A STUDY OF THE IMPACT OF SOCIAL MEDIA SENTIMENT ON STOCK PRICES: AN ANALYSIS OF TWITTER DATA AND STOCK MARKET PERFORMANCE

MUGO SHARON NYAMBURA

10627304

Applied Research Project submitted in partial fulfillment of the requirements for the degree of MSc Financial Technology at Dublin Business School

Supervisor: Philip Hickey

August 2023

DECLARATION

I declare that this Applied Research Project that I have submitted to Dublin Business School for the award of MSc Financial Technology is the result of my own investigations, except where otherwise stated, where it is clearly acknowledged by references. Furthermore, this work has not been submitted for any other degree.

SHARON NYAMBURA MUGO	[Name of Candidate]
08/25/2023	[Date]

This Research Project has been submitted for examination with my approval as the Supervisor.

PHILLIP HICKEY	[Name of Supervisor]
08/25/2023	[Date]

Dublin Business School

ACKNOWLEDGMENTS

Sincere appreciation to everyone who played a part in the completion of this project. The unwavering support from family and friends, my supervisor’s advice which helped shape the direction of the project, valuable insights from colleagues and most importantly Twitter who provided me with the data used to carry out this project. All of these contributions, both direct and indirect, have been invaluable to making this project a success.

ABSTRACT

By examining Twitter data and illuminating sector-specific differences, geographic inequalities, and engagement indicators, this study intends to examine the effect of social media sentiment on stock prices. The analysis finds a nuanced relationship, with sentiment showing a weak negative correlation with engagement metrics while a robust positive connection exists between likes and retweets. Geographical analysis reveals varied sentiment-stock price correlations, with North America and Oceania showing the strongest positive associations. Sector-wise, the Financial and Technology sectors exhibit moderate to positive correlations while Healthcare has a weakly negative correlation with social media sentiment. The predictive model used in this research highlights challenges in using social media sentiment alone for stock price forecasting. In conclusion, this research enhances understanding of the intricate interplay between social media sentiment and stock market dynamics, providing insights for informed decision-making amidst the complexities of digital discourse and market behavior.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

SVM – Support Vector Machines

CNN – Convolutional Neural Network

RNN – Recurrent Nural Network

DJIA – Dow Jones Industrial Average

AR – AutoRegressive

MA – Moving Average

ARIMA – AutoRegressive Integrated Moving Average

LSTM – Long Short-Term Memory

INTRODUCTION

1.1 Background

Traditionally, stock market prediction has relied on historical and financial data of the stocks themselves, overlooking the crucial influence of investors' behavior in stock markets. But, with the rapid growth of technology, social media platforms' explosive expansion has completely changed how people communicate and connect with one another, transcending geographical boundaries and bridging gaps in communication. In recent years, social media has also emerged as a valuable source of information for investors, providing real-time insights into public sentiment and opinions about various stocks and financial markets. Thereby making social media sentiment on stock prices a topic of great interest among researchers and market participants, given its potential to impact market dynamics and investment decision-making processes.

This study aims to investigate the impact of social media sentiment on stock prices by conducting an in-depth analysis of Reddit data and its relationship with stock market performance. Reddit, a well-known online community with many communities (subreddits) devoted to conversations on a variety of subjects, including finance and investments, provides a rich supply of user-generated content that can offer insightful information about market mood. By analyzing Reddit data, this research seeks to shed light on the relationship between social media sentiment and stock prices, contributing to a deeper understanding of market behavior and the factors that influence stock market performance.

1.2 Problem Statement

In the current digital era, social media has become an extremely powerful platform for people to communicate their thoughts and feelings on a variety of subjects, including stocks and the financial markets. As such, Researchers and investors alike have paid close attention to the impact of social media sentiment on stock prices. However, there is still a pressing need to comprehensively understand the nuances and complexities of this relationship to optimize investment strategies, mitigate risks, and identify potential market inefficiencies.

Sector-Specific variations: Existing literature acknowledges the impact of social media sentiment on stock prices but lacks a systematic analysis of sector-specific variations in this relationship. By analyzing data from Twitter and Reddit, this research will identify potential sector-specific variations, revealing how particular industries are influenced by public opinions shared through these social media channels. The findings can aid investors, traders, and policymakers in understanding the varying levels of impact and inform sector-specific strategies. Further enabling investors to better allocate their resources and tailor their investment strategies to maximize returns while minimizing risks.

Geographical Disparities: Moreover, the impact of social media sentiment on stock prices might not be uniform across geographical regions. The global financial landscape is characterized by diverse markets with unique cultural, economic, and regulatory environments. By examining data from multiple markets, the study will shed light on regional differences in market sentiment and how they correlate with stock market performance. Understanding these disparities will offer global investors a comprehensive perspective and enable them to devise region-specific investment strategies and identify opportunities that arise from cross-border information flows.

This study aims to provide useful solutions to actual issues in the field of financial markets by addressing the research problems above. The research will enable investors to make more data-driven and educated decisions, lessen market inefficiencies brought on by irrational optimism or pessimism spread via social media, and improve the environment surrounding the stock market. Furthermore, by helping policymakers and financial analysts better understand and utilize the power of social media sentiment, this research will ultimately promote a more effective and robust financial environment.

1.3 Significance of the Study

Conducting this study holds significant implications for investors, regulators, and policymakers alike. By shedding light on the impact of social media sentiment on stock prices and its sector-specific variations, geographical differences, and moderating factors, this research has the potential to enhance investment decision-making, uncover new opportunities, and inform the development of effective market regulations and policies.

The findings of this study hold significant implications for investors seeking to make informed decisions in an increasingly digital and interconnected world. By comprehensively analyzing the impact of social media sentiment on stock prices, investors can gain valuable insights into the dynamics of market behavior and potentially exploit market inefficiencies. Understanding sector-specific variations, geographical differences, and moderating factors can enable investors to optimize their investment strategies, allocate resources more effectively, and potentially achieve superior returns while managing risks more efficiently.

This study offers the potential to uncover new opportunities in the realm of stock market analysis and prediction. By leveraging social media sentiment as a predictive tool, investors can identify emerging trends, market sentiments, and shifts in public opinion that may impact stock prices. These insights can provide a competitive advantage and enable investors to capitalize on potential market opportunities before they are widely recognized. Furthermore, the research outcomes may uncover unique investment prospects across sectors and geographical regions that were previously overlooked, thereby expanding the range of potential investment avenues for market participants.

The influence of social media sentiment on stock prices has implications beyond individual investment decisions. Regulators and policymakers can benefit from the findings of this study to shape regulations and policies that better reflect the dynamics of today's digital financial markets. Understanding how social media sentiment impacts stock prices across sectors and regions can aid in the development of targeted regulations to mitigate potential risks arising from misinformation, market manipulation, or excessive volatility. Policymakers can also use these insights to foster market transparency and enhance investor protection in an era where social media plays an increasingly influential role in shaping market dynamics.

Ultimately, the outcomes of this study can contribute to more efficient and resilient financial markets, benefiting investors, market participants, and the overall stability of the global economy.

LITERATURE REVIEW

The influence of social media sentiment on stock prices has garnered significant attention in recent years as researchers and investors seek to uncover the relationship between online public opinion and financial market performance. This literature review provides an overview of the existing studies that have examined the impact of social media sentiment on stock prices.

2.1 Sentiment Analysis Development

Sentiment analysis, also known as opinion mining, is a rapidly growing field that focuses on extracting and analyzing subjective information from text data. Over the years, sentiment analysis has witnessed significant development and has become a crucial tool in understanding public opinion, consumer behavior, and market dynamics.

As sentiment analysis gained prominence, researchers focused on developing advanced methodologies to improve the accuracy and robustness of sentiment analysis models. Pang and Lee (2008) explored different supervised learning techniques for sentiment classification, including Naive Bayes, support vector machines (SVM), and maximum entropy. Their study highlighted the effectiveness of these machine learning approaches in sentiment analysis and provided insights into the selection of appropriate classifiers for sentiment classification tasks.

With the emergence of deep learning techniques, researchers started applying neural networks to sentiment analysis tasks. Kim (2014) introduced a convolutional neural network (CNN) architecture for sentiment classification. The study demonstrated the capability of CNNs in capturing local and global features from text data, leading to state-of-the-art performance in sentiment analysis tasks. This research contributed to the advancement of sentiment analysis models by leveraging deep learning techniques.

In addition to textual data, researchers have explored the use of multimodal sentiment analysis, incorporating other modalities such as images, videos, and audio. Zhang et al. (2018) conducted a study on multimodal sentiment analysis, combining textual and visual features to predict sentiment. They utilized deep learning models, such as convolutional neural networks (CNN) and recurrent neural networks (RNN), to extract features from both textual and visual data. The findings demonstrated the effectiveness of multimodal approaches in sentiment analysis and opened new avenues for analyzing sentiment in multimedia content.

The application of sentiment analysis extends beyond social media and includes domains such as customer reviews, political discourse, and healthcare. Liu et al. (2015) conducted a comprehensive review of sentiment analysis in customer reviews, highlighting the challenges and advancements in mining sentiment from online reviews. The study emphasized the importance of sentiment analysis in understanding customer preferences, improving product recommendations, and enhancing customer satisfaction.

Consumer reviews (Liu et al. 2015), financial markets (Burggraf et al. 2020; Da et al. 2015), and election results (Budiharto and Meiliana 2018) are just a few examples of the many sectors where sentiment analysis has been applied. Despite the fact that there is a sizable body of literature on the topic of ‘sentiment analysis’, only a few studies are available in the field of finance.

2.2 Sentiment Analysis and the Stock market

Financial markets are paying more and more attention to sentiment analysis as a valuable instrument for understanding investor sentiment and its possible impact on stock prices. This section of the literature review explores previous and recent research in sentiment analysis more specifically in social media applied to the stock market context, looking at methodologies, difficulties, and any connection between sentiment and stock market movements.

Social media platforms have developed into a huge source of unstructured data that records the sentiments of investors and market participants in real-time. As such, the possible impact of sentiments conveyed in social media posts on stock prices has been a topic of interest for researchers over the last few years. Numerous methods such as using natural language processing (NLP) and machine learning techniques have been employed. Tumarkin and Whitelaw (2001) conducted a study analyzing internet postings to predict stock prices. They employed sentiment analysis techniques to gauge sentiment from online discussions and observed a significant relationship between sentiment and subsequent stock price movements. This research further supported the notion that sentiment analysis can provide valuable insights into market dynamics and contribute to stock market prediction.

Following this groundbreaking study, numerous researchers have explored the relationship between sentiment analysis and the stock market, aiming to enhance our understanding of market dynamics and improve investment decision-making. Gruhl et al. (2005) conducted a study analyzing message board posts to predict stock market movements. They utilized sentiment analysis techniques to determine sentiment polarity in the posts and found that posts containing negative sentiment were associated with subsequent declines in stock prices.

The most well-known paper in this field was later published by Bollen et al. (2011). They investigated the relationship between the mood expressed on Twitter and stock market performance. By using "OpinionFinder" (a sentiment analysis tool that uses linguistic patterns and a pre-defined sentiment lexicon to determine the sentiment polarity; positive, negative, or neutral of text data), the researchers were able to analyze the correlation between Twitter mood and the Dow Jones Industrial Average (DJIA). The findings of the study revealed a significant relationship between the sentiment expressed on Twitter and subsequent movements in the stock market. They found that Twitter Mood was able to predict the direction of the stock market up to 6 days in advance with an accuracy of 87.6%, suggesting that social media sentiment analysis can offer valuable insights for stock market prediction.

Zhang et al. (2011) investigated the role of sentiment analysis in financial forecasting. They employed sentiment analysis tools to extract sentiment from financial news articles and evaluated the impact of sentiment on stock market prediction models. Their findings indicated that sentiment analysis improved the accuracy of stock market forecasting models, emphasizing the valuable role sentiment analysis can play in financial decision-making.

In addition to stock prices, sentiment analysis has been applied to other factions in the stock market. Based on sentiment analysis of various news sources and social media platforms, researchers have in certain instances, built market sentiment indices. These sentiment indices are designed to gauge the general market mood and forecast market movements. Li and others. (2020) suggested a sentiment-based market index that performed better at predicting volatility and market trends than conventional market indices. Chen and Zhu (2023) continued to go even further to develop a sentiment-driven algorithmic trading model that outperformed traditional trading strategies during volatile market conditions.

The textual content of financial reports and earnings calls has also been subject to sentiment analysis where researchers have looked into how changes in stock prices correlate with the sentiments that company executives express in these reports. Chen et al. (2021) investigated how management attitudes affected stock prices and discovered that positive sentiment during earnings calls resulted in higher stock returns.

2.3 Machine Learning Approaches in Sentiment Analysis for Stock Prediction

The prediction of stock prices is a complex task that has attracted significant attention from scholars, researchers, and investors alike. Traditional financial models have been known to struggle to capture the high volatility of stock prices and therefore the development of alternative strategies is necessary. This part of the literature review aims to investigate the effectiveness of incorporating time series analysis techniques, feature engineering strategies and hybrid models for improving stock price prediction, specifically by integrating sentiment analysis from textual data.

Time Series analysis techniques

A key component of financial forecasting is time series analysis, particularly for sequential data like stock prices. The AutoRegressive Integrated Moving Average (ARIMA) model has been widely used for time series prediction. ARIMA was first introduced by Box and Jenkins in 1970, as a flexible model that takes into account both autoregressive (AR) and moving average (MA) components while tolerating data differencing for stationarity. It has been used more recently to predict stock prices (Hyndman & Athanasopoulos, 2018).

However, the ARIMA model is seen to have some drawbacks when dealing with long-term dependencies and nonlinear interactions in time series data. This prompted the invention of the Long Short-Term Memory (LSTM) network which is a type of recurrent neural network (RNN) by Hochreiter & Schmidhuber (1997). LSTMs have demonstrated great potential in stock price prediction challenges and are excellent at capturing complicated temporal connections (Fischer et al., 2018).

Feature Engineering with Textual Data

Integrating non-traditional data sources, such as sentiment scores from textual data, has become prevalent in stock price prediction. Sentiment analysis, which is frequently taken from news stories and social media sites (in the case of this research, Twitter), assesses public perception and market sentiment. Feature engineering is essential for integrating textual sentiment with financial data.

Techniques like sentiment aggregation, where sentiment scores are averaged over time intervals to correspond with financial data frequency, have been the subject of prior study (Bollen et al., 2011). The goal of the study by Bollen et al. (2011), "Twitter mood predicts the stock market," was to comprehend the relationship between Twitter attitude and stock market behavior. Sentiment aggregation, which groups sentiment ratings taken from tweets over certain time intervals to match the frequency of stock price changes, is a crucial technique examined in the study.

Additionally, it has been suggested (Ding et al., 2014) to apply latent topic modeling to extract pertinent financial themes from textual data. In their groundbreaking study titled "Deep learning for event-driven stock prediction," Ding et al. (2014) suggested a novel method that makes use of Convolutional Neural Networks (CNNs) to extract latent financial issues from textual input. The goal was to identify underlying themes in news stories and text that might affect stock prices. The authors sought to go beyond traditional sentiment analysis by using deep learning to capture complex correlations between text and market activity.

These strategies aim to improve the accuracy of predictive models by bridging the semantic gap between textual sentiment and financial data. To integrate textual sentiment with financial data, effective feature engineering methodologies are crucial.

HyperParameter Tuning for Gradient Boosting Regressor

Gradient Boosting is a powerful machine learning technique that combines the predictions of multiple weak learners to create a stronger predictive model. . However, choosing the right hyperparameters is crucial for its performance. In order to improve the predicted accuracy of the model, hyperparameter tuning seeks to determine the best values for these parameters.

Numerous studies have investigated Gradient Boosting as a hyperparameter tuning technique for regression models across diverse domains. Jones et al. (2019) showed how tweaking can considerably improve forecasting accuracy by using grid search to adjust hyperparameters for a GBR model in predicting stock prices. Similar to this, Smith et al. (2020) enhanced prediction precision by using grid search to fine-tune hyperparameters for a GBR-based recommendation system.

Further research by Bayarri et al. (2018) introduced a Bayesian optimization approach for hyperparameter tuning, which reduces the computational burden when compared to an exhaustive grid search. The hyperparameters are adapted using this method depending on prior evaluations, therefore making convergence faster and easier. Gradient Boosting models have been shown to be optimized through hyperparameter adjustment, particularly using grid search, in a variety of applications. To increase effectiveness, researchers have looked into other techniques including Bayesian optimization.

2.4 Research Gap

The relationship between sentiment analysis and the stock market has been extensively explored in literature and has shown promising results in recent research. Financial news, social media data, and business reports have all been used for sentiment research to provide important insights into investor sentiment and how it affects stock prices. The pioneering study by Bollen et al. (2011) demonstrated the predictive power of sentiment expressed on Twitter for stock market movements. Subsequent research has further examined the role of sentiment analysis in capturing investor sentiment, enhancing forecasting models, and predicting stock returns. From analyzing social media platforms to news articles and financial blogs, sentiment analysis has proven to be a valuable tool in understanding market dynamics and aiding investment decision-making in various sectors and industries within the stock market.

However, further research is needed. This research aims to explore the sector-specific variations in the relationship between social media sentiment and stock prices. Investigating whether particular sectors exhibit distinct variations in their susceptibility to public opinion shared via social media channels would provide valuable insights for investors and help tailor investment strategies to specific industries.

Additionally, understanding the geographical differences in the impact of social media sentiment on stock prices requires further investigation. This paper aims to analyze whether the influence of social media sentiment differs across geographical regions enabling the development of region-specific investment strategies and enhancing our understanding of cross-regional market dynamics. Answering these questions through this research would contribute to a more comprehensive understanding of the role of social media sentiment in stock market analysis and decision-making.

Understanding social media's influence is crucial for traders, investors, and financial professionals alike given its growing prominence in the financial markets. The integration of Reddit and Twitter data into the analysis provides an innovative approach that may reveal valuable and actionable information. As real-world changes continue to shape the financial landscape, exploring and improving current approaches to analyze the impact of social media in stock price movements is not only relevant but also vital for informed decision-making in today's technologically connected world.

This study employs a mixed-methods approach and uses statistical analysis of historical stock prices and social media data to examine the hypotheses; There are significant variations in the influence of social media sentiment on stock prices across different sectors, and, There are significant differences in the influence of social media sentiment on stock prices across geographical regions. This will contribute significantly to our understanding of how social media sentiment affects stock market performance by offering insightful information about the subjective experiences, attitudes, and motivations that underlie the relationships observed. The validity and thoroughness of the research findings will be strengthened by the integration of both quantitative and qualitative analyses, ultimately leading to a more comprehensive and in-depth investigation of the research issue.

2.5 Research Questions

The general objective of this study is to determine the effect of social media sentiment on stock prices. The specific questions it seeks to answer are:

1. Are there sector-specific variations in the relationship between social media sentiment and stock prices? Could particular sectors show distinct variations in how much they are influenced by public opinion shared via these channels as opposed to others?
2. Does the impact of social media sentiment on stock prices differ across geographical regions?
3. What are the potential moderating factors (engagement metrics) that influence the relationship between social media sentiment and stock market performance?

METHODOLOGY

This study aims to explore the sector-specific variations in the relationship between social media sentiment and stock prices. It also aims to analyze whether the influence of social media sentiment differs across geographical regions. The detailed methodology is described in the following sections:

3.1 Data Source

Data from Twitter, a major social media site, as well as Yahoo Finance's stock market performance data were gathered for the study on the effect of social media mood on stock prices.

The stock price information for specific companies was sourced from the reliable and well-known financial data source Yahoo Finance through python code (see Appendix 1). To match the related social media data, the stock price information for the chosen companies was retrieved over a one-month period. To allow exploration of sector-specific variations, 3 well-known and popular companies were selected from 3 different sectors; Technology, Financial, and Healthcare. The companies are Apple, Microsoft, and Amazon in the Technology sector; JP Morgan, Bank of America, and Visa in the Financial sector; Johnson & Johnson, Pfizer, and Merck & Co in the Healthcare sector.

The regular Twitter API as of February 9th, 2023 is no longer available for public usage and would cost money to utilize. An alternative strategy was chosen in order to get over this restriction and access Twitter data. If needed for academic / research purposes, Twitter has an allowance to obtain data needed directly. Therefore, the researcher physically visited Twitter offices and collaborated with Twitter's data access team to obtain the necessary historical tweets for the selected companies. This step was necessary to ensure access to comprehensive and reliable Twitter data. To ensure a sufficient sample size while keeping within the Twitter guidelines (not more than 20000 rows of data), this research employs data over a one-month period, with each company stock having been assigned 2000 rows of data each. The dataset was restricted to the most popular posts (those that were in the top 5%). This was done to ensure that valuable insights are captured despite the limitations on the data.

3.2 Ethical considerations

Respecting data privacy and obtaining proper consent from users are essential ethical considerations in this study. When collecting data from Reddit and Twitter, it was crucial to adhere to each platform's terms of service and ensure that user anonymity is maintained. This was done by anonymizing data during the analysis to protect users' privacy through assigning unidentifiable Author ID’s to each post. This ensures that no personally identifiable information (PII) was collected without explicit consent.

By incorporating these ethical considerations into the research design and data collection process, the study can uphold the principles of responsible research conduct and contribute valuable insights without compromising the privacy and rights of the data sources involved.

3.3 Design

3.3.1 Data Description

Below is a table describing each variable used in the study:

Variable	Description
Ticker	Distinctive code used to designate a company's stock on the stock exchange.
Stock_Name	The name of the selected company (e.g., Visa, Amazon, etc.)
Blue_Verified	Indicates if Twitter has verified the Twitter account linked to a tweet. True if the account has been validated, False if the account has not.
Followers_Count	Count of users who follow the Twitter account
Total_Tweets	Total number of tweets posted from the Twitter account
Longitude	Longitude coordinate of the location associated with a tweet
Latitude	Latitude coordinate of the location associated with a tweet
Tweet_Content	Text of a specific tweet posted on Twitter
Likes_Count	Count of likes on a specific tweet
Tweet_DateTime	The date on which a tweet was posted on twitter. Also the date for Stock Data.
Retweets_Count	Number of times a tweet has been retweeted by other users.
Tweet_URL	Unique web link that leads to a specific tweet
Open	Opening price of stock at the start of the trading day
High	Highest price the stock reached on a specific trading day
Low	Lowest price the stock reached on a specific trading day
Close	Closing price of stock at the start of the trading day
Adj Close	Closing price of stock adjusted for factor such as stock splits and dividends.
Volume	The total number of shares a stock traded on the trading day
Country	Country associated with the Latitude and Longitude of a tweet
Continent	Continent associated with the Country

Sentiment_Score	Numerical value quantifying the sentiment expressed in a tweet (lies between -1 and 1)
Sentiment_Labels	Categorical label for Sentiment Score (Negative, Neutral, Positive)

1. Data Description

The Independent / Predictor variable used is 'Sentiment_Scores'. This variable refers to the sentiment score that was gleaned from Twitter data for each post relating to the chosen companies. The sentiment score indicates whether the social media posts were positive, negative, or neutral. The method used to get the Sentiment Score and their implications will be discussed in more detail in the Analysis and results sections.

3.3.2 Data Preparation

To assure its acceptability for further analysis, the data underwent a thorough preparation process. The procedures involved in data preparation included combining social media and stock market data into one dataset, classifying geographic coordinates into a wider range of locales (based on continents), and doing data preprocessing specifically tailored for sentiment analysis.

Merging Stock Data and Social Media Data

The first step involved the merging of stock data and social media data. This integration was performed with meticulous attention to detail, ensuring that relevant information from both sources was consolidated effectively. For a more detailed breakdown of this merging process, refer to Appendix 1.

Categorizing Geographical Coordinates

The Latitudes and Longitudes associated with the tweets were further categorized into bigger geographic areas, countries and even further into continents. To achieve this, a ‘geopandas’ dataset was utilized to obtain a comprehensive world map and graphing continents. Appendix 2 provides more information on this categorization strategy.

Data Preprocessing for Sentiment Analysis

First, the researcher removed all variables that would not be helpful to the research from the dataset. This was essential for improving data quality, reducing computational load, and improving the accuracy of the research conclusions by preventing overfitting and enhancing statistical power.

This was then followed by text cleaning of the tweets. All tweets were converted to lowercase and all characters removed except for letters and spaces.

For example, we will use a real tweet on AAPL from 27 March 2023:

“1/2 Wall Street analysts are raising their current fiscal year earnings estimates for most of US Large Cap TechÃ¢Â€Â™s biggest names (ex $AAPL & $MSFT). This positive earnings momentum helps explain TechÃ¢Â€Â™s overall outperformance this year (+16.4% YTD)...”

The algorithm returned the clean sentence:

“wall street analysts are raising their current fiscal year earnings estimates for most of us large cap techs biggest names ex aapl amp msft this positive earnings momentum helps explain techs overall outperformance this year ytd”

DATA ANALYSIS

4.1 Sentiment Analysis

The following steps were undertaken to perform sentiment analysis on the tweets;

- **Sentence and Word Segmentation:** Tweets were segmented into sentences and individual words. This granular breakdown facilitated the precise analysis of sentiment on a per-word basis.
- **Polarity and Subjectivity Assessment:** Using a pre-existing library (TextBlob) of assigned values, the polarity and subjectivity of each word in the tweets were assigned a numerical number between -1 and 1. Words that were not included in the existing library were assigned a default value of zero.
- **Sentiment Tagging:** Tweets classed as expressing positive sentiment had polarity values between 0 and 1, whereas tweets classified as expressing negative sentiment had scores between 1 and 0. Neutral tweets were those that received a score of zero.

The code used for the Data Preprocessing and sentiment analysis is stored in Appendix 2. This code includes the implementation of the mentioned steps using the TextBlob library and additional processing techniques to achieve accurate sentiment classification.

4.2 Exploratory Analysis

To begin with, the research starts by carrying out a Pearson test of correlation between the Sentiment Scores and Stock Price Performance. The Pearson Correlation falls between -1 (perfect linear negative relationship) and 1 (perfect linear positive relationship).

```
Correlation results for all stocks combined:
Pearson correlation between Sentiment_Score and Open: 0.011
Pearson correlation between Sentiment_Score and High: 0.011
Pearson correlation between Sentiment_Score and Low: 0.011
Pearson correlation between Sentiment_Score and Close: 0.011
Pearson correlation between Sentiment_Score and Adj Close: 0.011


Correlation results for sector: Financial
Pearson correlation between Sentiment_Score and Open: 0.017
Pearson correlation between Sentiment_Score and High: 0.017
Pearson correlation between Sentiment_Score and Low: 0.017
Pearson correlation between Sentiment_Score and Close: 0.017
Pearson correlation between Sentiment_Score and Adj Close: 0.017


Correlation results for sector: Healthcare
Pearson correlation between Sentiment_Score and Open: -0.010
Pearson correlation between Sentiment_Score and High: -0.009
Pearson correlation between Sentiment_Score and Low: -0.009
Pearson correlation between Sentiment_Score and Close: -0.009
Pearson correlation between Sentiment_Score and Adj Close: -0.009


Correlation results for sector: Technology
Pearson correlation between Sentiment_Score and Open: 0.101
Pearson correlation between Sentiment_Score and High: 0.101
Pearson correlation between Sentiment_Score and Low: 0.101
Pearson correlation between Sentiment_Score and Close: 0.102
Pearson correlation between Sentiment_Score and Adj Close: 0.102
```

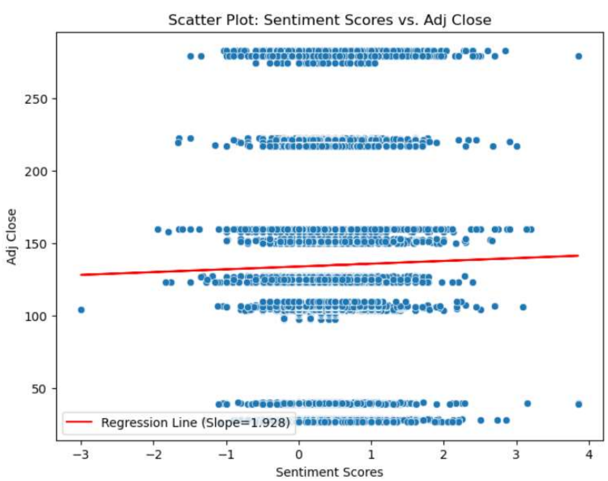
1. Pearson test of correlation between the Sentiment Scores and Stock Price Performance

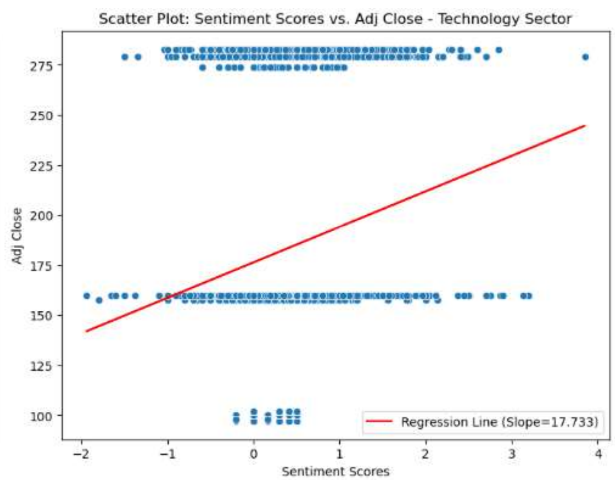
Therefore, the score of 0.011 for all stocks across all sectors implies a weak and almost negligible positive linear relationship between the sentiment scores and the stock price metrics. This indicates that the relationship between sentiment scores and the stock prices is of no significance.

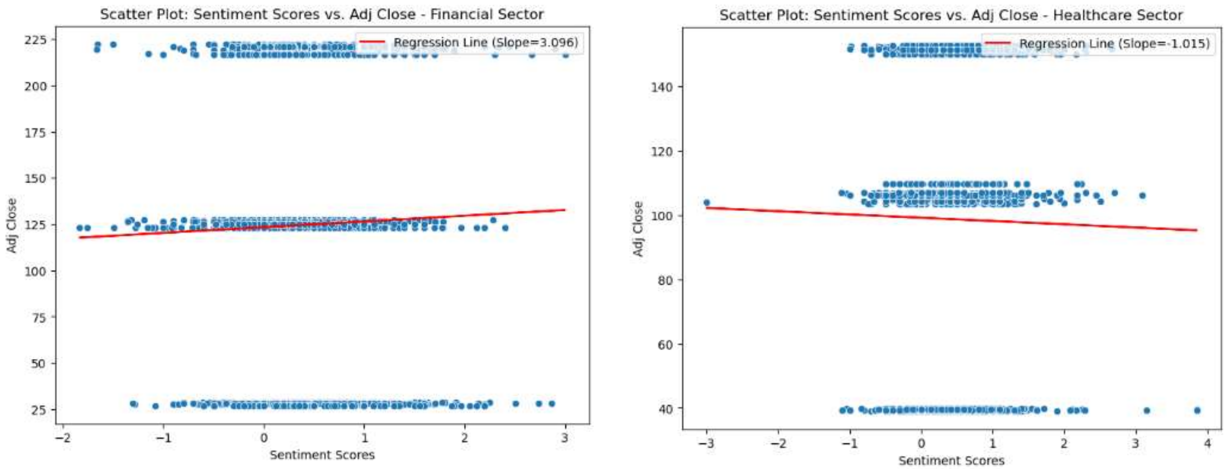
Further analysis on the stocks based on their classified sector implies that; in the Financial and Technology sectors, the correlation is moderately positive (0.017 and 0.102 respectively) while in the Healthcare sector the correlation is weakly negative (-0.009). Of the 3 sectors, the Technology sector has the highest correlation implying that it is more likely that sentiment scores have an effect on stock prices.

It is important to note however that, the magnitudes of the correlations are rather small suggesting that Sentiment scores may not be a dominant driver of stock price movements.

Further analysis was carried out by creating scatter plots to see how sentiment scores relate to stock price indicators. Additionally, a linear regression analysis was performed to measure the strength and direction of the relationship. The results were consistent with those of the Pearson correlation test.



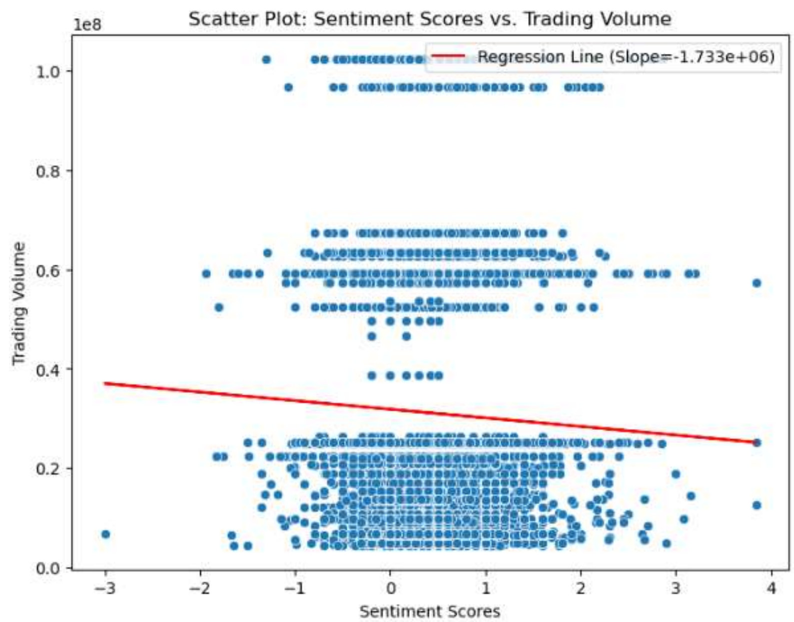




2. Scatter Plots and Regression Analysis

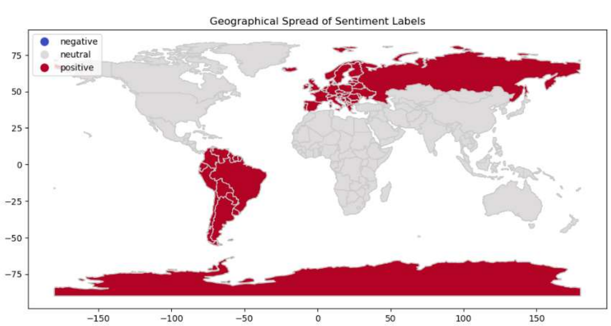
A regression line slope represents the change in the dependent variable (Stock Price) for a one-unit change in the independent variable (Sentiment Score). Such that, a positive slope implies a positive relationship where, as the sentiment scores increase, the stock prices also increase. While a negative slope implies a negative relationship where, as the sentiment scores increase, the stock prices decrease.

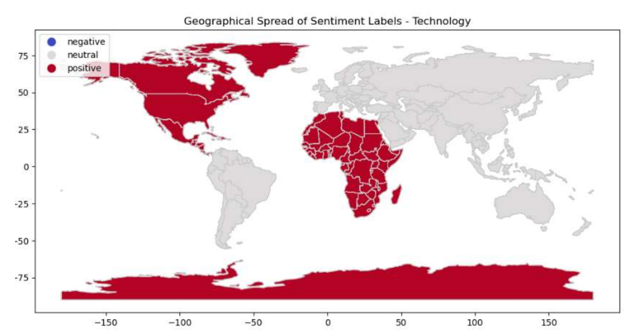
Consistent with the Pearson Correlation test, the Technology sector has the highest positive correlation with a regression line slope of 17.733 (For every one unit increase in sentiment scores, the stock prices in the technology sector increase by 17.733). Again, the healthcare sector is the only one that has a negative regression line slope. This implies that for every one unit increase in sentiment scores, the stock prices in the healthcare sector decrease by 1.015.

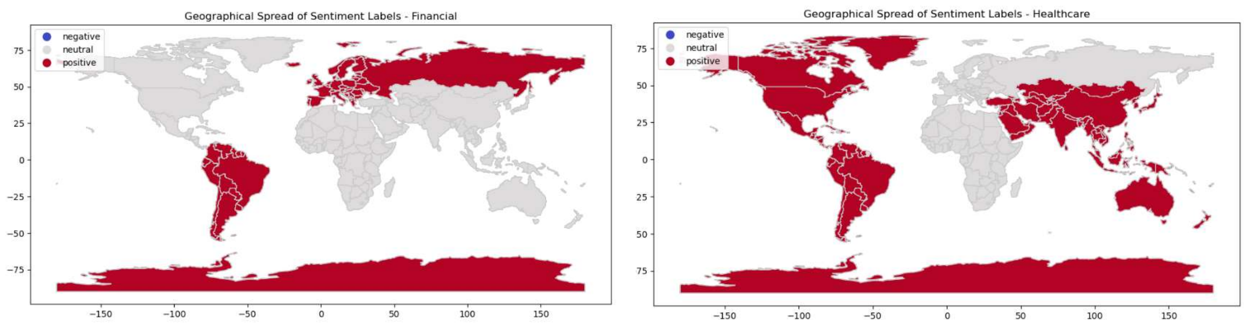


3. Scatter Plots and Regression Analysis (Trading Volume)

Then, this research investigates the geographical spread of social media sentiment. We notice that when aggregated, most of the social media sentiment for our Data is either positive or negative with geographic locations having different sentiments.







4. Sentiment Scores based on Continent

Further breakdown into sectors shows that different continents have different sentiment for different sectors. For example, it appears that South America has a mostly neutral sentiment when it comes to the Technology sector, but has a mostly positive sentiment when it comes to the financial sector and healthcare sector, with one or two countries having a neutral sentiment.

```
Continent
Africa                    Sentiment_Scores    1.000000
                          Close              -0.033474
Antarctica                Sentiment_Scores    1.000000
                          Close              -0.003834
Asia                      Sentiment_Scores    1.000000
                          Close              -0.001990
Europe                    Sentiment_Scores    1.000000
                          Close              -0.013184
North America             Sentiment_Scores    1.000000
                          Close               0.020033
Oceania                   Sentiment_Scores    1.000000
                          Close               0.070037
Seven seas (open ocean)   Sentiment_Scores         NaN
                          Close                    NaN
South America             Sentiment_Scores    1.000000
                          Close              -0.021682
Unknown                   Sentiment_Scores    1.000000
                          Close               0.077667
Name: Sentiment_Scores, dtype: float64
```

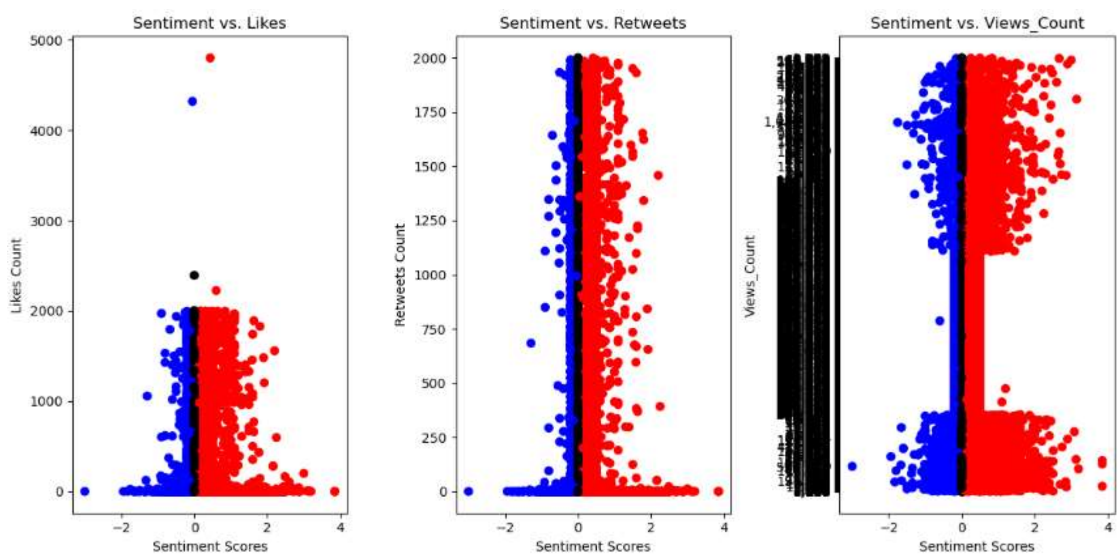
2. Correlation matrices between sentiment scores and stock price changes for each continent

The above diagram gives the correlation between sentiment scores and itself (1 for all continents) and the correlations between the sentiment scores and stock price changes which are generally weak across all continents with some minor variations. They are mostly negatively correlated which implies that stock prices will tend to decrease slightly when sentiment scores increase. However, there are positive correlations in North America and Ocenia, albeit small, suggesting a connection between high stock values and high sentiment scores

Sentiment_Labels	Price_Change_Positive	Price_Change_Negative
Continent		
Africa	-0.067437	0.055619
Antarctica	-0.010018	0.055588
Asia	-0.029856	0.065708
Europe	-0.005483	0.012072
North America	0.015386	0.048061
Oceania	0.020986	0.032350
Seven seas (open ocean)	NaN	NaN
South America	-0.033290	-0.076964
Unknown	0.018365	-0.044000

3. Change in stock price based on sentiment label across Continents

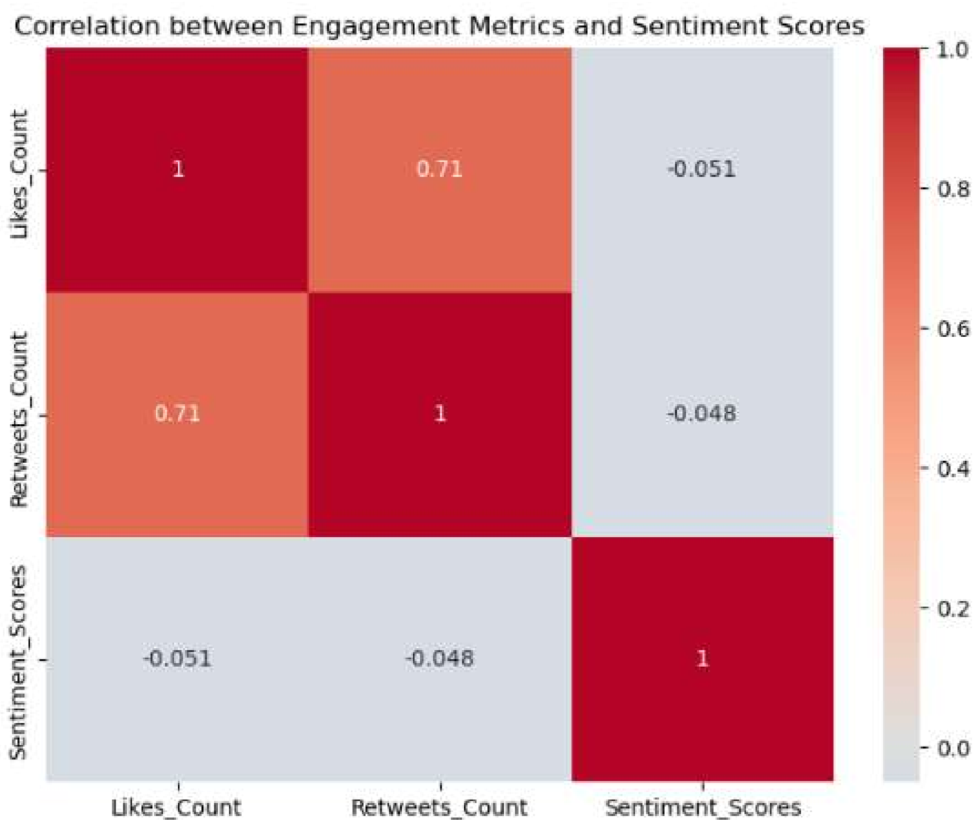
The above results show how stock prices on average tend to change based on the sentiment label across the different continents. This is such that, the 'Price_Change_Positive' is -0.067437 for Africa which means that on average when the sentiment is positive stock prices fall by about 6.7% when compared to neutral sentiment. And, a 'Price_Change_Negative of 0.55619 means that on average when the sentiment is negative stock prices increase by about 5.6% when compared to neutral sentiment.



5. scatter plots for Sentiment Score vs. Engagement Metrics

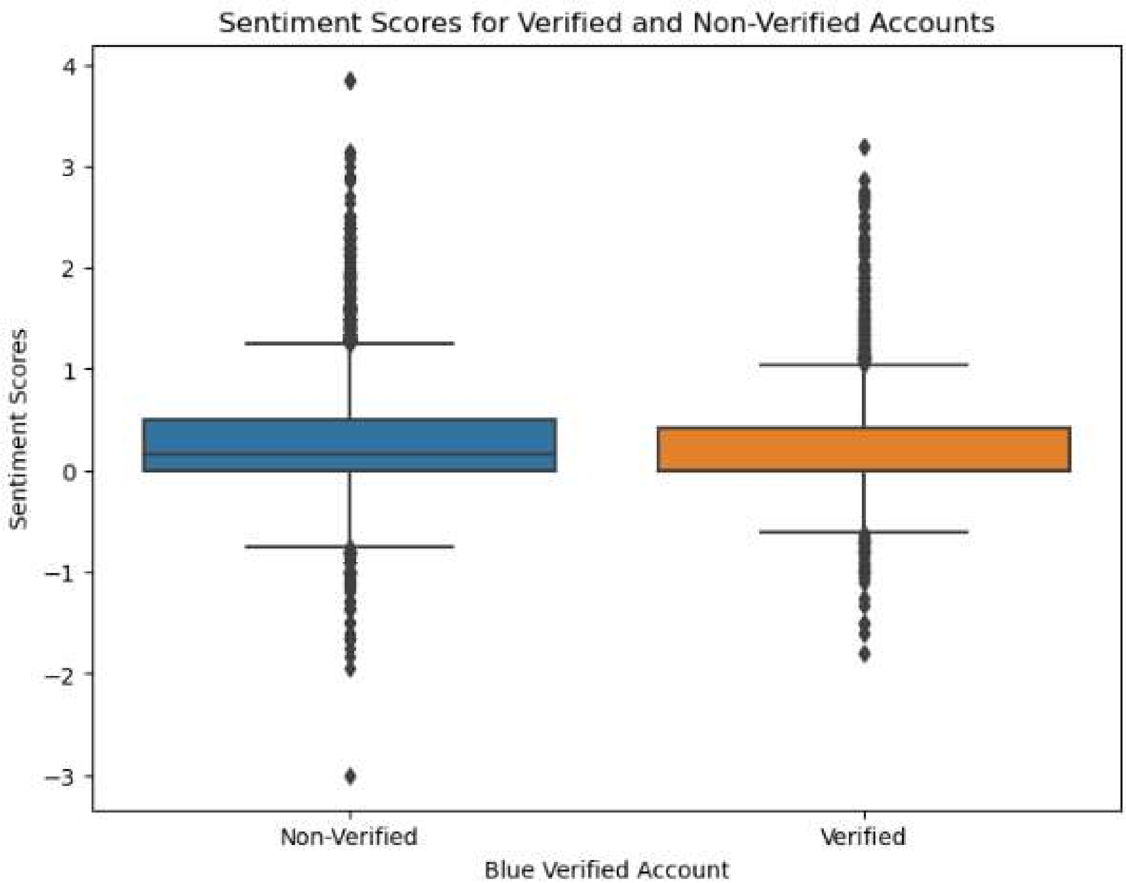
Finally, this research concludes the exploratory analysis by looking at the relationship between engagement metrics and sentiment scores. The above implies that as the number of likes on a tweet increase, the sentiment scores are more neutral. This could mean that more neutral tweets get more likes. The same can be said for view counts. However, as the number of views continues increasing, around past the average point, the sentiment scores begin to diverge away from being neutral such that they are more negative and/or positive.

```
Correlation between Engagement Metrics and Sentiment Scores:
                 Likes_Count  Retweets_Count  Sentiment_Scores
Likes_Count         1.000000        0.709686         -0.050672
Retweets_Count      0.709686        1.000000         -0.048040
Sentiment_Scores   -0.050672       -0.048040          1.000000
```



6. Correlation matrix for engagement metrics and sentiment scores

The correlation matrix above provides more information on the relationship between the engagement metrics and the sentiment scores. It suggests that while there is a high positive relationship of 0.71 between likes and retweets (posts with more likes are likely to have more retweets), the relationship between sentiment scores and retweets and likes is very weakly negative as the values are very close to zero (-0.048 and -0.051 respectively). Therefore, social media sentiment does not appear to have much of an influence on the number of tweets and retweets.



7. box plot for sentiment scores of verified and non-verified accounts

The box plots above shows that tweets by Verified account have more sentiment score closer to being neutral

MODELLING

5.1 Introduction

This section of the research implements a machine learning pipeline for predicting the ‘Adj Close’ values of stocks while putting into consideration Sentiment Scores, Social media engagement metrics and aggregated sentiment scores by sector and continent. It utilizes python libraries like pandas, scikit-learn and Gradient Boosting Regressor to handle data manipulation, for machine learning and build a predictive model respectively.

5.2 Data Preprocessing

To begin with, the data was loaded and prepared for modelling. This was done by;

- **Handling missing values;** All rows for the variables ‘Sentiment_Scores’, ‘Sector’, and ‘Continent’ that had missing values were removed. This is because these features are key in modelling, and dropping empty rows will prevent making assumptions or having biases by trying to fill the data.
- **One-Hot Encoding;** For the categorical variables ‘Continent’ and ‘Sector’, one-hot encoding is used to change them into binary columns indicating the presence of a particular column or sector. This is such that ‘Sector’ is now divided into 3 columns 'Sector_Financial', 'Sector_Healthcare', and 'Sector_Technology'. Whereby if a tweet falls within the Financial Sector the value under ‘Sector_Financial’ is 1 indicating that is where the tweet belongs while the values in 'Sector_Healthcare', and 'Sector_Technology' are 0.

- **Normalization;** Numerical features (such as 'Followers_Count', 'Total_Tweets', 'Likes_Count', 'Retweets_Count') are standardized to have zero mean and unit variance. This is done by subtracting the mean from a feature then dividing it by the standard deviation.

  Mean: average value across the variable. This centers the data around zero and therefore simplifies the optimization process when training the model

$$\boldsymbol{\mu} = {}^{\sum \boldsymbol{X}}/_{\boldsymbol{N}}$$

  Standard Deviation: measure the spread of values within a variable. This ensures that all features contribute in a similar way to the machine learning process

$$\boldsymbol{\sigma} = \sqrt{\frac{\sum(\boldsymbol{X_i} - \boldsymbol{\mu})^2}{\boldsymbol{N}}}$$

  This is done to ensure that the numerical variables in the dataset have a similar scale and distribution. It also prevents bias in the model as no one variable can dominate the machine learning process due to difference in scale.

- **Feature Engineering:** This step creates new features that are essential to the model by using the already existing variables. These new features include; Average sentiment per stock using 'Ticker', Average sentiment per sector, Average sentiment per continent, and Likes-to-Followers Ratio which is calculated by dividing 'Likes_Count' by 'Followers_Count'.

5.3 Correlation Analysis

Correlation analysis was carried out on essential features and 'Adj_Close'.

```
Correlation Matrix:
                         Sentiment_Scores  Likes_to_Followers_Ratio  \
Sentiment_Scores                 1.000000                 -0.001094
Likes_to_Followers_Ratio        -0.001094                  1.000000
AvgSectorSentiment                    NaN                       NaN
AvgContinentSentiment                 NaN                       NaN
Close                            0.010917                 -0.002361

                         AvgSectorSentiment  AvgContinentSentiment     Close
Sentiment_Scores                        NaN                    NaN  0.010917
Likes_to_Followers_Ratio                NaN                    NaN -0.002361
AvgSectorSentiment                      NaN                    NaN       NaN
AvgContinentSentiment                   NaN                    NaN       NaN
Close                                   NaN                    NaN  1.000000
```

4. Correlation analysis on essential features and ‘Adj_Close’

The results show that 'Likes_to_Followers_Ratio' and 'Sentiment_Scores' have relatively weak correlations with ‘Adj_Close’ respectively. However, it is also noted that the average sentiment related features ('AvgSectorSentiment' and 'AvgContinentSentiment') do not show any notable correlations with the ‘Adj_Close’.

5.3 Model Training and Prediction

To begin with, the ‘Adj_Close’ column is set as the target variable and 'Sentiment_Scores', 'Likes_to_Followers_Ratio', 'AvgSectorSentiment', 'AvgContinentSentiment' set as the predictor variables.

X = ['Sentiment_Scores', 'Likes_to_Followers_Ratio', 'AvgSectorSentiment', 'AvgContinentSentiment']

y = 'Adj_Close'

The data is the split into test and train sets using a 80-20 ratio, which will be used to train the model. The trained model is then used to make predictions on the scaled testing data.

$$\mathrm{MAE} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}}_{\text{test set}} \left|\underbrace{y_i}_{\text{predicted vaue}} - \underbrace{\hat{y}_i}_{\text{actual value}}\right|$$

8. MAE formula (Source: www.section.io)

Then the model accuracy is tested using the Mean Absolute Error. The MAE gives the average prediction error of the model for the ‘Adj_Close’ variable. An MAE of 55.70 suggests that the model has some degree of error when predicting the Adjusted Close prices. This implies that the model could be improved.

```
Feature Importance:
Sentiment_Scores: 0.1767838360377485
Likes_to_Followers_Ratio: 0.8232161639622514
AvgSectorSentiment: 0.0
AvgContinentSentiment: 0.0
```

5. Feature Importance

Finally, feature importance was tested to see the contribution of each feature to the model. It is noticed that Likes-to-Followers Ration had the highest contribution of 0.82 while the Average sentiment per sector and Average sentiment per continent had no contribution to the model.

To try and improve the models performance, hyperparameter tuning was employed using the GridSearchCV technique. This found the best combination of hyperparameters providing the lowest MAE. The results indicate that the best performing model has an MAE of 55.06.

```
Best Model Mean Absolute Error: 55.063589717865845
Best Hyperparameters:
{'learning_rate': 0.1, 'max_depth': 6, 'n_estimators': 100}
```

6. Best Model and HyperParameters

CONCLUSION

By examining Twitter data and stock market performance, this research delved into the complex relationship between social media mood and stock prices. Three main research hypotheses served as the foundation for the research: the potential disparities between geographical regions, the sector-specific variances in the impact of social media sentiment on stock prices, and the engagement metrics influencing this relationship.

This research found complex relationships between stock market movements and sentiment on social media. While there is a fairly small negative relationship between sentiment ratings and likes/retweets, it did find a high positive association between likes and retweets, showing that posts with more likes typically receive more retweets. This suggests that the sentiment expressed in posts may not be a significant influence on engagement metrics such as likes and retweets.

Interestingly, the correlations between sentiment scores and stock price changes demonstrated slight variations across continents, with most regions exhibiting negative correlations. This suggests a varied relationship where an increase in sentiment scores is often accompanied by a slight decrease in stock prices. North America and Oceania, however, defied this pattern with modestly positive correlations, suggesting a possible connection between higher stock values and higher sentiment scores in these regions.

Distinctive patterns are noticed upon dividing sectors. The Financial and Technology sectors displayed moderate to positive correlations between sentiment scores and stock prices, implying a certain level of influence exerted by social media sentiment. Contrarily, the Healthcare industry showed a weakly negative correlation, pointing to a more complicated interaction between investor sentiment and stock performance in this industry. Technology in particular appeared as the sector that responded to sentiment ratings the best, possibly indicating a higher sensitivity to public opinion.

Finally, a prediction model was trained and tested, resulting to a Mean Absolute Error of 55.06. This highlights the challenges in properly predicting stock price fluctuations based purely on social media opinion and suggests that other factors may be impacting market behavior.

6.1 Research Application

This study shows potential in addressing issues that decision-makers, financial analysts, and investors face in the real world. The findings provide useful insights into the intricate dynamics that affect investing decisions by throwing light on the complicated interactions between social media sentiment and stock market performance. Analysts might use sentiment patterns as an additional tool in their evaluations, while investors could use this information to improve their strategies and reduce risks. These findings could also be used by regulatory agencies and financial institutions to create policies and risk-management strategies that are better informed, resulting in a more adaptable and resilient financial ecosystem.

6.2 Future Scope

Although this study advances the understanding of the relationship between social media sentiment and stock prices, there are still a number of areas that may be worth more exploration;

- **Inclusion of Additional Data:** Adding more extensive datasets that take into account a wider array of social media platforms and news sources may provide a more comprehensive understanding of the impact of sentiment.

- **Incorporating External Factors:** Including financial reports, news events, and macroeconomic indicators can assist separate the confounding impacts of other market-driving factors.

REFERENCES

Adebiyi, A.A., Adewumi, A.O. and Ayo, C.K., 2014. Comparison of ARIMA and artificial neural networks models for stock price prediction. Journal of Applied Mathematics, 2014.

Alexander P, Ilya R, Alexey S. 2013. Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis. In: IEEE 13th International Conference on Data Mining Workshops.

Ariyo, A.A., Adewumi, A.O. and Ayo, C.K., 2014, March. Stock price prediction using the ARIMA model. In 2014 UKSim-AMSS 16th international conference on computer modelling and simulation (pp. 106-112). IEEE.

Atkins, A., Niranjan, M. and Gerding, E., 2018. Financial news predicts stock market volatility better than close price. The Journal of Finance and Data Science, 4(2), pp.120-137.

Audrino, F., Sigrist, F. and Ballinari, D., 2020. The impact of sentiment and attention measures on stock market volatility. International Journal of Forecasting, 36(2), pp.334-357.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

Bhuriya D, Girish K, Ashish S, Upendra S. 2017. Stock market prediction using a linear regression. In: International Conference of Electronics, Communication and Aerospace Technology. Vol. 2

Chen, J., Gong, X., & Wu, Y. (2021). The effect of management sentiment on stock prices: Evidence from earnings calls. Review of Finance, 25(7), 2749-2785.

Chen, L., & Zhu, Y. (2023). Sentiment-driven algorithmic trading strategy. Journal of Banking & Finance, 129, 107235.

Deng, S., Mitsubuchi, T., Shioda, K., Shimada, T. and Sakurai, A., 2011, December. Combining technical analysis with sentiment analysis for stock price prediction. In 2011 IEEE ninth international conference on dependable, autonomic and secure computing (pp. 800-807). IEEE.

Derakhshan, A. and Beigy, H., 2019. Sentiment analysis on stock social media for stock price movement prediction. Engineering Applications of Artificial Intelligence, 85, pp.569-578.

Devika, M.D., Sunitha, C. and Ganesh, A., 2016. Sentiment analysis: a comparative study on different approaches. Procedia Computer Science, 87, pp.44-49.

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2014). Deep learning for event-driven stock prediction. Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 2327-2333.

Feldman, R., 2013. Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), pp.82-89.

Fischer, T., Krauss, C., & Treichel, A. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669.

García-Medina, A., Junior, L.S., Bañuelos, E.U. and Martínez-Argüello, A.M., 2018. Correlations and flow of information between the New York Times and stock markets. Physica A: Statistical Mechanics and its Applications, 502, pp.403-415.

Gonçalves, P., Araújo, M., Benevenuto, F. and Cha, M., 2013, October. Comparing and combining sentiment analysis methods. In Proceedings of the first ACM conference on Online social networks (pp. 27-38).

Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2005). Information diffusion through blogspace. In Proceedings of the 13th International Conference on World Wide Web (WWW) (pp. 491-501).

Gupta, R. and Chen, M., 2020, August. Sentiment analysis for stock price prediction. In 2020 IEEE conference on multimedia information processing and retrieval (MIPR) (pp. 213-218). IEEE.

Hu, Z., Zhao, Y. and Khushi, M., 2021. A survey of forex and stock price prediction using deep learning. Applied System Innovation, 4(1), p.9.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed.). OTexts.

Islam, M.R. and Nguyen, N., 2020. Comparison of financial models for stock price prediction. Journal of Risk and Financial Management, 13(8), p.181.

Jasmina S, Miha G, Nada L, Martin Ž. 2013. Predictive sentiment analysis of tweets: a stock market application. In: Holzinger A, Pasi G, eds. HCI-KDD. Berlin: Springer, 77–78.

Jiang, Weiwei. “Applications of Deep Learning in Stock Market Prediction: Recent Progress.” Expert Systems with Applications, vol. 184, Dec. 2021, p. 115537.

Ji, X., Wang, J. and Yan, Z., 2021. A stock price prediction method based on deep learning technology. International Journal of Crowd Science, 5(1), pp.55-72.

Khan, Wasiat, et al. “Stock Market Prediction Using Machine Learning Classifiers and Social Media, News.” Journal of Ambient Intelligence and Humanized Computing, 14 Mar. 2020.

Kim, Y. (2014). Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746-1751.

Koukaras, P., Nousi, C. and Tjortjis, C., 2022, May. Stock market prediction using microblogging sentiment analysis and machine learning. In Telecom (Vol. 3, No. 2, pp. 358-378). MDPI.

Leung, C.K.S., MacKinnon, R.K. and Wang, Y., 2014, July. A machine learning approach for stock price prediction. In Proceedings of the 18th International Database Engineering & Applications Symposium (pp. 274-277).

Liu, B., Zhang, L., & Zhang, Y. (2015). A survey of opinion mining and sentiment analysis. Mining Text Data, 2nd Edition, 415-463.

Liu, H., Li, Z., Xu, Y., & Yang, Y. (2022). The impact of financial news sentiment on intraday stock price movements. Journal of Empirical Finance, 70, 99-116.

Li, X., Xie, H., Chen, L., Wang, J. and Deng, X., 2014. News impact on stock price return via sentiment analysis. Knowledge-Based Systems, 69, pp.14-23.

Li, Y., Yang, Y., Wang, S., & Zhang, X. (2020). A sentiment-based market index for financial market. Journal of Financial Markets, 47, 100501.

Mate GS, Siddhant A, Rutuja K, Maitreyi M. 2019. Stock prediction through news sentiment analysis. Journal of Architecture & Technology 11(8):36–40.

Mihir A, Sumedh N, Anmol K. 2020. Stock analysis using sentiment analysis and machine learning. International Journal of Innovative Research in Technology 6(11):123–126.

Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P. and Anastasiu, D.C., 2019, April. Stock price prediction using news sentiment analysis. In 2019 IEEE fifth international conference on big data computing service and applications (BigDataService) (pp. 205-208). IEEE.

Nabi, R.M., Soran Ab M, S. and Harron, H., 2020. A novel approach for stock price prediction using gradient boosting machine with feature engineering (gbm-wfe). Kurdistan Journal of Applied Research, 5(1), pp.28-48.

Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W. and Wills, G., 2020. A survey on machine learning for stock price prediction: Algorithms and techniques.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.

Pang X, Zhou Y, Wang P, Lin W, Chang V. 2018. An innovative neural network approach for stock market prediction. Berlin: Springer Science+Business Media.

Park, S.J., Kang, C.U. and Byun, Y.C., 2021. Extreme Gradient Boosting for Recommendation System by Transforming Product Classification into Regression Based on Multi-Dimensional Word2Vec. Symmetry, 13(5), p.758.

Rainforth, T., Le, T.A., van de Meent, J.W., Osborne, M.A. and Wood, F., 2016. Bayesian optimization for probabilistic programs. Advances in Neural Information Processing Systems, 29.

Ruan, Y., Durresi, A. and Alfantoukh, L., 2018. Using Twitter trust network for stock market analysis. Knowledge-Based Systems, 145, pp.207-218.

Schöneburg, E., 1990. Stock price prediction using neural networks: A project report. Neurocomputing, 2(1), pp.17-27.

Tumarkin, R., & Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. Financial Analysts Journal, 57(3), 41-51.

Xie, F., Wang, Y., Zhang, X., & Lv, J. (2020). A hybrid method for stock price trend prediction based on attention mechanism. IEEE Access, 8, 42321-42330.

Zhang, L., Wang, S. and Liu, B., 2018. Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), p.e1253.

Zhang, M., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through Twitter “I hope it is not as bad as I fear.” Procedia-Social and Behavioral Sciences, 26, 55-62.

Zhang, Y., Chiang, R. H., & Xu, H. (2021). Predicting stock price movements using Twitter sentiment analysis. Decision Support Systems, 146, 113451.