Ranjit Kumar Harijan, Yogendra Air

Impact of Social Media Sentiments on Stock Price and Trading Volume

Evidence from Google Trends, Reddit Data and Tweets

School of Accounting and Finance
Master's Degree Programme in Finance

Vaasa 2026

UNIVERSITY OF VAASA
School of Accounting and Finance
Author: Ranjit Kumar Harijan, Yogendra Air
Title of the thesis: Impact of Social Media Sentiments on Stock Price and Trading Volume: Evidence from Google Trends, Reddit Data and Tweets
Degree: Master of Science in Economic and Business Administration
Degree Programme: Master's Degree Programme in Finance
Supervisor: Timo Rothovius
Year: 2026 **Pages:** 68

ABSTRACT:

This research investigates the influence of social media sentiment on stock price and stock trading volume, recognising the increasing role of online information dissemination in stock markets. The advent of social media like Twitter (X), Reddit and online financial news and opinion sites has led to an increase in the influence of real-time information on investor sentiment. This research is based on the theories of behavioral finance and the Efficient Market Hypothesis (EMH), which explore the impact of social media sentiment on stock price and stock trading volume.

The study uses a quantitative research approach, with sentiment data from various social media and news websites, and stock price and trading volume data of a list of listed companies over a certain time period. Sentiment analysis is applied to categorize the sentiment polarity of the text data, and statistical linear and nonlinear machine learning models such as Artificial Neural Networks (ANNs) and random forests are used to examine the relationships and predictability of the data.

The findings show that the relationship between social media sentiment and stock prices is statistically weak but positive using linear methods. But the predictive value of sentiment is greatly enhanced by the nonlinear model, revealing the high explanatory power of machine learning models. This study also concludes that sentiment polarity has a greater impact than the number of mentions, i.e., the message's tone is more important than its volume. On the other hand, this study found no correlation between social media sentiment and trading volume and volatility.

The research suggests that social media sentiment is a complementary variable in financial prediction, rather than a standalone variable. It suggests that investors and financial institutions should use social media sentiment in conjunction with traditional financial data and apply sophisticated analytical methods for better predictions. Future studies should increase sample sizes, apply more sophisticated sentiment analysis models and study inter-market interactions to improve the validity and generalizability of the results.

KEYWORDS: social media, Stock Prices, Stock Trading Volume, Stock Volatility, Sentiment Analysis, Behavioral Finance, Efficient Market Hypothesis (EMH).

Contents

Figures

Tables

1 Introduction

1.1 Background of the study

The digitalisation of information and the arrival of social media have changed the way investors access, process and consume market-related information (Sul et al., 2017). Traditional financial theories, like the Efficient Market Hypothesis (EMH), assume that all publicly available information is reflected in security prices. However, empirical studies suggest that investor sentiment, and in particular the sentiment expressed in social media, play a significant role in stock market activity, stock-price volatility and volume (Bollen et al., 2011).

Historically, the quantitative analysis of investment and stock-market forecasting has relied on the analysis of price-volume data, fundamental data and traditional economic indicators. However, the world of financial forecasting has changed with the recent emergence and evolution of communication platforms. Today, social media platforms such as Twitter (X), Reddit, and even websites dedicated to investment, such as Stock Twits, are huge real-time aggregators of public opinion, investor sentiment, and economic activity (Agrawal et al., 2025; Zhou et al., 2026). These sites have changed the nature of information flow, allowing retail investors to connect with each other, to exchange emotional cues and to have an impact on the market in a new way. This has given rise to the field of behavioural finance and computational linguistics to study the importance of emotional sentiments in investment decisions and, ultimately, the stock returns.

Early research recognised the potential of social media and message boards as a source of investor sentiment and actions. Research on message boards (Raging Bull and Yahoo Finance) showed that investor sentiment can help predict stock returns and volatility (Antweiler & Frank, 2004; Tumarkin & Whitelaw, 2001). More recent studies found that negative sentiment in financial newspaper columns can lead to stock price drops (Tetlock,

2007) and that microblog sentiment, especially from the service Sina Microblog (Weibo) has statistically significant associations with short term stock price movements (Pang et al., 2012; Cheng & Lin, 2013). This research has paved the way for the current sentiment analysis, which measures sentiment cues in large volumes of text data.

Meanwhile, the study of leadership has also demonstrated that the sentiment of influential leaders (such as CEOs) affect the market. Online communication by CEOs, through platforms such as Twitter and Weibo is now a powerful medium for communication of corporate messages and sentiment signals to investors. The research on CEO posts demonstrates that CEO sentiment, key topic and virality significantly enhance volatility predictions and even affect market expectations (Zou et al., 2026). The high dimensional semantic features extracted by BERT and other deep learning semantic models suggest that the CEO's leadership style (transactional or transformational) affects the stability of the sentiment signals as well as their price generating power. For instance, stable and pragmatic styles of communication are in low entropy environments, whereas highly emotive or disruptive styles of communication create noise, instability and greater extremes in volatility (Zou et al., 2026). In conclusion, investor sentiment analysis needs to consider not just the text but also the patterns, styles and impacts of the generators.

There's evidence that markets are not purely rational or efficient; they are driven by a range of behavioural factors, biases and sentiment. The recent stock market anomalies (e.g. GameStop short squeeze) driven by retail investor sentiment on social media (e.g. Reddit) have demonstrated that the sentiment of the crowd on social media can influence markets; causing stocks to experience extreme price movements and spikes in the volume (Tarsi et al, 2024; Zhou et al, 2026). Such events show us that social media are no longer just a reaction to stock markets, but can influence them. There is now a trend for institutional investors and trading firms to incorporate social media information in their algorithmic trading strategies to capture such information (Zhou et al., 2026).

This development highlights the need to further explore the role of sentiment in social media content on investment and trading.

Although there have been some methodological advances in this area, there remains a need to better understand the effects of social media sentiment on stock returns and volume. First, although many studies have used microblogs (e.g. Sina Weibo), few have investigated the predictive ability of the combined sentiment, attention (e.g. discussion volume, portfolio sharing) and social factors. Some open source social data (e.g., stock attention and transaction sharing) may be more important than public sentiment in predicting the next day stock volatility (Wu et al., 2017), which challenges the traditional sentiment only models. Second, little is known about the longevity of sentiment effects. The impact of social media sentiment can last for three days, but quickly decays (Wu et al., 2017). Lastly, the growing sophistication of online communication stresses the importance of leadership driven sentiment signals and channels of transmission that impact the way that online narratives are transmitted into market sentiment (Zou et al., 2026).

For these reasons, I am motivated to conduct this study for both practical and theoretical purposes. Pragmatically, an increasing number of investors rely on social media (e.g. Twitter, Reddit) to support their investment decisions. I hope to be able to provide insights into the effects of sentiment in online media on investment decisions to better equip investors, analysts and financial firms to deal with market uncertainty and to recognise shifts in investor behaviour. Theoretically, the integration of knowledge in the field of behavioral finance, machine learning and information diffusion theory will provide a multifaceted view of the evolution of emotions in online media and their influence on stock prices. This research is thus important as it will critically examines the impact of social media sentiment on the stock market and the importance of sentiment signals.

1.2 Problem Statement

Investor-generated information has taken on a new life in the world of social media, with its unprecedented volume, velocity and accessibility. However, there remains a lack of clarity in this area of social media sentiment and its impact on market dynamics. While previous studies have shown that social media sentiment and attention measures affect short-term stock prices, we're not yet clear about the extent and nature of the relationship. For instance, although research has demonstrated that emotional polarity is related to stock price volatility, other evidence has found that social data such as the frequency of discussion, stock attention and sharing of transactions are more significant than sentiment (Wu et al., 2017). This calls for a better understanding of the roles of sentiment and social-interaction factors in the market.

Additionally, there is an unresolved question about the time frame of sentiment influence. The effects of public mood do not play out the same way across assets; highly speculative securities may have instantaneous price shocks from online sentiment but large diverse firms may have a lagged response (Tarsi et al., 2024). While spikes in negative sentiment have been linked to short term price fluctuations in technology stock and the financial sector (Khalil, 2025), there is a lack of evidence on the persistence and decay of sentiment signals. Sentiment signals may only have influence for one to three days and their predictive ability may decay quickly over longer time frames (Wu et al., 2017).

The lack of a unified approach to simultaneously filter noise in social media, extract signals from experts, account for different time lags and integrate "emotional" sentiment with technical indicators is a significant gap in the current set of predictive models in finance and a question of temporal persistence and tradability of sentiment signals. These issues result in an incomplete picture of the impact of social media sentiment on investment prices and volume. As such, a systematic, empirical study is needed to establish the role of social media sentiment signals in investment and market activities.

1.3 Research Aim and Research Questions

This study seeks to examine the effects of social media sentiments on stock prices and trading volumes by studying the role of social emotions, opinions, topics and online behavioral metrics in affecting investor behaviour and short-term price fluctuations. It aims to combine sentiment signals with social data indicators to assess their predictive capabilities and impact on stock price and trading volumes.

Based on this, the Key research question are as follows:

1. How much do sentiments, topics, news, economic and financial information from social media impact stock price forecasts and trading volumes?
2. How do investors' sentiments influence stock prices and trading volumes?

To achieve the research aim and research questions of this study, this current research will fulfil the following objectives

1. To investigate and examine the predictive values of users' sentiments on social media platforms in predicting stock prices and trading volumes using machine learning techniques.
2. To explore and study the predictive power of social media sentiments on stock prices and trading volumes.

1.4 Significance of the Study

This research has theoretical and practical importance, by adding to the body of knowledge in the disciplines of behavioral finance, computational linguistics and algorithmic trading. From a theoretical perspective, this research makes a number of important academic, methodological and practical contributions to the field of behavioral finance. First, it adds to the body of knowledge of behavioral finance by investigating the effect of social media sentiment in terms of emotional tone, discussion

clusters and user generated commentary, and online herd behaviour on investor behaviour and the momentum of the market (Zhou et al., 2026). While other research has identified links between internet sentiment and stock price volatility (Pang et al., 2012; Cheng & Lin, 2013), this research goes beyond to also include factors such as attention measures, transaction sharing behaviour and leadership sentiments.

Pragmatically, the research has important implications for investors, portfolio managers and regulators. Online information is becoming a key source of market information for investors, but the veracity of sentiment signals is questionable given potential noise, manipulation and biases of online platforms. The research enhances the transparency of sentiment signals, thus helping to manage risks, allocate resources and monitor markets. For policymakers, findings relating CEO driven volatility (Zou et al., 2026) in sentiment-based signals can guide future policy in the areas of digital communication and market integrity.

1.5 Research Structure

This study comprises six chapters. Chapter 1 provides an overview of the research, including the motivation, questions and importance. Chapter 2 provides a literature review of sentiment analysis, social media and financial markets. Chapter 3 describes the methodology, including sources of financial and social media data, processes of sentiment extraction and analysis techniques. Chapter 4 discusses empirical findings. Chapter 5 provides a discussion of these findings from a theoretical and practical perspective. Chapter 6 provides a conclusion and suggestions for future research.

1.6 Chapter Summary

This chapter of the present study has laid the foundation for the study. It discusses firstly, the background of the study and secondly, the problem statement, research question

and objectives for the study. Then, the justification for the study and the significance of the study are explained in this chapter.

2 Literature Review

This chapter presents a review of the literature about the effects of social media sentiment on stock prices and volume. The review revisits three major themes pertinent to the current study: sentiment analysis in financial markets, social media activities related to financial markets, and the links between sentiment signals and financial markets. The chapter starts with a review of theoretical background, including the Efficient Market Hypothesis (EMH) and behavioral finance that offer competing views on the impact of information on stock prices. The chapter then examines the influence of social media on the information structure of financial markets and reviews research that examines the forecasting ability of sentiment derived from social media. Finally, the chapter proposes a conceptual framework and hypotheses for the current research.

2.1 Sentiment Analysis in Financial Markets

2.1.1 Concept of Sentiment Analysis

Sentiment analysis is the computational approach for determining the emotional tone of textual information. It is commonly applied in natural language processing and machine learning and used to detect positive, negative or neutral opinions expressed in online media. Within the domain of financial markets, sentiment analysis aims to categorise textual data, from formal news reports to informal tweets, into various emotional categories, usually divided into bullish (positive), bearish (negative) and neutral. The idea is that these codified sentiment indicators are a proxy for investor sentiment and can be used to predict market movements (Agrawal et al., 2025).

Sentiment analysis has been conducted using various techniques, including lexicon-based and machine learning. These methods involve the use of dictionaries of words with positive and negative sentiment, while machine learning approaches train models to predict sentiment based on labelled data. Dictionary-based methods, such as the Valence Aware Dictionary and sentiment Reasoner (VADER) and Text Blob, use a list of

words that are given a certain polarity value. These methods are very effective in analysing the informal, short text often used on social media platforms, such as Twitter and Reddit. For example, Surulivel et al. (2025) successfully used VADER and TextBlob to assess public sentiment towards Tesla and found that VADER is especially effective to assess the nuanced emotional states and different levels of intensity that are common in informal social media messages.

On the other hand, Machine learning approaches (such as neural networks and deep learning) enable researchers to analyze large quantities of unstructured data and detect complex sentiment patterns. Machine learning approaches are extensively applied in financial sentiment analysis because they are able to classify large volumes of social media data (Skuza & Romanowski, 2013). But they also need significant data pre-processing and computational power.

Machine learning techniques seek to capture the meaning of words in a sentence. Derakhshan and Beigy (2019) contributed to this area by developing a part-of-speech (POS) graphical model, called LDA-POS. Their method uses Latent Dirichlet Allocation (LDA) topic models and the POS tags to separate words into different parts of speech (verbs, adjectives, nouns, prepositions) and then infer the topic distributions. The LDA-POS approach was tested using English and Persian language datasets to better predict stock movements than baseline approaches using exclusively human labeled sentiment, highlighting the importance of syntactic information in computational sentiment analysis (Derakhshan & Beigy, 2019).

2.1.2 Previous Studies on Sentiment and Financial Market

The impact of sentiment on the stock market has been the subject of research for decades. The initial studies mainly relied on traditional media outlets like newspapers and financial reports.

An early study in the area was Tetlock (2007), which analysed the mood of financial news stories and found a negative media sentiment led to downward pressure on stock prices.

This study showed that the emotional cues in words can affect market participants and market performance.

Other research suggests that the research into sentiment in financial markets began as a reaction to the Efficient Market Hypothesis (EMH). EMH is a theory proposed by Fama (1970) that says prices reflect all existing information and hence, forecasting is close to impossible (Piñeiro-Chousa et al., 2017). But the emergence of the field of behavioral finance suggests that the concept of "bounded rationality" suggests that human cognitive constraints and biases produce irrational market activity (Ren et al., 2020).

The first empirical research into online sentiment examined traditional message boards. Researchers who analysed internet message boards such as Yahoo! Finance and RagingBull.com produced a wide range of conclusions about their effectiveness. While some early studies, such as Wysocki (1998), reported that the volume of messages could explain the change in stock returns for the following day, Antweiler and Frank (2004) examined internet stock message boards and found that internet discussions were related to increased trading volume and volatility, some others, including Tumarkin and Whitelaw (2001), concluded that such internet postings were predominantly cheap talk and represented market noise rather than information (Piñeiro-Chousa et al., 2017). Notwithstanding these conflicting results, these early studies did underline the theoretical need to consider electronic communication as a potential source of market instability and investor herding, thus opening the way to the modern, high frequency social media analysis (Piñeiro-Chousa et al., 2017).

2.2 Social media and Investor Behavior

2.2.1 The Emergence of Social Media as an Information Source

The last ten years have seen social media platforms replace the mass media as the main vehicle for the rapid information and rumour sharing. Websites like Twitter, Reddit,

StockTwits and Sina Weibo have empowered the individual to broadcast financial information, discuss corporate earnings, and collectively move prices (Teti et al., 2019). The real-time nature and scale of this communication have transformed information flows.

Social media platforms have been shown to be "nowcasting" public mood. Through the analysis of millions of online messages, researchers can detect patterns in social mood that can affect the markets. Bollen et al. (2011), who used Twitter data to gauge social mood using mood tracking utilities and neural networks, found that some emotional states were very successful at predicting stock market fluctuations. This suggests that social media activity may capture the social mood that can affect economic activity.
Guo et al. (2017) show investor sentiment rapidly spreads through social networks to affect the market when attention is high. Their findings using the Thermal Optimal Path analysis shows that sentiment is not always a leading indicator of price with some cases the lagging indicator suggesting a dynamic relationship. Likewise, Ferraro and Sperlì (2024) note that sentiment, memes and group dynamics during GameStop short squeeze were amplified through Reddit and Twitter interactions, which were examples of real time collective coordination.

Finally, existing research shows that social media do not operate in isolation. It is not isolated but works in conjunction with the mass media. Ren et al. (2020) examined this phenomenon in the Chinese stock market, by analysing a large pool of 60 million Weibo posts and 6.2 million news articles from Sina Finance. They discovered the presence of a demand-driven media bias, where mass media outlets tilt the sentiment of their stock market news to align with the demand and sentiment of retail investors expressed on social media. Following the psychological concept of cognitive dissonance that people tend to consume information that supports their existing views, mass media outlets boost sentiment persistence to maximise viewership (Ren et al., 2020). As such, social media serves as a sentiment filter. The stability of mass media sentiment is greatly reinforced when social media responses are consistent with the news.

2.2.2 Social Media Sentiments and Market Predictions

A number of empirical studies have looked at the predictive power of social media sentiment. Studies using data from Twitter (X) have shown that sentiment derived from tweets enhances stock market forecasting models. For example, research has demonstrated that including social media sentiment in prediction models leads to better predictions than models based on historical stock prices alone (Hai et al., 2015). Likewise, Pagolu et al. (2016) found a significant relationship between social media sentiment and stock index returns. This research showed that positive sentiment expressed on Twitter indicates stock price rises.

Bollen et al. (2011) pioneered a model which demonstrated that certain mood dimensions derived from Twitter predict Dow Jones Industrial Average (DJIA) returns. Pang et al. (2012) and Cheng and Lin (2013) found predictive sentiment in Weibo for short-term stock price predictions in China. Ferraro and Sperlì (2024) demonstrate that herding by investors on Reddit and the associated Twitter commentary have led to price increases of more than 1700%. Their deep learning model shows that combining Reddit and Twitter sentiment enhances the understanding of stock price movements during such events, albeit to varying degrees as noisy, emotive data is involved. Therefore, social media sentiment can go beyond passively reflecting stock markets to driving speculation and herds.

But this positive outlook is heavily challenged by others who emphasise the noisiness of open social media platforms. Zhou et al. (2026) pointed out that most social media content is speculation, based on cognitive biases, or created by bots and spammers. Through empirical studies they showed that directly aggregating public opinion results in a prediction accuracy of around 47%, which is worse than chance. To fully leverage social media for prediction, Zhou et al. (2026) proposed a dynamic expert tracing algorithm to clean noise and discover market experts. Notably, their algorithm also uncovered inverse experts that are users who consistently provide the wrong predictions as a result of common decision-making biases. By using the forecasts of inverse experts as reliable negative signals and a Dynamic Graph Attention Network (DualGAT) to diffuse these sparse "expert" signals among correlated stocks, Zhou et al. (2026) were able to

increase the prediction accuracy to 72.8%. This research suggests that while social media can represent investors' actions, its effectiveness in predicting stock price movements is dependent on using sophisticated algorithms to filter unreliable information and evaluate source quality.

2.3 Social Media Sentiment and Stock Price Movement

The link between social media sentiment and stock prices is a common theme in both behavioral finance and computational stock market prediction studies. Existing research consistently demonstrates that investor sentiment expressed by online investors, be it on Twitter, forums, financial news websites, or search engines, contains valuable information, which has implications for expected returns, volatility, and volume. However, the magnitude, timing and direction of the impact vary across social media platforms, financial markets, and research methodologies.

Research consistently shows that online sentiment is positively associated with price momentum, and negatively associated with price decline. Teti et al. (2019) studied this phenomenon in the U.S. tech market, using Bloomberg's in-house Twitter sentiment index. Using Ordinary Least Squares (OLS) and Fixed Effects models, they demonstrated that the Twitter sentiment is a significant predictor of returns for a sample of 69 technology firms. Significantly, their findings also pointed to the time lag: while using a neutral sentiment index, the highest predictive power was found with a one-day time lag, suggesting that the irrational, sentiment-based response to the stock price is in the short term before the fundamentals take over again (Teti et al., 2019). Surulivel et al. (2025) also illustrated the differences in sentiment sources by comparing data from traditional news outlets and social media (Reddit) for Tesla (TSLA). Their polarity distribution analysis showed that while financial news articles are strictly neutral to avoid influencing the market, social media is highly polarised: overwhelmingly positive or negative with a quick reaction to short-term events. This suggests that the instantaneous nature of social media and its easy accessibility directly affect retail investors' attention

and investment decisions more so than news media, which in turn leads to swift price fluctuations in highly visible or popular stocks (Surulivel et al., 2025).

Nti et al. (2020) offer some of the most compelling evidence that social media sentiments positively impact future stock prices. Their research on the Ghana Stock Exchange shows that sentiment data from Twitter, Google Trends, forums and financial web news play a pivotal role in enhancing the accuracy of predictions, particularly when these multiple data sources are combined into a single sentiment-based input model. The artificial neural network model they developed, which delivered up to 77% accuracy for 90-day-ahead forecasts, suggests that the fusion of multi-platform sentiment signals improves the quality of the signal and noise reduction compared to single-source signals. A major takeaway is the strong correlation (0.9681) between public sentiment and volume, suggesting that investor conversations in social networks may impact price and volume.

On the other hand, Smith and O'Hare (2022) challenge the causal effect that social media, particularly Twitter, has on daily stock price changes. The research, which examined the sentiment of CEO tweets, regular news and financial news across large US companies, reveals only a weak or intermittent correlation between the sentiment of CEO tweets and stock price fluctuations, with most meaningful correlations occurring only in extreme market conditions (e.g., the COVID-19 pandemic). In their study, short-term sentiment-price correlations were weak, sometimes negative, and inconsistent across firms, industries and time frames, implying that social media sentiment may not be a good predictor of stock prices in normal market conditions. Moreover, their time-lag analysis shows that, in most cases, market prices seem to influence sentiment, particularly in the case of CEO tweets and news stories. This is in contrast to Nti et al. (2020) who found sentiment enhances predictive accuracy in an emerging market.

2.4 Social Media Sentiment and Trading Volume

Investor sentiment also affects an important measure of market activity - trading volume. When trading volumes are high, this can be indicative of increased market participation. When investors talk about a stock on social media, it can generate trading volume. Research has demonstrated that trading volume spikes occur in response to social media activity. The increased visibility the stock receives may prompt investors to view the stock as an investment opportunity (Teti et al., 2019).

There is a large body of empirical evidence on the direct effects of social media sentiment on stock prices and trading volumes. The earlier results were somewhat mixed. For example, while some studies found the volume of messages on Yahoo! Finance were associated with next-day returns, others did not find such an association, concluding that social media comments were primarily cheap talk (Piñeiro-Chousa et al., 2017). But with the evolution of Natural Language Processing (NLP) techniques, it has become possible to extract fine-grained sentiment polarity (optimism or pessimism), giving more consistent results.

Tarsi et al. (2024) examined the links between Twitter activities (volume, likes, and polarity) and next-day returns of large firms such as Tesla and Amazon. Their results show that the effect of social media sentiment varies across assets. Unstable and innovative firms such as Tesla demonstrated a sentiment-induced quick response. Tarsi et al. (2024) found that predictive models performed best for Tesla with a time lag of one or two days. In contrast, less volatile and well-diversified corporations, such as Amazon, showed a slower market reaction, with predictive models performing best with a four-day lag. This implies that although investor sentiment is affected by public sentiment, the persistence of these effects varies greatly and is dependent on the characteristics and market capitalisation of the company.

2.5 Theoretical Underpinnings

2.5.1 Efficient Market Hypothesis (EMH)

The Efficient Market Hypothesis (EMH), introduced by Eugene Fama in 1965, is the foundation for explaining stock market behaviour. The main idea behind the EMH is that markets are efficient with respect to information, in the sense that the stock prices at any time incorporate all available information (Malkiel, 2003; Sprenger & Welpe, 2010). Fama defined three versions of efficiency: weak, semi-strong and strong. The weak version of the efficient market hypothesis (EMH) holds that the price of traded securities reflects all publicly available information about the market. It holds that the past price and volume of the security are independent and do not forecast the future price of the security. The semi-strong form is the more popular notion as it adds to this by incorporating all publicly available information (including past and present information) and incorporating new public information into the current price of securities instantaneously (Malkiel, 2003; Sprenger & Welpe, 2010). For research purposes, the semi-strong version is the most important because it assumes that prices incorporate all publicly available information (economic news, social media) and that prices instantly adjust to new information (Sprenger & Welpe, 2010). As a result, the EMH suggests neither fundamental nor technical analysis can be used consistently to generate returns in excess of the risk-adjusted returns of the market (Malkiel, 2003).

The efficient market hypothesis (EMH) is strongly associated with random walk theory (RWT), which assumes that stock price fluctuations are random deviations from historical prices. An efficient market prices news immediately and, as news is unpredictable, the resulting stock price will be unpredictable (Malkiel, 2003). Saeedi et al. (2014) highlight if a daily return follows the random walk, the market is regarded as weak-form efficient, which means that prices and trends are not useful for predicting future returns. This study uses RWT to examine if social media sentiment creates a predictable pattern that would contradict the price independency assumption.

While the Efficient Market Hypothesis (EMH) has dominated research in financial markets for a long time, recent studies have shown mixed evidence for its validity. Some argue that while the market may not be efficient in the short term (such as during the 1987 stock market crash or the dot-com bubble in the 1990s), it is efficient in the long term (Malkiel, 2003). Also, Sprenger and Welpe (2010) argue that while news may be incorporated in prices, non-quantifiable data may not be fully or instantaneously incorporated into prices. This means that social media can be an important database, contradicting a semi-strong form of the EMH (Sprenger and Welpe, 2010).

Criticisms of the efficient market hypothesis (EMH) may be found in the field of behavioral finance, which focuses on psychological aspects of price setting. Behavioralists claim investors are prone to herd behavior and overreaction to information, which results in mean reversion or momentum (Malkiel, 2003). In emerging markets, empirical studies have rejected the random walk hypothesis and demonstrated that technical skills can lead to excess returns (Saeedi et al., 2014). These are used to interpret the social media data in this study

2.5.2 Behavioral Finance Theory

The main theoretical perspective that can be applied to understand the impact of social media sentiments on stock prices and trading volume is Behavioral Finance theory. In contrast to the Efficient Markets Hypothesis, Behavioral Finance asserts that investors exhibit irrational behaviour; cognitive biases, heuristics, emotions and social influences combine to distort investor decision making and behaviour (Baker & Nofsinger, 2010). Fundamental psychological processes like representativeness, availability, overconfidence and loss aversion affect investor judgement and decision-making processes, leading to systematic market anomalies such as underreaction, overreaction and excessive volatility (Schwartz, 2010; López Cabarcos et al., 2019).

Sentiment (the optimism or pessimism reflected in investors' beliefs) is an important variable within the Behavioral Finance theory. López Cabarcos et al. (2019) highlight that

sentiment, especially as expressed through microblogging sites, is a key behavioral force in stock markets, impacting on pricing, noise trading and market mispricing. The behavioral view also emphasises the impact of emotions, crowd and contagion effects on aggregate market behaviour.

Li et al. (2019) build on this work in showing how social media moods amplified by social influence forces on social networks like Tencent Weibo create behavioral feedback effects that can explain short term variations in high-frequency stock markets. Their research findings directly verify Behavioral Finance theories that investors' emotions and socially amplified opinions have an effect on return patterns, volatility and volume.

Following these insights, the current study assumes social media sentiment is a behavioral signal of cognitive biases and emotions. This approach therefore informs our investigation of how social opinions, debates and financial news collected from social media affect investor behaviour and impacts stock returns and volume.

2.6 Conceptual Framework

The theoretical framework of this research combines Efficient Market Hypothesis and Behavioral Finance theory with empirical evidence on social media mood to understand how online sentiment, moods and social interactions impact stock price and volume. Behavioral Finance theory argues that investors rely on cognitive biases, heuristics, emotions and social factors in their decision-making, rather than rational analysis (Baker & Nofsinger, 2010). These psychological factors create the basis for understanding the market-level implications of information expressed on social media.

Our model includes social influence factors, as investors' social media interactions create herd effects and contagion. Li et al. (2019) show that moods posted on Tencent Weibo can spread among users via retweets, comments, and social networks, resulting in

feedback loops that mirror herd effects. This has a direct implication for the high-frequency trading environment, where investors have less time to make decisions based on more heuristic-based processing.

In this current study, social media mood is viewed as the independent variable, with stock prices and trading volume as the dependent variables. This approach therefore conceptualises social media as an information environment impacting investors' psychology and, in turn, stock market activity.

2.7 Hypothesis Development

Drawing from the theoretical underpinnings of behavioral finance and the empirical evidence from the existing literature, the following hypotheses have been formulated to answer the key research questions of this study.

Social media is an information marketplace where investors share their views. The theory of behavioral finance posits that people are affected by emotional information (Teti et al., 2019). Previous research has demonstrated that investors are more likely to buy shares when there is positive sentiment in the market, and to sell shares when there is negative sentiment. This study, therefore, proposes that

H1*: Changes in social media sentiment have an impact on stock market investment.*

Li et al. (2019) show social influence on Tencent Weibo leads to the diffusion of mood, herd behaviour and mood contagion which in turn has an impact on short term stock indices. The study supports claims of a relationship between social media and the stock market. Therefore, positive sentiment cues can increase optimism, decrease risk perception and encourage speculations, causing price increases. In contrast, negative sentiment might increase pessimism and fear. Thus this study hypothesises that

H2: *Changes in social media sentiment correlates positively stock price.*

Sentiment can also impact trading volume by drawing investors' attention. There is evidence that negative sentiment is linked to higher trading volume and volatility as a result of panic selling and hedging (Teti et al., 2019). Thus, the third hypothesis is developed:

H3: *Negative change in social media sentiment has a positive effect on trading volume.*

Previous research shows that social media does not simply report news, but also creates and shapes market sentiments significantly impacting retail investors' sentiments. Ren et al. (2020) demonstrated that, because of the strong need for mass media to retain viewer interest, mass media often follows the trend of social media sentiment, known as demand-driven media bias. This results in a feedback loop that amplifies first impressions which influence group investment behaviour. Moreover, Piñeiro-Chousa et al. (2017) showed that the characteristics of the user and previous experience play a significant role in other market participants' responses, and ultimately affect market risk. As such, this research hypothesises that

H4: *Negative social media sentiment increases stock market volatility*

2.8 Chapter Summary

Literature review is a chapter that briefly outlined the literature on sentiment analysis, social media sentiments and stock market predictions. This chapter develops a critical review of the earlier studies which have dealt with some of the issues of this present research and a summary of the theoretical framework used in this research and the limitations of the models based on the research findings of previous studies.

3 Research Methodology

This chapter on research methodology provides an overview of the methods used for data gathering and hypothesis testing. The study was a deductive and causal quantitative descriptive study that detailed the correlation between social media sentiment, stock prices and stock volume. The time frame for the study is between January and December 2025. We downloaded secondary data for the three variables from public sources. The data seemed to be fairly reliable, as raw data was obtained. The text was then processed to enable quantitative analyses.

3.1 Research Paradigm

A research paradigm involves a system of beliefs, emotions and expectations about some aspects of the world and how it might be understood and studied (Collis & Hussey, 2021) in this instance, how scientific research is undertaken (Orlikowski & Baroudi, 1991). Myers & Klein (2011) notes that researchers can differ in their beliefs and values, and all research is based on initial assumptions about the nature of good research and what is deemed appropriate with regards to research methods. Researchers' field work is informed by a research paradigm which in turn has three interrelated main beliefs: ontology, epistemology, and methodology. Pozzebon (2004) adds that research paradigms assist researchers to discover their true selves and to provide them with a pathway for reaching the values they desire.

There are two distinct ways of determining research methods; one is positivism (Hussey & Hussey, 1997) and the other interpretivism (Mingers, 2001). The scientific method (also known as positivist methods) has tended to be based on quantitative methods; and interpretivist methods on qualitative methods. The philosophical underpinning of either approach can be beneficial or detrimental to research projects (depending on the situation) but is goal oriented (Bryman, 2016).

In the positivist approach, research is seen as value-free and researchers are neutral and impartial, or more accurately, neutral and objective (Collis & Hussey, 2013). Collis and

Hussey (2013) describe positivism, which is based on natural science methods, as derived from realist philosophy, which assumes reality exists independently from our knowledge of it. It seeks to inductively develop theories from empirical data through observation and experimentation.

The positivist paradigm is deductive from a philosophical point of view: it starts with a theory, develops hypotheses and then conducts data collection. Positivist research focuses on empirical verifiability of theories to explore the universal principles or laws that govern the natural and social world, and to strengthen our ability to predict events or phenomena (Orlikowski & Baroudi, 1991). For positivists, the question is how to use random sampling to measure the outcomes and build causal models with obvious and significant predictive variables (Myers & Avison, 2002).

The interpretivist paradigm is rooted in an ontological assumption that reality is socially constructed and interpreted by humans as social actors, based on their beliefs and values (Andrade, 2009; Saunders et al., 2011). interpretivists reject any research that assumes an objective perspective and stress human understanding and interpretation as part of valid knowledge (Gray, 2013; Saunders et al., 2011). The purpose of interpretivist research is not to test a hypothesis, but to understand, describe and explain the relationships and inter-relationships between all factors in a social context (Oates, 2006).

3.2 Research Approach

Research can be done in two ways: quantitative or qualitative. The roots of quantitative research can be found in the study of natural phenomena in the natural sciences (Saunders et al., 2011). The aim of this type of research is to quantify objects and their relationships. Also, quantitative research methods are not often concerned with the context of the data; they focus on large volumes of data without context, and look for statistical significance (David & Sutton, 2004; Neuman, 1994). Quantitative research methods include such techniques as questionnaires, laboratory experiments, simulations, mathematical models and econometrics (Myers, 2013; Neuman, 2005; Myers & Avison, 2002).

Unlike quantitative research, qualitative data uses words or pictures (Johnson & Harris, 2002; Miles & Huberman, 1994). The aim of qualitative research is to understand human behaviour, with an emphasis on defining the data collection methods. Qualitative methods are described as phenomenological, subjective or non-positivist. They can be applied when the problem or issue is not well defined. Qualitative researchers tend to study the meaning that people derive from their experiences with certain problems. Qualitative researchers usually have small sample sizes and perform in-depth analyses of the status quo (Berg, 2014). Interviewing respondents face-to-face will allow researchers to examine the behavior and responses of research subjects in their own environment (Creswell, 2012).

3.2.1 Choice of Methodology

This research adopts a quantitative research approach using positive methodology. This study aims to explore how sentiments on social media can impact stock prices and stock trading volumes. This study will generate quantitative evidence by assessing the gaps in the finance discipline by using some theories and models, and also to test some hypotheses that will be specified by using appropriate deductive reasoning. Furthermore, this current research will adopt data mining technologies in gathering secondary data such as historical financial data of chosen organisations, financial news and comments and tweets. For these reasons, a quantitative positivism paradigm is considered as the most suitable approach for this study.

3.3 Population

The primary factors in this study are the availability of social media posts and comments for each company between the period of January to December 2025, and availability of stock price and trading volume from January to December 2025. The target of the study

is considered as all companies worldwide with available social media posts and comments about their stocks' value from January to December 2025.

The population was restricted to the following. First, the companies to be included in the study must have been listed on an approved stock exchange for a certain period (January to December 2025). This enables the downloading of daily stock prices. We excluded companies that merged, or changed their stock symbol during this period to avoid any abnormal events that may affect their stock value, volume or social media sentiment. Second, it must be mentioned on Twitter between January and December 2025. Finally, the companies in the sample must have global sentiment data for at least 50% of the days. This means that they should have global sentiment data on social media for at least 183 days from a total of 365 days. This is to ensure the validity and reliability of the results.

3.3.1 Unit of Analysis

The unit of analysis of this study is companies. The choice of companies as unit of analysis is because it enables us to study the dependent and independent variables within the company. For instance, the sentiment of a company's social media can be used to predict its stock price, and stock trading volume.

3.4 Sampling Method and Sampling Size

This study uses purposive sampling. The reason for this is that following the inclusion criteria, it was impossible to obtain a full list of the population, to analyse and statistically register it to form a sampling frame. In this study, research subjects are selected based on whether the companies satisfy the selection criteria.

The automatic extraction and data processing capabilities of social media platforms, as well as the wealth of secondary data on stock prices and volume of stock trading, allowed purposive sampling. Based on the selection criteria, 10 publicly listed companies were selected. The research automatically gathered and transformed daily sentiment, stock prices and stock trading scores from January to December 2025.

3.5 Data Collection

The study started with the data collection in 2025 from different sources to achieve a complete picture of both the market and sentiment. The research examined the following ten prominent firms: Amazon (AMZN), Google (GOOG), Tesla (TSLA), J.P. Morgan (JPM), BAC, Netflix (NFLX), NVIDIA (NVDA), Apple (AAPL), MRK and Microsoft (MSFT).

Data from the stock market was collected to provide daily stock prices (open, close and previous close) and shares traded. This data was used to calculate stock returns and volume.

For sentiment data, we used three primary sources of text data. X (formerly Twitter) data gave short, live reactions from a wide variety of users. Reddit data provided more conversational forms of data, typically more in-depth and analytical content from a community perspective. Market news data (from FinnHub) was also used to provide more professional written commentary on what was happening in the market.

Further, Google Trends data was used to capture the interest of the public. This offered an alternative view, capturing the search volume for particular companies.

Each dataset was acquired using Python and stored in a structured format (CSV). Data from X, Reddit and financial news were analysed using natural language processing to

extract sentiment scores (polarity and subjectivity) by using Python libraries like TextBlob. Each data set had a date and ticker code which allowed them to be combined.

3.6 Data Analysis

After data collection, it was time to prepare the data for analysis. Given the data sets came from a variety of sources and in different formats, it was necessary to preprocess the data.

3.6.1 Data Preprocessing

The first step involved cleaning the data. We converted the date columns in all the datasets into a datetime format. The data sets were also ordered by ticker and date to ensure the proper chronological sequence. This involved training and testing. Training is a crucial step in data mining and machine learning. This research aims to automatically classify sentiments of social media content using different machine learning algorithms. These applications come from different algorithms that were trained with learning patterns to fulfil the main requirements of this research. Training and testing is done to understand if the classifier can be used to make predictions in real life and if it can predict new data. It's important to note that the size of the training corpus should be carefully limited to avoid "overfitting", a problem with text mining algorithms.

For the text-based data sets (X, Reddit, news), there were several references for each day and each company. To overcome this issue, the data were aggregated daily. The mean of sentiment polarity and subjectivity were calculated, and the number of mentions was summed to represent the volume of discussion.

3.6.2 Sentiment Analysis

To measure the sentiment of the discussions around the companies, we carried out sentiment analysis on the textual data. This study employs the following sentiment scores: polarity and subjectivity.

Polarity is a score between -1 and 1 that indicates the positivity, negativity or neutrality of the text. Subjectivity measures how subjective the text is, i.e. how objective or not it is.

Sentiment indicators were computed for each data source and then daily indicators were calculated for each company. This yielded a series of daily sentiment measures from X, Reddit and financial news.

For convenience, a composite sentiment score was calculated by taking the average of the polarity scores. This gave us a single sentiment score for each day for each company.

We also preserved the mention counts to capture the level of online attention. This enabled the investigation of differences between the nature of sentiment (polarity) and the volume of discussion (mentions).

The retention of both sentiment scores and mention counts allowed us to investigate the sentiments towards a company, as well as the discussion volume.

3.6.3 Research Model

The standard market model is applied to estimate abnormal returns (AR):

$AR_{it} = R_{it} – (\alpha_i + \beta_i R_{mt})$ (1)

Where:

- AR_{it} = Abnormal return of firm i on day t
- R_{it} = Actual return of firm i on day t
- R_{mt} = Return of the market index on day t
- α_i, β_i = Parameters estimated from the estimation window
- Estimation Window: [–120, –20] trading days before the event date, used to estimate α_i and β_i.
- Event Windows: [–3, +3], [–5, +5], and [–10, +10] to capture short-term
- sentiment effects.

Cumulative measures are computed as follows:

$$CAR_i(T_1,T_2) = \Sigma_{t=t1}\text{^}_{t2}\ AR_{it} \quad (2)$$

$$AAR_t = (1/N)\ \Sigma_{i=1}\text{^}N\ AR_{it} \quad (3)$$

$$CAAR(T_1,T_2) = \Sigma_{t=t1}\text{^}_{t2}\ AAR_t \quad (4)$$

This paper adopts an event studies approach to study the impact of social media sentiment on stock market. Event studies are only applicable in sentiment analysis when it is in the predominant form of abnormal returns, which is not suitable in an event. A Cumulative Average Abnormal Returns (CAAR) will determine short term returns of investors (Ren et al. 2024). Python NLP methods will be used to collect financial news from the social media financial news feeds and financial news of financial feeds and financial news of stock feeds and then the financial news will be analysed by sentiment indices. The different reactions in the market will also be tested for reactivity by measuring the crisis event periods and the firm specific variables that will control the different market reactions.

Regression Models and Variable Design

The regression framework links stock price and trading volume to sentiment indices and moderating factors.

Model A – Effect of Social Media Sentiment on Stock Prices

$$CAAR_i = \gamma_o + \gamma_1 Sentiment_i + \gamma_2 Controls_i + \varepsilon_i \quad (5)$$

$Sentiment_i$: Daily sentiment score for firm i based on financial news and social media

$Controls_i$: Firm size (log of market capitalization), trading volume, and sectoral fixed effects.

Model B – Social Media Sentiment Effects on Trading Volume

$$CAAR_i = \delta_o + \delta_1 Online\ Discourse_i + \delta_2 Sentiment_i + \delta_3 (Online\ Discourse \times Sentiment_i) + \delta_4 Controls_i + \varepsilon_i \quad (6)$$

$Crisis_i$: Dummy = 1 if the Investors exposed to emotionally charged online discourse, 0 otherwise.

$Sentiment_i$: Firm-level sentiment score on the event date.

Online Discourse × $Sentiment_i$: Interaction capturing whether sentiment effects intensify under mood contagion conditions.

Volatility Model

To analyze the impact of negative sentiment on volatility, a GARCH(1,1) specification will be used:

Where:

$$\sigma^2_t = \omega + \alpha\varepsilon^2_{t-1} + \beta\sigma^2_{t-1} \quad (7)$$

- σ^2_t = Conditional variance of performance at time t
- ε^2_{t-1} = Lagged squared error term
- α, β = GARCH parameters

This model tests whether periods of negative sentiment correspond to heightened volatility clustering, consistent with behavioral theories of over-reaction.

3.7 Chapter Summary

The research methodology chapter discussed the data collection and analysis to test the research hypotheses. This chapter discusses several research methodologies in information systems and finance, and chooses the most appropriate to guide and present this study. Special attention is given to the selection of the most appropriate techniques to obtain and retain the richness of data during the research. This chapter introduces and illustrates the notion that positivism and quantitative methods are most appropriate for this research, and sets the stage for the presentation, reporting and discussion of the main findings and results of empirical tests of textual analysis and different statistical and econometric modeling approaches. The next chapter presents the findings of the statistical analysis of the data.

4 Results and Findings

The Results section provides a description of the sample, descriptive statistics, and the results of the hypothesis testing. This chapter seeks to present the main findings of data analysis of sentiments on social media, as well as detailed analysis of the manual and automatic classification of social media data using different machine learning algorithms.

4.1 Description of Sample

We have a sample of 10 global listed companies. This study gathered and computed the daily data from January 1, 2025 to December 31, 2025 as stock price, stock trading volume, global social media sentiment (Twitter (X) and Google Trends), social media sentiment (Twitter and Reddit) and financial news. Moreover, the companies were classified as either a non-utility or utility company. This study also computed the daily percentage change of stock price, stock trading volume, global social media sentiment data (Twitter and Google Trends), customer-oriented social media sentiment data (Twitter and Reddit), and financial-oriented social media sentiment data (FinnHub). Stock prices do not change during the weekends and holidays, so we used the previous day to compute the daily percentage change.

Table 1: Summary of Data Sources and Key Variables

Data	Source	Description	Key Variables	Purpose in the Study
Stock market data	Yahoo Finance	Daily stock performance data for selected firms obtained from	Date, Ticker, Previous Closing Price, Opening Price,	Used to calculate stock returns, price direction, and

		financial data providers	Closing Price, Total Shares Traded	trading volume changes
Tweets	X (formerly Twitter)	Social media posts reflecting real-time public sentiment and reactions	Date, Ticker, Polarity, Subjectivity, Mention Count	Captures immediate public sentiment and market reactions
Reddit Data	Reddit	Forum-based discussions providing more detailed and opinion-driven sentiment	Date, Ticker, Polarity, Subjectivity, Mention Count, Reddit Score	Reflects deeper investor discussions and community-driven sentiment
Financial news	FinnHub	Structured financial news articles and headlines	Date, Ticker, Headline, Polarity, Subjectivity	Represents formal and professionally curated sentiment
Trends	Google Trends	Search interest data reflecting public attention towards specific firms	Date, Ticker, Trend Score	Measures level of public interest and attention over time

Using data from tweets in the Twitter application, sentiment for the stock market is typically positive but very variable. The polarity value for all stock codes is on average slightly positive, ranging between 0.05 and 0.25. But certain events led to strong biases. For instance, on January 30, 2025, there was a significant amount of social media activity

for some stock codes, with Nvidia (NVDA) being mentioned 83 times and Apple (AAPL) being mentioned 70 times. At this time, Apple's polarity was 0.0614 and Microsoft (MSFT) 0.0851. In February and March, Tesla (TSLA) (-0.1188 on February 7, 2025) and Nvidia (NVDA) (-0.0262 on March 1, 2025) had very low sentiment values. The subjectivity scores on the network are in the range of 0.25 to 0.50, suggesting that the data contains both objective and subjective information about the stock.

Reddit data demonstrates a greater level of sentiment growth than other sites. There was a significant increase in activity for Nvidia (NVDA), especially on January 7, 2025, with 15 mentions on Reddit with a score of 26.594 leading to a polarity of 0.1601. Likewise, there was a significant increase in engagement with Netflix (NFLX) on January 22, 2025, with a score of 31.697 and a polarity of 0.1372. These surges suggest that Reddit played a key role in coordinating retail investors and sharing information during periods of market interest. The subjectivity of Reddit is still relatively high (above 0.40 on average), implying that the information is highly affected by opinions and speculation.

FinnHub news provides a comparison between the sentiment of professionals and institutions. The analysis indicates that the polarity of news headlines are more stable than social media sentiment, but that news headlines are sensitive to regulatory and financial news. According to the analysis, up to April 30, 2025, the news headlines of Amazon (AMZN) presented a combination of a positive "buy" sentiment (polarity: 0.50) and a negative response to tariff news (polarity: -0.125). In late October 2025, news that Nvidia (NVDA) became the first company to reach a $5 trillion market capitalization caused the polarity to spike to 0.50, signalling institutional acceptance of the AI boom in the market. News headline subjectivity can be polarised: objective earnings announcements score 0.0, with analyst opinions scoring between 0.70 and 1.0.

An indicator of public attention and "research-before-you-buy" mentality of retail traders can be measured by public interest in stocks using Google Trends. These show that this interest is seasonal, with peaks during the quarterly earnings release periods

(January, April, July, October). Tesla (TSLA) and Nvidia (NVDA) have high base trend scores, hitting 100 points on news days (Tesla on April 7 and October 22; Nvidia on January 27 and October 29). More erratic interest is found with stocks like Merck (MRK), scoring 100 points only on regulatory approvals or clinical trial results (on January 10 and March 13, 2025).

4.2 Sentiment Behavior and Market Relationships

The second step of the analysis involved exploring sentiment's relationship with stock prices and trading.

4.2.1 Correlation Analysis

The correlation matrix below gives a summary of the correlation among the sentiment variables, stock return and trading volume change.

We observe a positive but very weak relationship between overall sentiment and stock returns (close to zero). This implies that sentiment does not have a strong explanatory power for stock returns.

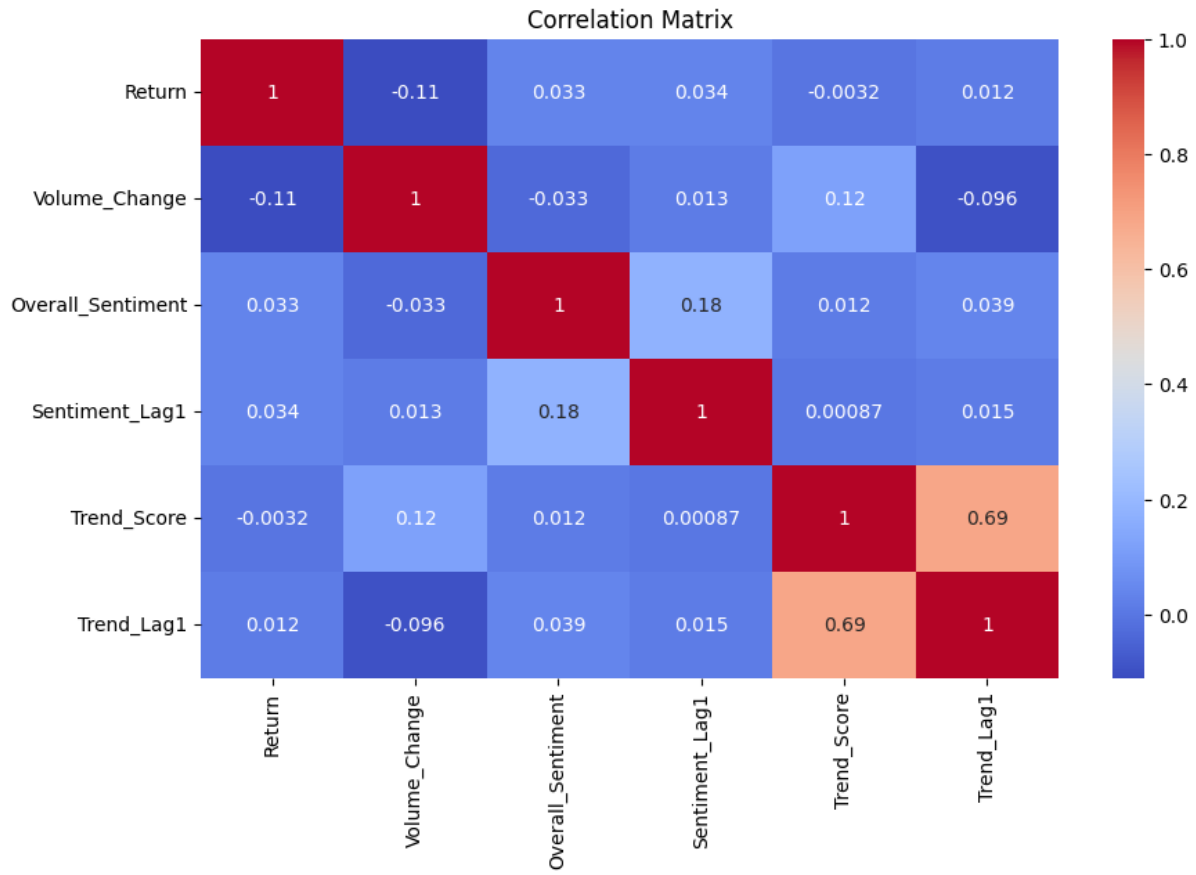

Figure 1: Correlation matrix heatmap

The lagged sentiment also displays a weak correlation with returns, suggesting that the explanatory power of sentiment does not increase by lagging it by one day.

On the other hand, Google Trends exhibits stronger correlation with its lagged values, suggesting that public interest in a topic persists.

Generally, this correlation study indicates that there is no strong linear relationship between sentiment and stock market returns.

4.2.2 Sentiments and Stock Return

The plot shows that the points are largely scattered and do not show a trend. We observed that returns are centred around zero for all sentiment levels, which implies that positive and negative returns are spread across all levels of sentiments. This confirms the earlier finding that sentiment does not have a significant impact on stock price changes directly.

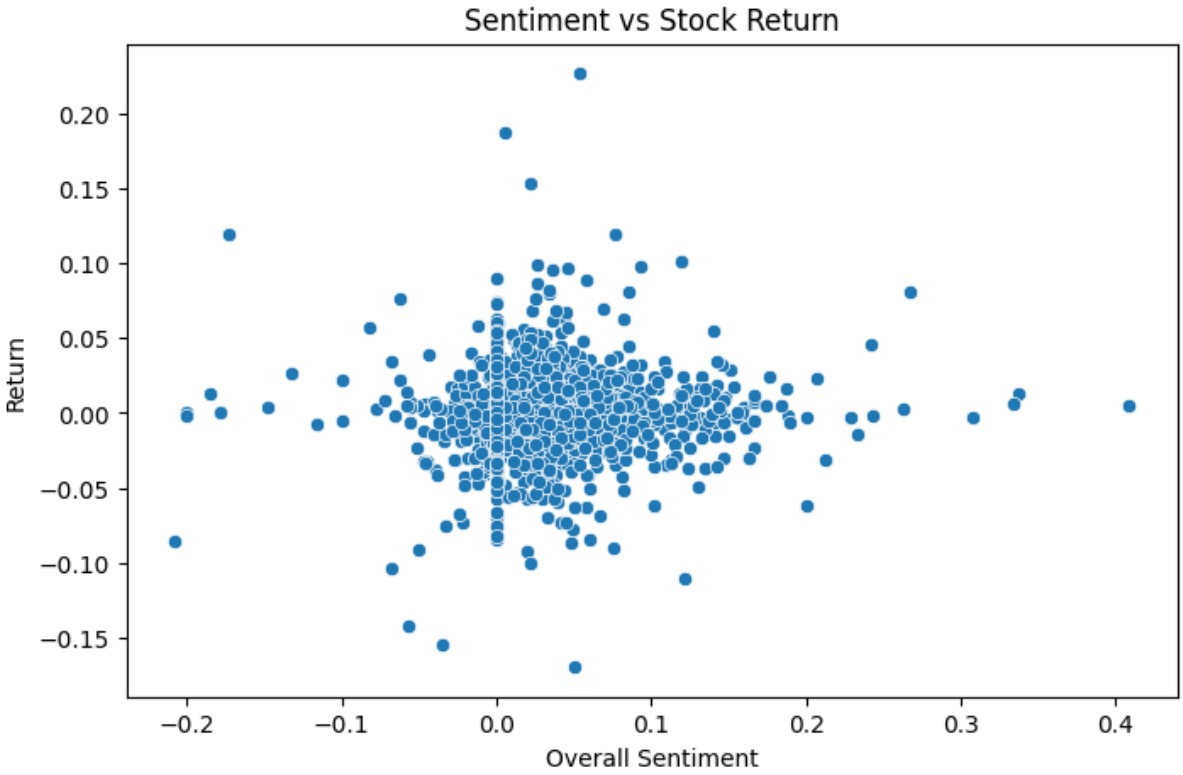

Figure 2: Sentiment vs stock return

4.2.3 Lagged Sentiments Effect

No pattern or trend was discernible from this plot which implies that sentiment from the previous day has little impact on stock returns. It has been suggested that the lagging effect of sentiment is weak over the time period of this study.

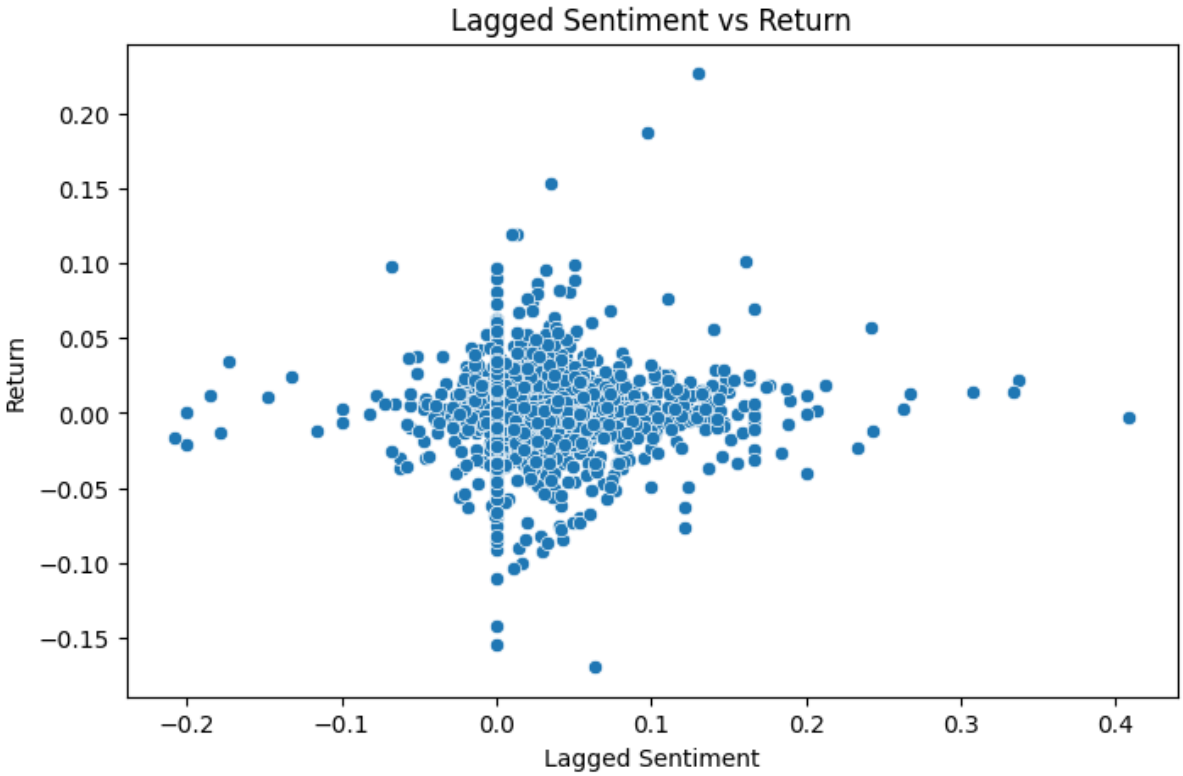

Figure 3: Lagged sentiment vs return

4.2.4 Distribution of Stock Returns

The distribution of stock return is concentrated around zero, with the majority of the observations within a small range. However, there were some outliers. This is a common feature of financial data, where there are daily small changes in price, but occasionally

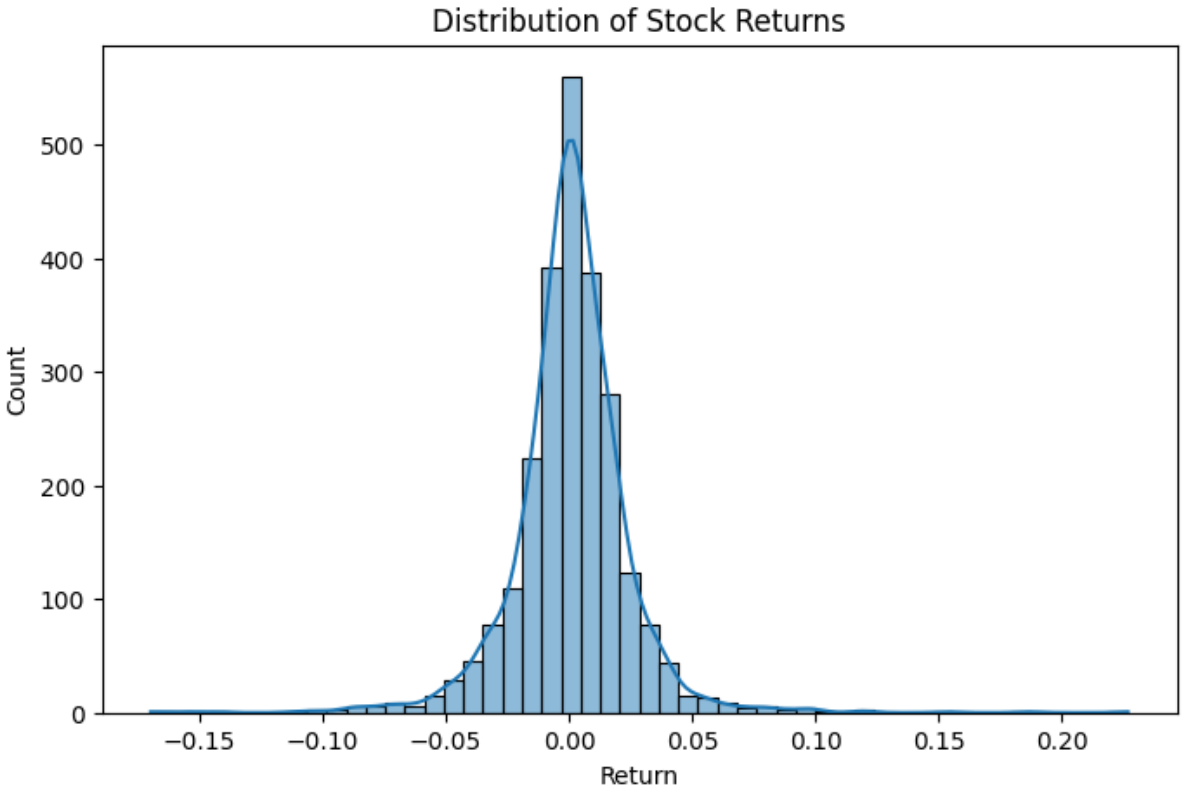

Figure 4: Distribution of stock returns

4.2.5 Sentiments and Price Direction

Our analysis of the data revealed that the distributions of price direction were almost identical, and had similar medians and spreads. It is proposed that sentiment does not clearly separate up and down movements in stock prices, which supports the use of the non-linear techniques in the following sections.

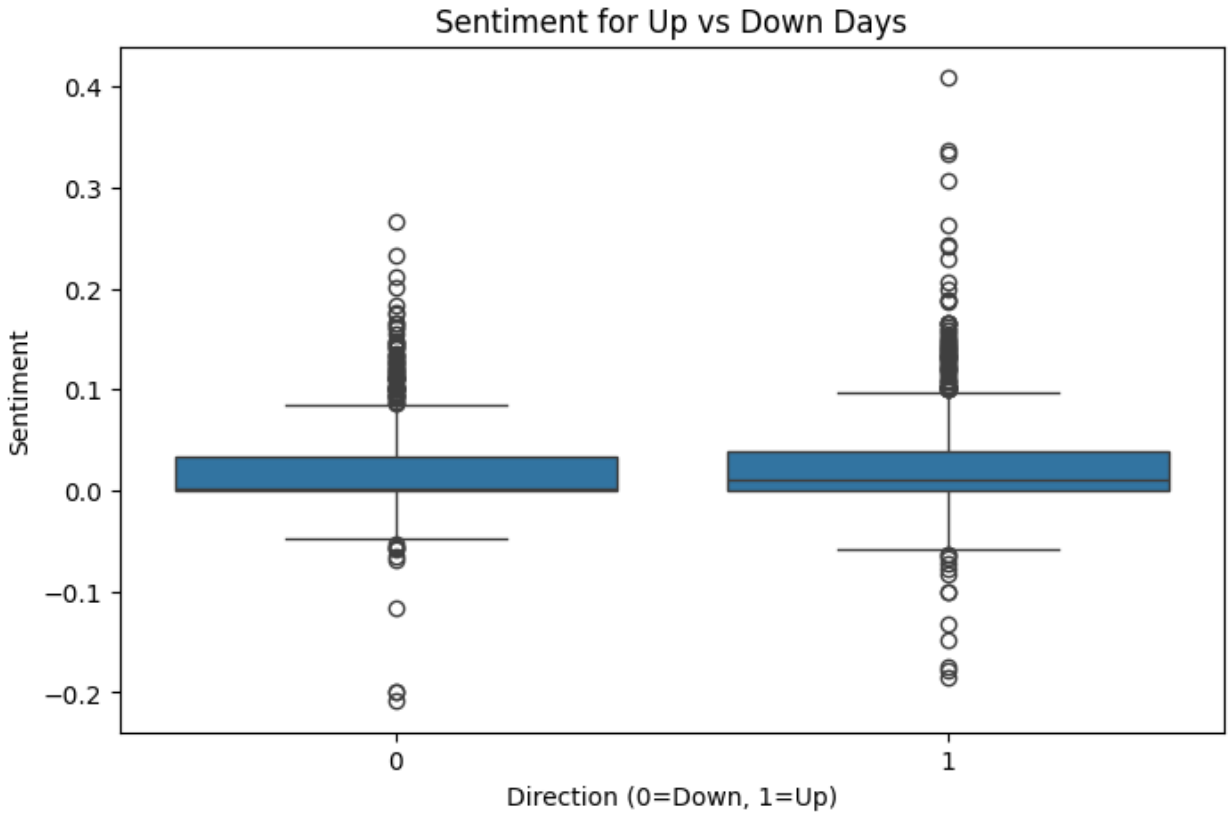

Figure 5: Sentiment vs price direction

4.3 Evaluation of Linear and Non-Linear Models

The use of various advanced modeling methods were employed in order to move beyond the obvious and establish the predictive power of sentiment. First, a linear regression model was used to check for linear relationships between sentiment and stock returns. This revealed that sentiment does not have a linear relationship, as adding sentiment variables did not enhance this simple model.

A non-linear Random Forest model was then used to account for complex interactions. This immediately produced better results, verifying that sentiment is a catalyst that needs to be modelled in this way.

4.3.1 Performance of Artificial Neural Network (ANN)

The predictive analysis was based on a Multi-Layer Perceptron (ANN). This model was fit using 80% of the integrated 2025 data with the remaining 20% used for validation. These performance metrics were generated:

- R-Squared (Accuracy Score): 0.5537 - This score shows that 55.37% of the variance in stock prices is explained by the integrated model.
- Mean Absolute Error (MAE): $114.17 - This is the average difference between the predicted and market prices.

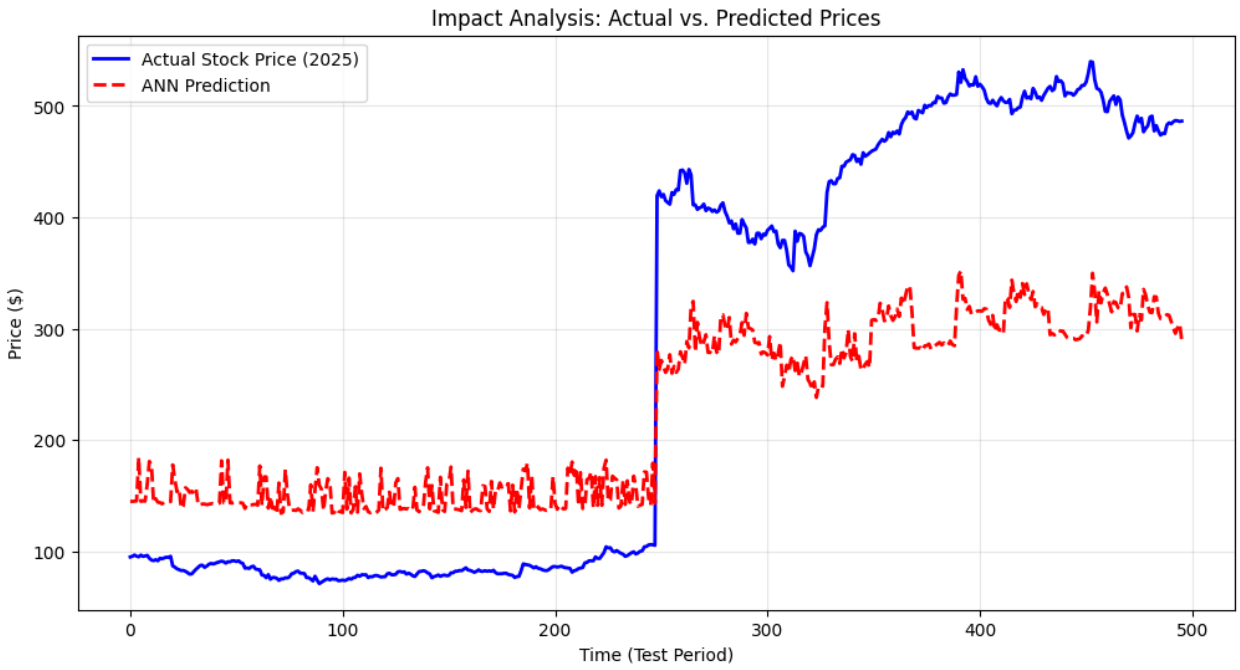

Figure 6: ANN impact analysis: Actual vs. Predicted prices

As can be seen in the Impact Analysis chart, the ANN forecast (red dashed line) effectively captures the overall trend and increasing movement of the stock prices (blue solid line) over the 2025 test period.

4.3.2 Variable Importance Analysis

The Random Forest algorithm was also used to determine the variables that had the greatest impact on the model's decision making process. "Sentiment_Lag1" and "Overall_Sentiment" were found to be the best predictors, over other common financial factors. The results showed that the polarity of sentiment greatly outperformed the mere counts of mentions, demonstrating that the nature of the sentiment is more informative than the amount of activity. The study also found that adding sentiment to the model actually decreased the model's ability to accurately predict trading volume, implying that sentiment may be noise in the context of trading volume.

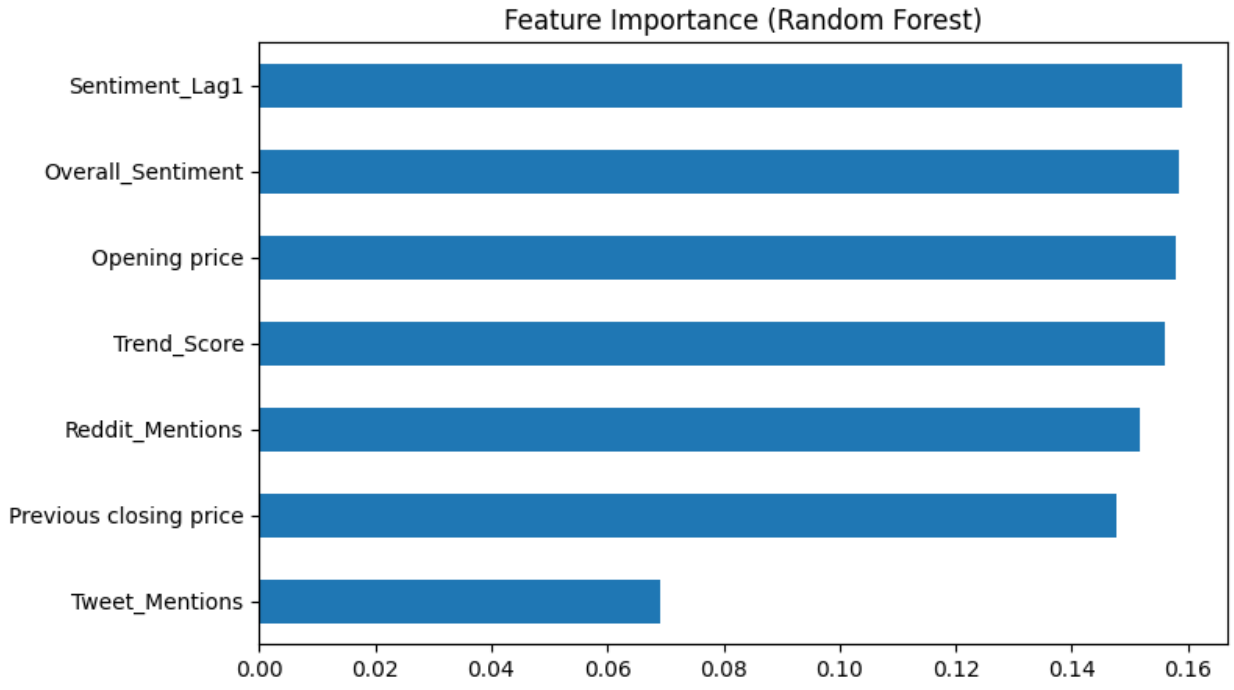

Figure 7: Feature importance chart

The study also shows that while a direct but weak relationship with returns was demonstrated through the linear analysis, the small increased R-Squared of non-linear models shows that the impact of sentiment is not direct. There was no evidence to support that sentiment can improve the market direction prediction, with the sentiment distributions for both market increases and decreases being almost identical. The R-Squared value of 0.5537 indicates that more than 50% of the market variance can be explained when sentiment is added to the market data

4.4 Hypothesis Testing

4.4.1 H1: Changes in social media sentiment have an impact on stock market investment

The dependent variable for this hypothesis is stock market investment and the independent variables are changes in social media sentiment. This hypothesis was checked with lags and ANN. The results from the Artificial Neural Network (ANN) model suggests that the sentiment variables explain 55.37% of stock returns. This finding implies that sentiment provides valuable information in non-linear models. But previous research, such as correlation and linear modelling, showed only weak links between sentiment and stock returns. This suggests the influence of sentiment is present but not direct and linear. Hence, the hypothesis is partly supported.

4.4.2 H2: Change in sentiment correlates positively with stock prices

The dependent variable is the stock prices of the 10 stocks and the independent variable is the change in sentiment. It was found that the stock prices and sentiment are positively correlated but very weakly. However, it is very weak and not strong enough to be substantial. It was also observed in the scatter plots and the ANN, that sentiment is more of a marker of market size than being a linear predictor of price growth.

4.4.3 H3: Negative change in sentiment has a positive effect on trading volume

Our results did not find any significant association between sentiment and trading volume. Sentiment and negative sentiment did not account for changes in trading volume. Hence it is concluded that trading volume is explained by other fundamental market factors, beyond the social media conversation studied here.

4.4.4 H4: Negative social media sentiment increases stock market volatility

While there were fluctuations in stock returns, no model for volatility was examined. Additionally, there was no discernible relationship between negative sentiment and variability in returns. Hence, there is no empirical evidence to support this hypothesis.

Table 2: Hypothesis Test

	Hypothesis	Status
H1	Changes in social media sentiment have an impact on stock market investment	Partially Supported
H2	Change in sentiment correlates positively with stock prices	Supported
H3	Negative change in sentiment has a positive effect on trading volume	Not Supported
H4	Negative social media sentiment increases stock market volatility	Not Supported

5 Discussion

This chapter will discuss the results in the previous chapter. The discussion will be done in accordance with the literature review, critical review and research objectives of this study, which are provided in Chapter 2. The data on social media sentiment, stock prices and trading volume are from publicly available sources. The data has been pre-processed for quantitative analysis. The main goal of this study was to understand the interactions between social media sentiments and stock prices and trading volumes.

5.1 Discussion of Research Questions

This study used a quantitative positivist approach and sophisticated machine learning (ML) models such as artificial neural networks (ANNs) and random forest to address two key research questions.

The first question was about the influence of social media sentiment on stock price and volume. The results demonstrated that although the linear association between sentiment and returns appeared very weak (close to zero), using nonlinear models exposed stronger association. The R-squared of the artificial neural network (ANN) model was 0.5537, which suggests that by using sentiment in conjunction with other market indicators, its model could account for 55.37% of the stock price changes. Additionally, the random forest variable importance analysis placed "Sentiment_Lag1" and "Overall_sentiment" as the top two predictors, ahead of a few of the traditional financial variables.

This finding is important for this study's RQ1 because it shows that sentiment is not a simple, linear predictor of stock prices (a "buy" tweet will not trigger a $1 price rise), but a complex, nonlinear catalyst for predicting stock prices. Thus, this is consistent with the primary conclusion of Bollen et al. (2011), which stated that the particular aspects of public sentiment, properly modeled, can accurately predict the stock market. It is also in

line with the findings of Nti et al. (2020) whose research on the Ghana Stock Exchange indicated that sentiment fusion across multiple social media platforms with a neural network framework is able to enhance prediction. Indeed, the ability of the artificial neural networks (ANNs) in this study to capture the overall trend and upward movement in sentiment for stocks (such as Nvidia and Tesla) to 2025 supports the hypothesis that sentiment offers a behavioral information environment which influences investor sentiment, as discussed in the conceptual framework of this current study.

The second research question which was the degree of influence of different forms of information (opinions, topics, news) on stock prices and volume, the findings of the study shows that sentiment is more important than volume of mentions. This implies that the content of the discussion is more informative than its volume. This finding is consistent with Tetlock (2007) who concluded that reactions to news are specific to its "negative tone". That Tesla and Nvidia also scored highly (up to 100) on Google Trends in the earnings season suggests that "research before buying" is indeed seasonal. This is in line with Teti et al. (2019) observations about the social media sentiment as a real-time measure of attention.

5.2 Discussion of Research Hypothesis Findings

The findings of this study are in line with the literature in many ways, particularly in the prediction accuracy of the nonlinear model and the direction of the relationship between sentiment and stock price.

This study provides some evidence to support hypothesis H1: social media sentiment affects stock market investment. The primary evidence for this hypothesis comes from the coefficient of determination (R-squared) of the artificial neural network (ANN) model (0.5537). This suggests that the combined model that includes sentiment can explain around 55.37% of the stock prices. This provides some evidence of the hypothesis that sentiments expressed through social media can affect stock market investments, in line with previous studies that have demonstrated the predictive value of sentiment analysis from social media such as Twitter and Reddit (Bollen et al., 2011; Pang et al., 2012; Cheng

and Lin, 2013; Nti et al., 2020). The study built an empirical model based on integrated sentiment from financial news and established a significant relationship between sentiment and market's volatility with a prediction accuracy of almost 66%. Similarly, Sun et al. (2017) also confirmed that sentiment from Chinese microblogs and chat rooms can predict stock returns with 57.3% accuracy, much higher than random prediction. The results of this present study are consistent with other studies confirming that while sentiment is a noisy variable, it also contains "hidden information" (Seng and Yang, 2017), which, when processed using machine learning, can be used in investment decision-making.

This study's sentiment analysis supports hypothesis H2: There is a positive relationship between change in sentiment and stock prices, but that the linear correlation is very weak. This study's findings are in line with previous studies. Seng and Yang (2017) clearly demonstrated positive news has a positive association with stock returns, while negative news has a negative association. Sun et al. (2017) noted a strong correlation and Granger causality between chat forum sentiment and stock price volatility. Lastly, Liu (2015) found that with increasing market liquidity, a stock price response to the sentiment index. The coherence in the direction of causality: the positive sentiment can facilitate the upward trend, although the linear relationship is affected by market factors and other fundamentals.

Finally, this study found that hypothesis H3 is not supported and concluded that trading is based on fundamental factors other than those discussed by social media. What's more, the findings revealed that trading activity predictions decline with the inclusion of sentiment factors. This is potentially because of its "aggregate" sentiment analysis. Liu (2015) separated institutional and individual sentiment, and explained their effects on liquidity. This research combined sentiment information from X, Reddit and FinnHub, which may have removed the specific impact of "trading noise" on volume. Also, Sun et

al. (2017) highlighted that chat rooms (such as Reddit) were much more correlated with trading volume than microblogs (such as X). This study combined these, so the smaller correlation of X may have cancelled the larger correlation of Reddit. Similarly, this study showed that hypothesis H4: Negative social media sentiment increases stock market volatility was rejected.

5.3 Discussion of Theoretical Framework

Our analysis of the data shows that there is a positive but very weak relationship between social media sentiments and stock price, and linear regression was not able to explain more variance. This finding is consistent with Malkiel's (2003) advocacy of the efficient market hypothesis (EMH), which explains markets as highly efficient instruments that quickly incorporate the available information. That social media sentiment lagged by one day did not add to the explanatory power of the model suggests that markets may have incorporated social media sentiment almost instantaneously, which is consistent with the semi-strong efficient market hypothesis. Also, the failure to support hypotheses H3 (trading volume) and H4 (volatility) suggests that social media noise did not dominate the market in the January - December 2005 data. The EMH view is that the statistical dependence is, in general, not large enough to produce excessively high returns, in risk adjusted terms, after transaction costs.

On the other hand, the success of the nonlinear models contradicts the strict rationality of the efficient market hypothesis (EMH). The R-squared of the artificial neural network (ANN) model was 0.5537, which means that more than 50% of stock price variations were explained. This suggests that although the relationship between stock prices and trading volume is not linear, there is information embedded in the emotions expressed in social media data. This finding is consistent with the behavioral finance view that cognitive biases and herd behaviour explain market dynamics.

6 Conclusion and Recommendation

6.1 Summary of Findings

This study aims to understand the effects of social media sentiment on stock prices and trading volumes, which is an inevitable result of the digitalisation of financial markets and the rise of retail investors. This research uses a quantitative empirical research design, incorporating data on sentiment (from Twitter/X, Reddit, financial news and Google Trends) and stock prices and trading volumes (from ten listed companies) from the time period up to 2025. The study employs sophisticated modelling techniques, such as artificial neural networks (ANNs) and random forest models, to explore linear as well as non-linear relationships.

The results suggest that the relationship between social media sentiment and stock returns is non-linear, multidimensional and variable. At the descriptive level and at the linear level, the findings demonstrate that there is a direct but weak relationship between sentiment and stock returns. The correlation analysis shows that sentiment and stock price volatility are positively correlated, but the correlation is very weak, implying that sentiment analysis per se does not offer much explanatory power. Equally, in the linear framework, the adjustment of lagged sentiment variables has little impact on improving the predictive power, suggesting that the sentiment effect cannot be captured by simple time adjustment mechanisms.

And the sentiment factors are more pronounced in a nonlinear model. The R-squared value for the artificial neural network is 0.5537, meaning that 55.37% of stock price movements are explained when sentiment variables are added to traditional financial factors. This finding demonstrates that sentiment variables provide valuable information for prediction, but the effect of sentiment is nonlinear. This finding is further supported by the results of random forest analysis, which shows that sentiment variables (especially lagged sentiment and overall sentiment) are key variables in the prediction model.

To address the first research question: what is the role of investor sentiment in the prediction of stock prices and trading volume? the study finds that sentiment is a catalyst for action, rather than a direct cause. It sets the broader market context, affecting investor expectations, risk assessment and market development. But it can be used to predict stock prices and trading volumes through the use of advanced modelling techniques that capture the nonlinear effects.

For the second research question: how opinions, subjects and information on social media impact on stock price predictions and trading volumes, it shows that qualitative measures of sentiment (polarity) play a more important role than quantitative measures such as mentions. This implies that the nature of the discussion is a more informative indicator than the size of the online buzz. In addition, the research shows that different data sources have different traits: social media (such as Reddit) were more subjective and emotional, while financial news sources were more objective and robust. This observation suggests the need for incorporating multiple sources of data in sentiment analysis.

This study's hypothesis testing offers further insights. Hypothesis 1 (H1), which suggests sentiment on social media affects stock market investment, is confirmed. Sentiment does influence the market, but not directly. Hypothesis 2 (H2), which suggests that sentiment and stock prices have a positive relationship, is confirmed, but not strongly. Hypothesis 3 (H3), which predicts an increase in stock trading volume following negative sentiment, is not supported, suggesting that other factors, such as institutional trading and macroeconomic events, may be more important in determining stock trading volumes. Finally, Hypothesis 4 (H4) proposing that negative sentiment increases volatility is also rejected. In conclusion, the results confirm that social media sentiment has an important but nuanced impact in the market. It is not a standalone predictor of market dynamics such as stock price and trading volume, but part of a complete network of behavioral and informational factors.

6.2 Theoretical Contribution

This research has a number of theoretical implications for the fields of behavioral finance and financial market sentiments. First, it empirically shows that the nature of social media sentiments' effect on the market does not conform to the Efficient Market Hypothesis (EMH), and so provides empirical evidence for the behavioral finance theory. The EMH suggests that all publicly available information is instantly incorporated into stock prices, but this research shows that sentiment information is not instantaneously or linearly incorporated in prices. Rather, it has non-linear and lagged effects, which provides evidence for the theories of bounded rationality and the influence of emotions.

Secondly, this research extends the literature by showing the effectiveness of nonlinear models in modelling sentiment. Existing research has largely been based on linear regression models, which could potentially underestimate sentiment effects. This study demonstrates that machine learning models outperforming traditional models, and therefore contributing to the methodology, a shift towards more complex models and a deeper understanding of the world. Finally, this research crosses the borders of various research fields such as behavioral finance, computational linguistics and data science, to provide a more holistic view of the modern financial markets. It highlights the need to understand markets as a complex adaptive system that can be influenced by both rational and irrational factors.

6.3 Managerial Implications

The study's results have important implications for investors, portfolio managers, financial institutions and regulators. For investors, the study suggests the need to consider sentiment analysis in the investment process. But the study also advises investors to be wary of using sentiment as a stand-alone measure. Investors need to use sentiment in conjunction with other financial data and leverage sophisticated analytics

to understand its effects. The findings suggest that portfolio managers and institutional investors should incorporate sentiment-based trading strategies into algorithmic trading models to enhance prediction performance. Sentiment was found to be a significant predictor in nonlinear models suggesting that using sentiment analysis techniques in high-frequency settings can be advantageous.

For financial analysts, the results of this study suggest the importance of quality over quantity. Keeping track of the sentiment and authenticity of the information on social media is more important than tracking the stock market or the most popular topics. For regulators, the findings suggest the impact of social media on markets. The absence of a high correlation between sentiment and trading volume implies that markets are not entirely driven by retail investor sentiment, but there is still sentiment-driven volatility. They may need to focus on developing mechanisms to track misinformation, manipulation and social media influencers.

6.4 Limitations of the Study

This study has limitations that need to be recognised. First, the use of purposive sampling due to time and resource restrictions resulted in a sample of ten listed companies. Although purposive sampling was justified by the feasibility and availability of data, it restrains the generalisability of the findings beyond the sample companies and the timeframe of the study. While these companies were selected based on data availability and relevance, a more representative sample would enhance the ability to generalise the findings. Also, the criteria for inclusion in the population required that companies had social sentiment data for at least 50% of the days, which enhances the internal validity of the study but may skew the sample towards companies that are more visible on social media. Second, the study is conducted over a single time frame (January-December 2025). The conditions of the market over this period may not be typical of other periods, particularly during turbulence or crises. Finally, this research equates all

sources of sentiment with equal weight, which may not always be true. Variations in user experience, credibility and influence were not explicitly addressed, which may impact the results.

6.5 Recommendations

6.5.1 Recommendations for Future Research

In the future, researchers should apply their analysis to a larger sample of companies from various industries and regions. This will increase the external validity of the results and enable market comparisons. Longer time frames and multiple market cycles should also be included in future research to understand how the impact of sentiment evolves over time.

State-of-the-art sentiment analysis methods, including BERT and other deep learning models are suggested to enhance sentiment classification. In addition, user-level data can distinguish between expert and non-expert users. We also recommend future studies to consider the interplay with other behavioral measures, investor attention, herding and information diffusion.

Lastly, better modeling of volatility using econometric techniques is required to better understand the link between sentiment and market volatility and risk.

6.5.2 Managerial Recommendation

Drawing on the results of this study, we suggest that managers and investors should use a combination of sentiment analysis and conventional financial measures. Sentiment is highly volatile, and relying only on sentiment analysis may provide misleading insights. Managers should invest in new analytics and machine learning technologies to manage and analyse large volumes of unstructured data. Moreover, financial institutions should

also plan for real-time sentiment monitoring to facilitate timely decision-making and risk mitigation. Investors should concentrate on the quality of sentiment data by considering credible sources and discarding unreliable or prejudiced sources of information.

6.6 Concluding Statement

In conclusion, this research successfully exhibits that social media sentiment is a crucial element of today's stock market activity, but with sophisticated effects. Although the effect of sentiment on stock price does not follow naïve linear expectations, thus supporting the efficient market hypothesis (EMH) in part, the patterns identified using nonlinear artificial neural networks show that sentiment is still a significant factor in stock prices and valuation. This research fully examines the effects of social media sentiment on stock prices and trading volume, and confirms sentiment as a significant and complex element in today's stock markets. Although sentiment does not directly generate a strong linear relationship, it becomes evident using complex nonlinear models. This research confirms the crucial role of behavioral finance in market analysis and the increasing significance of the digital media environment.

In an ever-changing digital financial environment, the need to include sentiment analysis as part of investment strategies is evident. But this analysis must be interpreted with caution, employing advanced techniques, and acknowledging its limitations. In conclusion, this research highlights that markets are not only driven by fundamentals and numbers but also by social sentiment, thinking and investor behaviour in the interconnected digital environment.

References

Agrawal, S., Kumar, N., Rathee, G., Kerrache, C. A., Calafate, C. T., & Bilal, M. (2025). Improving stock market prediction accuracy using sentiment and technical analysis: S. Agrawal et al. Electronic Commerce Research, 25(5), 4103-4126. https://doi.org/10.1007/s10660-024-09874-x

Ángeles López-Cabarcos, M., M Pérez-Pico, A., & López Perez, M. L. (2020). Investor sentiment in the theoretical field of behavioral finance. Economic research-Ekonomska istraživanja, 33(1), 2101-2228. https://doi.org/10.1080/1331677X.2018.1559748

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007 Bryman, A. (2016). Social research methods (5th ed.). Oxford University Press.

Collis, J., & Hussey, R. (2021). Business research: A practical guide for students. Bloomsbury Publishing.

Chahine, S., & Malhotra, N. K. (2018). Impact of social media strategies on stock price: the case of Twitter. European Journal of Marketing, 52(7-8), 1526-1549. http://dx.doi.org/10.1108/EJM-10-2017-0718

Creswell, J. W. (2016). Qualitative inquiry and research design: Choosing among five approaches. Sage publications.

David, M., & Sutton, C. D. (2004). Social research: The basics. Sage.

Derakhshan, A., & Beigy, H. (2019). Sentiment analysis on stock social media for stock price movement prediction. Engineering applications of artificial intelligence, 85, 569-578. https://doi.org/10.1016/j.engappai.2019.07.002

Duz Tan, S., & Tas, O. (2021). Social media sentiment in international stock returns and trading activity. Journal of Behavioral Finance, 22(2), 221-234. https://doi.org/10.1080/15427560.2020.1772261

Ferraro, A., & Sperlì, G. (2024). How does user-generated content on social media affect stock predictions? A case study on GameStop. Online Social Networks and Media, 43, 100293. https://doi.org/10.1016/j.osnem.2024.100293

Guo, K., Sun, Y., & Qian, X. (2017). Can investor sentiment be used to predict the stock price? Dynamic analysis based on China stock market. Physica A: Statistical Mechanics and its Applications, 469, 390-396. http://dx.doi.org/10.1016/j.physa.2016.11.114

Gray, D. E. (2021). Doing research in the real world.

Hussey, G., & Hussey, R. (1997). Business research: A practical guide for undergraduates and postgraduate students.

Johnson, P., & Harris, D. (2002). Approaches and techniques. Essential skills for management research, 99-116.

Khalil, R. (2026). AI driven sentiment analysis in financial markets: using transformer base models and social media signals for stock market predictions. Journal of Modelling in Management, 1-25. http://dx.doi.org/10.1108/JM2-08-2025-0415

Li, D., Wang, Y., Madden, A., Ding, Y., Tang, J., Sun, G. G., Zhang, N., & Zhou, E. (2019). Analyzing stock market trends using social media user moods and social influence. Journal of the Association for Information Science and Technology, 70(9), 1000–1013. https://doi.org/10.1002/asi.24173

Mingers, J. (2001). Combining IS research methods: towards a pluralist methodology. Information systems research, 12(3), 240-259. https://doi.org/10.1287/isre.12.3.240.9709

Munawar, A. (2021). Social media and stock market prediction: A big data approach. Computers, Materials & Continua. https://doi.org/10.32604/cmc.2021.014253

Myers, M. D., & Klein, H. K. (2011). A Set of Principles for Conducting Critical Research in Information Systems1. MIS quarterly, 35(1), 17-36.

Myers, M. D., & Avison, D. (Eds.). (2002). Qualitative research in information systems: a reader. Sage.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). Predicting stock market price movement using sentiment analysis: Evidence from Ghana. Applied Computer Systems, 25(1), 33-42.

Nueman, W. L. (1994). Social research methods, quantitative and qualitative approaches.

Nyakurukwa, K., & Seetharam, Y. (2023). The evolution of studies on social media sentiment in the stock market: Insights from bibliometric analysis. Scientific African, 20, e01596. https://doi.org/10.1016/j.sciaf.2023.e01596

Oates, B. J., McLean, R., & Griffiths, M. (2022). Researching information systems and computing.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. Information systems research, 2(1), 1-28. https://doi.org/10.1287/isre.2.1.1

Pozzebon, M. (2004). Conducting and evaluating critical interpretive research: Examining criteria as a key component in building a research tradition. In Information systems research: Relevant theory and informed practice (pp. 275-292). Boston, MA: Springer US.

Piñeiro-Chousa, J., Vizcaíno-González, M. and Pérez-Pico, A.M. (2017), Influence of Social Media over the Stock Market. Psychol. Mark., 34: 101-108. https://doi.org/10.1002/mar.20976

Ren, J., Dong, H., Padmanabhan, B., & Nickerson, J. V. (2021). How does social media sentiment impact mass media sentiment? A study of news in the financial markets. Journal of the Association for Information Science and Technology, 72(9), 1183-1197.

Rosati, P., Deeney, P., Cummins, M., Van der Werff, L., & Lynn, T. (2019). Social media and stock price reaction to data breach announcements: Evidence from US listed companies. Research in International Business and Finance, 47, 458-469. https://doi.org/10.1016/j.ribaf.2018.09.007

Saunders, M., Lewis, P., & Thornhill, A. (2011). Research Methods for Business Students (4th edn.)

Smith, S., & O'Hare, A. (2022). Comparing traditional news and social media with stock price movements; which comes first, the news or the price change?. Journal of big Data, 9(1), 47. https://doi.org/10.1186/s40537-022-00591-6

Sul, H. K., Dennis, A. R., & Yuan, L. (2017). Trading on Twitter: Using social media sentiment to predict stock returns. Decision Sciences, 48(3), 454–488. https://doi.org/10.1111/deci.12229

Surulivel, S. T., Sultan, M. A., Ajithkumar, K., & Selvabaskar, S. (2025, March). Sentiment Analysis in Financial Markets: Using NLP to Predict Stock Price from News and Social Media Data. In 2025 International Conference on Visual Analytics and Data Visualization (ICVADV) (pp. 25-29). IEEE.

Tarsi, M., Douzi, S., & Marzak, A. (2024). Forecasting financial market dynamics: An in-depth analysis of social media data for predicting price movements in the next day. Social Network Analysis and Mining, 14(1), 169. https://doi.org/10.1007/s13278-024-01338-2

Teti, E., Dallocchio, M., & Aniasi, A. (2019). The relationship between twitter and stock prices. Evidence from the US technology industry. Technological Forecasting and Social Change, 149, 119747. https://doi.org/10.1016/j.techfore.2019.119747

Wu, X., Wang, X., Ma, S., & Ye, Q. (2017). The influence of social media on stock volatility. Frontiers of Engineering Management, 4(2), 201–211. https://doi.org/10.15302/J-FEM-2017018

Zhou, W., Wang, S., Li, X., Qi, Y., Guo, J., & Chu, X. (2025). Unleashing Expert Opinion From Social Media for Stock Prediction. IEEE Transactions on Knowledge and Data Engineering, 38(2), 1380-1394.

Zou, Y., Tian, J., Zhang, Y., Shi, G., & Cai, X. (2026). Digital Leadership, Information Entropy, and Stock Price Volatility: Evidence from CEO Social Media Behavior. Entropy, 28(2), 200. https://doi.org/10.3390/e28020200

Appendics

Appendix 1 : Reddit Sentiment Sample (January 2025)

Date	Ticker	Polarity	Subjectivity	Mention_Count	Reddit_Score
01.01.2025	AMZN	0.10446654	0.462425172	3	194
1.1.2025	NVDA	0	0	1	0
1.1.2025	TSLA	0.114423077	0.512713675	1	0
1.2.2025	AAPL	0.10615942	0.317753623	1	4
1.2.2025	JPM	0.061111111	0.377777778	2	2
1.2.2025	TSLA	0.141220013	0.449791667	2	42
1.3.2025	MSFT	0.415625	0.5	1	135
1.3.2025	NVDA	-0.05	0.525	1	55
1.3.2025	TSLA	0.042222222	0.414920635	3	85
1.4.2025	AMZN	0.089697571	0.469095719	6	77
1.4.2025	NVDA	0.160326556	0.481913149	3	3
1.4.2025	TSLA	0.13613738	0.416410659	7	27
1.5.2025	GOOG	0.070918367	0.399603175	1	1498
1.5.2025	NVDA	0.172903912	0.442624717	1	0
1.6.2025	JPM	0.167560764	0.508628472	2	28
1.6.2025	NVDA	0.124316116	0.483005971	10	839
1.6.2025	TSLA	0.102002643	0.432045963	12	1099
1.7.2025	GOOG	0.118024793	0.523628099	8	46
1.7.2025	NVDA	0.160082451	0.484454586	15	26594
1.7.2025	TSLA	0.070595238	0.444642857	2	26
1.8.2025	NVDA	0.095599415	0.442777778	2	5
1.9.2025	MRK	0.057291667	0.295833333	14	85
1.9.2025	MSFT	0.008916571	0.377458109	19	2278
1.9.2025	TSLA	0.178181818	0.468787879	1	475
1.10.2025	TSLA	0.077121144	0.514269991	13	481
1.11.2025	AAPL	0.149789513	0.490866614	27	19993
1.11.2025	AMZN	0.127243211	0.413032927	7	244
1.11.2025	GOOG	0.168978046	0.51306174	3	15
1.11.2025	NVDA	0.020694444	0.483888889	1	1

Appendix 2 : Financial News Sentiment Sample (FinnHub)

Date	Ticker	Headline	Polarity	Subjectivity
4/30/2025	AMZN	Franklin Global Equity Fund Q1 2025 Com	0	0
4/30/2025	AMZN	Amazon.com, Inc. (AMZN): Will AI Spendi	0	0
4/30/2025	AMZN	Amazon and Apple earnings coverage, Eli	0	0
4/30/2025	AMZN	Amazon Plans to Invest $4 Billion by 2026	0	0
4/30/2025	AMZN	Top Stock Reports for Amazon.com, Johns	0.5	0.5
4/30/2025	AMZN	Amazon Plans to Build Dozens of US Ware	0	0
4/30/2025	AMZN	Longtime Investor Has Been ‚ÄúSelective	0	0
4/30/2025	AMZN	Stock Market Today: Dow, S&P 500 Turn It	0.145454545	0.35
4/30/2025	AMZN	Microsoft (MSFT), Amazon (AMZN), Meta	0	0
4/30/2025	AMZN	Amazon Stock Retreats Ahead of Thursday	0	0
4/30/2025	AMZN	Meta: An Exceptional Q1 2025 That Sprea	0.666666667	1
4/30/2025	AMZN	Amazon, Ford, Temu, and 7 other compan	-0.125	0.375
4/30/2025	AMZN	Trump's 145% China Tariffs Spark Seller B	0	0
4/30/2025	AMZN	Is Amazon.com Inc. (AMZN) the Best Dow	0.5	0.15
4/30/2025	AMZN	Why Amazon, Meta Platforms, and Alphab	0	0
4/30/2025	AMZN	Nvidia gets rare sell rating as Seaport say	0.3	0.9
4/30/2025	AMZN	Amazon.com (NasdaqGS:AMZN) Partners	0	0
4/30/2025	AMZN	Amazon to invest $4 billion to expand rura	0	0
4/30/2025	AMZN	Amazon com : investment in small towns	-0.125	0.575
4/30/2025	AMZN	Update: Market Chatter: Grindr Using Am	0.136363636	0.454545455
4/30/2025	AMZN	Amazon Investing Over $4 Billion To Expar	0	0
4/30/2025	AMZN	Temu and Shein customers in the US are r	0.0625	0.375
4/30/2025	AMZN	Amazon.com, Inc. (AMZN): Among Billion	0.2	0.95
4/30/2025	AMZN	10 Consumer Discretionary Stocks With V	0	0
4/30/2025	AMZN	Dow's 237-point fall led by losses in Chev	0	0

Appendix 3 : Correlation summary

Variable Pair	Correlation Coefficient	Interpretation
Overall_Sentiment vs Return	~0.03	Very weak positive relationship
Sentiment_Lag1 vs Return	~0.03	No meaningful lag effect

Trend_Score vs Lagged Trend	~0.69	Strong persistence over time
Volume_Change vs Return	~ -0.11	Slight negative relationship

Appendix 4 : ANN model performance

Metric	Value	Interpretation
R-Squared (R^2)	0.5537	Model explains ~55% of variance
Mean Absolute Error (MAE)	114.17	Average prediction deviation

Appendix 5 : Model comparison (stock returns)

Model	Features Used	Performance	Observation
Linear Regression	Financial Only	RMSE: 0.01658	Baseline
Linear Regression	+ Sentiment	RMSE: 0.01663	No improvement
Random Forest	Financial Only	RMSE: 0.01980	Baseline
Random Forest	+ Sentiment	RMSE: 0.01951	Slight improvement
ANN	Financial + Sentiment	R^2: 0.5537	Captures non-linear relationships

Appendix 6 : Trading volume model returns

Model	Features Used	Performance	Observation
Random Forest	Financial Only	RMSE: 0.4359	Baseline
Random Forest	+ Sentiment	RMSE: 0.4449	Performance declined

Appendix 7 : Summary of key findings

Area	**Finding**
Sentiment vs Returns	Weak linear relationship
Non-linear Modeling	Slight improvement with random forest and ANN
Price Direction	No predictive distinction
Trading Volume	No meaningful relationship
Sentiment Signal	Content more useful than volume