ABSTRACT

The study examines the dynamic interaction of financial news sentiment and market performance using advanced Natural Language Processing (NLP) techniques available today. The objective is to explore sentiment analysis in financial news and its correlation with market dynamics, focusing on stock price volatility and directional trends. In this study, a large corpus of financial news content would be processed using advanced Bidirectional Encoder Representations from Transformers (BERT) models in conjunction with other algorithms. The goal is to identify and quantify sentiments indicators. The sentiment scores generated from the news headlines and the content are compared against the historical market data primarily focusing on prevalent stocks. The objective is suggesting a more nuanced relationship between media sentiment and market performance, as it makes clear the impact of sector-specific news sentiment on the related companies, and overall market sectors. Results reveal that although the machine learning approaches are very effective with respect to sentiment mining from financial texts, their application in the context of predicting stock returns is complex which resulted in a profusion of nullified profits due to transaction costs associated with making trades. Nevertheless, such findings underscore the potential to use sentiment analysis as a tool for market prediction but with nuanced implications for the strategy. This report adds another dimension to a vast body of research by presenting the sentiment analysis in financial markets and this time providing insights into an emerging data-driven predictive news sentiment with its practical applications in finance.

TABLE OF CONTENTS

INTRODUCTION

Sentiment reflected in financial news has long been a key factor in shaping market dynamics. The following examples demonstrate the real influence of social media, and Twitter in particular, on the valuations made in stock markets. Corporate tweets, for instance, have shown to positively impact stock returns and trading volumes, as evidenced in a study involving over 1.2 million tweets from thirty Dow Jones Industrial Average companies (Ganesh and Iyer, 2021). When Elon Musk posted his tweet about Tesla's stock price being too high, the electric car manufacturer lost suddenly $15 billion off its market valuation. (Smith and O'Hare, 2022).

However, the emergence of data science and AI has transformed how we comprehend and utilize such sentiment. The effective deployment of NLP and Machine Learning (ML) for sentiment analysis in financial news has been a significant leap forward. For instance, Yadav (2020) demonstrated that the Noun-Verb approach with a level of accuracy of 79% was better than Turney's approach and Hybrid approach techniques respectively with an accuracy of 75% and 76% (Yadav, 2020).

Further, Wan et al. (2021) investigated the relationship between news sentiment and 2financial market by considering that media sentiment has non-linear dynamics with its positive or negative news sentiment, exerting non-monotonic effects on valuations. Similarly, Fazlija and Harder (2022), using BERT models predicted S&P 500 index movements, showing real-world applications of these sophisticated methodologies.

However, translating sentiment analysis to correct market prediction can be challenging. For instance, Azar (2009) shows that although machine learning algorithms vastly outdid word-counting approaches in sentiment analysis, they made costly errors when it came to classifying the outliers. In experiments, Decision Trees gave 10.5% of the annualized return but were not economically significant by trading costs that support a semi-strong form of market efficiency hypothesis (Azar, 2009).

This study develops these findings to analyse the predictive power of financial news sentiment on stock market moves. It intends to investigate the implications of these advanced techniques in practical financial decision-making and strategy development answering such research questions as:

How well do the NLP tools perform in identifying sentiment in financial news while predicting sentiment in stock returns, respectively?

What is the degree of correlation between sentiment metrics derived from financial news and the subsequent movements in stock market indices?

What are the implications of these findings for practical financial decision-making and strategy development?

METHODOLOGY

DATA COLLECTION AND PRE-PROCESSING

Data source:

The primary dataset (Tweets about the Top Companies from 2015 to 2020) was created for a paper presented at the 2020 IEEE International Conference on Big Data, focusing on identifying speculators and influencers in the stock market and available to download at Kaggle.

Description of Da ta:

The dataset contains over 3 million unique tweets, each with details like tweet ID, author, posting date, tweet content, and engagement metrics (comments, likes, retweets) spanning 5 years. This is divided into two parts on Kaggle, with a supplementary list of the companies associated with the tweets. It is particularly useful for those interested in analysing tweets related to Amazon, Apple, Google, Microsoft, and Tesla, using their respective stock tickers.

Algorithm and Procedures used:

In order to identify the most effective algorithm that yields the optimal result for sentiment analysis in this project, the performance of AFINN by Finn Årup Nielsen, VADER (Valence Aware Dictionary and Sentiment Reasoning), TextBlob, and FinBERT, a specialized NLP model designed for sentiment analysis of financial text were evaluated. While AFINN, VADER and TextBlob are more traditional sentiment analysis tools, FinBERT represents a more advanced approach, leveraging deep learning and large-scale data training specific to the financial domain (Araci, 2019).

For this analysis, the 'Stock Market Tweets Data' by Taborda et al. (2021) was employed. The dataset consists of 5,000 tweets randomly sampled out of a larger corpus of 943,672 tweets collected between April 9 and July 16, 2020. Of these 5,000 tweets, 1,300 were manually annotated - reviewed by a second independent annotator for precision reasons. Furthermore, the 'Financial PhraseBank' was utilized, which comprises of 4845 sentences excerpted from financial news headlines reflecting the perspective of a retail investors. The sentences were randomly selected out of the LexisNexis database and later annotated by 16 experts in finance and business administration (Malo et al., 2013).

The merged dataset result is a comprehensive record of sentiments in the financial domain with 2380 Positive, 1208 Negative and 3554 Neutral Tweets. Class weight balancing techniques have been utilized to treat the imbalance to ensure a more accurate analysis.

Python Environment: Google Colab

Google Colab, an online interactive development environment that enables users to write and execute Python code through the browser was the choice platform for the project. It possesses both large memory capacity and high-performance GPUs, essential for speed needed for computationally intensive deep learning models which helps in sentiment analysis tasks. Additional benefits of using Google Colab are that it offers a pre-configured environment along with pre-installed essential libraries, thus eliminating all the issues related to setup. Colab also integrates with Google Drive for easy data management especially since local files on Colab are removed periodically.

The assessment of the combined training dataset revealed that it was relatively clean and did not require any pre-processing. However, a function “TwitterTextProcessor" was designed for pre-processing and cleaning, specifically tailored for both Twitter and news headline content. Its key functions include removing non-ASCII characters and hyperlinks, handling hashtags and mentions, tokenizing and lemmatizing text, and cleaning for both general and VADER-specific sentiment analysis. Additionally, it offers the capability to process large batches of tweets efficiently using parallel processing and create CSV files containing pairs of original and cleaned tweets. Utilization of this function on the primary dataset hinged on which of the algorithms had the highest percentage match when the original labelled sentiments are compared with the AFINN, VADER and Finbert scores. A pre-processing step identified for the primary dataset was the removal of prevalent redundant "Read more" text, along with any prevalent hyperlinks, using regular expressions.

STATISTICAL ANALYSIS

On evaluating of the sentimental scores of all proposed tools, the following percentage is obtained:

AFINN	**VADER**	**TextBlob**	**Finbert**
55.82%	53.72%	47.07%	11.19%

Although Finbert had been specifically trained on domain-relevant data, equipping it with a robust capability to interpret and process financial language nuances (Yang et al., 2022), further training of the Finbert model with the merged labelled dataset was necessary to boost the percentage accuracy.

CLASS WEIGHTS (EFFICIENCY IN HANDLING IMBALANCED DATA)

Class weights were calculated using the compute_class_weight and utilized in the loss function (CrossEntropyLoss). It's a huge advantage while handling imbalanced training datasets wherein some classes showed under representation as opposed to the others (Androidkt.com, 2021).

The allocation of higher weights to the underrepresented classes, penalizes the model more for misclassifying these classes. This enhances the focus of the model and, thereby, may evoke balanced performance across all classes, which enhances generalization. By this means, it effectively introduces efficiency in the training time by reconciling the imbalance first without necessary adding additional data nor requiring more complex kinds of resampling techniques.

EARLY STOPPING (EFFICIENCY IN TRAINING TIME AND AVOIDANCE OF OVERFITTING)

Early stopping is implemented with a patience parameter that controls how many epochs to wait for an improvement in the validation loss before stopping the training process. This prevents the model from overfitting, in addition to saving computation time (Brownlee, 2020).

If the model does not improve over the performance on the validation set, training it further is unlikely to achieve better results and would only result in overfitting of the training data. Moreover, this helps in selecting a model that generalizes well since the point of stopping is defined in validation performance terms. Utilizing this approach boosted the percentage accuracy of FinBert model compared to the original labelled dataset to **88.15%.** and the Finbert model was backed up safely to Google drive.

DATA DISTRIBUTION

Understanding the distribution of data is fundamental in statistics and can have a variety of applications since a dataset's overall shape and behaviour, which is crucial for making informed decisions based on the data can be described visually (Polamuri, 2023). Multimodal, binomial and log-normal distribution are popular distributions with regards daily tweet volume (Eom et al., 2015). For distributions, particularly changing ones over time, an area chart is able to clearly demonstrate the overall shape and tendencies of it more fluently when compared to a line chart. Filling in the space under the line; may make patterns of change more apparent, and give a sense of the total volume. This can be particularly useful when displaying data that increases over time, such as tweet volumes where you may want to show the build-up of tweets over different periods.

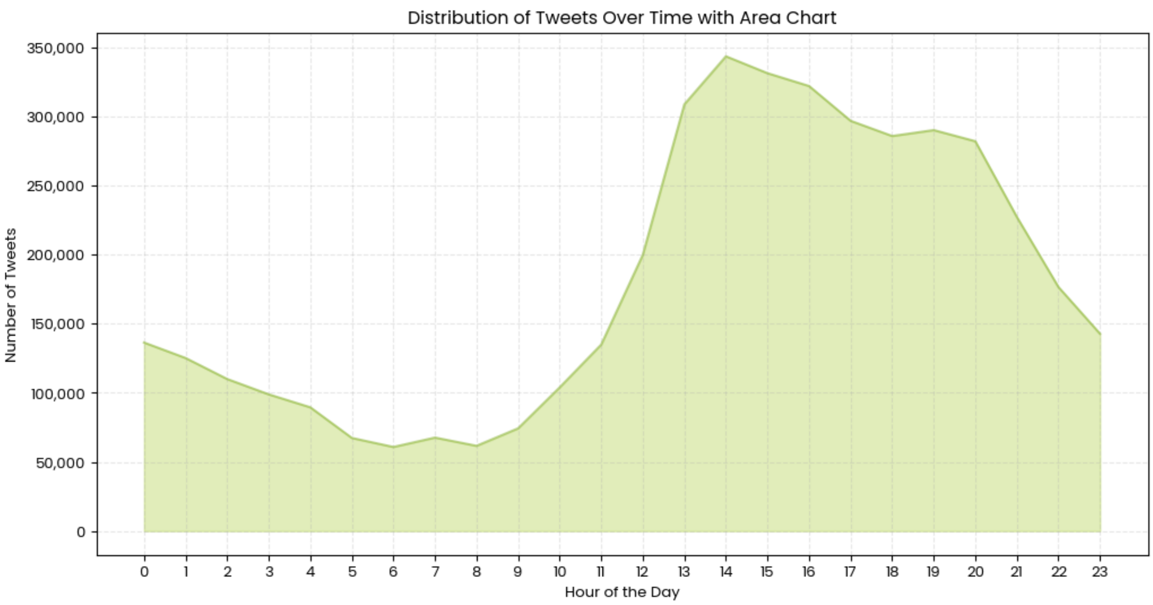

Figure 1: Tweets distribution over time (hour of day) with area chart

The distribution in Figure 1 appears to be multimodal, as there are several distinct peaks visible throughout the day, indicating multiple periods of high activity, unlike a bimodal distribution which would typically have two prominent peaks. This suggests that there are several different times when tweet activity is heightened, which could be due to various factors such as global user activity across different time zones, or peaks corresponding to specific events or announcements.

The time series data was shifted back by one using pandas' shift operator to indicate if tweet volume impacts a company's traded volume. This will then allow further analysis for comparison of the previous day tweet volume with the following day's company share price.

Spearman correlation, preferred to Pearson correlation due to its non-reliance on normal data distribution was employed with these hypotheses:

Null: No correlation between the tweet volume and stock trading volume. Alternate: Correlation exists between the tweet volume and stock trading volume. A Spearman correlation p-value of less than 0.05 will result in rejecting the null hypothesis in favour of a possibility that there might be a negative correlation between the stock trading volume and tweet volume, and this negative correlation might be statistically significant.

Additionally, the trading volume and tweet volume were plotted together on the same graph, to easily overview data. Following this plot line could have been confusing, hence a rolling average using a 30-day window was derived in order to clear any confusion brought about by the correlation of trends.

The 30-day rolling average is important time series analysis, especially for data related to stock market movements or social media activity, because of its ability to smooth out short-term fluctuations and at the same time highlight long-term trends. This strategy mitigates the effect that transient spikes or dips in data possess, making it easier to identify underlying patterns or trends. It is most useful for ongoing monitoring and analysis in dynamic environments, providing a clearer view of data trends by averaging out seasonal variations as well as noises. The 30-day window is a particularly good choice since: it helps striking a balance of smoothing the data and preserving enough information to be able to have meaningful analysis (Frost, 2020) (Kusawa, 2023).

In mathematical terms, the process described can be translated into a formulaic representation as follows:

1. Time Series Shift:

Let TV[t-1] be the tweet volume on day t-1 and SV[t] be the stock volume on day t.

The shifted tweet volume time series is represented as TV'[t] = TV[t-1], where each value is the tweet volume from the previous day.

2. Spearman Correlation Test:

The Spearman correlation coefficient is calculated as ρ(SV[t], TV'[t]), representing the correlation between the stock volume on day t and the shifted tweet volume.

3. Hypothesis Testing:

- Null Hypothesis (H0): ρ = 0 (No correlation)

- Alternative Hypothesis (H1): ρ ≠ 0 (Significant correlation)

- Decision Rule: Reject H0 if p-value < 0.05.

4. Visualization:

The 30-day rolling averages of the stock volume (SV[t,w]) and shifted tweet volume (TV'[t,w]) are plotted over time t, with a rolling window w of 30 days.

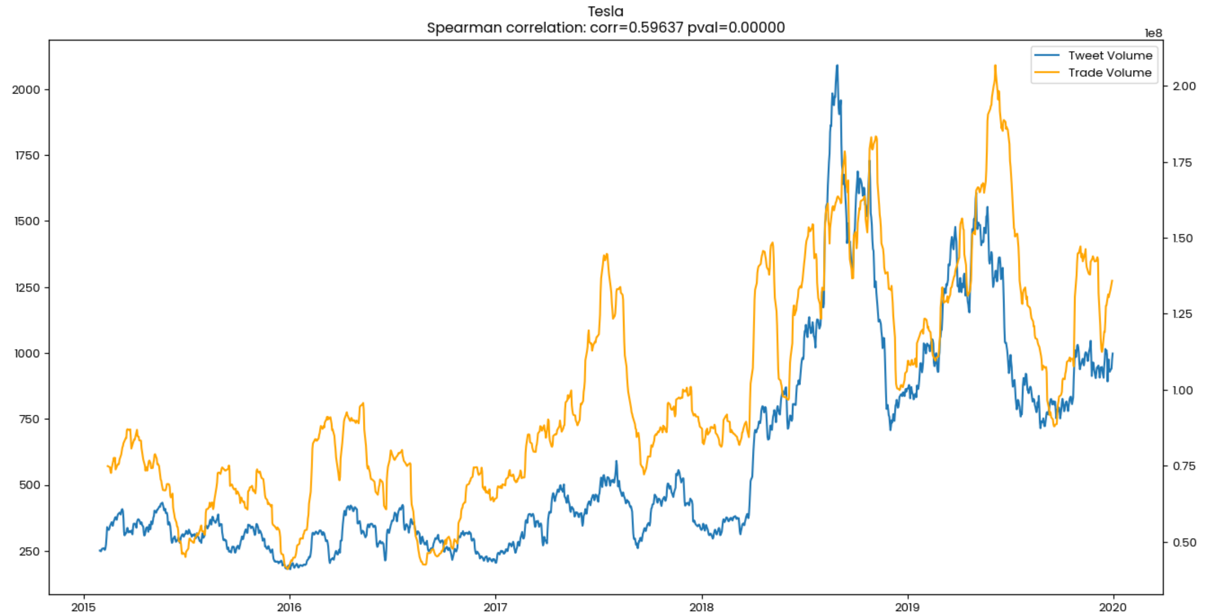

Figure 2: Spearman correlation of tweets and trade volume of Tesla’s stock from 2015 to 2020

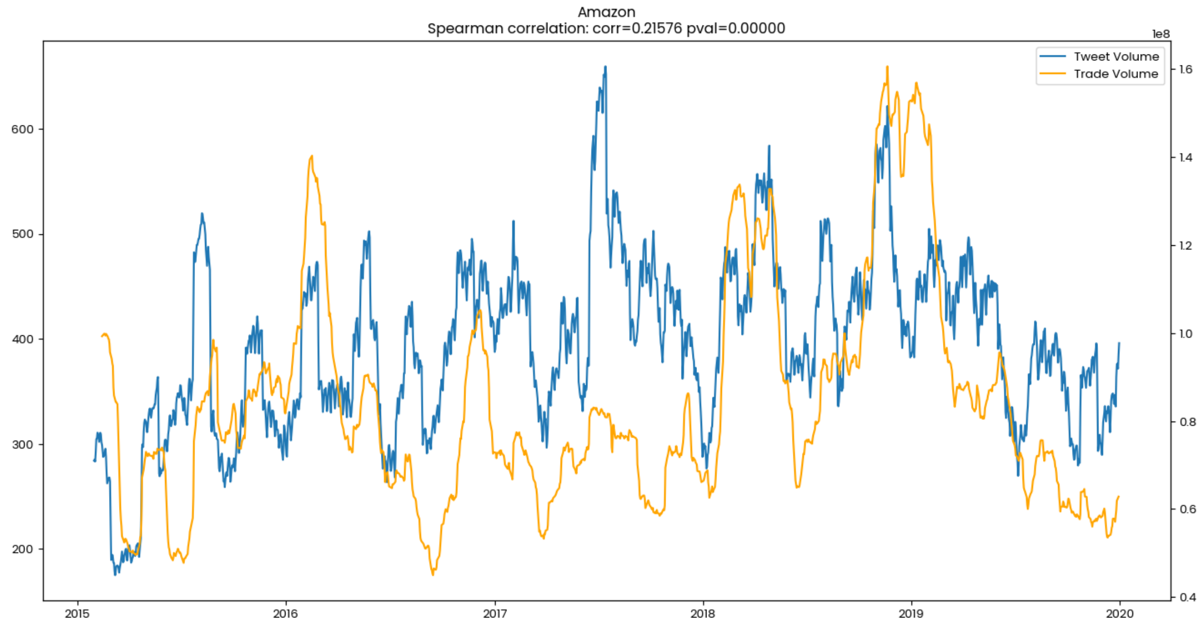

Figure 3: Spearman correlation of tweets and trade volume of Amazon’s stock from 2015 to 2020

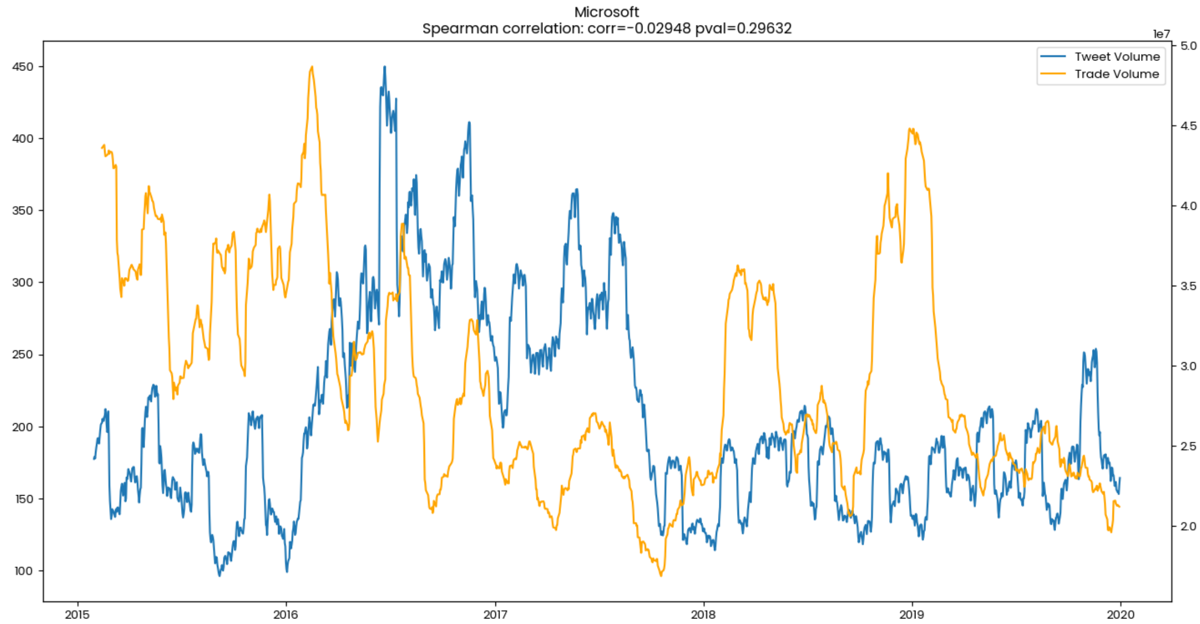

Figure 4: Spearman correlation of tweets and trade volume of Amazon’s stock from 2015 to 2020

RESULT

On close observation, there seems to be a positive correlation between the large number of tweets and trade volume in spite of having significant null/nan values in stock rolling mean. Nonetheless, the robustness of this correlation is highly debatable. Furthermore, it's not definitively established that the volume of tweets consistently correlates with the share price. This is evident from the case of Microsoft, where the p-value turned out to be higher than the prescribed cut-off points 0.05 and therefore raised questions on the statistical robustness of the correlation.

The next hypothesis to examine is as follows:

i. Null Hypothesis: There is no correlation between the sentiment expressed in tweets, and the share price of the company.

ii. Alternate Hypothesis: There exists a correlation between the sentiment in tweets, as determined by the FinBERT model, and the company's share price.

For that, the saved pretrained Finbert model is used to obtain the sentimental scores of the primary dataset segmented by their individual Company ticker symbols.

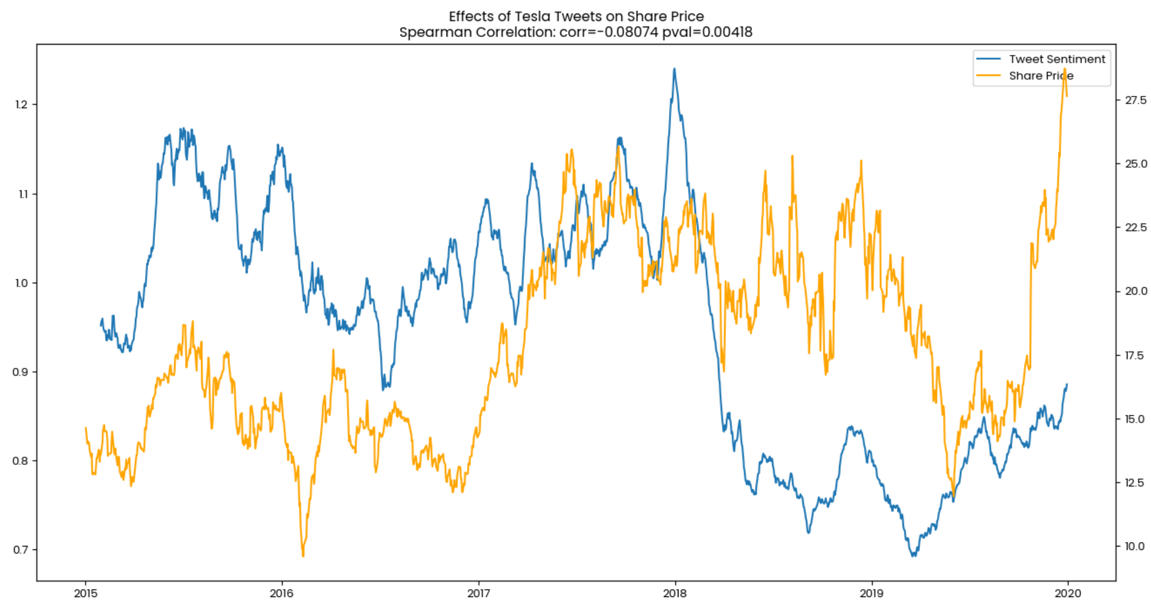

Figure 5: Spearman correlation of tweet sentiments and stock price of Tesla's stock from 2015 to 2020

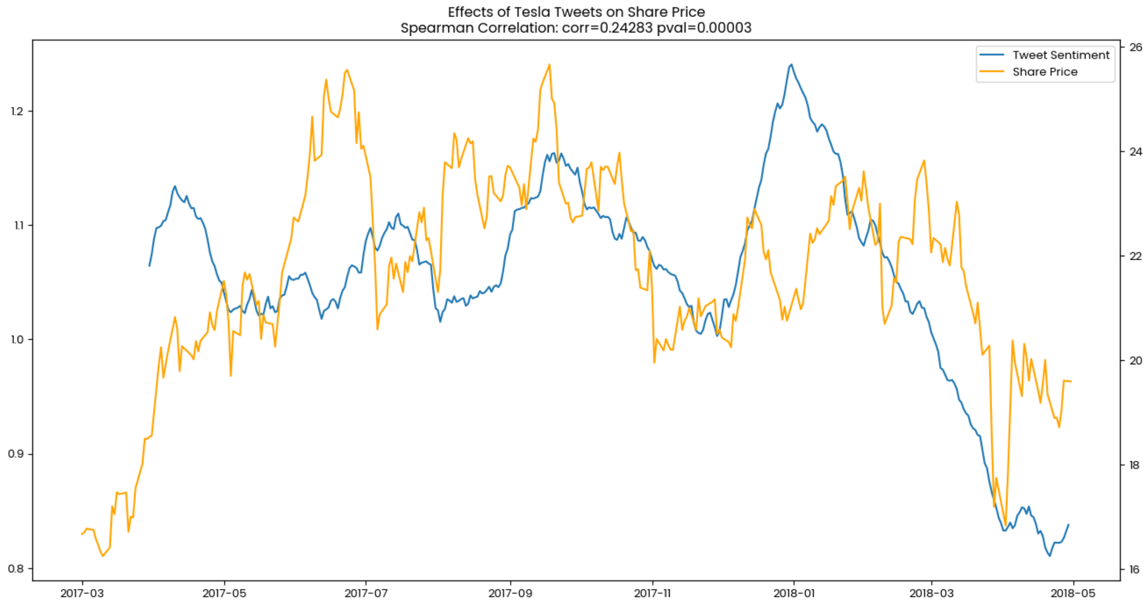

Figure 6: Spearman correlation of tweet sentiments and stock price of Tesla's stock from March 2017 to April 2018

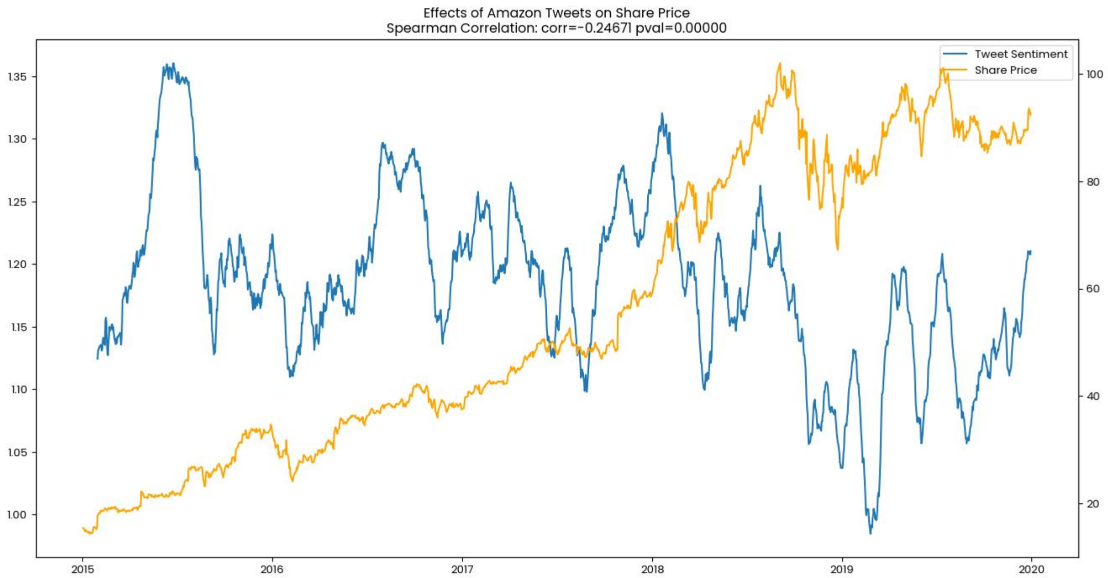

Figure 7: Spearman correlation of tweet sentiments and stock price of Amazon's stock from 2015 to 2020

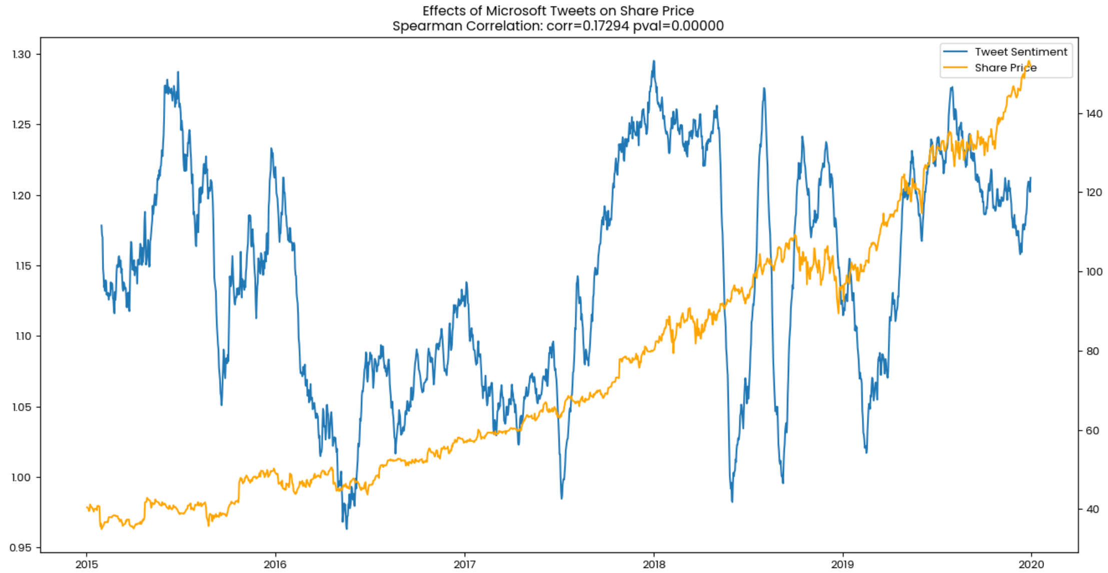

Figure 8: Spearman correlation of tweet sentiments and stock price of Microsoft's stock from 2015 to 2020

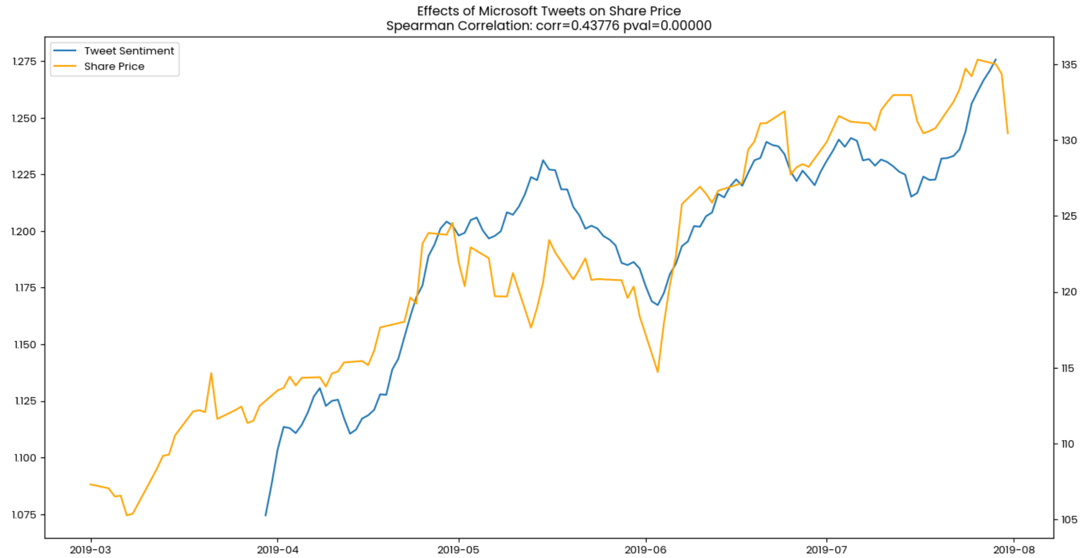

Figure 9: Spearman correlation of tweet sentiments and stock price of March 2019 to July 2019

Through the visualization with tweet sentiment and stock price movements for Tesla, Amazon and Microsoft, we aimed to figure out whether public sentiments posted in Twitter have a statistically significant association with the market valuation of these companies.

In this regard, the data was analysed in light of these two hypotheses:

i. A null hypothesis postulating the absence of a relationship between tweet sentiments and stock-prices.

ii. An alternate hypothesis postulating that a relationship does exist.

In this analysis, the method involved deployment of a function used to calculate the Spearman correlation of the daily aggregated tweet sentiment score with stock prices. Spearman's correlation is a nonparametric measure of statistical dependence between two variables, and describes a monotonic function for the relationship between the two variables.

Tesla's overall data -0.08074 presented an inverse relationship with sentiments negatively correlated with stock prices. However, when analysing a more focused period between 2017 and 2018, the correlation 0.24283 turned indicating how sentiment-stock correlations change with time frames.

Calculate the difference in correlation coefficients: |0.24283| - |-0.08074| = 0.24283 - 0.08074 = 0.16209

Calculate the percentage increase from the overall to the specific period: (0.16209 / 0.08074) * 100 = approximately 200.7%.

Amazon's analysis similarly yielded a negative Spearman correction of -0.24671. This counterintuitive result could be illustrative of investor behaviour where positive sentiment does not translate into increased stock demand, product of other confounding variables not accounted for during the analysis, or simply showing no correlation between the variables.

For Microsoft, the analysis showed Spearman's correlation coefficient over this extended time from 2015 to 2020 indicated a weak positive correlation between tweet sentiment and stock prices. This longer-term perspective suggests that while sentiment and price do not move in lockstep, there is a general tendency for an increase in stock price with more positive tweet sentiment. However, a stronger positive correlation was observed when the analysis was also narrowed down to a specific period in 2019. These findings suggest a potential temporal influence over the relationships between sentiment and stock prices, reflecting market conditions or particular events that may have increased the impact public sentiment had on stock valuation during this period.

Calculate the difference in correlation coefficients: |0.43776| - |0.17294| = 0.43776 - 0.17294 = 0.26482

Calculate the percentage increase from the overall to the specific period: (0.26482 / 0.17294) * 100 = approximately 153.13%

In conclusion, the data provides evidence against the null hypothesis and suggest that tweet sentiment does bear some correlation with stock prices, albeit the direction and strength of this correlation vary with time and across different stocks. The findings underscored the complexity in stock markets whereby investor sentiment interplays, perhaps via social media with a host of other financial, economic, and psychological factors that affect stock prices.

DISCUSSION

This research focuses on sentiment analysis employing, AFINN, VADER, TextBlob, and FinBERT to achieve accuracy of 55.82%, 53.72%, 47.07% and 11.19% respectively on a labelled dataset. Despite FinBERT's specialized training for financial contexts, it needed further refinement to achieve an improved accuracy of 88.15%.

The Pearson correlation study demonstrated significant enhancement of the values of pairwise correlation coefficients as smaller date range analysis was conducted for Tesla and Amazon Stocks. For Tesla, this positive correlation increased to about 200.7%, while impressive results were also derived for Amazon utilizing a 30-day rolling window. Similar research employing pretrained Finbert and Long Short-Term Memory (LSTM) model for a dataset spanning 2 years demonstrates a consistent predictive capability of 98.59% even on a rolling window of 10 days (Halder, 2022)! While the Pearson analysis in this research highlights the impact of specific time frames on sentiment-stock correlations, the overall performance of utilizing an LSTM model suggests a general predictive performance across a wider range of data. The lack of usage of LSTM and other machine learning models like Recurrent Neural Networks (RNNs) for predicting stock prices was a limitation of this research.

Further limitations include the time required to obtain the primary dataset necessary for stock market prediction, as many websites offered only a paid service for large historical data. The inability to utilize web scrapping techniques for most social media websites like Twitter was also a hindrance. Additionally, free resources such as the New York Times API often yielded datasets of subpar quality, characterized by significant noise and less reliability for accurate predictions.

Processing of sentimental scores of the primary dataset using the pretrained FinBert Model took considerable time averaging 5 hours per selected stock indices even when optimised with batch processing methods (100 batches) using intensive GPU resources on Colab. This hindered the research from analysing other scenarios that could potentially boost correlation between sentiments and stock prices like tweets with high retweets, actual numerical sentimental scores not limited to labelled negative (0), neutral (1) and positive (2) scores.

CONCLUSION

Several studies have been made with regards the relationship between Twitter volume activity and sentiments with stock prices of various indices.

Ranco et al. (2015) focused on the connection between the volume of tweets and financial markets with analyses of future indicators of the stock market, like S&P 500. The research found a correlation between the increasing number of tweets in Twitter and stock market performance. The rapid volume of tweets may signal an increased level of market interest or even an increased level of awareness of a particular company, which may impact trading.

Our study equivocally agrees with the research which acknowledges the vastly complicated task of determining tweet sentiment and how important expert human annotation continues to be in achieving a high degree of accuracy which a Finbert model trained on a corpus of labelled dataset provides in contrast to other tools like AFINN, VADER and TextBlob.

In reality, though, intensity of tweet volume and sentiment both show correlations with stock price, this however never connotes causality. The sentiments conveyed through the tweets could statistically significantly be related to the stock's returns, especially if tested on shorter horizons or event window as the data visualization figures indicates.

Overall, this study suggests that while there is a relationship between Twitter activity (in terms of volume and sentiment) and stock prices, this can only be contextualized via the complexity's presence in social media communication, market psychology, and the specific context of the analysed data (Ranco et al., 2015).

The stock market is influenced by more factors beyond social media trends like tweet volumes and sentiments. Further work ahead includes exploring economic indicators like interest rates, supply-demand dynamics which are important drivers of investment idea and stock attractions. Changes in political events and policy, external factors like trade wars and natural calamities can sway investor confidence, affecting market volatility that impact economies and business operations. These influences are part of a broad array of datasets, sourced both internally and externally to any given organization. Therefore, understanding stock market movements requires considering these diverse and interconnected factors (Remesh, 2023).

REFERENCES

Androidkt.com (2021). Class Weight in CrossEntropyLoss for an imbalanced dataset. [online] PyTorch & Keras. Available at: https://androidkt.com/how-to-use-class-weight-in-crossentropyloss-for-an-imbalanced-dataset/ [Accessed 14 Dec. 2023].

Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. arXiv:1908.10063 [cs]. [online] Available at: https://arxiv.org/abs/1908.10063.

Azar, P.D. (2009). Sentiment Analysis in Financial News. www.academia.edu. [online] Available at: https://www.academia.edu/9826354/Sentiment_Analysis_in_Financial_News [Accessed 30 Nov. 2023].

Brownlee, J. (2020). Use Early Stopping to Halt the Training of Neural Networks At the Right Time. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/how-to-stop-training-deep-neural-networks-at-the-right-time-using-early-stopping/.

Eom, Y.-H., Puliga, M., Smailović, J., Mozetič, I. and Caldarelli, G. (2015). Twitter-Based Analysis of the Dynamics of Collective Attention to Political Parties. PLOS ONE, 10(7), p.e0131184. doi:https://doi.org/10.1371/journal.pone.0131184.

Fazlija, B. and Harder, P. (2022). Using Financial News Sentiment for Stock Price Direction Prediction. Mathematics, 10(13), p.2156. doi:https://doi.org/10.3390/math10132156.

Frost, J. (2020). Using Moving Averages to Smooth Time Series Data. [online] Statistics By Jim. Available at: https://statisticsbyjim.com/time-series/moving-averages-smoothing/.

Ganesh, A. and Iyer, S. (2021). Impact of Firm-Initiated Tweets on Stock Return and Trading Volume. Journal of Behavioral Finance, pp.1–12. doi:https://doi.org/10.1080/15427560.2021.1949717.

Halder, S. (2022). FinBERT-LSTM: Deep Learning based stock price prediction using News Sentiment Analysis. doi:https://doi.org/10.48550/arxiv.2211.07392.

Kusawa, S. (2023). What is Rolling Statistics in Time Series Analysis? - Data Magic AI Blog. [online] datamagiclab.com. Available at: https://datamagiclab.com/what-is-rolling-statistics-in-time-series-analysis/ [Accessed 12 Dec. 2023].

Malo, P., Sinha, A., Korhonen, P., Wallenius, J. and Takala, P. (2013). Good debt or bad debt: Detecting semantic orientations in economic texts. Journal of the Association for Information Science and Technology, 65(4), pp.782–796. doi:https://doi.org/10.1002/asi.23062.

Polamuri, S. (2023). 8 Most Popular Data Distribution Techniques - Dataaspirant. [online] Dataaspirant.com. Available at: https://dataaspirant.com/data-distribution-techniques/.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M. and Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. PLOS ONE, 10(9), p.e0138441. doi:https://doi.org/10.1371/journal.pone.0138441.

Remesh, A. (2023). Factors that Affect Stock Market Prices. [online] www.strike.money. Available at: https://www.strike.money/stock-market/price-factors [Accessed 14 Dec. 2023].

Smith, S. and 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). doi:https://doi.org/10.1186/s40537-022-00591-6.

Taborda, B., Almeida, A. de , Dias, J.C., Batista, F. and Ribeiro, R. (2021). Stock Market Tweets Data. ieee-dataport.org. [online] doi:https://dx.doi.org/10.21227/g8vy-5w61.

Wan, X., Yang, J., Marinov, S., Calliess, J.-P., Zohren, S. and Dong, X. (2021). Sentiment correlation in financial news networks and associated market movements. Scientific Reports, [online] 11(1), p.3062. doi:https://doi.org/10.1038/s41598-021-82338-6.

Yadav, A., Jha, C.K., Sharan, A. and Vaish, V. (2020). Sentiment analysis of financial news using unsupervised approach. Procedia Computer Science, 167, pp.589–598. doi:https://doi.org/10.1016/j.procs.2020.03.325.

Yang , Y., UY , M.C.S. and Huang, A. (2022). FinBERT: A Pretrained Language Model for Financial Communications. [online] ar5iv. Available at: https://ar5iv.labs.arxiv.org/html/2006.08097 [Accessed 11 Dec. 2023].