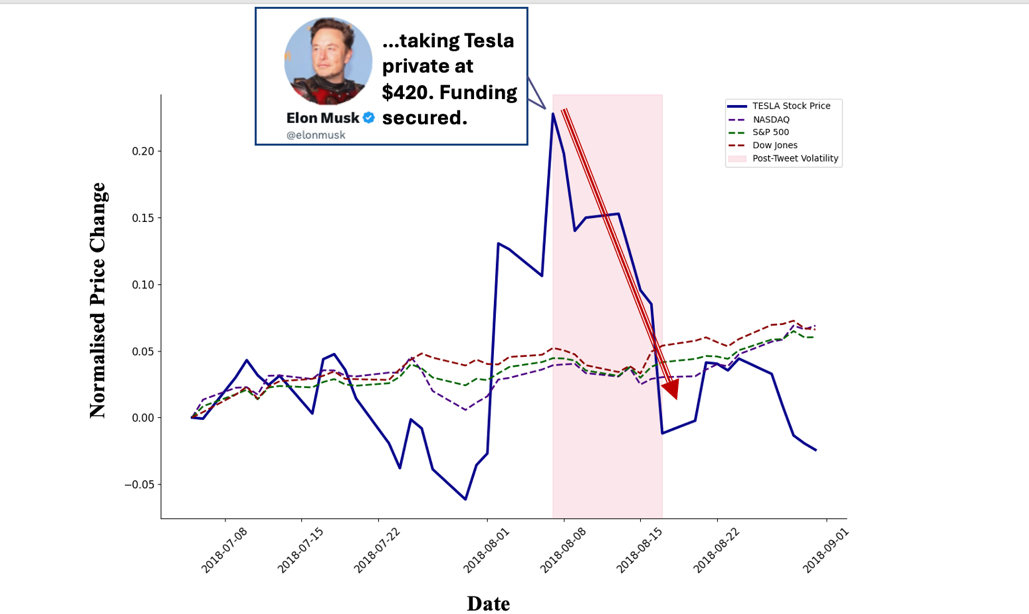
**Stock Market Analysis informed by CEO Social Media Communications**

**Efsa Sema AKTAY**

1. **Introduction**
   1. **Background**

Traditionally, the companies have been evaluated for their fair pricing through fundamental analysis. Fundamental analysis utilises multiple factors by evaluating corporate performance , interest rates and discount rates, to forecast to free cash flow and arrive at a fair value price. Although, the financial markets have been generally influenced by a variety of factors, including economic indicators, geopolitical events, and corporate performance; in recent years, social media has emerged as a force capable of changing market sentiment and influencing the movement of stock prices. Social media communication especially from influential CEOs, have shown to have an impact on market behaviour. To exemplify, the instance such as Elon Musk's tweets about Tesla, provide a context for exploring this phenomenon. Specifically, Elon Musk's infamous tweet about taking Tesla private at $420 per share led to massive market reaction and SEC scrutiny. Another significant shock on the market is the case of GameStop stock price had received significant increase in both 2020 and 2024 through the Reddit community called r/wallstreetbets through the simultaneous purchase of the stock. High-profile cases, such as the Tesla and GameStop stock surges, highlight the significant impact of social media activity on stock prices. For instance, the GameStop short squeeze in early 2021 was heavily influenced by discussions and coordination on the subreddit r/wallstreetbets, illustrating the power of collective social media behaviour in driving market dynamics.



*Figure 1: Displaying the stock price decrease following Elon Musk’s tweet.*

A graph of a price fall

Description automatically generated with medium confidence

*Figure 2: Displaying the stock price increase in the case of GameStop.*

Despite the growing influence of social media communications, the quantitative impact of CEO tweets on market movement have not been extensively researched and documented. Specifically, the mechanism for which these communications affect stock market volatility and price movements provide an opportunity for further research. The objectives include firstly, gauging whether there is a quantifiable impact of CEO tweets on stock market volatility and price movements of the stocks. Secondly, developing models that can efficiently utilise the text input within the predictive models for stock market forecasting.

* 1. **Importance**

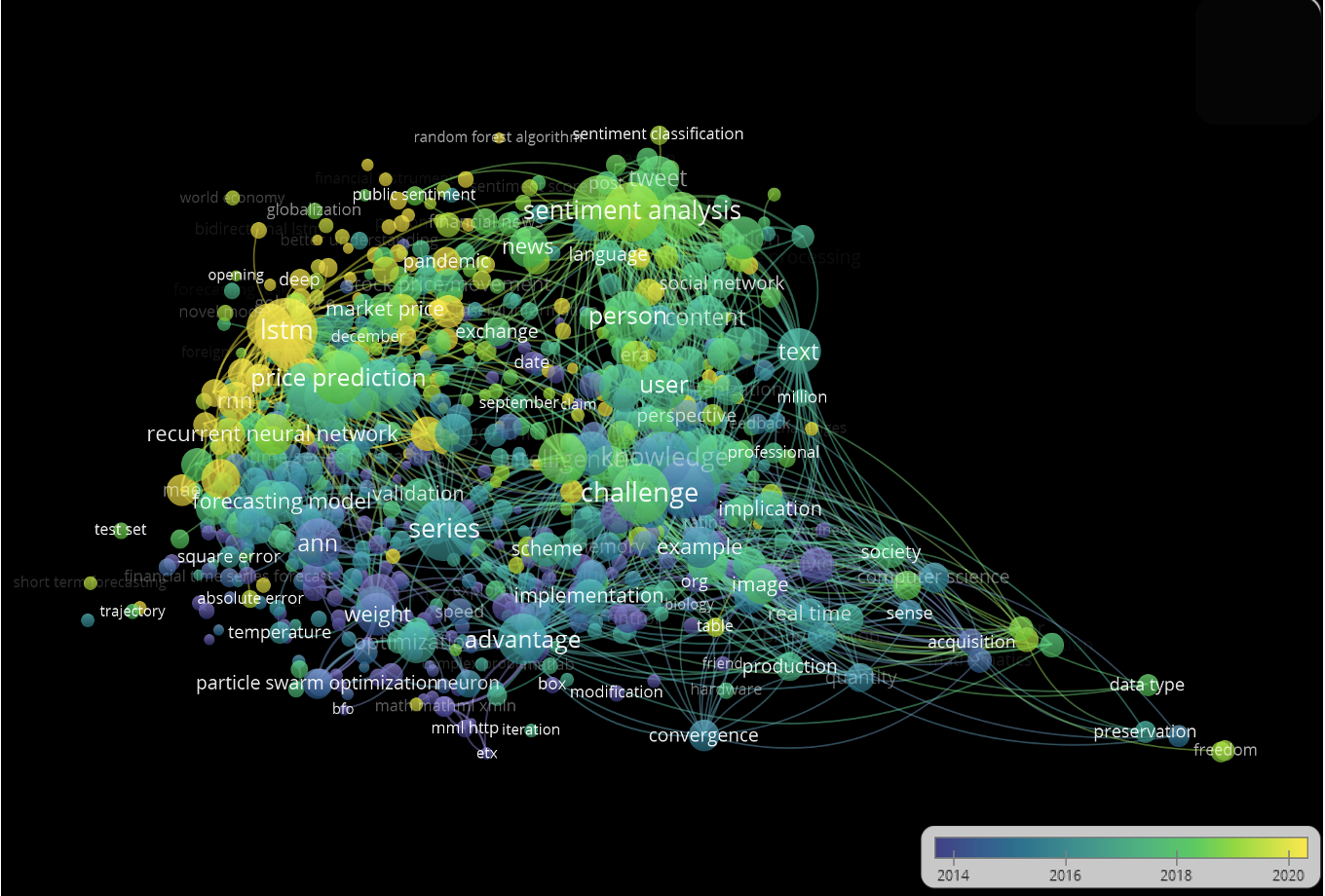
Exploration of the influence of non-traditional market, such as social media communication can help enhance the accuracy of predictive financial models, provide windows of opportunity for profit or help reduce the downside risk. This analysis could provide stakeholders such as investors, financial analysts and managers with new tools and insights for market analysis. By integrating social media analytics into financial models, these stakeholders can gain a slightly better understanding of market dynamics and create possible opportunities for achieving better investment outcomes.

1. **Literature Review & Theories**
   1. **Efficient Market Hypothesis and Random Walk Theory**

There are two major theories that make up the foundation of the financial stock forecasting. Firstly, the Random Walk Theory suggests that stock price changes are random and do not follow a predictable pattern. Secondly, the Efficient Market Hypothesis, proposed by Fama (1970), reflects that stock prices reflect all the available information, proving it quite impossible to generate returns that are consistently higher than the overall market. According to EMH, any new information is quickly incorporated into stock prices, rendering them unpredictable. The study is exploring the opportunity for discerning the market sentiment through social media analysis and recognises the randomness of the stock market data.

**B. Literature Review**

Existing sentiment analysis techniques have been applied to financial predictions with varying degrees of success. These methodologies often involve natural language processing (NLP) to analyse the sentiment of social media posts and correlate them with stock market data. Previous studies have employed various machine learning models, including linear regression, support vector machines, and neural networks, to predict stock price movements based on sentiment analysis.



*Figure 3: Displaying the existing research in a cluster algorithm.*

1. **Methodology**
   1. **Data Collection**

A diagram of a data processing process

Description automatically generated

*Figure 4: Displaying the flowchart of the overall process.*

The data collected can be grouped into two, the synthetic data that is collected to validate the model and the real company and tweet data. The synthetic data was created by the help of GENAI specifically, GPT-4 to simulate the tweeting patterns and corresponding profiles of companies. In the end, three different stock and index prices alongside the corresponding CEO text tweet were generated systematically. By assigning the random seed number it was ensured that the data can be reproduced for verification purposes. The generated tweets had a randomised correlation with the tweet.

For the real data, the stock price data of Apple and Tesla were retrieved from Yahoo Finance between 2022-01-01 to 2024-04-15 and the corresponding tweets were sourced from X ApiV2 through the Api subscription. Certain rate limits and the limits on the scraped tweets is a possible constraint for the comprehensive analysis. Tweets were collected corresponding CEOs which are Elon Musk and Tim Cook.

A screenshot of a computer

Description automatically generated

*Figure 5: Displaying the scaled data frame of stock price containing tesla, apple and major US stock indexes.*

A white background with black text

Description automatically generated

*Figure 6: Displaying the scraped tweet data frame of Tim Cook.*

**B. Data Processing**

BERT (Bidirectional Encoder Representations from Transformers) was utilised for processing the text data which was embedded as an input into the deep learning model. The output from the embedder was integrated with stock price data to develop the stock price prediction model. The machine learning algorithms LSTM was utilised train the model. The BERT embedder and the LSTM models were selected because of their ability to handle timeseries data and capture complex patterns in text, making them well-suited for analysing the temporal dynamics of tweet sentiments and stock prices.

**A diagram of a graph

Description automatically generated with medium confidence**

*Figure 7: Displaying the embedding in a Pairwise Similarity Matrix*

**C. Validation**

The model was validated using both real and synthetic datasets through two different sets of tests. In both cases, to ensure robustness and reliability of the result, the set of stock and index data is trained both with and without the tweet data and the results are compared. The actual validation process is completed through comparing the model's predictions with actual stock price movements and evaluating its performance through the root mean square error (RMSE).

1. **Analysis**
   1. **Model Performance & Results**

The predictive model was compared with traditional financial models, such as the Autoregressive Integrated Moving Average (ARIMA) models, which primarily rely on historical stock price data. The integration of social media analytics into these models showed improvements at certain key moments. Specifically, the models incorporating the analysis data from CEO tweets outperformed the traditional models in past case scenarios with moments that could be pinpointed to a tweet. However the overall performance of sentiment integration was not robust enough to perform better at the real stock market data. In that case, some limitations were noted, in such where the tweet data provided some noise, reducing the robustness of the overall prediction. The research can be furthered by filtering the tweets posted by the CEO’s prior to their incorporation into the model. As Elon Musk tweets very frequently, the additional tweet data introduces noise into the model that do not have much impact on the stock price.

A graph of a graph of a graph

Description automatically generated with medium confidence

*Figure 8: Displaying the loss result from the model training in synthetic data.*

A graph of a graph of a graph

Description automatically generated with medium confidence

*Figure 9: Displaying the loss result from the model training for the real data.*

1. **Discussion**
   1. **Limitations and Challenges**

The study faced potential inaccuracies due to the influence of external factors. For example, geopolitical events, regulatory changes, and macroeconomic indicators can all impact stock prices independently of social media sentiments especially the post covid recovery could be one factor that has not been considered in the analysis. Additionally, the sentiment analysis models may misinterpret sarcasm or context-specific language used in tweets. Further research is needed to refine the models and account for these variables, incorporating more sophisticated NLP techniques and more robust data sources for a real time analysis.

**B. Future Research Directions**

Future studies could explore the impact of other social media platforms, such as Reddit and LinkedIn, on stock market trends. Additionally, deeper linguistic analyses, including the detection of sarcasm, irony, and context-specific language, could improve the accuracy of sentiment analysis. The development of more sophisticated models and real-time data processing, could further enhance predictive accuracy.

**C. Technological Advancements**

The future of AI and NLP in financial analytics is promising. With the continued advancements in these technologies are likely to provide even greater insights and predictive power. For example, improvements in deep learning algorithms, real-time data processing capabilities, and the integration of diverse data sources can significantly enhance the ability to predict market movements based on social media activity.

**Bibliography**

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>

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. Journal of Finance, 25(2), 383-417. <https://doi.org/10.2307/2325486>

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. PloS One, 10(9), e0138441. <https://doi.org/10.1371/journal.pone.0138441>

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. Journal of Finance, 62(3), 1139-1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through Twitter “I hope it is not as bad as I fear”. Procedia-Social and Behavioral Sciences, 26, 55-62. <https://doi.org/10.1016/j.sbspro.2011.10.562>

**Appendix**

A graph of a stock market

Description automatically generated

*Figure 1: Tesla Stock Price plotted together with S&P 500 and 9 and 21 day the moving averages.*

A graph of stock prices

Description automatically generated

*Figure 2: Apple Stock Price plotted together with S&P 500 and 9 and 21 day the moving averages.*