size, earnings-price, debt-equity and book-to-market ratios. They created Three-Factor Model for expected returns of stocks with the inclusion of difference between the returns on diversified portfolios of small and big stocks (SMB: small minus big) and the difference between the returns on diversified portfolios of high and low book to market ratios of stocks (HML: high minus low):

$$[R_i - R_f] = c + \beta_{i,m} [R_m - R_f] + \beta_{i,s} (SMB) + \beta_{i,h} (HML) + \varepsilon$$

(c should be zero for Three Factor Model to hold)

Three factor model, however, is claimed to fail in explaining the momentum effect of stock prices. “Stocks that do well relative to the market over the last three to twelve months tend to continue to do well for the next few months, and stocks that do poorly continue to do poorly. This momentum effect is distinct from the value effect captured by book-to-market equity and other price ratios. Moreover, the momentum effect is left unexplained by the three-factor model, as well as by the CAPM” (Fama and French (2004)).

As for the impact of news, Shiller (2000) discusses its role in the stock markets by stating following:

> As we shall see, news stories rarely have a simple, predictable effect on the market. Indeed, in some respects, they have less impact than is commonly believed. However, a careful analysis reveals that the news media do play an important role both in setting the stage for market moves and in instigating the moves themselves. (p.71)

In the following sections, we bring the hypothesis of Market Models and CAPM to show the effect of financial news on stock prices.

4.2. EMPRICAL DATA AND METHODOLOGY

The economic data is partly described in Chapter 2. In this section, the “News Data” is created as a new dummy (binary) variable. That is, if a stock has a news on a given day, the News Data is 1, otherwise, it is 0. We generate another variable from the news data. This time, we drive a “negative” and a “positive” variable, based on the learner that we build in Chapter 3, not on the stock’s performance. Negative/Positive News variable consists of binary variables (either 0 or 1) depending on the Multinomial Naïve Bayes learner estimates.

A portfolio is created from companies that have the most news for the time period of 1/8/2006-2/28/2018. Some restrictions are applied to the companies while selecting the portfolio. The first one is about the establishing date of the firms. We choice only those firms established before 2006. The second restriction is about the number of news that firms have. This time, we set a bottom line of 300 number of news. As a result of these restrictions, the remaining 15 companies is shown in Table 10.

Table 10- Portfolio stocks

Stock Code	**Company**	**Number of News**
AKBNK	Akbank T. A.Ş.	682
ARCLK	Arçelik A.Ş.	356
ASELS	Aselsan Elektronik Sanayi ve Ticaret A.Ş.	348
BJKAS	Beşiktaş Futbol Yatırımları Sanayi ve Ticaret A.Ş.	310
DENIZ	Denizbank A.Ş.	552
FENER	Fenerbahçe Futbol AŞ.	310
FROTO	Ford Otomotiv Sanayi A.Ş.	437
GARAN	T. Garanti Bankası A.Ş.	754
KCHOL	Koç Holding A.Ş.	570
SKBNK	Şekerbank T. A.Ş.	374
TCELL	Turkcell İletişim Hizmetleri A.Ş.	1025
THYAO	Türk Hava Yolları A.O.	1215
TOASO	Tofaş Türk Otomobil Fabrikası A.Ş.	428
ULKER	Ülker Bisküvi Sanayi A.Ş.	379
VESTL	Vestel Elektronik Sanayi ve Ticaret A.Ş.	320

The other economic data variable is the risk-free rate. In this work, the return of 2-year government bond of Turkey is taken as the risk-free rate for the models. The figure below shows the risk-free rate (in logarithmic scale) between the years 2006 and 2018.

Figure 24- Risk free return rate (logarithmic scale)

4.3. MARKET MODELS

CAPM, despite of its unsuccessful explanatory power for empirical studies, still exists as one of the simple risk-expected return relationship measuring model. In this section, single factor Market Model, CAPM and Market Model (with Panel Data) are estimated for the expected returns of the firms that are subject to this study, but with the inclusion of dummy variables representing the news as explanatory variables. The news about a firm are considered one of the causes of the change in its own stock prices. Investors do not only follow the changes in the total market return and related macro indicators, but they also strictly and instantly follow the news of the firms to get the immediate profits realized due to the news distorting the existing equilibrium.

To make the model estimations, return values of the stock variables are required and they are calculated as given below:

$$R_i = \ln\left[\frac{P_{i,t}}{P_{i,t-1}}\right] \qquad R_m = \ln\left[\frac{P_{m,t}}{P_{m,t-1}}\right] \qquad R_f = \frac{\ln[1 + i/100]}{360}$$

i : interest rate as a percentage value.

$P_{i,t}$: daily price of the stock i at time t and $R_{i,t}$ is the daily return of stock i at time t.

$P_{m,t}$: daily price of BIST100 as a market proxy at time t and $R_{m,t}$ is the daily return of BIST100 as a market proxy at time t.

The models are estimated with Ordinary Least Squares (OLS) in which the return variables should be stationary. Augmented Dickey Fuller Test results show that all return values of the stocks and the market proxy are stationary with the intercept term and zero lag, but the return of the government bond is nonstationary.

Table 11- Augmented Dickey Fuller Test Results

R_{AKBNK}	-52.753	$R_{AKBNK} - R_f$	-52.700
R_{ARCLK}	-50.587	$R_{ARCLK} - R_f$	-50.565
R_{ASELS}	-52.827	$R_{ASELS} - R_f$	-52.817
R_{BJKAS}	-51.808	$R_{BJKAS} - R_f$	-51.804
R_{DENIZ}	-48.971	$R_{DENIZ} - R_f$	-48.953
R_{FENER}	-49.951	$R_{FENER} - R_f$	-49.920
R_{FROTO}	-49.343	$R_{FROTO} - R_f$	-49.336
R_{GARAN}	-55.220	$R_{GARAN} - R_f$	-55.195
R_{KCHOL}	-52.227	$R_{KCHOL} - R_f$	-52.192
R_{SKBNK}	-51.010	$R_{SKBNK} - R_f$	-51.017
R_{TCELL}	-51.738	$R_{TCELL} - R_f$	-51.756
R_{THYAO}	-53.975	$R_{THYAO} - R_f$	-53.966
R_{TOASO}	-50.306	$R_{TOASO} - R_f$	-50.296
R_{ULKER}	-52.435	$R_{ULKER} - R_f$	-52.919
R_{VESTL}	-51.844	$R_{VESTL} - R_f$	-51.803
R_m	-53.167	$R_m - R_f$	-53.151
R_f	-1.992		
Note: The critical values for ADF test are -3.432, -2.862, -2.567 for the confidence levels of 99%, 95% and 90%, respectively.			

The correlations of the daily return variables of 15 stocks, the market proxy and the return of risk free asset for the time period of 01/08/2006- 02/28/2018 are given in the table below. The return of BIST100 has correlations higher than 40% for the returns of 10 stocks. The highest correlation of the return of the market proxy index is observed with R_{GARAN} and it is 68%. On the other hand, the highest correlation in the table is 72% and it is between R_{GARAN} and R_{AKBNK}.

Table 12- Correlation Matrix (%)

	R_m	R_f	R_{THYAO}	R_{FROTO}	R_{TOASO}	R_{ARCLK}	R_{DENIZ}	R_{VESTL}	R_{TCELL}	R_{ULKER}	R_{SKBNK}	R_{BJKAS}	R_{ASELS}	R_{FENER}	R_{GARAN}	R_{KCHOL}	R_{AKBNK}
R_m	100	-3	26	42	47	50	26	43	47	41	45	24	23	21	68	58	63
R_f	-3	100	-3	-1	-2	-4	-4	-8	2	-3	1	-1	-3	-6	-4	-5	-9
R_{THYAO}	26	-3	100	18	21	19	13	22	18	19	22	13	11	12	29	26	25
R_{FROTO}	42	-1	18	100	50	40	17	34	31	34	33	21	19	22	43	45	39
R_{TOASO}	47	-2	21	50	100	46	20	39	33	37	38	24	21	21	47	51	44
R_{ARCLK}	50	-4	19	40	46	100	21	42	36	36	37	21	20	21	50	52	46
R_{DENIZ}	26	-4	13	17	20	21	100	24	19	26	23	13	12	14	27	23	27
R_{VESTL}	43	-8	22	34	39	42	24	100	32	35	38	23	20	19	48	44	43
R_{TCELL}	47	2	18	31	33	36	19	32	100	30	33	19	18	13	45	42	41
R_{ULKER}	41	-3	19	34	37	36	26	35	30	100	35	20	19	18	42	41	38
R_{SKBNK}	45	1	22	33	38	37	23	38	33	35	100	21	17	19	49	43	44
R_{BJKAS}	24	-1	13	21	24	21	13	23	19	20	21	100	12	28	25	22	23
R_{ASELS}	23	-3	11	19	21	20	12	20	18	19	17	12	100	11	22	20	20
R_{FENER}	21	-6	12	22	21	21	14	19	13	18	19	28	11	100	22	20	20
R_{GARAN}	68	-4	29	43	47	50	27	48	45	42	49	25	22	22	100	60	72
R_{KCHOL}	58	-5	26	45	51	52	23	44	42	41	43	22	20	20	60	100	58
R_{AKBNK}	63	-9	25	39	44	46	27	43	41	38	44	23	20	20	72	58	100

4.3.1. Single Factor Market Model

MM:

$$R_i = c + \beta_{i,m} R_m + \beta_{i,n} D_n + \beta_{i,p} D_p + \varepsilon$$

D_n : dummy variable for the negative news and D_p is the dummy variable for the positive news.

$\beta_{i,m}$: market return coefficient of stock i

$\beta_{i,n}$: coefficient of the dummy of negative news for stock i.

$\beta_{i,p}$: coefficient of the dummy of positive news for stock i.

The regression results for the Single Factor MM in table below clearly show that the expected returns of all stocks are significantly affected by the expected market return. The coefficients of dummies of negative and positive news are mostly statistically significant and theoretically consistent in explaining the changes in expected return of individual stocks. In general, negative news is effective in reducing the expected returns of the stocks while positive news increases the expected returns of the stocks.

R^2 values in Single Factor MM models express the portion of the expected return of stock *i* explained by the expected market return and the dummy variables of the news. The highest R^2 is 47% and it belongs to Single Factor MM of R_{GARAN}. In fact, R^2 also represents the systematic risk of the return of the stock stems from the changes and the dynamics of the market return. (1 - R^2) shows the unsystematic risk of the stock which can be eliminated with a construction of diversifiable portfolio in which negatively correlated stocks are preferred to reduce the total risk of the portfolio. MM estimations given in the table below display that the systematic risks of expected returns of individual stocks are considerably low except R_{AKBNK} and R_{GARAN}. The unsystematic or diversifiable risks are high in general for these stocks and it is possible to create portfolios with lower risks with diversification of the stocks.

Table 13- Regression Results for Single Factor MM[32]

AKBNK	Ri =	0.0001	+ 0.704 Rm	+ 0.001 Dp	- 0.004 Dn	R^2 =0.408
	stdev	(0.0003)	(0.0166)	(0.0008)	(0.0009)	
	t ratio	[0.302]	[42.388]	[1.600]	[-4.686]	
ARCLK	Ri=	0.0002	+ 0.51 Rm	+ 0.006 Dp	- 0.005 Dn	R^2 =0.266
	stdev	(0.0003)	(0.0166)	(0.0011)	(0.0012)	
	t ratio	[0.649]	[30.706]	[5.191]	[-4.372]	
ASELS	Ri=	0.0007	+ 0.491 Rm	+ 0.007 Dp	- 0.008 Dn	R^2=0.055
	stdev	(0.0007)	(0.0408)	(0.0025)	(0.0033)	
	t ratio	[1.044]	[12.04]	[2.887]	[-2.316]	
BJKAS	Ri=	0.0023	+ 0.381 Rm	+ 0.004 Dp	- 0.016 Dn	R^2=0.083
	stdev	(0.0005)	(0.0282)	(0.0019)	(0.002)	
	t ratio	[4.637]	[13.495]	[2.088]	[-8.068]	
DENIZ	Ri=	0.0012	+ 0.363 Rm	+ 0.01 Dp	- 0.007 Dn	R^2=0.089
	stdev	(0.0005)	(0.0264)	(0.0014)	(0.0016)	
	t ratio	[2.499]	[13.763]	[6.991]	[-4.3]	
FENER	Ri=	0.0006	+ 0.263 Rm	+ 0.006 Dp	- 0.008 Dn	R^2=0.059
	stdev	(0.0004)	(0.0226)	(0.0016)	(0.0018)	
	t ratio	[1.52]	[11.616]	[4.098]	[-4.606]	
FROTO	Ri=	0.0006	+ 0.45 Rm	+ 0.005 Dp	- 0.005 Dn	R^2=0.184
	stdev	(0.0003)	(0.0188)	(0.0011)	(0.0012)	
	t ratio	[1.911]	[23.91]	[4.67]	[-4.085]	
GARAN	Ri=	0.0012	+ 0.733 Rm	+ 0.001 Dp	- 0.002 Dn	R^2=0.470
	stdev	(0.0003)	(0.015)	(0.0007)	(0.0008)	
	t ratio	[4.281]	[48.74]	[0.84]	[-2.384]	
KCHOL	Ri=	-0.0001	+ 0.589 Rm	+ 0.002 Dp	- 0.003 Dn	R^2=0.343
	stdev	(0.0003)	(0.0157)	(0.0008)	(0.0009)	
	t ratio	[-0.198]	[37.51]	[2.211]	[-2.985]	
SKBNK	Ri=	0.0015	+ 0.569 Rm	+ 0.004 Dp	- 0.007 Dn	R^2=0.210
	stdev	(0.0004)	(0.0216)	(0.0013)	(0.0016)	
	t ratio	[4.041]	[26.389]	[2.829]	[-4.43]	
TCELL	Ri=	0.0001	+ 0.445 Rm	+ 0.003 Dp	- 0.005 Dn	R^2=0.244
	stdev	(0.0003)	(0.0163)	(0.0007)	(0.0008)	
	t ratio	[0.326]	[27.238]	[4.223]	[-6.471]	
THYAO	Ri=	0.0004	+ 0.508 Rm	+ 0.006 Dp	- 0.004 Dn	R^2=0.074
	stdev	(0.0008)	(0.0387)	(0.0015)	(0.0017)	
	t ratio	[0.452]	[13.131]	[3.649]	[-2.42]	
TOASO	Ri=	0.0006	+ 0.54 Rm	+ 0.007 Dp	- 0.006 Dn	R^2=0.244
	stdev	(0.0003)	(0.0191)	(0.0011)	(0.0013)	
	t ratio	[1.826]	[28.281]	[6.414]	[-4.837]	
ULKER	Ri=	0.0013	+ 0.438 Rm	+ 0.007 Dp	- 0.005 Dn	R^2=0.188
	stdev	(0.0003)	(0.0183)	(0.0011)	(0.0014)	
	t ratio	[4.154]	[23.914]	[6.429]	[-3.817]	
VESTL	Ri=	0.0019	+ 0.557 Rm	+ 0.004 Dp	- 0.003 Dn	R^2=0.192
	stdev	(0.0004)	(0.0218)	(0.0014)	(0.0017)	
	t ratio	[4.983]	[25.585]	[2.784]	[-1.893]	

[32] There are three insignificant coefficients of dummies and they are shown in red colour in the table.

4.3.2. Capital Asset Pricing Model

The table below displays CAPM regression results for each firm. All constants in the regressions are statistically insignificant, therefore, CAPM holds for all stocks. The excess return on market proxy index $[R_m - R_f]$, is found to be statistically significant in explaining excess return on each stock $[R_i - R_f]$. The dummies of the negative and positive news are mostly significant in explaining the change in excess return on each stock (except the ones of which t ratios are shown in red color). News are affecting the returns on the same day they are reported without any day lag and the negative news decreases the stock return while the positive news increases the stock return for the dummies of which coefficients are statistically significant.

CAPM:

$$[R_i - R_f] = c + \beta_{i,m} [R_m - R_f] + \beta_{i,n} D_n + \beta_{i,p} D_p + \varepsilon$$

CAPM regressions for each stock return are estimated with daily data but time periods differ for some of the stocks. In general, regressions are estimated for the time period of 8/01/2006-2/28/2018, but some of them are estimated with different time periods which are specified below Table 10. Moreover, insignificant coefficients of dummies are shown in red color in the table.

Table 14- Regression Results for CAPM

AKBNK	(Ri-Rf)=	0.0000	+ 0.705 (Rm-Rf)	+ 0.001 Dp	- 0.004 Dn	R^2=0.409
	stdev	(0.0003)	(0.0166)	(0.0008)	(0.0009)	
	t ratio	[0.003]	[42.427]	[1.598]	[-4.678]	
ARCLK	(Ri-Rf)=	0.0000	+ 0.511 (Rm-Rf)	+ 0.006 Dp	- 0.005 Dn	R^2=0.266
	stdev	(0.0003)	(0.0166)	(0.0011)	(0.0012)	
	t ratio	[0.135]	[30.728]	[5.183]	[-4.375]	
ASELS	(Ri-Rf)=	0.0006	+ 0.491 (Rm-Rf)	+ 0.007 Dp	- 0.008 Dn	R^2 =0.056
	stdev	(0.0007)	(0.0408)	(0.0025)	(0.0033)	
	t ratio	[0.824]	[12.054]	[2.889]	[-2.313]	
BJKAS *	(Ri-Rf)=	0.0013	+ 0.597(Rm-Rf)	+ 0.000 Dp	- 0.014 Dn	R^2=0.096
	stdev	(0.0008)	(0.0608)	(0.0031)	(0.0037)	
	t ratio	[1.581]	[9.812]	[0.49]	[-3.908]	
DENIZ**	(Ri-Rf)=	0.0012	+ 0.372 (Rm-Rf)	+ 0.01 Dp	- 0.003 Dn	R^2=0.073
	stdev	(0.0007)	(0.0433)	(0.0022)	(0.0026)	
	t ratio	[1.753]	[8.587]	[4.491]	[-1.244]	
FENER	(Ri-Rf)=	0.0004	+ 0.264 (Rm-Rf)	+ 0.006 Dp	- 0.008 Dn	R^2=0.059
	stdev	(0.0004)	(0.0226)	(0.0016)	(0.0018)	
	t ratio	[0.942]	[11.645]	[4.105]	[-4.59]	
FROTO	(Ri-Rf)=	0.0005	+ 0.45 (Rm-Rf)	+ 0.005 Dp	- 0.005 Dn	R^2=0.184
	stdev	(0.0003)	(0.0188)	(0.0011)	(0.0012)	
	t ratio	[1.405]	[23.922]	[4.673]	[-4.083]	
GARAN ***	(Ri-Rf)=	0.0005	+ 1.121 (Rm-Rf)	+ 0.001 Dp	- 0.001 Dn	R^2=0.679
	stdev	(0.0004)	(0.0281)	(0.0012)	(0.0012)	
	t ratio	[1.375]	[39.835]	[0.558]	[-0.531]	
KCHOL	(Ri-Rf)=	-0.0002	+ 0.59 (Rm-Rf)	+ 0.002 Dp	- 0.003 Dn	R^2=0.343
	stdev	(0.0003)	(0.0157)	(0.0008)	(0.0009)	
	t ratio	[-0.642]	[37.534]	[2.213]	[-2.984]	
SKBNK**	(Ri-Rf)=	0.0008	+ 0.546(Rm-Rf)	+ 0.005 Dp	- 0.005 Dn	R^2=0.204
	stdev	(0.0005)	(0.0324)	(0.0018)	(0.0023)	
	t ratio	[1.74]	[16.837]	[2.841]	[-2.203]	
TCELL	(Ri-Rf)=	-0.0001	+ 0.445 (Rm-Rf)	+ 0.003 Dp	- 0.005 Dn	R^2=0.244
	stdev	(0.0003)	(0.0163)	(0.0007)	(0.0008)	
	t ratio	[-0.201]	[27.244]	[4.235]	[-6.459]	
THYAO	(Ri-Rf)=	0.0002	+ 0.508 (Rm-Rf)	+ 0.006 Dp	- 0.004 Dn	R^2=0.074
	stdev	(0.0008)	(0.0387)	(0.0015)	(0.0017)	
	t ratio	[0.26]	[13.143]	[3.656]	[-2.412]	
TOASO	(Ri-Rf)=	0.0005	+ 0.54 (Rm-Rf)	+ 0.007 Dp	- 0.006 Dn	R^2=0.244
	stdev	(0.0003)	(0.0191)	(0.0011)	(0.0013)	
	t ratio	[1.411]	[28.296]	[6.409]	[-4.837]	
ULKER **	(Ri-Rf)=	0.0008	+ 0.44 (Rm-Rf)	+ 0.009 Dp	- 0.006 Dn	R^2=0.142
	stdev	(0.0005)	(0.0311)	(0.0017)	(0.0019)	
	t ratio	[1.793]	[14.146]	[5.711]	[-3.327]	
VESTL ***	(Ri-Rf)=	0.0016	+ 1.083 (Rm-Rf)	+ 0.010 Dp	- 0.009 Dn	R^2=0.286
	stdev	(0.0008)	(0.0641)	(0.004)	(0.004)	
	t ratio	[1.947]	[16.907]	[2.394]	[-2.254]	

Note: * is estimated for the time period of 01/01/2014-2/28/2018, ** is estimated for the time period of 01/02/2013-2/28/2018, *** is estimated for the time period of 01/02/2015-2/28/2018, **** is estimated for the time period of 01/02/2012-2/28/2018. Insignificant coefficients of dummies are shown in red color.

4.3.3. Single Factor Market Model (Panel Data)

In this section, cross section data for the 15 stocks is used for the analysis. Total number of observation is 43,725. We add cross-section (fixed) dummy to the regression. The regression results for MM (Panel Data) below clearly and consistently show that the expected returns of all stocks are significantly affected by the expected market return. The coefficients of dummies of negative and positive news are statistically significant and theoretically consistent in explaining the changes in expected return of stocks.

Single Factor Model with Positive and Negative Sentiment Dummies:

$$R_i = \beta_0 + \beta_{i,m} R_m + \beta_{i,n} D_n + \beta_{i,p} D_p + \varepsilon$$

Table 15- Single Factor Model with Positive and Negative Sentiment Dummies

R_i=	0.0008	+0.4996 R_m	-0.0055 $D_{i,n}$	+0.0047$D_{i,p}$	R^2 = 0.1504
stdev	(0.0001)	(0.0061)	(0.0003)	(0.0003)	
t ratio	[7.630]	[81.140]	[-14.832]	[14.089]	

4.4. TRADE ON CLASSIFIED NEWS DATA

In this section, the classified news data is used for trade action on different scenarios. All the news stories (688 stories) for the time period of 1/1/ 2017 – 28/02/2018 are used for trade with different scenarios based on estimation of the learners generated in Section 3.6[33]. All the actions (short or long) are taken based on these three machine learning classifiers: Multinomial Naïve Bayes, SVM (Linear kernel) and DT. That is, if a learner classifies the news as positive (negative), then the action is taking a long (short) position.

[33] The training set does not consist of any news published in this time period.

Also, it is supposed that all the positions close on the day t, with the close price of the stock. It is also assumed that the news is published at the beginning of market opening and the transaction cost is zero. So, Strategy 1 states that trade on the opening session and close the position same day with close price. Strategy 2 and Strategy 3 assumes that the news published on day t has already known on day t-1 and t-2, respectively.

Table 16- Cumulative Payoffs for Different Classifiers on Different Scenarios

Strategy	Time/Action	Multinomial NB	SVM	DT
1	**Trade on t (open price)**	39%	42%	**67%**
2	**Trade on t-1 (close price)**	245%	**246%**	240%
3	**Trade on t-2 (close price)**	427%	**449%**	410%

Table 17- Average Returns for Different Classifiers on Different Scenarios

Strategy	Time/Action	Multinomial NB	SVM	DT
1	**Trade on t (open price)**	0,05%	0,06%	**0,09%**
2	**Trade on t-1 (close price)**	0,35%	**0,36%**	0,34%
3	**Trade on t-2 (close price)**	0,62%	**0,65%**	0,59%

It is clear that all strategies generate significantly positive returns on each scenario. However, if the trade cost is taken into account, average returns may not cover the trade cost for the trade on the day t. This result verifies a form of semi-strong efficiency in the market: one cannot make superior profits based on public information.

Strategy-2 and Strategy-3 give significantly positive returns but they are infeasible since the positions open before the news is announced. These strategies may feasible under one condition: this is called insider information. As Table 18 and Table 19 clearly show that a trader with insider information can realize profits based on these machine learning algorithms up to 449%.

In conclusion, the news stories (either negative or positive) are statistically significant and theoretically consistent in explaining the performance of individual stocks. However, the trade strategies based on machine learning algorithms do not generate significant profit unless there is insider information. This finding confirms *the semi-strong form of EMH* holds in BIST.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1. CONCLUSION

This thesis examines the relationship between the price data of companies in different sectors in Borsa Istanbul (BIST) and the influence of how financial news on these companies is expressed through wording. In this work, sentiment analysis is formed by automatically labelling the news for companies publicly traded in BIST as positive or negative on the basis of the daily performance of stocks with different methods in machine learning.

As result of our novel approach, we achieved to get valuable information from unlabeled data. We made a dictionary with effective words for financial news. Each word has also label as negative or positive. One valuable thing is that the entire process is managed automatically. Thus, the automatic labelling approach that we built in this work gives some incentives about generating a large corpus easily. The other important thing added to this process is the collection of market sense as a determiner of the polarity. Hence, it has been gathered more reliable judgement. We have found that our classifiers can learn and detect sentiment in the news stories with moderate success. As we compared several learners, we got better results from Naive Bayes classifiers about 70% of accuracy and 80% score for F-Score.

As for market concerns, it is clear that both negative and positive news has an effect on its related stock prices. On the other hand, it is hard to gain profit from this information unless there is insider information. It is very likely, though, that labelling news stories would be an important input as far as it is implemented in a very short period of time after the stories published.

5.2. FUTURE WORK

First of all, we have room for optimism that more sophisticated feature set might improve the accuracy scores of the model. As Boudoukh et al. (2013) proposes, enhancing algorithms so that it can detect relevant and irrelevant news may improve these scores. Likewise, adding subjects of the news as a new class to the algorithm could be the other point for better results.

We built our sentiment analysis based on firm-specific news stories. Instead of firm-specific sentiment analysis, industry-specific level of news may give more appropriate results. Categorizing firms and their related news may affect score positively as well. With the help of this classification, some words may both be negative or positive markers based on its category. Also, we calculate rate of return by the closing prices of stocks. With this method, we accumulate all the news published in each day as one body and one label for all. Thus, instead of taking daily performance of the stocks, using tick-by-tick data may reflect the sign of each news stories, which in turn, would help to get better accuracy rate.

One of the possible application of the learner built in this thesis is to track and detect opinions in online discussions. Advisors, brokers and traders could benefit from this by shaping their financial opinions about the current event; participants could gather general idea about trends and so on.

The other feasible application for sentiment analysis is to establish a news sentiment index. Sentiment scores calculated across newspapers, magazines and other media, could be used to construct a monthly sentiment index and to assess how it reflects economic outcomes (Shapiro et al. (2018)). Also, it can be questioned whether or not the sentiment index has similar pattern with consumer confidence index.

REFERENCES

Akbas, F., Boehmer, E., Erturk, B., & Sorescu, S. M. (2013). Short interest, returns, and fundamentals. Returns, and Fundamentals (August 31, 2013). http://doi.org/10.2139/ssrn.2216919

Akın, A. A., & Akın, M. D. (2007). Zemberek, An Open Source Nlp Framework for Turkic Languages. *Structure*. https://doi.org/10.1.1.556.69

Argamon-Engelson, S., Koppel, M., & Avneri, G. (1998). Style-based text categorization: What newspaper am I reading? *Proceedings of AAAI Workshop on Learning for Text Categorization,* 1–4. Retrieved from https://www.aaai.org/Papers/Workshops/1998/WS-98-05/WS98-05-001.pdf

Azar, P. D. (2009). *Sentiment Analysis in Financial News*. Harvard College.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*. https://doi.org/10.1037

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: "Analyzing text with the natural language toolkit"--Cover*. Beijing: O'Reilly.

Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., & Weber, I. (2012). Web search queries can predict stock market volumes. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0040014

Boudoukh, J., Feldman, R., Kogan, S., & Richardson, M. (2013). Which News Moves Stock Prices? A Textual Analysis. http://doi.org/10.3386/w18725

Boynukalin, Z. (2012). *Emotion analysis of turkish texts by using machine learning methods*. Middle East Technical University. Retrieved from http://etd.lib.metu.edu.tr/upload/12614521/index.pdf

Brigham, E. F. , & Ehrhardt, M. C. (2011). *Financial Management: Theory and Practice*.

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed)*. *Ecological Modelling* (Vol. 172). http://doi.org/10.1016/j.ecolmodel.2003.11.004

Catal, C., & Nangir, M. (2017). A sentiment classification model based on multiple classifiers. *Applied Soft Computing Journal*. https://doi.org/10.1016/j.asoc.2016.11.022

Chang, C., & Lin, C. (2013). LIBSVM : A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*. http://doi.org/10.1145/1961189.1961199

Chen, M. Y., & Chen, T. H. (2017). Modeling public mood and emotion: Blog and news sentiment and socio-economic phenomena. *Future Generation Computer Systems*. http://doi.org/10.1016/j.future.2017.10.028

Çoban, Ö., Özyer, B., & Özyer, G. T. (2015). Türkçe Twitter Mesajlarının Duygu Analizi (Sentiment Analysis for Turkish Twitter Feeds). *23rd Signal Processing and Communications Applications Conference, (SIU)*, 2388–2391. http://doi.org/10.1109/SIU.2015.7130362

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. Management Science. https://doi.org/10.1287/mnsc.1070.0704

Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery. *Proceedings of the twelfth international conference on World Wide Web - WWW '03*. doi:10.1145/775152.775226

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*. https://doi.org/10.1111/j.1911-3846.2011.01130.x

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*. https://doi.org/10.1007/s11142-014-9309-4

Dehkharghani, R., Saygin, Y., Yanikoglu, B., & Oflazer, K. (2016). SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation*. https://doi.org/10.1007/s10579-015-9307-6

Eliacik, A. B., & Erdogan, N. (2018). Influential user weighted sentiment analysis on topic based microblogging community. *Expert Systems with Applications*, *92*, 403–418. https://doi.org/10.1016/j.eswa.2017.10.006

Engelberg, J. (2008). Costly Information Processing: Evidence from Earnings Announcements. *Working Papers*. https://doi.org/10.2139/ssrn.1107998

Eroğul, U. (2009). *Sentiment Analysis in Turkish*. *Middle East Technical University, Ms Thesis, Computer Engineering*.

Eyuboglu, K. (2015). The effects of financial news on stock performance: case of BIST-30 companies. *Pressacademia*, *2*(1), 75-75. doi:10.17261/pressacademia.201519780

Fama, E. F., & French, K. R. (1992). the Cross-Section of Expected Stock Returns. *Journal of Finance*. https://doi.org/10.2307/2329112

Fama, E. F., & French, K. R. (2004). The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives*. http://doi.org/10.1257/0895330042162430

García, D. (2013). *Sentiment during recession*. *Journal of Finance*. https://doi.org/10.2139/ssrn.1571101

Généreux, M., Poibeau, T., & Koppel, M. (2008). Sentiment Analysis Using Automatically Labelled Financial News Items. *LREC 2008 Workshop on Sentiment Analysis Emotion Metaphor Ontology and Terminology*. http://doi.org/10.1007/978-94-007-1757-2_9

Hastie, T. J., Tibshirani, R. J., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction*(2nd ed.). New York, NY: Springer.

Henry, E. (2006). Market Reaction to Verbal Components of Earnings Press Releases: Event Study Using a Predictive Algorithm. *Journal of Emerging Technologies in Accounting*. https://doi.org/10.2308/jeta.2006.3.1.1

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *Journal of Business Communication*. https://doi.org/10.1177/0021943608319388

Heston, S. L., & Sinha, N. R. (2016). News versus Sentiment: Comparing Textual Processing Approaches for Predicting Stock Returns. *Finance and Economics Discussion Series*. http://doi.org/10.2139/ssrn.2311310

Huang, A., Zang, A., & Zheng, R. (2013). Large Sample Evidence on the Informativeness of Text in Analyst Reports. SSRN Electronic Journal. 10.2139/ssrn.1888724.

Kasparov, G. (1996, March 25). The day that I sensed a new kind of intelligence. *TIME*.

Kaya, M., Fidan, G., & Toroslu, I. H. (2012). Sentiment analysis of Turkish political news. In *Proceedings - 2012 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2012*. http://doi.org/10.1109/WI-IAT.2012.115

Kecman, V. (2005). Support Vector Machines – An Introduction, (May 2005), 1–47. http://doi.org/10.1007/10984697_1

Koppel, M., & Shtrimberg, I. (2006). Good News or Bad News? Let the Market Decide.*The Information Retrieval Series*, 297-301. doi:10.1007/1-4020-4102-0_22

Kothari, S. P., Li, X., & Short, J. E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *Accounting Review*. https://doi.org/10.2308/accr.2009.84.5.1639

Li, F. (2006). Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports ? *Gene*. https://doi.org/10.2139/ssrn.898181

Li, F. (2010). The Information Content of Forward-Looking Statements in Corporate Filings—A Naïve Bayesian Machine Learning Approach. *Journal of Accounting Research Journal of Accounting Research Journal of Accounting Research*, *48*(5), 1049–1102. http://doi.org/10.HH/j.1475-679X.2010.00382.x

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics,47*(1), 13-37. doi:10.2307/1924119

Liu, B., & McConnell, J. J. (2013). The role of the media in corporate governance: Do the media influence managers' capital allocation decisions? *Journal of Financial Economics*. https://doi.org/10.1016/j.jfineco.2013.06.003

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

https://doi.org/10.1080/15427560.2015.1000335

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

http://doi.org/10.1111/1475-679X.12123

MacKay, D. J. C. (1992). Bayesian Interpolation. *Neural Computation*. http://doi.org/10.1162/neco.1992.4.3.415

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments. Cowles Foundation Monograph No. 16. New York: John Wiley & Sons, Inc.*

Martindale, C. (1975). *Romantic progression: : The psychology of literary history*. Washington: Hemisphere Publ. Co.

Martindale, C. (1990). *The clockwork muse: The predictability of artistic change. Personality and Individual Differences*. https://doi.org/10.1016/0191-8869(92)90144-E

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*. https://doi.org/10.1145/219717.219748

Oflazer, K., Say, B., Hakkani-Tür, D. Z., & Tür, G. (2003). Building a Turkish Treebank. *Invited chapter in Building and Exploiting Syntactically-annotated Corpora, Anne Abeille Editor, Kluwer Academic Publishers*.

Özsert, C. M., & Özgür, A. (2013). Word polarity detection using a multilingual approach. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*(Vol. 7817 LNCS, pp. 75–82). https://doi.org/10.1007/978-3-642-37256-8_7

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. *Empirical Methods in Natural Language Processing (EMNLP)*, *10*(July), 79–86. http://doi.org/10.3115/1118693.1118704

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. https://doi.org/10.1007/s13398-014-0173-7.2

Price, S. M. K., 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 and Finance*. https://doi.org/10.1016/j.jbankfin.2011.10.013

Raschka, S. (2014). *Python Machine Learning*. *Igarss 2014*. http://doi.org/10.1007/s13398-014-0173-7.2

Rogers, J. L., Van Buskirk, A., & Zechman, S. L. C. (2011). Disclosure tone and shareholder litigation. *Accounting Review*. https://doi.org/10.2308/accr-10137

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news. *ACM Transactions on Information Systems*. http://doi.org/10.1145/1462198.1462204

Shapiro, A. H., Sudhof, M., & Wilson, D. (2018). Measuring News Sentiment. *Federal Reserve Bank of San Francisco Working Paper Series*. http://doi.org/10.24148/wp2017-01

Sharpe, W. F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *The Journal of Finance*. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Shiller, Robert J. (2000). Irational ExuberancePrinceton University Press. ISBN 1400824362

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., … Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*. http://doi.org/10.1038/nature16961

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., … Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*. http://doi.org/10.1038/nature24270

Smith, M. S., Ogilvie, D. M., Stone, P. J., Dunphy, D. C., & Hartman, J. J. (1967). The General Inquirer: A Computer Approach to Content Analysis. *American Sociological Review*. https://doi.org/10.2307/2092070

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*. http://doi.org/10.1111/j.1540-6261.2007.01232.x

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *Journal of Finance*. http://doi.org/10.1111/j.1540-6261.2008.01362.x

Tirunillai, S., & Tellis, G. J. (2012). Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance. *Marketing Science*. https://doi.org/10.1287/mksc.1110.0682

Tomayko, J. E. (2003). Behind Deep Blue: Building the Computer that Defeated the World Chess Champion (review). *Technology and Culture*. http://doi.org/10.1353/tech.2003.0140

Türkmenoğlu, C., & Tantuğ, A. (2014). Sentiment Analysis in Turkish Media. *(International Conference on Machine Learning*. http://doi.org/10.13140/2.1.1502.1125

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic Orientation applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. http://doi.org/10.3115/1073083.1073153

Wang, L. (2005). *Support Vector Machines: Theory and Applications*. Berlin Heidelberg: Springer-Verlag.

Wiebe, J., Wilson, T., & Bell, M. (2001). Identifying Collocations for Recognizing Opinions. In *Proceedings of the ACL/EACL Workshop on Collocation: Computational Extraction, Analysis, and Exploitation*. https://doi.org/10.1097/ACO.0b013e32833025d9

Zhang, H. (2004). The Optimality of Naive Bayes. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference FLAIRS 2004*. https://doi.org/10.1016/j.patrec.2005.12.001

APPENDICES

A. SAMPLES OF NEWS ITEM

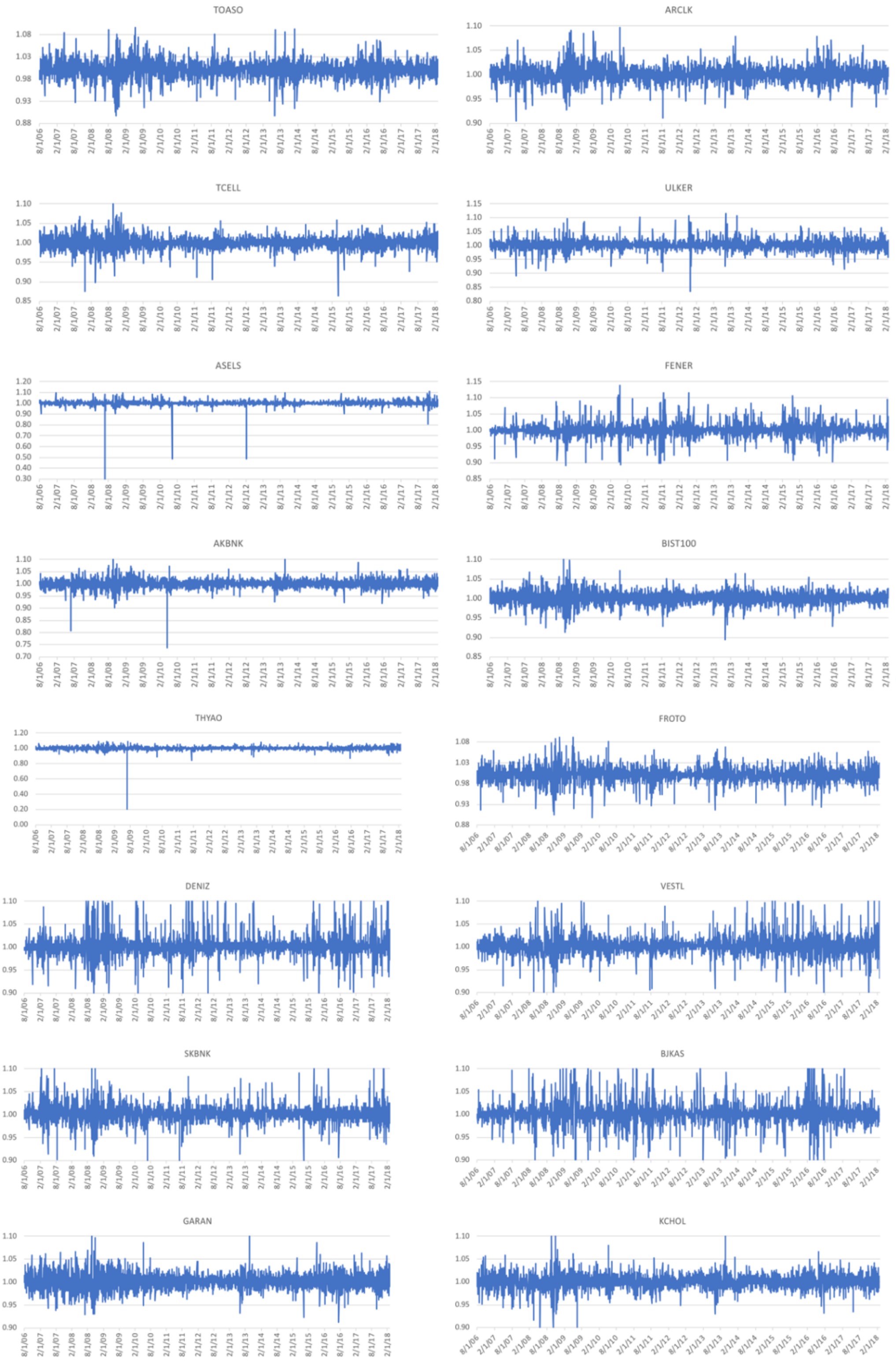
ASELS, Published on 11-Oct-01.

Aselsan füze gibi. ABD'nin Afganistan harekâtı sonrası savunma sanayi hisselerine yönelik artan ilgiyi kullanan bir spekülatör şirketin aldığı ihalenin de etkisiyle Aselsan'ı ateşledi. ABD'nin Afganistan'a yönelik başlattığı askerî harekât yüzde 83.15'i Türk Silahlı Kuvvetleri Güçlendirme Vakfı'na ait olan Aselsan'a yaradı. ABD'ye yapılan saldırıların ardından 7 Ekim'de Afganistan'da Taliban yönetimine karşı başlatılan 'Sonsuz Özgürlük' harekâtı, tüm dünyada olduğu gibi Türkiye'de de savunma sanayini ateşledi. Ancak bir spekülatörün çıkardığı kıvılcımla. Mali piyasaların kriz ve savaşın etkisiyle sarsıldığı bir dönemde, hisseleri borsada işlem gören Aselsan'da yaşanan olağanüstü çıkış hareketini, 35 yaşında broker'lıktan spekülatörlüğe hızlı geçiş yapan genç bir oyuncunun başlattığı öne sürülüyor. Genç yaşında hatırı sayılır bir servetin sahibi olan bu spekülatörün, Kanlıca'daki tripleks villasından verdiği emirlerle Aselsan hisselerindeki operasyonu başlattığı söyleniyor. Genç spekülatörün şirket hisselerine ilgi göstermesinin arkasında ise, dünyada savunma sanayiinde yaşanan hareketlilikten ziyade Aselsan'ın aldığı ihalenin etkili olduğu belirtiliyor. Aselsan, 1 Ekim'de Savunma Sanayii Müsteşarlığı'ndan 265 milyon dolarlık bir ihale aldığını açıklamıştı. Borsa açıklama istedi Türk Silahlı Kuvvetleri ile kamu ve özel kuruluşlara elektronik haberleşme ve savunma cihazları üreten Aselsan'ın hisseleri, son sekiz işlem gününde 6 bin liradan 12 bin liraya fırlayarak yüzde 100 prim yaptı. Dün seans sonunda gelen satışlarla 11 bin 750 liraya inmesine karşın, Aselsan'ın kısa sürede elde ettiği prim İMKB'nin de dikkatini çekti. Aselsan yaptığı açıklamada, olağanüstü fiyat hareketine neden olacak özel bir durumun olmadığını, ancak yurtdışı borsalarda işlem gören savunma sanayi firmalarının hisse fiyatlarının arttığı gibi Aselsan'ın fiyatının da bundan etkilendiğini belirtti. Ancak sekiz işlem gününde elde edilen yüzde 100'lük kazanç, dış borsalarda kısa sürede ulaşılacak bir prim değil. Bu arada Aselsan'da yaşanan çıkış Otokar'ı da etkiledi. Zırhlı ve arazi aracı üreticisi Otokar, son 11 işlem gününde 5100 liradan 7900 liraya fırlayarak yüzde 54 değer kazandı.

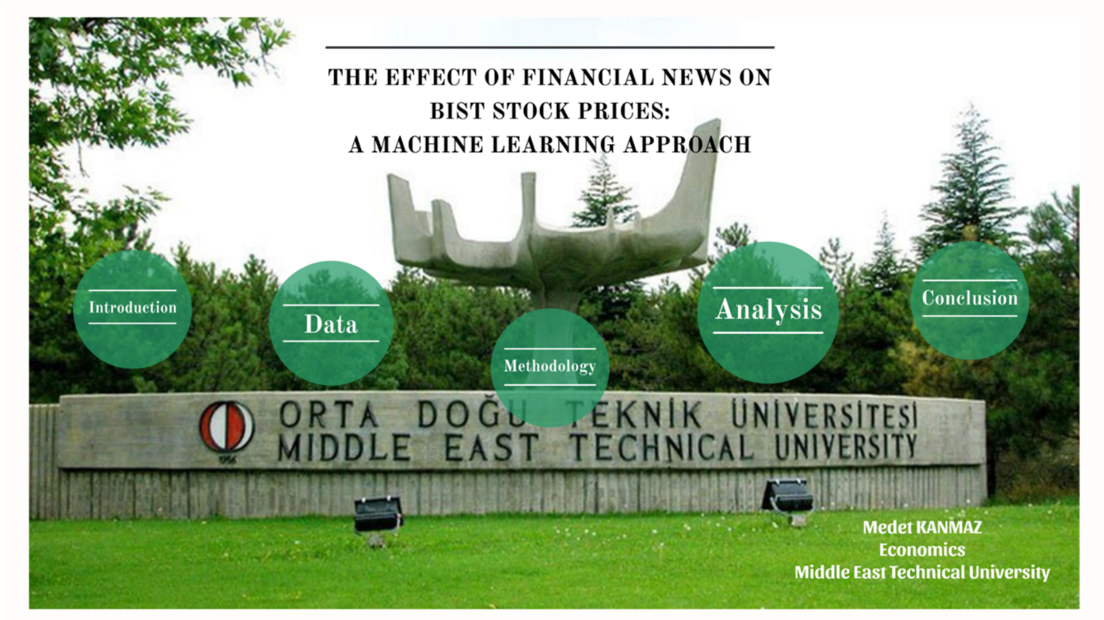
BIMEKS, Published on 5-Feb-18.

Bir süredir mali açıdan zor günler yaşayan ve birçok mağazasını kapatmak zorunda kalan Bimeks, kurtuluş için alacaklılarına iki alternatifli bir plan sundu. Buna göre ya borçlu firmalar alacakları ölçüsünde şirkete ortak olacak ya da alacaklarını bir şirkete devredip o şirket vasıtasıyla hissedar olacaklar. Bimeks, içinde bulunduğu ticari ve finansal darboğazı aşmak ve normal ticari kapasitesine tekrar dönmek ve sorunlara bir çözüm oluşturmak maksadıyla, ticari alacaklıların, finansal kuruluşların ve tahvil yatırımcılarının davet edildiği bir toplantı düzenledi. Şirketten KAP'a gönderilen açıklamaya göre, bu toplantıda faaliyetlerinin devam edebilmesine ve bu sayede finansal/ticari tarafların alacaklarını tahsil edebilmesine imkan sağlamak amacıyla iki alternatifli bir çözüm paketi sunuldu. Açıklamaya şöyle devam edildi: “1. alternatifte; ticari alacaklılar, finansal kuruluşlar ve tahvil yatırımcıları, Şirketten alacaklarını, Şirketin gerçekleştirmeyi planladığı tahsisli sermaye artışına iştirak ederek, nominal bedel üzerinden sermaye payına çevireceklerdir. Bir diğer ifade ile şirketin ticari ve finansal borçları özsermayeye dönüşecek ve alacaklı firmalar Şirket ortağı olacaklardır. 2. alternatifte; İlk alternatife teknik veya idari sebeplerle dahil olamayan alacaklılar, alacaklarını, kurulacak bir Özel Maksatlı Şirkete (ÖMŞ) devredeceklerdir. ÖMŞ, 1.alternatifte izah edilen tahsisli sermaye artışına katılacak ve Bimeks hissedarı olacaktır. Bilahare ÖMŞ, sahibi olduğu hisseleri, alacaklarını devreden alacaklılara, prorata olarak rehin verecektir.” Bu alternatifli çözüm planı ayrıca yazılı olarak ticari alacaklılarla, finansal kuruluşlarla ve tahvil yatırımcılarıyla paylaşıldığı belirtilerek, “kendilerinden hangi alternatifi tercih ettiklerini yazılı olarak bildirmeleri talep edilmiştir. Gelen cevaplardan sonra oluşacak duruma göre, Sermaye Piyasası Kuruluna müracaat ve diğer yasal prosedürler planlanacak ve bütün taraflarla birlikte şeffaf bir biçimde kamuoyu ile paylaşılacaktır.” denildi.

B. RETURN VALUES OF STOCKS AND MARKET



C. THESIS DEFENSE PRESENTATION AND PYTHON CODES



The thesis defense presentation and the python codes can be shared upon request.

D. TURKISH SUMMARY / TÜRKÇE ÖZET

FİNANSAL HABERLERİN BİST HİSSE SENEDİ FİYATLARINA ETKİSİ: MAKİNE ÖĞRENMESİ YAKLAŞIMI

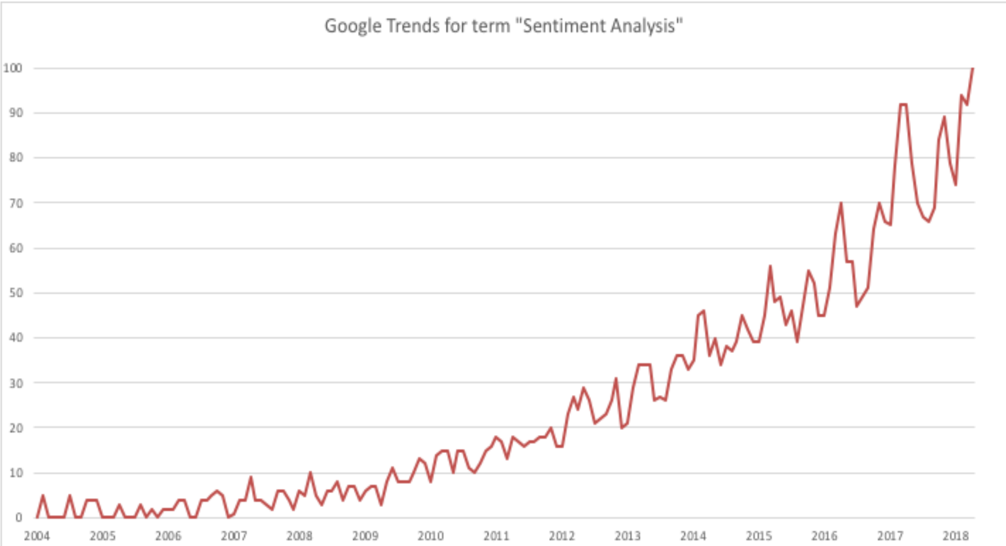
1. GİRİŞ

Günlük yaşamımızda, veriler en önemli kaynaklardan biri haline gelmiştir. Üstelik kaynakların doğasına aykırı olarak sayısı ve boyutu da her geçen gün artmaktadır. Güçlü makine öğrenimi ve algoritmalarının yardımıyla, bu veriler zekaya dönüştürülebilmektedir. Hem veri kaynaklarının bol olması hem de bilgisayar bilimindeki araçların gelişmesinin bir yansıması olarak makine öğrenmesi alanına adım atmak ve bu bilgiyi farklı alanlarda nasıl kullanabileceğini öğrenmek için muhtemelen daha iyi bir zaman olmamıştır. Bu nedenle nitel analiz, daha spesifik olarak metinsel analiz konusu, birçok araştırmacı tarafından ilgi görmüştür. E-posta kutunuzu bir düşünün. En son ne zaman spam postalarla karşılaştınız? Hepimizin hatırladığı gibi, ilk zamanlarda posta kutuları istenmeyen maillerle doluydu ve bu postaları silmek oldukça uzun zaman alırdı. Ancak günümüzde bir e-postadaki işaretleyicileri ve özellikleri tanımlamak ve spam algılama yapmak, güçlü algoritmalar (spam filtreleri) sayesinde basit bir özellik haline gelmiştir. Burada atılan önemli adım, şüphesiz metinsel analizdir.

Muhasebe ve finans alanına gelindiğinde, gazete, televizyon, sosyal medya, konferans görüşmeleri, mali tablolar, kamuya açık açıklamalar ve hatta forumlar yatırım kararlarının alınmasında temel bilgi kaynağı olmaktadır. Bu nedenle, bu metin verilerinden içerdikleri duyguları ortaya çıkaracak herhangi bir araç ilgi çekici nitelikte olacaktır.

Bir metnin içinde yer alan bilgiler duygu olarak adlandırılır. Bu çalışma, finansal haberlerden firmalara yönelik (olumlu veya olumsuz) duyguları otomatik olarak almayı amaçlamaktadır. Aşağıdaki metin verisi, NETAS firması için olumlu bir haber hikâyesi örneğidir.

> "Fatih Projesi ihalesini kazandı, hisseleri uçtu. Netaş Telekom'un bağlı ortaklığı Probil Bilgi, Fatih Projesi 2. Faz Yerel Alan Ağ Kurulum ihalesine toplam 249,94 milyon TL ile en iyi teklifi verdi. Açıklama sonrası Netaş hisseleri 9.79 liraya kadar çıktı."



Şekil D.1-Google Trendler

Şekil 1, 2004'ten 2018'e kadar olan dönemde "duygu analizi" terimi için arama trendini göstermektedir. 100 değeri, terimin en yüksek popülerliğe sahip olduğu anlamına gelirken ve 0 değeri, duygu analizi terimi için yeterli veri olmadığı anlamına gelmektedir. Şekilden de anlaşılacağı üzere 2004'ten bu yana duygu analizi için artan bir ilgi ve Mayıs 2018 itibarıyla da en yüksek popülerlik bulunmaktadır.

Duygu analizi terimi her ne kadar bu günlerde popüler olsa da Türkçe dilinde sınırlı bir çalışma olduğunu söylemek üzücüdür. Bu problemin ana nedeni, çoğunlukla İngilizce dilinde bulunan duygu sözlükleri, kelime listeleri, etiketli ve kategorilere ayrılmış veri kümelerinin, bu dil için geniş çapta ve yeterli miktarda kaynak sağlarken Türkçe dilinde bu tür kaynakların bulunmamasıdır. Bu nedenle, bu tezin amacı, Türkçe kaynaklardan verilerinin toplanması ve etiketlenmesine yönelik yeni bir yaklaşım önermek, oluşturulacak bu veri kümesi üzerinde çalışmalar yapmak ve gelecek araştırma alanlarına ışık tutmaktır.

İlgili çalışmalar / Litaratür incelemesi

Metinsel analiz ve sınıflamaya yoğunlaşan birçok çalışma vardır. Bu çalışmalar 1990'ların sonundan günümüze kadar uzanmaktadır. "Verinin kendisinden" ve "sözlük tabanlı" öğrenme, duygu analizindeki iki farklı yaklaşımdır.

Verinin kendisinden öğrenilmesi birçok farklı teknikten oluşur. Argamon-Engelson, Koppel ve Avneri (1998), sözcüklerin ve konuşma dilinde sıklıkla tekrar eden üçlü kelimelerin özelliklerini baz alarak çalışma yürütmüştür. Daha sonra New York Time News, NY Times Editorial, NY Daily news ve Newsweek'ten hikayeleri sınıflandırmak için makine öğrenme tekniklerini kullanmışlardır. Turney (2002) finansal olmayan alanlara odaklanan diğer duygu temelli sınıflandırma makalesidir. Yazar, restoran ve otomobil değerlendirmelerini pozitif (başparmak yukarı) veya negatif (başparmak aşağı) olarak sınıflandırır. Bu incelemede, metin içerisinde yer alan sıfatlar ve zarflar, incelemenin anlamsal yönelimini tahmin etmek için kullanılmıştır.

Bu yaklaşımda, veriler elle veya otomatik olarak etiketlenmelidir. Dave, Lawrence ve Pennock (2003), manuel olarak etiketlenmiş verilerle ürün incelemelerini pozitif veya negatif olarak sınıflandırmaktadır. Pang, Lee ve Vaithyanathan (2002), bir yorumun olumlu veya olumsuz olup olmadığını belirlemeyi amaçlamaktadır. Eğitim verileri olarak manuel olarak etiketlenmiş film/sinema incelemeleri kullanmışlardır. Koppel ve Shtrimberg (2006), belirli bir stoktaki getirilere oranlarına göre otomatik etiketlemeyi denemişlerdir. Haber hikayelerinde yer alan duyguları tespit etmek için Destek Vektör Makineleri (SVM), Karar Ağaçları ve Naive Bayes metotlarını kullanmışlardır. Lineer SVM modelleri %70.3 oranında bir doğruluk sağlamıştır. Benzer şekilde, Généreux, Poibeau ve Koppel (2008) "bir haberin piyasadaki tepkisinin finansal haberleri etiketlemek için iyi bir gösterge olduğunu ve fiyat haberlerini etkin bir şekilde algılayan modeller oluşturmak için bu haberlerle bir makine öğrenimi algoritmasının eğitilebileceğini" öne sürmüşlerdir. Makine öğrenimi algoritmaları aracılığıyla fiyat değişikliklerini tespit etmek için bir özellik havuzu önermektedirler. Çalışma sonucunda elde ettikleri modellerinin doğruluk oranı %69'dur.

Sözlük temelli öğrenme yöntemleri göz önüne alındığında, Henry (2006), Henry (2008), Li (2006) önemli makalelerdir. Bu yaklaşım esas olarak kelime sayılarına odaklanmaktadır. Bu yöntemde verilerin önceden etiketlemesine gerek bulunmamaktadır. Her sözlük belirli bir aralıkta farklı özelliklere (kelime listeleri) sahiptir. Belirli bir metnin negatif / pozitif sayı puanı o haber öyküsünün sonucunu (etiketini) gösterir. Pek çok araştırmacı duygu analizi için kelime veya sözlük listesini kullanır. Li (2006), örneğin, şirket yıllık raporları içeriklerinden risk algılarını tespit etmek için risk ve belirsizlik için özel hazırlanmış kelime listesini kullanarak çalışmasını yürütmüştür.

General Inquirer Dictionary (GI), Smith ve ark. (1967), birçok araştırmacı tarafından yaygın olarak kullanılmaktadır. Tetlock (2007), Engelberg (2008) bu sözlüğü içerik analizi yöntemi için kullanmıştır. Tetlock (2007), medya kötümserliğinin yüksek değerinin fiyatlar üzerindeki aşağı yönlü baskıya yol açtığını ve medya karamsarlık tahminlerinin piyasadaki oynaklıktaki artışın arttığını öne sürmüştür. Finansal literatürde GI'ye dayalı duyarlılık analizi önemli sonuçlar vermektedir. Engelberg (2008), firmaların finansal raporlarındaki hem nicel hem de nitel bilgilerin gelecekteki getiriler üzerinde bir etkisi olduğunu açıklamaktadır. Tetlock, Saar-Tsechansky ve Macskassy (2008) haber makalelerinde olumsuz kelimeleri saymak için GI sözlüğünü kullanmışlardır. Borsa fiyatlarının olumsuz bilgiyi bir günlük gecikmeyle yansıttığını; firma temel değerleri hakkındaki haber hikayelerinde yer alan olumsuz kelimelerin kazançları etkilediği ve diğer öykülere nazaran getirilerle daha fazla ilintili olduğu görülmüştür.

Sözlük tabanlı kategorileştirme, olumlu ve olumsuz kelimeleri saymaya odaklandığından, Boudoukh, Feldman, Kogan ve Richardson (2013) "daha sofistike bir metin analizi metodolojisi kullanmanın sonuçları daha da geliştireceğini" öne sürmektedir. Bu nedenle, duygu analizi için bir başka kategori daha tanıtmışlardır. "Haberi tanımlamak ve tonunu daha doğru olarak değerlendirmek" ile ilgili olarak haberler için yeni bir kategori önermektedirler ve hisse senedi fiyatı değişiklikleri ile bilgi arasında güçlü bir ilişki olduğuna dair fazla kanıt bulunmaktadır (s. 4). Bu yöntemde öncelikle haber hikayeleri şirket temelleri hakkında "ilgili" bilgiler içeriyorsa, "ilgili" oldukları kabul edilir, aksi takdirde bahsekonu haber "ilgisiz" haber olarak addedilir.

Benzer şekilde hem sözlük tabanlı yöntem hem de verinin kendisinden öğrenme metotlarının performanslarını karşılaştıran başka çalışmalar da vardır. Azar (2009), duyguları tespit etmek ve haberleri pozitif / negatif olarak iki sınıfa ayırmak için makine öğrenimi algoritmalarını ve GI sözlüğünü kullanmıştır. Yazar, SVM modelinin insanlara göre dahi, daha iyi bir performans sergilediğini göstermiştir.

Zengin morfolojisi nedeniyle, doğal dil işlemesi Türkçe için zahmetli bir iştir. Bu nedenle Stemming (bir kelimenin kök halini ortaya çıkarmak) ana handikaptır. Kaya, Fidan ve Toroslu (2012), iyi bilinen denetimli makine öğrenimi algoritmalarının (Naive Bayes (NB), Maksimum Entropi (ME), Destek Vektör Makinesi (SVM) ve karakter tabanlı N-Gram Dil Modeli) performansını karşılaştırmıştır. Türk siyasi yazılarının duygu analizi yapılmıştır. Kök aracı olarak Zemberek'i kullanılmıştır. Modelleri, %65 ile %77 arasında bir hassasiyete ulaşmıştır. Boynukalin (2012) Türkçe metnin duygu analizini sunar. İki farklı kaynaktan (ISEAR veri kümesinin ve Türk masallarının çevirisi) yeni bir veri seti oluşturmuştur. Tezinde, farklı yaklaşımlar önermiş ve doğruluk sonuçlarını karşılaştırmıştır. Modelleri %76 ile %81 arasında bir hassasiyete ulaşmıştır.

Firmalar ile ilgili birçok bilgi kaynağı vardır. Tweetler gibi sosyal medya metinleri duygu analizi için diğer kaynaklardır. Bu metin verisindeki sınırlı sayıda karakter araştırmacıları hem sözlük tabanlı hem de kendi kendine öğrenme metotlarını kullanarak duyarlılığı tespit etmeye yönelik çalışmalara odaklanmayı teşvik etmektedir. Türkmenoğlu ve Tantuğ (2014), yaptıkları çalışmada, Türk sosyal medyası için iki farklı duyarlılık analiz çerçevesini (twitter ve sinema yorumlarını) karşılaştırmıştır. Karşılaştırma için sözcükleri oluşturmak için doğal dil işleme aracını kullanmışlardır. Sinema eleştirileri veri seti ile yapılan çalışmanın doğruluğunun hem sözcük tabanlı hem de ML tabanlı duyarlılık analiz yöntemlerinde twitter veri kümesinin doğruluğundan daha iyi olduğunu bulmuşlardır.

Bu tez, Généreux ve ark. (2008) çalışma hipotezini esas alarak BİST bünyesindeki firmalara ait finansal haberlere Türkçe olarak uygular. Bu çalışmada, duygu analizinin çıkarılmasında verinin kendisinden öğrenme yöntemi tercih edilmiştir. Ayrıca, makine öğrenmesi yöntemiyle elde edilen sonuçların teorik ve istatistiksel olarak anlamlı olup olmadığı ve piyasada kâr elde etmek için teşvik sağlayıp sağlamadığı hususu da ele alınmıştır.

Ana Katkı

Yeni yaklaşımımızın bir sonucu olarak, etiketlenmemiş verilerden değerli bilgilere ulaşılmış; finansal haberler için etkili kelimelerden oluşan bir liste oluşturulmuş; her kelimeye ait etiketler negatif veya pozitif olarak elde edilmiştir. Değerli olan bir diğer husus, tüm bu sürecin otomatik olarak yönetilmesidir. Böylece, bu çalışmada yaptığımız otomatik etiketleme yaklaşımı, büyük bir kelime listesi oluşturma konusunda bazı ipuçları vermektedir. Bu sürece eklenen diğer bir önemli husus, duygu belirleyicisi olarak piyasa fiyat performansının derlenmesidir. Bunun bir sonucu olarak, kutupluluğun belirlenmesinde daha objektif bir karar mekanizması kurulması sağlanmıştır.

Tezde oluşturulan makine öğrenmesi yöntemlerinin, haber metinleri içindeki duyguları orta seviyede bir başarı ile tespit etmeyi başardığı tespiti yapılmıştır. Birkaç yöntemin karşılaştırılması neticesinde, Naive Bayes sınıflandırıcılarının görece başarılı olduğu ve bu yöntemin doğruluk oranının %70 civarında (F-Puanı için %80) bir skor olduğu görülmüştür. Daha sofistike özellik setinin bu skoru iyileştirebileceği hususunda ise beklenti mevcuttur.

Öte yandan hem olumsuz hem de olumlu haberlerin ilgili hisse senedi fiyatları üzerinde etkili olduğu istatistiksel olarak gösterilmiştir. Çeşitli piyasa modellerinde haber kukla değişkenlerinin istatistiksel olarak anlamlı olduğu gösterilmiştir. Bununla birlikte, haberlerin yayınlanmasından sonra çok kısa bir süre içinde hayata geçirilebildiği takdirde, haberlerinin etiketlenmesinin piyasa oyuncuları açısından önemli bir girdi olacağı çok muhtemeldir. Bu bağlamda, oluşturulan algoritmaların ticaret teşviki sağlayıp sağlamadığı sorgulanmış içeriden bilgi (insider information) bulunmadığı sürece kamuya açık bu bilgilerden kâr elde etmenin zor olduğu gösterilmiştir.

2. VERİ SETİ VE TANIMLARI

Bu bölümde hem duygu analizinin gerçekleştirilmesinde hem de getirilerin hesaplanmasında kullanılan veri kümeleri kısaca tanıtılmaktadır. Bunlar, finansal haberler, BIST100 endeksi, bu piyasada işlem gören/görmüş firmaların günlük fiyat performansı ve iki yıllık devlet tahvili getirilerinden oluşmaktadır.

Haber kaynağı olarak FINNET verileri esas alınmıştır. Gazeteler, aylık dergiler, internet haberleri, bültenler ve kamuyu aydınlatma platformu haberleri gibi 100'den fazla farklı kaynaktan oluşan bu veri setinden 1996-2018 yılları arasında BİST (İMKB) bünyesinde işlem gören halka açık şirketlere ilişkin yaklaşık 75000 haber elde edilmiştir. Bu veri seti ayrıca bazı önemli unsurları da içerir: haber yayım tarihi, şirketin adı, haber başlığı ve haber metni.

Bu noktada, haberler hakkında bazı varsayımlar yapılarak veri setinde bazı kısıtlamalara gidilmiştir. İlki, ağırlıklı olarak reklam içeren haberlere ilişkindir. Genel olarak, dergilerde yer alan makalelerin şirketler için reklam barındırması ihtimaline binaen bu verileri kullanmanın makine öğrenmesinde olumlu kelimelerin derlenmesinde hataya yol açabileceği değerlendirilmiştir. Ayrıca, aylık dergilerdeki haberlerin yayınlanma tarihi ve içeriğinin hisse senedinin fiyatındaki değişikliğe karşılık gelemeyeceğinden hareketle, aylık ve haftalık dergilerdeki makaleler veri setinden ayrıştırılmıştır. Haberlerin günlük bazda gözlemlenen hisse senedi getirisine etkisini incelemek için bir şirkete ait aynı gün yayınlanan haberler tek bir vücutta birleştirilmiştir. Haber veri setinde yapılan diğer bir değişiklik ise, hafta sonu ya da tatil günlerinde yayınlanan haber hikâyesinin, referans tarihinin bir sonraki iş günüyle değiştirilmesidir. Piyasanın kapanış zamanından sonra yayınlanan haber hikayeleri için de aynı yaklaşım ele alınmıştır. Bu nedenle, haber hikayelerinin yayım tarihi yerine oluşturulan referans günü kullanılmıştır.

Bundan noktadan sonra, 38808 haber hikayesi veri setinde kalmıştır. Tablo 1 bu veri kümesine genel bir bakış sunmaktadır. 347 farklı halka açık şirkete ait 38808 haber hikayesi bulunmaktadır. Bir şirket için maksimum hikaye sayısı 1687 iken, sadece bir hikayesi olan şirketler de bulunmaktadır. Şirket başına ortalama hikaye sayısı ise yaklaşık 110'dur.

Haber öyküleri için son sınırlama, haber öğelerinin uzunluğudur. Bu nedenle, sadece haber özeti 18'den fazla kelime barındıran haberleri ele alınmıştır. Bu kısıtlama, firma hakkında daha ayrıntılı haberler almak ve makine öğrenmesinde daha bilgilendirici bir özellik oluşturmak için belirlenmiştir. Tüm kısıtlamalardan sonra, 29302 haber maddesinden oluşan bir veri seti elde edilmiştir.

Bu çalışmada kullanılan diğer kaynak, stokların performans verileridir. Bu amaçla Borsa İstanbul Data Store'dan günlük hisse senedi fiyat verileri satın alınmıştır. Veriler, 1996 ve 2018 (şubat ayına kadar) mali yıllarında BIST'teki tüm hisse senetlerinin açılış ve kapanış fiyatlarından oluşmaktadır. BIST 100 endeksinin günlük performansı için, invesing.com verilerinden ayrıca bir set oluşturulmuştur.

Haber verilerinin bir etiketi mevcut değildir. Öncelikle, haberlerin etiketlemesi amacıyla her hikaye kendi hisse senediyle eşleştirilmiştir. Her bir etkinlik günü (t) için (haber öyküleri için referans günü), ilgili hisse senedinin kapanış fiyatı ve önceki kapanış fiyatı seans bazında eşleştirilmiş ve getiri oranı belirlenmiştir.

Ocak 1996-Kasım 2015 döneminde, işlemler iki işlem seansında yürütüldüğünden bu hesaplamalar her seans için ayrı ayrı yapılmıştır. Mutlak değer olarak en büyük etki, o günkü haber hikayesinin bir göstergesi olarak ele alınmaktadır. Aynı prosedür bir önceki güne de (t-1) uygulanmıştır. Otomatik etiketleme yaklaşımında (t) ve (t-1) getiri değerlerinin mutlak değer olarak en büyük olanının yönü, haber öğesinin işareti (kategorisi) ile ilgili olarak nihai karara ulaştırmaktadır.

Kasım 2015-Şubat 2018 dönemi, seans ayrımı yerine, ilgili hisse senetleri için gün içi fiyat değişimlerini (kapanış fiyatı – açılış fiyatı) kapsamaktadır. Mutlak değer olarak en büyük etki, haber öyküsünün bir göstergesi olarak dikkate alınmaktadır. Kısacası, (t) ve (t-1) günlerine ait getiri oranları, mutlak değerler açısından karşılaştırılarak ve haberlerin kutupluluğu belirlenmektedir.

3. METODOLOJİ

Sınıflandırma Yöntemleri

Makine öğrenmesinde sınıflandırıcılar, girdi verilerini gruplara, sınıflara veya kategorilere ayıran algoritmalar veya matematiksel işlevlerdir. Bu tezde, haberlerin duygularını belirlemek için üç ana sınıflandırıcı kullanıyoruz. Bu çalışmada kullanılan makine öğrenmesi algoritmaları şunlardır: Naïve Bayes Sınıflandırma, Karar Ağacı Sınıflandırma ve Destek Vektör Makine Sınıflaması.

Naïve Bayes Sınıflandırma (NB):

Naive Bayes, denetimli makine öğreniminde en etkili ve etkili algoritmalardan biridir. Temel olarak, Naïve Bayes Classification, her bir özelliğin (yani, haber öğesinin her bir sözcüğü) kendi konumundan bağımsız olarak üretildiğini varsayar.

Karar Ağacı Sınıflandırma (DT):

Karar Ağaçları makine öğrenme yöntemlerindeki yaygın sınıflandırıcılardan biridir. Basit olarak, DT özellik dizisi için etiketler gösteren bir akış şemasıdır. DT, kök düğümü ile başlar ve karar düğümleri ve yaprak düğümleri ile devam eder. Bu sırada ayrıca bir karar kütüğü de bulunabilir. Karar kütüğü, etiketi tek özellikli olarak gösteren düğümdür. Özelliğin metinde olup olmadığına karar verir. DT'lerin en büyük avantajı karar modelini görselleştirmek açısındandır. Öte yandan, özellik kümesinin yüksek sayıda öğeye sahip olması DT'lerin zaman açısından etkin olmaması ile sonuçlanır.

Destek Vektör Makinesi:

Destek Vektör Makinesi (SVM), sınıflandırılmış ve belirlenmiş etiketli bir eğitim setinden giriş-çıkış haritalama algoritmaları üreterek sınıflandırma ve aykırı saptama için kullanılan denetimli bir öğrenme yöntemidir. SVM'ler, en yakın sınıflara en uzak mesafeye sahip olan hiperdüzlemler kullanır. SVM, ayırma çizgisi için "çekirdek" terimini kullanır. Ayırma çizgisi doğrusal ise, doğrusal çekirdekli SVM olarak adlandırılır. Bu konuya özel okuyucular Kecman (2005) bakabilir.

Özellik Kümesi

Metinsel analiz için özellik kümesi, başlangıçta, haber metinlerinden elde edilen kelime seti içinde en az 50 kez görünen tüm kelimelerden oluşturulmuştur. Bu işlem 2100 farklı özellik ögesi ile sonuçlanmıştır. Metnin vektörel gösterimi için bu özellik seti tüm öğrenme modellerinde kullanılmıştır. Aşağıdaki tabloda, kelime setinde en sık görülen ilk 50 sözcük gösterilmiştir. Özellik kümesindeki ögeler, naive bayes yöntemi ile elde edilen sonuçlara göre özelleştirilip sayısı düşürülecektir.

Tablo D.1-Özellik Seti

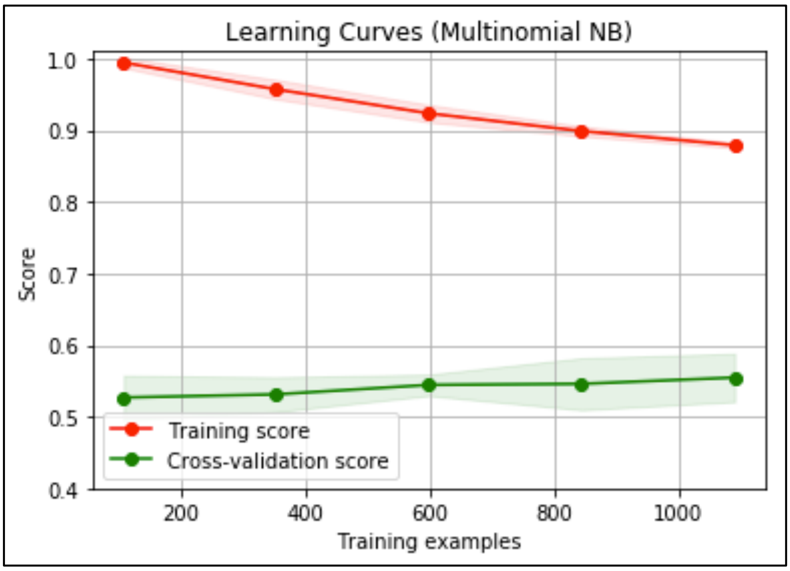
Kelime	Sıklık	Kelime	Sıklık
milyon	14292	büyük	2464
yüzde	9483	aldı	2241
bir	6718	trilyon	2206
için	6481	şirket	2117
milyar	6305	grubu	2096
yeni	5137	yönetim	2070
ilk	4849	geçen	2027
yıl	4719	ciro	1843
genel	4475	daha	1812
bin	4403	satış	1757
en	4149	enerji	1702
türk	4059	yapı	1609
kredi	4020	hava	1571
holding	3636	üretim	1516
dolar	3548	hedefliyor	1499
yatırım	3426	etti	1495
türkiye	3366	ytl	1463
net	3094	anadolu	1458
lira	2975	kadar	1440
müdür	2958	satın	1418
olarak	2905	halka	1406
tl	2902	şirketi	1388
iş	2654	son	1386
kurulu	2521	toplam	1385
dolarlık	2511	aylık	1373

Veri Eğitimi ve Test Sonuçları

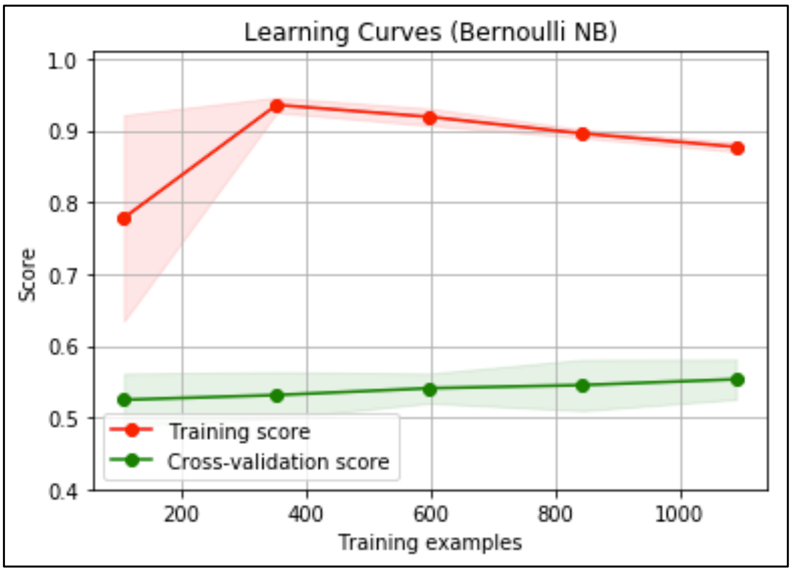
Bölüm 2'de açıklandığı gibi, haber veri kaynağındaki ilk kısıtlamalar, 1996-2018 yılları için 29302 farklı haber hikayesi ile sonuçlanmaktadır. Bu bölümde ilk olarak, haber verilerinin ilgili stokun performansına göre pozitif veya negatif olarak otomatik etiketlenmesi yapılmıştır. Etiketlemede objektif olmak için hisse senedi performanslarını bir ölçüt olarak alınmıştır. Ayrıca, hikayelere karşı başka kısıtlar getirilmiştir. Şöyle ki, bir şirkete ait bir haber, söz konusu hisse senedinin en az %9,3'lük bir artış göstermesi ve borsanın genel performansından daha iyi bir performans göstermesi durumunda pozitif olarak etiketlenir. Benzer şekilde, bir hisse senedi piyasadan daha kötü bir performansa sahipse ve fiyatı %7,6 veya daha fazla düşerse, öykü, negatif olarak etiketlenir. Bu yüksek eşikler, piyasadaki fiyat değişimlerinin, piyasa dinamikleri veya rastgele dalgalanmaların sonuçları değil, haberlerin bir sonucu olarak değişmesi olasılığını güçlendirmek için seçilmiştir. Olumlu olanlara kıyasla yakın sayıda öykü almak için olumsuz haberler için daha düşük eşik kullanılmıştır. Ek olarak, haber hikayeleri arasından sadece kelime sayısının 18 ve daha üstü olan hikayeler ele alınmıştır. Özetle, tüm bu sınırlamalar, 1.436 haber öyküsü ve %47 oranında negatif etiketli haberlerle sonuçlanmıştır.

Bu etiketli verilerden "Eğitim" ve "Test" ler için iki ayrı set düzenlenmiştir. Eğitim seti, rastgele seçilen 1.436 haber öyküsünün %80'inden oluşmaktadır. Öğrenme modelimizi bu eğitim veri seti ile eğitiyoruz. Ayrıca, haber yayım tarihlerine ilişkin olarak 1436 öyküyü başkaca iki bölüme ayırılmıştır. Bu kez, "eğitim" verilerini 1996 ve 2015 yılları arasında yayınlanan haber öykülerine göre, "test" verileri ise bu tarihten sonra yayımlanan hikayelere göre oluşturulmuştur.

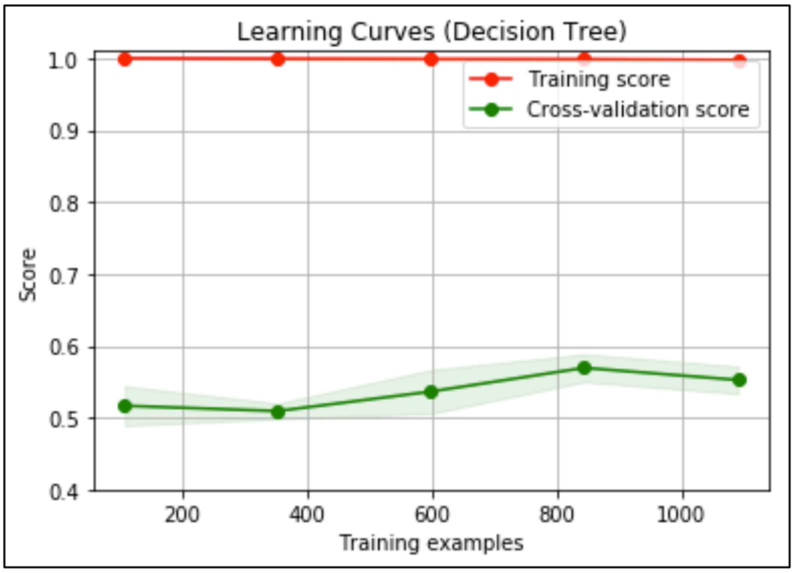
Şekil 2-5'deki grafiklerden görülebileceği gibi, tüm modeller benzer doğruluk oranlarına sahiptir. Bu sonuç ne kadar iyi ya da kötü olduğunu bu noktada önemli bir sorudur. Etiketlenmiş haber verileri popülasyonunda %53'lük bir oranda pozitif olması nedeniyle, her zaman pozitif sonuç veren bir algoritma %53 doğruluk oranıyla sonuçlanacağı açıktır. Bu nedenle, herhangi bir sınıflandırıcı bu kritik noktadan daha fazla doğruluk sağladığı takdirde, en basit anlamda iyi bir sınıflandırıcı olarak kabul edilir. Bu noktadan bakıldığında, tüm modellerin bu eşiği aştığı açıktır.



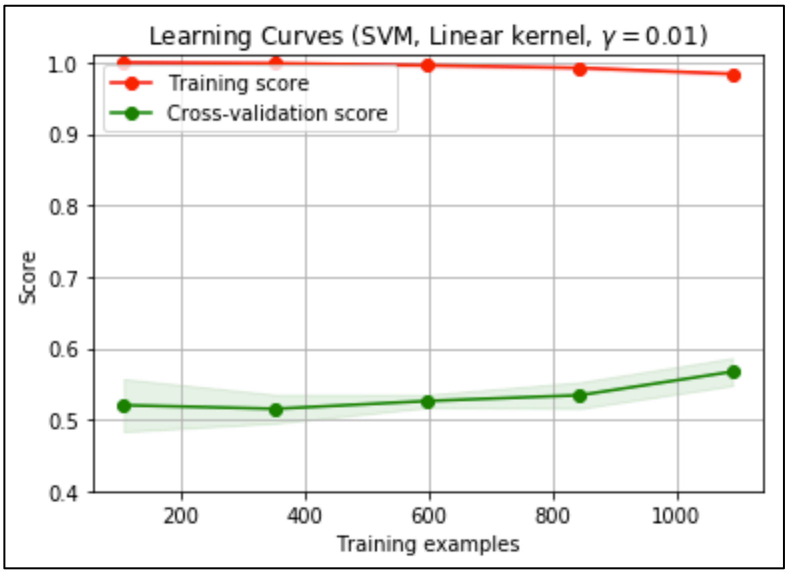
Şekil D.2- Multinomial Naïve Bayes Öğrenme Eğrisi



Şekil D.3- Bernoulli Naive Bayes Öğrenme Eğrisi



Şekil D.4- Karar Ağacı Öğrenme Eğrisi



Şekil D.5- Destek Vektör Makinesi Öğrenme Eğrisi

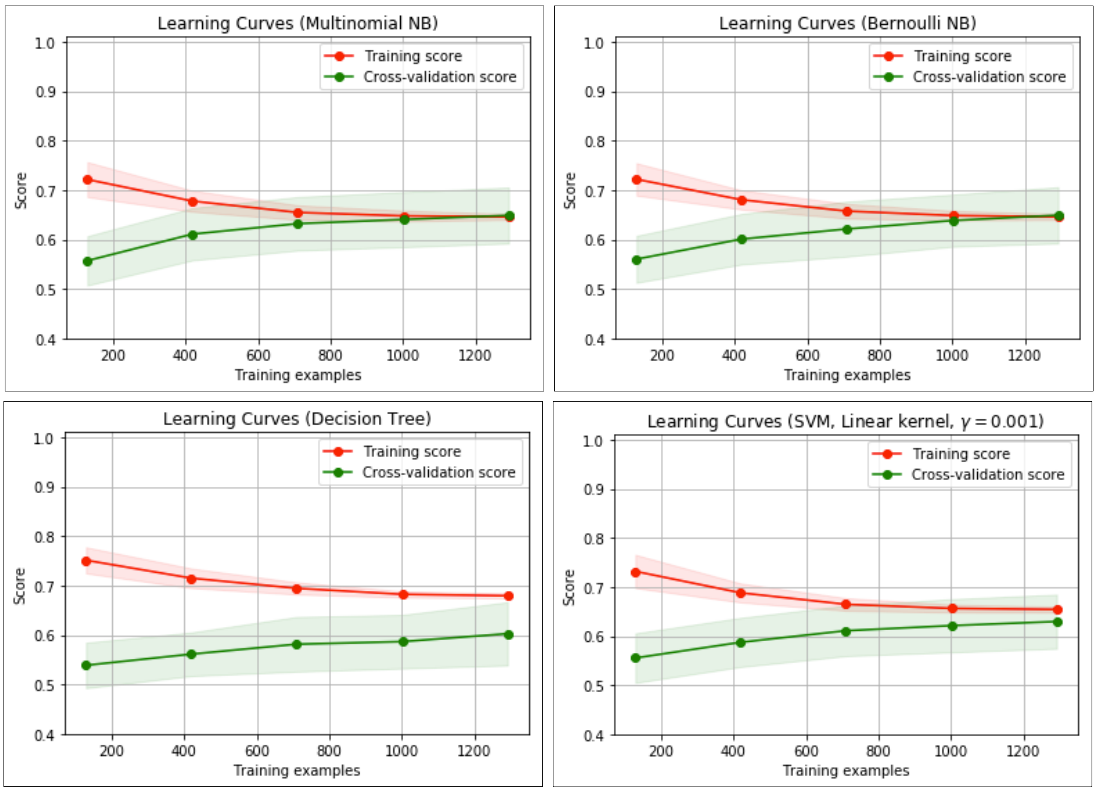
Şekil D.2-5'de çapraz doğrulama puanının yeni veriler ortaya çıktıkça dalgalanma gösterdiği açıktır. Bunun nedeni, modellerin eğitim setindeki verilere aşırı uyması (overfit), fakat iyi bir şekilde genelleştirilmemesidir. Eğitim setinde yüksek doğruluk oranı olmasına rağmen, model yeni verilerle iyi çalışmayabilir. Modeller, az miktarda eğitim verisi için karmaşık yapıdaki (yüksek sayıda) özellik ögesine bağlı olarak yüksek bir varyansa sahip gibi görünmektedir. İki eğri arasındaki boşluk bu sorunu göstermektedir. Daha fazla eğitim verisi veya daha basit bir model kullanmak bu sorunun basit bir çözümüdür. Bu nedenle, önce rastgele seçim yöntemindeki eğitim seti oranını %80'den %90'a yükseltilmiş, hem de özellik kümesindeki öge sayısı 2100'den 100'e indirmek suretiyle modellerin skorlarını yeniden test edilmesi kararlaştırılmıştır. Yeni özellik seti, Naïve Bayes yönteminin en yüksek bilgi kazanımı içeren sözcüklerinden seçilmiştir. Sonuç olarak, yeni özellik seti en fazla bilgi içeren 100 kelimeyle sınırlandırılmıştır.

Tablo D.2- Yeni Özellik Seti

Kelime	İşaret	Etki
sürüyor	Negatif	8.1
sanayi	Pozitif	6.6
vurdu	Negatif	6.5
geriledi	Negatif	6.5
çıkan	Negatif	6.5
görmeye	Pozitif	6.1
teknoloji	Pozitif	6.1
düştü	Negatif	5.8
çıkardı	Pozitif	5.4
artırım	Negatif	5.0
bedelsiz	Negatif	5.0
karşı	Negatif	5.0
ağırlıklı	Negatif	5.0
az	Negatif	5.0
ihale	Pozitif	4.9

Tablo D.2, Naive Bayes metoduna göre başlangıç eğitim veri setimizde yer alan en bilgilendirici özellikleri göstermektedir. Örneğin, "sürüyor" özelliği, özellik setinde maksimum etkiye sahiptir. Bu kelime, olumsuz haber maddelerinde olumlu olanlardan 8 kat daha fazla görülmektedir. Benzer şekilde, “sanayi” olumsuz haberlere kıyasla olumlu haberlerde 6.6 kat daha fazla kullanılan bir özelliktir.

Modellerin karmaşıklığına yönelik değişiklikler yapıldıktan sonra, aşırı benzeme probleminin (overfitting) çözüldüğü ve doğruluk puanlarının önemli ölçüde arttığı görülmüştür. Aşağıdaki şekilde görüleceği üzere, eğitim ve çapraz doğrulama puan çizgileri birbirine yaklaştığından, modeller ilk durumla karşılaştırıldığında düşük varyans ortaya koymaktadır. Aşağıdaki grafikler, her modelin yeni öğrenme eğrilerini göstermektedir. Multinomial Naive Bayes sınıflandırıcısının göreceli olarak daha iyi skorlar sağladığı ve bazı doğrulama gruplarında (açık yeşil alan) doğruluk oranının% 70'ine ulaştığı açıktır.



Şekil D.6- Yeni Öğrenme Eğrileri

Genel olarak, yeni veriler eğitim setine eklendikçe, bir modelin daha iyi sonuçlar vermesi beklenir. Bu nedenle veri setine koyulan kısıtlamalar değiştirilerek çapraz performanslar yeniden ölçülebilir. Yani, “kelime sayısı” kısıtlamasının (halihazırda 18) değiştirilmesi veya etiketleme kısıtlamalarının farklılaştırılması (halihazırda -%7.6; %9.3), negatif haber popülasyonunun aynı seviyede (% 47) kalmasını sağlandığı taktirde farklı doğruluk sonuçlarının elde edilebileceği açıktır.

Aşağıdaki özet tabloda, Multinomial NB ve SVM sınıflandırıcılarının doğruluk puanının belirli kısıtlamalara göre nasıl değiştiğini gösterilmiştir. Bu analizde, test seti 2016-2018 yılları arasındaki haberler sabit tutulmuştur. Tablodan görüleceği üzere, Multinomial NB modelinin, haber kelimelerinin 18 kelime ve daha fazla olması ve ilgili hisse senedinin performansına ilişkin sınırların -%9,10 ve %10,30 olması halinde % 71.43'lük bir doğruluk oranı sağladığı görülmektedir. Karar Ağacı ve Bernoulli NB dahil olmak üzere diğer modeller, Multinomial NB sınıflandırıcıya kıyasla nispeten daha düşük puanlar vermektedir.

Tablo D.3- Farklı Kısıtlar Altında Modellerin Doğruluk Oranları

Negatif Popülasyon	Haber Sayısı	Eğitim Seti	Test Seti	Kelime Sayısı	Negatif Sınır	Pozitif Sınır	Multinomial NB	SVM (Lineer)
48.18%	909	818	91	18	-9.10%	10.30%	**71.43%**	53.57%
47.61%	899	809	90	18	-9.20%	10.30%	70.91%	67.27%
48.58%	883	794	89	19	-9.10%	10.30%	70.91%	56.36%
48.19%	855	769	86	20	-9.10%	10.30%	70.37%	**68.52%**
47.89%	831	747	84	21	-9.10%	10.30%	70.37%	68.52%
47.02%	889	800	89	18	-9.30%	10.30%	70.37%	66.67%
48.38%	833	749	84	18	-9.40%	10.40%	70.37%	66.67%
48.00%	873	785	88	19	-9.20%	10.30%	70.37%	55.56%
47.43%	818	736	82	18	-9.50%	10.40%	69.81%	67.92%
48.02%	756	680	76	18	-9.70%	10.50%	69.81%	67.92%

Sınıflandırıcıların, haber öykülerindeki duyguları ılımlı bir başarı ile tespit etmeyi öğrenebildiklerini görülmüştür. Birkaç öğrenme yöntemine kıyasla, Naïve Bayes sınıflandırıcılarından daha iyi sonuçlar elde edilmiştir: doğruluk oranında yaklaşık %70 ve F-Puanı için %80. Doğruluk puanları çeşitli yollarla geliştirilebilir. En basit olanı haber veri setini arttırmaktır. Bu husus, finans alanındaki diğer dış kaynaklardan veri alarak sağlanabilir. Ayrıca, finansal metin verileri sosyal medya ve forumlar aracılığıyla da elde edilebilir. İkinci yol, etiketleme işlemiyle ilgilidir. Bu çalışmada kurduğumuz otomatik etiketleme yaklaşımı, pazar algısını polaritenin belirleyicisi olarak ele almaktadır. Bu yönteme ek olarak insan yorumcular yardımıyla daha tutarlı etiketler toplanabilir.

4. FİNANSAL HABERLERİN BIST STOK FİYATLARINA ETKİSİ

Bu bölümde, ilgili literatür incelenerek ve etkin piyasa hipotezi (EMH) ve formlarına değinilmiştir. Ardından araştırmada kullanılan veri yapısı ve metodoloji açıklanmıştır. Bir önceki bölümde kurduğumuz makine öğrenimi algoritmalarının bir sonucu olarak, haber materyalini matematiksel formata dönüştürülmüş ve polaritesi hesaplanıp bu haber hikayeleri piyasa analizine bir girdi olarak uygulanmıştır. Hisse senedi getiri verilerinin haberlerden nasıl etkilendiğini inceleyen farklı piyasa modelleri kurularak olumsuz ve olumlu haberlerin kukla değişkenlerinin istatistiksel olarak anlamlı olup olmadığı ve bireysel stokların beklenen getirilerini açıklarken teorik olarak tutarlı olup olmadığı tartışılmaktadır. Son bölümde, haberleri olumlu veya olumsuz olarak kategorize eden algoritmanın bu piyasada ticaret teşviki sağlayıp sağlamadığı sorgulanmaktadır.

Ekonomik veriler Bölüm 2'de kısmen açıklanmıştır. Bu bölümde, “Haber Verileri” adı altında yeni bir kukla (ikili) değişken oluşturulmuştur. Yani, bir hisse senedine ait belirli bir günde bir haber bulunuyorsa, Haber Verisi 1’dir, aksi halde 0’dır. Haber verisinden başka bir değişken daha türetilmiştir. Bu kez, hisse senedi performansına bağlı olarak değil, Bölüm 3'te oluşturduğumuz makine öğrenmesi yöntemine dayalı olarak “negatif” ve “pozitif” olarak başka bir değişken kullanıyoruz. Negatif / Olumlu Haber değişkeni, Multinomial Naïve Bayes öğrenen tahminlerine bağlı olarak ikili değişkenlerden (0 veya 1) oluşmaktadır.

1/8/2006-2/28/2018 dönemi için en çok haberi bulunan firmalardan portföy oluşturulmuştur. Portföyü seçerken şirketlere bazı kısıtlamalar uygulanmıştır. Bunlardan ilki, firmaların kuruluş tarihi ile ilgilidir. 2006’dan önce kurulan firmalar seçilmiştir. İkinci kısıtlama ise firmaların sahip olduğu haber sayısı ile ilgilidir. Bu sefer, 300 adet haber ve üstü olan firmalar seçilmiştir. Bu kısıtlamalar sonucunda kalan 15 şirket aşağıda gösterilmiştir.

Tablo D.4-Portföy firmaları

Hisse Senedi	Şirket	Haber Sayısı
AKBNK	Akbank T. A.Ş.	682
ARCLK	Arçelik A.Ş.	356
ASELS	Aselsan Elektronik Sanayi ve Ticaret A.Ş.	348
BJKAS	Beşiktaş Futbol Yatırımları Sanayi ve Ticaret A.Ş.	310
DENIZ	Denizbank A.Ş.	552
FENER	Fenerbahçe Futbol AŞ.	310
FROTO	Ford Otomotiv Sanayi A.Ş.	437
GARAN	T. Garanti Bankası A.Ş.	754
KCHOL	Koç Holding A.Ş.	570
SKBNK	Şekerbank T. A.Ş.	374
TCELL	Turkcell İletişim Hizmetleri A.Ş.	1025
THYAO	Türk Hava Yolları A.O.	1215
TOASO	Tofaş Türk Otomobil Fabrikası A.Ş.	428
ULKER	Ülker Bisküvi Sanayi A.Ş.	379
VESTL	Vestel Elektronik Sanayi ve Ticaret A.Ş.	320

Risksiz getiri oranı olarak kabul ettiğimiz 2 yıllık devlet tahvili getirisi, diğer bir değişken olarak modellerde kullanılmıştır.

<u>Tek Faktör Piyasa Modeli</u>

MM:

$$R_i = c + \beta_{i,m} R_m + \beta_{i,n} D_n + \beta_{i,p} D_p + \varepsilon$$

D_n : negatif haber kukla değişkeni

D_p : pozitif haber kukla değişkeni

$\beta_{i,m}$: hisse i için market getiri oranı katsayısı

$\beta_{i,n}$: hisse i için negatif haber kukla değişkeni katsayısı

$\beta_{i,p}$: hisse i için pozitif haber kukla değişkeni katsayısı

Aşağıdaki tabloda, Tek Faktör MM için regresyon sonuçları, tüm hisse senetlerinin beklenen getirilerinin beklenen piyasa getirisinden önemli ölçüde etkilendiğini göstermektedir. Olumsuz ve olumlu haber kukla değişkenlerinin katsayıları, istatistiksel olarak anlamlı ve teorik olarak bireysel stokların beklenen getirilerindeki değişimleri açıklarken tutarlılık göstermektedir. Genel olarak, olumsuz haberler stokların beklenen getirilerinin azaltılmasında etkili olurken, olumlu haberler stokların beklenen getirisini artırmaktadır. CAPM ve panel data MM de benzer sonuçlara ulaşmıştır.

Tablo D.5- Tek faktör model regresyon sonuçları

AKBNK	Ri =	0.0001	+ 0.704 Rm	+ 0.001 Dp	- 0.004 Dn	R^2 =0.408
	stdev	(0.0003)	(0.0166)	(0.0008)	(0.0009)	
	t ratio	[0.302]	[42.388]	[1.600]	[-4.686]	
ARCLK	Ri=	0.0002	+ 0.51 Rm	+ 0.006 Dp	- 0.005 Dn	R^2 =0.266
	stdev	(0.0003)	(0.0166)	(0.0011)	(0.0012)	
	t ratio	[0.649]	[30.706]	[5.191]	[-4.372]	
ASELS	Ri=	0.0007	+ 0.491 Rm	+ 0.007 Dp	- 0.008 Dn	R^2=0.055
	stdev	(0.0007)	(0.0408)	(0.0025)	(0.0033)	
	t ratio	[1.044]	[12.04]	[2.887]	[-2.316]	
BJKAS	Ri=	0.0023	+ 0.381 Rm	+ 0.004 Dp	- 0.016 Dn	R^2=0.083
	stdev	(0.0005)	(0.0282)	(0.0019)	(0.002)	
	t ratio	[4.637]	[13.495]	[2.088]	[-8.068]	
DENIZ	Ri=	0.0012	+ 0.363 Rm	+ 0.01 Dp	- 0.007 Dn	R^2=0.089
	stdev	(0.0005)	(0.0264)	(0.0014)	(0.0016)	
	t ratio	[2.499]	[13.763]	[6.991]	[-4.3]	
FENER	Ri=	0.0006	+ 0.263 Rm	+ 0.006 Dp	- 0.008 Dn	R^2=0.059
	stdev	(0.0004)	(0.0226)	(0.0016)	(0.0018)	
	t ratio	[1.52]	[11.616]	[4.098]	[-4.606]	
FROTO	Ri=	0.0006	+ 0.45 Rm	+ 0.005 Dp	- 0.005 Dn	R^2=0.184
	stdev	(0.0003)	(0.0188)	(0.0011)	(0.0012)	
	t ratio	[1.911]	[23.91]	[4.67]	[-4.085]	
GARAN	Ri=	0.0012	+ 0.733 Rm	+ 0.001 Dp	- 0.002 Dn	R^2=0.470
	stdev	(0.0003)	(0.015)	(0.0007)	(0.0008)	
	t ratio	[4.281]	[48.74]	[0.84]	[-2.384]	
KCHOL	Ri=	-0.0001	+ 0.589 Rm	+ 0.002 Dp	- 0.003 Dn	R^2=0.343
	stdev	(0.0003)	(0.0157)	(0.0008)	(0.0009)	
	t ratio	[-0.198]	[37.51]	[2.211]	[-2.985]	
SKBNK	Ri=	0.0015	+ 0.569 Rm	+ 0.004 Dp	- 0.007 Dn	R^2=0.210
	stdev	(0.0004)	(0.0216)	(0.0013)	(0.0016)	
	t ratio	[4.041]	[26.389]	[2.829]	[-4.43]	
TCELL	Ri=	0.0001	+ 0.445 Rm	+ 0.003 Dp	- 0.005 Dn	R^2=0.244
	stdev	(0.0003)	(0.0163)	(0.0007)	(0.0008)	
	t ratio	[0.326]	[27.238]	[4.223]	[-6.471]	
THYAO	Ri=	0.0004	+ 0.508 Rm	+ 0.006 Dp	- 0.004 Dn	R^2=0.074
	stdev	(0.0008)	(0.0387)	(0.0015)	(0.0017)	
	t ratio	[0.452]	[13.131]	[3.649]	[-2.42]	
TOASO	Ri=	0.0006	+ 0.54 Rm	+ 0.007 Dp	- 0.006 Dn	R^2=0.244
	stdev	(0.0003)	(0.0191)	(0.0011)	(0.0013)	
	t ratio	[1.826]	[28.281]	[6.414]	[-4.837]	
ULKER	Ri=	0.0013	+ 0.438 Rm	+ 0.007 Dp	- 0.005 Dn	R^2=0.188
	stdev	(0.0003)	(0.0183)	(0.0011)	(0.0014)	
	t ratio	[4.154]	[23.914]	[6.429]	[-3.817]	
VESTL	Ri=	0.0019	+ 0.557 Rm	+ 0.004 Dp	- 0.003 Dn	R^2=0.192
	stdev	(0.0004)	(0.0218)	(0.0014)	(0.0017)	
	t ratio	[4.983]	[25.585]	[2.784]	[-1.893]	

Sınıflandırılan Haber Verileri ile İşlem:

Bu bölümde, sınıflandırılmış haber verileri farklı senaryolarda ticari işlem için kullanılmaktadır. Tüm eylemler (kısa veya uzun pozisyon), bu üç makine öğrenimi sınıflandırıcısına dayalı olarak alınır: Multinomial Naïve Bayes, SVM (Lineer kernel) ve DT. Yani, eğer sınıflandırıcı model bir haberi olumlu (olumsuz) olarak sınıflandırırsa, o zaman strateji olarak uzun (kısa) bir pozisyon alınmaktadır. Ayrıca, tüm pozisyonların, gün sonunda (t günü) ilgili hissenin kapanış fiyatı ile kapanacağı; haberlerin piyasa açılışının başında yayınlandığı ve işlem maliyetinin sıfır olduğu varsayılmaktadır. Özetle, Strateji 1, t günü açılış fiyatı ile işleme başlamayı ve gün sonunda pozisyonu kapatmayı öngörmektedir. Strateji 2 ve Strateji 3, gün t'de yayınlanan haberlerin, (t-1) ve (t-2)'de bilindiğini varsayarak pozisyon alır.

Tablo D.6- Strateji ve işlem başına ortalama getiri

Strateji	**Zaman/Eylem**	**Multinomial NB**	**SVM**	**DT**
1	**Ticarete başlama (t) (açılış fiyatı)**	0,05%	0,06%	**0,09%**
2	**Ticarete başlama (t-1) (kapanış fiyatı)**	0,35%	**0,36%**	0,34%
3	**Ticarete başlama (t-2) (kapanış fiyatı)**	0,62%	**0,65%**	0,59%

Tüm stratejilerin her senaryoda önemli ölçüde pozitif getiri sağladığı açıktır. Bununla birlikte, eğer ticaret maliyeti hesaba katılırsa, ortalama getiriler strateji 1 için işlem maliyetini karşılamayabilir. Bu sonuç piyasada kamuya açık bilgilerle yüksek kar elde edilemeyeceğini öngören yarı-güçlü bir form (semi-strong form of EMH) doğrular.

Strateji-2 ve Strateji-3, anlamlı bir şekilde pozitif getiri sağlar, ancak haberler açıklanmadan önce pozisyonlar alındığından, bunlar mümkün değildir. Bu stratejiler tek bir şart altında uygulanabilir: bu durum içeriden bilgi olması halidir. Yukarıdaki tablodan, içeriden bilgiye sahip bir tüccarın, bu makine öğrenme algoritmalarına dayanarak bir yıllık süre zarfında %449’a varan karları gerçekleştirebileceğini açıkça göstermektedir.

Sonuç olarak, haberler (olumlu ya da olumsuz), bireysel stokların performansını açıklarken istatistiksel olarak anlamlı ve teorik olarak tutarlıdır. Bununla birlikte, makine öğrenimi algoritmalarına dayanan ticaret stratejileri, içeriden öğrenilen bilgiler olmadıkça önemli bir kazanç sağlamamaktadır.

5. SONUÇ

Bu çalışmada, BIST'de işlem gören şirketlere ait finansal haberler, ilgili hisse senedinin günlük stok performanslarına göre pozitif veya negatif olarak otomatik olarak etiketlenerek çeşitli makine öğrenmesi yöntemleriyle duygu analizi yapılmaktadır.

Çalışmadaki yeni yaklaşımın bir sonucu olarak, etiketlenmemiş verilerden değerli bilgiler elde edilmiş ve finansal haberler için etkili kelimelerden oluşan bir liste oluşturulmuştur. Değerli olan bir diğer şey, tüm sürecin otomatik olarak yönetilmesidir. Bu çalışmada yapılan otomatik etiketleme yaklaşımı, büyük bir veri seti oluşturma konusunda bazı teşvikler sunmaktadır. Bu sürece eklenen diğer önemli husus, kutupsallığın belirleyicisi olarak pazar algısının (hisse senedinin getiri yönü) derlenerek objektif kararlar elde edilmesidir.

Sınıflandırıcıların, haber öykülerindeki duyguları ılımlı bir başarı ile tespit etmeyi öğrenebildiklerini görülmüştür. Birkaç öğrenme yöntemine kıyasla, Naïve Bayes sınıflandırıcılarından daha iyi sonuçlar elde edilmiştir: doğruluk oranında yaklaşık %70 ve F-Puanı için ise %80.

Hisse senedi performansına gelindiğinde hem olumsuz hem de olumlu haberlerin ilgili hisse senedi fiyatları üzerinde etkili olduğu görülmüştür. Öte yandan, içeriden öğrenilen bilgiler dışında, kamuya açık bu bilgiden kâr elde etmek zor olduğu, bununla birlikte, haberlerin yayınlanmasından sonra çok kısa bir süre içinde haber duygusunun hayata geçirilebildiği taktirde, haber etiketlerinin finansal açıdan önemli bir girdi olacağı çok muhtemeldir.

Her şeyden önce, daha sofistike özellik kümesinin modellerin doğruluk puanlarını artırabileceği konusunda iyimserlik bulunmaktadır. Boudoukh ve diğ. (2013), haberler arasından şirkete ilişkin ilgili ve ilgisiz haberleri ayırt edebilmek için algoritmaları geliştirmeyi önermektedir. Benzer şekilde, haber konu başlıkların algoritmaya yeni bir sınıf olarak eklenmesi diğer bir iyileştirme noktası olacaktır.

Bu çalışmadaki duygu analizi, sektör ayrımı gözetmeksizin şirketlere özgü haberlere dayanarak oluşturulmuştur. Firmaların sınıflandırılması (sektörlere ayrılması) ve modele başka bir parametre olarak eklenmesi de modeli geliştirecek diğer bir husustur. Böylece bazı kelimeler, kategorisine göre hem negatif hem de pozitif işaretçi olarak modelin doğru tahmin yapmasına fayda sağlayacaktır.

E. TEZ FOTOKOPİSİ İZİN FORMU

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences ☐

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences ☑

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics ☐

Enformatik Enstitüsü / Graduate School of Informatics ☐

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences ☐

YAZARIN / AUTHOR

Soyadı / Surname : KANMAZ
Adı / Name : MEDET
Bölümü / Department : İKTİSAT

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English) : THE EFFECT OF FINANCIAL NEWS ON BIST STOCK PRICES: A MACHINE LEARNING APPROACH

TEZİN TÜRÜ / DEGREE: Yüksek Lisans / Master ☑ **Doktora** / PhD ☐

1. **Tezimin tamamı dünya çapında erişime açılsın ve kaynak gösterilmek şartıyla tezimin bir kısmı veya tamamının fotokopisi alınsın.** / Release the entire work immediately for access worldwide and photocopy whether all or part of my thesis providing that cited. ☑

2. **Tezimin tamamı yalnızca Orta Doğu Teknik Üniversitesi kullancılarının erişimine açılsın. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.)** / Release the entire work for Middle East Technical University access only. (With this option your work will not be listed in any research sources, and no one outside METU will be able to provide both electronic and paper copies through the Library.) ☐

3. **Tezim bir (1) yıl süreyle erişime kapalı olsun. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.)** / Secure the entire work for patent and/or proprietary purposes for a period of one year. ☐

Yazarın imzası / Signature .... **Tarih** / Date ....