INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY

Using Sentimental Analysis on Social Media to Predict Stock Market Fluctuation

Nathaniel Mizzi

Supervisor: Mr Silvio Abela

May, 2021

Submitted to the
Institute of Information and Communication Technology
in partial fulfillment of the requirements for the degree of
B.Sc. (Hons.) Software Development

Authorship Statement

This dissertation is based on the results of research carried out by myself, is my own composition, and has not been previously presented for any other certified or uncertified qualification.

The research was carried out under the supervision of Mr Silvio Abela.

N. Mizzi

Nathaniel Mizzi

June 6, 2022

Copyright Statement

Nathaniel Mizzi

102, Santa Rita,
Triq il-Gudja,
Hal Ghaxaq, GXQ2177

June 6, 2022

Dedication

I would like to dedicate this dissertation to my family and friends who gave me continuous support and encouragement throughout my studies.

Acknowledgements

I'd want to convey my appreciation to a few individuals for their assistance. I'd like to thank Mr. Silvio Abela, my mentor, for his guidance and support throughout this research. Finally, I would want to thank my family for their continuous encouragement and support.

Abstract

This study explored that the use of sentiment analysis on social media platforms combined with historical stock market data integrated with machine learning can forecast stock market fluctuations. Three separate stock businesses that provide services in three distinct areas were hand-picked for this study. Microsoft, McDonald's, and Tesla were the three companies selected from the Fortune 500 list.

Two social media networks, Reddit and StockTwits, were used to collect public opinion. Posts from the previous two years were gathered and saved, and the data was later pre-processed to be cleaned before being employed for model training. For Historical stock prices, four different data sets were gathered from two sources: Yahoo! Finance for daily and weekly prices and AlphaVantage for hourly and bi-hourly prices.

The sentiment of public opinions was classified using VADER, a lexicon-based sentiment analysis classification method. The classifier categorizes the posts as positive, neutral, or negative. This information was eventually used for training the stock fluctuation predictor. These two unique data sets were eventually aggregated to generate four distinct data sets: hourly, bi-hourly, daily, and weekly.

Following that, the aggregated data sets were used to train two distinct machine learning models, one based on Bidirectional LSTM and the other on convolution neural networks. Many models had been created in order to fine-tune the hyper parameters for time series prediction. The optimal number of epochs across all platforms and stocks investigated was discovered to be 25 epochs, and the optimal number of entries in the time window for each interval was discovered to be nine hours for hourly, five bi-hours for bi-hourly, one or two days for daily, and one or two weeks for weekly. For this study, two days and two weeks were used, giving the model more data to predict the following.

When the models were trained using the identified hyper parameters, the accuracy varies per company, and the platform from which the data is collected also has an effect. Although it has been discovered that the sentiment analysis feature set does help boost accuracy when compared to the baseline feature set (Open Price, Close Price, and Volume) alone. The results also reveal that the accuracy significantly improves when the data sets were aggregated into a longer interval.

The highest accuracy on various stocks and platforms was found to be 52 percent for hourly using the sentiment analysis feature set, 59 percent for bi-hourly, 61 percent for daily, and 74 percent for weekly predictions. Bidirectional LSTM had an average accuracy of 52 percent and a maximum accuracy of 74 percent, while convolution neural networks had an average accuracy of 52 percent and a maximum accuracy of 54 percent. Bidirectional LSTM is, therefore, the optimal model algorithm for this situation.

To conclude, the study was influenced by two global events that shocked the world: the COVID-19 global pandemic and the Russian invasion of Ukraine. These events had an impact on global sentiment and the stock market. Another limitation is the accuracy of the sentiment classifier used in this investigation, which was 67.6

percent accurate. A larger data set is recommended for future research because weekly data sets get smaller when aggregated and would also reduce overfitting. Another suggestion was to use more accurate NLP sentiment analysis tools, as the sentiment classifier is the study's limitation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Purpose Statement

This research study aims to determine if using sentiment analysis on social media can be used to predict the direction of the stock market fluctuations. The predicted results are then compared with the actual results to determine the accuracy of the investigated and utilized models in this study. The independent variables used in this study are the data sets, model hyper-parameters, number of epochs, number of entries in the time window, sentiment analysis, and machine learning models. On the other hand, the dependent variables in this study are the Accuracy and F1 Score of the stock fluctuation prediction models and the model's accuracy compared with the baseline data set. Another dependent variable is the accuracy of the sentiment analysis classifier tool which affects the outcome of the other dependent variables.

1.2 Hypothesis

Using the purpose statement given in Section 1.1, it is hypothesized ***"that analyzing social media platforms one can infer whether fluctuations will happen on a stock market price by analyzing the public comments"***.

With this hypothesis, the following research questions are identified:

1. “What type of algorithms allow us to identify fluctuations based on a data set gathered from social media?”

2. “Can we predict future fluctuations based on the information gathered from Social Media Comments and Stock Price history?”

1.3 Motivation

With the current trends, the stock market is attracting much interest, according to a CNBC report (Cox 2021), but having many people investing is not always a good thing because it can make the stock market more volatile. With today’s technology breakthroughs, a plan was devised to construct and develop such algorithms to detect fluctuations in stocks.

1.4 Research Outline

This section gives an overview of the remaining chapters and sections of this research beginning with the Literature Review (Chapter 2), which includes a backdrop of the stock market, sentiment analysis, and market prediction. Various methodologies employed by other researchers are also presented.

The Methodology (Chapter 3) begins with a Figure 3.1 pipeline which displays a diagram and a description of how the pipeline works and then moves on to the tools and packages used to develop and test this approach. Data sources and methods of collection are examined, followed by pre-processing procedures. Finally, a summary of the methods utilized and the hyper-parameters employed.

The Analysis of Results & Discussion (Chapter 4) begins with the results of the sentiment analysis classification approaches followed by a comparison of such methods. The outcomes of the stock market fluctuation prediction are then reported in tables. The optimal number of epochs and the optimal number of entries in the time window are displayed for the time series. The successive section discusses models training which displays the outcomes of each time period for each stock under consideration and each algorithm. Models that employ both the sentiment analysis feature set and the baseline feature set are compared to models that use the baseline feature set. To conclude the chapter, a description of the sentiment analysis approach and stock market fluctuation predictions is presented.

The Conclusion (Chapter 5) chapter brings the research to a close, by answering the research questions. Following that, limitations and recommendations for future research are presented.

Chapter 2

Literature Review

The stock market can be influenced by various conditions, making its value fluctuate. These conditions fall under three distinct categories: Fundamental Factors, Technical Factors, and Market Sentiment (Sindhu et al. 2014). Each of these three important factors will be discussed. Furthermore, different types of sentiment analysis algorithms such as Lexicon-based, Machine Learning, and Hybrid will be discussed. Finally, a discussion about stock price fluctuation prediction will be discussed along with different approaches and algorithms involving sentiment analysis and machine learning.

2.1 Conditions that influence the stock market

This section presents and discusses stock market-related factors that can have an affect or influence the stock value to fluctuate.

2.1.1 Quantitative Factors

According to a study conducted by Sindhu et al. in 2014, quantitative factors such as "The Earnings base" are set on measures such as "Earnings Per Share" (EPS), "Cash Flow Per Share" (FCF), and "Dividends Per Share" (DPS). They are utilized to convey information about the performance of the company's shares.

Other essential elements include the expected expansion in the earning base and the stock discount rate as a result of inflation (Sindhu et al. 2014).

2.1.2 Technical Factors

Technical factors also play a significant role in influencing the stock market. A study conducted by Zulkarnaen et al. found that technical factors impact the stock market prices significantly. These technical factors such as Inflation, "Return On Assets", "Debt To Equality", "Net Profit Margin" ,"Total Asset Turnover", Stock Volume, and Stock prices in the past are used by investors to determine the stock market's performance and to project it's performance and growth for the future before investing or to keep the ownership of current stock shares (Zulkarnaen et al. 2016). Inflation is a significant factor that has the potential to create shocks in the market that affects the market. In 2008, Bordo et al. researched inflation shocks and discovered that inflation shocks leave a sizeable negative impact on the stock market, which results in affecting the stock market prices (Bordo et al. 2008).

2.1.3 News

News plays a significant factor in the stock market as it can influence the investors' sentiment.

As early as 2009, Schumaker & Chen found that stock prices were affected following the announcement of a piece of news even after only 20 minutes (Schumaker & Chen 2009). Later on, research conducted in 2011 (Ormos & Vázsonyi 2011) on the impact of news on the stock market confirmed that which had been affirmed by (Schumaker & Chen 2009) that there is a change in price after news announcements. Furthermore, the authors also found that news about a particular company, such as earning call reports, is able to influence the price of the stock. News that is shared on social media can trigger fluctuations in trends of the stock market, whether it is real news or fake news, as mentioned by Jiao et al. (2020).

2.1.4 Market Sentiment

The Market Sentiment is another significant factor in the stock market industry which gives us an insight into how the investors feel about a particular stock psychologically, individually, and collectively. The sentiment changes based on the feelings and emotions of investors, some factors which might affect it are company performance, news related to the company, events, and social media. Other variables are found in market sentiments, such as Fear Index, High-Low Index, and Bullish Percent Index (BPI). Information shared publicly on platforms such as social media produces sentiment in investors that would influence their decision on whether to buy or sell (Kearney & Liu 2014).

2.2 Sentiment Analysis

Sentiment Analysis uses natural language processing (NLP) to analyze a word, sentence, or a paragraph's sentiment to determine whether the data analyzed is either positive or negative. Sentiment Analysis has multiple uses; according to Alessia et al. this technique can be used in businesses, politics, public actions, and finances. The authors also said that the sentiment analysis data is retrieved from social media sites such as Twitter and Facebook (Alessia et al. 2015). According to Medhat et al. Sentiment Analysis algorithms fall into three different types: Rule-based/Lexicon-based approach, Machine Learning approach, and Hybrid approach (Medhat et al. 2014).

2.2.1 Lexicon/Rule Based

The lexicon-based approach uses a labeled pre-compiled collection of sentiment lexicons, which are later separated into two classes: dictionary-based, and corpus-based. Later, the sentiment polarity is calculated using various statistics and semantic methods.

This method depends on a dictionary of sentiment words labeled either positive or negative. The collection of words is later used to determine if a word, sentence, or paragraph contains these sentiment words. A large data corpus must do lexicon-based sentiment analysis depending on the subject. Most lexicon-based methods analyze common words. However, new words such as misspellings, foreign words, abbreviations, and colloquialisms are not in the dictionary, so these words cannot be identified.

Some of these models, such as LIWC, according to Dhaoui et al. include frequently used words used in social media and messaging apps such as abbreviations (e.g., *"LOL", "IMHO", "TTYL", "TBH"*). Punctuation based emoticons are also identified (e.g. *":)", ":D", ":(" and ";)"*). Although this method does not support emojis (Dhaoui et al. 2017).

2.2.2 Machine Learning

Machine learning can be used to classify text messages. A dataset is required for this approach to work. The dataset needs to be split into two separate datasets for training and testing the model. There are a variety of machine learning methods that may be used to classify text, including Naïve Bayes (NB), Support Vector Machines (SVM), Random Forests (RF), Multilayer Perceptrons (MP), and RoBERTa, all of which are commonly used for text classification.

The gathering of a labeled data set is the first step in the approach. The data set needs to be labeled correctly with the values they correspond to sentimentally. Once the dataset has been collected, each entry is subjected to a variety of Natural Language Processing (NLP) techniques, such as the removal of stop words, symbols, and numeric characters and lower-casing the text in the text body.

According to Symeonidis et al., a study was carried out in which he compared several pre-processing strategies for Twitter sentiment analysis. Based on their findings, the authors found that lemmatization, replacing repetitions of punctuation, eliminating contradictions, and removing numerals were the most effective strategies for analyzing the sentiment on Twitter.

This procedure is carried out to eliminate the background noise that specific sentences include. Afterward, the cleaned data is processed by the model to be trained, and the correctness of the model is tested on a test data set that the model has never encountered before.

After that, feature extraction is performed to identify the relevant characteristics that will be used in the sentiment analysis to extract and classify them.

Machine learning models algorithm such as Naïve Bayes (NB), Support Vector Machines (SVM), Random Forests (RF), and RoBERTa are to be selected at this stage to train and test.

2.2.3 Hybrid

The hybrid approach relies on Lexicon and Machine Learning. In research conducted by Zhang et al. the sentiment of the text was classified by using a model trained with a data set of opinion lexicons. The model was then used on a data set to be later used to train a binary classifier model to predict the sentiment polarity. A binary classifier is then used to predict the sentiment polarity of the text given (Zhang et al. 2011).

2.3 Stock Fluctuation Prediction

As was previously mentioned in Section 2.1, the stock market may be anticipated in various methods, including news, patterns, sentiment analysis, and machine learning. A mixed approach can also be explored whenever it comes to prediction. As Zhang et al. suggested, it is the optimal choice. Sentiment has such an impact on the stock market that changes are unavoidable. Investor sentiment could be monitored from various sources to obtain intelligence into how the stock market responds.

2.3.1 Pattern Recognition

Pattern Recognition is a technique that uses technical analysis to gather data and identify specific patterns. Technical analysis is formed from years of stock market observations, and the analysis can assist investors in making the best decision on whether to make a purchase or sale choice.

In their research Chen & Chen, the authors gathered a data set containing six distinct attributes: date, opening price, closing price, highest price, lowest price, and the trading volume for that day. Two types of template grids were utilized, charting patterns and technical indicators representing the "*bull-flag template*". Stock Patterns and other technical indicators were required to be produced beforehand to forecast these attributes.

Additionally, the authors stated that the stock price pattern is established by the closing price of each day and that nine various technical indicators are generated by six different values: MA, RSI, STOD, OBV, ROC, VR, PSY, AR, and DIS. These indicators were chosen because the authors believe they are highly predictive of future stock prices. To lessen the complexity of computing Chen & Chen used a cumulative probability distribution approach to segment the values of each indicator into linguistic values: low, medium, and high.

During the same research, the authors trained the model with a technique known as "PIP-bull pattern matching," which recognizes the "bull-flag pattern" in the training data. Additionally, they highlighted that the model was given three experimental parameters: the fitting window, the holding period, and the distance threshold. The authors then proceeded with template matching, which is utilized to generate a suitable value for each "bull-flag pattern," then employed in the subsequent step. The fitting value is calculated via template matching. The authors then constructed the criteria for determining whether a stock fits the pattern or not by utilizing a minimum and average criterion. Finally, Chen & Chen stated that their proposed method returned a negative TIR% for predicting the fourth period of the NASDAQ, but the average TIR% was still high at 454.29%.

(Chen & Chen 2016)

Another study, carried out by Cervelló-Royo et al., made use of a flag pattern with a 10x10 grid of weights. The pattern is used to identify the presence of a bull flag pattern. According to the authors, the first seven columns of weights indicate the Downes and Goodman consolidation process, while the last three columns represent the breakout process. According to the authors, the breakout phase is typically accompanied by a dramatic increase in price. The bear flag variation is generated by reflecting the horizontal axis in the opposite direction.

By placing the template atop the price window, the researchers said it was possible to detect the presence of these separate flags. The highest price in the window is then made to coincide with the top of the grid, and the lowest price is made to match with the bottom of the grid. The price range is defined as the difference between the maximum and the lowest price possible. The pricing range is used to determine the optimal fit value that best reflects the degree of matching. The authors then explain that when the price in the cell labeled one falls within a specified time period, the fit value is increased by one unit, and when the price in the cell labeled 0.5 falls within a set time frame, the fit value is increased by another 0.5 units.

This procedure is repeated for each unique value that appears in the template. Rather their bodies, because the candlestick might fall in more than one cell per column in the template and carries more information on price evolution. Additionally, it is stated that not all of the candlestick's high and low ranges were utilized, as it may contain extreme price levels that were already reached during the opening and rejected before the close. Finally, the authors mentioned that with their proposed method, the best results were obtained when considering the adjusted levels of stop-loss between 0.2 and 0.4 times the price (Cervelló-Royo et al. 2015).

2.3.2 Sentiment Analysis

A study conducted by Bollen et al. considered using Twitter to gather sentiment data to try and predict stock movements. The data collected was then analyzed using Google Profile of the Mood States and Opinion Finder to find correlations and predict the closing stock price. The author mentioned that they had chosen a time period that included socio-cultural events. A 2-month period was chosen from October 5$^{\text{th}}$, 2008, till December 5$^{\text{th}}$, 2008. This time period was selected because it encompasses several events that significantly impacted the public's sentiment. During those two months, two events, such as U.S presidential election and Thanksgiving, were being held. These events were used to cross validate the public's sentiment. Later the author mentioned that a Granger causative analysis was performed to determine the correlation of a time period with another. The data collected from the sentiment analysis is later applied to a Self-Organizing Fuzzy Neural Network model with a history of stock market prices. It is mentioned that 9,853,498 tweets by 2.7 million users data set were used for the sentiment data set. A data set containing closing prices from February to December 2008 was collected for the stock price history. Bollen et al. 's technique managed to achieve an accuracy score of 87.6% while predicting the up and down movement(Bollen et al. 2011).

Mittal & Goel continued their study based on Bollen et al.'s approach. The authors mentioned that a much larger data set was used of 476 million tweets from 17 million different users. While pre-processing the stock market prices data set collected from *Yahoo!* the author noticed that the data set was missing some data for weekends and holidays when the stock market happened to be closed. The values were approximated using a concave function to fill this missing data. The authors observed that the stock market movement has a few sudden jumps or falls that are hard to predict. To prevent these actions from affecting the model, the authors adjusted the stock values from shifting up or down to steep falls or jumps without disturbing the trends of that time period. Although the authors stated that that change still contained volatile activity that is difficult to predict, this data was pruned in the final training and testing to prevent it from affecting the model.

For the sentiment analysis Mittal & Goel decided to classify only four different types of moods precisely: Alert, Clam, Happy and Kind. The researchers were not sufficiently using ready-made tools such as *"SentiWorldNet"* or *"OpinionFinder"* as they found them to be inefficient. Because of that, they opted to make their analysis tools. A world list generation was used to map the moods and later used to analyze the tweets. The tweets had to be filtered because of the immense amount of data they had, and it would take them a long time to be processed if used without any filtering. They mentioned that they only considered tweets that express a feeling while the rest were stripped out. Later a daily score computation algorithm was used to calculate the score of every word in the world list generation data set said per day against the total words. When it came to model learning and prediction, the authors tried to use four different machine learning algorithms Linear Regression, Logistic Regression, Support vector machine, and Self Organizing Fuzzy Neural Networks. Self Organizing Fuzzy Neural Networks gave the best results with an accuracy of 75.56% (Mittal & Goel 2012).

Alkubaisi et al. proposed an approach to a stock market classification model by using sentiment analysis on a data set collected from Twitter. Then the author mentioned that it is categorized using Hybrid Naïve Bayes classifiers. The author mentioned that experts labeled the Twitter data set with experience in the stock market to determine which polarity (negative, positive, or neutral) the tweets fall. The data set classified by experts got an accuracy of 90.38%. The author made another data set consisting of 10% of tweets labeled by experts and 90% of tweets labeled by lexical-based auto labeling. For this data set, the author mentioned an accuracy of 82.53% was achieved (Alkubaisi et al. 2018).

Pant et al. have researched using sentiment analysis and historical price history on crypto-currency. The same technique can be applied to the stock market. The authors used a Recurrent Neural Network approach to gather knowledge about a certain crypto-currency. When it comes to data collection, the authors gathered the information needed from Twitter accounts related to Crypto Currency from January 1st 2015 to December 31st 2017. The dataset containing over 7500 tweets was labeled manually into three groups: positive, negative, and neutral or irrelevant. The authors removed irrelevant tweets from the dataset by using the FuzzyWuzzy method, checking for similarities to the predefined irrelevant or repeated tweets. Later a regex search was performed to remove hyperlinks and emojis. The authors then used a combination of Regex and a Named Entity Recognizer algorithm to extract information such as Names, Organizations, and Counties present in tweets which later are used to give the sentiment weight. For feature extraction Pant et al. used Word2Vector and Bag-of-Words methods to classify text.

For Sentiment Analysis, the authors used the manually labeled data set tweets and trained against five different machine learning algorithms Naïve Bayes, Bernoulli Naïve Bayes, Multinomial Naïve Bayes, Linear Support Vector Classifier, and Random Forest. It is mentioned that the output of each of these algorithms (positive or negative) is used to classify a new tweet to that class. A Recurrent Neural Network approach was used as a predictor with LSTM and GRU variations. It uses a three-layer model, an input layer, a GRU layer, an LSTM, and a dense layer, the output layer. When it comes to results for sentiment analysis, the model with Word2Vector got a classification accuracy of 69.82%, while the Bag-of-words approach managed to get an accuracy of 78.49%. The author later mentioned that the overall accuracy of the sentiment classification with a voting classifier and a validation split of 1:3 managed to get an 81.39% accuracy and 82.90% precision. When it comes to price prediction, the accuracy for the Recurrent Neural Network resulted in 77.62% (Pant et al. 2018).

Research on the prediction of stock closing price during the COVID-19 pandemic was carried out by the researchers Chou et al. It is mentioned that the authors collected 289,308 tweets from StockTwits and another 437,868 tweets directly from Twitter. The researchers also mentioned that the tweets from StockTwits were ultimately chosen to train the sentiment analysis classifier as the StockTwits platform allows the investors to label comments with a "bearish" or "bullish" tag. Therefore, according to the researchers, the comments are labeled accurate. A one-to-one ratio of bullish and bearish comments was collected, 65,783 tweets for each. For the sentiment analysis, an LSTM-based approach was used. The authors mention that their chosen approach is to predict whether a comment is bullish or bearish and not the polarity of the tweets. The Long short-term memory (LSTM) model incorporates the GloVe embeddings in the Embedding layer. A 200-dimensional vector representation of words was used, and then trained the bidirectional LSTM recurrent neural network model, which returns the prediction of the sentiment using a *'sigmoid'* activation function. Another LSTM model is used when it comes to price prediction. Four separate layers were used: the input layer, LSTM layer, Attention layer, and Output layer. For the input layer, the researchers stated that input contained the opening stock price, highest price, lowest price, closing price, trading volume, and the sentiment index gathered from the previous LSTM sentiment analysis model. The LSTM layer for predicting the closing price was given 128 neurons. The attention layer consisted of three parts: attention weights, context vector, and attention vector. The output layer then returns the result of the attention layer through the *'relu'* activation function. When it comes to results, the sentiment analysis LSTM model achieved an accuracy of 74%. The LSTM model for prediction had an R squared of 0.737, which means it has a high level of correlation (Chou et al. 2021).

Stenqvist & Lönnö published a paper in which they investigated the use of Twitter sentiment analysis to forecast the price volatility of a cryptocurrency, in this case, Bitcoin. In their investigation, the authors used two distinct datasets, one consisting of historical bitcoin prices in USD and another dataset containing tweets scraped on a dedicated server to prevent interruptions while gathering data. Daily bitcoin price data was acquired from a public API provided by CoinDesk, a provider of crypto-currency news and price data. For the other dataset, tweets were acquired in real-time from Twitter. To narrow the scope and incorporate only valuable data, the researchers selected particular keywords such as *Bitcoin, BTC, XBT* and *satoshi*. A total of 2.2 million Bitcoin-related tweets were gathered. The sentiment of the gathered data was then analyzed using VADER, a lexicon-based sentiment analysis tool. The sentiment analysis results are then pooled to create a time series. The average sentiment score for each created group is then calculated. After the data is collected, analyzed, and aggregated Stenqvist & Lönnö try to derive predictions from the data gathered. The authors attempt to establish a correlation between the change in bitcoin's price and the change in Twitter sentiment. To do so, they applied frequency lengths and shifts to establish a link. Discrete-time intervals, such as time series, were used for short-term forecasting. A five-minute to four-hour time series range was used to try and find a relation. The time series is examined over four distinct shifts of 1, 2, 3, and 4. The authors discovered that in their tests, hour frequency with 3-hour shifts had the most accuracy with 83%

In a research by Bahceci & Alsing, the use of social media to predict the stock market was investigated. A set of machine learning methods were utilized to determine the optimal performer. Twitter was utilized to collect public opinion, classified using a sentiment dictionary by assigning a score of -1 to every negative term and a score of 1 to every positive word. The total score was then determined by adding all positive and negative terms in the text. Later, the data was aggregated on a per-day basis. However, they were given a definition. Tweets with more than two hundred thousand followers were classed as having considerable influence, with a positive score as positive and a negative score as negative. Five different machine learning models were trained using Naive Bayers, SVM, Decision Trees, Random Tree, and Artificial Neural network. The Artificial Neural Network model achieved the highest accuracy score of 68 percent. When the model is deployed and trained on stocks such as Walmart, Netflix, and Microsoft, the accuracy for each stock differs. Walmart achieved 80 percent accuracy, followed by Netflix with 60 percent and Microsoft with 55 percent. The author stated that Microsoft's accuracy could be compromised due to the company's size and the volatility of its shares. The authors also discovered that public opinion has little impact on the stock price and that only accounts with a significant number of followers have an effect. According to the author, the majority of these accounts are news or reporters covering the organization.

2.4 Chapter conclusion

This section has discussed several market situations that can affect the stock market and the various ways in which sentiment analysis can process, analyze, and recognize a word or phrase. Finally, various approaches to stock market prediction were studied utilizing machine learning techniques in conjunction with sentiment analysis.

Chapter 3

Research Methodology

This chapter aims to go into greater detail, explain, and walk through the phases of implementing a machine learning approach for forecasting stock market volatility by combining sentiment analysis and other machine learning techniques. This study examines the stock of three different companies: McDonald's, Microsoft and Tesla. The companies were hand-selected from a pool of hundreds of thousands of Fortune 500 candidates.

3.1 Aim

This aim of this research is to predict the fluctuating movement of a stock by using social media platforms to identify the sentiment of people's public opinions and gather knowledge about a specific stock to predict changes in the stock market before it happens.

3.2 Pipeline

The pipeline begins by gathering two distinct data sets: opinion data acquired from various sources and historical stock prices. The data sets are then cleansed, and only the values that are used are retained. The opinion data set is then classified

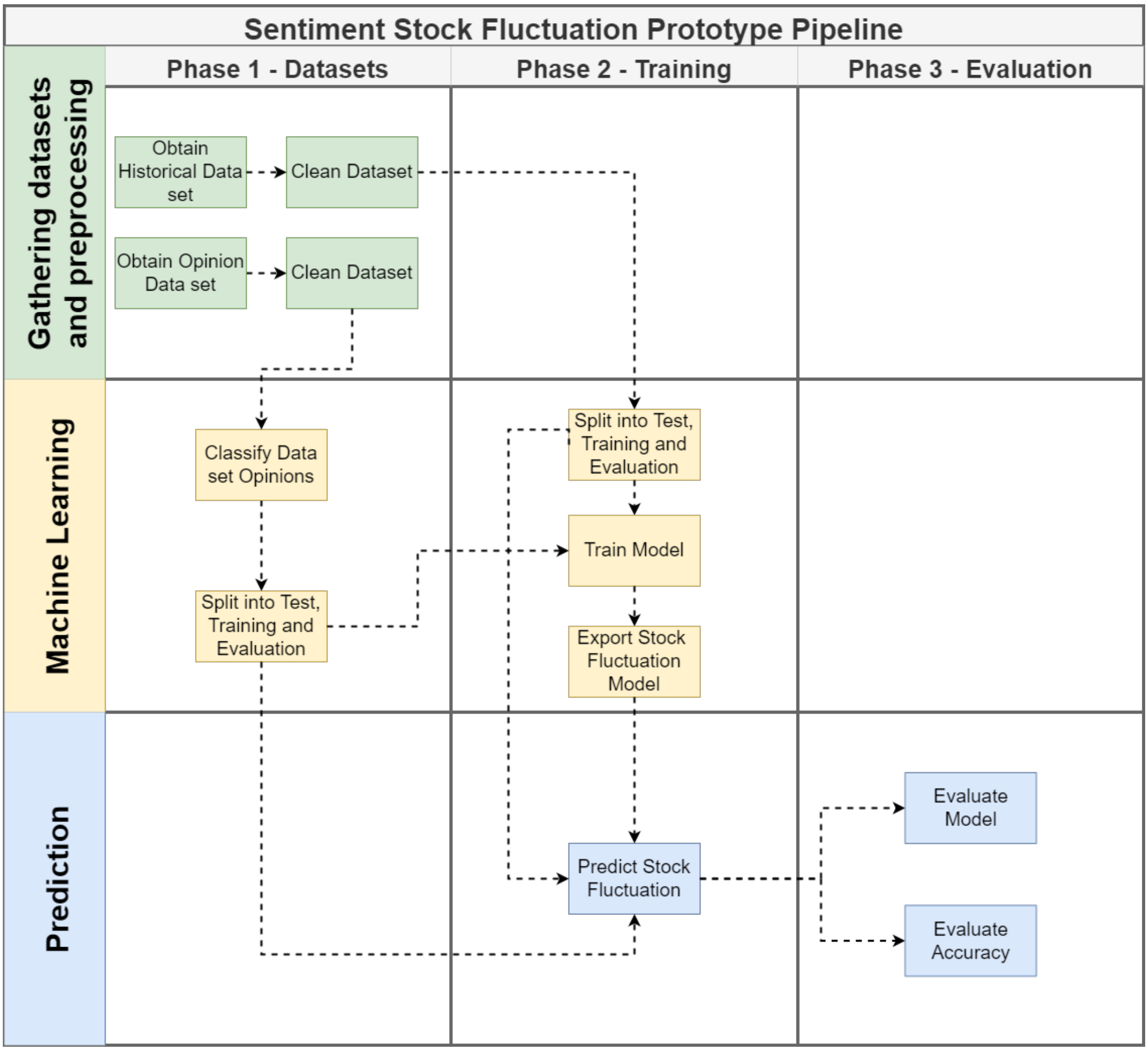


Figure 3.1: The different phases of the prototype pipeline

using a sentiment classifier. The two data sets are then combined and divided into training and testing data sets. The models are then trained using the previously created data sets, and the model is exported, saved and evaluated.

3.3 Tools and Packages

This study's research used a variety of software packages to aid the process, including a programming language, an integrated development environment, and libraries/packages to be used with the programming language.

3.3.1 Programming Language

To develop such system, most of the work was done by using a programming language called Python. As stated in a book about Python Machine Learning by Raschka, Python is one of the most popular programming languages for data science. It is powerful and benefits from an abundance of helpful add-on libraries built by its thriving community. It provides us with an environment where we can quickly jot down our ideas and put concepts into action [1].

3.3.2 Integrated Development Environment(IDE)

The prototype development for this research was developed on a IDE called "Visual Studio Code," as it can be customized for case by case scenarios. Thanks to this functionality, it was used for this study because it enables the installation of plugins such as Jupyter notebook, which allows the processing of Python code in individual cells. Additionally, it is free, extremely fast, lightweight, and cross-platform compatible [2].

3.3.3 Virtual Environment

To keep things separate from other Python environments, a virtual environment tool called "Anaconda" was utilized to isolate the Python environment of this study separate from other projects, thereby minimizing conflicts and keeping the modules in use separate [3].

---

[1] Python: https://www.python.org
[2] Visual Studio Code: https://code.visualstudio.com
[3] Anaconda: https://www.anaconda.com

3.3.4 Packages

Packages help in making our tasks easier and rapidly prototyping an idea. For this research, two major packages were used to create the machine learning models:

Scikit Learn

Scikit-learn is a free Python library mainly used for machine learning purposes. It includes a variety of models for classification, regression, and clustering techniques, such as support vector machines, random forests, gradient boosting, and k-means [4].

Keras

Keras is an open-source high-level neural network API intended to be simple to use and allow for rapid experimentation. It enables users to train and test their models without worrying about the underlying algorithms' low-level specifics. *Keras* is a Python library that can be used with other frameworks such as TensorFlow, Theano, or the Microsoft Cognitive Toolkit (CNTK)[5].

3.4 Data Collection of public opinion data sets

Three large multi-national companies were chosen for the purposes of this study, each of which focuses on a distinct market segment. Microsoft is primarily known for providing software, McDonald's is known for selling foods, and Tesla is known for selling automobiles.

3.4.1 Microsoft

Microsoft is an American multinational technology company that makes computer software, consumer electronics, personal computers, and other products and ser-

[4]Scikit: https://scikit-learn.org
[5]Keras: https://keras.io

vices. Windows, Microsoft Office, and Microsoft Edge are the most well-known software products. Microsoft is also well-known for its hardware products, including the Xbox video game system and the Microsoft Surface touchscreen computer[6].

3.4.2 Tesla

Tesla is an electric car and sustainable energy corporation based in Austin, Texas. Tesla designs and manufactures electric vehicles, solar panels, solar roof tiles, and associated goods and services. Tesla is one of the most valuable corporations globally and continues to be the most valuable carmaker in the world, with a market capitalization of about US$1 trillion [7].

3.4.3 McDonalds

McDonald's is an international fast food corporation founded by Richard and Maurice McDonald in 1940 as a restaurant. They later rebranded their company as a hamburger shop and later on franchised it. McDonald's is the world's biggest restaurant chain in sales, serving more than 69 million people daily in more than a hundred countries[8].

3.5 Social Media Platforms for data harvesting

For this study, two social media platforms were used StockTwits[9] and Reddit. Opinion data about a stock are gathered by using the stock ticker symbols "MCD", "MSFT", and "TSLA".

Twitter was initially intended to be utilized for this research but had to be excluded due to Twitter's API being limited to only retrieving data of the previous 30 days, meaning that only a limited amount of tweets could be obtained. To

[6] Microsoft: https://www.microsoft.com
[7] Tesla: https://www.tesla.com
[8] McDonald's: https://www.mcdonalds.com
[9] StockTwits: https://stocktwits.com

circumvent this restriction, one has to apply for the *Academic Research level*. This level was applied for and unfortunately got refused by Twitter as the research was considered to be ineligible. A predefined date range is set for this experiment from 07-01-2020 until 22-03-2022.

3.5.1 Reddit

Reddit was picked as a social media platform for harvesting posts via Pushshift's API[10]. Since 2015, Pushshift has harvested and made available Reddit data to researchers as part of its social media data gathering, analysis, and archiving platform. Pushshift's Reddit data set is constantly updated and contains historical data dating all the way back to Reddit's founding (Baumgartner et al. 2020). The data was gathered from the Pushshift's API using a library called *PMAW: Pushshift Multithread API Wrapper*[11]. This package enables us to obtain faster results through the use of multi-threading, as the data retrieval process requires thousands of API calls. Other libraries, such as *PRAW: The Python Reddit API Wrapper* and *PSAW: The Python Pushshift.io API Wrapper*, can accomplish the same task but slower due to their sequential execution, which would make the task process much longer to complete.

3.5.2 StockTwits

StockTwits was picked as it is a social media platform similar to Twitter but targeted at stock and crypto-currency investors. StockTwits limits its search that it can only be used by entering the stock or crypto ticker. To gather data from StockTwits, a custom Python script was created to collect data from their API. Due to the fact that each API request returns back only 1000 posts, a loop had to be created to retrieve data up to the chosen start date.

[10]Pushshift API: https://api.pushshift.io
[11]PMAW: Pushshift Multithread API Wrapper: https://github.commattpodolak/pmaw

3.5.3 Data Storage

Finally, the collected opinion data is saved in comma-separated values (CSV) files named after the company and then imported into a MongoDB database. It is then sanitized and categorized further down the pipeline.

MongoDB

MongoDB is a NoSQL database management program that provides: high performance, high availability, and easy scalability. *MongoDB* stores documents in a JSON-like format with optional schemas[12].

Additionally, historical stock price data is kept in CSV files labeled after the company's ticker and used to train the fluctuation model later.

For StockTwits, entries from the specified date range were scraped. However, the data was decreased due to cleaning as some entries were blank or contained only photos. For Reddit the same date range was set between January 7$^{\text{th}}$, 2020 to March 22$^{\text{nd}}$, 2022.

Table 3.1: Number of entries collected

Dataset	Total	Microsoft	Tesla	McDonalds
Reddit	3429248	222782	2524183	682283
StockTwits	8728974	2912843	2873511	2942620

From Reddit, 3,429,248 records were collected using PushShift's public API service, and 8,728,974 records were managed to be collected from the StockTwits platform. For model training and testing, these data sets were split into 80% for training and 20% for testing.

[12]MongoDB: https://www.mongodb.com

3.6 Stock Price History Data set collection

The Yahoo Finance API is frequently used to obtain historical stock price data because it provides various types of information, including opening and closing prices, among other elements. The information obtained from this service is critical for forecasting fluctuations. A Python package called "yfinance" [13] was utilized since it simplifies the process of retrieving historical stock data. Yahoo Finance gives daily stock information but not hourly. A free service named "AlphaVantage" was used to download the last two years' worth of hourly stock price fluctuations to obtain hourly data. A series of API calls had to be called to retrieve two whole years of data by using a python script to automate the process. The data gathered is then saved into CSV files to be later used for training the fluctuation model.

3.7 Pre-processing of data opinion data sets

Before data can be used for training the model, it first must be pre-processed. This entails converting raw data received from various sources into a format that the model can interpret.

3.7.1 Data cleaning of data sets

The following techniques are accomplished in preparation for the model to classify the entries, including a few techniques advocated by Symeonidis et al.: Table 3.2

[13] yfinance: https://pypi.org/project/yfinance

Table 3.2: Pre-Processing Techniques employed

Number	Pre-Processing Technique
1	Convert to Lowercase
2	Remove URLs
3	Remove Special Characters
4	Remove Numbers
5	Remove Mentions
6	Remove Social media Abbreviations (RT, FAV, Via,...)
7	Remove Duplicates
8	Remove Entries that are not in English
9	Replace Emotes with words
10	Replace Abbreviations with full word

3.7.2 Sentiment Labeling

Valence Aware Dictionary and sEntiment Reasoner (VADER) is a state-of-the-art sentiment analysis tool that is built on utilizing a vocabulary and rule-based sentiment analysis tool that is specifically developed to analyze sentiment in social media, and it was used to label the data set in this study. The VADER system was chosen for this project since it is quick, efficient, and reliable. VADER also does not require any additional training or fine-tuning because it is ready to use right out of the box (Hutto & Gilbert 2014).

Confusion matrices containing True Positive(TP), True Negative(TN), False Positive(FP), and False Negative(FN) values are used to determine the accuracy. The accuracy of the approaches is calculated by using the classified value and the actual value.

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + f_n + f_p + T_n} \tag{3.1}$$

3.7.3 Data Overview and Aggregation

For this research, data aggregation is required to train the model. For every hour, a new entry is created in a new database collection that consists of eleven fields: date, hour, company name, stock volume, open, close, total entries, total sentiment polarity, positive, negative, and neutral. These fields are to be utilized by the machine learning model to train.

Table 3.3: Data set of Microsoft from Reddit before aggregation

Original Text	**Cleaned Text**	**Sentiment Polarity**
DUMPING MSFT IS TREASON	dumping msft is treason	-0.6369
Can't go wrong with some good ol' $msft 200c exp 3/20	cannot go wrong with some good ol c exp	0.6656
Anyone who doesn't know about ALLY 3/20 32c or MSFT...	anyone who doesn t know about ally c or msft....	0.0

Table 3.4: Reddit Hourly aggregated data set sample for Microsoft stock

Hour	**Volume**	**Close**	**Open**	**# Entries**	**Sentiment Compound**	**# +Ve**	**# -Ve**	**# Neu**
05	64320	149.057	145.512	16	0.203	8	4	4
06	32428	149.548	149.057	2	0.4455	1	0	1
07	33135	148.448	149.263	3	-0.295	1	2	0

Table 3.5: Reddit Daily aggregated data set for the Microsoft stock

Date	**Volume**	**Close**	**Open**	**# Entries**	**Compound Sentiment**	**# +Ve**	**# -Ve**	**# Neu**
2020-03-03	71030800	169.669	162.324	1948	0.258	1169	452	327
2020-03-04	71677000	161.539	170.661	1921	0.291	1196	396	329
2020-03-05	49814400	167.470	165.447	1901	0.287	1159	434	308

Table 3.6: Reddit Weekly aggregated data set for the Microsoft stock

Date	**Volume**	**Close**	**Open**	**# Entries**	**Compound Sentiment**	**# +Ve**	**# -Ve**	**# Neu**
2020-03-07	31193500	185.331	199.744	4116	0.117	2004	1301	811
2020-03-14	25997100	189.808	184.988	3612	0.120	1761	1135	716
2020-03-21	36370700	169.067	179.127	3271	0.109	1590	1061	620

3.8 Stock Price Fluctuation Classification

Several machine learning methods are implemented and evaluated to get the most accurate model for the part on stock price fluctuation classification. It was decided that the model would have four predictor fields: positive and negative counts, as well as both positive and negative counts and the stock volume. The closing price serves as the target field. The data is normalized before the models use it, and this is accomplished by using a function named "StandardScaler" from the Scikit learn Library (Pedregosa et al. 2011). Two different scalers are created, one for the predictors and the other for the target variables. After that, the two scalers are fitted so that they can be normalized and converted. The data is then transformed into a time series, in which the number of entries in the previous series is used to forecast the next target in the series. After the data has been transformed, the data set is divided into a training portion and a testing portion with a ratio of 80 and 20.

Table 3.7: Normalized sentiment analysis feature set example

PosNeg Polarity	Pos + Neg	Change %	Main %	Target
-0.14328	-0.09999	-0.48999	0.70668	0
-0.27406	1.28662	-0.22271	0.51732	1
-0.13576	0.69235	-0.35853	0.69081	1

The machine learning algorithms to be compared are Recurrent neural networks, Bidirectional Long short-term memory, and Convolutional Neural Networks.

3.8.1 Bidirectional LSTM

Scikit-Learn and Keras were used to develop the Recurrent neural network Bidirectional Long short-term memory model because it is simple to use and implement the framework. The model comprises four layers with 50 neurons each, with a dropout of 0.2 between each layer, and is finally condensed into a single neutron for output. Afterward, the model is compiled using Adam as an optimizer and a learning rate of 0.001.

3.8.2 CNN

For the development of the Convolutional Neural Network, the Scikit-Learn and Keras were both also utilized. The model comprises a single 1D convolution with a filter of 32 units, a kernel size of 1, and activation of relu. Then by using a sigmoid activation, the model is compacted into a single unit. Finally, the model is compiled, and the loss is set to "binary_crossentropy", optimized with adam. For the metrics, accuracy is used. Finally, the model is compiled, the loss is set to "binary crossentropy", and the optimization is done with adam. The metrics are calculated using accuracy as a measure of precision.

3.8.3 Metrics

Several metrics are used to evaluate the models trained with different data sets and time window ranges: Accuracy, Precision, Recall, and F1 Score.

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + f_n + f_p + T_n} \tag{3.2}$$

$$Precision = \frac{T_p}{T_p + F_p} \tag{3.3}$$

$$Recall = \frac{T_p}{T_p + F_n} \tag{3.4}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{3.5}$$

A confusion matrix Table including True Positive(TP), True Negative(TN), False Positive(FP), and False Negative(FN) values are used to determine the Accuracy, Precision, Recall, and F1 Score.

Table 3.8: Example of Confusion Matrix

		Predicted	
		Positive	**Negative**
Actual	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (FN)

Figure 3.2: LSTM Model

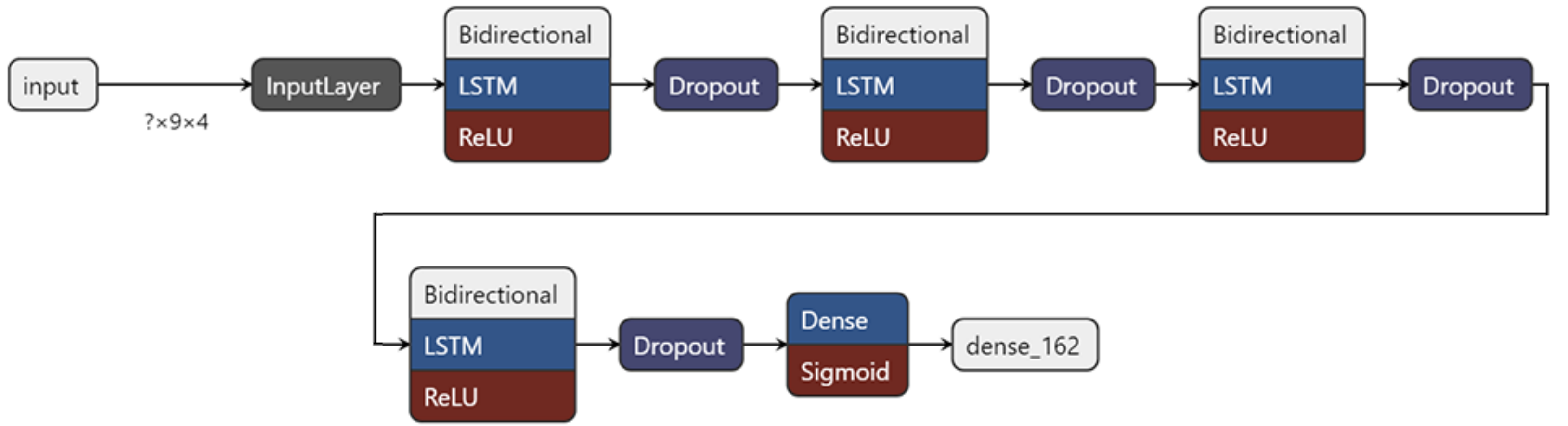
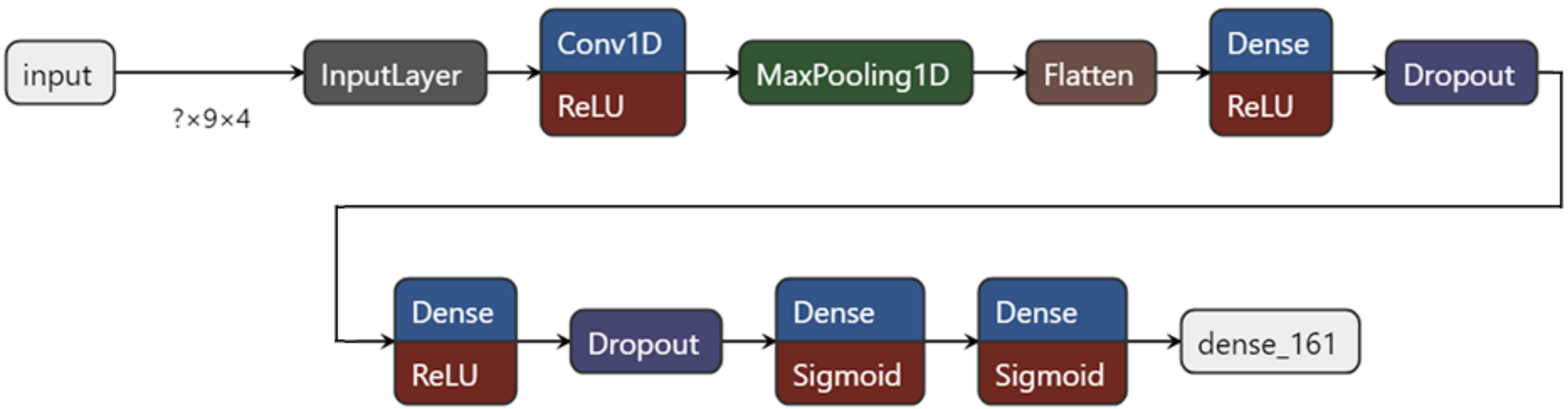


Figure 3.3: CNN Model



3.9 Conclusion

Different research approaches for sentiment analysis classifier and stock market fluctuation prediction were explored in this section. Two distinct data sets are to be collected, one for the opinion data set and another for the historical price history. Before training the model, the opinion data set must be pre-processed. Following that, the sentiment of each entry is classified using VADER, a lexicon-based sentiment analysis classifier. Furthermore, both data sets must be aggregated into hourly, daily, and weekly time periods to discover the best accurate window for predicting the subsequent fluctuation. The data set is then split to train and test the models. Finally, the data is trained using both bidirectional LSTM and CNN.

Chapter 4

Analysis of Results & Discussion

4.1 Introduction

In Chapter 3, the various research methodologies were discussed and explained in some detail. Throughout this chapter, the analysis and discussion of the findings will be presented. In Section 4.2, the results of the sentiment analysis approaches are presented, followed by a discussion in Section 4.4, where the various approaches are compared and the findings are analyzed in order to determine the most accurate and efficient sentiment analysis classifier. In Section 4.3, the findings of the models used to predict the fluctuations of the stock market are presented for each of the time intervals within the study, as well as a baseline for comparison purposes. The models that were used in this study are then discussed and compared in Section 4.5, which includes a discussion about which time range is the most accurate and which platform delivers the most accurate data from the acquired data sets.

4.2 Results of Sentiment Analysis

Two methods are tried and tested for sentiment analysis: VADER, a lexicon-based sentiment analysis tool, and roBERTa sentiment analysis model, a roBERTa-based sentiment analysis tool trained on Twitter tweets.

4.2.1 Comparison of sentiment analysis methods

A pre-labeled IMBD sentiment data set by Maas et al. is utilized to compare the techniques. The data set contains 40,000 entries in total, but only 10,000 samples were used for this benchmark. The entries selected for sampling were evenly distributed, with a total of 10,000 entries divided between negative and positive categories. (Maas et al. 2011)

To calculate the accuracy of both models, a confusion matrix was used, and by using formula 3.1, the accuracy can be calculated by the sum of the two correct predictions (True Positive and True Negative) divided by the total number of observations in the data set.

Table 4.1: Comparison of sentiment analysis classifiers on 10,000 pre labeled data set

Method	**Text Cleaned**	**Time Taken (Seconds)**	**Accuracy**
VADER	No	163.2	67.6%
VADER	Yes	253.4	67.5%
roBERTa	No	1471.1	48.2%
roBERTa	Yes	1301.6	48.9%

Table 4.2: Confusion Matrix of Vader without Clean

		Predicted	
		Positive	**Negative**
Actual	**Positive**	4305	2542
	Negative	695	2458

Table 4.3: Confusion Matrix of Vader with Clean

		Predicted	
		Positive	**Negative**
Actual	**Positive**	4306	2560
	Negative	694	2440

Table 4.4: Confusion Matrix of roBERTa without clean

		Predicted	
		Positive	**Negative**
Actual	**Positive**	3427	3538
	Negative	1581	1462

Table 4.5: Confusion Matrix of roBERTa with clean

		Predicted	
		Positive	**Negative**
Actual	**Positive**	3419	3419
	Negative	1581	1402

4.3 Results of Stock Market Fluctuation Prediction

In order to find the optimal number of epochs and the optimal time frame for the time series fluctuation prediction, a number of models were created and tested to find the optimal combination. A significant number of models were created for each stock for every time period (Hourly, Daily, Weekly).

The accuracy of the models is calculated by plotting the results in a confusion matrix and calculating the accuracy by using the Formula 3.2 by first summing up the True Positives and True Negatives together and finally is divided by the total number of entries in the data set.

4.3.1 Optimal Number of Epochs

The number of epochs also affects the model's accuracy. To accomplish this, each model is calculated using the default optimal parameters.

The results obtained indicate that the average best epoch for all models is 25 across all platforms and stocks.

4.3.2 Optimal number of entries in the time window for time series

When using a time series model, the time window range is critical, and in order to find the optimal window range, a variety of models were generated for each platform. An average of all outcomes was calculated to determine the optimal window range.

Table 4.6: Comparison of average accuracy grouped by time window (Hours, Days, Weeks)

Time Window	**Hourly**	**Daily**	**Weekly**
1	48%	**51%**	**50%**
2	48%	**51%**	**50%**
3	49%	50%	49%
4	48%	50%	47%
5	48%	50%	47%
6	48%	49%	47%
7	49%	49%	47%
8	51%	49%	47%
9	**54%**	49%	47%
10	51%	48%	46%

As shown in Table 4.6, the time window range for hourly fluctuations predictions makes a significant difference; based on the results, the optimal time window range for hourly across all stocks and social media platforms was determined to be by using the previous 9 hours to predict the following fluctuation with a 54 percent accuracy. One can observe an incremental increase in accuracy as one approaches the ninth hour and, subsequently, a decline in accuracy as more than nine hours are used.

The time window range for daily variations forecasts in Table 4.6 reveals that having a single or two-day time range provides the highest average accuracy compared to the others, as the average accuracy decreases as the time range increases.

Table 4.6 shows that weekly predictions exhibit behavior that is similar to that of daily predictions. When a time period of one or two weeks is chosen, the accuracy appears to be the highest on a weekly basis. When there are more than two elements in the time range, the accuracy of the prediction begins to decline.

4.3.3 Training the Models

After determining the appropriate number of epochs and the appropriate number of entries in the time window for the time series, the models are trained using the optimal number of epochs and the best time window for each time range.

Each model is trained and tested five times in order to get the average accuracy of the models. A few features from the Stock Price History Data Set were used as a baseline, including the open, close, and volume features. These attributes serve as a baseline against which sentiment analysis features are compared.

Tables 4.7, 4.11 and 4.13 display the accuracy of each stock company under investigation for this study, as well as the social media sites from which the data set was gathered displayed hourly, daily, and weekly. In order to predict market movements over the following hour, day, and week, aggregated sentiment analysis results (positive and negative change, total posts count, sentiment polarity change between entries, difference between positive and negative) and stock market features (open, close and volume) are used to produce this accuracy.

Tables 4.8, 4.12 and 4.14 further demonstrate the accuracy of each stock company that was utilized in this research, given the fact that only three features were used to predict the target: open, close, and volume. This was done so that it would be possible to determine whether the sentiment analysis data provided in Tables 4.7, 4.11 and 4.13 is enhancing the model's accuracy or if it has no effect.

Table 4.7: Hourly Fluctuation Prediction Accuracy using Sentiment Analysis and Baseline Features

Company	Source	Model	Min Accuracy	Max Accuracy	Accuracy
McDonald's	Reddit	BiLSTM	50%	50%	50%
McDonald's	Reddit	CNN	50%	50%	50%
McDonald's	StockTwits	BiLSTM	52%	52%	52%
McDonald's	StockTwits	CNN	52%	52%	52%
Microsoft	Reddit	BiLSTM	50%	50%	50%
Microsoft	Reddit	CNN	50%	50%	50%
Microsoft	StockTwits	BiLSTM	50%	50%	50%
Microsoft	StockTwits	CNN	50%	50%	50%
Tesla	Reddit	BiLSTM	52%	52%	52%
Tesla	Reddit	CNN	52%	52%	52%
Tesla	StockTwits	BiLSTM	52%	52%	52%
Tesla	StockTwits	CNN	52%	52%	52%

As can be seen in Table 4.7, the accuracy of all stocks on all platforms in the model accuracy is very low, and they might be making random guesses at either 1 or 0.

Table 4.8: Hourly Fluctuation Prediction Accuracy using Baseline Features

Company	Model	Min Accuracy	Max Accuracy	Accuracy
McDonald's	BiLSTM	50%	50%	50%
McDonald's	CNN	50%	50%	50%
Microsoft	BiLSTM	50%	50%	50%
Microsoft	CNN	50%	50%	50%
Tesla	BiLSTM	**52%**	**52%**	**52%**
Tesla	CNN	**52%**	**52%**	**52%**

The accuracy of Table 4.8 is similar to the accuracy of Table 4.7 when only the baseline attributes are considered. This shows that the sentiment analysis features were not a substantial help in anticipating the volatility of the next hour.

Table 4.9: Bi Hourly Fluctuation Prediction Accuracy using Sentiment Analysis and Baseline Features

Company	Source	Model	Min Accuracy	Max Accuracy	Accuracy
McDonald's	Reddit	BiLSTM	48%	52%	51%
McDonald's	Reddit	CNN	46%	54%	51%
McDonald's	StockTwits	BiLSTM	39%	54%	48%
McDonald's	StockTwits	CNN	45%	**59%**	**53%**
Microsoft	Reddit	BiLSTM	49%	54%	51%
Microsoft	Reddit	CNN	49%	55%	52%
Microsoft	StockTwits	BiLSTM	49%	53%	51%
Microsoft	StockTwits	CNN	47%	52%	50%
Tesla	Reddit	BiLSTM	**52%**	52%	52%
Tesla	Reddit	CNN	50%	54%	**53%**
Tesla	StockTwits	BiLSTM	45%	53%	50%
Tesla	StockTwits	CNN	50%	51%	52%

The accuracy of the models when trained on numerous distinct equities and social media sites can be seen in Table 4.9. The Table shows that there was a slight improvement as compared to the hourly accuracy. Although the minimum accuracy is lower than the minimum hourly accuracy, the average and maximum accuracy achieved were greater.

Table 4.10: Bi Hourly Fluctuation Prediction Accuracy using Baseline Features

Company	Model	Min Accuracy	Max Accuracy	Accuracy
McDonald's	BiLSTM	49%	51%	50%
McDonald's	CNN	46%	**53%**	50%
Microsoft	BiLSTM	50%	**53%**	**52%**
Microsoft	CNN	**52%**	53%	**52%**
Tesla	BiLSTM	50%	51%	50%
Tesla	CNN	46%	53%	50%

When comparing the baseline feature set to the baseline and sentiment analysis feature sets, it is clear that the sentiment analysis feature set significantly improves model accuracy compared to the baseline feature set findings presented in Table 4.10.

Table 4.11: Daily Fluctuation Prediction Accuracy using Sentiment Analysis and Baseline Features

Company	Source	Model	Min Accuracy	Max Accuracy	Accuracy
McDonald's	Reddit	BiLSTM	50%	55%	51%
McDonald's	Reddit	CNN	51%	51%	51%
McDonald's	StockTwits	BiLSTM	48%	51%	49%
McDonald's	StockTwits	CNN	48%	48%	48%
Microsoft	Reddit	BiLSTM	51%	55%	51%
Microsoft	Reddit	CNN	51%	51%	51%
Microsoft	StockTwits	BiLSTM	49%	61%	54%
Microsoft	StockTwits	CNN	50%	50%	50%
Tesla	Reddit	BiLSTM	49%	55%	52%
Tesla	Reddit	CNN	52%	52%	52%
Tesla	StockTwits	BiLSTM	52%	**66%**	**57%**
Tesla	StockTwits	CNN	**53%**	53%	53%

On the two different platforms, the accuracy of daily predictions is higher than the accuracy of hourly and bi-hourly predictions when predicting the volatility of the stock market in selected companies for the following day. Although the accuracy of the daily prediction results, as seen in Table 4.11 they are not exceptionally high, they have been improved when compared to the bi-hourly forecast results in Table 4.9.

Table 4.12: Daily Fluctuation Prediction Accuracy using Baseline Features

Company	Model	Min Accuracy	Max Accuracy	Accuracy
McDonald's	BiLSTM	48%	52%	49%
McDonald's	CNN	48%	48%	48%
Microsoft	BiLSTM	50%	50%	50%
Microsoft	CNN	50%	50%	50%
Tesla	BiLSTM	**53%**	**53%**	**53%**
Tesla	CNN	**53%**	**53%**	**53%**

When Tables 4.11 and 4.12 are compared, it becomes clear that the accuracy of the Table that incorporates sentiment analysis, as well as baseline features, is substantially more significant than the accuracy of the Table that contains baseline features alone. Therefore, the sentiment analysis feature set improves the accuracy of predicting whether the next day's movements will be positive or negative.

Table 4.13: Weekly Fluctuation Prediction Accuracy using Sentiment Analysis and Baseline Features

Company	Source	Model	Min Accuracy	Max Accuracy	Accuracy
McDonald's	Reddit	BiLSTM	**57%**	**74%**	**65%**
McDonald's	Reddit	CNN	43%	43%	43%
McDonald's	StockTwits	BiLSTM	33%	58%	46%
McDonald's	StockTwits	CNN	50%	50%	50%
Microsoft	Reddit	BiLSTM	43%	65%	51%
Microsoft	Reddit	CNN	48%	48%	48%
Microsoft	StockTwits	BiLSTM	46%	66%	55%
Microsoft	StockTwits	CNN	54%	54%	54%
Tesla	Reddit	BiLSTM	39%	57%	49%
Tesla	Reddit	CNN	48%	52%	49%
Tesla	StockTwits	BiLSTM	46%	71%	56%
Tesla	StockTwits	CNN	50%	53%	50%

As can be seen in Table 4.13, the accuracy is significantly improved when data is aggregated into weekly intervals rather than hourly, bi-hourly and daily intervals. In general, when predicting the next event, the BiLSTM model outperformed the CNN model. The accuracy of BiLSTM was 54 percent on average, while CNN's accuracy was 49 percent in the same test.

Table 4.14: Weekly Fluctuation Prediction Accuracy using Baseline Features

Company	Model	Min Accuracy	Max Accuracy	Accuracy
McDonald's	BiLSTM	50%	63%	51%
McDonald's	CNN	48%	50%	49%
Microsoft	BiLSTM	48%	52%	48%
Microsoft	CNN	48%	48%	48%
Tesla	BiLSTM	48%	**65%**	**53%**
Tesla	CNN	**53%**	**53%**	**53%**

When comparing the results of the sentiment analysis with the baseline feature set and the other with just the baseline feature set, the one with sentiment analysis helped improve the accuracy of the models. Analysing Table 4.13, it can be seen that the average accuracy improved when compared to Table 4.14. In some instances, the overall accuracy is enhanced by 14 percentage points. However, it had the opposite effect on others and decreased the accuracy by six percentage points.

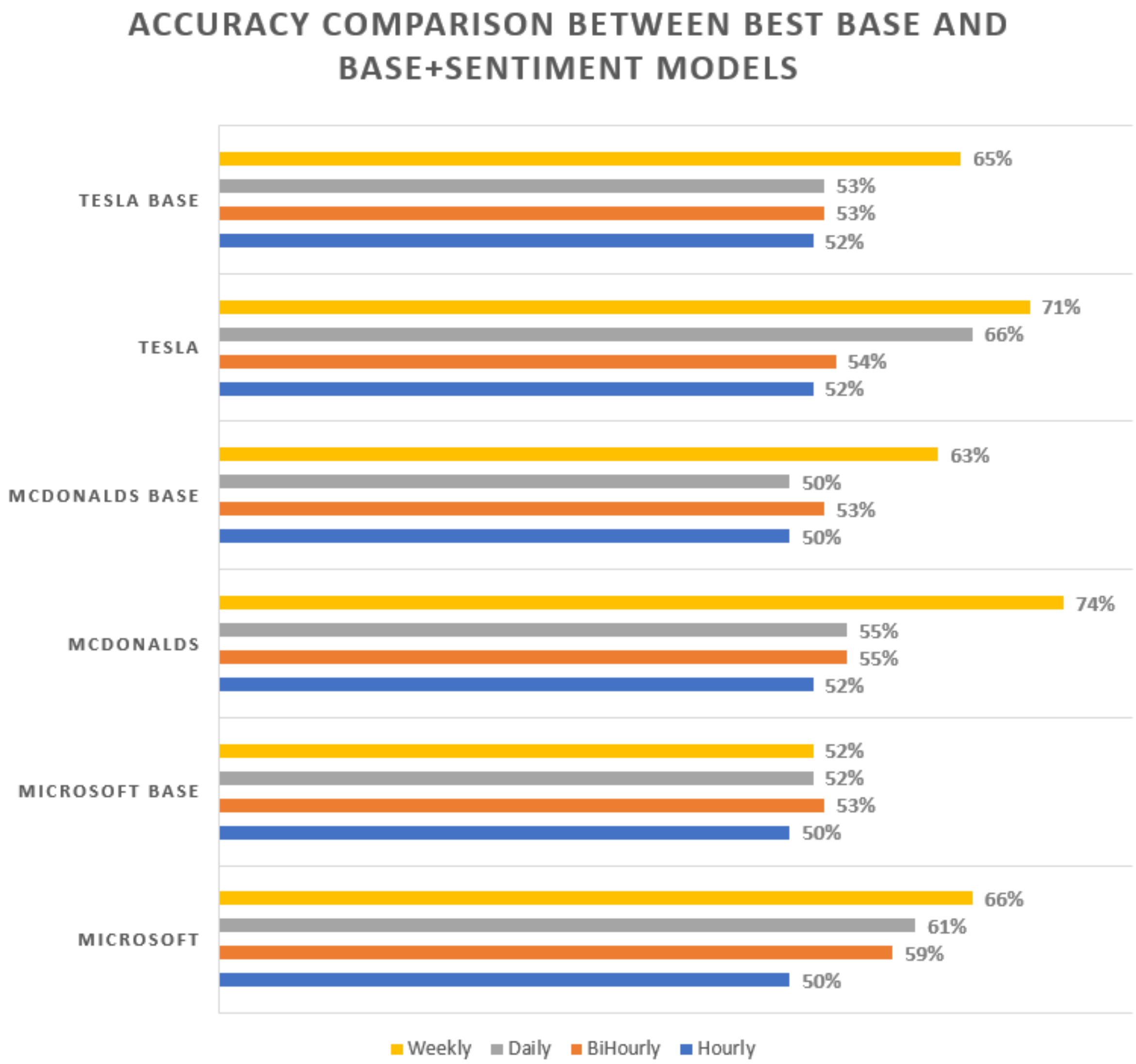


Figure 4.1: Accuracy comparison between best models using baseline and baseline+sentiment data sets

Table 4.15: F1 Scores of best performing models

Company	**Platform**	**Model**	**Hourly**	**Bi Hourly**	**Daily**	**Weekly**
McDonald's	Reddit	BiLSTM	**0.6619**	**0.6349**	0.6375	**0.7143**
McDonald's	Reddit	CNN	**0.6619**	0.5855	0.6335	0.6061
McDonald's	StockTwits	BiLSTM	0.6351	0.5893	0.6627	0.6667
McDonald's	StockTwits	CNN	0.6351	0.507	**0.6509**	0.6667
Microsoft	Reddit	BiLSTM	**0.6887**	0.6662	0.6747	0.6667
Microsoft	Reddit	CNN	**0.6887**	0.6019	0.6747	0.6471
Microsoft	StockTwits	BiLSTM	0.6661	**0.7046**	**0.7143**	**0.7429**
Microsoft	StockTwits	CNN	0.6661	0.6006	0.6667	0.7027
Tesla	Reddit	BiLSTM	0.5093	0.5207	0.6914	0.6471
Tesla	Reddit	CNN	0.5013	0.5751	0.6826	0.6471
Tesla	StockTwits	BiLSTM	**0.6605**	**0.5825**	**0.7101**	**0.6667**
Tesla	StockTwits	CNN	**0.6605**	0.5824	0.6897	**0.6667**

Table 4.15 shows the F1 Scores of the best performing models. From the collected results, the lowest F1 Score in the entire Table was 0.5013, which occurred when predicting hourly using CNN on the Tesla Stock using the Reddit data set. The highest F1 Score was 0.7429, which occurred when predicting weekly using BiLSTM on the McDonald's stock using the StockTwits data set.

Table 4.16: Comparison of stock prediction models

Model	Min Accuracy	Max Accuracy	Avg Accuracy	Avg F1 Score
BiLSTM	33%	**74%**	**52%**	0.5521
CNN	**43%**	54%	51%	**0.5731**

In Table 4.16, all of the experiments that were conducted for this test have been compiled into one Table, and an average has been determined in order to establish which of the two models is the more correct one. The findings established that the bidirectional LSTM model had one percent greater accuracy, which is not a huge difference. In comparison, CNN was only able to achieve maximum accuracy of 54 percent, while BiLSTM successfully achieved a maximum accuracy of 74 percent. The final Score that CNN achieved was 0.5731 for the F1 test, whereas the Score that the bidirectional LSTM model achieved was 0.5521.

Table 4.17: Comparison of social media platforms to predict stock market fluctuation

Platform	Min Accuracy	Max Accuracy	Avg Accuracy	Avg F1 Score
StockTwits	33%	71%	**54%**	0.5597
Reddit	**39%**	**74%**	50%	**0.5961**

The social media platforms StockTwits and Reddit that were selected for this investigation are compared head-to-head in Table 4.17. From the data collected and presented in the Table, the social media network StockTwits achieved an overall accuracy that was 4 percent higher than that of Reddit.

4.4 Discussion on Sentiment Analysis

From Table 4.1 it can be observed that the two different methods performed differently. VADER took the lead with 67.6 percent accuracy when not cleaned and 67.4 percent when cleaned compared to the roBERTa based model, which got a low accuracy of 48.8 percent when using a not cleaned data set and 48.2 percent when using the cleaned data set. There is an 18.8 percent difference in accuracy when comparing both approaches. The roBERTa technique managed to take 1471.1 seconds to process 10,000 entries, while VADER managed to process and classify the same data in 253.4 seconds. It takes 5.8x longer to process and classify than VADER.

Since 12 million entries were collected, classifying the entire data set obtained beforehand would require a significant amount of time. Additionally, processing, cleaning, and analyzing each input would demand substantial time and resources complete. For this method to be used in an active prediction system, the data must first be gathered, pre-processed, and then analyzed to predict the subsequent stock market fluctuation. This kind of processing would take a considerable amount of time and resources.

From the results shown in Table 4.1 one can see that the data cleaning does not help the accuracy but makes it slightly worse.

4.5 Discussion on Stock Market Fluctuation Prediction

A significant number of models were created, trained, and tested due to the experiment, and metrics were recorded. For every stock a different model is trained as it uses time series to forecast the next fluctuation. The results of these studies were used to create an average across all stock variants and social media platforms.

The data in the results section shows that when the data is aggregated, the models perform better. The hourly data, shown in Table hourly, reveal that the accuracy is so low that the model could be chosen at random. The models are having difficulty learning with such minor changes to the hourly data set, with a maximum accuracy of 52 percent across all stocks and models. When the data is aggregated bi-hourly, the accuracy climbs to a maximum of 59 percent, more significant than the hourly prediction accuracy. Taking a step back and combining the data set into daily, the most remarkable accuracy achieved was 66 percent, an increase over the Hourly and Bi-Hourly. Finally, when the data set is aggregated into weekly intervals, the maximum accuracy climbs to 74 percent, a significant improvement over hourly, bi-hourly, and daily intervals.

4.5.1 Comparison of input time window

To find the best accurate time frame for predicting price movements when it comes to hourly predictions, the results for both platforms and all stock firms were averaged together in Table 4.6 of hourly prediction results.

Using these findings, it was established that using the most recent 8 to 10 hours resulted in the maximum accuracy compared to the other 1 to 7 hours, which resulted in an accuracy of less than 50 percent. The most accurate result was nine hours, which was three percent more accurate than eight or ten.

After looking over the results of time window accuracy for daily predictions to determine the optimal time window for daily forecasts, it was discovered that utilizing the most recent two days was the most accurate when combined, and an average was calculated. When using more than two days, it was discovered that the accuracy gradually of the predictions decreased by every few days.

From the results gathered in Table 4.6 for average accuracy for predicting the following week, it was discovered that using the most recent two weeks of data to predict the fluctuations of the following week was determined to be the most accurate based on the collected data. Using more than two weeks affected the accuracy negatively.

Based on the findings obtained, it was determined that when predicting the stock movement fluctuation of the following hour, using the last nine hours was the best time range to predict the following hour's movement. It was also determined that using data from the previous one or two days was the most accurate method of projecting stock movement the next day. In order to anticipate week-to-week fluctuations, it was discovered that the same scenario as for predicting daily applied, using the previous one or two weeks was the most accurate method.

From the four different time interval windows, it is noticed that predicting the following week was more accurate than predicting the following hour, bi-hour, or day.

4.5.2 Comparison between models

From the various experiments performed using different stocks and data sets from different social media platforms, the results show that the Bidirectional LSTM models always managed to be the most accurate compared to the other CNN models with the configuration that was mentioned in Section 3.8.2.

Based on the findings presented in Table 4.16, the Bidirectional LSTM model achieved the highest level of accuracy, scoring an average of 52 percent with an F1 score of 0.5521 across the board when all stocks and social media sites were taken into consideration. The model's accuracy ranged from a high of 74 percent to a low of 33 percent.

When all stocks and social media platforms are taken into account, the CNN models reached an average accuracy of 51 percent. The models could only achieve a maximum of 54 percent and a minimum of 43 percent of the possible outcomes. The total number of F1 Scores yielded an average of 0.5731.

When placing both models side by side and comparing them in Table 4.16, it is clear that the bidirectional LSTM models performs slightly better than the CNN models. A similar result can be found in similar study such as (Critien et al. 2022) where the BiLSTM model had the best mean accuracy compared to CNN. The BiLSTM models had the highest maximum accuracy overall and had the highest average accuracy, but it also had the lowest accuracy.

4.5.3 Comparison between social platforms

How social media users behave and interact differs from one social media platform to another. How users engage on the platforms can impact the accuracy of the models used in this research.

From the results displayed in Table 4.17, one can see that when taking into account all experiments, the social media platform StockTwits got an average of 54 percent with an average F1 score of 0.5597. Using StockTwits's opinion data set, a maximum accuracy of 71 percent was achieved, although it had a minimum accuracy of 33 percent in some scenarios.

It had an average accuracy of 50 percent across the board for the other social media platform known as Reddit, earning an average F1 score of 0.5961. A similar result occurred in Gui Jr's study, where the average accuracy turned out to be 50 percent (Gui Jr 2019). The platform achieved a maximum accuracy of 74 percent in certain conditions, while in others, it achieved a minimum accuracy of 39 percent.

When compared side by side, the accuracy of both platforms reveals that they are comparable to one another. On average, Reddit achieved both the highest minimum and highest maximum accuracy, but StockTwits achieved the most accuracy. In addition, Reddit achieved a higher F1 Score compared to StockTwits.

4.5.4 Comparison of Baseline and Sentiment Feature sets

When comparing the baseline feature set to the sentiment feature set, it is evident that the sentiment feature set increases the accuracy of the prediction, however as shown in Tables 4.8 and 4.7, the sentiment feature set is ineffective when hourly predictions are made though this only happens for hourly predictions when it comes to daily the accuracy of the baseline and sentiment feature set differs. As shown in Tables 4.12 and 4.11 when using only the baseline information, McDonald's was only able to achieve maximum accuracy of 52 percent, however, when using both the baseline information and sentiment features. The company was able to achieve maximum accuracy of 55 percent. When sentiment features are added, the accuracy increases by 61 percent when using the StockTwits data set and the BiLSTM model and by 55 percent when using the Reddit data set and the BiLSTM Model. The baseline features for Microsoft only achieved an accuracy of 50 percent, but when sentiment features are included, the accuracy increases significantly. The model achieved an accuracy of 53 percent when used to predict daily changes in Tesla's stock using only the baseline feature set. However, the model's accuracy increased to 66 percent on the StockTwits data set and 55 percent on Reddit after the sentiment feature set was introduced.

When it comes to weekly predictions, the accuracy also increases compared to daily predictions. As one can see in Tables 4.14 and 4.13 With the baseline feature set, the best model was able to forecast the following week's stock fluctuations for McDonald's with an accuracy of 63 percent. However, when the sentiment feature set was added to the baseline, the accuracy increased to 74 percent. Microsoft was only able to get a maximum accuracy of 52 percent using just baseline features. However, when sentiment features were added, the accuracy increased to 66 percent on the StockTwits data set and 65 percent on Reddit using the BiLSTM model. When sentiment features are added, Tesla's accuracy increases to 71 percent for the StockTwits data set but declines to 57 percent for the Reddit data set.

4.5.5 Comparison of stock fluctuation prediction performance

Looking at Tables 4.7, 4.11 and 4.13 the accuracy differs from one stock to another. According to the Tables, McDonald's stock was the most predictable overall, followed by Tesla and Microsoft. When it comes to hourly predictions, all stocks performed poorly, and the accuracy was as low as fifty percent, indicating that the predictions could be random. Although daily predictions are more accurate, Tesla's stock fluctuation prediction had an average accuracy of 57 percent and a maximum accuracy of 66 percent, followed by Microsoft with a maximum of 61 percent and an average of 54 percent. Comparing the Microsoft stock prediction to another study, (Bahceci & Alsing 2015), the author used an artificial neural network (ANN) to predict daily fluctuations and got an accuracy rate of 55 percent. Compared to this study's methodology and data set, the best model for the stock achieved an accuracy rate of 61 percent. Taking a step back to weekly intervals, the accuracy continues to increase. The accuracy of the Bidirectional LSTM models trained for each stock varies from stock to stock. When trained for McDonald's, the model attained a maximum accuracy of 74 percent using Reddit as a data set.

When trained for Tesla, a 71 percent maximum was obtained using StockTwits as a data set, while when trained for Microsoft, a 66 percent maximum was obtained using StockTwits as a data set.

4.6 Conclusion

In this chapter, a series of results were presented, and a comparison of two sentiment analysis methods was discussed in Section 4.2.1. The results of such methods are presented in Table 4.1. The tests conducted concluded that the VADER method was the best method of the methodologies tested, resulting in a 67.6% accuracy when classifying an already classified data set. For the stock market fluctuation prediction in Section 4.2.1, another set of test results was conducted and presented. The optimal number of epochs was evaluated, and it was found that 25 epochs provided the best results without overfitting even further. Several models were created to determine the optimal number of necthe time window for a time series. For hourly rates, the best time window was to use the last nine hours, one or two days daily and one or two weeks weekly. Later on, the Tables compare the Baseline feature set and Sentiment Analysis with the Baseline feature sets to determine if the sentiment analysis feature set improved the accuracy. It was found that it does indeed help improve the accuracy.

A further comparison was made between the two techniques under consideration in this study. Table 4.16 reveals that the BiLSTM models had an average accuracy of 52%, a maximum accuracy of 74%, and a minimum accuracy of 33%. In contrast, the CNN models had an average accuracy of 51%, a maximum accuracy of 54%, and a minimum of 43%. Making Bidirectional LSTM the best algorithm between the two.

Finally, a comparison of the two social media platforms was performed to establish accuracy. Table 4.17 shows that when an average is gathered, StockTwits had the highest overall accuracy of 54 percent, with an average F1 score of 0.5597.

Reddit only managed to gather 50 percent with an average F1 score of 0.5961.

Chapter 5

Conclusions & Recommendations

This study aimed to examine the application of sentiment analysis on social media platforms such as Reddit and StockTwits to predict stock market price fluctuations.

Before predicting stock market fluctuations, two NLP sentiment analysis tools VADER and a roBERTa based sentiment analysis classification model were applied and tested to determine the best tool for the job.

To predict stock market movements based on sentiment analysis, two natural language processing (NLP) sentiment analysis methods were deployed and tested, VADER and a roBERTa based sentiment analysis classification model. The performance of the technologies was measured by using an IMDB data set that had already been annotated with sentiment labels. It was discovered that the VADER model had achieved 18.8 percent greater accuracy than the roBERTa model, although at the expense of a longer processing time. From the benchmark test, the roBERTa model took x5.8 longer than VADER.

For the stock market fluctuation predictor, when the two models, Bidirectional LSTM and CNN, were compared, it was determined that the suggested approach that used Bidirectional LSTM was the superior algorithm as it outperformed the CNN approach in almost all of the different stocks, platforms and time ranges.

This study investigated three different stocks: Microsoft, McDonald's, and Tesla. The three companies each supply their services in a different sector, and the three stocks were examined in the same way. It was found that hourly fluctuations predictions are extremely low. The model could be selected at random. Although the same models are run bi-hourly, daily, or weekly, the accuracy is significantly increased. When the hourly, bi-hourly, and daily forecasts were compared to the weekly predictions, it was found that the weekly forecasts were the most accurate.

5.1 Research Questions

In Chapter 1, Section 1.2 the research questions were identified for this research study. To verify whether research question one was answered through this research, taking a look at the results in Table 4.16 it was discovered that while both CNN and Bidirectional LSTM can provide significant knowledge, they are not always dependable as they can be influenced by many variables. The model's accuracy varies according to platform, company, and time period selected (hourly, bi-hourly, daily, or weekly). However in this study Bidirectional LSTM performed the best compared to CNN.

To answer the second and final research question it was discovered that one can predict the subsequent possible fluctuation but cannot forecast more than the following fluctuation since the prediction of the future fluctuation depends on the prior fluctuations. Social media platforms such as Reddit and StockTwits can provide valuable insights into investor sentiment.

5.2 Limitations

This study has a number of limitations that impact the findings. The opinion data set is influenced by a number of causes, including the Covid-19 global pandemic and the Russian invasion of Ukraine from the 24^{th} February 2022 to the 22^{nd} March 2022, which affects the final portion of the data set. These two occurrences affect the opinion data sets. In a study about the effect of COVID-19 on social media (Li et al. 2020) the research shows that at the start of the pandemic the emotions of posts on social media involved more sadness, anger, disgust, worry and surprised while happiness decreased. This shows that it indeed affects the sentiment analysis. Another study which focuses on the Russian invasion of Ukraine (Polyzos 2022) the author mentioned that the event caused a fluctuation on social media sentiment and majority of stock market indices.

VADER, the sentiment analysis method employed in this study, was only 67.6 percent accurate compared to a pre-labeled IMDB data set. Due to its mediocre accuracy of 67.6 percent, it has a negative impact on the accuracy of the stock market fluctuation prediction, which relies on the sentiment data provided by VADER. The method's accuracy utilized in this study varies from one stock to another. Therefore the method may perform exceptionally well on one stock and data set but poorly on other stocks and data sets. Another limitation of the VADER, a sentiment analysis classifier tool, is that it identifies words as positive or negative based on a lexicon of positive and negative words. It has been observed that these platforms build their vocabulary that the sentiment analysis tool cannot comprehend or interpret.

Initially, it was planned to include Twitter in the research because it is one of the most prominent social media platforms for expressing opinions and ideas. However, not enough data could be acquired, and even if a script were created to begin gathering data, it would still be insufficient. In order to get over this restriction, one would have to apply for the "Academic Research level" by filling out an application form explaining the research idea. It was applied for and nevertheless turned down because, according to Twitter, the research that was being conducted was deemed ineligible.

5.3 Recommendations

Future research should use a more extensive data set when predicting weekly results because the data set gets smaller as more time is aggregated together. The greater the data set size, the more likely it is that the model will avoid overfitting and gain more insight into forecasting the next step. Addition of new platforms, such as Twitter, and any attempt to integrate the data sets to see whether this makes a difference would be another recommendation. In addition, more accurate NLP sentiment analysis tools should be utilized in future research since the accuracy of the NLP sentiment analysis tools being used in this study is limiting the accuracy of the fluctuation forecast accuracy in this study. The tool must be tailored to the platform. It has been observed that some communities, such as Reddit, have established their vocabulary, which current tools are unable to read, resulting in data being mislabeled appropriately. Another idea would be to filter out noise by restricting the sources of data collection to a small number of locations, such as sub-Reddit communities.

References

Alessia, D., Ferri, F., Grifoni, P. & Guzzo, T. (2015), 'Approaches, tools and applications for sentiment analysis implementation', *International Journal of Computer Applications* **125**(3).

Alkubaisi, G. A. A. J., Kamaruddin, S. S. & Husni, H. (2018), 'Stock market classification model using sentiment analysis on twitter based on hybrid naive bayes classifiers.', *Comput. Inf. Sci.* **11**(1), 52–64.

Bahceci, O. & Alsing, O. (2015), 'Stock market prediction using social media analysis'.

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. & Blackburn, J. (2020), The pushshift reddit dataset, *in* 'Proceedings of the international AAAI conference on web and social media', Vol. 14, pp. 830–839.

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

Bordo, M. D., Dueker, M. J. & Wheelock, D. C. (2008), 'Inflation, monetary policy and stock market conditions'.

Cervelló-Royo, R., Guijarro, F. & Michniuk, K. (2015), 'Stock market trading rule based on pattern recognition and technical analysis: Forecasting the djia index with intraday data', *Expert systems with Applications* **42**(14), 5963–5975.

Chen, T.-l. & Chen, F.-y. (2016), 'An intelligent pattern recognition model for supporting investment decisions in stock market', *Information Sciences* **346**, 261–274.

Chou, C., Park, J. & Chou, E. (2021), 'Predicting stock closing price after covid-19 based on sentiment analysis and lstm', **5**, 2752–2756.

Cox, J. (2021), 'Investors have put more money into stocks in the last 5 months than the previous 12 years combined'.
URL: *https://www.cnbc.com/2021/04/09/investors-have-put-more-money-into-stocks-in-the-last-5-months-than-the-previous-12-years-combined.html*

Critien, J. V., Gatt, A. & Ellul, J. (2022), 'Bitcoin price change and trend prediction through twitter sentiment and data volume', *Financial Innovation* **8**(1), 1–20.

Dhaoui, C., Webster, C. M. & Tan, L. P. (2017), 'Social media sentiment analysis: lexicon versus machine learning', *Journal of Consumer Marketing* .

Gui Jr, H. (2019), Stock prediction based on social media data via sentiment analysis: a study on reddit, Master's thesis.

Hutto, C. & Gilbert, E. (2014), Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, *in* 'Proceedings of the international AAAI conference on web and social media', Vol. 8, pp. 216–225.

Jiao, P., Veiga, A. & Walther, A. (2020), 'Social media, news media and the stock market', *Journal of Economic Behavior & Organization* **176**, 63–90.

Kearney, C. & Liu, S. (2014), 'Textual sentiment in finance: A survey of methods and models', *International Review of Financial Analysis* **33**, 171–185.

Li, X., Zhou, M., Wu, J., Yuan, A., Wu, F. & Li, J. (2020), 'Analyzing covid-19 on online social media: Trends, sentiments and emotions', *arXiv preprint arXiv:2005.14464* .

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y. & Potts, C. (2011), Learning word vectors for sentiment analysis, *in* 'Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies', Association for Computational Linguistics, Portland, Oregon, USA, pp. 142–150.

Medhat, W., Hassan, A. & Korashy, H. (2014), 'Sentiment analysis algorithms and applications: A survey', *Ain Shams engineering journal* **5**(4), 1093–1113.

Mittal, A. & Goel, A. (2012), 'Stock prediction using twitter sentiment analysis', *Standford University, CS229* **15**.

Ormos, M. & Vázsonyi, M. (2011), 'Impacts of public news on stock market prices: Evidence from s&p500'.

Pant, D. R., Neupane, P., Poudel, A., Pokhrel, A. K. & Lama, B. K. (2018), 'Recurrent neural network based bitcoin price prediction by twitter sentiment analysis', pp. 128–132.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011), 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research* **12**, 2825–2830.

Polyzos, E. S. (2022), 'Escalating tension and the war in ukraine: Evidence using impulse response functions on economic indicators and twitter sentiment', *Available at SSRN 4058364* .

Raschka, S. (2015), *Python machine learning*, Packt publishing ltd.

Schumaker, R. P. & Chen, H. (2009), 'Textual analysis of stock market prediction using breaking financial news: The azfin text system', *ACM Transactions on Information Systems (TOIS)* **27**(2), 1–19.

Sindhu, M. I., Bukhari, S. M. H., Sub-Campus, B. & Hussain, A. (2014), 'Macroeconomic factors do influencing stock price: A case study on karachi stock exchange', *J. Econ. Sustain. Dev* **5**, 114–124.

Stenqvist, E. & Lönnö, J. (2017), 'Predicting bitcoin price fluctuation with twitter sentiment analysis'.

Symeonidis, S., Effrosynidis, D. & Arampatzis, A. (2018), 'A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis', *Expert Systems with Applications* **110**, 298–310.

Zhang, L., Ghosh, R., Dekhil, M., Hsu, M. & Liu, B. (2011), 'Combining lexicon-based and learning-based methods for twitter sentiment analysis', *HP Laboratories, Technical Report HPL-2011* **89**.

Zulkarnaen, A. H., Syamsun, M. & Maulana, T. N. A. (2016), 'Analysis of fundamental and technical factors to stock price on residential property sector companies listed in indonesia stock exchange', *International Journal of Scientific and Research Publications* **6**(12), 315–319.