CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Stock Market Prediction: A Multiclass Classification on Emotions and Sentiment Analysis for Tweets and News Headlines

Master's thesis

Author: B.Sc. Dejan Lazeski

Study program: Economics and Finance

Supervisor: prof. Ing. Evžen Kočenda M.A., Ph.D., DSc.

Year of defense: 2020

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes. The thesis is submitted electronically to SIS UK.

Prague, July 30, 2020 Dejan Lazeski

Acknowledgments

I am incredibly grateful to my supervisor, prof. Ing. Evžen Kočenda M.A., Ph.D., DSc., for his valuable comments and time dedicated to this thesis. Also, I would like to express gratitude towards my family and Yevgenia for their support, encouragement, and patience while writing this thesis.

This thesis is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 681228.

Bibliographic Record

Lazeski, Dejan: *Stock Market Prediction: A Multiclass Classification on Emotions and Sentiment Analysis for Tweets and News Headlines.* Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2020, pages 78. Advisor: prof. Ing. Evžen Kočenda M.A., Ph.D., DSc.

Abstract

In this thesis, we look beyond extracting binary sentiment in regards to News Headlines and Tweets. As a data source, we target tweets and headlines from well-known financial newspapers, explicitly addressing the top 5 Big Tech companies. To examine the effectiveness of sentiment and Ekman's emotions in predicting future stock price movements, we develop multiclass emotion and sentiment classifiers utilizing a supervised learning approach. Moreover, we manually annotate our corpora for positive, negative, and neutral sentiment as well as one of Ekman's emotions: anger, joy, surprise, sadness. We did not confirm any robust correlation between daily stock price movements and the distribution of sentiment and emotions. However, we did observe that tweets are less neutral than news headlines. Finally, we implement a simple investing strategy by extracting sentiment polarity scores using VADER and other metrics such as followers and shares. Two classifiers, SVM and ANN, delivered robust predictions for Google and Amazon compared to weak predictions for the rest of the companies. Nevertheless, the results suggest that sentiment polarity can effectively predict future stock price movements compared to finer-grained emotion classification.

JEL Classification	C53, G41, G17, C61
Keywords	News Headlines, Tweets, Sentiment Analysis, Emotions
Title	Stock Market Prediction: A Multiclass Classification on Emotions and Sentiment Analysis for Tweets and News Headlines

Abstrakt

Tato práce zkoumá využití sentimentu na základě titulků zpráv a tweetů. Hlavním zdrojem dat jsou tweety a novinové titulky z dobře známých finančních novin, speciálně cílené na top 5 "Big Tech" firem. Abychom prozkoumali užitečnost sentimentu a emocí dle Ekmana v odhadu budoucích cen akcií, vytvořili jsme vícetřídní klasifikátory emocí a sentimentu za použití přístupu strojového učení. Zkoumané zdroje dat byly manuálně ohodnoceny pro pozitivní, negativní a neutrální sentiment a také k nim byly přiřazeny primární emoce podle Ekmana, jako jsou hněv, radost, překvapení a smutek. Nepotvrdila se nám žádná významná korelace mezi denním pohybem akcií a rozložením sentimentu. Bylo však zjištěno, že tweety jsou méně neutrální než novinové titulky. Nakonec jsme zavedli jednoduchou investiční strategii extrakcí skóre polarity za použití VADER a dalších metrik jako počet sledujících a sdílení. Dva klasifikátory, SVM a ANN, se vyznačovaly silnou predikcí u akcií Googlu a Amazonu, ale slabou predikcí u ostatních firem. Výsledky práce naznačují, že polarita sentimentu může lépe předpovídat budoucí výkyvy cen akcií než vícetřídní klasifikace emocí.

Klasifikace JEL	C53, G41, G17, C61
Klíčová slova	Titulky zpráv, Tweety, Analýza Sentimentu, Emoce
Název práce	Prognóza Vývoje Akciových Trhů: Vícetřídní Klasifikace Emocí a Analýza Sentimentu na základě Tweetů a Titulků Zpráv

Contents

List of Tables

List of Figures

Acronyms

API Application Programming Interface

NLP Natural Language Processing

BoW Bag of Words

TF-IDF Term Frequency, Inverse Document Frequency

MCC MultiClass Classification

SVM Support Vector Machine

KNN K Nearest Neighbors

VADER Valence Aware Dictionary and sEntiment Reasoner

DJIA Dow Jones Industrial Average

ANN Artificial Neural Network

Master's Thesis Proposal

Author	B.Sc. Dejan Lazeski
Supervisor	prof. Ing. Evžen Kočenda M.A., Ph.D., DSc.
Proposed topic	Impact of Media on Financial Markets

Motivation ”A Sunday New York Times article on a potential development of new cancer-curing drugs caused EntreMed's stock price to rise from 12.063 at the Friday close, to open at 85 and close near 52 on Monday. It closed above 30 in the three following weeks”, more than a 400% return in a day (Huberman & Regev 2001). What makes this story more interesting is that the research had been already published, but in a scientific journal Nature, and in other newspapers including the Times (Campbel et al 2001). Obviously, this public attention led to a permanent rise in share prices. Reading such articles and more recently about a small-town Veles in Macedonia (known in the press as the ”factory of fake news”) from Macedonia and the large influence that these fake news had on the USA presidential elections, I was inspired to choose my topic, which is also informed by a number of recent studies that demonstrate strong correlations between media and stock market. The distribution of information plays a crucial role in shaping financial markets (Da, Engelberg, & Gao, 2011, Strycharz et. al 2018). In particular, financial news, market announcements, corporate news, or analyst forecasts are reflected in volatile stock market reactions (Tetlock, 2014).

The article above is a clear contradiction of the efficient markets hypothesis (EMH) (Malkiel & Fama, 1970), which is also confirmed by many scholars who find that financial news in particular influences the stock market to varying extents (Kleinnijenhuis et al. 2015). The article shows us that the financial markets are not efficient such that all available information is instantaneously integrated in prices as assumed by proponents of the EMH (Strycharz et. al 2018). ”If we assume that not all investors are equally well informed, having enough time and attention to process and evaluate information, it is relevant that scholars from behavioral economics are questioning the EMH that financial markets are solely governed by rational market participants. Thus, trading decisions are said to be shaped and influenced by emotions, herd, and irrational behavior. In this regard, the media have been identified to play a significant role in shaping the consensus market opinion and evoking this ”herdlike” behavior” (Kleinnijenhuis et al. 2015, Strycharz et. al 2018).

The Internet today is integrated into every aspect of our lives and has undoubtedly improved our ability to access information in real time. A particular aspect of the internet with substantial growth is social media. An example is provided by Twitter, a blogging and message sharing service which started in 2006. Today Twitter has more than 330 million monthly users and over 500 million daily tweets, and is used globally by a broad demographic to publicly broadcast. According to (Zheludev et al. 2014 see pp. 01) ”For the first time in human history, it is arguably possible to monitor the moods, thoughts and opinions of a large part of the world's population in an aggregated and real-time manner with almost negligible data-collection costs. Of present focus is the prediction of financial markets via the analysis of Tweets and other comparable data sources such as Google Trends, Yahoo! search engine data and Wikipedia articles.”. What is also important to note is that social media content is updated rapidly and spreads virally with an unprecedented pace, bringing first-hand information to investors ahead of other sources.

Hypotheses

> Hypothesis #1: Fake news on Twitter during U.S. 2016 Presidential Elections had no impact on future stock price movements.
>
> Hypothesis #2: Fake news on Twitter and its volume during U.S. 2016 Presidential Elections had no impact on future stock price movements.
>
> Hypothesis #3: Fake news on Facebook during U.S. 2016 Presidential Elections had no impact on future stock price movements.

Methodology ”We will folow almost identical methodology from (Strycharz et. al 2018 see pp. 75). starting with a vector autoregression model (VAR) as it takes the interdependence of variables into account. In that way, both media variables and the stock market fluctuation variable are considered as dependent and independent variables at the same time in the estimation models. I intend to construct a VAR model for the variable measuring stock market fluctuation as well as one of the respective media variables (e.g., attention, sentiment, emotionality, topics) based on the studies from (Scheufele et al. 2011, Vliegenthart 2014, Strycharz et. al 2018). where, Augmented Dickey-Fuller (ADF) test will be estimated to assure that all time series were stationary. And if this is not the case, I plan to difference the series until stationarity is achieved. Second, in order to select the optimal number of lags for each VAR model, selection-order statistics will be consulted (e.g., Akaike's information criterion). Third, Granger-causality tests will be performed to identify whether one series predicts another series above and beyond the past values of its own series.

In the final steps, cumulative impulse response functions (CIRF) and forecast error variance estimations (FEV) will be carried out" (Vliegenthart 2014, Strycharz et. al 2018).

Expected Contribution From a hypothetical perspective, the discoveries will ideally affirm the discoveries of past exploration that media consideration is a significant factor in understanding financial exchange responses and feelings and assumes a critical job in advertise developments. Second, this study will give significant insights for strategic financial communication and the role of news media in dealing with the assessment of firms on the stock market (Vliegenthart 2014, Strycharz et. al 2018).

Outline

1. Introduction:
2. Literature review:
3. Data:
4. Methodology:
5. Empirical analysis:
6. Conclusion:

Core bibliography

Campbell, J. Y., M. Lettau, B. G. Malkiel, & Y. Xu (2001): "Have indi-vidual stocks become more volatile? an empirical exploration of idiosyncratic risk." The journal of finance 56(1): pp. 1-43. _of_Twitter.pdf, 2011.

J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. Journal of Computational Science, 2011

Da, Z., J. Engelberg, & P. Gao (2011):"In search of attention." The Journal of Finance 66(5): pp. 1461-1499.

Kleinnijenhuis, J., F. Schultz, S. Utz, & D. Oegema (2015): "The medi-ating role of the news in the bp oil spill crisis 2010: How us news is influenced by public relations and in turn influences public awareness, foreign news, and the share price." Communication research 42(3): pp. 408â€“428.

E.F. Fama. The behavior of stock-market prices. The journal of Business, 38(1):34-105, 1965.

Tetlock, Paul, 2007, Giving content to investor sentiment: The role of media in the stock market, Journal of Finance 62, 1139-1168.

Tetlock, P. C. (2014). Information transmission in finance. Annual Review of Financial Economics, 6(1), 365-384. doi:10.1146/annurev-financial-110613-034449

Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. Academy of Management Review, 20(1), 65-91. doi:10.5465/AMR.1995.9503271992

Strauss, N., Vliegenthart, R., & Verhoeven, P. (2016). Newspapers lagging behind? Emotions in newspaper articles and stock market prices in the Netherlands. Public Relations Review, 42(4), 548-555. doi:10.1016/j.pubrev.2016.03.010

Strauss, N., Vliegenthart, R., & Verhoeven, P. (2017). Intraday News trading: The reciprocal relationships between the Stock Market and Economic News. Communication Research.

Strycharz, J., N. Strauss, & D. Trilling (2018): "The role of media coverage in explaining stock market fluctuations: Insights for strategic financial communication." International Journal of Strategic Communication 12(1): pp. 67-85.

Vliegenthart, R. (2014): "Moving up. applying aggregate level time series analysis in the study of media coverage." Quality & quantity 48(5): pp.2427-2445.

Huberman, G. & T. Regev (2001): "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar." The Journal of Finance 56(1): pp. 387-396

Zheludev, I., R. Smith, & T. Aste (2014):"When can social media lead financial markets?" Scientific reports 4: p.4213.

Author Supervisor

Chapter 1

Introduction

The 2016 U.S. Presidential Elections was by far one of the most exciting elections in history. In the preceding two months of election day, a plethora of events was the center of attention. Social media was flourishing of comments and discussions. According to Facebook, election-related content generated 716.3 million likes and 643 million views (Guynn 2016). Twitter was no exception. More than 1 billion tweets were sent since the start of the critical debates, with Hillary Clinton sending the famous tweet ”Delete your account” addressing Donald Trump (Coyne 2016). However, misinformation and false stories - ”fake news” were also a big part of the elections on both social platforms. Allcott & Gentzkow (2017) in their work does not provide any evidence that ”fake news” influenced the election score, however they do provide a well documented overall assessment of fake news circulating the election period. It would be fascinating to investigate the direct impact of fake news on the 2016 election and stock market returns. However, fake news had such a significant influence during the 2016 U.S. Presidential Elections, that many of those tweets and accounts were either deleted or suspended from Twitter and Facebook. For standard accounts, Twitter limits a sampling of recently published tweets in the past seven days and has a strict policy that tweets ids can be publicly shared, however, tweets cannot. Due to this fact, free and big enough dataset of verified fake news on which we can conduct analysis was almost impossible to obtain. However, this episode highlights the importance of the topic of our thesis that media play a vital role in creating public opinion. That being said, we would like to emphasize that the idea of this thesis has changed from what was suggested in the original proposal. Instead we will examine the effectiveness of News Headlines and Tweets in predicting future stock price returns through multiclassification using supervised machine learning approach.

The effect of news media in various aspects of the financial markets is already well documented for which many studies exist. Tetlock (2007); Engelberg (2008); Garcia (2013) focus on the correlation between news articles and future return of the major stock market indices, Li *et al.* (2014b) investigated the relation of company specific news and their stock price movements, Heston & Sinha (2017) conducted a study on whether news are affecting mergers and acquisitions. Social media, in particular Twitter, and the relation with stock price movements is also well documented through many studies. Zhang *et al.* (2011) are among the first to look for correlation of a wider spectrum of emotions in tweets and price movements of Dow Jones, NASDAQ and S&P 500. Mao *et al.* (2012) neglect sentiment analysis and focus on correlation between volume and stock prices of tweets mentioning S&P 500, both for sector and company specific level. Bollen *et al.* (2011) measure six dimensions of mood and binary sentiment polarity in tweets.

The number of technological innovations for delivering information to people has grown drastically since the creation of the internet. The costs of collecting, aggregating, and consuming information have decreased dramatically over time. This has resulted in the quantity of people's information flow, accuracy of their beliefs, and investor's access to information increasing exponentially at the same time. Historically when there is technological change, it is usually surrounded by excitement and enthusiasm about potential improvements and benefits that the technology can bring. This is no different for investors, who have embraced this technological progress and implemented various algorithmic models of tracking public opinion through sentiment analysis. The traditional analyst model is slowly fading away. Today, analysts do not need to read every earnings announcement or transcript. On the contrary, they can rely on sentiment analysis, instead highlighting the most important news.

At the same time concerns about unintended consequences and potential downsides of those technologies are justified by the fact that the overall effect over long periods of time is a combination of indeed sentiment analysis makes markets more efficient but also interacts with biases and inefficiencies in the markets. It is understood that people exhibit a wide range of emotions comparing to the more frequently binary classification between positive and negative as majority of sentiment analysis approaches have. Moving beyond only negative and positive sentiments towards a fine-grained emotion classification scheme may improve the information investors receive about future business valuation. If we observe words like Fear and Anger from a semantic and emotional perspective we can claim that both express negative sentiment, however Lerner & Keltner (2000) in their research concluded that fearful people tend to make more pessimistic risk assessments, whereas angry people tend to make more optimistic risk assessments, leading to important implications in the financial domain of stock price movement prediction.

In this thesis we will evaluate to which extent we can predict a known dimension of stock price movements from the top five Big Tech companies by extracting features from Tweets and News Headlines of well known professional newspapers. Since both tweets and news headlines are sentence level documents, we argue that they are comparable and the semantic importance of particular event will be expressed in a few words. We develop finer-grained emotion and sentiment classifiers utilizing a supervised machine learning approach. In addition, we create simple investing strategy by extracting sentiment polarity using Valence Aware Dictionary and sEntiment Reasoner (VADER). To the best of our knowledge, study predicting future price movements using as a proxy headlines and tweets from the same source in parallel was not conducted. Furthermore, most of the studies examine binary sentiment analysis or finer-grained classification of emotions, however not comparing them.

The rest of the thesis is structured as follows. Chapter 2 covers the previous scholarly work investigating links between news, twitter and capital markets. In addition, previous work on finer-grained emotion classification and datasets in covered. Chapter 3 describes the data we use and how we build and evaluate the classifiers for both sentiment analysis and emotion classification. In Chapter 4 we examine the correlation between news headlines and tweets daily distribution and future stock price returns. In Chapter 5 we propose an investing strategy based upon the extraction of sentiment polarity using VADER. Chapter 6 is an overall discussion. Finally, Chapter 7 provides an conclusion of our findings and proposals for future work.

Chapter 2

Literature review

In this chapter, we summarize the growing set of literature that addresses our research methodology of extracting emotions and sentiment from tweets and news headlines. More specifically, we group and explain the latest research in the field of sentiment analysis and emotion multiclassification applicable for stock market prediction.

2.1 Stock Market Performance and Investor Sentiment

Academic research such as Tetlock (2016) has shown that media play at least three somewhat interrelated roles in the stock market. First, media attracts attention to important current events, followed by the fact that media conveys information. The last of these roles is probably the most controversial, i.e., media influence individuals beliefs about current events by providing compelling interpretations of these events. In the financial markets thousands of events are occurring daily and investors as human beings can obviously notice and recall only a limited amount of this information. We all have imperfect memories, hence media focus our attention by selecting certain news that are important among thousands and promotes those particular news. In doing so the media also aids our memory by exploiting various cognitive heuristics, such as that investors attend salient stimuli which stand out against the background and that investors can recall only certain information (see Tetlock 2016, pp. 03)

The efficient market hypothesis by Fama (1970) indicated that all available information is instantaneously integrated into prices. However, many scholars such as Malkiel (2003) contradict this by pointing out the fact that media influences the stock market to vary extents, Malkiel (see 2003, pp. 01) says, "But news is by definition unpredictable and thus, resulting price changes must be unpredictable and random". Capital markets by far are one of the most information-sensitive markets, and when provides information in an unbiased way, investors can still misinterpret the information, or for some reason, media can be biased when interpreting certain events, thus creating misconception for investors. Shiller (see 2003, pp. 90) also contradicts Fama (1970) by saying that, "The efficient markets model, for the aggregate stock market, has still never been supported by any study effectively linking stock market fluctuations with subsequent fundamentals".

Brown & Cliff (2005) as part of the behavioral finance in their paper, provides evidence that mispricing of stock valuation can be explained by sentiment. They use investor sentiment as a proxy variable and conclude that when there is optimism among investors about a specific stock, that particular stock will obtain a premium price, and the market, in general, will tend to be overvalued. However, in their earlier work Brown & Cliff (2004) using a signal extraction approach for investor sentiment across direct measures find no evidence of short term predictions of future stock returns.

Baker & Wurgler (2007) take behavioral economics approach to explore which stocks are most affected by sentiment and conclude that the effect of investor sentiment should not be questioned anymore but on the contrary we should ask how to determine its effects since there are plenty of historical events with sensational stock price changes, e.g., "the Great Crash of 1929, the 'Tronics Boom of the early 1960s, the Go-Go Years of the late 1960s, the Nifty Fifty bubble of the early 1970s, and the Black Monday crash of October 1987". Baker & Wurgler (2006; 2007) note that stocks with speculative character tend to be more sensitive to investor sentiment. Also, they examine low capitalized, unprofitable, highly volatile, and non-dividend paying companies where they determine that such companies yield much higher returns in contrast to low investor sentiment. Furthermore, Baker & Wurgler (2006; 2007) conclude that stocks tend to be overvalued before a crash because of extensive optimism, opposite to that pessimism leads stock prices to be undervalued.

2.2 News and Stock Market Performance

When discussing financial news, the majority of previous influential work is focusing on sentiment in whole articles rather than the headlines of those particular articles. Tetlock (see 2007, pp. 1143) summarizes that "The sentiment theory predicts short-horizon returns will be reversed in the long run, whereas the information theory predicts they will persist indefinitely.". Therefore, it is important to make a distinction between permanent and transient news impact, whereas we can conclude that permanent is the information itself, and transient news impact will indicate the sentiment. In general, the majority of previous findings have contemporaneous or transient consequences on stock price returns and trading volumes (Tetlock 2007; Engelberg 2008; Schumaker *et al.* 2012; Garcia 2013).

One of the most influential papers in the specific field of sentiment analysis is the work of Tetlock (2007). He examines general financial aspects and news in the daily column "Abreast of the Market" from the U.S. newspaper "The Wall Street Journal" over some time between 1984 and 1999. For content analysis, he is using General Inquirer developed by Stone *et al.* (1966) and Harvard IV-4 dictionary containing most of the word lists, furthermore he applies the Vector Auto-Regression method. Tetlock's main discovery is that excessive negative sentiment prompts market prices to fall. Traders start to be more active throughout the trading day, causing increased volatility and market volume in the short term. Consequently, lower prices create even more negative news attention, which influences investor psychology.

Engelberg (2008) investigates news articles mentioning earning announcements of approximately 5000 companies over a period between 1999 and 2005. Similarly, to Tetlock, for deriving sentiment, he is using dictionary-based General Inquirer/Harvard IV-4. His methodology approach is a linear regression model using an event study where for the dependent variable, he takes the cumulative abnormal return, which is the difference between actual return and a standard return. He concludes that earning announcements published in news articles containing qualitative information has additional predictability for future returns.

Schumaker *et al.* (2012) in their research uses a system called Arizona Financial Text (AZFinText) specifically designed to collect financial news articles from Yahoo! Finance and predict almost live price movements by assigning particular sentiment polarity. AZFinText employes the Sequential Minimal Optimization, a machine-learning algorithm, and ten-fold cross-validation. Their first main finding is that subjective articles are easier for prediction, the second finding is that articles with negative sentiment are most comfortable determining in which direction the price will move. And lastly, they correlate downward movement in price with news articles labeled with positive sentiment, which contradicts Tetlock (2007) findings. Moreover, worth mentioning is that none of their findings has a higher accuracy score of 60%.

Similarly to Tetlock (2007), another research focusing on financial articles from newspapers such as the New York Times is Garcia (2013). However, Garcia (2013) in his work for that time utilizes a novel financial lexicon developed by Loughran & McDonald (2011) considered as the founding fathers of financial sentiment. This lexicon factors in alternative negative and five other word list that better reflects tone in the financial text. Garcia (2013) examines the relation of exact market responses and sentiment polarity in news articles during both recession and non-recession periods, where he concludes that during recessions, only for daily frequencies news content can predict stock returns.

Our datasets of tweets and news headlines are gathered for five particular S&P 500 companies. The same to our approach of company-specific news has Li *et al.* (2014b), where they obtain news for companies listed on the Hong Kong Stock Exchange. Their study differs from the earlier mentioned since they integrate stock price prediction framework one step before using both previously mentioned Harvard IV-4 and Loughran-McDonald sentiment dictionaries. After they compare the results from their approach with the already well-known measure for word frequency Bag of Words (BoW), they conclude that implementing their model with sentiment analysis can lead to higher prediction accuracy, and it outperforms BoW. However, when assigning only positive and negative labels, the model has just moderate results. Lastly, there is almost no difference in the results comparing both used dictionaries.

A recent study by Heston & Sinha (2017) differs from the previously mentioned work by additionally aggregating news for a week. They use a broader and larger dataset of more than 900,000 news stories already containing various metrics like sentiment or staleness of news provided by the neural-network engine of Thomson Reuters NewScope Data. Their principal findings compared to previous work are that news daily produces short-term predictions on price direction for no more than a week. However, when the news is aggregated for more than a week, the price prediction is positive for one quarter. More precisely, positive news predict returns for one week, negative news can give predictions for at least one quarter.

The last paper we will discuss in this section is one very recent work from Liao *et al.* (2019) who is focusing on whether the news is affecting mergers and acquisitions. Their central hypothesis is whether more media coverage can influence some firm decision to merge. They use a sample of approximately 78,000 mergers and acquisitions from around 200 countries together with news sentiment scores provided by RavenPack. They find that the more optimistic sentiment embedded in the news, the higher chance of becoming an acquirer. However, they also conclude that if the acquirer receives high media coverage, the more likely, it will experience negative post-acquisition returns.

2.3 Twitter and Stock Market Performance

Over the past decade, the rise of social media has enabled millions of people to share their opinions and react to current events in real-time. Regarding social media, Twitter as one of the most popular micro-blogging platforms has attracted several streams of researches to investigate its financial forecasting role (Arias *et al.* 2014; Kordonis *et al.* 2016). In general, researchers apply a variety of sentiment analysis on this massive data source to obtain public opinion and try to leverage its predicting power or correlation with the stock markets, however, we can generalize and say that most of the studies are focusing on how twitter volume affects the financial markets (Mao *et al.* 2012; Ranco *et al.* 2015) or focus on semantic context in tweets (Bollen *et al.* 2011).

Zhang *et al.* (2011) are among the first who try to extract a wider spectrum of emotions from tweets and investigate if a correlation exists with all three major indices (Dow Jones, NASDAQ, and S&P 500). They found no correlation between outbursts of negative or positive emotions in tweets and the stock market. Liu (2017), in her work, has a multiclassification approach for tweets into Ekman's six basic emotions using supervised machine learning. Again, she confirms no robust correlation in any of the emotions and future stock market performance. However, when she applies a combination of Twitter volume and sentiment polarity, she captures earnings announcements and their impact on future stock returns. Similarly, a previous study of (Ranco *et al.* 2015), confirm that Twitter volume is a useful variable for forecasting abnormal returns for a sample of Dow Jones companies.

Mao *et al.* (2012) in their paper neglect the sentiment and focus only on whether there is a correlation between daily volume and stock prices of tweets mentioning S&P 500 divided on between sector and company-specific level. They apply linear regression using Twitter data as exogenous input. They confirm the correlation between the daily volume of tweets and future stock returns and suggest that twitter can be a useful data source for future stock prediction. Another work confirming Twitter volume as relevant for the forecasting stock price movements on S&P 500 is Oliveira *et al.* (2017), however, on firms with lower market capitalization and only some particular industries.

A breakthrough in identifying emotions from tweets was made by Bollen *et al.* (2011). They measure six dimensions (Calm, Alert, Sure, Vital, Kind, and Happy) of mood and binary sentiment polarity (positive or negative) in tweets. Speaking about polarity worth mentioning is that Ranco *et al.* (2015) did find a positive correlation with future stock returns of Down Jones, however not just because of the polarity of the tweets but the calmness dimension precisely. Another research confirming correlation with Dow Jones, happiness and calmness is the work of Mittal & Goel (2012). They use a similar approach to ours in Chapter 5 where they employ a neural network algorithm and then develop a portfolio management strategy of buy and sell decisions using as a proxy the mean values of future stock prices.

Similar to our approach of whether the sentiment expressed for selected companies mentioned on Twitter can indicate their stock movements examine Smailović *et al.* (2014). They develop Support Vector Machine (SVM) classifier, which shows that few days in advance sentiment in tweets mentioning related companies can predict stock price movements, especially when they move from binary to multiclassification of positive, negative, and neutral category. Furthermore, Kordonis *et al.* (2016) use a same approach, explicitly targeting tweets for popular tech companies, according to Yahoo! Finance and extracting sentiment polarity. Their results are auspicious and indicates that stock price movements can be predicted with a high chance of using various machine learning techniques. Another companion paper worth mentioning is the work of Pagolu *et al.* (2016). They collect 2,500,000 tweets on Microsoft for almost one year. Word2vec, together with N-grams, is used as textual representations for sentiment derivation. They conclude that positive news and tweets would drive the stock price upward due to the higher magnitude of encouragement in positive sentiment. Furthermore, they present a strong correlation in both direction of stock price swings due to public opinions or emotions in tweets.

2.4 Emotion Classification

Sentiment analysis is a well-researched methodology in the field of Natural Language Processing (NLP), however, this is not the case of fine-grained emotion classification. Nowadays, without a high cost in collecting data, we can monitor individuals mood, reactions, and emotions live and on an aggregated level through social media and the internet as a whole. Discussing about categorization of sentiment data into emotions, most of the computational linguistics are focusing on the discrete emotion categorization, since it observes the categorization of the natural language. Ekman (1992) identified six basic emotions (anger, fear, sadness, joy, disgust, and surprise). He argues that these primary emotions share nine characteristics, some interconnected, some unique to each emotion. Plutchik (1984) identified eight, which he paired in four opposites (trust-distrust, joy-sadness, surprise-anticipation, anger-fear). In the field of NLP and multiclassification tasks, most commonly used are Ekman's six basic emotions. In our research, we will follow the trend and use Ekman's emotions, however, based on Liu (2017) findings, disgust had the lowest F1 score and is the least represented among the other emotions in the majority of datasets she explored. Therefore, for our multiclassification task, we will neglect disgust and focus only on the other five.

One of the pioneers in-text annotation are Strapparava & Mihalcea (2007). During SemEval 2007 held in Prague, they presented a task for news headlines annotation in one of Ekman's basic emotions. This corpus is widely used in many studies, however it is not suitable for our research since the headlines are not from financial domain. One example of a manually annotated training dataset for further classifier training is (Roberts *et al.* 2012). The corpus is based on Ekman's six emotions. Additionally, they add love believing it is very closely related to joy, and they chose 14 topics believing would show emotion on Twitter.

Mohammad (2012b), in his paper, emphasizes that today in regards to emotion classification, we need to distinguish between supervised machine learning where the classifiers use features such as unigrams and bigrams to learn from (Alm *et al.* 2005) and affect lexicon-based methods which are a list of words expressing some emotional trait or sentiment. However, this approach is practical to some extent. For example, the sentence ”I wish it was not raining today” clearly expresses sadness. However, none of the words can be classified as ’sad’ words. In the early days, researchers were employing lexicon-based approaches more often because of its memory and computational efficiency. However, in recent years the technology has improved significantly, therefore, in our thesis, we will use supervised machine learning methods.

In Klinger *et al.* (2018) thorough analysis of annotated corpora for emotion classification in text, it is visible that emotion classification still has room for improvements comparing to the success of sentiment analysis. Mohammad (2012a) Twitter Emotion Corpus contains more than 21 000 tweets tagged with one of Ekman’s six emotions. Mohammad (2012a) main idea was to test if emotion-word hashtags can be used as emotion labels in creating emotion labeled corpus. Liu (2017) was using this corpus and concluded that the corpus is not suitable for analysis in the financial domain since the corpus represents only a small portion of all tweets not contain company-specific tweets nor financial jargon. Mohammad *et al.* (2015) also created the Electoral-Tweets The dataset based on the 2012 US presidential election tweets. Since emotional responses during election season are highly relevant, they approached crowdsourcing where more than 100,000 responses from two online questionnaires are obtained. Worth mentioning is that all Mohammad datasets are free for download for research purposes[1]. Schuff *et al.* (2017) during SemEval 2016 created The Stance Sentiment Emotion Corpus to contribute to closing the gap of scarce emotion annotated datasets and provide further research. The corpus contains approximately 5,000 tweets following Plutchik’s emotions, however, none of them are related to the financial domain.

Two very recent datasets were released by Bostan *et al.* (2019) on news headlines and Chatterjee *et al.* (2019) on textual dialogues. Both datasets are labeled in one of Ekman’s emotions using various machine learning techniques with respectful accuracy results, however, none of them again are related to finance.

To the best of our knowledge, most of the previous work is focusing on collecting tweets with specific queries using ’cashtags’ across the whole twitter

[1] http://saifmohammad.com/

domain but not narrowing to specific users. Furthermore, as described, financial training datasets labeled with emotions are almost non-existent, therefore by downloading tweets and news headlines from well-established newspapers and quantifying their impact on the financial markets this thesis represents a useful contribution to the existing literature.

Chapter 3

Creating Emotion and Sentiment Classifiers

3.1 Data Sources

Can we use the content of Tweets and News Headlines to predict stock price performance? According to previous work, we can agree on a short term basis. Bollen *et al.* (2011) using tweets was able to find a correlation between calmness and market index, Garcia (2013) documents that news content can predict daily stock returns, specifically during recessions, and Mittal & Goel (2012) also confirmed correlation with Dow Jones Industrial Average (DJIA), calmness, and happiness using tweets. Wuthrich *et al.* (1998) and more recently Wong & Ko (2016) use daily news articles from well-known business newspapers for market index prediction, however, their accuracy results are only moderate. The abundance of today's unstructured data is undeniable, and text is one of the leading sources. However, the challenge is how we digest and interpret the signals from the data to furthermore help investors make better investment decisions.

Our first data source is Twitter, an online micro-blogging social network with over 330 million active monthly users and more than 500 million tweets sent per day as of May 2019. Users on Twitter create short messages called tweets shared with other users who interact by retweeting and responding. Since 2017 Twitter employed a message size restriction of 280 characters when it expanded its character count from 140 to 280, however historically, only

5% of tweets are longer than 140 and only 2% more than 190 characters[1]. A limited number of characters guides users to express their opinions, emotions, and message they wish to communicate with their followers or other users. In 2006 Twitter officially introduced Application Programming Interface (API), and together with the short text messages, it became a massive data source of public opinion, news, and most recent trends. All this makes Twitter one of the most popular platforms for investors to explore its predicting power through a variety of machine learning techniques for sentiment analysis.

News Headlines, our second data source, is one of the primary origins for financial information wherein correlation with the unprecedented availability of technology provides us the opportunity to advance the state of research in understanding the predictive power of the news in stock price movements. According to Schumaker *et al.* (2012), news is treated as an authentic and realistic source, hence this power, if harnessed correctly, could help predict financial outcomes and produce a significant economic impact on the world. It is impossible to overestimate the importance of news as a thermometer for people's inclinations today.

Both headlines and tweets deliver the message in a Sentence-Level and short format. However, Wong & Ko (2016) suggests that the majority of Twitter users are young people, and their messages do not reflect sentiment adequately. Hence, we will compare whether this hypothesis stands when the headlines and tweets come from a more relevant source, a financial news outlet.

We decided to narrow our search to seven well-known business newspapers and news websites such as CNBC, The New York Times, The Wall Street Journal, Reuters News, Market Watch, Financial Times, and Business Insider based on unbiased and independent web metrics extracted from three different search engines such as Google Page Rank, Alexa Traffic Rank and Ad Fontes Media where all mentioned sources provide high daily information flow. Furthermore, we are focusing on Big Tech companies and choose the top five Standard & Poor S&P 500 Index companies: Facebook, Apple, Amazon, Microsoft, and Google(Alphabet). These companies are highly discussed and shared on Twitter and in the news since they make up to 13% of the total S & P 500 value and have a market capitalization of over five trillion dollars.

[1]https://www.statista.com/topics/737/twitter/

All computations, data processing and visualization in this thesis are written in Python language.

3.1.1 Twitter

Downloading data from Twitter with a standard account has limitations regarding how far back in time and how many requests can be made. For non-paying developer access, Twitter limits a sampling of recently published tweets in the past seven days. Per Twitter's Developer Policy[2], tweet ids can be publicly shared. However, tweets cannot. Thus, if you have the tweets ids, you can retrieve a tweet using the GET statuses/user with the Twitter REST API. This method has downsides as the Twitter API will not return deleted tweets nor tweets from suspended accounts.

Another way to download data from Twitter is by using the Streaming API for data being produced in real-time. As we are certain which keywords and hashtags for each of the five companies (e.g., Amazon or Jeff Bezos) we want to have in our search criteria, we use the Streaming API to obtain tweets on a daily basis. To automate the tweets collection, we use a Python library called Tweepy. For connecting to Twitter API and download the tweets, a periodically running script was developed and launched between April 1st and December 31st, 2019. Every time we run the script, a query would go seven days historically and download tweets posted at different points in time, resulting in 18,647 tweets. Table 3.1 shows the three most retweeted tweets in our dataset.

Table 3.1: Top Retweeted Tweets

Tweet	**Source**	**Retweets**
This is why Apple products are so expensive.	Business Insider	2,721
U.S. lawmakers urge Apple to restore HKMap app used in Hong Kong	Reuters	2,215
Facebook and its other apps including Instagram, WhatsApp, Messenger are all experiencing outages.	Business Insider	1,675

Twitter considers these tweets as distinct tweets, even though they may have the same text. Most of the news websites employ software to periodically

[2]https://developer.twitter.com/en/developer-terms/policy#id8

post their articles for high exposure. It is good practice to post the same article every six to eight hours since not all their followers are active on Twitter at the same time. Consequently, after removing the tweets with identical text, a dataset of 13,658 tweets is obtained. Table 3.2 shows the distribution of tweets from each source and their number of followers after removing duplicates.

Table 3.2: Tweets by source and followers

Source	**Tweets**	**Followers**
Business Insider	4,646	2,985,548
CNBC	3,233	3,655,163
Reuters	1,959	22,047,219
New York Times	1,033	799,865
Market Watch	1,002	3,739,035
Financial Times	897	6,628,434
Wall Street Journal	887	561,239

3.1.2 News Headlines

For gathering our news headlines dataset, instead of web scraping we are using BuzzSumo.com. The reason is that many websites, especially financial newspapers, do not allow their data to be scraped if you are not subscribed. BuzzSumo is a powerful online tool that allows finding what the top-performing content across the internet is. BuzzSumo offers a comprehensive search engine that allows us to filter by specific data and keywords (e.g., Amazon or Jeff Bezos) to obtain the most relevant headlines by their total engagement across the web. Table 3.3 shows the top three headlines by total shares and engagements across the web.

Table 3.3: Top News Headlines and their share across the web

Headline	**Source**	**Facebook**	**Twitter**	**Reddit**
Microsoft's 4 day workweek led to 40% boost in productivity	Business Insider	1,128,600	8,059	1,584
Jeff Bezos would pay $9 billion a year in taxes under Sanders' plan	CNBC	673,774	3,753	21,895
Profitable Giants Like Amazon Pay $0 in Corporate Taxes. Some Voters Are Sick of It.	New York Times	539,489	14,428	5,597

To cover three quarterly announcements and capture visible peaks in volatility and news activity, headlines were obtained for some time between April 1st and December 31st, 2019. Resulting in a dataset of 9,676 different headlines, their source, date when it was published, and total engagement across social networks such as Facebook, Twitter, and Reddit. Table 3.4 shows the distribution of headlines from each source after removing duplicates.

Table 3.4: News Headlines by source

Source	**Headlines**
Business Insider	4,140
CNBC	2,335
Reuters	1,154
New York Times	567
Wall Street Journal	562
Market Watch	511
Financial Times	406

3.1.3 Training Data

Predicting stock movement using tweets and news headlines through fine-grained emotion classification and sentiment analysis means we need to use two separate datasets. To train our emotion and sentiment classifiers first, we need previously labeled tweets and headlines with one of the five emotions, and second, we need sentiment pre-labeled tweets and headlines, either positive, negative, or neutral.

"Corpus linguistics is defined as a methodology for studying the use of language" (Stubbs 2006). According to Bowker & Pearson (see 2002, pp. 12), a field "specific corpus is one that focuses on a particular aspect of language" of a particular subject field. Obtaining a field-specific corpus can be a challenge, and as it is presented in the literature review Section 2.4, a financial training dataset labeled with emotions is almost non-existing. Many studies with promising results on sentiment analysis for stock price changes were conducted using their specific corpora and approach. Kordonis *et al.* (2016) obtained an average accuracy of 87% in predicting future stock price movement using positive and negative emoticons. Pagolu *et al.* (2016) is another example where a promising result of 70% was obtained using a human-annotated training corpus. However, when we speak about fine-grained emotion classification in correlation with stock price movements, limited research has been done. Liu (2017) in her research concludes that using general tweets as previously labeled training data can lead to poor results when the classifier is used on tweets with financial jargon.

That being said, we manually annotate our training data for both emotion and sentiment classifiers. To obtain our training data, we use the same sources and approaches for the period between January 1st and March 31st, 2019. Manually labeling text messages is labor-intensive, error-prone, and can lead to semantic ambiguity. However, to our knowledge, an open-source of labeled financial corpus with one of the Ekman six basic emotions is not available (Klinger *et al.* 2018). Identical datasets were given to three individuals with financial, trading and academic background, however without any semantic or psychological experience to label the tweets and headlines for positive, negative and neutral sentiment or else in five emotions: Anger, Fear, Joy, Surprise, Sadness and Neutral (Strapparava & Mihalcea 2007; Roberts *et al.* 2012). The highest scoring emotion or sentiment across all individuals was chosen in the training dataset. Table 3.5 shows the moderately balanced and similar distribution of sentiment labeled tweets and headlines within both datasets.

Table 3.5: Distribution of Sentiment labeled tweets and headlines

Tweets	
Sentiment	**Number**
Negative	1,729
Neutral	333
Positive	2,059

Headlines	
Sentiment	**Number**
Negative	2,009
Neutral	265
Positive	2,316

Fear was discarded from the emotions since, in the tweets dataset, it represented only 2.58% tweets and 2.37% in the headlines dataset. This behavior is intuitive since all our sources are professional newspapers, and they tend not to propagate emotions of fear or hatred and instead tend to be more neutral and informative as possible. Therefore, we decided to remove Fear considering the number of samples is low, and the classifier will always miss classifying in other emotions resulting in a lower accuracy score. Table 3.6 shows examples and the total number of tweets labeled in one of the emotions, where within both datasets, we have a relatively imbalanced distribution of classes. However, since both emotions, Anger and Sadness represent negative sentiment and Surprise can be either positive or negative, the datasets reflect very similar sentiment comparing with Table 3.5.

Table 3.6: Distribution and examples of labeled Tweets

Tweet	**Emotion**	**Tweets**
Supreme Court sends consumers suing Google over privacy issues back to court https://t.co/9mekek7zKb	Anger	532
Amazon is hiring 3,000 remote workers in 18 states- Here is what the company is looking for. via @CNBCMakeIt https://t.co/yLualInl8Z	Joy	1,826
Here is everything Apple isn't telling you about its new credit card https://t.co/obWpSFbUgH https://t.co/hxrGgBxWGo	Surprise	1,171
Saudis hacked Amazon CEO phone, says Bezos security chief https://t.co/fh9rVOEYdl	Sadness	399
Microsoft CEO on the 3 qualities that make a great leader via @CNBCMakeIt https://t.co/nt7hf4fxrX	Neutral	293
Headlines	**Emotion**	**Headlines**
San Bernardino shooting lawsuits v Facebook, Google, Twitter dismissed	Anger	680
Microsoft pledges $500 million to tackle Seattle housing crisis	Joy	1,971
Apple largest holders could lose $10B if stock opens down 8%	Surprise	1,156
Amazon CEO Jeff Bezos and MacKenzie Bezos are getting a divorce	Sadness	438
10 new books Amazon editors say are must-reads this January	Neutral	345

3.2 Text Pre-processing

In this section, we will describe several text processing techniques we use and why this is a pivotal step in developing our classification pipeline. Building a prediction model should be based on user data in order to produce meaningful results. To choose correct methods, knowing your data type plays a significant role. Our data sources are well-established news outlets. The text they produce, whether on Twitter or in the headlines, is well written and barely contains any noise.

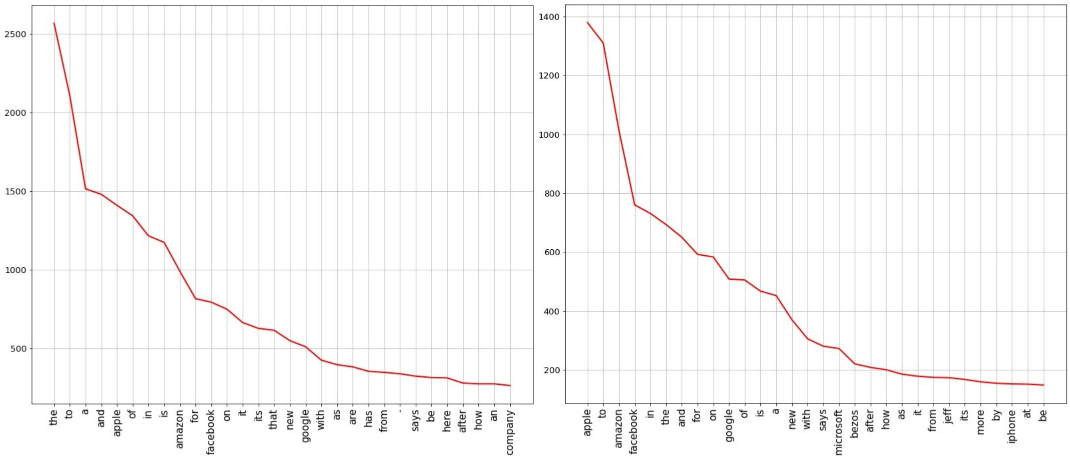
Our pre-processing method is divided into two stages. First is the preparation stage consisting of Lowercasing and Noise removal, removing Stop Words, and Tokenization. To turn our text into numeric vectors, we use encoding techniques such as BoW and Term Frequency, Inverse Document Frequency (TF-IDF), which is our second stage, vectorization.

1. **Lowercasing and Noise removal** is highly domain-dependent. The majority of NLP is based on feature extraction, and this step will improve the features we obtain by minimizing information loss. News Headlines are significantly less noisy than tweets. Thus, we only remove numbers, convert to lowercase, and apply contractions. Regarding tweets, we remove all the special characters, usernames, tickers, numbers, hyperlinks, hashtags, punctuations, convert to lowercase, and apply contractions.

2. **Stop Words** are the most frequently used words in any language because they form the functional skeleton of any sentence (e.g., 'the,' 'and,' 'is,' 'for'), yet they do not carry a lot of informational content. Removing stop words without controlling them can become critical since it can affect the context of the sentence and the sentiment polarity. Since we are using BoW i.e., CountVectorizer and TF-IDF methods, both work on frequency and counting words, removing stop words with little meaning to the text is beneficial step because it lowers the dimensional space. Nevertheless, we need to be careful of which stop words are being removed and what consequences they can have. Such examples are the negation words (e.g., 'not,' 'nor,' 'never'), where if we remove them, the semantic meaning of a sentence can significantly change. In Figure 3.1 we visualize 30 most frequent words across both training sets.

Figure 3.1: Distribution of most frequent words

(a) Tweets - 30 most frequent words (b) Headlines - 30 most frequent words



What is noticeable across both datasets is that Apple, Amazon, and Facebook are mentioned twice more than Google and Microsoft. Jeff Bezos, the founder of Amazon is the only person who is part of the 30 most frequent words in the headlines dataset but not in the tweets. Nevertheless, in both datasets same stop words are repeating, thus, we remove them.

3.Tokenization is the process of parsing a string and segmenting it by spaces into individual words, which we call them tokens. Extracting tokens is important because we can count the number and frequency of words present in a particular document and analyze the meaning of those words. Table 3.7 shows an example of original tweet and after processing.

Table 3.7: Example of Tweet processing

Before Processing	#BREAKING Google fined $1.7 billion over a third breach of EU antitrust rules in as many years https://t.co/E98JJ5FO1s.
Noise removal	breaking google fined billion over a third breach of eu antitrust rules in as many years
Stop Words removal & Tokenization	'breaking', 'fined', 'billion', 'over', 'third', 'breach', 'eu', 'antitrust', 'rules', 'many', 'years'

4.Feature Extraction or feature encoding is the second stage of our text pre-processing method. After normalizing the text, we need to transform it into vectors and quantify it into features which we can use in our machine learning algorithms (Goldberg 2017).

(a) **BoW** is a text representation in numbers that describes the appearance of the individual word within a document, i.e., each word represents a feature. Separating each unique word from every document and creating a vocabulary is the first step in BoW. Followed by assigning scores through counts and frequency of each word in every document, i.e., vector creation, where each dimension represents a different word. However, when working with large corpus increasing the number of words leads to increasing vocabulary, consequently the vector representations. This caveat can be mitigated by creating a vocabulary of paired words as one feature. Pairing a words is a technique called *"n-gram"*, where each word is called *"gram"* and *"n"* is the number of words. For example, a 2-gram (more commonly referred to as bigram) is a two-word sequence of words. Although this approach can bring more meaning from the document, increasing n can lead to over-fitting as the dimensionality of the vectors increases exponentially (Jurafsky 2000).

Another problem in the scoring word frequency is despite the fact we removed the frequent stop words with less valuable information. There can still exist words that are frequently appearing and do not bring much of a value, hence to avoid this, we apply TF-IDF (Schütze *et al.* 2008).

(b) **TF-IDF** is one approach to how we can rescale the appearance of word frequency in all documents. *Term Frequency* computes the score of a word in a particular document, whereas *Inverse Document Frequency* is computed as a logarithm of the number of the documents divided by the number of documents where the particular word appears. In other words, weight is assigned depending on how common the word is, hence the rare the word is, the more critical it becomes. This implies that IDF will be high on rare words and *vice versa*, while computing TF, all terms are considered equally important (Schütze *et al.* 2008; Wu *et al.* 2008).

TF-IDF formula is stated as:

$$tfidf_{t,d,D} = tf_{t,d} \cdot idf_{t,D} \tag{3.1}$$

Where **TF** is:

$$tf_{t,d} = log(1 + freq(t, d)) \tag{3.2}$$

And **IDF** is:

$$idf_{t,D} = log(\frac{N}{count(d \in D : t \in d)}) \tag{3.3}$$

Where t denotes the terms (words), d denotes each document, D denotes the collection of documents.

3.3 Emotion and Sentiment Classifiers

There are generally two main approaches in NLP for sentiment extraction. The first is to use a rule-based approach, where we use lexicon, which is a pre-recorded database of positive and negative opinions and expression words (Ding *et al.* 2008; Taboada *et al.* 2011). The other approach is to use machine learning by extracting features. We can model this as a classification problem in which an algorithm is given some text and returns a matching polarity during the training process. The algorithm learns to associate an input with an output using an optimization strategy (Pang *et al.* 2008). Considering the nature of our data is Sentence-Level, and we want to predict stock price movement using sentiment and fine-grained emotion derivation, most of the studies confirm we should use a supervised machine learning approach for both of our methods (Kim & Hovy 2007; Feldman 2013).

Sentiment analysis with a great extent of text polarity has been widely researched for what many studies are accessible. On the contrary, studies that are exploring the categorization of human emotions at finer-grained levels are not present to such an extent. In order for us to categorize our headlines and tweets in five emotions (Anger, Fear, Joy, Surprise, and Sadness) or sentiment polarity (positive, negative, or neutral), we will develop our classifiers. This section condenses how we train and derive sentiment and emotions using supervised learning with MultiClass Classification (MCC) approaches and afterward depict and assess the performance of our classifiers.

3.3.1 MultiClass Classifiers

There are numerous types of classification tasks that you can experience in machine learning. However, when we need text classification into certain emotions or sentiment, it comprises two major classifications: Binary and MCC. Binary classification algorithms definitely can be adjusted and used for multiclass problems such as classification of movie reviews to "positive" or "negative". One of these examples is the so-called One-vs-Rest and One-vs-One methods, which are best utilized by using Logistic Regression and SVM both binary classification algorithms. Nevertheless, our goal is to classify the headlines and tweets into five emotions and sentiment, hence despite the classifiers mentioned above, we will use MCC algorithms.

MCC is a classification task with more than two class labels, and it assumes that each sample is assigned to labels with a set of input examples, where there are more than two classes. MCC is one of the most common supervised learning tasks after regression. In classification, the main idea is that we have training examples separated into K classes on which we train a machine learning model to further predict unseen data in one of the training K classes (Ahuja & Yadav 2012; Robert 2014).

Choosing the right classifier is always correlated with the problem we want to solve. Wang & Manning (2012) in their paper are evaluating many variants of Naïve Bayes and SVM where their results are outperforming other researches on sentiment analysis. However, since we need finer-grained emotion and sentiment classification, we will use the following classifiers: Naïve Bayes, Random Forest, Logistic Regression, SVM, and K Nearest Neighbors (KNN), which are summarized below.

3.3.1.1 Naïve Bayes

In-text classification, the Naïve Bayes model is widely used. It is a probabilistic model that is mainly used for classification tasks. The Naïve Bayes classifier is based on the Bayes Theorem, where the joint probability of membership is divided into classes of conditional probabilities. There are three types of Naïve Bayes classifier: Multinomial, Bernoulli and Gaussian. In our thesis, we will use Multinomial since it fits best for document classification tasks. It makes a naive assumption that all features in the model, which are the frequency of the words present in a document, are mutually independent. Even though this is not correct, for us is a good estimation to obtain the necessary results. Naïve Bayes formula stated in equation 3.4 calculates the posterior probability of an outcome given another one, using the inverse of that relationship (Kuhn *et al.* 2013).

$$P\left(c|x\right) = \frac{P\left(x|c\right) \cdot P\left(c\right)}{P\left(x\right)} \tag{3.4}$$

Where c and x are events and $P(x) \neq 0$. $P(c|x)$ is the posterior probability of class (c, target) given predictor (x, attributes). $P(c)$ is the prior probability of class. $P(x|c)$ is the likelihood which is the probability of predictor given class. $P(x)$ is the prior probability of predictor (Friedman *et al.* 1997; Brownlee 2020; Sayad 2020).

3.3.1.2 Random Forest

Random Forest is an algorithm that can perform regression and classification tasks using a collection of decision trees and a statistical technique called Bagging or Bootstrap Aggregation. Each decision tree is constructed of branches, nodes, and leaves, representing a potential class. Hence, this algorithm is just an extension of the decision tree algorithm.

This algorithm, instead of taking all the data for each decision tree it takes a random subset of the training data for each random decision tree as mentioned before this technique is called bagging, or bootstrap aggregating. By observation and combining the results from each decision tree, we make our decision by vote (classification) or taking the average (regression). If we observe the predictions individually, they might not be so accurate, however averaging several random trees will significantly lower the risk of over-fitting (Breiman 2001; Rokach & Maimon 2008).

3.3.1.3 Support Vector Machine

The basic principle behind SVM classifier is to separate the dataset in two classes and maximize the margin as optimal as possible using decision boundary or hyperplanes. Margins are the (perpendicular) distances between the hyperplane and the closest data points. The data points closest to the hyperplane are called support vectors.

Hyperplane is an (n-1) dimensional subspace for an n-dimensional space. If we visualize this in two-dimensions, the hyperplane will represent a line where the line can completely separate the input points.

Where the Hyperplane is:

$$\beta_0 + (\beta_1 \cdot x_1) + (\beta_2 \cdot x_2) + \cdots + (\beta_n \cdot x_n) = 0 \tag{3.5}$$

For two-dimensional space is:

$$\beta_0 + (\beta_1 \cdot x_1) + (\beta_2 \cdot x_2) = 0 \tag{3.6}$$

Where (β_1 and β_2) are coefficients which are determining the slope of the line and together with the intercept (β_0) are found by the learning algorithm, x_1 and x_2 are the two input variables (Cortes & Vapnik 1995).

3.3.1.4 Logistic Regression

Linear regression, as one of the binary classification models, investigates a linear relationship between the input variables and a target variable. A linear prediction function in equation 3.7 is well explained by Murphy (2012) were looking from the right-hand side we have x as a vector of our training samples, θ alludes to the model parameters, and y will be our vector for the labels we want to predict.

$$y = h_\theta(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T x \tag{3.7}$$

However, linear regression is not the most suitable for classification problems, where we are interested in the probability of an outcome occurring. Range of a probability is from [0,1], where 1 denotes something certain to happen, and 0 denotes unlikely to happen. The problem with linear regression is that an absolute number can range outside [0,1], which brings logistic regression into the picture. The sigmoid function depict in equation 3.8 solves the problem by mapping the output in range strictly between [0,1] (Ng 2000).

$$P(y = 1|x) = h_\theta(x) = \frac{1}{1 + e^{-\theta^t x}} \tag{3.8}$$

Where $P(y = 1|x; \theta) = h_\theta(x)$ will give us the probability that x is a positive sample, and $P(y = 0|x; \theta) = 1 - h_\theta(x)$ will give us the probability that x is a negative sample (Menard 2002; Murphy 2012).

3.3.1.5 K Nearest Neighbors

KNN classifier makes an assumption that similar data points are near to each other and classifies the new data points based on a similarity measure (e.g., distance functions) only valid for continuous variables. The Euclidean equation 3.9 is the most used equation explaining the distance function:

$$\sqrt{\sum_{i=1}^{k}(x_1 - y_1)^2} \tag{3.9}$$

A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by one of the distance functions (Madhumathi & Rajan 2014). If $K = 1$, then the case is assigned to the class of most similar instances (neighbors). KNN is used for *regression* where the prediction is based on the mean or median of the K most common instances and *classification* where the output is calculated based on the class with highest frequency among the K most common instances (Altman 1992; Sayad 2020).

3.4 Models Training and Evaluation

After building our classification models next step is to train and evaluate their performance in predicting the outcome of observations that have not been used in training the models. To estimate our model's accuracy and error predictions, we will split our labeled datasets in train and test sets with ratio 70/30, where the models will learn based on the train test and afterward compare the predicted outcomes against know values in the test set. We will follow the methodology of the most commonly used assessment metrics for MCC models.

Average classification accuracy represents the proportion of correct predictions from all the observations. Tables 3.8 and 3.9 shows the median accuracy from all our classifiers. We can notice that LinearSVC produces the most accurate results with a small margin than Logistic Regression but outperforms the rest of the classifiers significantly on all our datasets. Because of the Sentence Level similarity, we experiment by combining both datasets for tweets and headlines as one. Figure 3.2 shows the distribution of words where most of the sentences are in a range between 10 and 25 words. This experiment improved the accuracy in the majority of our classifiers since there is more data to learn from. More precisely, it improved LinearSVC for an additional 5.6265% in the sentiment dataset and 11.4805% in the emotion dataset.

Table 3.8: Models Accuracy for Sentiment

Tweets	
Model	**Accuracy**
KNN	0.466391
LSVC	**0.835479**
LR	0.814857
NB	0.755894
RF	0.549627

Headlines	
Model	**Accuracy**
KNN	0.695861
LSVC	**0.831373**
LR	0.792375
NB	0.754902
RF	0.513943

Combined	
Models	**Accuracy**
KNN	0.514291
LSVC	**0.891744**
LR	0.872228
NB	0.807365
RF	0.526460

Table 3.9: Models Accuracy for Emotions

Tweets	
Model	**Accuracy**
KNN	0.500363
LSVC	**0.711965**
LR	0.664165
NB	0.595727
RF	0.492114

Headlines	
Model	**Accuracy**
KNN	0.541612
LSVC	**0.707190**
LR	0.638344
NB	0.582571
RF	0.437908

Combined	
Models	**Accuracy**
KNN	0.517165
LSVC	**0.826770**
LR	0.741362
NB	0.649753
RF	0.459075

Figure 3.2: Words count in combined datasets

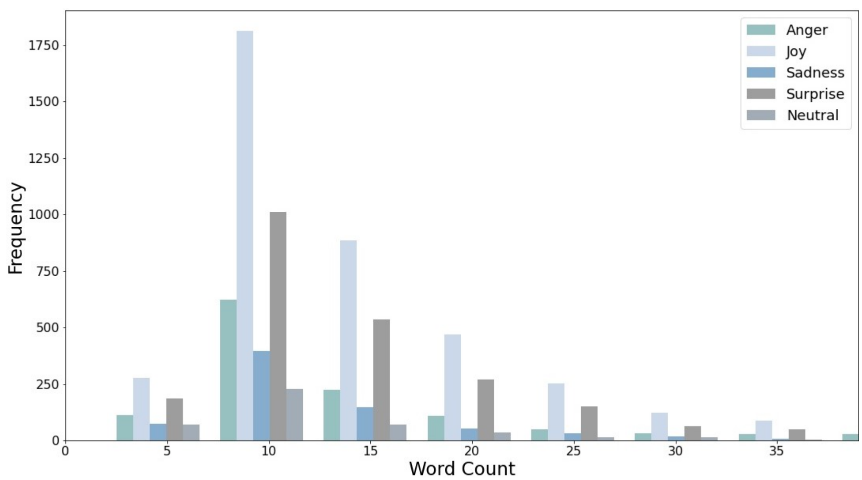
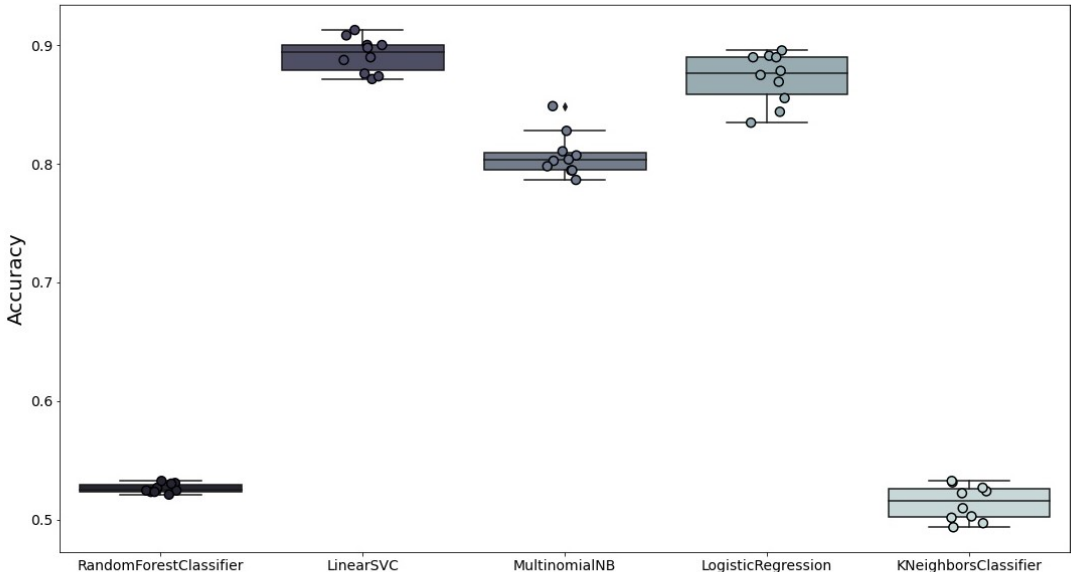
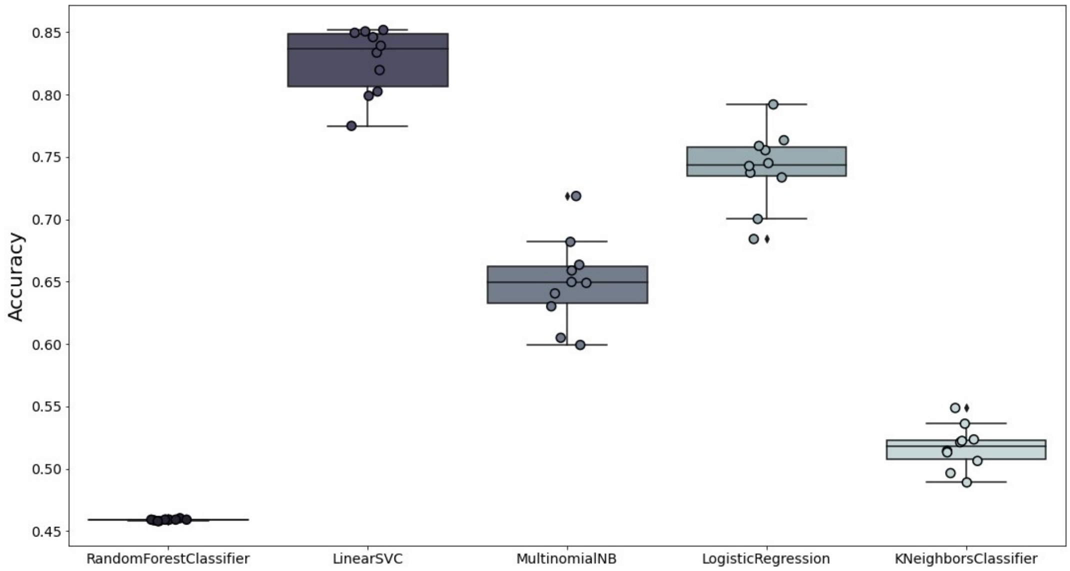


Figure 3.3: Classifiers Accuracy

(a) **Combined Sentiment**



(b) **Combined Emotion**



Precision, Recall, and F1 scores are three primary performance metrics alternative to using classification accuracy. They are essential since our datasets are imbalanced, and the majority of classes can overcome the minority classes, misleading into high accuracy scores.

Precision, equation 3.10 quantifies how many of the predicted positive are actually positive. Through precision we can determine when we have high number of false positives.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{3.10}$$

Recall, equation 3.11 determines the percentage of actual positives from all the positive predicted examples in the dataset. Recall is the metric to determine when there is a high cost of false negative.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3.11}$$

F1 equation 3.12 provides a single score of precision and recall and is needed when we want to find a balance between precision and recall, since artificially is possible to build classifiers with high precision and recall scores.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{3.12}$$

Table 3.10 and 3.11 summarizes the results for precision, recall, and F1 for all our classifiers.

Table 3.10: Classifiers metrics comparison for Sentiment

Model	Precision	Recall	F1
KNeighborsClassifier	0.737	0.477	0.45
LinearSVC	**0.89**	**0.837**	**0.857**
LogisticRegression	0.9	0.697	0.747
MultinomialNB	0.887	0.59	0.62
RandomForestClassifier	0.51	0.347	0.257

Table 3.11: Classifiers metrics comparison for Emotion

Model	Precision	Recall	F1
KNeighborsClassifier	0.678	0.31	0.32
LinearSVC	**0.784**	**0.698**	**0.734**
LogisticRegression	0.83	0.546	0.61
MultinomialNB	0.794	0.402	0.436
RandomForestClassifier	0.092	0.2	0.126

Finally, Tables 3.12 and 3.13 shows classification reports for each class in emotion and sentiment for our best performing classifier LinearSVC.

Table 3.12: Classification report for LinearSVC for Sentiment

Sentiment	**Precision**	**Recall**	**F1**
Positive	0.90	0.89	0.89
Negative	0.89	0.70	0.78
Neutral	0.89	0.92	0.90

Table 3.13: Classification report for LinearSVC for Emotion

Emotion	**Precision**	**Recall**	**F1**
Joy	0.76	0.65	0.70
Neutral	0.80	0.90	0.85
Surprise	0.77	0.51	0.61
Anger	0.76	0.62	0.69
Sadness	0.83	0.81	0.82

Expectantly looking in the metrics comparison, all the classifiers have better results for sentiment comparing to emotion since there are only three classes. Nevertheless, LinearSVC outperforms all other classifiers, which was also confirmed by the average classification accuracy. Therefore, throughout this thesis, we will use LinearSVC as our primary classifier in all further classification problems.

Confusion matrix, is a 2x2 matrix of four parameters: true positives (TP), true negatives (TN), false negatives (FN) and false positives (FP), where for TP we predicted yes and is actually yes, for TN we predicted no, and it is actually no, for FN we predicted yes but it is actually no, and for FP we predicted no, but it is actually yes. This way, we count the correct and incorrect predictions by evaluating how our model is being confused. It gives us insights into the errors being made and, more importantly, summarizes the type of error that is being made.

Figures 3.4 and 3.5 show the confusion matrices from our best model SVM for both sentiment and emotions. It is noticeable that in both cases, most of the predictions are on the diagonal (actual = predicted). However, there are misclassifications.

Figure 3.4: Confusion Matrix for Sentiment

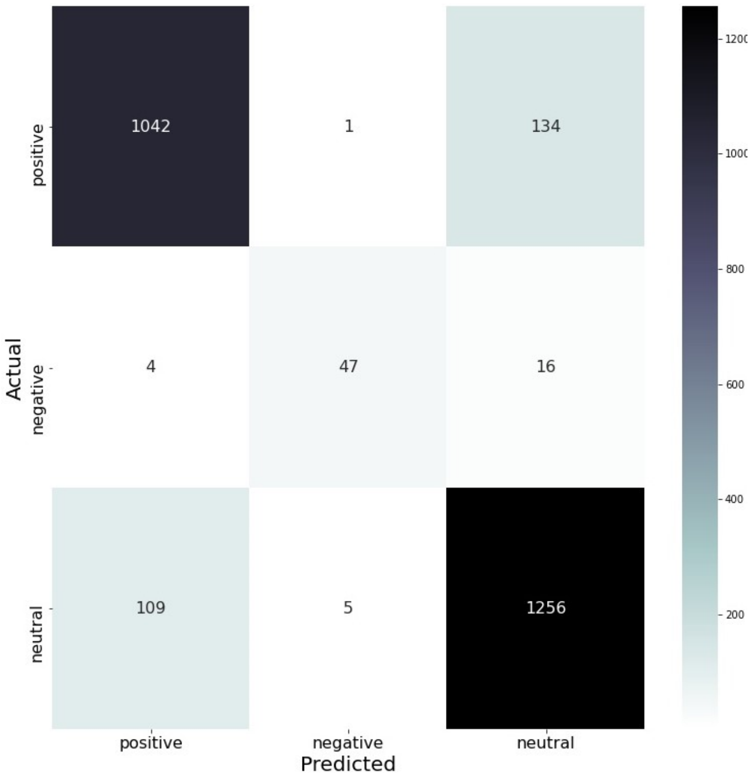
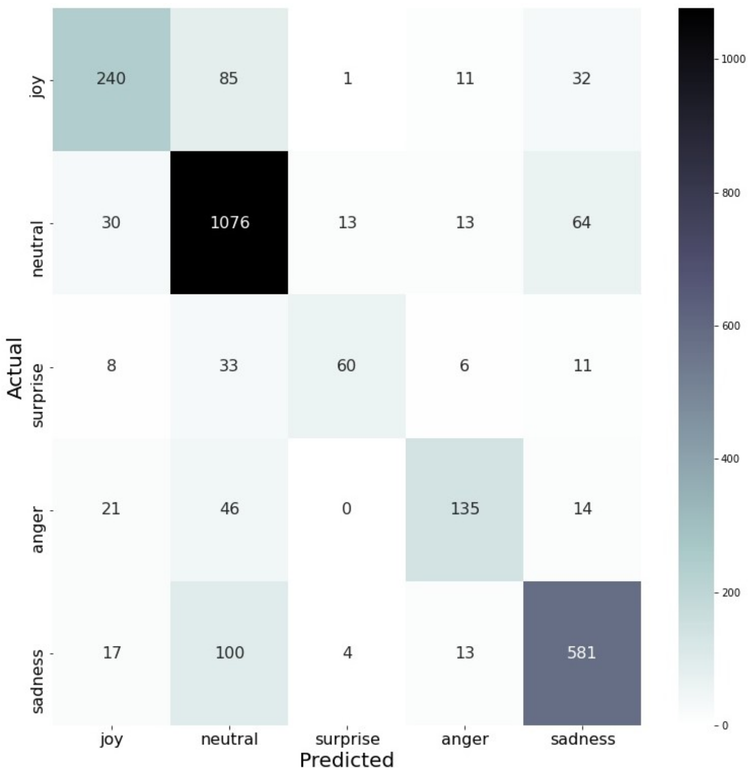


Figure 3.5: Confusion Matrix for Emotion



In Table 3.14 we can see that sentiment is relatively well categorized, only 9.844% from the test sample was misclassified. Very positive is the fact that our classifier does not misclassify from negative to positive and *vice versa*, which can be critical. On the contrary, the majority of the misclassifications are coming from neutral to else and *vice versa*. Also, neutral is the least represented emotion in the dataset, hence enhancing our dataset with more neutral examples can potentially improve the classifier accuracy.

Regarding emotions, our model makes more mistakes, 19,650% from the test samples were misclassified. This number is relatively higher comparing to 9.844% in sentiment. However, we need to consider that there are more classes, and assigning a correct label to emotion can be a challenge even for humans. For example, distinguishing between Sadness and Anger or Joy and Surprises can be difficult. The biggest misclassifications are coming from Sadness to Neutral, Joy to Neutral, and Neutral to Sadness. This is again positive sign since the misclassifications for stronger emotions such as Joy to Sadness or Anger to Joy are moderately low. As stated above, adding more neutral examples in the emotion dataset can lead to a higher accuracy score for our classifier.

Table 3.14: Misclassified Predictions

Sentiment		
Actual	**Predicted**	**Missclassified**
neutral	positive	109
positive	neutral	134
negative	neutral	16
Tweets		
Actual	**Predicted**	**Missclassified**
anger	joy	21
anger	neutral	46
anger	sadness	14
joy	neutral	85
joy	anger	11
joy	sadness	32
neutral	joy	30
neutral	surprise	13
neutral	anger	13
neutral	sadness	64
sadness	joy	17
sadness	neutral	100
sadness	anger	13
surprise	joy	8
surprise	neutral	33
surprise	anger	6
surprise	sadness	11

Chapter 4

Predictive Analytics on Unseen Data

This chapter will analyze how our leading classifier LinearSVC labeled both unseen headlines and tweets into emotions and sentiment. Furthermore, we will examine whether there is a correlation between emotions or sentiment and future stock price returns for the five Big Tech companies.

4.1 Classification Analysis

In Figure 4.1, we can observe the distribution of emotion and sentiment for headlines across all five companies. The first (Emotion distribution) graph explains the percentages distribution for each emotion and company. Noticeable is that Joy dominates across all four companies except Facebook, with only 23%. Microsoft has the highest Joy ratio of 63% comparing to the rest of the companies. This can be because in 2019 Microsoft had two very positively perceived conferences, Microsoft Inspire in July and Microsoft Ignite in November, where a plethora of new products and applications were announced. The second most dominant emotion is Surprise, where Apple, with 28%, has the highest proportion comparing to the rest of the companies. For Apple 2019 was a year full of announcements and some indeed surprising. Worth mentioning is the June announcement of Sabih Khan as senior vice president of operations, followed by the acquisition of Intel's 5G modem business in July, and lastly, in September, Apple confirmed its redesigned Mac Pro would be manufactured in Texas (Apple 2020). For Sadness and Anger, both expressing the negative sentiment, Facebook has the biggest portion of 20% and 14%, respectively, compared to the rest of the companies. The percentages are no surprise since Facebook had plenty of scandals in the preceding years, and 2019 was no different. In May, Facebook co-founder Chris Hughes publicly said it is time for a Facebook break-up, in July, Facebook was charged with the largest ever fine of $ 5 billion by the Federal Trade Commission, last in December 267 million phone numbers were exposed (Sanders 2020).

Figure 4.1: Distribution of Emotions and Sentiment for Headlines

(a) Emotion distribution (b) Sentiment distribution

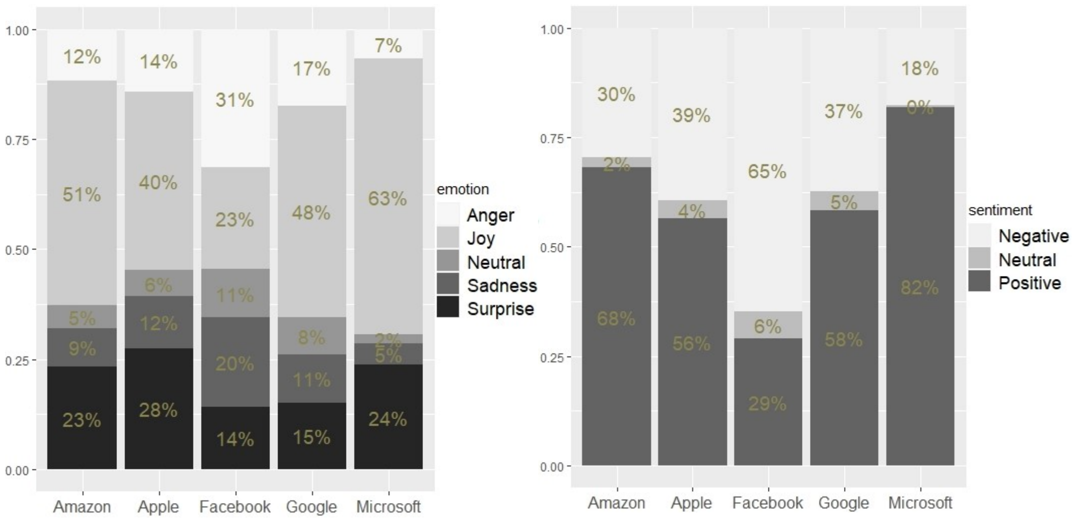
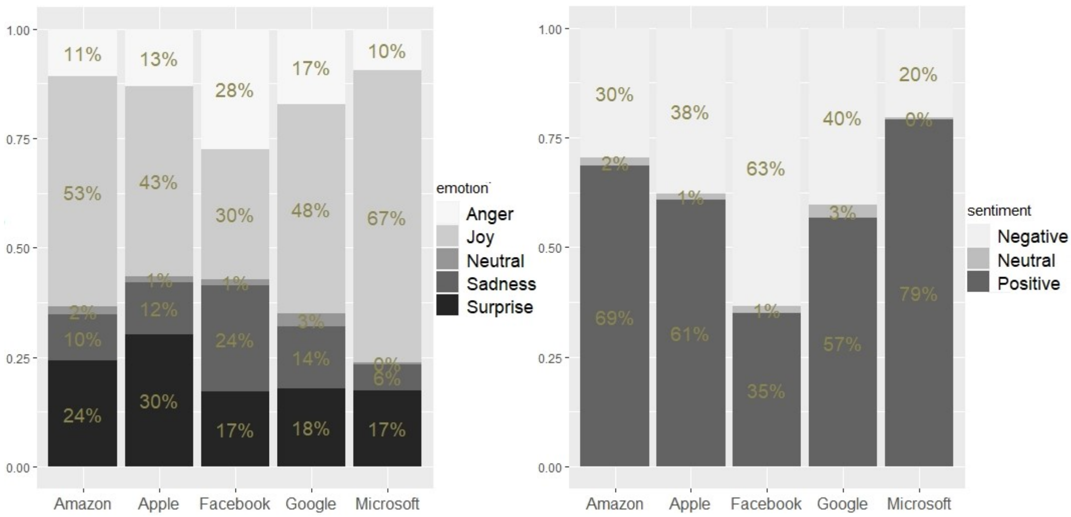


Figure 4.1, the second (Sentiment distribution) graph explains the percentages distribution for each sentiment and company. The first observation we can see is that again Facebook with 65% negative sentiment is greater than the rest of the companies. This is very straightforward because, compared to the Emotion graph Facebook has the biggest portion of negative emotions (Anger and Sadness). We can draw a similar conclusion by observing Microsoft with 82% of positive sentiment very similar to Joy in the emotion graph. This pattern between emotions and sentiment is present in the rest of the companies. Thus, we can conclude that our classifier performed well and did not misclassify, for example, joy as negative or anger as positive emotion or sentiment, respectively.

In Figure 4.2 we can observe the distribution of emotion and sentiment for tweets across all five companies. Comparing both emotion and sentiment graphs for headlines, the distribution is almost identical, with Neutral emotion and sentiment almost being not present for tweets. Thus, we can draw two conclusions. First, the content shared in the headlines and tweets from the same source is just about the same. However, the content shared on twitter is less neutral in order to attract more attention, which will lead to sharing the tweets and gaining more engagements.

Figure 4.2: Distribution of Emotions and Sentiment for Tweets

(a) Emotion distribution (b) Sentiment distribution



Our classifier has made some mistakes, and Table 4.1 and 4.2 shows some examples of wrongly classified headlines and tweets after we subjectively evaluated the classifier. In Table 4.1 obviously, both headlines express joy and anger, but they are classified as anger and joy. Furthermore, for the first headline, the sentiment should be positive, however, it is classified as negative, and the second headline is classified as positive, yet it clearly express a negative sentiment.

Table 4.1: Missclasified Examples of Headlines

Headlines	**Emotion**	**Sentiment**
Apple cuts iPhone prices in China along with Macs, iPads, and AirPods	Anger	Negative
Facebook still not doing enough to prevent ethnic hate in Myanmar, U.N. investigator says	Joy	Positive

Similarly, in Table 4.2, both tweets express joy and sadness, but they are classified as anger and joy. The same goes for sentiment where the first tweet should not have been classified as negative but rather to positive, and the second tweet is classified as positive, however, clearly should have been classified as negative.

Table 4.2: Missclasified Examples of Tweets

Tweets	**Emotion**	**Sentiment**
Elizabeth Warren praises Mark Zuckerberg's call for tech regulation after she campaigned for Facebook's breakup	Anger	Negative
After an employee backlash, Google has cancelled its AI ethics board a little more than a week after announcing it	Joy	Positive

4.2 Correlation Analysis

The majority of previous research on the financial domain focuses on aggregated or simple classification into positive or negative sentiment (Smailović *et al.* 2014; Li *et al.* 2014a). However, the number of studies examining the influence of human emotions on the stock market is somehow limited (Mittal & Goel 2012; Bollen *et al.* 2011). Both streams of research do not adequately address our research question of whether emotions and sentiment in headlines and tweets can predict a market movement. Zhang *et al.* (2011) in his work obtain weak negative correlations between emotions and significant large-cap index prices. However, he did not use specific financial, rather general tweets.

Before attempting to predict stock price movements by extracting emotions and sentiment in text, we will follow a similar methodology to Liu (2017) work where she performed linear correlation between the distribution of emotions on each day for financial tweets tagged by 'cashtags' and NASDAQ-100 return on the next day.

To obtain our market data Yahoo Finance was used for the period between April 1st and December 31st, 2019. Daily adjusted closing prices were downloaded for the top five S&P 500 companies: Facebook, Apple, Amazon, Microsoft, and Google(Alphabet). Stock markets are closed on weekends, and we cannot obtain prices for these days, while tweets and headlines were obtained for every day. To cover this gap in days we will follow Mittal & Goel (see 2012, pp. 02) approach, using a concave function. In theory, this approach is justified by the fact that stock prices generally have a concave function. However, a sudden rise or drop is possible. Thus we manually investigated whether there are many cases of this kind in our datasets, and we did not find any. That being said, if the price on the day x and y is given and there is a missing price in between we approximate the missing data by estimating the first day after x to be $(y + x)/2$, following the same methodology till all gaps are filled.

We will use Pearson correlation coefficient, equation 4.1 to find strength between pairs of daily emotions or sentiment percentages and future stock price returns on the day $t + 1$ and $t + 2$ for each company separately.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4.1}$$

Where r is measuring the strength between two variables and can span between -1 perfect positive linear correlation, 0 no linear correlation and 1 perfect positive linear correlation (Benesty *et al.* 2009). In our case x represents daily emotion percentages and y is representation of stock return for $t + 1$ and $t + 2$.

Figure 4.3 shows the average correlation for headlines between emotion or sentiment percentages and stock return on day $t + 1$ and $t + 2$ for each company separately. We can observe that none of the emotions or sentiment have statistically significant correlations with the five companies' returns for a day $t + 1$ and $t + 2$, where all coefficients are below 20%. However, interesting observation is on day $t+2$, where Amazon has the highest correlation coefficient with $r = .17$ and $p - value = .004675$ for positive sentiment and second-highest $r = .14$ and $p - value = .016792$ for emotion joy. This is not such a surprise since Amazon has one of the highest percentages of headlines classified as joy and positive sentiment. On the contrary, Facebook among all companies has the lowest percentage of headlines classified in joy and positive sentiment, however after Amazon has second-highest correlation coefficient with $r = .14$ and $p - value = .020507$ for positive sentiment and highest correlation coefficient with $r = .18$ and $p - value = .002951$ for emotion joy. This can be because Facebook had a three-year string of scandals starting with fake news problems in 2016, Russian election interference in 2017 then followed by the Cambridge Analytica scandal that broke in March 2018. At the beginning of April 2019, Senator Elizabeth Warren, a front-runner presidential candidate at that time, published a blog post calling for breaking up of several Big Tech companies, specifically accusing Facebook of acquiring Instagram and WhatsApp to limit competition (Sanders 2020). However, Facebook announced strong Year-over-Year second and third quarters with total revenue growth of 28% and 29%. This could result in investors trading more when there is periodically positive news on Facebook, thus increasing the price. Nevertheless, since both companies have a higher correlation coefficient for a day $t+2$, it implies that positive news has a bigger effect on the second day when more investors are being familiarized with the positive news.

Figure 4.3: Correlation Between Stock Returns and Headlines

(a) Headlines t+1

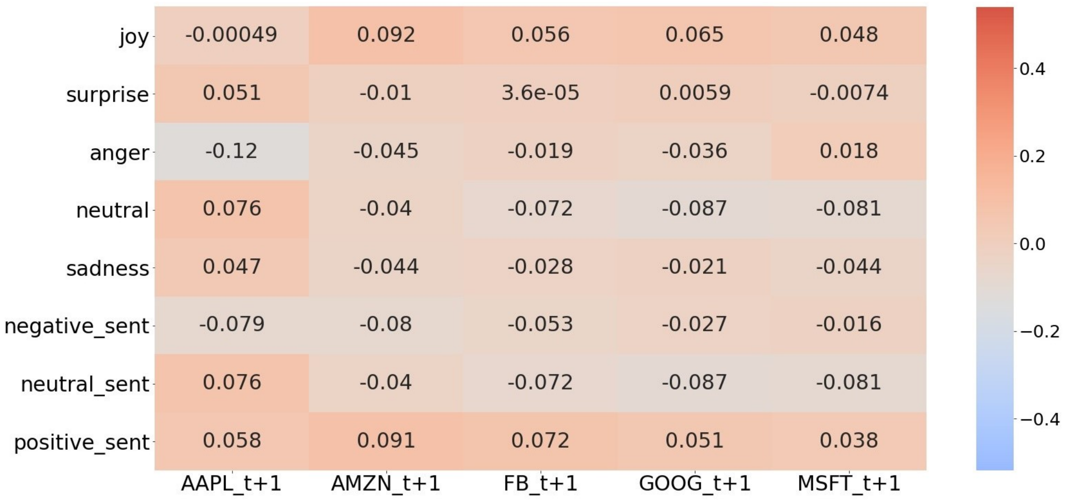
(b) Headlines t+2

Figure 4.4 shows average correlation for tweets between emotion or sentiment percentages and stock return on day $t+1$ and $t+2$ for each company separately. Again we can see that none of the emotions or sentiment has statis-

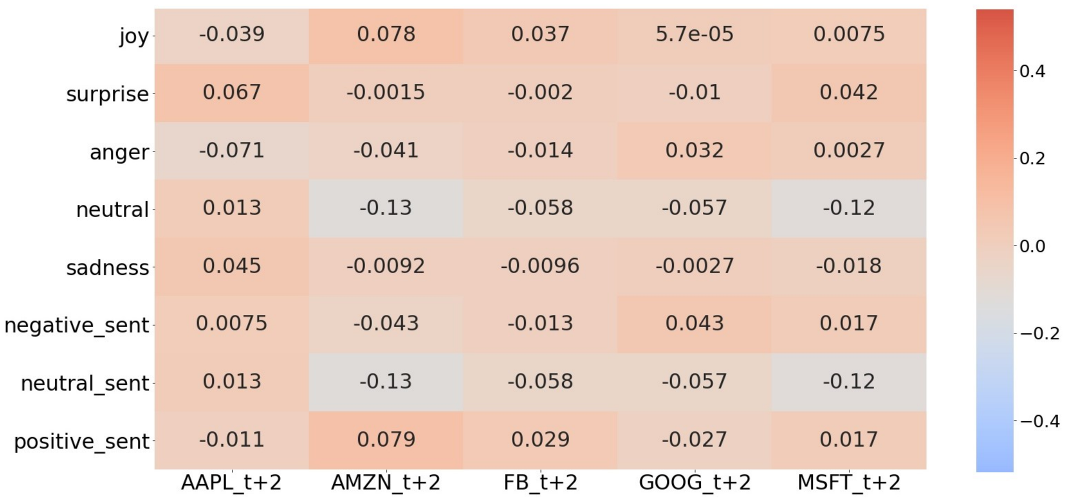
tically significant correlations with the returns for a day $t+1$ and $t+2$, where now all coefficients are below 10%. However, for both Amazon and Microsoft at day $t+2$, we have a negative correlation for neutral in both emotions and sentiment. Based on AdFontesMedia (2020), a web page for reliability and bias of news, all our newspaper sources are reporting neutral or balanced bias of the news. Therefore it is not unlikely that these newspapers have weaker neutrality when publishing news on Twitter. This confirms our previous conclusion when we analyzed both Figures 4.3 and 4.4.

Figure 4.4: Correlation Between Stock Returns and Tweets

(a) Tweets t+1



(b) Tweets t+2



Chapter 5

Stock Market Prediction

In Chapter 4, we did not find any significant correlation between automatically classified emotions or sentiment and future stock price returns. Understanding emotions in text, which can have more emotions, is not an easy task. Manually labeling emotions is labor-intensive, and humans are sometimes error-prone. Therefore, one of the reasons we did not find correlation can be imperfectly annotated training data from which the classifiers were learning. Nevertheless, our corpora are domain-specific, referring to the top five S&P 500 companies. In this chapter, we propose creating buy and hold value investing strategy following a similar methodology proposed by Noah Mukhtar & Chandra (2020) with extracting sentiment polarity using VADER and various machine learning algorithms.

5.1 VADER

VADER is a lexicon and rule-based sentiment analysis tool created by Hutto & Gilbert (see 2014, pp. 01). In their study, they use qualitative and quantitative methods to construct a list of lexical features, specifically attuned to the sentiment expressed in social media. They compare VADER[1], among other sentiment classifiers on various product reviews on Amazon, New York Times editorials and movie reviews, where VADER outperforms all benchmark classifiers, even human raters with an F1 classification accuracy of 0.96%.

We see using VADER as a suitable tool for our trading strategy since our

[1]https://github.com/cjhutto/vaderSentiment#citation-information

datasets are tweets and headlines, both sentence-level types of documents and VADER has proven as an excellent performer to this particular type of short documents. VADER is intelligent enough to understand the polarity and intensity in emotion, and it does not require text pre-processing nor training data to learn from. Also, it understands the emphasis of capitalization and punctuation such as "HAPPY!!!" to be the positive sentiment with increased magnitude and the basic context in features such as "not happy", which is expressing negative sentiment. VADER's SentimentIntensityAnalyzer() function assigns scores of (-1) negative, (0) neutral, (1) positive and compound (by normalizing previous scores). Tables 5.1 and 5.2 show differences between classification and Vader polarity sentiment scores for same tweets and headlines.

Table 5.1: Comparing vader polarity and classification in Headlines

Headlines	**Emotion**	**Sentiment**	**VADER polarity**
Facebook denies it is to blame after Russian political advertising accusation	Joy	Positive	-0.7351
Goldman rival pulled out of Apple Card on fears it was money loser	Surprise	Positive	-0.7461
Watch Apple Watch users share how heart rate feature saved their lives	Anger	Negative	0.8481
Supreme Court to Consider Google Appeal of Oracle Win in Java Case	Neutral	Negative	0.8126

Table 5.2: Comparing vader polarity and classification in Tweets

Tweets	**Emotion**	**Sentiment**	**VADER polarity**
Facebook endured a punishing 2 years of political hell. 2020 will be even worse	Joy	Positive	-0.9062
Apple's week has gone from bad to worse and it points to even more potential problems down the road	Joy	Positive	-0.875
5 years ago, Google gave away a cloud computing project for free. Now people love it so much they're celebrating its anniversary in Spain	Sadness	Negative	0.9093
Facebook's new cryptocurrency system Libra will be counting on strong growth from emerging markets like India	Anger	Negative	0.8126

5.2 Creating the Investing Strategy

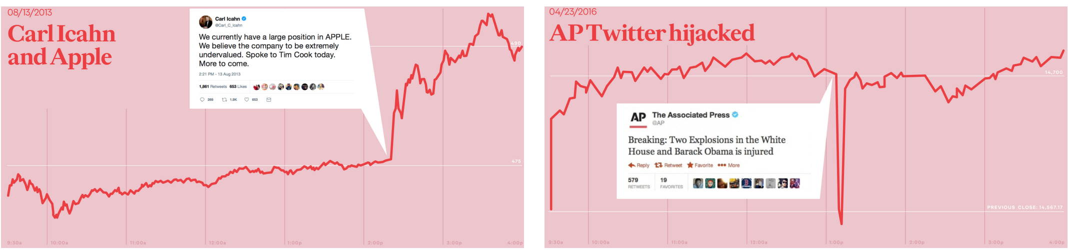
Our investment strategy is simple and based on a several steps. First, we obtain a sentiment score and then capture the volatility of each company stock. Next, the process of our strategy is creating a buy-hold signal, which is a crucial step to proceed using various machine learning classification algorithms, which should predict the future stock price. We will use the already available tweets and headlines with the historical prices for all Big Tech companies downloaded from Yahoo Finance, assuming that there is no after-hours price change. Fundamentally, the various machine learning algorithms will label a tweet or headline into positive, negative, or neutral by factoring the sentiment of the words, hence creating a signal of whether a stock price will move upwards or downwards.

5.3 Calculating Sentiment Score

Obtaining a sentiment score is our first step in creating the investing strategy. To create a sentiment score and associate it with the correct magnitude of importance and impact, we will use metrics such as followers for tweets and shares for headlines, which we already have in the metadata of our datasets. So far, these metrics proved crucial and excellent indicators to measure how broad auditorium reacts to certain tweets related to a specific company. Billionaire Carl Icahn on August 13th, 2013, will tweet that Apple stock is undervalued. Instantaneously, Apple stock gained 17$ billion in market cap. The second example is a hacked account of Associated Press announcing that two explosions in the White House injured President Obama, S&P 500 lost more than 130$ billion within minutes. Such examples are illustrated in Figure 5.1.

Figure 5.1: Example of Tweets Impacting the Stock Market

(a) Carl Icahn on Apple (b) Hacked Associated Press



Source: https://www.linkedin.com/in/adamkornblum/detail/recent-activity/posts/

To quantify the magnitude of impact from tweets or headlines and obtain the sentiment score, we multiply the number of followers and shares, respectively, with the compound polarity value. Table 5.3 and 5.4 shows sentiment score example for tweet and headline from all our sources.

Table 5.3: Sentiment score examples for Headlines

Headline	Source	Shares	Compound Score	Sentiment Score
Jeff Bezos would pay $9 billion a year in taxes under Sanders' plan	CNBC	699,436	-0.3027	-211,719
Distorted Videos of Nancy Pelosi Spread on Facebook, Helped by Trump	NY Times	145,386	-0.4019	-58,431
Microsoft's 4 day workweek led to 40% boost in productivity	Business Insider	1,138,265	0.4019	457,469
Exclusive: Amazon rolls out machines that pack orders and replace jobs	Reuters	50,249	0.8126	40,832

Table 5.4: Sentiment score examples for Tweets

Tweet	Source	Followers	Compound Score	Sentiment Score
U.S. allies urge Facebook not to encrypt any messages as they fight child abuse and terrorism	Reuters	22,047,219	-0.9081	-20,021,079
Facebook, Google and Apple are some of the most profitable companies the world has ever seen. But after several scandals on privacy, hate speech and more, regulators around the world are considering a crackdown.	Financial Times	6,628,434	-0.8502	-5,635,494
Mark Zuckerberg said new regulations are needed to protect society from harmful content, ensure election integrity, protect people's privacy, and to guarantee data portability in a departure from what he's said on regulation in the past.	CNBC	3,655,163	0.8885	3,247,612
$AMZN smashes earnings expectations with its first quarter earnings results making it one of the best performing big tech stocks this year (up 28% in 2019), and the third most valuable company in the world, behind Microsoft and Apple. Are we ready for #amazon takeover?!	Wall Street Journal	561,239	0.9019	506,181

5.4 Preparing the Market data

We begin preparing our market data by creating a time-series data of daily averages for values calculated in the sentiment score section. As already mentioned in Chapter 4, we fill the gap of missing market data for weekends and non-working days by applying the Mittal & Goel (2012) approach, using a concave function. Furthermore, we calculate the daily stock change using equation 5.1. Another essential step is standardizing the values so we can represent the volatility instead of the price change.

$$P_{i,t} = \frac{C_{i,t} - O_{i,t}}{O_{i,t}} \cdot 100 \tag{5.1}$$

Where $P_{i,t}$ is the change in stock price at day t for stock i. $C_{i,t}$ is the closing price on day t for stock i and $O_{i,t}$ is the opening price for stock i at day t.

The last step in preparing our market data is to artificially create a buy-hold signal depending on if an investor would have made a positive return on a certain day. In order to do this, we will use dataframe.shift(), a function representing shifting an index axis by desired number of periods in positive or negative direction. This function is beneficial for the manipulation of time-series data. Our desired number of periods will be three, and the daily stock price will become an input for the predicted stock price change. In other words if $P_{i,t}$ is a positive or negative value for day t it will signal 1 for Buy or -1 for Hold for day $t-3$. Using the sentiment score in correlation with a three-day lag of actual stock price change, our models can learn and predict the future movement of a stock price. Figure 5.2 illustrates an example of a prepared data frame with a three-day lag.

Figure 5.2: Example of Prepared Market Data

Open	Volume_of_stock	Adj_Close_stock	stock_val_change	stock_val_change_scaled	sentiment_score	stock_val_change_pred	buy_sell
282.230011	68994500.0	278.025757	-0.988558	-1.158783	6.124526e+05	1.787092	1
280.529999	24643000.0	282.562683	1.236945	1.047494	7.537192e+05	-0.453424	-1
284.690002	12119700.0	282.831299	-0.147534	-0.325024	2.226271e+05	0.711669	1
284.820007	23280300.0	288.442780	1.787092	1.592889	1.214060e+06	1.283069	1
291.119995	36566500.0	288.333313	-0.453424	-0.628272	3.779663e+05	NaN	
289.459991	36028600.0	290.044617	0.711669	0.526756	1.125283e+06	NaN	
289.929993	25201400.0	292.163818	1.283069	1.093220	4.046867e+05	NaN	

5.5 Training Classifiers and Accuracy Scores

We use the same classifiers described in Chapter 3 except we will replace Random Forest with Artificial Neural Network (ANN) since Random forest performed worst among all the classifiers.

For the classifiers to learn, we will split the data in training and testing sets with ratio 70/30, which will allow the classifiers to learn based on the train test and afterward compare the predicted outcomes against known values in the test set. Then the classifiers will predict the buy-hold signal given the sentiment score and the possibility of price change. Tables 5.5, 5.6 and 5.7 show results for all classifiers and companies for both Headlines and Tweets.

Table 5.5: Prediction scores for Apple and Amazon

Accuracy Scores Apple		
Classifier	**Headlines**	**Tweets**
KNN	0.5193	0.4852
SVM	0.5678	**0.7309**
Log Reg	0.5193	0.4852
ANN	**0.6797**	0.6642

Accuracy Scores Amazon		
Classifier	**Headlines**	**Tweets**
KNN	0.5073	0.5555
SVM	0.6011	0.6583
Log Reg	0.4338	0.5555
ANN	**0.6297**	**0.7702**

Table 5.6: Prediction scores for Facebook and Google

Accuracy Scores Facebook		
Classifier	**Headlines**	**Tweets**
KNN	0.5151	0.5757
SVM	0.5476	**0.6880**
Log Reg	0.5151	0.5757
ANN	**0.5726**	0.6630

Accuracy Scores Google		
Classifier	**Headlines**	**Tweets**
KNN	0.5365	0.4692
SVM	**0.6309**	**0.7178**
Log Reg	0.4146	0.4692
ANN	**0.6309**	0.6928

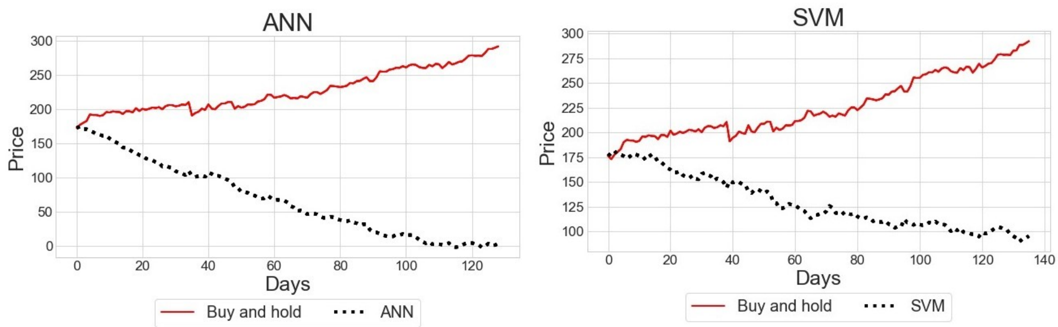
Table 5.7: Prediction scores for Microsoft

Accuracy Scores Microsoft		
Classifier	**Headlines**	**Tweets**
KNN	0.5882	0.5258
SVM	**0.6869**	0.6309
Log Reg	0.5882	0.4913
ANN	**0.6869**	**0.6547**

It is noticeable that two classifiers are outperforming the rest of the classifiers with ANN having accuracy scores between 57.26% and 68.69% for Headlines and 65.47% and 77.02% for Tweets. SVM accuracy scores are moving in a range between 54.76% and 68.69% for Headlines and 63.09% and 73.09% for Tweets. Finally, we will create trend charts to visualize and understand how the classifiers' predictions are compared with the buy-hold signals representing the historical trend lines of a company stock price movement. As observed for Apple, in Figure 5.3, both classifiers in both datasets predict stock movement in the opposite direction than the buy-hold signal.

Figure 5.3: Apple

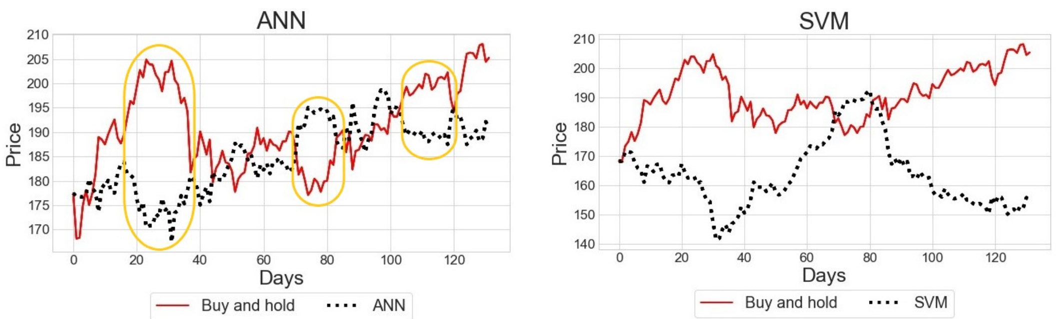
(a) Headlines (b) Tweets



In Figure 5.4 for Facebook both classifiers are returning not so accurate predictions. ANN periodically predicts almost the same price with buy-hold, however, it struggles to predict big spikes accurately. Reason can be due to VADER wrongly assign polarity sentiment. Let's assume that if tweets were wrongly labeled with negative polarity, and the user had many followers, this would lead to a deficient sentiment score. Since the sentiment score is the magnitude for creating a buy-hold signal, the classifier will inevitably predict the wrong signal.

Figure 5.4: Facebook

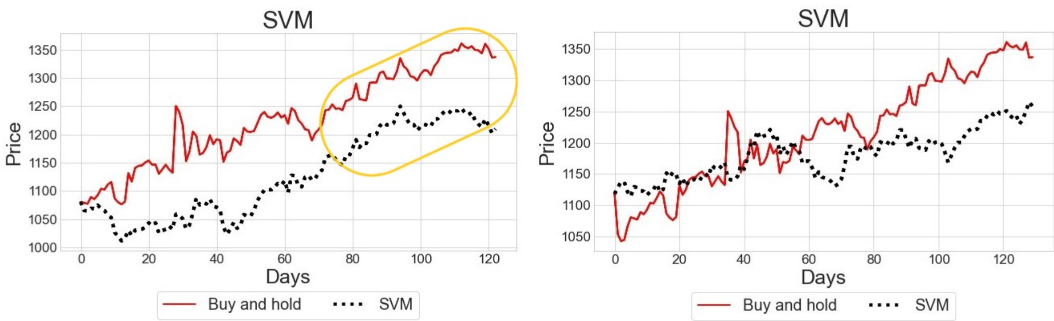
(a) Headlines (b) Tweets



In Figure 5.5 for Google, we observe auspicious results from SVM, especially for headlines. The difference between buy-hold and the classifier trend is only in the price with not so big margin. If we observe in the right upper corner of the headlines graph, the classifier manages to predict exact spikes in volatility and an almost identical downward trend in the last days. It may be interesting to understand why we have such a better result for Google compared to Facebook and Apple. As previously suggested, including additional information outside of newspapers, the classifier will only benefit since if more sources provide similar sentiment polarity higher the chance, the classifier will converge predictions in the correct direction.

Figure 5.5: Google

(a) Headlines (b) Tweets



In Figure 5.6 for Amazon, we can observe significant improvement in ANN performance for tweets where after day 100 of the time series, the classifier is predicting the almost identical trend in price movement with buy-hold, however with slightly lower price. However, the results are encouraging further investigation on how we can achieve positive predictions for a more extended period. Also, tweets from professional investors could lead to even better results.

Figure 5.6: Amazon Tweets

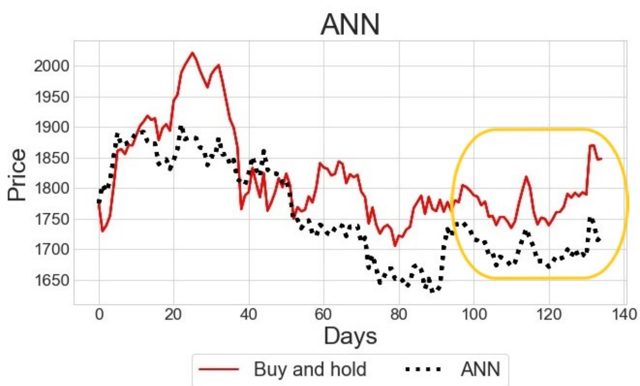
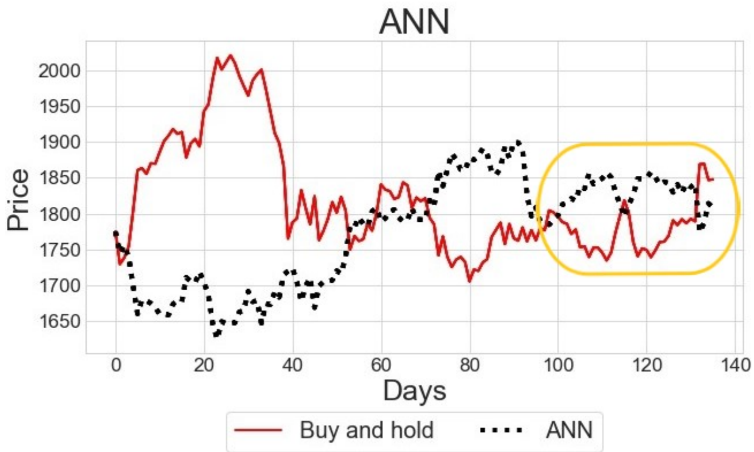


Figure 5.7 shows an exciting trend for headlines since, after day 100, the classifier mirrors the buy-hold signal, which contradicts the positive predictions in the tweets dataset. Therefore, we will accurately evaluate this particular time-period and compare both datasets to gain more insights from the data as to why we have such a big contrast in the predictions.

Figure 5.7: Amazon Headlines



This particular event is occurring in the second half of October and the first week of December 2019. There were a series of events that triggered professional and retail investors to speculate significantly more. On November 8th Reuters reported the first event that Steve Kessel, a senior vice president, is leaving the company. On November 12th almost all newspapers reported that Amazon would open and employ people for its new grocery stores who will not be part of the Whole Food chain. Then, on November 15th CNBC published that Nike's products will not be available on Amazon anymore, instead, Nike will sell its product directly to the customers. From November 15th to November 22nd, there were two adverse events. First, Amazon did not manage to secure a $10 billion JEDI contract with Pentagon at the end of October, losing to Microsoft for which in November Amazon filed a suit to the Court of Federal Claims. Then Amazon responded to the antitrust inquiry, where they claimed to host more than 384,000 individuals and over 500,00 professional sellers, however denying collecting aggregated data to manipulate the markets. Finally, positive news published on December 3rd where Reuters reported that Goldman Sachs would use Amazon cloud services to launch its new products. This positive news was followed by next day double announcements that Amazon and Novartis are entering into a multiyear partnership, and British Petroleum will migrate its entire data and over 900 applications to Amazon Cloud infrastructure. These series of news were shared across all newspapers and garnered lots of attention. However, we noticed that the sentiment tone for the same topic expressed through headlines and tweets varies. This clearly impacted VADER sentiment analysis and predictions of our classifiers. We believe there are many more events like this where VADER can assign positive sentiment or the total opposite for the same event. In Figure 5.8, we will illustrate a few examples of differences between headlines and tweets on the same topic.

Table 5.8: Distribution and examples of labeled Tweets

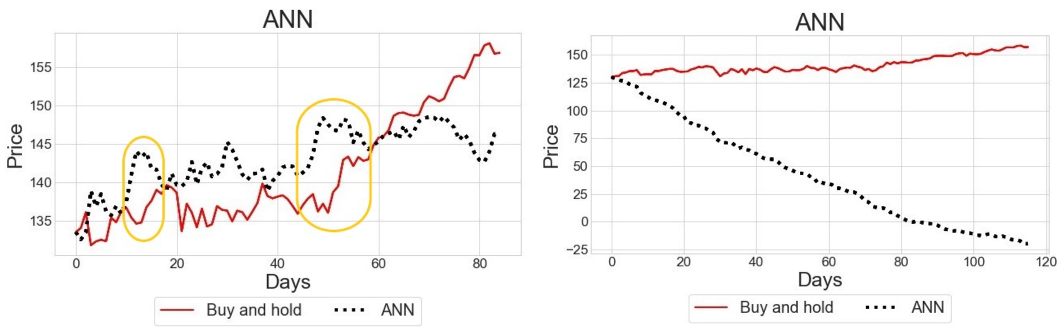
Tweet	Source	Sentiment Score
'Brands don't need Amazon' Nike's departure will prompt others to go apart	CNBC	-178.87
Amazon files suit protesting Microsoft's JEDI cloud contract with Pentagon	CNBC	-11043.54
Amazon uses 'aggregated' seller data to 'help' business and its own products, it tells lawmakers	Reuters	-238.65
Headlines	**Source**	**Sentiment Score**
Nike just 'tip of the iceberg' of companies ditching Amazon and selling directly to consumers	CNBC	553.42
Amazon cites 'unmistakable bias' in Microsoft's military cloud contract win	CNBC	654.80
Google, Facebook, Amazon and Apple offer defense in congressional antitrust probe	Reuters	356.51

In Figure 5.8 for Microsoft at first sight, the predictions look pessimistic. It seems that ANN is predicting inadequately comparing to buy-hold signals, either going in the opposite direction or capturing volatility based on some news, however, with the opposite sentiment. There is a pattern carried on by other companies explained previously for Amazon. Nevertheless, here is different since tweets predictions are not even close to what we observed for Amazon. We can notice two events in the headlines. First, there is a $\bigcup$ shape recovery in the price for the first 20 days, and the classifier is capturing the event, however, in opposite $\bigcap$ shape. This type of transient movement on future price returns and trading volumes was already confirmed in the work of Tetlock (2007), and Garcia (2013). The second event is between 40th and 50th days when there is a W-shaped recovery with an extended positive trend in the stock price.

Similarly, to the first event, the classifier captures this event too, however in opposite $\Lambda\Lambda$ shape again. As previously explained, our classifier's predictions are based on the magnitude in the sentiment score computed with the compound polarity assigned by VADER and the shares for headlines or the number of followers for twitter. For example, if some big event is happening like earnings or product announcement and VADER assigns negative sentiment, but this event is positive, it will receive many engagements on twitter, thus instead of $\bigcup$ shape it will converge to $\bigcap$ shape.

Figure 5.8: Microsoft

(a) Headlines (b) Tweets



The proposed strategy from Noah Mukhtar & Chandra (2020) is conducted only on tweets, our strategy additionally included news headlines where we show different behavior in the predictions of the classifiers. Arguably from some of the predictions, there is a need for greater sophistication because we are dealing with increasingly complex language, not just good or bad buzzwords. Moreover, we cannot process every publicly available primary source of communication for each company. As shown, changing a title for the same news creates near-duplicates, leading to redundant information flow. Nevertheless, increasing the number of news sources can help to improve the accuracy of the predictions and, as shown in Figures 5.6 and 5.5 can lead to robust results, particularly in cases in which multiple sources agree about the future direction of the market.

Chapter 6

Dicussion

In this Chapter, we would like to discuss the potential empirical caveats we came across in our analysis for finer-grained emotion classification.

At the beginning we would like to emphasize that our results are consistent with other previous work of no significant correlation between tweets and headlines labeled in one of the Ekman's emotions and future stock returns (Liu 2017; Zhang *et al.* 2011). If we observe Table 3.12 and 3.13 we can notice that LinearSVC for sentiment has higher accuracy scores comparing to the same classifier for emotions. Expectantly, the fewer classes we have, the more distant they become, indicating that it is easier for the classifier to learn and recognize the sentiment in contrast to the emotion labels. Therefore, we assume that finding no correlation might be because of this caveat when dealing with multiclass classification problems.

The training corpus we have was labeled manually and collected from professional financial newspapers. Creating manually annotated corpora is labor-intensive and difficult to predict the correct emotions. We believe that obtaining a professional opinion from psychologists and linguists can improve the annotation process, particularly for emotions. Moreover, as shown in Table 3.6, we have a relatively imbalanced distribution of emotions in our training datasets. Joy is the most frequently represented emotion. Therefore, we assume that many unseen tweets that are likely to be neutral are labeled as a joy even though they do not contain positive or negative words. If we observe Figures 4.1 and 4.2 it is evident that neutral is the least represented in both cases. In the future, this can be alleviated by extending the datasets with additional financial tweets from the less represented emotions and tweets from professional investors as undoubtedly their opinion influences retail investors. According to (Liu 2017) findings additionally extending the training dataset with particular words and phrases from financial or lexicon associated with Ekman's emotions improved her trading strategy accuracy.

A different aspect of our findings of no correlation might be because we used only a limited number of Ekman's emotions. For example, Bollen *et al.* (2011) in their study uses six dimensions (Calm, Alert, Sure, Vital, Kind, and Happy) of mood in tweets. They observed a positive correlation only with calm. Therefore, extending the dataset beyond Ekman's basic emotions might reveal a correlation with future stock returns. In this thesis, we focus only on Big Tech companies such as Facebook, Apple, Amazon, Microsoft, and Google, however extending our approach to companies from other sectors might show significantly better results.

Chapter 7

Conclusion

In this thesis, we have assessed the possibility of predicting future stock price movements of the top 5 Big Tech companies by applying multiclass classification models. As a proxy variable, we use emotions and sentiment extracted from tweets and news headlines. For all our classifiers, we use training datasets manually annotated by humans. The tweets and headlines were obtained based on keywords such as (e.g., Apple, Tim Cook, $aapl) originating from professional financial newspapers and web pages such as (e.g., CNBC, The New York Times, Reuters News, Market Watch).

While fear and anger are both negative emotions, according to Lerner & Keltner (2000), fearful people make more pessimistic risk assessments, whereas angry people make more optimistic risk assessments. We see finer-grained emotion classification as the first important contribution to the existing literature of the financial domain in predicting stock price movements. As a second important contribution from this thesis is the comparison of both headlines and tweets from same source in parallel, which to the best of our knowledge was not investigated so far. From our point of view, understanding whether the semantic context, financial jargon and sentiment polarity differs between tweets and headlines is a crucial event.

We did not confirm any robust correlation between daily stock price movements and distribution of sentiment and Ekman's basic emotions. However, we did identify a pattern that positive news has more effect on the second day after some event happened, in contrast to Fama (1970) findings that all information is immediately integrated with the next day price (Malkiel 2003; Shiller 2003). Furthermore, we observed that even though the source is the same, the content published on tweets is less neutral than the headlines. We assume this is done in order to attract more attention and engagements on Twitter. Nevertheless, since tweets contain weaker neutrality and we can capture more emotions as a future work would be interesting to expand the datasets with tweets from professional investors, financial institutions and firms, since there are studies (Ranco *et al.* 2015; Liu 2017) confirming Twitter volume can be effective in predicting future stock price movements. The majority of previous work is conducted relatively on a short period of time. However, expanding the time period may result in higher correlation, since Engelberg (2008) in his work observes approximately 5000 companies over a period between 1999 and 2005 and concludes that earning announcements published in news articles containing qualitative information has additional predictability for future returns.

Since we did not find significant correlation between automatically classified emotions or sentiment and future stock price returns but our corpora are domain-specific, referring to the top five S&P 500 companies, we decided to create a simple buy-hold investing strategy. As a proxy variable, we create sentiment score polarity using VADER in junction with other metrics such as followers and shares for tweets and headlines, respectively. In contrast to the finer-grained classification, our investing strategy has shown promising predictions from the classifiers, both on headlines and tweets. However, we observed that just by changing a title for the same news can lead to dramatically opposite predictions from the classifiers. Nevertheless, through VADER we have shown that both SVM and ANN delivered robust predictions for Google and Amazon, respectively. Implying that sentiment polarity together with other metrics can be effectively used in predicting future stock price movements.

On the other hand, the initial approach of multiclass classification did not result in robust correlation. One of the reasons is that the training datasets utilized in the analysis were manually annotated by individuals with no psychological or linguistic background whatsoever. An important future work would be to create a training corpus by obtaining a professional opinions using crowdsourcing from Amazon Mechanical Turk as Mohammad & Turney (2010) did when creating their corpora. Second reason is that we do have a relatively imbalanced distribution of emotions in both training sets. This is likely to improve by including additional resources such as tweets from professional investors or headlines from more newspapers for the less represented emotions.

Bibliography

AdFontesMedia (2020): "Media bias chart: Version 5.0." https://www.adfontesmedia.com/interactive-media-bias-chart/, Last accessed on 2020-07-13.

Ahuja, Y. & S. K. Yadav (2012): "Multiclass classification and support vector machine." *Global Journal of Computer Science and Technology Interdisciplinary* **12(11)**: pp. 14–20.

Allcott, H. & M. Gentzkow (2017): "Social media and fake news in the 2016 election." *Journal of economic perspectives* **31(2)**: pp. 211–36.

Alm, C. O., D. Roth, & R. Sproat (2005): "Emotions from text: machine learning for text-based emotion prediction." In "Proceedings of human language technology conference and conference on empirical methods in natural language processing," pp. 579–586.

Altman, N. S. (1992): "An introduction to kernel and nearest-neighbor nonparametric regression." *The American Statistician* **46(3)**: pp. 175–185.

Apple, A. (2020): "Everything about apple's products." https://igotoffer.com/apple/category/history-of-apple, Last accessed on 2020-07-13.

Arias, M., A. Arratia, & R. Xuriguera (2014): "Forecasting with twitter data." *ACM Transactions on Intelligent Systems and Technology (TIST)* **5(1)**: pp. 1–24.

Baker, M. & J. Wurgler (2006): "Investor sentiment and the cross-section of stock returns." *The journal of Finance* **61(4)**: pp. 1645–1680.

Baker, M. & J. Wurgler (2007): "Investor sentiment in the stock market." *Journal of economic perspectives* **21(2)**: pp. 129–152.

Benesty, J., J. Chen, Y. Huang, & I. Cohen (2009): "Pearson correlation coefficient." In "Noise reduction in speech processing," pp. 1–4. Springer.

Bollen, J., H. Mao, & X. Zeng (2011): "Twitter mood predicts the stock market." *Journal of computational science* **2(1)**: pp. 1–8.

Bostan, L., E. Kim, & R. Klinger (2019): "Goodnewseveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception." *arXiv preprint arXiv:1912.03184* .

Bowker, L. & J. Pearson (2002): *Working with specialized language: A practical guide to using corpora.* Routledge.

Breiman, L. (2001): "Random forests machine learning. 45: 5–32." *View Article PubMed/NCBI Google Scholar* .

Brown, G. W. & M. T. Cliff (2004): "Investor sentiment and the near-term stock market." *Journal of empirical finance* **11(1)**: pp. 1–27.

Brown, G. W. & M. T. Cliff (2005): "Investor sentiment and asset valuation." *The Journal of Business* **78(2)**: pp. 405–440.

Brownlee, J. (2020): "Can twitter help predict stock prices." `https://machinelearningmastery.com/bayes-optimal-classifier/`, Last accessed on 2020-07-10.

Chatterjee, A., K. N. Narahari, M. Joshi, & P. Agrawal (2019): "Semeval-2019 task 3: Emocontext contextual emotion detection in text." In "Proceedings of the 13th International Workshop on Semantic Evaluation," pp. 39–48.

Cortes, C. & V. Vapnik (1995): "Support-vector networks." *Machine learning* **20(3)**: pp. 273–297.

Coyne, B. (2016): "How election 2016 was tweeted so far." `https://blog.twitter.com/official/en_us/a/2016/how-election2016-was-tweeted-so-far.html`, Last accessed on 2020-07-10.

Ding, X., B. Liu, & P. S. Yu (2008): "A holistic lexicon-based approach to opinion mining." In "Proceedings of the 2008 international conference on web search and data mining," pp. 231–240.

EKMAN, P. (1992): "An argument for basic emotions." *Cognition & emotion* **6(3-4)**: pp. 169–200.

ENGELBERG, J. (2008): "Costly information processing: Evidence from earnings announcements." In "AFA 2009 San Francisco meetings paper," .

FAMA, E. F. (1970): "Efficient capital markets: A review of theory and empirical work." *The journal of Finance* **25(2)**: pp. 383–417.

FELDMAN, R. (2013): "Techniques and applications for sentiment analysis." *Communications of the ACM* **56(4)**: pp. 82–89.

FRIEDMAN, N., D. GEIGER, & M. GOLDSZMIDT (1997): "Bayesian network classifiers." *Machine learning* **29(2-3)**: pp. 131–163.

GARCIA, D. (2013): "Sentiment during recessions." *The Journal of Finance* **68(3)**: pp. 1267–1300.

GOLDBERG, Y. (2017): "Neural network methods for natural language processing." *Synthesis Lectures on Human Language Technologies* **10(1)**: pp. 1–309.

GUYNN, J. (2016): "Forget trump: Election's big winner was twitter." `https://eu.usatoday.com/story/tech/news/2016/11/08/election-winner-twitter/93509896/`, Last accessed on 2020-07-20.

HESTON, S. L. & N. R. SINHA (2017): "News vs. sentiment: Predicting stock returns from news stories." *Financial Analysts Journal* **73(3)**: pp. 67–83.

HUTTO, C. J. & E. GILBERT (2014): "Vader: A parsimonious rule-based model for sentiment analysis of social media text." In "Eighth international AAAI conference on weblogs and social media," .

JURAFSKY, D. (2000): *Speech & language processing.* Pearson Education India.

KIM, S.-M. & E. HOVY (2007): "Crystal: Analyzing predictive opinions on the web." In "Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)," pp. 1056–1064.

KLINGER, R. *et al.* (2018): "An analysis of annotated corpora for emotion classification in text." In "Proceedings of the 27th International Conference on Computational Linguistics," pp. 2104–2119.

Kordonis, J., S. Symeonidis, & A. Arampatzis (2016): "Stock price forecasting via sentiment analysis on twitter." In "Proceedings of the 20th Pan-Hellenic Conference on Informatics," pp. 1–6.

Kuhn, M., K. Johnson *et al.* (2013): *Applied predictive modeling*, volume 26. Springer.

Lerner, J. S. & D. Keltner (2000): "Beyond valence: Toward a model of emotion-specific influences on judgement and choice." *Cognition & emotion* **14(4)**: pp. 473–493.

Li, Q., T. Wang, P. Li, L. Liu, Q. Gong, & Y. Chen (2014a): "The effect of news and public mood on stock movements." *Information Sciences* **278**: pp. 826–840.

Li, X., H. Xie, L. Chen, J. Wang, & X. Deng (2014b): "News impact on stock price return via sentiment analysis." *Knowledge-Based Systems* **69**: pp. 14–23.

Liao, R. C., X. Wang, & G. Wu (2019): "The role of media in mergers and acquisitions." *Available at SSRN 3456718* .

Liu, C. H. (2017): *Applications of twitter emotion detection for stock market prediction.* Ph.D. thesis, Massachusetts Institute of Technology.

Loughran, T. & B. McDonald (2011): "When is a liability not a liability? textual analysis, dictionaries, and 10-ks." *The Journal of Finance* **66(1)**: pp. 35–65.

Madhumathi, V. & T. S. Rajan (2014): "Review on: Nearest neighbor search with keywords." *International Journal of Science, Engineering and Computer Technology* **4(10)**: p. 263.

Malkiel, B. G. (2003): "The efficient market hypothesis and its critics." *Journal of economic perspectives* **17(1)**: pp. 59–82.

Mao, Y., W. Wei, B. Wang, & B. Liu (2012): "Correlating s&p 500 stocks with twitter data." In "Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research," pp. 69–72.

Menard, S. (2002): *Applied logistic regression analysis*, volume 106. Sage.

MITTAL, A. & A. GOEL (2012): "Stock prediction using twitter sentiment analysis." *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)* **15**.

MOHAMMAD, S. (2012a): "# emotional tweets." In "* SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)," pp. 246–255.

MOHAMMAD, S. (2012b): "Portable features for classifying emotional text." In "Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies," pp. 587–591.

MOHAMMAD, S. & P. TURNEY (2010): "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon." In "Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text," pp. 26–34.

MOHAMMAD, S. M., X. ZHU, S. KIRITCHENKO, & J. MARTIN (2015): "Sentiment, emotion, purpose, and style in electoral tweets." *Information Processing & Management* **51(4)**: pp. 480–499.

MURPHY, K. P. (2012): *Machine learning: a probabilistic perspective.* MIT press.

NG, A. (2000): "Cs229 lecture notes." *CS229 Lecture notes* **1(1)**: pp. 1–3.

NOAH MUKHTAR, R. H. & V. CHANDRA (2020): "Can twitter help predict stock prices." `http://localhost:8888/notebooks/OneDrive%20-%20Anheuser-Busch%20InBev/Thesis/Data/Tweet_Stock_price_prediction.ipynb`, Last accessed on 2020-07-10.

OLIVEIRA, N., P. CORTEZ, & N. AREAL (2017): "The impact of microblogging data for stock market prediction: Using twitter to predict returns, volatility, trading volume and survey sentiment indices." *Expert Systems with Applications* **73**: pp. 125–144.

Pagolu, V. S., K. N. Reddy, G. Panda, & B. Majhi (2016): "Sentiment analysis of twitter data for predicting stock market movements." In "2016 international conference on signal processing, communication, power and embedded system (SCOPES)," pp. 1345–1350. IEEE.

Pang, B., L. Lee *et al.* (2008): "Opinion mining and sentiment analysis." *Foundations and Trends® in Information Retrieval* **2(1–2)**: pp. 1–135.

Plutchik, R. (1984): "Emotions: A general psychoevolutionary theory." *Approaches to emotion* **1984**: pp. 197–219.

Ranco, G., D. Aleksovski, G. Caldarelli, M. Grčar, & I. Mozetič (2015): "The effects of twitter sentiment on stock price returns." *PloS one* **10(9)**: p. e0138441.

Robert, C. (2014): "Machine learning, a probabilistic perspective."

Roberts, K., M. A. Roach, J. Johnson, J. Guthrie, & S. M. Harabagiu (2012): "Empatweet: Annotating and detecting emotions on twitter." In "Lrec," volume 12, pp. 3806–3813. Citeseer.

Rokach, L. & O. Z. Maimon (2008): *Data mining with decision trees: theory and applications*, volume 69. World scientific.

Sanders, J. (2020): "Shame never sleeps:updating the facebook timeline of scandal and strife." `https://creativefuture.org/facebook-scandal-timeline/`, Last accessed on 2020-07-13.

Sayad, S. (2020): "Can twitter help predict stock prices." `https://www.saedsayad.com/k_nearest_neighbors.htm`, Last accessed on 2020-07-10.

Schuff, H., J. Barnes, J. Mohme, S. Padó, & R. Klinger (2017): "Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus." In "Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis," pp. 13–23.

Schumaker, R. P., Y. Zhang, C.-N. Huang, & H. Chen (2012): "Evaluating sentiment in financial news articles." *Decision Support Systems* **53(3)**: pp. 458–464.

Schütze, H., C. D. Manning, & P. Raghavan (2008): *Introduction to information retrieval*, volume 39. Cambridge University Press Cambridge.

Shiller, R. J. (2003): "From efficient markets theory to behavioral finance." *Journal of economic perspectives* **17(1)**: pp. 83–104.

Smailović, J., M. Grčar, N. Lavrač, & M. Žnidaršič (2014): "Stream-based active learning for sentiment analysis in the financial domain." *Information sciences* **285**: pp. 181–203.

Stone, P. J., D. C. Dunphy, & M. S. Smith (1966): "The general inquirer: A computer approach to content analysis." .

Strapparava, C. & R. Mihalcea (2007): "Semeval-2007 task 14: Affective text." In "Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)," pp. 70–74.

Stubbs, M. (2006): "Corpus analysis: The state of the art and three types of unanswered questions." *System and corpus: Exploring connections* pp. 15–36.

Taboada, M., J. Brooke, M. Tofiloski, K. Voll, & M. Stede (2011): "Lexicon-based methods for sentiment analysis." *Computational linguistics* **37(2)**: pp. 267–307.

Tetlock, P. (2016): "The role of media in the stock market." *News and Finance* .

Tetlock, P. C. (2007): "Giving content to investor sentiment: The role of media in the stock market." *The Journal of finance* **62(3)**: pp. 1139–1168.

Wang, S. I. & C. D. Manning (2012): "Baselines and bigrams: Simple, good sentiment and topic classification." In "Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)," pp. 90–94.

Wong, C. & I.-Y. Ko (2016): "Predictive power of public emotions as extracted from daily news articles on the movements of stock market indices." In "2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)," pp. 705–708. IEEE.

WU, H. C., R. W. P. LUK, K. F. WONG, & K. L. KWOK (2008): “Interpreting tf-idf term weights as making relevance decisions.” *ACM Transactions on Information Systems (TOIS)* **26(3)**: pp. 1–37.

WUTHRICH, B., V. CHO, S. LEUNG, D. PERMUNETILLEKE, K. SANKARAN, & J. ZHANG (1998): “Daily stock market forecast from textual web data.” In “SMC’98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218),” volume 3, pp. 2720–2725. IEEE.

ZHANG, X., H. FUEHRES, & P. A. GLOOR (2011): “Predicting stock market indicators through twitter â€śi hope it is not as bad as i fearâ€ť.” *Procedia-Social and Behavioral Sciences* **26**: pp. 55–62.