Heriot-Watt University

Thesis Report On

Twitter-Based Financial Sentiment Analysis: Enhancing SentiStrength for Improved Stock Market Prediction

Author:
Omana Prabhakar

Supervisor:
Jessica Chen-Burger

A thesis submitted in fulfilment of the requirements for the degree of MSc.Data Science

School of Mathematical and Computer Sciences

April 2025

Student Declaration of Authorship

HERIOT WATT UNIVERSITY
UK | DUBAI | MALAYSIA

Course code and name:	F21MP / Masters Project and Dissertation
Type of assessment:	**Individual**
Coursework Title:	A project report on: Twitter-Based Financial Sentiment Analysis: Enhancing SentiStrength for Improved Stock Market Prediction
Student Name:	Omana Prabhakar
Student ID Number:	H00459314

Declaration of authorship. By signing this form:

- **I declare** that the work I have submitted for individual assessment OR the work I have contributed to a group assessment, is entirely my own. I have NOT taken the ideas, writings or inventions of another person and used these as if they were my own. My submission or my contribution to a group submission is expressed in my own words. Any uses made within this work of the ideas, writings or inventions of others, or of any existing sources of information (books, journals, websites, etc.) are properly acknowledged and listed in the references and/or acknowledgements section.
- I confirm that I have read, understood and followed the University's Regulations on plagiarism as published on the University's website, and that I am aware of the penalties that I will face should I not adhere to the University Regulations.
- I confirm that I have read, understood and avoided the different types of plagiarism explained in the University guidance on Academic Integrity and Plagiarism

Student Signature *(type your name):* *Omana Prabhakar*

Date: *09/04/2025*

"Your emotions are the slaves to your thoughts, and you are the slave to your emotions."

Elizabeth Gilbert

Abstract

Today, we know that social media platforms are powerful tools for shaping public opinion, influencing behavior, and impacting various industries. Among them, Twitter has proven to be an effective space for research compared to other social media platforms due to its real-time nature and diverse user base. Retail investors often react to trends, news, and opinions shared on this platform, which directly affects market sentiment, making it an ideal venue for analyzing stock market trends that can influence investor behavior. The public's positive or negative sentiments expressed on Twitter can lead to market fluctuations, which can be utilized for stock market prediction. This research explores the potential of financial sentiment expressed on Twitter to predict stock market trends of the S & P 500. The study employs the SentiStrength tool to extract sentiment from financial-related tweets. The main goal of this work is to enhance sentiment classification accuracy and refine it with financial-specific words extracted from the dataset. The study compare the performance of three different sentiment analysis methods: (1) the original lexicon; (2) a custom financial lexicon that includes domain-specific terms; and (3) a context-based approach that focuses on tweets related to U.S.–specific financial events. Machine learning models such as LogisticRegression, XGBoost, and LSTM are then used to evaluate the predictive power of sentiment. To improve the predictive power of the models even more, feature engineering was performed to create new features to include within the machine learning models. Experimental results show the effectiveness of this hybrid approach, which combines sentiment analysis with machine learning, improving its prediction accuracy by 96.42%. Additionally, the novel financial sentiment analysis approach achieved an accuracy of 97.89%. This indicates the potential of financial sentiment analysis and financial-sentiment- driven stock market prediction, as its applications will help investors gain insights that can assist in decision-making, trading strategies, and financial analysis regarding whether to buy, sell, or hold.

Keywords: Sentiment Analysis, SentiStrength, Financial Lexicon, Context-Based Sentiment Analysis, Machine Learning models, Stock Market Prediction, Twitter Sentiment, S&P500.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr. Jessica Chen Burger, for her invaluable guidance, support, and encouragement throughout this research. Her insightful feedback and expertise have been instrumental in shaping this study, from refining the research methodology to analyzing the results. Her patience and constructive advice have helped me navigate challenges and enhance the quality of my work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Predicting stock market trends has become a challenging task in the present world, unlike traditional prediction methods. Social media networks generate and store massive volumes of data on a variety of topics on a regular basis Pagolu [2016]. Social media networks generate and store massive volumes of data on a variety of topics on a regular basis. This pool of data represents the modern equivalent of a goldmine, filled with important information. Baruah and Changkakati [2024]. So, the majority of stock information is available online via various social media channels, with Twitter being one of the most prominent. Twitter (now known as X) is popular for sentiment analysis. With more than 192 million active users and over 500 million tweets each day, Twitter effectively executes sentiment and time series analysis. Rakshit [2024]. In 2021, Twitter's global user base was around 429.79 million. Furthermore, it is expected that by 2025, there will be 497.48 million users globally Mokhtari [2023]. This microblogging and social networking service has grabbed the interest of scholars from a range of sectors, including politics Ibupoto et al. [2022], health care, and finance Mokhtari [2023]. As a result, Twitter is regarded as an essential resource for forecasting stock market moves among all social media platforms.

With this growing influence of public sentiments on stocks as explained in the study by Figà-Talamanca and Patacca [2022], this project aimed to perform sentiment analysis on financial-related tweets using the SentiStrength tool to generate the sentiment score from the sentiment from the tweet and present the distribution of the feelings as positive, negative, or neutral. Following sentiment extraction, supervised machine learning algorithms such as Random Forests and Logistic Regression were applied to build a model that evaluated and analyzed tweets, which helped to understand the correlation between stock price movements and Twitter sentiments. This project also compared the effectiveness of SentiStrength tool and machine learning techniques to determine the most accurate method for stock market prediction to analyse stock movements combined with sentiment analysis on Twitter allows us to tap into this psychological influence on stock prices.

As most of the studies of Divya and Menaka [2025],Samuel et al. [2020] have used the Twitter data, suggesting the correlation of sentiments and stock market fluctuations. This study also uses Twitter data, especially tweets containing hashtags denoted to the financial data, collected from June 2020, as this was an interesting year of study due to the ongoing pandemic in that year and several studies have been conducted by Goel et al. [2020], Simanjuntak and Pramana [2021] and Samuel et al. [2020] have showed the impact of COVID-19 on stocks. So with the help of the enhanced SentiStrength tool, the stock market trends were analysed, which can be fed into the recommendations system that will collect, process, and analyse tweets related to finance, offering recommendations such as "Buy," "Hold," or "Sell," based on aggregated sentiment data. Recommendations often promote a long-term, more disciplined investment strategy, which aligns well with the needs of less experienced investors helping them make more informed investment decisions. This sentiment-driven stock tweet empowers beginner investors by offering insights into market sentiment, allowing them to make more informed and data-driven investment decisions. It aimed to bridge the gap between investor sentiment influence and stock market performance, contributing to the broader field of sentiment analysis in financial forecasting.

1.1 Current Knowledge and Gaps

Sentiment analysis is increasingly applied to financial markets, particularly for stock market prediction. Sentiment analysis methods shown in the existing studies that used VADER, lexicon-based approaches, CNN Sina and Setiawan [2023] RNN Zhang et al. [2022], LSTM Pathan and Prakash [2021], and traditional machine learning techniques like Naive Bayes Maulidiana et al. [2024], SVM Putri et al. [2024] and Logistic Regression, and Sentiment analysis tools such as VADER are popular for handling general sentiment analysis but is limited when dealing with nuanced language, sarcasm, and domain-specific jargon. Also, there is a trend toward studies often using Twitter or Reddit sentiment data to predict stock trends. Still, they typically encounter issues due to linguistic subjectivity, cultural nuances, and limited language support Bjork and Kiran [2019]. Accuracy rates vary, with some methods reporting improvements in sentiment prediction. Still, overall reliability is constrained by small datasets, limited to single languages,a often English, and lack of robustness in real-time applications. While some studies achieve accuracy improvements of ASBA, which shows a 1.07% accuracy improvement Yili [2024], the features are often limited to standard sentiment scores without deeper financial indicators, so it isn't easy to select which model to predict. Most sentiment models are general-purpose and struggle to capture financial sentiment specifically. Sentiment algorithms like VADER Youvan [2024] and SentiWordNet are not fine-tuned for financial terminology, leading to inaccurate predictions when applied directly to financial contexts Bjork and Kiran [2019].

1.2 Motivation

This research aimed to tackle these gaps by utilizing SentiStrength sen [2017], a sentiment analysis tool that could potentially be customized for financial lexicons discussed in the study of Yekrangi and Abdolvand [2020] combined with an advanced feature engineering approach specifically tailored to financial-domain-specific sentiment analysis by customizing the SentiStrength to account for financial terms and expressions that can lead to more accurate sentiment capture within financial discussions or indicators like historical price data on Twitter. This approach improved the sentiment scores, increasing the stock price predictions reliability. By using a tool tailored for the financial domain, this study improved the existing model's ability to interpret sarcasm and slang that are common in market discussions. Using longer and sufficient data can make the model more effective and applicable in various markets, potentially increasing its utility for international investors. A strong foundation in stock movement forecasting is also an inspiring future research area direction to pursue Nti et al. [2020].

1.3 Aims and Objectives

Aim: To develop an enhanced version of the SentiStrength sentiment analysis model, optimized with a financial-specific lexicon based on Twitter data, thereby improving prediction accuracy for stock market movements.

- **Objective 1:** To analyse the performance of SentiStrength and the machine learning approach in predicting stock market movements.
- **Objective 2:** To develop financial context-specific enhancement by incorporating a financial-specific lexicon and stock-related features.
- **Objective 3:** To implement source credibility analysis to give more weight to tweets from financial experts, where these lexicons will be annotated with appropriate sentiments from an investor's point of view.

- **Objective 4:** To evaluate the enhanced SentiStrength model on a real-world dataset of financial tweets and stock price data and evaluate the results with present models to measure the improvement in prediction accuracy.

1.4 Research questions

Research Question 1: How can the SentiStrength sentiment analysis tool be enhanced? and which financial context-specific features may be used to improve the accuracy of predicting stock movements and providing actionable investment recommendations (buy, hold, sell) based on Twitter data.

Research Question 2: How can noise reduction techniques be applied to Twitter data to improve the reliability of the sentiment analysis model for stock market prediction?

Research Question 3: How can sentiment analysis feed into a prediction model that predicts stock movements, and helps investors to decide on trading strategies?

Research Question 4: Does Financial Discussion on Twitter impact stock market movements?

Chapter 2

Literature Review

2.1 Overview

2.1.1 Introduction to Sentiment Analysis

Sentiment analysis is a natural language processing (NLP) technique that extracts sentiment from a large volume of text data. It investigates the author's remarks, opinions, feelings, beliefs, views, queries, preferences, attitudes, and demands in a string of text directed at entities such as services, issues, individuals, products, events, themes, organisations, and their attributes. It analyses the writers feelings expressed in the form of a text, which can be comments Qian [2022], tweets, product reviews, social media data, and articles, journals and analyses the feelings of the writer in two different forms of that is objectivity and subjectivity, polarity such as negative, positive, and neutral, as emotions Lamba and Madhusudhan [2021]. Multiple studies show how sentiment analysis is applied for prediction scalability Gouthami and Hegde [2023] and used in people's sentiment classifications based on tweet posts Siregar et al. [2021].

2.1.2 Sentiment Analysis Workflow

In this section we will discuss the workflow of how the sentiment analysis categorizes the sentiments based on the subjectivity of the text; for example "The TV is 65 inches wide and 60 inches long". Subjective sentiments are more useful for assessing customer product reviews, social media comments, or the sentiments of writers in song lyrics, among other things. On the other hand, when dealing with sentiments such as "I love the dress, but I did not like the material," we cannot determine whether it is completely positive, negative, or neutral because it contains both sentiments. In such cases, we define this as the polarity of a text. Polarity is assessed on a scale of -1 to $+1$, with -1 representing very negative, 0 representing neutral, and $+1$ indicating very positive.

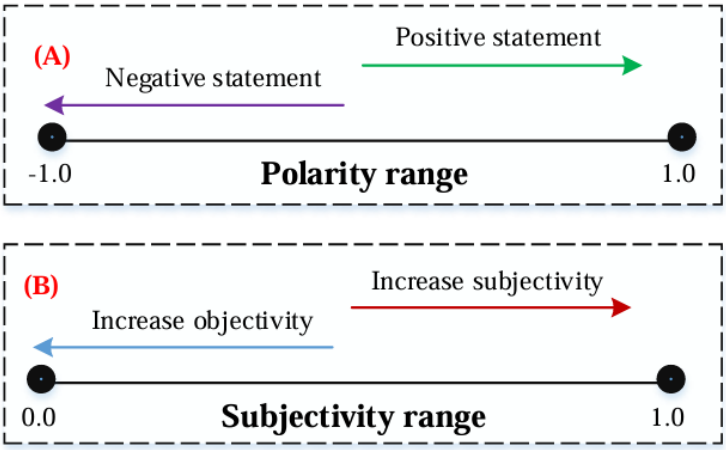


Fig 2.1 Polarity, Neutrality and Subjectivity Range Source: Nti et al. [2020]

For analysing the polarity and subjectivity the sentence can be broken down into,

1. "Opinions or feelings: This is also known as polarity, where emotions can be qualitative, such as upset, happiness, wrath, surprise, disgust, or happy, or quantitative such as rating a movie on a scale of one to ten Khan and Srivastava [2024]".

2. "Subject: It refers to the subject of the debate where one perspective might address more than one aspect of the same issue, for example, "the camera of the phone is great, but the battery life is disappointing Khan and Srivastava [2024]".

3. "Opinion holder: It refers to the author or person who presents the opinion Khan and Srivastava [2024]".

The figure below explains the general procedure of sentiment analysis explained by Khan and Srivastava [2024]. The paper also explains about types and approaches of sentiment analysis techniques that will be discussed in the next section.

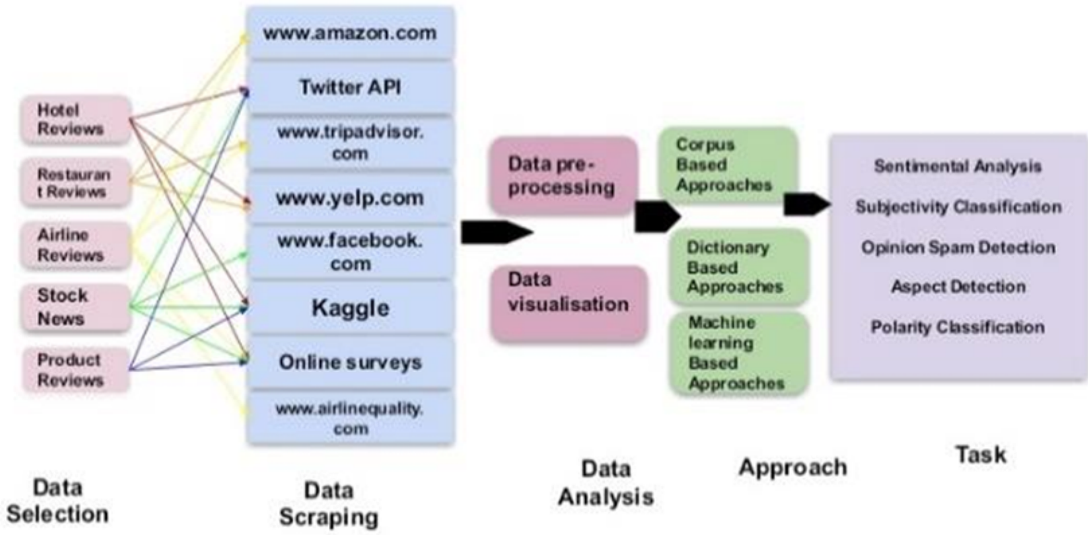


Fig 2.2 Working of Sentiment Analysis Source: Khan and Srivastava [2024]

2.1.3 Types of Sentiment Analysis

1. Fine-Grained Sentimental Analysis: Fine-grained sentimental analysis provides a more exact level of polarity by splitting it down into smaller groupings, which are frequently highly positive or very negative. In my opinion, this is akin to a five-star rating system Khan and Srivastava [2024].

2. Aspect-Based Sentiment Analysis (ABSA): The greatest utility emerges when it is associated with a specific quality or characteristic specified in the text. The ABSA approach involves determining these features and how they make you feel example, rweviews of products Al-Ayyoub et al. [2017]. At Thematic, these aspects are known as "themes" as explained in the studies of Khan and Srivastava [2024] and Jiang et al. [2024]. ABSA investigates further to assess attitudes towards the individual traits or features described. ABSA analyses text at this granular level, providing substantial insights into specific qualities of products, services, or entities, helping businesses to better understand client opinions and address specific areas of concern or satisfaction Zhang et al. [2022] and the current challenges faced by this approach which is explained in Yin [2024]

3. Emotion Detection: Emotion detection recognises specific emotions rather than just positive and negative ones. Joy, displeasure, excitement, rage, and sorrow are among examples Buckley [2013].

4. Intent-based: In a text, intent-based analysis differentiates between facts and opinions. For example, negative online feedback regarding changing a battery may motivate customer service to contact you to resolve the issue Khan and Srivastava [2024].

2.1.4 Sentiment Analysis Techniques

1. Lexicon-based Approach: The Lexicon-based technique uses sentiment scores to divide tweets into negative, neutral, and positive. The word-by-word approach, which takes raw text as a processed structural representation, assumes that the presence of a word or phrase in a record correlates with sentiment. A lexicon is a set of attributes that have been ascribed sentiment values. It acts essentially as a dictionary a predetermined set of words—with each word having several synonyms that connect to it. SentiNet, WordNet, and other popular lexicons are examples. Another example of vocabulary in the tweet is the emoticon, which is a type of emoji used to portray emotion as explained in the study of Khan and Srivastava [2024] and Raees and Fazilat [2024].

2. Machine Learning Approach: The machine learning technique approaches sentiment classification using syntactic and linguistic properties in the same way that it would a typical text classification problem. The categorisation model assigns each record's properties to a unique class identifier. The framework is then applied to anticipate an appropriate class label for a previously unknown class instance. This study has problems categorising data, and each case is assigned only one label. When an instance is assigned a probabilistic value, the soft categorisation issue occurs Khan and Srivastava [2024].

3. Hybrid Approach: Hybrid Machine Learning originated as a result of the merger of classical machine learning and current AI technologies. This strategy combines the best of both worlds, resulting in more lasting and adaptable models. Rather than pitting one against the other, Hybrid Machine Learning combines their strengths to overcome individual limitations and create better outcomes. A hybrid technique is adopted, combining machine learning with lexicon-based methods. Hybrid sentiment analysis combines machine learning with lexicon-based approaches. The hybrid technique combines the two and is widely utilised, with sentiment lexicons playing an important role in a variety of systems. In most additional evaluations and comparisons, the hybrid model beat both models Khan and Srivastava [2024]Gurusamy [2024] Baltezarevic and Baltezarevic [2024].

2.1.5 Applications of Sentiment Analysis

As social media are used widely these days, such as Twitter, Instagram, and YouTube, the amount of comments conveying people's thoughts and information has skyrocketed; as a result, sentiment analysis has arisen as a prominent and trendy issue among academics and professionals. Sentiment analysis also allows us to make conclusions about numerous topics based on the opinions of others, and sentiment orientation is important in anticipating user behaviour and market trends. Thus, assessing opinions and detecting sentiments is crucial Mokhtari [2023].

For example, SA can be applied in brand reputation management as discussed in Bell et al. [2024] paper, which shows the case study of Samsung Product Launches and Market Response where Samsung uses this sentiment analysis approach to monitor customer reviews to its newly lauched products and promoting its business. Sentiment analysis is employed to assess public reaction to new product releases and advertising campaigns. This includes monitoring social media posts, reviews, and tech forums. Based on sentiment data, Samsung can adjust its marketing strategies, address any product-related concerns, and refine future product developments. The capability to interact with market feedback in real time has improved the reception of new products and bolstered Samsung's competitive position in the tech industry.

It also provides actionable insights that help brands enhance their reputation Olusegun [2024], engage with their audience, and achieve strategic goals. In conclusion, social media sentiment analysis offers a powerful means for brands to understand and manage their reputation. By applying best practices, staying updated with emerging trends, and addressing challenges effectively, brands can make use of sentiment analysis to strengthen their market position, foster positive relationships, and drive long-term success. This orientation plays a key role in predicting user behavior and market trends Bell et al. [2024]. Also the study of Khaira et al. [2020] explains how sentiment analysis helps to identify cyberbullying across the social media platforms.

Another observation from the study of Baruah and Changkakati [2024] and Dilik [2024] discovered that sentiments had a positive impact on the daily returns of the stock, meaning that stronger positive sentiment is linked with larger returns. However, existing SA models struggle with domain-specific jargon, leading to misclassification in areas such as financial sentiment.

2.2 Stock Market Prediction using Sentiment Analysis

There has been an increase in research into sentiment analysis to anticipate stock market trends. For example: Pagolu [2016] Sentiment analysis was used to predict stock market movements by examining how closely changes in a firm's stock price, both up and down, connect with public opinions discussed in tweets about that company. This work used two alternative representations, Word2vec and N-gram, to analyse public sentiment in tweets. In this research, the authors applied sentiment analysis and supervised machine learning methods to tweets retrieved from Twitter, examining the relationship between a company's stock market movements and sentiments in a tweet. Using a Logistic regression approach, the classifier achieved an accuracy rate of 69.01%, which varied based on the training set. When the model with LibSVM is trained with 90% of the data, Bjork and Kiran [2019]. Another paper has shown observable correlations between sentiments in Reddit comments and the stock market price. Their findings suggest a noticeable connection between Reddit comment sentiments and stock prices, showing that past sentiments can better predict future market trends. The highest cross-validation accuracy reached was 65.47%, with test accuracy at 64.87% The authors utilized a large-scale machine learning platform, integrating sentiment evaluation with SentiWordNet 3.0, dimensionality reduction via Principal Component Analysis (PCA), and logistic regression for modeling.

The study by Sina and Setiawan [2023] primarily focused on Bank Central Asia which is a well-known Indonesian financial institution, to forecast stock price swings using sentiment trends derived from social media, specifically Twitter, resulting in considerable increases in prediction accuracy by 77.75% by implementing the CNN-LSTM model. It was seen that a strong connection developed between sentiment predictions and actual stock price fluctuations. Positive sentiment correlates with stock price fluctuations at 0.745%, while negative sentiment resulted in a correlation of 0.691%. This concludes that sentiment-driven predictions improves the accuracy using deep learning models.

The recent study Baruah and Changkakati [2024] investigates the efficacy of sentiment research in predicting stock price fluctuations, with a specific focus on the Indian stock market. It was mainly focusing on the dependency of financial markets on the sentiments expressed on social media platforms experimenting with VADER suggested that sentiment analysis alone obtained roughly 60% accuracy in predicting stock price movement with a focus on the "SBI stock which improved when merged with technical indicators in the period between January 2021 to February 2024. The accuracy of the three machine learning models combining approaches with technical indicators increased their effectiveness by an average of 19.6%.

According to some experts, sentiment influences the risk of a stock market drop. However, this study focusses mostly on media sentiment and company sentiment. In an examination of Yu [2024] A growing body of research has shown that market sentiment can have an impact on stock returns and volatility, as well as cause asset values to vary from their intrinsic value over time. When bad news reaches a specific threshold, it is immediately exposed to the market, resulting in a dramatic decline in stock prices. According to this study, there is a positive association between investor sentiment and stock price collapse risk. Investors base their decisions not just on bank-level or macroeconomic facts, but also on risk perceptions. As crisis sentiment rises, a large part of investors will begin to use stop-loss tactics and even sell at very low levels to minimise losses.

According to some academics, evaluating Twitter user opinions in the context of financial markets enables anyone to obtain relevant information about the stock market and use it to forecast price movements Mokhtari [2023]. The stock market is critical to any country's economy. Predicting how a stock will perform in the future allows us to make smarter investments with our money. Sentiment analysis gathers opinions about stocks and attempts to determine the relationship between the sentiments and the stock by collecting and processing various datasets. We could use emotional analysis to anticipate a stock's trajectory following data collection and processing. Thus, forecasting stock movements is always the most critical step in the investment process. However, anticipating stock trends precisely and accurately is always a difficult task.

In the past, due to the lack of advanced computer equipment and a significant information gap, individuals mostly relied on word of mouth or newspapers to learn about a company's development and decide whether to hold its stocks. This situation changed with the advent of the information revolution in the late 20th century. Thus, collecting and understanding public opinion may also be a method to anticipate stock market trends. Given the crucial role of emotion in investment decisions, there is a pressing need to understand how collective sentiment expressed online influences stock prices. Sentiment analysis provides a potential and great solution by enabling the systematic examination of public emotions and opinions shared on digital platforms, by providing insights into general market sentiment and helping to predict market trends more accurately and effectively. Eventually, after the outcome is generated, the comparison between the outcome and actual market movement is conducted, and the connection between them is found at the end Lau [2024]. The analysis of user comments reveals distinct sentiments toward the post: 212 comments are neutral, 196 express positivity, and 82 are negative. Notably, a significant 83% of commenters display either a neutral or positive attitude. This substantial majority suggests an openness and optimism toward the subject of the post, which in this context, relates to investing in Reddit stock. It is well documented that emotions significantly influence human decision-making, especially in financial investments using TextBlob.

The table below shows the comparison of various studies, and the accuracy column represents the experiment results of the studies; and some do not show accuracy as the research involved the introduction of the tools.

Table 2.1: Comparision Grid of Literature Review of Sentiment Analysis Studies

SL. No	Year	Authors	Findings	Algorithm	Accu -racy	Limitations
1.	2024	David Nonan and Upar Kiran	The noticeable connection between Reddit comment sentiments and stock prices	SentiWordNet 3.0, PCA and Logistic Regression	64%	Fine-tuning model and Limited to Reddit and Logistic regression model
2.	2024	Amlan Baruah Banajit Changkakati	Sentiment analysis in predicting stock price movements	VADER	-	Linguistic subjectivity, cultural nuances, and data scarcity limit the reliability of sentiment analysis
3.	2024	Chris Bell, Ayeolu Olukemi, Peter Brokyln	sentiment analysis tools and techniques, offering practical recommendations for brands	-	-	Just a review report no analysis was done or implemented
4.	2024	Manissha Khan and Ankita Srivastava (2024)	Review of Twitter Data Analysis Using ML Techniques	ML algorithms like random forest, Logistic regression, Naive Bayes, SVM, and decision tree classifiers	-	Understanding emotions and context with language hurdles is difficult
5.	2024	Douglas C. Youvan	Understanding VADER in Sentiment Analysis	VADER	-	Difficulty handling sarcasm and language support
6.	2024	Gowtham Kumar Sandaka (2024)	sentiment analysis on the Twitter data regarding the COVID-19 vaccine and get out the sentiment's polarity from the data	Uses Vader Sentiment Analyzer	-	time series analysis based on months and SA limited to only the English language
7.	2024	TszYenLau (2024)	relationship between the sentiments of the public and the trend of SPX	Natural Language Processing (NLP)	-	Lack of data (one or two days were collected)
8.	2024	Xiangkui Jiang1, Binglong Ren1 Qing Wu1, Wuwei Wang1, Hong Li1	-	ASBA, Context-Aware Sentiment Analysis Model	improvement in accuracy by 1.07%	Domain limitations e.g. Politics, customer review
9.	2024	Haonan Yu	Relationship between Investor sentiment and stock crash risk	-	-	Not all models could run. Endogeneity Issues

Table 2.2: Comparision Grid of Literature Review of Sentiment Analysis Studies

SL. No	Year	Authors	Findings	Algorithm	Accu -racy	Limitations
10.	2024	Md Shahidul Amin	Twitter Sentiment Analysis in Forecasting Stock Market Trends	Random Forests, Decision Trees, Gradient boosting	-	ML models
11.	2024	Pranati Rakshi, Pronit Sarkar, Shubhankar Roy	Sentiment Analysis on Twitter Data	Bi-Direction RNN LSTM	83.50%	Limited to Sentiment Analysis
12.	2024	Muhammad Raeesa, Samina Fazilata	application of lexicon-based sentiments	lexicon-based, Text Blob and Vader Sentiment, Random Forest, SVM, Naive Bayes	81%	Feature selection for analysis
13.	2022	M. Lamba, M. Madhusudhan	Review of Sentiment Analysis	Python	-	Sarcasm, irony, and implication are common and hard to decipher
14.	2022	Yili Wang, Jiaxuan Guo, Chengzheng Yuan and Baozhu Li	Comprehensive overview of TSA techniques and related fields	Linear Classifier, Rule-Based Classifier, Decision Tree Classifier, Lexicon-Based Approach	-	size of the training dataset, domain dependence, and Context independence
15.	2020	Isaac Kofi Nti, Benjamin Asabam Weyori and Adebayo Adekoya	Predicting stock movements based on sentiment analysis	ANN, NLP, RMSE, MAPE	-	Limited Stock market data (only Ghana Stocks)
16.	2016	Venkata Sasank Pagolu, Ganapati Panda and Babita Majhi	Correlation between public sentiment on a company and the changes in the stock price of that company	Word2Vec and N-gram, NLP	-	Limited to Twitter Data

2.3 SentiStrength as an SA Tool

In this section, we shall focus specifically on the sentiment analysis tool SentiStrength which we are going to utilize in our research. We will also be discussing its lexicon-based approach and how it is used in stock market prediction and let's also look into its advantages and limitations in financial sentiment analysis.

2.3.1 Introduction to SentiStrength

The paper by Khan [2019] has introduced the SenstiStrength tool and its complete working in his paper. Based on social data related to sentiment it classifies based on lexicons and evaluates sentiment strength using ratings, where positive emotions of the writer refers to the good feeling and negative emotions refers to a bad feeling. So basically this algorithm of SentiStrength examines a text and divides it into words, separating emoticons and punctuation. Each word in the text is then searched against the lexicon dictionaries to see if it matches the terms present in the sentence. If the lexicon and the term is matched then the term is assigned with the corresponding weight assigned to the lexicon in the dictionary. The overall score for a sentence is further calculated by adding the highest positive and negative values for its words. So, In conclusion SentiStrength follows a "lexical approach". At its centre is a vocabulary comprising "I, 125 words, and 1,364-word stems, each with a score for positive or negative" emotion. When these match a term in a text, it indicates the presence of feeling and its intensity. There are standards for enquiries, idioms, spelling correction, and punctuation, as well as norms for conveying sentiment via computer-mediated communication. SentiStrength includes a collection of emoticons and sentiment strength scores for them "(e.g., smiling faces, score +2)".

There is also two modes that SentiStrength operates supervised and unsupervised modes. In the unsupervised mode, the lexicon is weighted based on predefined sentiment strength. Buckley [2013]. Although both the modes increase sentiment detection accurately in few circumstances, therefore they are used analysing social media material with a specific areas.

2.3.2 Lexicon Structures and Sentiment Rules

Sentiment lexicons are useful tools for research that includes opinion mining and sentiment analysis. Du et al. [2023]. SentiStrength has numerous rules to deal with exceptional scenarios. SentiStrength incorporates the following rules as mentioned in the studies by Thelwall [2016] Buckley [2013] Pagolu [2016].

- **Idiom List Usage:** An idiom list aids in identifying the sentiment of some often used idioms, taking precedence over the sentiment associated with individual words. This collection has been updated to include phrases that explain the emotional context behind commonly used terminology.
- **The Case of "Miss":** The word "miss" is an unusual case because it conveys both a positive (+2) and a negative (-2) attitude. This dual character reflects its frequent use to indicate both sadness and tenderness, as in the sentence "I miss you," which becomes neutral. A positive example for the miss case would be "I did not miss a single opportunity that life offered me," and a negative example would be "I missed the bus".
- **Spelling Correction:** A specialised algorithm corrects spelling by deleting repeated letters that occur more frequently than is common in English, or when this adjustment results in a recognised term. For example, "hellp" would be changed to "help."
- **Intensity from Repeating Letters:** When at least two more letters are added to a word, the sentiment intensity rises by one. For example, "haaaappy" is considered more positive than "happy." This repetition gives neutral words a positive emotion score of +2.
- **Booster Words:** Certain words, known as boosters, are employed to either magnify (e.g., "very") or lower (e.g., "somewhat") the power of the sentiment in the word that comes after them.
- **Negating Words:** A list of negating keywords is employed to cancel out the feeling of words that occur immediately afterward, assuming no intervening boosting words. For example, "I don't hate him" does not imply a negative attitude.
- **Emoticons:** Sentiment analysis includes an emoticon set that has predefined polarity scores, such as ":)" contributing a score of +2.
- **Exclamation Marks:** Exclamation marks add a minimum positive sentiment of +2 unless combined with negative terms, as in "hello Kaaaalix!!!".
- **Repeated Punctuation:** Consecutive punctuation marks, especially exclamation points, increase the sentiment intensity of the preceding word by one.
- **Negative Phrase Pairing:** When two successive moderately or severely negative terms with a strength of at least -3 appear together, the second term receives a -1 negativity boost. For example, "nasty[-3] hateful[-4]" produces a cumulative negative score of -5.

- **Capitalised Sentiment Phrases:** Phrases written in uppercase letters gain +1 strength.
- **Positive Phrase Pairing:** Two consecutive positive terms, each with an emotion strength of at least +3, raise the intensity of the second word by one.
- **Irony Detection:** Sentences marked as ironic have their positive sentiment reduced to +1, while their negative sentiment is altered to be one less than the positive value. For example when a person says "what an amazing movie, I would pull out my eyes," in this example "pull out my eyes" refers to the ironic expression of the person, which means the person does not appreciate the movie. Irony is recognised using a user-defined set of suggestive terms or idioms, such as political allusions or popular derogatory phrases. Additional rules in the system can be customised or turned off as needed for individual applications Thelwall et al. [2011].

2.3.3 Applications of SentiStrength

According to the literature survey, SentiStrength can be applied to the following as specified in the paper Thelwall [2016].

- To demonstrate a simple use of SentiStrength to examine sentiment trends in social web messages.
- The significance of sentiment in YouTube comments was examined, and a small percentage of viewers write comments during or after a YouTube video, and they are interested in the information they can offer about how viewers are responding to the movie or its subject.

2.3.4 Advantages of using Sentistrength

It is said that the SentiStrength is fast and cn process more than 13000 tweets per second on a basic computer and is transparent in showing details on how each ratings were derived and supports multiple languages Thelwall et al. [2011] and this capability of SentiStrength in identifying positive and negative sentiments is done by testing it with human coded texts from different domains such as "Myspace, YouTube, Twitter, the Runners World marathon discussion forum, the Digg news identification site, and the BBC Forum news discussion site." SentiStrength has been applied to comments on the social networking site MySpace and is intended to identify the positive and negative sentiment in brief, casual social web writing.

2.3.5 Limitations of SentiStrength

One of SentiStrength's drawbacks is that it doesn't differentiate between different word meanings using grammatical parsing techniques like part-of-speech tagging. This is due to the fact that, in contrast to conventional language parsers, it does not depend on standard grammar for optimal performance and is developed for comparatively informal information from the socials. Thelwall [2016].

The dictionary's shortcomings originate mostly from domain-specific variances in the meanings of technical terms in casual writing. We must create domain-specific dictionaries to improve the accuracy of text SA in a specific domain.Islam and Zibran [2017]. Several ways to build a dictionary have been tried in the past, with varying degrees of success. The production of a domain lexicon dictionary is a tiresome operation, and it is frequently hard to tell in advance which technique to dictionary building can suit better for a specific domain.

2.3.6 Tool Comparison of SentiStrength with other Approaches

This section compares SentiStrength with other Lexicon-Based Approaches. The comparision shows that the SentiStrength dictionary has outperformed AFINN and VADER. AFINN performs somewhat better than SentiStrength in detecting positive and neutral sentiments, but substantially worse in detecting negative sentiments. Thus, AFINN lags behind the SentiStrength dictionary in overall accuracy. Islam and Zibran [2017].

The SentiStrength dictionary combines "LIWC and GI dictionaries, similar to VADER, and adds lists of emoticons, negations, and intensifiers" Islam and Zibran [2017]. This lexical tool is designed to be modular, allowing to replace the default dictionary with another one. SentiStrength-SE is considered to be significantly more accurate in SA in Software engineering text than the most widely used tool in the software engineering community.

According to Buckley [2013], SentiStrength outperforms baseline accuracy for negative sentiment strength on all datasets and positive sentiment strength on all datasets, with the exception of "Digg and BBC forums". The correlation is the most relevant measure since it essentially measures the level of accuracy of each forecast; as a result, matches that are more wrong are faulted everely. SentiStrength generates a positive correlation of 0.3 or higher for all data sets, whereas a random forecast would have a correlation of zero and a bad prediction would have a negative correlation. SentiStrength may therefore be used to find sentiment patterns in any type of data. This gives some confidence that SentiStrength is a trustworthy method for determining the strength of sentiment in social media data. Buckley [2013].

2.4 Techniques and Tool Performance Comparison

In this section, we will look at the various sentiment analysis techniques and their comparisons as indicated by the literature review conducted using various sentiment analysis methodology. Sentiment analysis strategies can be classified into three types: machine learning-based, lexicon-based, and hybrid approaches. Sentiment analysis taxonomy number seven out of fifteen. Yili [2024].

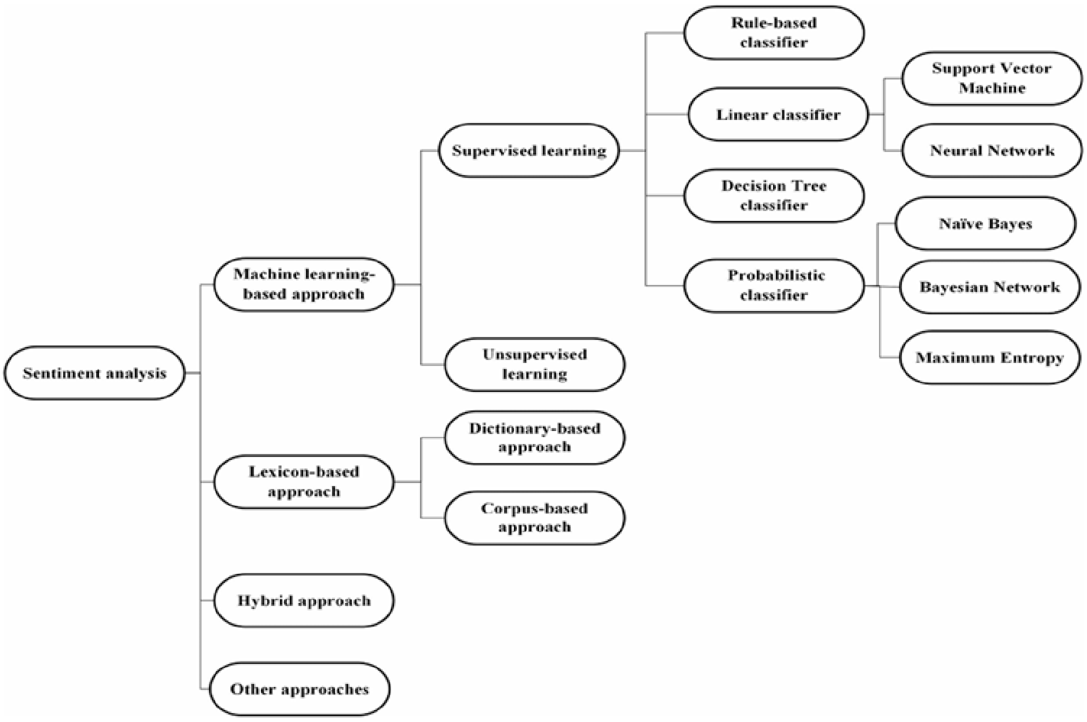


Fig 2.3 Taxonomy for Sentiment Analysis by Source: Yili [2024] .

1. Lexicon-Based Approaches:

- **Overview:** Lexicon-based approaches rely on predefined lists of sentiment-laden words (lexicons) and their associated sentiment scores. These methods calculate the overall sentiment of a text by aggregating the sentiment scores of individual words Yili [2024].
- **Advantages:** Simplicity, interpretability, and ease of implementation. They do not require labeled training data Yili [2024].

- **Disadvantages:** Limited ability to handle context, sarcasm, and nuanced language. Performance depends heavily on the quality and coverage of the lexicon Yili [2024].

- **Examples:** SentiWordNet, AFINN, VADER Yili [2024].

2. **Machine Learning-Based Approaches:**

- **Overview:** Machine learning approaches involve training algorithms on labeled datasets to classify text based on sentiment. Common algorithms include Naive Bayes, Support Vector Machines (SVM), and logistic regression Yili [2024].

- **Advantages:** Ability to learn from data and capture complex patterns. Higher accuracy compared to lexicon-based methods Yili [2024].

- **Disadvantages:** Requires substantial amounts of labeled data for training. May not generalize well to unseen data. Interpretation of models can be challenging Yili [2024].

- **Examples:** Naive Bayes classifiers, SVM, Random Forests Yili [2024].

3. **Deep Learning-Based Approaches:**

- **Overview:**Neural networks, including transformers, long short-term memory (LSTM), and recurrent neural networks (RNN), are used in deep learning techniques to simulate text sentiment. Yili [2024].

- **Advantages:** High accuracy and ability to capture context and long range dependencies in text. Can handle complex language nuances and idiomatic expressions Yili [2024].

- **Disadvantages:** Requires large datasets and significant computational resources for training. Complex models can be difficult to interpret Yili [2024].

- **Examples:** RNNs, LSTMs, BERT, GPT-3 Yili [2024].

2.4.1 Tool Comparison

This section compares SentiStrength with other Lexicon-Based Approaches.

The SentiStrength dictionary appears to have performed the best, followed by AFINN and Vader. AFINN outperforms SentiStrength in detecting positive and neutral feelings, but fared poorly in detecting negative attitudes. As a result, AFINN's total accuracy is lower than that of the SentiStrength dictionary. Islam and Zibran [2017]. The table below shows the tool comparison of SentiStrength, AFFIN and VADER.

Feature/Criteria	SentiStrength	AFINN	VADER
Overall Performance	Best overall performance, especially in detecting negative sentiments Islam and Zibran [2017].	Performs slightly better for positive and neutral sentiments but worse for negative sentiments.	Reliable overall; good accuracy on social media data.
Dictionary Composition	Combines LIWC and GI dictionaries; includes emoticons, negations, and intensifiers.	Word-based scoring; lacks modularity and contextual nuances.	Incorporates contextual valence shifters like negation, intensification, and punctuation.
Modularity	Modular; default dictionary can be replaced Islam and Zibran [2017].	Not designed for modularity.	Limited modularity; general sentiment focus.
Sentiment Strength Accuracy	High correlation with sentiment strength, robust across datasets Buckley [2013].	Lower accuracy for negative sentiments.	Good general accuracy but less emphasized in SE text.

TABLE 2.3: Comparison of SentiStrength, AFINN, and VADER Part 1

Feature/Criteria	SentiStrength	AFINN	VADER
Specialized Use in SE Text	SentiStrength-SE variant shows higher accuracy in SE text.	Not optimized for SE text.	Not specifically optimized for SE text.
Baseline Accuracy	Exceeds baseline accuracy for negative and most positive sentiment strength Buckley [2013].	Lower baseline accuracy for overall sentiment strength.	Comparable baseline accuracy for social media data.
Correlation with Predictions	Positive correlation (0.3) across datasets; robust predictions Buckley [2013].	Weak or negative correlations in some cases.	Positive correlation; reliable but not strong in negative sentiment detection.
Strengths	High adaptability, robust in diverse datasets, excels in negative sentiment detection.	Slightly better for positive and neutral sentiment detection.	Strong contextual analysis, suitable for social media analysis.
Weaknesses	Requires tuning for specialized domains like SE.	Poor negative sentiment detection; lacks modularity.	Less robust in specialized contexts.

TABLE 2.4: Comparison of SentiStrength, AFINN, and VADER Part 2

Chapter 3

Research Methodology

3.1 Research Methodology

This section explains the steps taken to conduct the literature survey as it is one of the critical steps in the research.

1. Identification of the Research Area- The research began with a wider understanding of Sentiment Analysis, its tools, and its applications in different fields such as customer reviews, brand reputation management, movie reviews, recommendation systems, prediction models, etc, by reading various papers.

2. Collection of Relevant Papers- Various papers, articles, and journals relevant to the research were collected from different resources such as ResearchGate, IEEE, Springer, Wiley Online Library, etc using precise keywords from the Research topic breaking down the search into subtopics such as "Sentiment analysis", "Stock market prediction", "Twitter SA", "Twitter SA for Stock market prediction"

3. Selection and Evaluating the Paper- After reading various papers it was found that not all the papers were relevant so this step was carried to narrow down the focus on " SA in stock marketing" as a key application to understand the impact of sentiments on stock price movements and how a positive or negative sentiment may impact stock performance. It was understood that SA can provide insights into the stock market trends. Also made sure the selected papers were recent and made sure that they answered all my questions and aligned with my study.

4. Investigation of Twitter's Role in SA- Investigated why Twitter was considered one of the most widely used social media platforms for SA. Some papers helped confirm its popularity as it has the most active users who express their emotions on different events. Also Validated that Twitter provides real-time datasets relevant to stock market analysis.

5. Comparison of Different Approaches and Tools of SA- Studied various tools and methods used in SA for stock market prediction and addressed that Lexicon-based Approaches were outperforming unlike traditional ML Models, as they were not able to understand the polarity, idioms and sarcasm. Whereas, lexicon-based tools such as VADER, TextBlob and SentiStrength outperformed with their ability to handle nuanced language and multilingual abilities.

6. Defining Research Objectives- By this step, I had a clear understanding of the future scopes and gaps present in the existing studies which led me to expand my research and helped me identify my aim and objectives to conduct my research. This study mainly focuses on enhancing SentiStrength to increase the accuracy of predicting stock market trends.

3.1.1 Requirement Analysis

1. Data Requirements

To build a sentiment analysis model for stock market prediction, the historical Twitter data from a reputable source, Financial Times, of June 2020 for prediction and February 2020 dataset from manually extracting lexicons to integrate it with the original lexicon was collected. A historical Historical stock market data of S&P 500 for the same timestamp as Twitter data, including information such as Open, Close price, was collected. Financial lexicons designed to improve sentiment classification in financial contexts and using a different sentiment analysis approach to cross-validate the prediction.

2. Literature Review

For implementing and conducting the study, the research papers mainly focusing on sentiment analysis, twitter based sentiment analysis, sentiment analysis using different approaches, sentiment analysis using different sentiment analysis techniques, domain-specific sentiment analysis using SentiStrength, Vader, and their study on financial applications were collected through IEEE, Springer, and financial journals. Research papers on Stock market prediction using different approaches, and also papers including machine learning models for financial forecasting, financial-Lexicon Development for financial keyword weighting, and lexicon-based sentiment analysis were gathered.

3. Analytical Methods and Tools

Python was used for data pre-processing and libraries such as pandas, Scikit-learn, XGBoost, and TensorFlow. SenitStrength was used for sentiment analysis for all three approaches and considered the baseline tool for stock market prediction. Further predictive models included logistic regression, XGBoost, and LSTM for evaluation and comparison with the SentiStrength tool, and evaluation metrics included confusion matrix, accuracy, sensitivity, specificity for the SentiStrength tool and Precision, Recall, F1-score for machine learning models.

3.2 Methodology and Proposed Framework

3.2.1 Introduction

The primary goal of this research is to determine the efficiency of financial sentiment analysis in predicting stock movements using Twitter data. This study examines the accuracy and performance of a sentiment analysis tool when applied to financial data. The study compares the outcomes provided by a generic vocabulary and a specialised financial lexicon to evaluate which technique produces more reliable sentiment-based forecasts. A systematic methodology is used to analyse the function of sentiment analysis in stock market prediction. The approach combines Twitter-based sentiment research with machine learning algorithms, allowing for a data-driven assessment of how financial sentiment drives stock price fluctuations. The methodology includes data collecting, preprocessing, sentiment analysis, predictive modelling, and evaluation to ensure a thorough examination of the research hypothesis.

The Twitter data for June 2020 was taken from the Financial Times, a trustworthy and authoritative source of financial news and opinions, and they mostly focused on stock market debates, as opposed to random tweets. The continued economic conditions induced by the COVID-19 outbreak made June 2020 a critical period in global financial markets. Governments and central banks undertook a variety of initiatives, which boosted speculation and market volatility. The S&P500 witnessed tremendous volatility, making it a good moment for studying how Twitter attitudes connect with stock price fluctuations. Historical Stock Market Data for S&P500 were collected and aligned with Twitter timestamps to permit direct comparison between sentiment patterns and stock movements.

SentiStrength, which is a lexicon-based tool, was used as a baseline sentiment analysis approach. it was used to extract sentiment due to its ability to assign positive and negative sentiment scores to short texts like tweets. Followed by lexicon customization, the original lexicon was enahnced with the financial-specific words to improve the accuracy of sentiment classification for stock-related discussions. Furthermore to that a context-based sentiment filtering approach was selected to focus on U.S market-related discussions, ensuring that only relevant financial sentiments were considered in prediction models.

Predictive machine learning models, including Logistic Regression, XGBoost, Random Forest, and LSTM, were chosen to compare different predictive capabilities. These models range from simple interpretable models Logistic Regression to complex deep learning approaches LSTM for capturing time-dependent trends.

Feature Engineering was implemented to generate additional sentiment features to evaluate model performance before and after feature engineering. A hybrid approach, which combined sentiment analysis and machine learning, was implemented to utilize both text sentiment insights and numerical data-driven learning and improving the predictions. Evaluation metrics included creation of confusion matrix to determine the accuracy, specificity, sensitivity and F1-score to evaluate and compare the different model performances.

3.2.2 Data Collection

- **Collection of Twitter Data-** The tweets were collected from June 2020, sourced from Financial Times, which ensures high-quality financial discussions, and June 2020 was selected due to ongoing Covid-19 pandemic making it suitable period for sentiment analysis. The data was collected using the Twitter API. The dataset includes fields such as timestamp and tweet text's full content. Hashtags: Stock-related hashtags like #StockMarket, #Investing. Mentions: mentions of companies or financial influencers. The collected data was saved and stored in the ".xlsx" for easier processing and analysis.
- **Collection of Historical Stock Data** The historical stock market data was individually gathered to confirm that it corresponded to the timestamps in the Twitter stream. This dataset covers key stock market indicators like the S&P 500 index, which serves as a benchmark for overall market performance. The gathered parameters include the date and time that correlate to the tweets' timestamps, allowing for a direct comparison of social media sentiment and market movements. The S&P500 index's open and close prices were set at the start and end of each trading day, respectively. By studying these variations alongside sentiment data, we may infer whether social media sentiment drives daily stock movements.

3.2.3 Data Pre-Processing

- **Data Cleaning-** This step involves unnecessary characters, HTML tags, and special symbols from tweets. Eliminate URLs, mentions (username), and hashtags (#tag), as they may not contribute to sentiment analysis.
- **Case Folding** All the texts in the tweet were converted to lowercase to standardize data.
- **Stop Word Removal** Sentistrength can handle stop words but does not perform optimally especially in the character return some common words like (e.g., "is," "and," "the") and punctuation marks are explicitly droppped using Python's NLTK and spaCy library.
- **Noise Reduction** For context-based sentiment analysis custom filtering method was used to remove spam, irrelevant content, and promotional tweets, including identifying unnecessary information, hashtags, and keywords on the context of analysis. non-linguistic content such as emojis, icons or graphical symbols are either removed or converted into a textual description eg:(:) into "happy" or :(into "sad").

3.2.4 Exploratory data Analysis (EDA)

Exploratory data analysis involved an understanding of the June 2020. Explored the number distribution of total number tweets and thei like counts, sources from which the tweets came, percentage of likes, quote counts, reply counts, retweet counts by the sources and visualizing the distribution to understand the importance of user engagements to the tweets. Correlation analysis to check the relation between these attributes. Word frequency to analyse the most common words in the dataset.

3.2.5 Sentiment Analysis Methodology

Sentiment Analysis was carried out using SentiStrength Tool sen [2017] to capture the sentiment scores for each word and generate positive and negative sentiment scores. Further a complex compound sentiment analysis score was calculated using Hao.

3.2.5.1 Sentiment Analysis Using SentiStrength's Original Lexicon

This step included setting up of SentiStrength tool by customizing SentiStrength with the default lexicon sen [2017]. The data was pre-processed, and tweets were fed into

SentiStrength for polarity scoring. The output scores were captured for positive and negative sentiments for each tweet, to produce compound sentiment analysis values, individual sentiment scores was aggregated within predetermined time hourly intervals as 9:30 am to 10:30 am, 10:31 am to 11:30 am, 11:31 am to 12:30 pm, 12:31 pm to 1:30 pm, 1:31 pm to 2:30 pm, 2:31 pm to 3:30 pm and tweets tweeted after trading hours was added to the after hours that is from 3:30 pm to next trading opening hour 9:30 am. These aggregated scores were calculated using metrics such as the mean, sum, or weighted average of sentiment scores for all tweets during the specified time frame, guaranteeing consistency with stock data alignment. The mathematical formula used for the calculation of SA_w is as follows, provided by Hao et al. [2022].

$$SA_w = \frac{\bar{N}_{pos} * N_1 + \bar{N}_{neg} * N_3}{N - N_2}$$

where,

- **N1** = Number of positive tweets published within that time-window
- **N2** = Number of neutral tweets published within that time-window
- **N3** = Number of negative tweets published within that time-window
- **N** = Total number of tweets published within that time-window
- **N_pos** = Means of positive and Negative sentiment scores.
- **N_neg** = Means of negative and positive sentiment scores.
- **SA_c** = Compound sentiment score of that time-window, which is determined by the strength of weighted SA_w of that time-window.

and SA_c was computed as,

$$SA_c = \begin{cases} if\ SA_w > 0, SA_c = \bar{N}_{pos} \\ if\ SA_w < 0, SA_c = \bar{N}_{neg} \\ if\ SA_w = 0, SA_c = 0 \end{cases}$$

3.2.5.2 Sentiment Analysis Using Financial Lexicon

Lexicons were customized by adding stock-related terms and financial jargon (for example, 'bullish', 'bearish', 'downgrade' and the process was repeated for polarity scoring and compound sentiment scoring for the financial sentiment analysis approach. as the existing financial sentiment lexicons, which were manually or mechanically produced, were predominantly made up of single-word entries, despite the fact that finance jargon, terminologies, and collocations are frequently multi-word phrases. To fill this gap, SentiStrength's lexicon phrases was customised to domain-specific terms, creating a concept-level domain-specific lexicon specifically intended for financial sentiment analysis. Du et al. [2023] and the process of polarity scoring and generation of compound scoring was repeated after integrating the financial lexicon.

3.2.5.3 Context-Based Sentiment Analysis

the study by Kartbayev [2024] introduces a method to develop an intelligent classification to predict the stock market using context-based sentiment analysis. This was further implemented to enhance the accuracy of sentiment classification by filtering tweets relevant to the U.S. stock market. Instead of analyzing all financial tweets, only those directly related to U.S. market discussions were considered. This approach helped eliminate irrelevant or global market sentiments that might not influence U.S. stock prices. By combining domain-specific financial lexicons with contextual filtering, this method improved the precision of sentiment analysis, ensuring that only relevant financial discussions were incorporated into stock market prediction models.

3.2.6 Correlation with Stock Data

Alignment of data: To maintain consistency in analysis, Twitter sentiment data was aligned with stock price data using fixed time frames as 9:30 am to 10:30 am, 10:31 am to 11:30 am, 11:31 am to 12:30 pm, 12:31 pm to 1:30 pm, 1:31 pm to 2:30 pm and 2:31 pm to 3:30 pm and one more window that is after trading hours that is the tweets tweedted from the closing price of the day 3:30 pm to the next day's opening price 9:30 am will be considered as after hour window. Sentiment scores will be computed using the mean, sum, and weighted average of each trading day within these time ranges. This guaranteed that the datasets were synchronised and enabled accurate correlation analysis.

Exploring Relationships:

Exploratory Data Analysis (EDA) was conducted to gain insights into the Twitter sentiment data and historical stock market data, ensuring data quality and identifying potential relationships between sentiment trends and stock price movements.

- **Time-Series Analysis:** Analyzed sentiment variations over time to observe trends in market sentiment before and after significant stock movements. The research explored the potential lagged effects, such as whether sentiment from a one-time window influences stock prices in subsequent time windows (for example, "Does today's sentiment affect tomorrow's stock price?") using,
- **Correlation Analysis:** Assessed the relationship between sentiment scores and stock price changes to determine if sentiment has predictive value for stock movements.
- **Stock Price Trends:** Visualized S&P 500 price fluctuations and compared them with aggregated sentiment scores to identify potential sentiment-driven market reactions.

3.3 Machine Learning Approach for Stock Market Prediction

3.3.1 Model Selection and Training

To evaluate the impact of sentiment analysis on stock market prediction, several machine learning models were chosen based on their capacity to detect various patterns in financial data. Logistic Regression was chosen for its interpretability, whilst XGBoost and Random Forest were picked for their ability to deal with complex, nonlinear interactions. Furthermore, LSTM (Long Short-Term Memory) was examined for detecting temporal dependencies in stock price movements.

The models were trained with sentiment ratings, rolling averages, and lagged values. The training method included hyperparameter adjustment and cross-validation to improve model performance and provide robust and reliable predictions.

3.3.2 Feature Engineering:

To improve predictive capabilities beyond the present compound sentiment score SA_c, a full feature engineering methodology was used, which created the following types of designed features, such as lagged sentiment features, moving average of sentiment scores, normalized tweet counts, sentiment proportions, and lag feature sentiment. The lagged sentiment features was used to identify the dependency of sentiment Sa_c and the positive tweet count N1. The previous hour's compound sentiment (SAc_lag1 = SA_c [t1]) and positive tweet count (N1_lag1 = N1 [t1]) were considered. This feature enabled the models to use immediate past sentiment to forecast future moves. To determine short-term price changes and sentiment trends, a moving average of sentiment scores using the rolling mean for Sa_c was calculated during the previous hours. This moving average provided a more consistent view of the current sentiment trend. Experimentation was performed to determine the appropriate window size for this rolling mean of 2 and 3 hours.

To account for varying tweet volumes (N) and prevent disproportionate influence from positive, and negative tweets, counts were normalised by the total number of financial-related tweets within the timeframe. Normalised features were:

- $N_{1norm} = N1 / N$
- $N_{2norm} = N2 / N$

To illustrate the balance of positive and negative sentiment, the fraction of positive and negative tweets relative to the total number of tweets was calculated as,

- $Pos\_Ratio = N_1 / N$
- $Neg\_Ratio = N_3 / N$

To identify sentiment shifts and trends, a sentiment momentum feature was created by combining the previous day's compound sentiment value (SA_c [t1]) and positive tweet count (N1 [t1]). This tried to account for longer-term sentiment carry-over effects. The goal of incorporating engineering features was to provide the machine learning models with a deeper and more complex representation of the changing financial sentiment on Twitter, potentially boosting the accuracy of stock market movement forecasts.

3.3.3 Hybrid Approach for Stock Market Prediction

The hybrid technique combines sentiment analysis with the enhanced sentiment features and machine learning algorithms to improve stock market prediction accuracy. This strategy combines lexicon-based sentiment scores with machine learning-driven predictions.

Enhanced sentiment features, such as compound sentiment values, rolling sentiment averages, and sentiment ratios, were added to models such as XGBoost, Random Forest, and LSTM, enhancing their capacity to detect market trends. The combination of sentiment research and data-driven modelling produced a more complete and accurate forecasting approach for stock price changes.

3.4 Evaluation framework

To ensure a systematic evaluation, the study follows a ground truth generation, tool evaluation, data comparison and evaluation metrics.

3.4.1 Ground Truth Generation

This analysis is based on actual stock price changes. If the stock price rose, it was labelled as one (upward movement), and if it fell, it was labelled as zero (downward movement). This method establishes an objective and data-driven benchmark for assessing sentiment analysis outcomes. Stock price changes are utilised as the ground truth because they provide a quantitative and unbiased reference point for evaluating sentiment's impact on the market. Unlike subjective manual labelling of tweets, this strategy is consistent with real-world financial results.

3.4.2 Tool Evaluation for Different Sentiment Analysis Approaches

The evaluation of the tool was carried out by comparing the two different lexicon-based sentiment analysis approaches. The first approach uses the generic lexicon where the tool uses its default lexicon to generate the positive and negative sentiment scores. The second approach uses the financial lexicon, which is customized lexicon incorporating financial sentiment terms for the financial domain.

3.4.3 Data Comparison

The sentiment scores generated by the tool under both lexicon settings were compared against the ground truth, which is the stock price movements benchmarked as 1 for stock price up and 0 for stock price down. The key comparisons included comparing tool performance using the original lexicon and the financial lexicon, and context-based filtering sentiment analysis with ground truth as the benchmark for comparison.

3.4.4 Evaluation Metrics for Model Comparison

To assess the performance of different sentiment analysis approaches and machine learning models for stock market prediction, various evaluation metrics were utilized and mentioned below:

1. Confusion Matrix: Confusion Matrix was created to determine the accuracy of the tool to compare the sentiment score and the ground truth. The stock price movements price up and price down are treated as ground truth and tool generated sentiment scores values are prediction. The ground truth determined if the tool has generated correct predictions. The table below is the method to create a confusion matrix where,

		Predicted Condition	
	Total Population = P+N	*Positive (PP)*	*Negative (PN)*
Actual Condition	*Positive (P)*	*True Positive (TP)*	*False Negative (FN)*
	Negative (N)	*False Positive (FP)*	*True Negative (TN)*

In the table, **TP** indicates if the prediction is positive sentiment, and if the ground truth is indeed positive sentiment, then this condition satisfies the true positive condition. **FP** indicates if the prediction is positive sentiment, but the ground truth is not positive sentiment, which is wrong, then it becomes a false positive condition. **TN** indicates if the prediction is not positive sentiment, and is correct and the ground truth was negative too then it is considered a true negative case. **FN** indicates the prediction is not positive sentiment, but it is wrongly predicted and the ground truth is indeed positive, then it is considered as false negative condition. With these determined values of the TP/FP/TN/FN values for each tweet, they were summed and filled in the confusion table. The hypothesis for the confusion matrix was considered as if the sentiment is positive, then the price goes up, and if negative, then the price goes down. For instance, if the market price has gone up, and the tool has predicted a positive sentiment, then the confusion matrix value of actual and predicted was agreed and counted as TP. If the stock market price goes down, but the tool prediction is positive, then the confusion matrix value is FP. If the stock market price goes down, but the tool prediction is negative, then the confusion matrix value is TN. If the market price has gone up, but the tool has predicted a negative sentiment, implying that the stock price will go up The confusion matrix value is FN.

1. Overall Accuracy: This evaluation metrics measures the percentage of correctly predicted stock's up and down movements, indicating the general reliability of the model. The formula used to calculate the overall accurcay,

$$Accuracy = \frac{No.\ of\ results\ predicted\ correctly}{Total\ no.\ of\ predicted\ results}$$

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

2. Sensitivity (Recall): The sensitivity evaluates how well the model detects actual stock price increases which can be considered as accuracy for predicting the price going up which the true positive rate. If the Sensitivity is very high, it means the model predicts price increases (stock upward movement) well.

$$Sensitivity = \frac{No.\ of\ pos\ results\ predicted\ correctly}{No.\ of\ pos\ results\ predicted\ correctly + No.\ of\ neg\ results\ predicted\ incorrectly}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

3. Specificity: Assesses the model's ability to correctly identify stock price declines which means accuracy for predicting price going down called the true negatives rate. If the Specificity is very low that means the model struggles to predict price decreases (stock upward movement).

$$Specificity = \frac{No.\ of\ neg\ results\ predicted\ correctly}{No.\ of\ neg\ results\ predicted\ correctly + No.\ of\ pos\ results\ predicted\ incorrectly}$$

$$Specificity = \frac{TN}{TN + FP}$$

The machine learning models used an evaluation metric, precision, recall, and F1-score, where precision.

1. Precision : Where the precision calcultes the total number of true positives to the total number of true positive and false positives.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

2. F1- Score: This metric checks the balance between both Precision and Recall by calculating by the formula below,

$$F1\text{-}Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

3.4.5 Time series comparison

The studies by Sandaka [2024] and Zhang [2024] shows the impact of time-series approach. Hence this study also uses the time-series evaluation to check the correlation of sentiments and stock price changes.

- **Model Comparison:** The performance evaluation involved comparing the enhanced SentiStrength model to the original tool by looking at the correlation between their sentiment scores and subsequent stock price movements over predetermined time periods. This investigation used lag durations of 0, 1, 2, 3, and 4 hours to determine how time delay affected the accuracy and as well as compared the different model's sensitivity and specificity at each hour in forecasting stock movements.
- **Interpretability:** By analysing the aggregated data, it was determined which sentiment elements and time windows contribute the most to stock movement forecasting. These insights will give investors practical information to help them make decisions on whether to buy, hold, or sell stock.

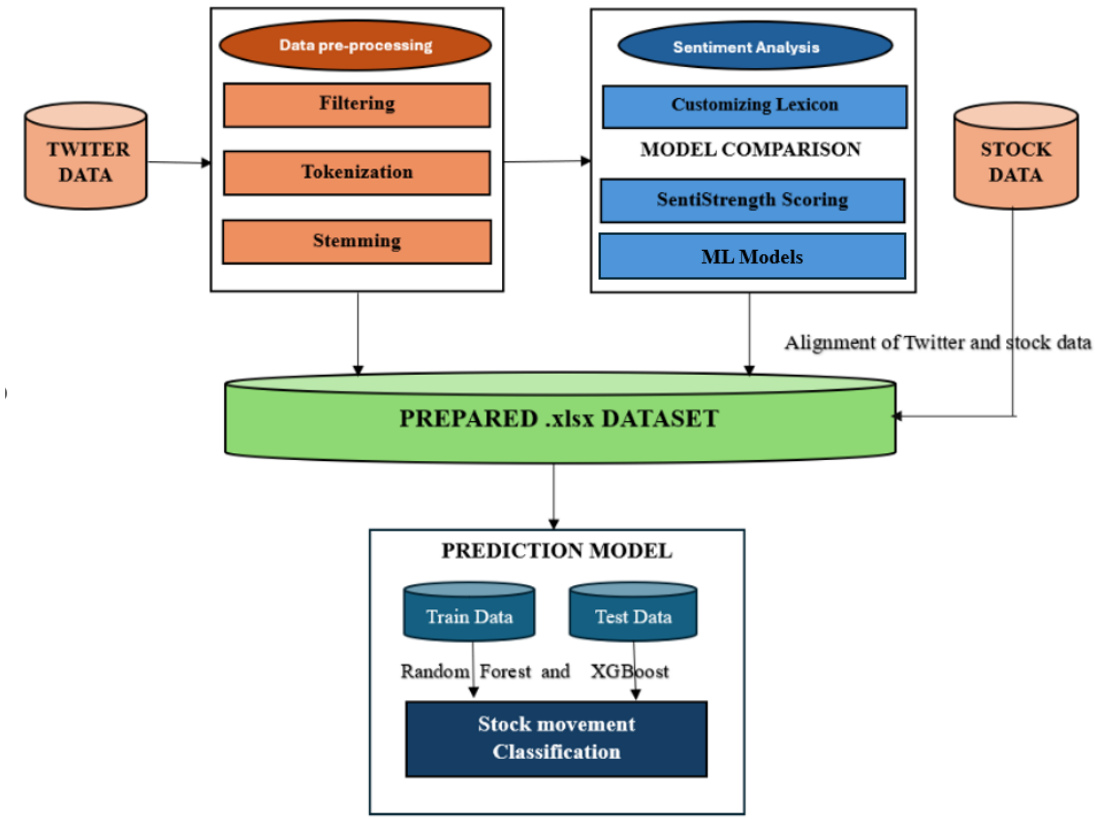


Figure 3.1: Proposed framework for the study

Chapter 4

Implementation

4.1 Introduction

This chapter presents an end-to-end implementation of the research, detailing how sentiment from financial tweets is extracted, processed, and incorporated into predictive models for stock price movement. The implementation began with data collection, data preprocessing, and exploratory data analysis to ensure the quality of the data. The SentiStrength tool was used to extract sentiments from the tweets for sentiment analysis. To evaluate the effectiveness of sentient analysis in stock market prediction, three different sentiment analysis approaches were implemented. These approaches ranged from using a default lexicon to enhancing sentiment analysis through financial lexicon integration with domain-specific enhancements and contextual filtering. The approaches are as follows:

1. Original Lexicon Approach: This method utilized the default sentiment lexicons without incorporating any financial domain-specific words. The goal was to establish a baseline sentiment analysis and evaluate its impact on stock market movement predictions. However, since the general sentiment lexicons are not specifically designed for financial contexts, they may not accurately capture the sentiment behind financial discussions, which led to misinterpretation.

2. Integration of Financial Lexicon: To refine sentiment classification, an enhanced financial lexicons were created by integrating domain-specific financial terms from the February 2020 dataset with enhancement of Hao et al. [2022]. This modification improved the alignment of sentiment analysis with financial contexts, ensuring that sentiment scores more accurately reflected real-time market discussions.

3. Context-Based Financial Sentiment Analysis: Recognizing that financial sentiment may vary based on geographical context, a filtering approach was implemented where only tweets related to the U.S market were considered. Using the enhanced financial lexicon, this approach aimed to improve sentiment accuracy by ensuring that only relevant stock-related discussions were included in the analysis.

Machine Learning for Stock Market Prediction Following these sentiment analysis approaches, machine learning models were implemented to determine their predictive performance when combined with sentiment-based features. The models, included logistic regression, XGBoost, LSTM, and Random forests, were trained using the enhanced financial sentiment scores to predict stock price movements. However, to further improve predictions accuracy.

Feature Engineering for Improved Accuracy To further enhance predictive performance, feature engineering techniques were applied. Addictional sentiment features such as sentiment ratios, lagged sentiment values, and rolling sentiment averages were introduced to provide a more robust representation of sentiment trends.

Hybrid Approach: Combining Sentiment Analysis and Machine Learning

Finally, a hybrid approach was developed by combining lexicon-based sentiment analysis with machine learning models. This approach leveraged both textual sentiment insights and data-driven learning techniques, resulting in an impressive prediction accuracy. The findings highlight the effectiveness of integrating sentiment analysis with machine learning for stock market forecasting. By these systematical implementation and analysing these methods, the thesis aims to uncover whether Twitter sentiment can be a reliable indicator of stock market trends. The results of this implementation will provide insights into the importance of sentiment-driven market predictions and the effectiveness of integrating financial specific sentiment analysis into predictive modelling.

4.2 Data Collection

4.2.1 Twitter Data Collection

The dataset was acquired from the Financial Times utilizing the Twitter API as the tweets from Financial Times provide high-quality financial news and opinions relevant to the stock market. Most tweets in the dataset originated from SocialFlow, Twitter Web App, Media Studio, Ads Uploader, and Echobox platforms primarily used by financial news agencies, institutional investors, and media outlets. This indicd that the dataset is heavily focused on broad market discussions rather than individual small-cap companies. The dataset comprises tweets from June 2020, encompassing 15 columns and 1,000 rows, characterized by attributes such as tweet information, which included columns "created_at," representing the timestamp of tweet creation, "tweet," the textual content of the tweet, and "lang," indicating the language of the tweet, all of which appear to be in English, denoted by 'en.' Additionally, engagement metrics are documented, including "like_count," which reflects the number of likes received by the tweet; "quote_count," denoting the frequency with which the tweet was quoted; "reply_count," indicating the total number of replies to the tweet; and "retweet_count," which reveals the count of retweets. Other pertinent details include 'author id', 'geo', and 'source' providing metadata about the author and the origin of the tweet.

4.2.2 Historical Stock Data Collection

The S&P 500 was chosen as the stock market index for this study because it represents a broad measure of the U.S. market and aligns with the nature of the Twitter dataset. Additionally, historical stock data for the S&P 500, specifically the open and close prices of trading days, was manually collected. By calculating stock price changes from the open and close prices, a new attribute was added to analyse stock movements about sentiment scores. The historical price changes of the S&P 500 served as a benchmark for market performance, making it a suitable index for evaluating sentiment-driven market trends.

4.3 Data Pre-processing and Exploratory Data Analysis

4.3.1 Data Processing

To perform sentiment analysis on the tweets, the tweet texts were cleaned by removing URLs, mentions, hashtags, and unnecessary characters from the tweets, duplicate tweets were dropped, using text normalization method texts were standardized by converting all text to lowercase using importing "regular expression" and "nltk" modules in Python. This cleaned dataset was saved into .xlsx file and used as input for sentiment analysis using SentiStrength.

	Source.Name	author id	created_at	geo	id	lang	like_count	quote_count	reply_count	retweet_count	source	tweet	cleaned_tweet
	ITimesv3_0603.csv	4898091.0	01-06-2020 01:00		1.270000e+18	en	36.0	2.0	4.0	15.0	SocialFlow	Hello, Hong Kong. While you were sleeping, thi...	hello hong kong sleeping read story
	ITimesv3_0603.csv	4898091.0	01-06-2020 01:27		1.270000e+18	en	66.0	3.0	8.0	17.0	SocialFlow	Coronavirus latest: US fatalities decline for ...	coronavirus latest us fatalities decline third...
	ITimesv3_0603.csv	4898091.0	01-06-2020 04:20		1.270000e+18	en	123.0	11.0	20.0	55.0	SocialFlow	The US will deliver 2m hydroxychloroquine dose...	us deliver hydroxychloroquine doses brazil uk ...
	ITimesv3_0603.csv	4898091.0	01-06-2020 04:41		1.270000e+18	en	64.0	3.0	5.0	29.0	SocialFlow	Curfews were in force and the National Guard d...	curfews force national guard deployed cities a...
	ITimesv3_0603.csv	4898091.0	01-06-2020 05:00		1.270000e+18	en	103.0	2.0	5.0	27.0	SocialFlow	Elon Musk's SpaceX launched a new era of space...	elon musks spacex launched new era space activ...

FIGURE 4.1: Cleaned dataset

time	Translation
2020-06-01 01:00:00	hello hong kong sleeping read story
2020-06-01 01:27:00	coronavirus latest us fatalities decline third straight day
2020-06-01 04:20:00	us deliver hydroxychloroquine doses brazil uk hit daily testing target hong kongs hang seng jumped first trading day donald trump threatened rescind tr
2020-06-01 04:41:00	curfews force national guard deployed cities across us sunday country braced violence response killing george floyd police officer
2020-06-01 05:00:00	elon musks spacex launched new era space activity sunday becoming first company carry astronauts international space station
2020-06-01 05:40:00	ft view perhaps since year urban riots widespread opposition involvement vietnam war us presidential election taken place fraught context
2020-06-01 06:00:00	edward luce america faces spectre long summer unrest president stoking polarisation
2020-06-01 06:30:00	rt swine flu made qin yinglin richest farmer world coronavirus added another bn fortune yet people know
2020-06-01 06:44:00	primary schools across england open pupils monday everyone returning class parents planning keep children home according new survey
2020-06-01 08:00:00	senior ft trade writer alan beattie explains true environmental cost trading flowers meat fruit globally becoming clearer debate focuses carbon footpr
2020-06-01 11:00:00	small cluster hedge funds managed women outperformed run men coronavirus crisis new data show highlighting industrys longrunning lack progress fi
2020-06-01 12:30:00	uk banks warning half bn bounce back coronavirus loans unlikely repaid lobbying chancellor prepare collapse hundreds thousands small businesses
2020-06-01 13:00:00	nissan renault planning survive coronavirus pandemic
2020-06-01 13:30:00	hello new york sleeping one mostread stories
2020-06-01 14:30:00	mr trump posted facebook twitter would respond violent protests military force looting starts shooting starts twitter added warning post hid view fac
2020-06-01 15:00:00	free read small cluster hedge funds managed women outperformed run men coronavirus crisis new data show highlighting industrys longrunning lack p
2020-06-01 16:31:00	definitely feel markets way ahead reality really telling every client tap market think pricing couldnt get better says manolo falco investment banking cc
2020-06-01 16:31:00	businesses counting cost coronavirus pandemic years every one ailing ft spoke six companies benefiting changes ways work talk eat shop
2020-06-01 16:32:00	although government guarantee spares banks credit risk worried pursuing hundreds thousands familyrun businesses courts would pr disaster
2020-06-01 16:32:00	hong kong condemned donald trumps threat revoke territorys special trade privileges beijing moved impose antisubversion laws asian financial hub
2020-06-01 16:32:00	kim jong un demanding sharp increase cash north koreas moneyed class counter dual threats coronavirus sanctions
2020-06-01 16:32:00	could one day trust dollar trust america reconverge rana foroohar writes
2020-06-01 16:32:00	hsbcs dominant lucrative business position hong kong taken granted former hong kong chief executive warned pressure grows bank declare support ne
2020-06-01 16:32:00	wolfgang mnchau eus emergency budget increase seems germany still convert fiscal union
2020-06-01 16:32:00	want move city stray far ft staff share tips best rural escapes near london new york berlin hong kong
2020-06-01 16:32:00	world joins solve crisis heavy odds writes jared diamond current pandemic might thus represent beginning bright era worldwide cooperation
2020-06-01 16:32:00	singapore become battleground chinese tech us tech see springboard region said ashley swan executive director property group savills
2020-06-01 16:32:00	swine flu made qin yinglin richest farmer world coronavirus added another bn fortune yet people know name
2020-06-01 16:50:00	ft view one world platform match us presidents misuses consequences ordinary americans live
2020-06-01 18:07:00	global pandemic forcing reappraisal way work socialise travel death
2020-06-01 18:07:00	broken piece jade turbulent future hong kong

FIGURE 4.2: Input Tweets for SentiStrength

4.3.2 Exploratory Data Analysis

The dataset was loaded, and exploratory data analysis was conducted on the raw dataset for a basic understanding. It was observed that most tweets received few or nearly 0 likes, a very small number of tweets had more than 1000 likes, and most tweets had no engagement. It was also seen that a small subset of tweets was significantly more popular depending upon the content, timing, or the influence of the user who posted them. Overall exploration showed that the data was highly imbalanced in terms of engagements; most of the tweets in the dataset were likely from regular users, while few were from highly influential accounts. So, understanding which tweets get higher engagement could help in correlating sentiment strength with stock market movements. The important attributes "created_at" and the "tweet" were selected as important features for sentiment analysis. The below figure shows the statistical representation of the user engagements towards the tweets.

	Unnamed: 0	author id	id	like_count	quote_count
count	1000.000000	1000.0	1.000000e+03	1000.000000	1000.000000
mean	1073.660000	4898091.0	1.273090e+18	48.737000	4.163000
std	618.191871	0.0	4.623124e+15	254.805482	31.214336
min	1.000000	4898091.0	1.270000e+18	0.000000	0.000000
25%	544.750000	4898091.0	1.270000e+18	11.000000	0.000000
50%	1085.000000	4898091.0	1.270000e+18	23.000000	1.000000
75%	1610.500000	4898091.0	1.280000e+18	44.000000	3.000000
max	2107.000000	4898091.0	1.280000e+18	7624.000000	949.000000

	reply_count	retweet_count
count	1000.000000	1000.000000
mean	5.523000	25.065000
std	49.720132	77.317998
min	0.000000	0.000000
25%	0.000000	5.000000
50%	2.000000	12.000000
75%	4.000000	23.000000
max	1556.000000	1057.000000

Figure 4.3: Statistical Analysis of User Engagemnet with the Tweets

The descriptive statistical results of mean shows that the average likes engagement per tweet is 48.98% and retweets of 25%. The standard deviation show the distribution of engagement where retweet and replies vary more than likes. The min and max shows that some tweets has 0 engagement whereas most of the tweets had 7624 likes and 1057 retweets. Finally the quartiles 25%, 50% and 75% shows that 75% of the tweets had less than 44 likes and it also implies that a few tweets had very high engagemnt which may indicate those tweets were from an influential accounts or key events.

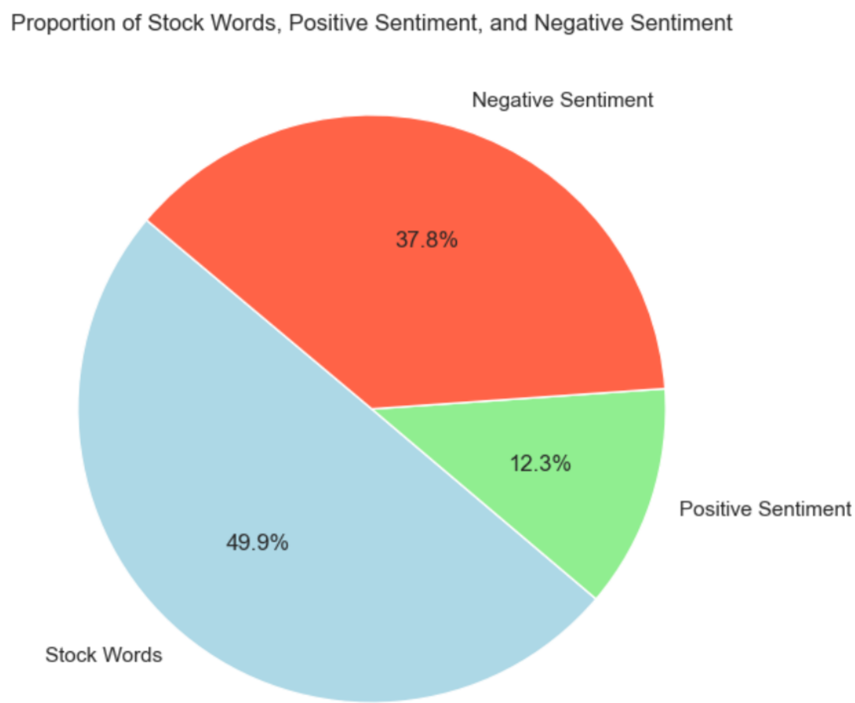


FIGURE 4.4: Distribution of stock words, positive sentiments, and negative sentiments in the June 2020 tweet dataset

It was found that there were a 49.9% proportion of stock-related words, 37.8% Negative sentiment, and 12.3% of positive sentiment in June 2020 dataset.

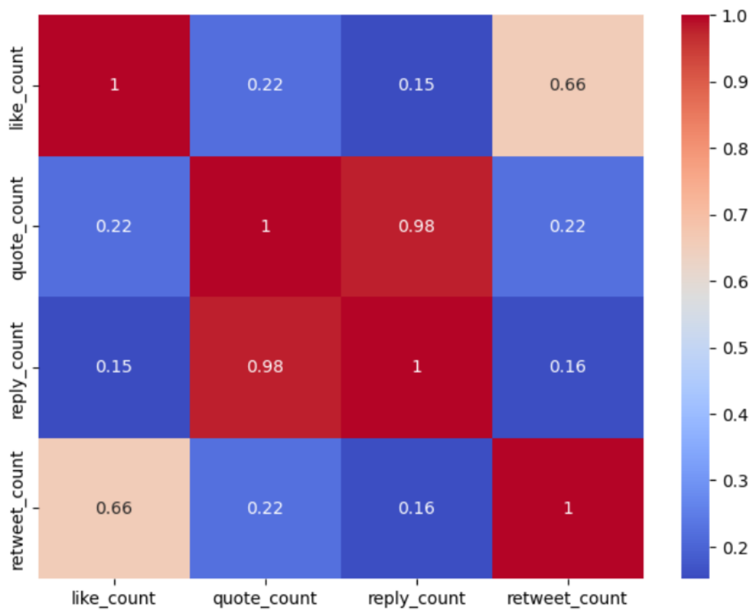


FIGURE 4.5: Correlation Analysis Heatmap

The figure above shows the correlation heatmap shows that quote reply and reply counts correlates well indicating that those tweets fuels the discussion trend to be both quoted and replied to. Retweet and like counts show engagement that supports a message, while replies might indicate differences types of sentiment. This analysis helps which metrics might be helpful for further exploration.

4.4 Sentiment Analysis Implementation

4.4.1 Setting up the SentiStrength Tool

The SentiStrength tool was downloaded from the official SentiStrength Website for Windows. The zip files were extracted and converted into .txt files. The files that contained the lexicons were data files registered as a new file path and the sentiment strength analysis was set to analyze all texts in the file each line separately reporting all the information and save the output file as a .csv file.

4.4.2 Generating Sentiment Analysis Values

The tweets were passed to the SentiStrength tool with the default lexicons, and positive, negative and EmotionRationale values were generated. The output .csv file was cleaned and saved into .xlsx format for better reading and analysis.

Translation	Positive	Negative	EmotionRationale
coronavirus latest us fatalities decline third straight day	1	-1	alities[0] decline[0] third[0] straight[0] day[0] [[Sentence=-1,1=word max, 1-5]][[[1,-1 max of sentences]]]
zil uk hit daily testing target hong kongs hang seng jumped first trading day donald trump threatened rescind trade privileg	2	-4	eng[0] jumped[0] first[0] trading[1] day[0] donald[0] trump[0] threatened[-3] rescind[0] trade[0] privileges[0] follo
e national guard deployed cities across us sunday country braced violence response killing george floyd police officer	1	-2	y[0] country[0] braced[0] violence[-1] response[0] killing[0] george[0] floyd[0] police[0] officer[0] [[Sentence=-2,
spacex launched new era space activity sunday becoming first company carry astronauts international space station	1	-1	/[0] becoming[0] first[0] company[0] carry[0] astronauts[0] international[0] space[0] station[0] [[Sentence=-1,1=
nce year urban riots widespread opposition involvement vietnam war us presidential election taken place fraught contex	1	-3	volvement[0] vietnam[0] war[-2] us[0] presidential[0] election[0] taken[0] place[0] fraught[0] context[0] [[Sentenc
edward luce america faces spectre long summer unrest president stoking polarisation	1	-1	ng[0] summer[0] unrest[0] president[0] stoking[0] polarisation[0] [[Sentence=-1,1=word max, 1-5]][[[1,-1 max of
rt swine flu made qin yinglin richest farmer world coronavirus added another bn fortune yet people know	1	-1	ld[0] coronavirus[0] added[0] another[0] bn[0] fortune[0] yet[0] people[0] know[0] [[Sentence=-1,1=word max, 1-
across england open pupils monday everyone returning class parents planning keep children home according new survey	1	-1	turning[0] class[0] parents[0] planning[0] keep[0] children[0] home[0] according[0] new[0] survey[0] [[Sentence=
ntal cost trading flowers meat fruit globally becoming clearer debate focuses carbon footprint contribution global warmin	2	-1	coming[0] clearer[0] debate[0] focuses[0] carbon[0] footprint[0] contribution[0] global[0] warming[0] long[0] sup
d women outperformed run men coronavirus crisis new data show highlighting industrys longrunning lack progress fixing	1	-2	sis[-1] new[0] data[0] show[0] highlighting[0] industrys[0] longrunning[0] lack[0] progress[0] fixing[0] gender[0] im
bn bounce back coronavirus loans unlikely repaid lobbying chancellor prepare collapse hundreds thousands small busine	1	-4	epaid[0] lobbying[0] chancellor[0] prepare[0] collapse[-3] hundreds[0] thousands[0] small[0] businesses[0] [[Se
nissan renault planning survive coronavirus pandemic	1	-5	0] survive[0] coronavirus[0] pandemic[-4] [[Sentence=-5,1=word max, 1-5]][[[1,-5 max of sentences]]]
hello new york sleeping one mostread stories	1	-1	eeping[0] one[0] mostread[0] stories[0] [[Sentence=-1,1=word max, 1-5]][[[1,-1 max of sentences]]]
would respond violent protests military force looting starts shooting starts twitter added warning post hid view facebook le	1	-2	ng[0] starts[0] shooting[0] starts[0] twitter[0] added[0] warning[-1] post[0] hid[0] view[0] facebook[0] left[0] mess
naged women outperformed run men coronavirus crisis new data show highlighting industrys longrunning lack progress fi	3	-2	s[0] crisis[-1] new[0] data[0] show[0] highlighting[0] industrys[0] longrunning[0] lack[0] progress[0] fixing[0] gend
ead reality really telling every client tap market think pricing couldnt get better says manolo falco investment banking coh	2	-2	g[-1] couldnt[0] get[0] better[1][NegatedDueToPreviousWord] says[0] manolo[0] falco[0] investment[0] banking[
ting cost coronavirus pandemic years every one ailing ft spoke six companies benefiting changes ways work talk eat shop	3	-5	ling[0] ft[0] spoke[0] six[0] companies[0] benefiting[2] changes[0] ways[0] work[0] talk[0] eat[0] shop[0] [[Senten
nt guarantee spares banks credit risk worried pursuing hundreds thousands familyrun businesses courts would pr disast	2	-2	pursuing[0] hundreds[0] thousands[0] familyrun[0] businesses[0] courts[0] would[0] pr[0] disaster[0] [[Sentence
d donald trumps threat revoke territorys special trade privileges beijing moved impose antisubversion laws asian financia	1	-2	0] trade[0] privileges[0] beijing[0] moved[0] impose[0] antisubversion[0] laws[0] asian[0] financial[0] hub[0] [[Sen
ng un demanding sharp increase cash north koreas moneyed class counter dual threats coronavirus sanctions	2	-1	oreas[0] moneyed[0] class[0] counter[0] dual[0] threats[0] coronavirus[0] sanctions[0] [[Sentence=-1,2=word m
could one day trust dollar trust america reconverge rana foroohar writes	3	-1	2] america[0] reconverge[0] rana[0] foroohar[0] writes[0] [[Sentence=-1,3=word max, 1-5]][[[3,-1 max of senten
hong kong taken granted former hong kong chief executive warned pressure grows bank declare support new national se	1	-1	f[0] executive[0] warned[0] pressure[0] grows[0] bank[0] declare[0] support[0] new[0] national[0] security[0] law
wolfgang mnchau eus emergency budget increase seems germany still convert fiscal union	2	-1	increase[1] seems[0] germany[0] still[0] convert[0] fiscal[0] union[0] [[Sentence=-1,2=word max, 1-5]][[[2,-1 ma
want move city stray far ft staff share tips best rural escapes near london new york berlin hong kong	2	-1	st[1] rural[0] escapes[0] near[0] london[0] new[0] york[0] berlin[0] hong[0] kong[0] [[Sentence=-1,2=word max, 1
risis heavy odds writes jared diamond current pandemic might thus represent beginning bright era worldwide cooperatio	1	-5	0] pandemic[-4] might[0] thus[0] represent[0] beginning[0] bright[0] era[0] worldwide[0] cooperation[0] [[Senten
me battleground chinese tech us tech see springboard region said ashley swan executive director property group savills	1	-1	board[0] region[0] said[0] ashley[0] swan[0] executive[0] director[0] property[0] group[0] savills[0] [[Sentence=-
vine flu made qin yinglin richest farmer world coronavirus added another bn fortune yet people know name	1	-1	oronavirus[0] added[0] another[0] bn[0] fortune[0] yet[0] people[0] know[0] name[0] [[Sentence=-1,1=word max,
ft view one world platform match us presidents misuses consequences ordinary americans live	1	-1	sidents[0] misuses[0] consequences[0] ordinary[0] americans[0] live[0] [[Sentence=-1,1=word max, 1-5]][[[1,-1
global pandemic forcing reappraisal way work socialise travel death	1	-5	aisal[0] way[0] work[0] socialise[0] travel[0] death[-2] [[Sentence=-5,1=word max, 1-5]][[[1,-5 max of sentences
broken piece jade turbulent future hong kong	1	-2	0] turbulent[0] future[0] hong[0] kong[0] [[Sentence=-2,1=word max, 1-5]][[[1,-2 max of sentences]]]
ds friends farmville making biggest acquisition date paying billion peak games maker popular puzle titles ton blast toy blas	2	-1	0] paying[0] billion[1] peak[1] games[0] maker[0] popular[0] puzle[0] titles[0] ton[0] blast[0] toy[0] blast[0] [[Sente
olidarity protesters including taking knee condemning death floyd others taken actions inflamed tensions pointed systemic	3	-4	ng[-1] death[-2] floyd[0] others[0] taken[0] actions[0] inflamed[0] tensions[-1] pointed[0] systemic[0] police[0] bru

Figure 4.6: Output Tweets with Positive, Negative and EmotionRationale

The output dataset file consisted of Translation column that contained the original tweet or text that was analyzed by SentiStrength. Positive and Negative columns that contained the sentiment scores indicated how positive and negative the text was and lastly the The Emotion Rationale column contained the reasoning behind the sentiment analysis for each word in the sentence and an overall sentiment summary which is essentially the explanation of how the individual words contribute to the overall sentiment score.

Example 1:

- **Tweet:** coronavirus loans unlikely repaid lobbying chancellor prepare collapse hundreds thousands small businesses
- **Sentiment Score:** [1 -4]
- **Sentiment score breakdown of the tweet:** coronavirus[0] loans[0] unlikely[0] repaid[0] lobbying[0] chancellor[0] prepare[0] collapse[-3] hundreds[0] thousands[0] small[0] businesses[0] [[Sentence=-4,1=word max, 1-5]][[[1,-4 max of sentences]]]

In this example, 1 This means the text has a positive sentiment, where 1 represents a mildly positive sentiment, the scale can vary depending on how SentiStrength defines its scoring. -1 This means the text has a negative sentiment, with -1 indicating a mildly negative sentiment.

Example 2:

- **Tweet:** coronavirus latest us fatalities decline third straight day
- **Sentiment Score:** [1 -3]
- **Sentiment score breakdown of the tweet:** coronavirus[0] latest[0] us[0] fatalities[-2] decline[-1] third[0] straight[0] day[0] [[Sentence=-3,1=word max, 1-5]][[[1,-3 max of sentences]]]].

In this example the positive sentiment score of 1, meaning it's slightly positive, The negative sentiment score is -3, indicating the sentence carries a stronger negative sentiment and In the EmotionRationale words like "fatalities" and "decline" have negative sentiment scores (-2 and -1), which heavily influence the overall sentiment of the sentence. The final sentence sentiment is calculated as -3 by SentiStrength because it takes the maximum negative sentiment.

In conclusion, SentiStrength analyzes the sentiments of each word in a sentence and combines them to derive the overall sentiment score for the entire sentence. The positive and negative pair indicates the final sentiment score, while the EmotionRationale provides a detailed breakdown of how each word contributed to the overall sentiment.

4.4.3 Generating compound SA Values

The compound SA values were generated based on the original sentiment pair of positive and negative scores. Since this thesis focuses on working on Twitter sentiment analysis for stock market prediction, on an hourly basis time window the negative and positive values must be converted into a single compound sentiment score which is the general mood of the hour which is SA_w i.e., weighted SA score based on SA of all tweets that have expressed an emotion. Thus, the compound score is used to track overall market sentiment during the hour and find trends that may impact stock prices.

The compound score sentiment analysis is the method used to represent the overall sentiment polarity of the text by combining individual word sentiments into a single normalized score. This ranges from -5 to +5, where negative sentiment means less than -1, neutral sentiment is between -1 and +1, and positive sentiment is greater than +5. It is calculated by assigning sentiment scores to words as positive, negative, or neutral, considering intensifiers like "very good" is stronger than "good", negating words like "not happy" changes the meaning in the context, handling punctuations and emojis like "amazing!!!" has stronger sentiment than "amazing." It normalizes the scores between -5 and +5 using a mathematical formula.

$$SA_{\mathrm{w}} = \frac{\bar{N}_{pos} * N_1 + \bar{N}_{neg} * N_3}{\mathrm{N} - N_2}$$

where,

- **N1** = Number of positive tweets published within that time-window
- **N2** = Number of neutral tweets published within that time-window
- **N3** = Number of negative tweets published within the time-window
- **N** = Total number of tweets published within that time-window
- **N_mean_pos** = Means of positive and Negative sentiment scores.
- **N_mean_neg** = Means of negative and positive sentiment scores.
- **SA_c** = Compound sentiment score of that time-window, which is determined by the strength of weighted SA_w of that time-window.

  In the formula the N - N2 removes neutral tweets, as they don't carry emotional weight. This gives a weighted average sentiment of all positive and negative tweets.

$$SA_c = \begin{cases} if\ SA_w > 0, SA_c = \bar{N}_{pos} \\ if\ SA_w < 0, SA_c = \bar{N}_{neg} \\ if\ SA_w = 0, SA_c = 0 \end{cases}$$

In the above formula, the SA_c is used to determine the dominant sentiment strength. So, while the SA_w captured a balance of all sentiment signals the SA_c captured if positive or negtaive score was dominant and as SA_c is less noisy than SA_w which is less noisy than SA_c that was considered to check the correlation with the stock price change.

Step 1: Categorized Tweets by Sentiment, for each predefined time window and classified tweets based on sentiment as

- **N1** = Number of positive tweets
- **N2** = Number of neutral tweets
- **N3** = Number of negtaive tweets
- **N** = Total Number of tweets (N = N1 + N2 + N3)

Day	time_range	N1	N2	N3	N
2020-06-01 00:00:00	10:31-11:30	0	0	1	1
2020-06-01 00:00:00	11:31-12:30	0	0	1	1
2020-06-01 00:00:00	12:31-1:30	0	1	1	2
2020-06-01 00:00:00	1:31-2:30	0	0	1	1
2020-06-01 00:00:00	2:31-3:30	1	0	0	1
2020-06-01 00:00:00	After Hours	7	16	13	36
2020-06-02 00:00:00	10:31-11:30	0	0	1	1
2020-06-02 00:00:00	11:31-12:30	0	1	0	1
2020-06-02 00:00:00	12:31-1:30	0	0	2	2
2020-06-02 00:00:00	2:31-3:30	0	2	0	2
2020-06-02 00:00:00	9:30-10:30	0	1	0	1
2020-06-02 00:00:00	After Hours	2	12	14	28
2020-06-03 00:00:00	10:31-11:30	0	1	0	1
2020-06-03 00:00:00	11:31-12:30	0	0	2	2
2020-06-03 00:00:00	12:31-1:30	0	1	0	1
2020-06-03 00:00:00	9:30-10:30	0	1	1	2
2020-06-03 00:00:00	After Hours	0	9	8	17
2020-06-04 00:00:00	11:31-12:30	1	1	0	2
2020-06-04 00:00:00	1:31-2:30	0	2	1	3
2020-06-04 00:00:00	2:31-3:30	1	2	4	7
2020-06-04 00:00:00	After Hours	3	18	6	27
2020-06-05 00:00:00	10:31-11:30	0	1	0	1
2020-06-05 00:00:00	11:31-12:30	0	2	0	2
2020-06-05 00:00:00	12:31-1:30	0	2	0	2
2020-06-05 00:00:00	1:31-2:30	1	1	1	3
2020-06-05 00:00:00	2:31-3:30	0	1	0	1

FIGURE 4.7: Output of Tweets grouped into predefined time window and classified based on N1, N2 and N3

Step 2: Computed the Sentiment Scores to determine the average sentiment scores for positive and negative tweets as

- **N_mean_pos** = Mean positive sentiment score (computed from all positive tweets)
- **N_mean_neg** = Mean negative sentiment score (computed from all negative tweets)

Day	time_range	N1	N2	N3	N	Ñ_pos	Ñ_neg
2020-06-01 00:00:00	10:31-11:30	0	0	1	1	0	-2
2020-06-01 00:00:00	11:31-12:30	0	0	1	1	0	-4
2020-06-01 00:00:00	12:31-1:30	0	1	1	2	0	-5
2020-06-01 00:00:00	1:31-2:30	0	0	1	1	0	-2
2020-06-01 00:00:00	2:31-3:30	1	0	0	1	3	0
2020-06-01 00:00:00	After Hours	7	16	13	36	2.142857	-3.53846
2020-06-02 00:00:00	10:31-11:30	0	0	1	1	0	-5
2020-06-02 00:00:00	11:31-12:30	0	1	0	1	0	0
2020-06-02 00:00:00	12:31-1:30	0	0	2	2	0	-4
2020-06-02 00:00:00	2:31-3:30	0	2	0	2	0	0
2020-06-02 00:00:00	9:30-10:30	0	1	0	1	0	0
2020-06-02 00:00:00	After Hours	2	12	14	28	3	-2.71429
2020-06-03 00:00:00	10:31-11:30	0	1	0	1	0	0
2020-06-03 00:00:00	11:31-12:30	0	0	2	2	0	-2
2020-06-03 00:00:00	12:31-1:30	0	1	0	1	0	0
2020-06-03 00:00:00	9:30-10:30	0	1	1	2	0	-3
2020-06-03 00:00:00	After Hours	0	9	8	17	0	-3.125
2020-06-04 00:00:00	11:31-12:30	1	1	0	2	3	0
2020-06-04 00:00:00	1:31-2:30	0	2	1	3	0	-3
2020-06-04 00:00:00	2:31-3:30	1	2	4	7	3	-3
2020-06-04 00:00:00	After Hours	3	18	6	27	2.333333	-3.33333
2020-06-05 00:00:00	10:31-11:30	0	1	0	1	0	0
2020-06-05 00:00:00	11:31-12:30	0	2	0	2	0	0
2020-06-05 00:00:00	12:31-1:30	0	2	0	2	0	0
2020-06-05 00:00:00	1:31-2:30	1	1	1	3	2	-3
2020-06-05 00:00:00	2:31-3:30	0	1	0	1	0	0

FIGURE 4.8: Output of generated average sentiment scores for positive and negative tweets

Step 3: Calculated the Weighted Sentiment Score (SA_w)

Multiplied the mean positive sentiment score by the number of positive tweets (N1) Multiplied the mean negative sentiment score by the number of negative tweets (N3) Computed the weighted sum and divide by the total number of tweets (N) Subtracted the number of neutral tweets (N2) to adjust for neutral sentiment impact. Determined the Compound Sentiment Score (SA_c) The SA_c was derived based on the strength of SA_w within that time window. A higher SA_w indicates stronger sentiment polarity, influencing SA_c accordingly. This approach ensures that the SA_c score reflects the overall sentiment intensity and distribution within the time window, rather than relying solely on raw positive/negative tweet counts.

Day	time_range	N1	N2	N3	N	Ñ_pos	Ñ_neg	SA_w	SA_c
2020-06-01 00:00:00	10:31-11:30	0	0	1	1	0	-2	-2	-2
2020-06-01 00:00:00	11:31-12:30	0	0	1	1	0	-4	-4	-4
2020-06-01 00:00:00	12:31-1:30	0	1	1	2	0	-5	-5	-5
2020-06-01 00:00:00	1:31-2:30	0	0	1	1	0	-2	-2	-2
2020-06-01 00:00:00	2:31-3:30	1	0	0	1	3	0	3	3
2020-06-01 00:00:00	After Hours	7	16	13	36	2.142857	-3.53846	-1.55	-3.53846
2020-06-02 00:00:00	10:31-11:30	0	0	1	1	0	-5	-5	-5
2020-06-02 00:00:00	11:31-12:30	0	1	0	1	0	0	0	0
2020-06-02 00:00:00	12:31-1:30	0	0	2	2	0	-4	-4	-4
2020-06-02 00:00:00	2:31-3:30	0	2	0	2	0	0	0	0
2020-06-02 00:00:00	9:30-10:30	0	1	0	1	0	0	0	0
2020-06-02 00:00:00	After Hours	2	12	14	28	3	-2.71429	-2	-2.71429
2020-06-03 00:00:00	10:31-11:30	0	1	0	1	0	0	0	0
2020-06-03 00:00:00	11:31-12:30	0	0	2	2	0	-2	-2	-2
2020-06-03 00:00:00	12:31-1:30	0	1	0	1	0	0	0	0
2020-06-03 00:00:00	9:30-10:30	0	1	1	2	0	-3	-3	-3
2020-06-03 00:00:00	After Hours	0	9	8	17	0	-3.125	-3.125	-3.125
2020-06-04 00:00:00	11:31-12:30	1	1	0	2	3	0	3	3
2020-06-04 00:00:00	1:31-2:30	0	2	1	3	0	-3	-3	-3
2020-06-04 00:00:00	2:31-3:30	1	2	4	7	3	-3	-1.8	-3
2020-06-04 00:00:00	After Hours	3	18	6	27	2.333333	-3.33333	-1.44444	-3.33333
2020-06-05 00:00:00	10:31-11:30	0	1	0	1	0	0	0	0
2020-06-05 00:00:00	11:31-12:30	0	2	0	2	0	0	0	0
2020-06-05 00:00:00	12:31-1:30	0	2	0	2	0	0	0	0
2020-06-05 00:00:00	1:31-2:30	1	1	1	3	2	-3	-0.5	-3
2020-06-05 00:00:00	2:31-3:30	0	1	0	1	0	0	0	0

FIGURE 4.9: Output of generated weighted average and compound sa scores

4.4.4 Aggregation of Twitter data and Historical stock data

The Twitter data with the calculated compound SA scores and the historical stock data with the Price Change values were merged on an hourly basis based on 7 different time windows that are 9:30 to 10:30, 10:31 to 11:30, 11:31 to 12:30, 12:31 to 1:30, 1:31 to 2:30, 2:30 to 3:30 and as the trading hours are only from 9:30 am to 3:30 pm the last window was calculated as after-hours where all the tweets being posted the trading hours that is "3:31 pm" of that day to the next opening hour of the next day which is "9:29 am" were grouped into "After hours" time window for further analysis.

Day	time_range	N1	N2	N3	N	Ñ_pos	Ñ_neg	SA_w	SA_c	Price_Change_%
2020-06-01 00:00:00	10:31-11:30	0	0	1	1	0	-2	-2	-2	-3.29115E-05
2020-06-01 00:00:00	11:31-12:30	0	0	1	1	0	-4	-4	-4	-0.000295349
2020-06-01 00:00:00	12:31-1:30	0	1	1	2	0	-5	-5	-5	0.117235761
2020-06-01 00:00:00	1:31-2:30	0	0	1	1	0	-2	-2	-2	-0.201425871
2020-06-01 00:00:00	2:31-3:30	1	0	0	1	3	0	3	3	0.088288194
2020-06-01 00:00:00	After Hours	7	16	13	36	2.142857	-3.53846	-1.55	-3.53846	-0.0185362
2020-06-02 00:00:00	10:31-11:30	0	0	1	1	0	-5	-5	-5	0.006698446
2020-06-02 00:00:00	11:31-12:30	0	1	0	1	0	0	0	0	0.009965654
2020-06-02 00:00:00	12:31-1:30	0	0	2	2	0	-4	-4	-4	0.016384167
2020-06-02 00:00:00	2:31-3:30	0	2	0	2	0	0	0	0	0.006413905
2020-06-02 00:00:00	9:30-10:30	0	1	0	1	0	0	0	0	-0.018036203
2020-06-02 00:00:00	After Hours	2	12	14	28	3	-2.71429	-2	-2.71429	-2.29115E-05
2020-06-03 00:00:00	10:31-11:30	0	1	0	1	0	0	0	0	0
2020-06-03 00:00:00	11:31-12:30	0	0	2	2	0	-2	-2	-2	-0.019116068
2020-06-03 00:00:00	12:31-1:30	0	1	0	1	0	0	0	0	0.018081781
2020-06-03 00:00:00	9:30-10:30	0	1	1	2	0	-3	-3	-3	0.026031498
2020-06-03 00:00:00	After Hours	0	9	8	17	0	-3.125	-3.125	-3.125	-3.39115E-05
2020-06-04 00:00:00	11:31-12:30	1	1	0	2	3	0	3	3	0.012682512
2020-06-04 00:00:00	1:31-2:30	0	2	1	3	0	-3	-3	-3	-0.020967651
2020-06-04 00:00:00	2:31-3:30	1	2	4	7	3	-3	-1.8	-3	0.018114122
2020-06-04 00:00:00	After Hours	3	18	6	27	2.333333	-3.33333	-1.44444	-3.33333	-0.021457803
2020-06-05 00:00:00	10:31-11:30	0	1	0	1	0	0	0	0	0.016632433
2020-06-05 00:00:00	11:31-12:30	0	2	0	2	0	0	0	0	0.009271505
2020-06-05 00:00:00	12:31-1:30	0	2	0	2	0	0	0	0	0.290175425
2020-06-05 00:00:00	1:31-2:30	1	1	1	3	2	-3	-0.5	-3	0.088557791
2020-06-05 00:00:00	2:31-3:30	0	1	0	1	0	0	0	0	0.015627692

FIGURE 4.10: SentiStrength's accuracy over different lag hours using original lexicon approach

4.4.5 Stock Market Trend Prediction Using SentiStrength's Original Lexicon

This approach examines the effectiveness of an original lexicon-based approach in predicting stock price movements. The objective of this analysis was to assess the relationship between SentiStrength sentiment scores and percentage changes in stock prices by measuring accuracy and Pearson correlation across different time lags. The baseline accuracy was established at a 0-hour lag, meaning sentiment scores were evaluated against immediate stock price movements to determine how often they correctly predicted upward or downward trends. To explore potential lagged effects, the analysis introduced time delays of 1 to 4 hours, shifting sentiment scores forward in time to examine whether earlier sentiment could predict subsequent price movements. Accuracy was calculated at each cumulative delay 1-hour, 2-hour, 3-hour, and 4-hour to determine the alignment between sentiment at time t and stock price changes at time t + lag.

Initially, a Python-coded accuracy produced a different result due to a miscalculation of accuracy. So, upon manual calculation using the TP, TN, FP, and FN the SentiStrength's default lexicons model achieved an final accuracy of 65.74% at a 0-hour lag, increasing to 69.44% at a 1-hour lag, 74.07% at a 2-hour lag, 77.77% at a 3-hour lag, and reaching a peak accuracy of 79.62% at a 4-hour lag. These results indicate that market reactions to sentiment may be delayed, with the strongest predictive alignment occurring approximately four hours after sentiment is recorded.

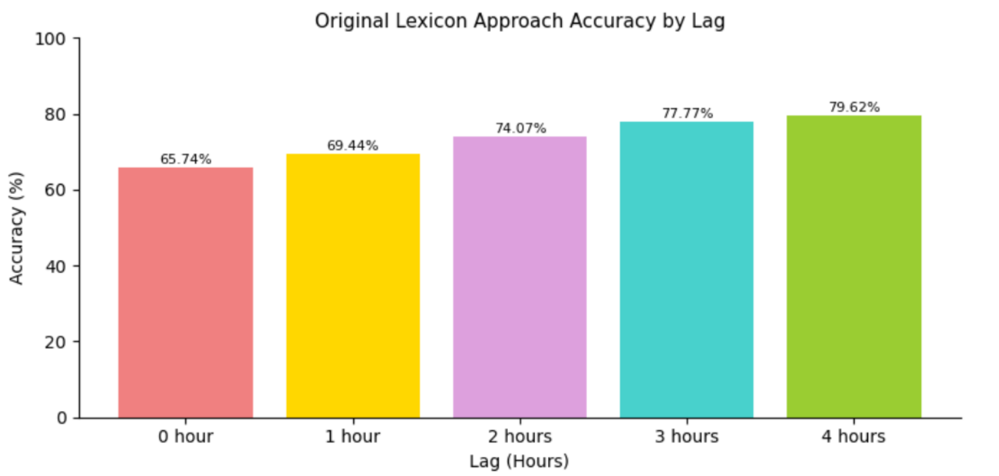


FIGURE 4.11: SentiStrength's accuracy over time using original lexicon approach

In conclusion, while the original lexicon approach is effective for general sentiment, it lacks financial domain specificity.

4.4.6 Integrating Customized Financial Lexicon into SentiStrength

Integration of financial lexicons was done by filtering out the non-financial-specific terms, and map stock-specific and negative sentiment-related words from the February 2020 dataset was chosen as it was already available and to save time from exploring another dataset, as this dataset was already available. A strategy of exploratory data analysis was implemented to analyze the most frequently occurring financial-specific, stock-related, positive, and negative sentiment.

Step 1 : Identify the most common words in the dataset Initially the top 50 common words from the dataset were identified (for eg. us: 143, coronavirus: 113, new: 92, ft: 55, pandemic: 53, covid: 51, uk: 50, trump: 50, one: 41, donald: 39, read: 34, lockdown: 34, first: 32, hong: 31, economic: 31.)

```
    # Count most common words
    word_counts = Counter(words).most_common(100)

    print("\nMost Common Words:")
    for word, count in word_counts:
        print(f"{word}: {count}")
else:
    print("No 'Translation' column found in dataset.")
```

```
Most Common Words:
us: 143
coronavirus: 113
new: 92
ft: 55
pandemic: 53
covid: 51
uk: 50
trump: 50
one: 41
donald: 39
read: 34
lockdown: 34
first: 32
hong: 31
economic: 31
```

FIGURE 4.12: Most common words from February 2020 dataset

Step 2: Identify and categorize the most common words into stock-related, positive sentiment, and negative sentiment with their word count.

```
# Load the first few rows of the dataset
df = pd.read_excel(xls, sheet_name="Sheet1")
df.head()
```

	Stock_Word	Stock_w_Count	Positive_Sentiment	Positive_Sentiment_w_count	Negative_Sentiment	Negative_Sentiment_w_count
0	pandemic	53	recovery	13.0	pandemic	53.0
1	economy	27	increase	7.0	protests	25.0
2	government	27	trust	6.0	crisis	22.0
3	financial	27	pay	5.0	war	11.0
4	crisis	22	hope	5.0	struggling	9.0

FIGURE 4.13: Stock related words, negative and positive sentiments

These words were checked with the default and the lexicons provided by Hao et al. [2022] and the existing terms were enhanced, and non-existing terms were added and assigned correct sentiment scores. These non-existing terms were designed from the investor and financial-specific point of view, and updated files were passed to SentiStrength for further analysis.

FIGURE 4.14: Positive and negative sentiment words in the February 2020 dataset

The above wordcloud shows the positive and negative sentiments in the February 2020 dataset.

Step 3: Categorized the words to assign weights.

As stock fluctuations usually depend upon different sentiment factors, a strategy for integrating new lexicons was carried out by focusing on extracting new lexicons from different contexts, like geopolitics. For instance, there was an ongoing conflict between different geographic areas, event-driven, and domain-specific. The study focused on implementing specialized lexicons from different aspects. These lexicons were manually collected from the February 2020 dataset to predict the June 2020 stock market. The lexicons were collected based on different contexts, which included company-based such as "Apple", "Zoom", etc., sector-based including "technology", "Finance" and

"Healthcare", time-sensitive lexicons that trigger words during specific times, like earnings and losses, and geopolitical-event lexicon (e.g. wars) and event-driven lexicon such as COVID-19.

Below are some examples for the manually collected lexicons from different categories of context terms from the dataset.

- **Financial-Lexicon Terms:** "market," "economic," "economy," "trade," "companies," "group",
- **Investor-Related-Lexicon Terms:** "wirecard," "bn" (billion), "financial", Emotions: "pandemic," "lockdown," "crisis",
- **Sector Based Lexicon Terms:** "global," "trade," "government," "protests", "Apple", "Zoom"
- **Event-impact Lexicon Terms:** coronavirus, covid, pandemic, lockdown.
- **Events and People-driven Lexicons:** trump, Donald, protests, government, war, america, china, europe, hong kong, wirecard.

The SentiStrength uses a general sentiment lexicon, but financial terms (e.g., bullish, bearish, downgrade, rally) may not be recognized correctly. The solution to the above problems of context misinterpretations and limitations of financial lexicons is to enhance SentiStrength's lexicon and refine sentiment weights to enhance its performance and improve the stock market predictions. While assigning the right sentiment scores, the sentiment scores assigning criteria were as followed:

- 5 (significantly positive): significantly positive influence on the stock market, such as economy policies changes, interest rate change, e.g. stock price will increase more than 5%.
- 4 (quite positive): strong positive influence on the stock market, such as well-known company (e.g., APPLE) statements, e.g. stock price will increase 4-5%.
- 3 (moderate positive): moderate positive influence on stock market, such as dividends announcements. For example, 2-3% increase on the stock market comparing with the previous trading day?
- 2 (slightly positive): slightly positive influence on stock market, such as demands increase of companies. For example, 0.3-1% increase of price on the stock market with the previous trading day?

- 1 (barely positive): a little positive influence on stock market, such as some popular investors predictions.
- 0 (neutral): no effects on stock market, such as nonrelated areas and advertisements.
- -1 (barely negative): a little negative influence on stock market, such as rumour from medias.
- -2 (slightly negative): weak negative influence on stock market, supply demands changes
- -3 (moderate negative): moderate negative influence on stock market, such as products failures and amassing debt
- -4 (quite negative): strong negative influence on stock market, such as negative political factors, exchange rates.
- -5 (significantly negative): significantly negative influence on stock market and it may cause trading curb etc, such as covid-19, war, and natural disasters. This was implemented by applying the strategy of identifying and adding the frequently occurring financial-specific terms and customizing the SentiStrength's lexicon files like EmotionLookupTable.txt, NegatingLookupTable.txt, and IdiomsLookuptable.txt, and BoosterWordLookupTable.txt, assigning correct sentiment scores for financial sentiment analysis. The updated input lexicon files were passed to SentiStrength.

4.4.7 Stock Market Trend Prediction Using Customized Financial Lexicon Approach

SentiStrength now includes a financial domain-specific lexicon, which improves sentiment analysis for stock market prediction. This approach attempted to improve sentiment categorisation of financial tweets by using domain-relevant phrases and assigning sentiment weights based on the stock market context. By refining sentiment detection, the goal was to better capture market sentiment and its relationship with stock price movements. Upon manual calculation of TP, TN, FP, and FN, the enhanced financial lexicon produced an initial accuracy at a 0-hour lag of 62.11%, which was lower than the baseline accuracy observed with SentiStrength's default lexicons. However, after adjusting for lag effects, the financial lexicon showed a considerable improvement in prediction performance. The accuracy increased with time delays, reaching 83.16% at a 1-hour lag, 89.47% at a 2-hour lag, 94.74% at a 3-hour lag, and peaking at 97.89% with a 4-hour delay.

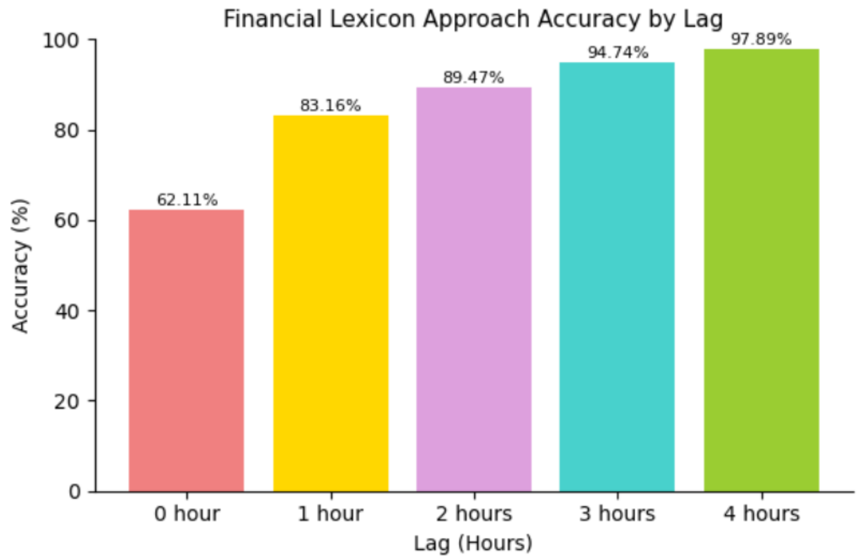


FIGURE 4.15: SentiStrength's accuracy over time using customized and enhanced financial lexicon approach

These findings imply that, while initial stock price fluctuations may not always correspond to sentiment scores, the market's responsiveness to financial mood grows stronger with time. These results suggest that while immediate stock price movements may not always align with sentiment scores, the market's reaction to financial sentiment becomes more pronounced over time. The enhanced lexicon likely captures more nuanced sentiment shifts relevant to financial contexts, allowing for stronger predictive alignment as the time lag increases.

The sharp increase in accuracy with longer time delays supports the hypothesis that sentiment-driven market reactions are not instantaneous but instead develop over a period of hours. This delayed response could be due to traders analysing sentiment trends, institutional decision-making procedures, or the distribution of financial news among market players. Compared to the default SentiStrength lexicons, which obtained a peak accuracy of 79.62% at a 4-hour latency, the financial lexicon approach demonstrated a much greater accuracy of 97.89% for the same lag. This improvement highlights the importance of domain-specific sentiment lexicons in capturing financial market sentiment and improving stock movement prediction. It was observed that many tweets mentioned events and contexts outside the USA that have a lesser impact on the S&P500. To increase the accuracy and performance of sentiment analysis further, a context-based financial sentiment analysis approach was implemented futher.

4.4.8 Context-Based Sentiment Analysis using Financial Lexicon Approach

The final approach was the context-based sentiment analysis approach, implemented to reduce noise in the data to increase the performance of SentiStrength's SA. This method included re-processing the June 2020 dataset to develop a focused analysis by filtering irrelevant tweets that were referring to events or contexts outside the USA that did not predict S&P500 and diluted the quality of sentiment analysis and its relationship with the stock market trends. The Tweets were manually extracted and filtered to ensure clean and relevant data and alignment with financial-domain-specific tweets only. This approach, after manual calculations of accuracy resulted in a decreased accuracy of accuracy of 56.66% for a 0-hour delay, 76.66% for a 1-hour delay, 81.66% for a 2-hour delay, 88.33% for a 3-hour delay, and 93.33% for a 4-hour delay. This suggests that removing irrelevant tweets was helpful, but some information might have been lost.

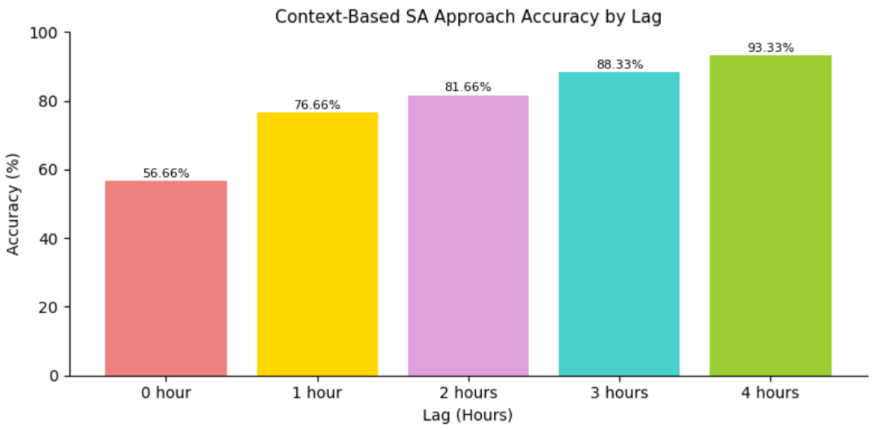


FIGURE 4.16: Accuracy of Context-Based Sentiment Analysis using Financial Lexicon Approach

Conlusion: The results from the original and financial lexicon approaches showed that the accuracy of financial lexicons improved compared to the original lexicon because the original lexicon had the presence of extreme sentiment scores frequently which captured general sentiment intensity but introduced more noise in the data whereas the financial lexicon is more domain-specific which led to less extreme sentiment scores which may be suitable for financial text. Other reasons might also include a mismatch between the lexicons and the tweets that were being worked on. For example, the language used in financial news might be different from what the lexicon was built on. Financial sentiment lexicons are often limited to focus on general financial jargon, so it might have not captured the nuances of the evolution of terminology used in June 2020 which also led to contextual misinterpretation where some words in financial news might have different meanings based on the context e.g., "hit" can mean both positive and neutral depending upon the context and may have different sentiment meaning in a financial context. This misinterpretation of sentiment suggests inverse sentiment effects where negative sentiment leads to price increases and positive sentiment leads to declines. These two approaches served as baseline methods to compare the tool's performance with customized lexicons.

4.5 Machine Learning Models for Stock Prediction

This section explores various machine learning models like Logistic Regression, XGBoost, and Random Forests in stock market predictions based on sentiment scores.

4.5.1 Model Selection and Training

Since the goal is to classify the price increases and decreases, models such as Logistic Regression, XGBoost, and LSTMs were developed for predictions to check which model best captures the relationship between sentiment and stock price movement. Random forests will be used to understand which feature contributes the most to predictions. The performance of these model will be then compared with the results of SentiStrength's Predictions to evaluate if sentiment analysis improves forecasting. The dataset was split into 80% for training and 20% as testing sets. The dataset was trained and evaluated using Logistic regression, XGBoost. Feature importance analysis using Random Forest.

Logistic regression achieved an accuracy of 58.06%, XGBoost achieved an accuracy of 45.16%.

4.5.2 Feature Engineering

Since only SA_c was being used to predict the movements, which may not be enough feature engineering was implemented by adding features like lagged sentiment features such as SA_c from previous hours, moving average of sentiment scores by computing a rolling mean of SA_c over past few hours, and time-based features like including hour of the day and day of the week. The features like N1, N2, N3 and N were normalized to prevent large values from dominating the model as introduced in the study of Sandaka [2024] N1norm = N1 / N, N2norm = N2 / N, N3norm = N3/N Computed the sentiment proportions to the ratio of positive, neutral, and negative tweets using, Pos_Ratio = N1 / N, Neg_Ratio = N3 / N Calculated the Lag Feature Sentiment Momentum by adding previous day's sentiment values to capture trends as, SA_c_lag1 = SA_c[t-1] (previous day's compound sentiment) and N1_lag1 = N1[t-1] (previous day's positive tweet count).

```python
# Ensure no division by zero
df['N'] = df['N'].replace(0, 1)  # Avoid division by zero

# 1. Defining normalizing features
df['N1_norm'] = df['N1'] / df['N']
df['N2_norm'] = df['N2'] / df['N']
df['N3_norm'] = df['N3'] / df['N']

# 2. Defining sentiment proportions
df['Pos_Ratio'] = df['N1'] / df['N']
df['Neg_Ratio'] = df['N3'] / df['N']

# 3. Defining lag features
df['SA_c_lag1'] = df['SA_c'].shift(1)
df['N1_lag1'] = df['N1'].shift(1)
df['N3_lag1'] = df['N3'].shift(1)

# 4. Defining rolling sentiment trends
df['SA_c_rolling3'] = df['SA_c'].rolling(window=3).mean()
df['SA_c_rolling5'] = df['SA_c'].rolling(window=5).mean()

# Drop NaN values introduced by lag and rolling operations
df.dropna(inplace=True)

# Save updated dataset
df.to_excel("C:\\Users\\omana\\Downloads\\Finance Lexicon for SentiStrength\\enhanced files\\enhanced_features.xlsx", index=False)

print("Feature engineering complete! New dataset saved as enhanced_features.xlsx")
```

FIGURE 4.17: Implementation of enhanced sentiment features

N1_norm	N2_norm	N3_norm	Pos_Ratio	Neg_Ratio	SA_c_lag1	N1_lag1	N3_lag1	SA_c_rolling3	SA_c_rolling5
1	0	0	1	0	-2	0	1	-1.333333333	-2
0.194444444	0.444444444	0.361111111	0.194444444	0.361111111	3	1	0	-0.846153846	-2.307692308
0	0	1	0	1	-3.538461538	7	13	-1.846153846	-2.507692308
0	1	0	0	0	-5	0	1	-2.846153846	-1.507692308
0	0	1	0	1	0	0	0	-3	-1.907692308
0	1	0	0	0	-4	0	2	-1.333333333	-2.507692308
0	1	0	0	0	0	0	0	-1.333333333	-1.8
0.071428571	0.428571429	0.5	0.071428571	0.5	0	0	0	-0.904761905	-1.342857143
0	1	0	0	0	-2.714285714	2	14	-0.904761905	-1.342857143
0	0	1	0	1	0	0	0	-1.571428571	-0.942857143
0	1	0	0	0	-2	0	2	-0.666666667	-0.942857143
0	0.5	0.5	0	0.5	0	0	0	-1.666666667	-1.542857143
0	0.529411765	0.470588235	0	0.470588235	-3	0	1	-2.041666667	-1.625
0.5	0.5	0	0.5	0	-3.125	0	8	-1.041666667	-1.025
0	0.666666667	0.333333333	0	0.333333333	3	1	0	-1.041666667	-1.225
0.142857143	0.285714286	0.571428571	0.142857143	0.571428571	-3	0	1	-1	-1.825
0.111111111	0.666666667	0.222222222	0.111111111	0.222222222	-3	1	4	-3.111111111	-1.891666667
0	1	0	0	0	-3.333333333	3	6	-2.111111111	-1.266666667
0	1	0	0	0	0	0	0	-1.111111111	-1.866666667
0	1	0	0	0	0	0	0	0	-1.266666667
0.333333333	0.333333333	0.333333333	0.333333333	0.333333333	0	0	0	-1	-1.266666667
0	1	0	0	0	-3	1	1	-1	-0.6
0.103448276	0.551724138	0.344827586	0.103448276	0.344827586	0	0	0	-2	-1.2
0	1	0	0	0	-3	3	10	-1	-1.2
0	0	1	0	1	0	0	0	-1.666666667	-1.6
0.181818182	0.545454545	0.272727273	0.181818182	0.272727273	-2	0	1	0.5	-0.3
1	0	0	1	0	3.5	2	3	1.5	0.3
0	1	0	0	0	3	1	0	2.166666667	0.9

FIGURE 4.18: Enhanced Sentiment Features

Computed the Rolling Sentiment Trends to capture trends using the code snippet below above and the enhanced features were then saved into .xlsx file which was then used for implementing to the hybrid appraoch.

4.6 Implementation of Hybrid Approach

This work takes a hybrid method, combining a financial lexicon sentiment analysis approach of SentiStrength, with machine learning models to improve stock market prediction. While typical sentiment analysis systems like SentiStrength provide sentiment scores using predetermined lexicons, their accuracy in financial contexts is limited due to the absence of domain-specific terminology and the inability to identify nuanced sentiment changes.To address this, an enhanced financial lexicon was developed, allowing for more accurate sentiment detection in financial news and discussions. The sentiment scores obtained from SentiStrength, including SA_c (compound sentiment score), lagged sentiment values, and sentiment ratios, were used as features in machine learning models such as Logistic Regression, XGBoost, and along with one more advance LSTM LSTM. By incorporating these sentiment-driven features along with stock price movements, the hybrid approach used both text-based sentiment insights and data-driven predictive modeling, and model Logistic Regression, XGBoost and LSTM model acquired an accuracy of 96.42%, 96.42% and 75% respectively. Although the accuracy increased after feature engineering compared to the baseline machine learning model accuracy, but still did not perform as well as the financial lexicon sentiment analysis approach.

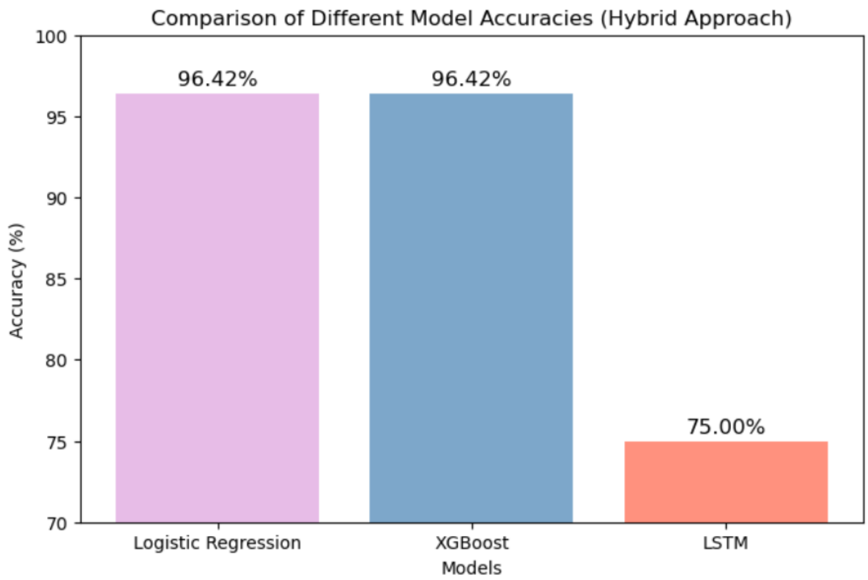


Figure 4.19: Accuracy of different machine learning models after implementation of sentiment features to predict stock market trends

Advantages of Hybrid Model Approach shown in the study of Rakshit [2024] are as follows:

- Enhanced Sentiment Representation: By refining SentiStrength's lexicon, the sentiment scores better reflect financial market sentiment, leading to more meaningful input features for ML models.
- Combining Expert Knowledge with Data-Driven Insights: Rule-based sentiment analysis encodes linguistic expertise, while ML models identify patterns that are not explicitly captured by lexicons.

- Improved Stock Movement Prediction: Sentiment alone is insufficient for market prediction, but when combined with machine learning, it enhances the model's ability to detect patterns in historical stock trends and sentiment shifts.
- Flexibility and Adaptability: ML models can learn from historical data and adjust their weightage on sentiment scores dynamically, improving generalization across different market conditions.
- Empirical Performance Gains: Experimental results showed that machine learning models alone outperformed SentiStrength, but the best performance was achieved when sentiment features were used alongside ML models, achieving an accuracy of 96.4%.

Results The results demonstrated that the first approach achieved an accuracy of 65.74%, which increased to 79.62% with a 4-hour lag feature. The second approach attained 62.11% accuracy, and accuracy rose to 97.89% with a 4-hour lag effect, enhancing performance. The third approach had an initial accuracy of 56.66%, but after a 4-hour lag effect, integration accuracy improved to 93.33%. Stock movements were then predicted using machine learning models, including Logistic Regression and XGBoost attaining 58.06% and 45.16% accuracies, respectively. Further to enhance the predictive power, feature engineering was implemented, generating new features integrated into machine learning models. This hybrid approach, combining sentiment analysis with machine learning included Logistic regression, XGBoost, and LSTM accuracies by 96.42%, 96.42%, and 75%, respectively. Future research can explore deep-learning-based sentiment models like FinBert in real-time trading applications.

Chapter 5

Evaluation Framework

5.1 Introduction

The evaluation in this research played an important role as it helped understand and analyse the effectiveness of financial sentiment analysis compared with the generic sentiment analysis using default lexicons of the tool. The enhanced SenitStrength model performed effectively in predicting stock market movements using sentiments compared to the traditional machine learning models. This chapter compares the original SentiStrength tool with financial lexicon-enhanced version of SentiStrength to determine whether the domain-specific modifications introduced meaningful improvements.

The central objective of this evaluation is not only to assess overall prediction accuracy, but to examine the model's ability to correctly identify both upward and downward stock price movements. Given that sentiment-driven investment strategies depend on reliable signals for both buying and selling decisions, a balanced evaluation framework that includes sensitivity (true positive rate) and specificity (true negative rate) is essential.

Unlike traditional sentiment evaluation, where precision or recall alone may suffice, this project emphasizes a dual-perspective analysis. Therefore, each version of the tool was evaluated using key performance metrics—including overall accuracy, sensitivity, and specificity—across multiple hourly intervals. The evaluation also involved computing confusion matrices at each time step to track the distribution of true positives, true negatives, false positives, and false negatives.

To ensure the evaluation remains grounded in real-world stock behaviour, the ground truth was defined based on actual stock price movements: a label of 1 represented an upward trend and 0 represented a downward trend. These values were then compared to the sentiment-based predictions generated by both the generic and enhanced models. Through this rigorous approach, the chapter aims to demonstrate the practical value of sentiment analysis in financial prediction and establish whether domain-specific enhancements truly offer a significant performance advantage.

5.2 Evaluation of Sentiment Analysis

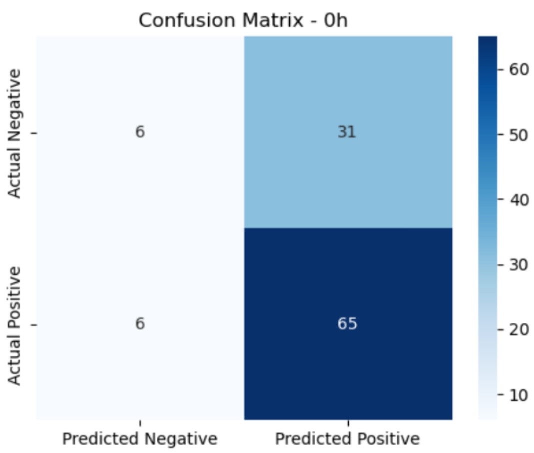
5.2.1 Introduction

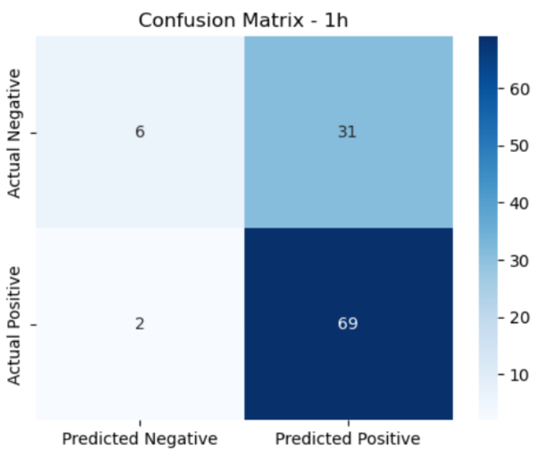
To ensure correct results and evaluation, the study involved manual identification of TP, TN, FP, and FN to further calculate the accuracy, sensitivity, and specificity. FThese values were identified for each approach and every hour delay to acheive accurate results. This section evaluated and compared the two different sentiment analysis approaches to check which is more effective and reliable for stock market prediction, which used performance metrics, specificity, and recall calculations. for different time intervals from 0h to 4h based on true positives (TP), true negatives (TN), false negatives (FN), and false positives (FP). The first two sections compare the evaluation of the original and financial lexicon-based sentiment analysis.

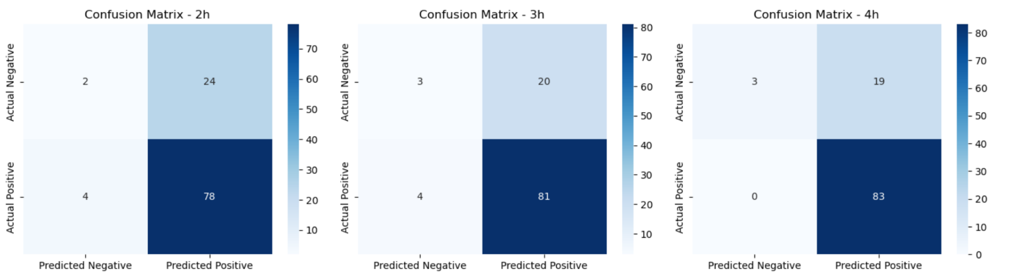
5.2.2 Original Lexicon Sentiment Analysis Approach

The original sentiment analysis approach acquired an overall accuracy of 65% at a 0-hour steadily increased to 81.90% below is the confusion matrix that shows the number of TP, TN, FP and FN to further calculated the specificty, sensitivity and accuarcy metrics.

Confusion Matrix:







The figure below shows the evaluation metrics for the 0, 1, 2, 3, and 4 hour time interval accuracies, sensitivity and specificity, respectively.

Time Interval	Overall Accuracy	TP Accuracy (Sensitivity)	TN Accuracy (Specificity)
0h	0.6574	0.9155	0.1622
1h	0.6944	0.9718	0.1622
2h	0.7407	0.9512	0.0769
3h	0.7778	0.9529	0.1304
4h	0.8190	1.0000	0.1364

FIGURE 5.1: Original lexicon's Accuracy, Sensitivity, and Specificity

Results : Sensitivity begins at 0.9155 at 0h and improves with time as the highest recall occurs at 4 hours 1.0, indicating that all genuine positives have been correctly identified at that time. This demonstrates that the model is getting better at identifying positives, but at the expense of producing more false positives since specificity remains low. The model prioritises recall above specificity, thus, it finds more true positives while misclassifying many negatives as positives.

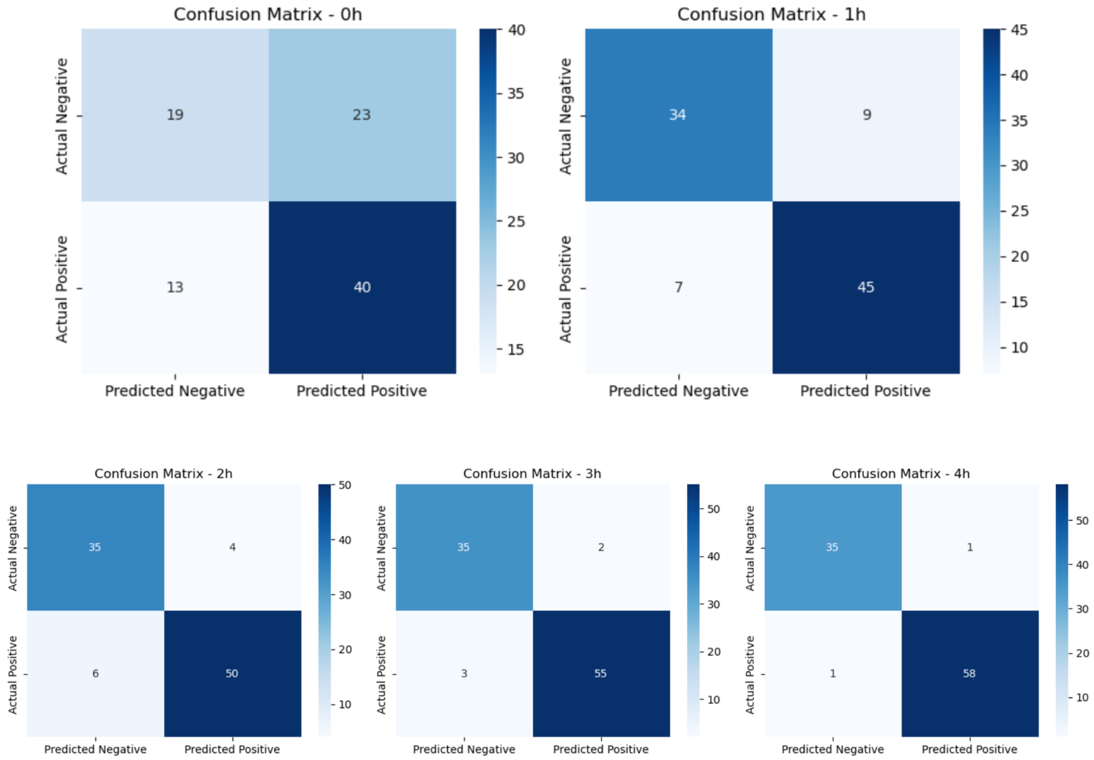
Sensitivity increases with time, reaching 100% after 4 hours, demonstrating the model correctly recognises all positives.. Specificity remains poor, indicating that the model incorrectly classifies many negatives as positives. This could indicate that as time passes, the model forecasts more bullish sentiment (positives) but struggles to correctly identify negative sentiment.

5.2.3 Financial Lexicon approach

The financial lexicon approach acquired an accuracy of 62.11% at 0 hours and increased its accuracy by 97.89% at 4 hours, which showed that the model was enhanced and performed better than the original lexicon approach.

Confusion Matrix

The figure below shows the evaluation metrics, overall accuracy, sensitivity, and specificity of 0, 1, 2, 3, and 4 hour time intervals.



Time Interval	Overall Accuracy	TP Accuracy (Sensitivity)	TN Accuracy (Specificity)
0h	0.6211	0.7547	0.4524
1h	0.8316	0.8654	0.7907
2h	0.8947	0.8929	0.8974
3h	0.9474	0.9483	0.9459
4h	0.9789	0.9831	0.9722

FIGURE 5.2: Original lexicon's Accuracy, Sensitivity, and Specificity

Results: TP & TN increase over time, showing financial lexicon performs better with delay. FP & FN decrease, meaning mistakes reduce as time progresses. High Sensitivity & Specificity of the model indicated that it strongly predicted both price increases and decreases. At 0h, sensitivity jumps from 75.47% to 86.54%, indicating the financial lexicon is much better at detecting when prices go up. At 4h, sensitivity reaches 98.31%, showing that over time, the financial lexicon becomes extremely effective at predicting price increases and improves specificity which is the TN Rate too. At 0h, from 45.24% to 79.07%, proving that false alarms for price increases (FP) are lower. At 4h, specificity reaches 97.22%, meaning it's also excellent at identifying price decreases correctly.

Conclusion In conclusion the financial lexicon becomes more reliable over time with each hour of delay, and both sensitivity & specificity increase steadily. This suggests that the financial lexicon's predictions get more accurate as the model has more data to analyze. Financial Lexicon has significantly higher specificity (TN accuracy) at all time intervals. Original Lexicon struggles to predict stock price decreases (low TN accuracy), while the Financial Lexicon corrects this, reaching 97.22% specificity at 4h. This means that the Financial Lexicon is much better at predicting price drops, which is crucial for trading strategies. Financial Lexicon maintains strong sensitivity while improving specificity. The original lexicon is biased toward predicting price increases (high sensitivity, low specificity). The financial lexicon balances both, ensuring good accuracy for both up and down movements.

5.2.4 Comparison of Accuracy Sensitivity, and Specificity

The plot clearly visualizes the performance differences between the original lexicon and the financial lexicon.

Overall Accuracy: The financial lexicon consistently outperforms the original lexicon across all time intervals. The accuracy gap widens as the time interval increases. Sensitivity (TP Accuracy - Correctly Predicting Price Increases): The original lexicon starts with a high sensitivity but plateaus near 1.0 at 4 hours.The financial lexicon also maintains high sensitivity, but it starts lower and gradually improves. Specificity (TN Accuracy - Correctly Predicting Price Decreases):The biggest difference is in specificity: the financial lexicon performs significantly better. The original lexicon struggles with predicting price decreases, staying below 0.2 across all time intervals. The financial lexicon achieves balanced performance, improving specificity close to 1.0 by 4 hours.

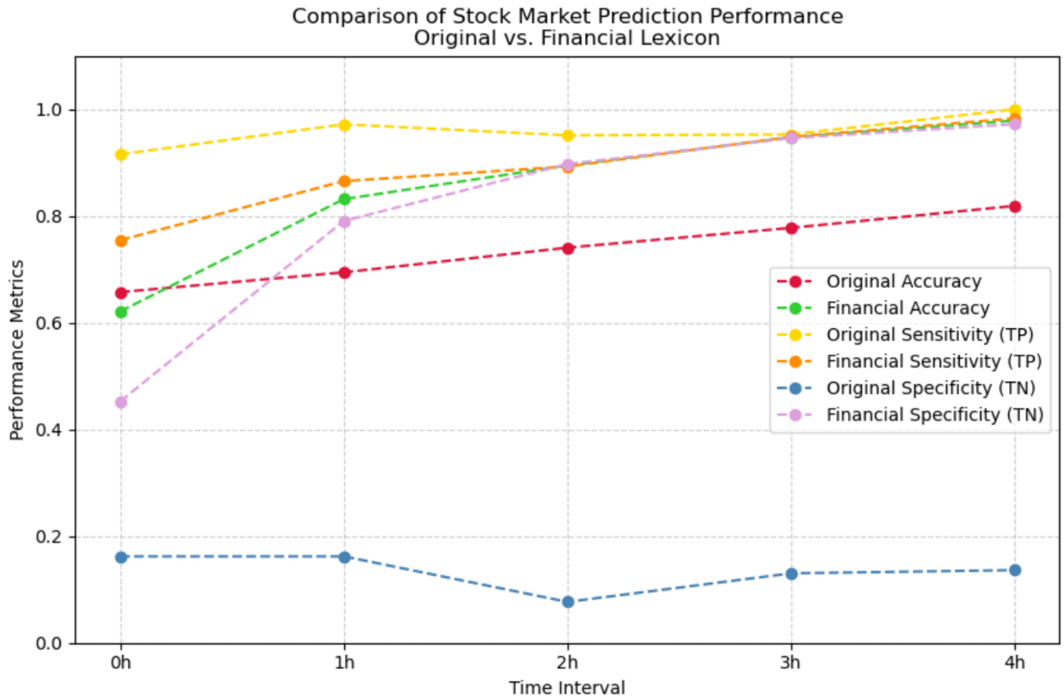


FIGURE 5.3: Comparison of Stock Market Prediction Performance of Original vs. Financial Lexicon

Results: Enhancing the financial lexicon significantly improved the accuracy of sentiment analysis for stock market prediction. The default SentiStrength lexicon struggled to interpret financial terms correctly, often misclassifying words like "bullish" (positive sentiment) or "crash" (negative sentiment). By integrating a custom financial lexicon, sentiment classification became more aligned with market-specific language, reducing misinterpretations and improving the quality of sentiment-driven features used in predictive models. The results showed that models trained with enhanced sentiment features from the financial lexicon achieved higher accuracy, precision, and recall compared to those using the original lexicon. This confirms that domain-specific sentiment analysis is crucial for effectively linking social media sentiment with stock market trends.

Cnclusion: The original lexicon is biased toward predicting price increases, with high sensitivity but very low specificity.The financial lexicon achieves a better balance between sensitivity and specificity, meaning it can predict both price increases and decreases more accurately. This suggests that the financial lexicon provides a more reliable method for stock market sentiment prediction. Financial Lexicon Leads to More Reliable Predictions Over Time. At 0h (immediate prediction), the original lexicon is overly optimistic about price increases (high sensitivity, low specificity). However, as time progresses (1h to 4h), the financial lexicon's performance stabilizes, achieving nearly perfect accuracy in both TP and TN. This suggests that financial sentiment signals take time to reflect in actual stock movements—making delayed predictions more accurate. Financial Lexicon

Reduces False Alarms (False Positives & False Negatives) The original lexicon frequently misclassified stock movements (high false positives and false negatives). The financial lexicon reduces these errors, meaning fewer false buy/sell signals—this is essential for traders to minimize risk. Financial Lexicon Improves the Balance Between TP (Price Up) & TN (Price Down) Predictions The original lexicon has very high sensitivity (TP accuracy), meaning it predicts price increases well but struggles to detect price decreases (low specificity). The financial lexicon significantly improves specificity (TN accuracy) at all time intervals, meaning it better detects price drops.

5.2.5 Evaluation of Machine Learning Models Approaches

As the financial lexicon approach was outperforming the original lexicon approach. The study further evaluates the comparison of different models using the financial lexicon approach. This section covers the results of the classification report comparison of different machine learning models. Primarily, the models were tested without implementing new features. It is seen that after feature engineering and implementing new features to the machine learning models improved their prediction accuracy.

1. Results of Logistic Regression accuracy before feature engineering 58.06%.

Confusion Matrix:

Actual vs Predicted	**Predicted 1**	**Predicted 0**
Actual 1 (Price up)	0 (TP)	13 (FN)
Actual 0 (Price down)	0 (FP)	18 (TN)

TABLE 5.1: Confusion Matrix for Stock Price Movement Prediction

Prediction	**Precision**	**Recall**	**F1-Score**	**Support**
1 (Price up)	0.00	0.00	0.00	13
0 (Price down)	0.58	1.00	0.73	18

TABLE 5.2: Logistic Regression classification Report for Stock Price Movement Prediction

This model did a good job predicting when the stock price would go down, it got all of those cases right (100% recall), and was correct most of the time when it guessed, with 58% precision. But when it came to predicting price increases, the model didn't get a single one right. It simply never predicted an "up" movement, which led to zero scores across precision, recall, and F1 for that class. This shows the model is biased toward one direction, likely due to imbalanced data or limited signal for upward trends. It was a clear sign that further tuning or feature improvements were needed to make the model fairer and more accurate.

2. Results of XGBoost accuracy before featuring 45.16%.

Confusion Matrix:

Actual vs Predicted	**Predicted 1**	**Predicted 0**
Actual 1 (Price up	3 (TP)	10 (FN)
Actual 0 (Price down)	7 (FP)	11 (TN)

Table 5.3: Confusion Matrix for Stock Price Movement Prediction

Prediction	**Precision**	**Recall**	**F1-Score**	**Support**
1 (Price up)	0.30	0.23	0.26	13
0 (Price down)	0.52	0.61	0.56	18

Table 5.4: Classification Report for Stock Price Movement Prediction

This time, the model tried to predict both directions—price going down and up. It was better at catching the "down" cases, with decent recall 61% and a balanced F1-score 56%. For "up" movements, it struggled more. It only caught a few, with lower precision and recall, resulting in a modest F1-score of 26%. Still, this was a step forward compared to the earlier version—it started recognizing both classes, even if not perfectly. It showed signs of learning, but there was still room to improve balance and accuracy. XGBoost is less biased than compared to logistic regression, but still struggles to predict the movements.

Results after the implementation of feature engineering

1. Logistic Regression Accuracy: 0.9642.

Confusion Matrix:

Actual vs Predicted	**Predicted 1**	**Predicted 0**
Actual 1 (Price up	13 (TP)	0 (FN)
Actual 0 (Price down)	1 (FP)	14 (TN)

Table 5.5: Confusion Matrix for Stock Price Movement Prediction

Prediction	**Precision**	**Recall**	**F1-Score**	**Support**
1 (Price increase)	0.93	1.00	0.96	13
0 (Price decrease)	1.00	0.93	0.97	15

Table 5.6: Classification Report for Stock Price Movement Prediction

The model's performance is measured using precision, recall, and F1-score. For price decrease class 0, the model has perfect precision 1.00 but a recall of 0.93, meaning it misses 7% of actual price decreases. The F1-score is 0.97. For price increase class 1, it correctly identifies all price increases recall 1.00 but has a precision of 0.93, meaning it makes some false positives. The F1-score is 0.96. The support values show the number of instances in each class 15 for price decrease, 13 for price increase.

2. XGBoost Accuracy: 0.9642

Confusion Matrix:

Actual vs Predicted	**Predicted 1**	**Predicted 0**
Actual 1 (Price up)	15 (TP)	0 (FN)
Actual 0 (Price down)	1 (FP)	12 (TN)

Table 5.7: Confusion Matrix for Stock Price Movement Prediction

Prediction	**Precision**	**Recall**	**F1-Score**	**Support**
1 (Price increase)	1.00	0.92	0.96	13
0 (Price decrease)	0.94	1.00	0.97	15

Table 5.8: Classification Report for Stock Price Movement Prediction

The model's performance shows a precision of 0.94 for price decrease class 0, meaning it correctly predicts 94% of price decreases, with a perfect recall of 1.00, correctly identifying all actual price decreases. The F1-score is 0.97. For price increase class 1, the model has perfect precision 1.00 but a recall of 0.92, meaning it misses 8% of actual price increases. The F1-score is 0.96. The support values indicate 15 instances for price decrease and 13 for price increase.

3. LSTM Accuracy: 0.75

Prediction	**Precision**	**Recall**	**F1-Score**	**Support**
1 (Price increase)	0.69	0.85	0.76	13
0 (Price decrease)	0.83	0.67	0.74	15

Table 5.9: Classification Report for Stock Price Movement Prediction

The model shows a precision of 0.83 for price decrease (class 0), meaning it correctly predicts 83% of price decreases, but its recall is 0.67, indicating it misses 33% of actual price decreases. The F1-score is 0.74. For price increase (class 1), the precision is 0.69, meaning it makes some false positive predictions, while the recall is 0.85, correctly identifying 85% of actual price increases. The F1-score is 0.76. The support values show 15 instances for price decrease and 13 for price increase.

5.3 Time-Series Comparison of Different Approaches in Stock Market Prediction

Time series analysis is essential in evaluating the relationship between sentiment scores and stock price movements over time. By analyzing the predictive performance of different sentiment approaches across various time lags, we can assess their effectiveness in capturing market sentiment dynamics and forecasting stock price trends. Comparison of Sentiment-Based Approaches over multiple time lags (0h, 1h, 2h, 3h, and 4h) to determine how well sentiment scores aligned with stock price movements. The three original lexicon, enhanced financial lexicon, and context-based approaches were examined.

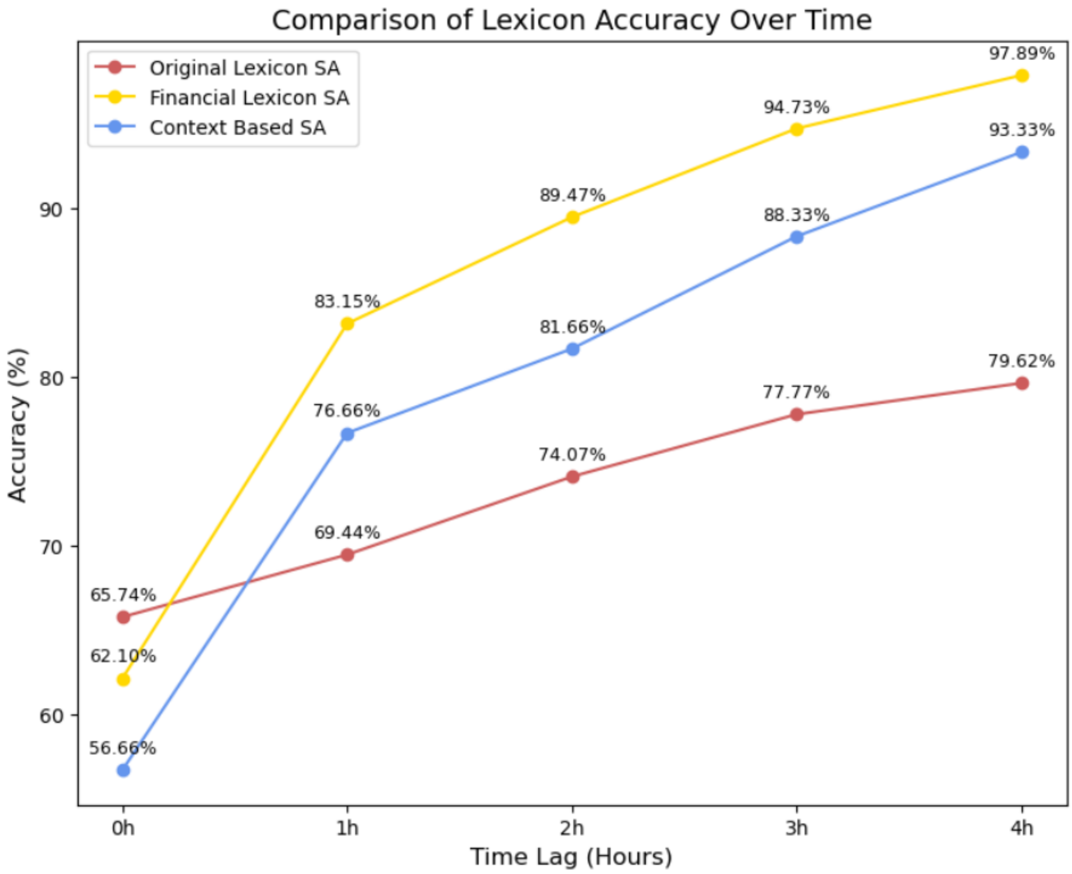


Figure 5.4: SentiStrength's accuracy comparison between different Sentiment Analysis approaches

1. Default Lexicon-Based Approach – Utilizing SentiStrength's original lexicons showed the accuracy increase gradually over time, starting at 65.74% with no lag and peaking at 79.62% at a 4-hour lag. This suggests that the market does not react to sentiment immediately but exhibits delayed responses, with the strongest alignment occurring after several hours.

2. Financial Lexicon-Based Approach – Incorporating an enhanced financial lexicon designed for stock market sentiment. The model initially achieved a lower accuracy (61.0%) at 0-hour lag, indicating that financial sentiment alone may not immediately reflect stock price changes. However, accuracy increased significantly with longer delays, reaching 97% at a 4-hour lag. This suggests that financial sentiment may have a more substantial predictive influence on stock movements when accounting for delayed market reactions.

3. Context-Based Approach - Using financial enhanced lexicons and a noise reduction method of removing unnecessary tweets that were not related to the U.S context achieved an initial accuracy of 56.66% to an improved accuracy of 93.33%, which performed better than the original lexicon approach but not as well as the financial lexicon approach.

Since the enhanced financial approach performed better than the context-based sentiment analysis approach. Further comparison of the evaluation of different machine learning models used a financial lexicon approach for prediction of the stock movements, and comparison of the original and financial lexicons was considered.

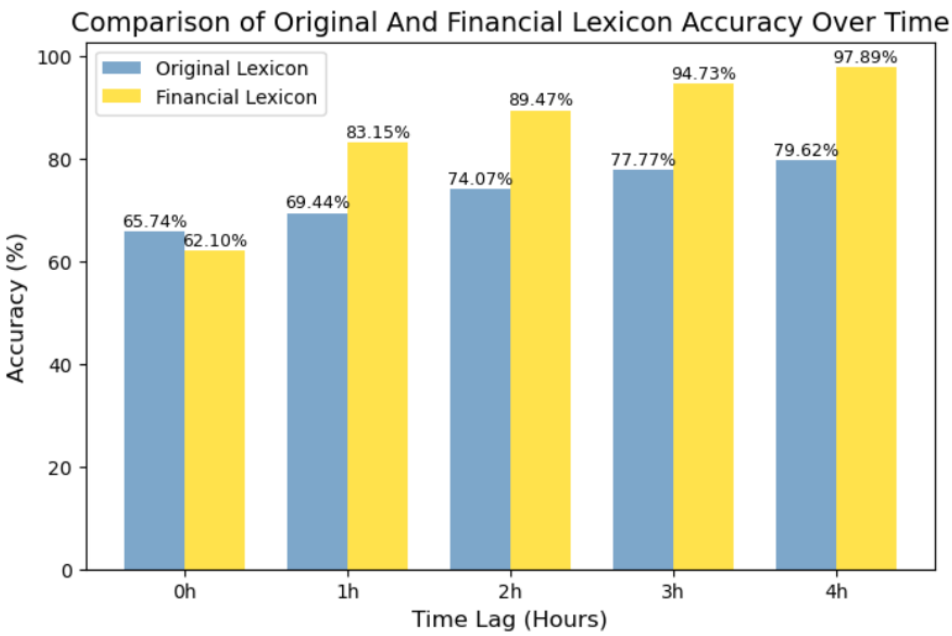


FIGURE 5.5: Comparison between the SentiStrength's original and integrated financial lexicon approach increasing accuracies with lagged sentiment effect

The results revealed distinct patterns in their predictive performance over time which was further explored for trend Analysis of accuracy across each time intervals.

After analyzing the time series comparison, the results showed that the common best prediction accuracy performance was during the time windows 10:31 to 11:30, 11:31 to 12:30, and After Hours, and the worst prediction accuracy performance was during

TABLE 5.10: Original Model Accuracy Across Different Time Ranges

Sl. No	**Time Range**	**Accuracy**
Best Performing Hours		
1	9:30 – 10:30	0.611111
2	10:31 – 11:30	0.545455
3	11:31 – 12:30	0.534524
4	After Hours	0.833333
Worst Performing Hours		
5	1:31 – 2:30	0.529412
6	2:31 – 3:30	0.500000
7	12:31 – 1:30	0.473684

TABLE 5.11: Financial Model Accuracy Across Different Time Ranges

Sl. No	**Time Range**	**Accuracy**
Best Performing Hours		
1	10:30 – 11:30	0.571429
2	11:31 – 12:30	0.666667
3	1:31 – 2:30	0.600000
4	After Hours	0.758621
Worst Performing Hours		
5	9:30 – 10:30	0.500000
6	2:31 – 3:30	0.470588
7	12:31 – 1:30	0.090909

12:31 to 1:30 and 2:31 to 3:30 for both the cases. So, we can conclude that the best performance is mostly during the After Hours which may lead to market reaction time in the morning, which is highly active trading hours, and the trading activity goes down after 2:31 pm. The peak performance of the model can be seen during specific intervals like 9:30 to 10:30, 11:31 to 12:30, and 1:31 to 2:30 which could be due to higher market activity and more predictable patterns during these times. Low accuracy during 12:30 to 1:30, 10:31 to 11:30 and 2:31 to 3:30 which could be due to lagged sentiment effects, or these might be the low trading activity or noise and randomness in the data that makes it harder for model to predict accurately. This shows that there is no specific trend in all trading hours due to high variability and its performance may heavily depend upon the specific time of the day.

Conclusion: The sharp rise in predictive accuracy for the financial lexicon approach at extended time lags indicates that market participants process financial sentiment over time rather than reacting instantaneously. This aligns with the concept of delayed price discovery, where investors gradually integrate new information into trading decisions. This may due to Financial news and social media discussions may influence trading behavior over time rather than triggering immediate price changes, due to Financial news and social media discussions may influence trading behavior over time rather than triggering immediate price changes and retail and institutional traders often analyze sentiment trends before making investment decisions, contributing to a lag in stock price responses.

5.4 Model Performance Evaluation

5.4.1 SentiStrength-based predictions vs ML-based predictions.

Different models' accuracy comparison using financial lexicon approach, showing the SentiStrength accuracy at 0 hour without considering the lag effect, and the other machine learning models, Logistic Regression, and XGBoost accuracies before implementing sentiment features.

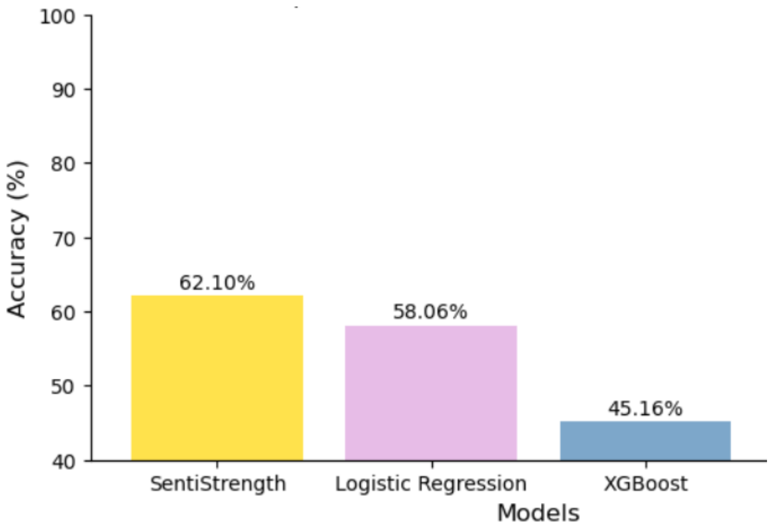


FIGURE 5.6: Comparison between SentiStrength's accuracy and different machine learning models at 0 hours

5.4.2 SentiStrength vs Hybrid Approach Comparison

This comparison shows the comparison using the financial lexicon approach, between the SentiStrength model shows the increase in accuracy after calculating it for lag sentiment shift effect, and machine learning models Logistic Regression, XGBoost, along with one more advanced model, LSTM was worked on and the accuracy increases from the baseline machine learning models after integrating the models with enhanced features.

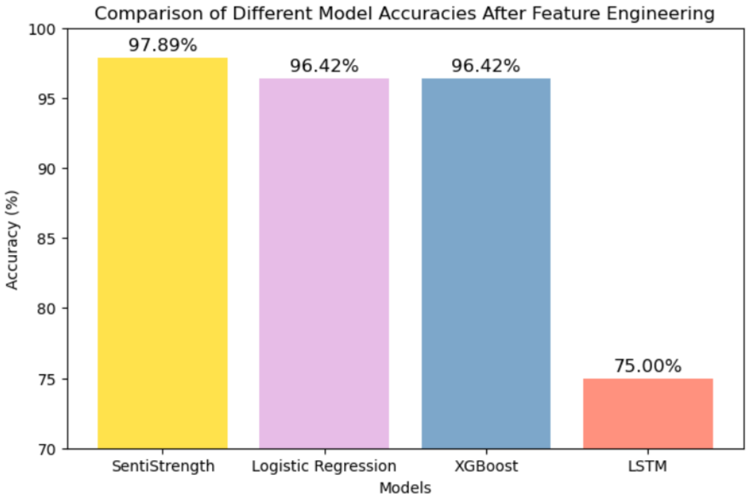
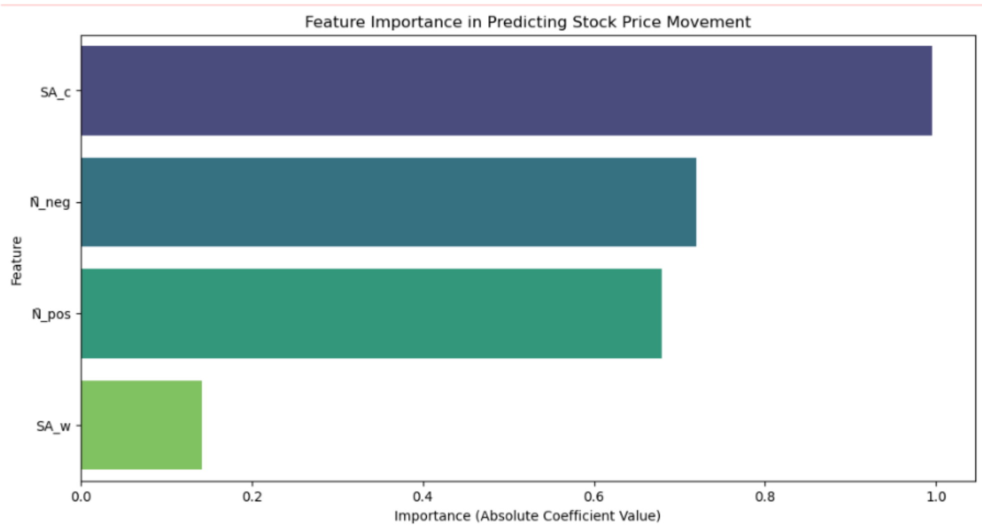


FIGURE 5.7: Accuracy comparison between the baseline SentiStrength Model and the Hybrid Approach for stcok market prediction

5.4.3 Correlation and Feature Analysis Importance

Random Forest was used to identify the most influential features, and the findings revealed that sentiment scores (SA_c) had a moderate influence as a stock market predictor, as it works best with the stock price changes, where extreme sentiment has more predictive power than average sentiment.



While sentiment contributes to stock market movements, external factors must also be considered, such as Stock Market Features, such as price volatility: (High - Low) / Open and features from the correlation analysis of user engagement with tweets as shown in figure 4.4 the additional features such as like count, retweet count can be further used for analysis.

5.4.4 Evaluation of Research with Objectives and Research Questions

- **Research Goal 1**: Financial-specific lexicons were developed and integrated into SentiStrength, which improved sentiment analysis and enhanced stock market prediction accuracy for stock market predictions. This enhancement included domain-specific terms relevant to finance, ensuring better sentiment classification. While the lexicon improved sentiment accuracy, fine-tuning weight assignments for financial terms required multiple iterations. Further optimization could be explored to refine accuracy further and reduce overfitting.
- **Research Goal 2:** Noise Reduction and Data Preprocessing. Various preprocessing techniques were implemented, including stopword removal, and financial keyword filtering. These steps effectively reduced noise in the Twitter dataset, improving sentiment classification reliability. While filtering out irrelevant tweets and spam helped, distinguishing between sarcastic and genuine tweets remained difficult. More advanced deep learning techniques could further refine data quality.
- **Research Goal 3:** Integrating Sentiment Analysis into Stock Prediction Models. The impact of sentiment on stock price movements was analyzed using a predictive model. Sentiment scores from SentiStrength, both generic and financial-enhanced versions, were compared against actual stock movements, revealing key insights. Logistic Regression and XGBoost were tested as potential prediction models. Incorporating real-time market sentiment dynamics remains an open area for improvement. Future work could involve deep learning models like FinBert for better temporal predictions.

5.4.5 Additional Research Question

Though the model predictions were accurate, it was seen that although there were so many negative sentiments yet the stock market prices went up. Due to this further investigation of was conducted by a survey of news and articles. This section further explores some additional questions and answers found from the survey.

Why did stock market prices go up even though there were negative sentiments in June 2020?

This could be due to several factors, as it was seen that in the month of June 2020, there was a growing recovery to the economy as the business were begining to reopen follwing the COVID-19 lockdown. The federal band started purchasing the corpus funds which boosted the confidence of investors to invest in the stocks and government also contributed to market support. Some companies announced earnings rates increased, which helped improve people's sentiments and the overall market better than expected. Also, due to the COVID-19 lockdown, tech stocks performed exceptionally well, as people increased their use of the online applications for working remotely, such as Zoom and Teams, during the month, their stock prices surged. Apple Continued to perform well with its strong balance sheet and innovative ecosystem, and Microsoft saw growth driven by its Azure cloud platform. Neufeld [2021]

5.5 Error Analysis

5.5.1 Types of Errors:

The errors observed in the model can be categorized as False Positives(FP) errors where instances in the model incorrectly predicted a stock price increase when it actually decreased. False Negatives (FN) errors, instances where the model incorrectly predicted a stock price decrease when it actually increased. These errors directly affect the specificity and sensitivity of the model, influencing its reliability in real-world stock trading scenarios.

5.5.2 Causes of Errors

The external factors (news events, political announcements, economic data releases) significantly impact stock prices but are not always reflected in tweets because of the influence of market noise. Some financial terms evolve rapidly (e.g., "short squeeze", "hot stocks" in retail trading) resulting in lexicon limitations in handling emerging trends: Regular updates to the financial lexicon with new market terminology is necessary. The analysis of False Positive and False Negative rates revealed were computed incorrectly, which led to the manual calculation of the accuracy, which impacted on prediction accuracy errors. The original Lexicon had higher sensitivity, where it correctly predicted stock increases, but low specificity suggested that it struggled with stock decreases. Whereas, Financial Lexicon significantly improved specificity, reducing False Positives, leading to better-balanced predictions.

5.6 Limitations

5.6.1 Limitations in Data Collection

- Data Collection of Twitter Data: The dataset used for sentiment analysis consisted of financial tweets from verified sources such as the Financial Times and stock-related discussions on Twitter. While these sources are relevant, the dataset may not be fully representative of the entire stock market sentiment, as retail investor opinions and informal financial discussions were not comprehensively included. Although the tweets were collected based on financial keywords, there is potential selection bias since tweets from influential accounts (e.g., financial analysts, economists) may dominate the dataset, while retail investors' sentiments could be underrepresented.

- Ground Truth Bias: The stock price movement up/down was used as the ground truth, ensuring an objective measure. However, the classification of stock movement into binary categories 0 for down, 1 for up may oversimplify market behavior, as it does not consider neutral or stable movements.

5.6.2 Limitations in Analysis and Conclusions

- Biased in Sentiment Scoring: The financial lexicon enhancement aimed to improve sentiment classification, but assigning sentiment weights to financial terms was a subjective process. While guidelines were followed, different annotators might have assigned slightly different sentiment scores.

- Assumptions in Predictive Models: The study assumes that sentiment has a direct and measurable impact on stock price movement. While correlation was analyzed, other macroeconomic factors (e.g., interest rates, geopolitical events) were not incorporated, which could affect stock movements independently of sentiment.

Chapter 6

Discussion

6.1 Key Findings

The study revealed that while SentiStrength struggled to achieve high predictive accuracy, integrating it with financial lexicon significantly improved performance. Furthermore, incorporating a time-lagged analysis demonstrated that sentiment's predictive power evolves, with certain delays leading to marginal improvements. Filtering out irrelevant tweets and focusing on financial events within the U.S. also showed the improvement from the original lexicon approach but not as well as the financial lexicon approach. These findings suggest that refining sentiment-based approaches and integrating them with more advanced models could further improve stock market prediction. In conclusion, while lexicon-based sentiment analysis remains valuable for qualitative insights.

6.2 Interpretation of Results

The results of the correlation analysis show that sentiment cannot immediately predict the stock movement accurately and is very ineffective at 0 hours. However, the lag effect proved that the market takes time to react to the sentiments. By implementing the lagged sentiment features from 1 to 4 hours significantly improved the predictive accuracy, especially when using a domain-specific financial lexicon. Financial sentiment has a delayed yet stronger impact on stock prices, suggesting that predictive models should incorporate time-dependent sentiment trends rather than relying solely on instantaneous sentiment scores.

6.3 Implications

This study can be used as a practical application for traders, analysts, and investors. The results underscore the importance of incorporating time-lagged features in predictive models to capture the full impact of sentiment on stock prices. While short-term sentiment fluctuations may not always translate into immediate price changes, analyzing sentiment shifts over extended periods provides a more robust framework for stock price prediction.

6.4 Limitations

Although sentiment analysis is useful for predicting stock market trends, as shown in this study, there are certain limitations, such as a limited financial vocabulary. While the expanded financial lexicon improved accuracy, manual fine-tuning can introduce bias, and the model would likely lag in response to evolving financial language, world events, and market sentiment. Moreover, the dataset was limited to Financial Times tweets and S&P 500 stock data, narrowing the approach to one news source and one market index. Despite the improvement of lagged sentiment analysis, there are other factors than sentiment that drive stock movements, according to the data. And in addition to that, this study shows that if stock price swings are mostly driven by sentiment, they cannot consider the influences of others, which is that the law of interest rates, geopolitical events, and macroeconomic factors, which play a critical role. Future work could integrate multi-source sentiment data, explore alternative ground truth definitions, and also focus on user engagement with the tweets too.

6.5 Future work and suggestions

Future research can improve sentiment-based stock prediction by using deep learning models like as FinBERT and Transformer-based architectures to refine sentiment classification. Expanding datasets to incorporate more financial indices, longer economic periods, and other indicators like trading volume, market volatility, and macroeconomic factors will improve model reliability. Integrating sentiment analysis with fundamental and technical indicators can result in a more complete forecasting framework. Exploring alternative financial lexicons and real-time sentiment tracking to enhance predictive accuracy in financial markets. Advanced time-series models, such as LSTMs with attention mechanisms, can help capture delayed sentiment effects. Furthermore, building a real-time sentiment tracking system that integrates live financial news and social media data will give investors actionable information, hence improving decision-making and market forecasting.

Chapter 7

Conclusion

This study demonstrates that sentiment analysis conducted on social media platforms, particularly sentiments extracted from Twitter, aids investors in decision-making based on the sentiments expressed on this platform. Consequently, this study significantly addresses a critical gap in the sentiment analysis systems utilized by SentiStrength for financial microblogs, thus enhancing the accuracy of its results when detecting sentiments related to finance alongside sentiments expressed across various sectors, including healthcare, technology, geopolitical events, and influential individuals rather than general users. This assists investors in making informed decisions. Furthermore, with the application of machine learning models such as Logistic Regression, XGBoost, and LSTM, the efficiency of stock market prediction improves when implemented with feature engineering, resulting in a more robust hybrid approach. However, comparisons indicate that this hybrid approach did not perform as well as the SentiStrength model. Ultimately, time-series analysis helps identify the best and worst performing hours, which play an important role in informed investment decision-making by enabling investors to incorporate sentiment-backed data into their decisions regarding a company's stock: whether to buy, sell, or hold. The increasing influence of social media in financial markets underscores the importance of valuing financial sentiment analysis research and employing it to forecast changes in stock prices, thereby affecting the future of financial analysis and trading practices. This thesis explored the impact of financial sentiment analysis on stock market prediction, emphasizing the enhancement of sentiment accuracy through the integration of lexicons from various sectors that influence stocks to refine financial lexicons and incorporate improved sentiment features into machine learning models.

Appendix A

Appendix 1

A.1 Financial Lexicons

The below shown lexicons were used to integrate with the default lexicons of SentiStrength, focusing on domain specificity to improve the performance of the model. These words are mostly chosen from various sectors based on the major factors that impact the stock price movements that including technology, healthcare, finance, geopolitical events, and event-specific factors. The chosen words were checked with **?** and added accordingly.

A.1.1 Emotion Words List

Below is the set of words used in the EmotionLookUpTable.txt file. Some of the extracted words were already present. The words that were already present in the file were added, and the rest of them were added. These emotions play a major role in the sentiment scoring of the sentence.

- bankruptcy -3
- depression -4
- unemployment -4
- racism -3
- brutality -4
- pandemic -5

- mass -2
- riots -4
- slump -2
- slumped -3
- inequality -3
- threatened -4
- sanctions -1
- failing -3
- defund -4
- boycott -4
- backslash -3
- fines -3
- volatility -1
- sinking -3
- saddled -3
- downturn -3
- plummet -3
- struggling -3
- hikes -2
- eviction -3
- debt -3
- controversial -3
- cuts -3
- warning -2
- skepticism -2
- nationalism -2

- tensions -2
- pricing -2
- risk -2
- blame -2
- sensitive -2
- caution -2
- fall -2
- exposing -2
- colonialism -1
- dismissed -2
- love 3
- market 0
- correction -1
- quarantine -1
- budget 0
- cautious -1
- regulation -1
- procrastination -1
- lowered -1
- discount -1
- downgrade -2
- outflow -1
- resistant -1
- stock 0
- trading 0
- economy 0

- banking 0
- hedge 0
- fund 1
- funds 0
- investment 0
- loan -2
- asset 0
- assets 0
- wealthy 1
- capital 0
- federal 0
- interest 0
- bonds 0
- borrowing 0
- acquisition 0
- mortgage -2
- partnership 0
- licensing 0
- earnings 1
- negotiation 1
- payout 1
- insurance 0
- support 1
- stability 1
- revenue 1
- wage 0

- growth 3
- dividends 3
- digital 1
- surplus 1
- expansion 2
- increase 2
- upturn 2
- gains 2
- positive 2
- trust 2
- higher 2
- highest 2
- booming 3
- bullish 3
- takeover 3
- performance 1
- merger 2
- surge 2
- job 2
- boom 4
- highs 2
- equity 2
- luxury 3
- surpassing 3
- profits 4
- buyback 4

- gains 4
- billion 2
- dollar 1
- deal 1
- ipo 3
- government 0
- financial 0
- hit -2
- stocks 0
- credit 2
- investors 0
- buy 4
- bailout -2
- target 1
- pay 2
- offered 2
- highest 3
- solidarity 3
- boosted 3
- benefitting 3
- freedom 3
- beautiful 3
- strongest 2
- wage 1
- wages 1
- lucrative 1

- impact 1
- piles 1
- upsurge 2
- free 3
- multiplying 2
- charitable 2
- rescind -1
- shameful -1
- shrinking -1
- layoffs -2
- gambling -3
- fell -2

A.1.2 Booster Words list

Booster words are such words that strengthen the meaning of the statement made. So these increases the scoring of the overall sentence by adding more weights to the related sentiment.

- absolutely 2
- astoundingly 2
- breathtakingly 3
- catastrophic -3
- consistently 3
- critical -2
- crucial -1
- dazzlingly 1
- deeply 2

- definitely 1
- enormous 1
- exceptionally 3
- excessive 1
- extraordinarily 3
- extraordinary 3
- extreme 2
- extremely 2
- hard -1
- highest 2
- highly 1
- huge 1
- huge 2
- just -1
- major 1
- massive 2
- most 1
- outrageous 2
- outstandingly 2
- overwhelmingly 2
- phenomenally 2
- strong 2
- substantial 1
- worst -3

A.1.3 Idiom Words lists

The Idiom word list is used to give a meaning to the context with a different reference, which adds value to the sentiment.

- bail out 1
- bear market -3
- conservative economic climate -2
- curfews force -3
- economic downturn -2
- emerging market 4
- economic reaction -3
- hit big 1
- hot stocks 3
- hottest stocks 4
- increase cash 2
- inflation surge -2
- job loss -3
- job market 1
- killed it 4
- mass production 3
- pay cut -2
- paying billion 3
- stock crash -3
- stock exchange 1
- stock rally 3
- stockpiles 2
- toy blast -2

A.1.4 Negating Words List

The collection of negating words is used in the context to express the denial. These words make the sentence even stronger conveying a sense of clear negative sentiment.

- absent
- against
- avoided
- barely
- cant
- couldn't
- couldnt
- decline
- declined
- denied
- denies
- deny
- didn't
- didnt
- disallowed
- disapprove
- doesn't
- doesnt
- don't
- dont
- eliminate
- fail
- failing

- fails
- hardly
- hasn't
- hasnt
- insignificant
- isn't
- isnt
- lack
- lacking
- little
- missing
- neither
- never
- no
- nor
- not
- nothing
- nowhere
- refuses
- reject
- rejection
- scarcely
- shouldn't
- shouldnt
- unacceptable
- unprofitable

- unsuccessful
- withdraw
- withdrew
- withheld
- without
- won't
- wont
- wouldn't
- wouldnt

Bibliography

(2017). Sentistrength - sentiment strength detection in short texts - sentiment analysis, opinion mining. Accessed: 2025-04-09.

Al-Ayyoub, M., Gigieh, A., Al-Qwaqenah, A., Al-Kabi, M., and Alsmadi, I. (2017). Aspect-based sentiment analysis of arabic laptop reviews. *The International Arab Journal of Information Technology*. Accessed: 2025-04-09.

Baltezarevic, R. and Baltezarevic, I. (2024). The advantages of sentiment analysis driven by artificial intelligence (ai).

Baruah, A. and Changkakati, B. (2024). Vader sentiment analysis on twitter: Predicting price trends and daily returns in india's stock market. *Bangladesh Journal of Multidisciplinary Scientific Research*, v9i2.2226:45–54.

Bell, C., Olukemi, A., and Broklyn, P. (2024). Social media sentiment analysis for brand reputation management. *Preprints.org*, 1.

Bjork, S. and Kiran, U. (2019). Predicting sp 100 index movement by sentiment analysis.

Buckley, K. (2013). Topic-based sentiment analysis for the social web: The role of mood and issue-related words. *Journal of the American Society for Information Science and Technology*.

Dilik, B. (2024). The relationship of social media messages with stock returns and volatility. *ResearchGate*. https://doi.org/10.13140/RG.2.2.16452.44163.

Divya, K. and Menaka, M. P. (2025). Twitter sentiment analysis. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 11(2):1305–1310.

Du, K., Xing, F., Mao, R., and Cambria, E. (2023). Finsenticnet: A concept-level lexicon for financial sentiment analysis.

Figà-Talamanca, G. and Patacca, M. (2022). An explorative analysis of sentiment impact on s&p 500 components returns, volatility, and downside risk. *Annals of Operations Research*. https://doi.org/10.1007/s10479-022-05129-w.

Goel, R., Ford, L., Obrizan, M., and Sharma, R. (2020). Covid-19 and the stock market: Evidence from twitter. arXiv (Cornell University). Accessed: 2025-04-09.

Gouthami, S. and Hegde, N. P. (2023). Automatic sentiment analysis scalability prediction for information extraction using sentistrength algorithm. *Lecture notes in networks and systems*, pages 21–30.

Gurusamy, B. M. (2024). The social media sentiment analysis framework: deep learning for sentiment analysis on social media.

Hao, Z., Yun-Heh, and Chen-Burger, J. (2022). An investigation into influences of tweet sentiments on stock market movements. Unpublished manuscript or date not specified.

Ibupoto, N. A., Rasheed, Z., and Ashraf, S. (2022). Sentiment analysis on current political topics in pakistan's twitter user bases. *ResearchGate*.

Islam, M. R. and Zibran, M. F. (2017). A comparison of dictionary building methods for sentiment analysis in software engineering text.

Jiang, X., Ren, B., Wu, Q., Wang, W., and Li, H. (2024). Dcasam: advancing aspect-based sentiment analysis through a deep context-aware sentiment analysis model. *Complex Intelligent Systems*, 10.

Kartbayev, A. (2024). Development of context-based sentiment classification for intelligent stock market prediction.

Khaira, U., Johanda, R., Utomo, P. E. P., and Suratno, T. (2020). Sentiment analysis of cyberbullying on twitter using sentistrength. *Indonesian Journal of Artificial Intelligence and Data Mining*, 3(1):21.

Khan, L. (2019). Sentiment analysis with sentistrength.

Khan, M. and Srivastava, A. (2024). A review on sentiment analysis of twitter data using machine learning techniques. *International Journal of Engineering and Management Research*, 14:186–195.

Lamba, M. and Madhusudhan, M. (2021). Sentiment analysis. *Text Mining for Information Professionals*, pages 191–211.

Lau, T. Y. (2024). Application of sentiment analysis to explore the connection between public sentiment and stock price movement by python.

Maulidiana, P. D. R., Vitianingsih, A. V., Kacung, S., Maukar, A. L., and Hermansyah, D. (2024). Comparative analysis of svm and nb algorithms in evaluating public sentiment on supreme court rulings. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 13(2):230–236.

Mokhtari, M. (2023). The impact of twitter sentiments on stock market trends.

Neufeld, D. (2021). Visualizing s&p performance in 2020, by sector. Accessed: 2025-04-09.

Nti, I. K., Adekoya, A. F., and Weyori, B. A. (2020). Predicting stock market price movement using sentiment analysis: Evidence from ghana. *Applied Computer Systems*, 25:33–42.

Olusegun, J. (2024). The role of ai in enhancing sentiment analysis for brand management.

Pagolu, S. (2016). Sentiment analysis of twitter data forpredicting stock market movements.

Pathan, A. F. and Prakash, C. (2021). Attention-based position-aware framework for aspect-based opinion mining using bidirectional long short-term memory. *Journal of King Saud University - Computer and Information Sciences*.

Putri, N. K., Vitianingsih, A. V., Kacung, S., Maukar, A. L., and Yasin, V. (2024). Sentiment analysis of brand ambassador influence on product buyer interest using knn and svm. *Indonesian Journal of Artificial Intelligence and Data Mining*, 7(2):327.

Qian, H. (2022). Review of comment sentiment analysis based on emoticons. *Advances in Social Science, Education and Humanities Research*.

Raees, M. and Fazilat, S. (2024). Lexicon-based sentiment analysis on text polarities with evaluation of classification models.

Rakshit, P. (2024). Hybrid deep learning approach for sentiment analysis on twitter data.

Samuel, J., Md. Nawaz Ali, G. G., Mokhlesur Rahman, M., Esawi, E., and Samuel, Y. (2020). Covid-19 public sentiment insights and machine learning for tweets classification. *Information*, 11(6):314.

Sandaka, G. K. (2024). Sentiment analysis and time-series analysis for the covid-19 vaccine tweets.

Simanjuntak, T. N. and Pramana, S. (2021). Sentiment analysis on overseas tweets on the impact of covid-19 in indonesia. *Indonesian Journal of Statistics and Its Applications*, 5(2):304–313.

Sina and Setiawan, E. B. (2023). Stock price correlation analysis with twitter sentiment analysis using the cnn-lstm method. *Sinkron*, 8(4):2190–2202. Accessed: 2025-04-09.

Siregar, B., Misyuari, M. N., Nababan, E. B., and Fahmi (2021). Person's multiple intelligence classification based on tweet post using sentistrength and processed on the apache spark framework. In *Journal of Physics: Conference Series*, volume 1882, pages 012125–25.

Thelwall, M. (2016). The heart and soul of the web? sentiment strength detection in the social web with sentistrength. *Understanding Complex Systems*, pages 119–134.

Thelwall, M., Buckley, K., and Paltoglou, G. (2011). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63:163–173.

Yekrangi, M. and Abdolvand, N. (2020). Financial markets sentiment analysis: Developing a specialized lexicon. *Journal of Intelligent Information Systems*, 57(1):127–146.

Yili, W. (2024). Sentiment analysis of twitter data.

Yin, S. (2024). The current state and challenges of aspect-based sentiment analysis. *Applied and Computational Engineering*, 114(1):25–31.

Youvan, D. C. (2024). Understanding sentiment analysis with vader: A comprehensive overview and application.

Yu, H. (2024). Analysis of the impact of investor sentiment on stock price crash risk.

Zhang, H. (2024). Time-series analysis and forecasting of s&p500 index based on arima and ets model. https://www.researchgate.net/publication/383485021_Time-Series_Analysis_and_Forecasting_of_SP500_Index_Based_on_ARIMA_and_ETS_Model. Accessed: 2025-04-09.

Zhang, W., Li, X., Deng, Y., Bing, L., and Lam, W. (2022). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *arXiv.org*.