Analysis of Stock Market Trends and Sentiment on Twitter: Successful Individuals vs. General Public

Akanksha Cheeti, Meghana Chava

Department of Computer Science, Purdue University

I. INTRODUCTION

Social media platforms such as Twitter is found to be a valuable source for analyzing the public opinion and sentiment regarding various topics. Such analysis has been utilized before to examine the relationship between stock prices and market sentiment. However, it is interesting to explore the efficacy of tweets from different group, such as famous individuals and random individuals who have knowledge in technology but are not well known, and see how their sentiment affects the stock prices in separate groups (or) which group's sentiment aligns better with the stock market.

In this project, we are trying to explore the co-relation between social media sentiment and stock market trends for tech companies like Microsoft and Amazon. Our Analysis will be done separately for each company. We extracted relevant data using twitter API and analyzed the tweets during the period of 2022-2023 for Microsoft and from 2021-2023 for Amazon. We examined the impact of influential and random individuals social media sentiment on stock. We observed that there was a better co-relation between successful individuals tweets and stock prices compared to random individuals.

To further understand why a certain group's co-relation was better, we did some analysis on these tweets to identify the specific topics being discussed. Using BERT, we generated embeddings for the tweets and applied k-means clustering to identify the keywords, topics in the tweets, and also tweets were categorized into specific topics. Through this study, we found that for a company,successful individuals tweets were higher in count for topics related to company's performance, financial outlook, and business strategy compared to the random individuals.

Therefore, the analysis suggests that the properties of successful people tweets play a more significant role in the co-relation between social media sentiment and stock prices in the technology industry.

II. TECHNICAL DETAILS

The technical implementation will involve collecting tweets from twitter that contains specific keywords related to input stock ticker and the company name, with atleast few favorites, from Jan 2022 to April 2023 for Microsoft, and from Jan 2021 to April 2023 for Amazon. Next, we identified the most successful people list in the tech industry using articles like Forbes and Business Insider, and we manually found the twitter usernames for these people.Using these twitter usernames, we extracted data from the successful people usertimelines using keywords like #MSFT and microsoft for the Microsoft dataset, and and amazon for the Amazon dataset.

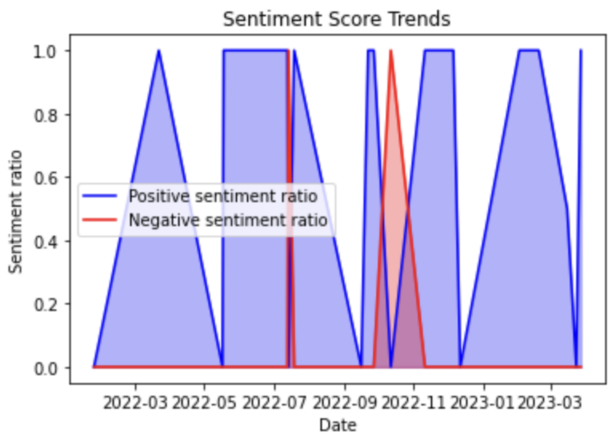

Figure 1: *Successful Category Sentiment Ratio- Microsoft*

Similarly, we collected the twitter userhandle's of random people by extracting the usernames that had tech related words in their bio. We scrapped the data from these user timelines in the same way we did for successful people dataset. This allows us to analyze the difference in corelation between social media sentiment and real-time stock market changes between these two groups during the period analyzed.

We then used NLTK for preprocessing the data, by removing punctuation's, stop words,tags,emoji's and regular expressions to remove urls, hashtags.

In the methodology to find the sentiment, we

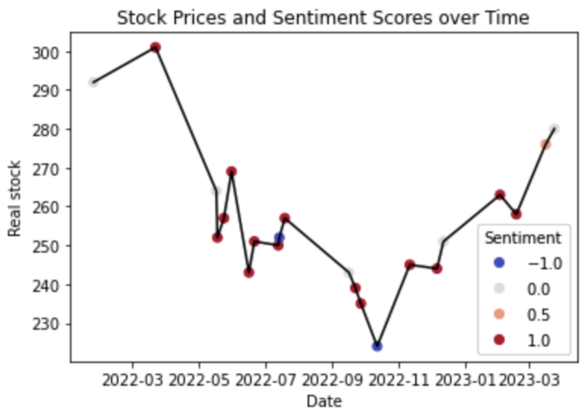

Figure 2: *Sentiment Score plotted against Stock prices*

used lexicon based approach using python libraries such as TextBlob and VADER method. These were used for calculating the sentiment scores and values for each tweet and nltk for tokenization and lemmatization. After getting the sentiment scores from both the models, we compared it with the real-time stock prices of the companies during the same period, using y finance package. We plotted the sentiment scores against the stock prices over time and analyzed the trend to see if there were any differences in the corelation between sentiment and stock among these two groups of people.
Next, to determine the specific topics that were discussed in the tweets, we used BERT to generate embeddings for the tweets and applied k means clustering to group them into topics. Then, we identified the most frequent keywords for each cluster. As an additonal analysis, we used an other method LDA for topic modeling and categorized tweets into specific topics. Finally, we plotted a histogram to visualize the distribution of tweets among few topics for successful and random individuals.
The obtained results suggest that the topics discussed is a property that affects the corelation between the sentiment expressed by successful people and random people towards stock prices.

III. EXPERIMENTAL EVALUATION

To start the experimental evaluation, we compared the positive and negative sentiment count among tweets extracted from successful and random tech people handles.

For the company Microsoft, in the successful people category, during the beginning of 2022, the positive sentiment count was higher and later around September to November 2022, there was an increase in negative sentiment count(shown in fig1). The trend in sentiment score we see in fig1 follows a similar pattern when compared to the stock price trend, with positive sentiment scores peaking at the highest stock price and decreasing as the stock price decreases.(shown in fig2). This observation suggests a possible corelation between sentiment and stock price trends.

Figure 3: *Random Category Sentiment Ratio- Amazon*

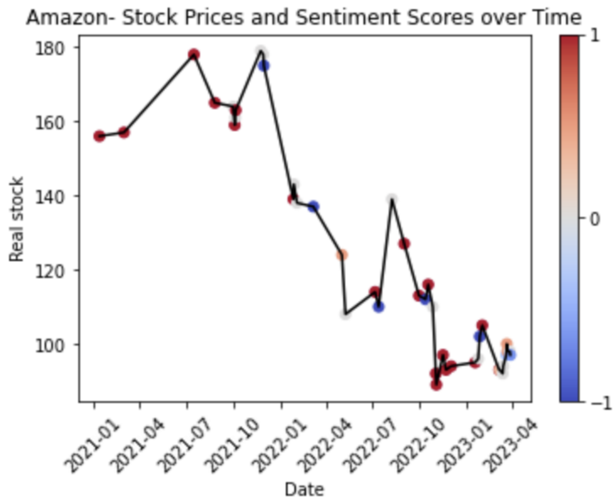

Figure 4: *Amazon-Random people stock price and sentiment*

For Amazon, a similar trend can be observed for the successful people group.
For the random people group, we conducted some similar experiments for both the companies and plotted graphs to observe the positive and negative sentiment count and plotted sentiment score against stock prices. However, unlike the successful people results, the trend here doesn't follow a similar distribution. The graphs for Amazon can be found in Fig3 and Fig4.
We quantified the corelation between sentiment and

stock trends using Pearson correlation coefficient. For the successful people category, we found a better positive corelation of around 0.4. However,the corelation for random people category was negative for both the companies. To understand, why is the corelation little better between tweets collected from successful people user timelines and stock prices than random people tweets, what might be the properties that affect these?

The difference in properties between successful

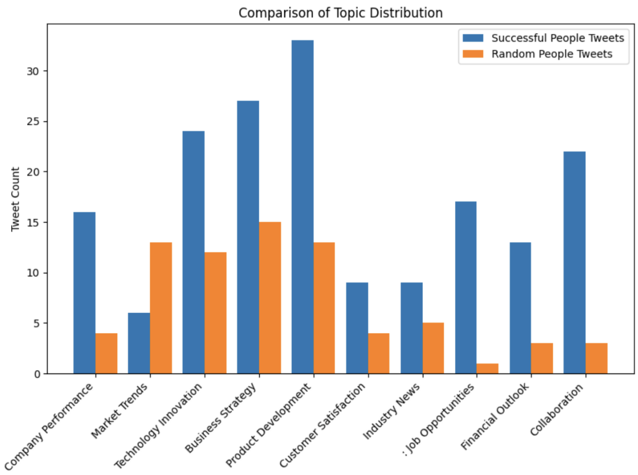

Figure 5: *Microsoft- Comparison of Topic Distribution in Tweets about Stock Market Trends for Successful People and Random People*

people and random people tweets towards stock prices could be due to many factors. To get a better understanding of the factors that affect the difference in corelation, we did further analysis of the tweets to identify the specific topic that were discussed. After analyzing the tweets, we found that there is a difference in distribution in topics that were discussed in these two groups. To identify the topics, we used BERT model to get the embeddings of tweets on both the datasets for a company, and applied k means with k=10 to cluster topics and extracted frequent keywords for a topic. As a next step, we plotted a histogram, to see how the count differs for 10 specific topics like financial outlook,company performance. By categorizing the tweets among these topics, we found that successful people had a higher count of tweets related to these topics like company performance, financial outlook and business strategies, while random people had less count, their topics or keywords were more related to daily market trends, current events. The difference in distribution of these topics (shown in Fig 5.) between these two groups could explain why successful people tweets have a better corelation with stock trends compared to random people tweets.

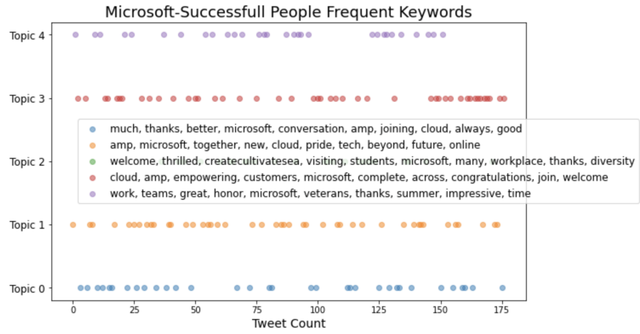

Figure 6: *Topic cloud*

IV. RESULTS

For Microsoft:
a. Analyzing Successful people corelation with stock market.
b. Analyzing Random people corelation with stock market.

For Amazon:
c. Analyzing Successful people co relation with stock market.
d. Analyzing Random people co relation with stock market.

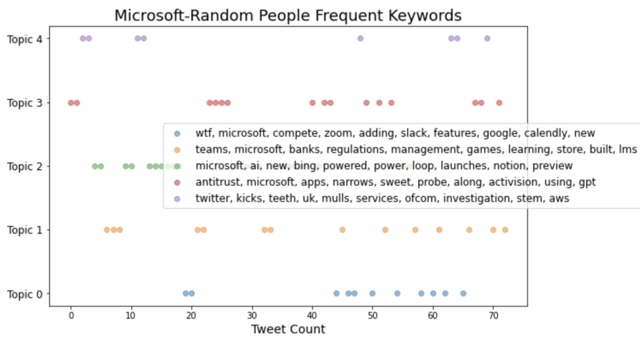

Figure 7: *cloud*

1. Analyzing Successful people corelation with stock market.
Based on the analysis, we got a moderate corelation between sentiment on twitter and stock trends. However, when we extracted the datasets using different timelines by changing the period, the co relation kept changing. When we included some more username handles, the tweet samples increased and the corelation increased very slightly. So various parameters like timezone and userhandles being used affected the results.

The corelation was not strong enough hovering around 0.4,besides the factors that were mentioned above, there may be other factors that could affect the stock market.

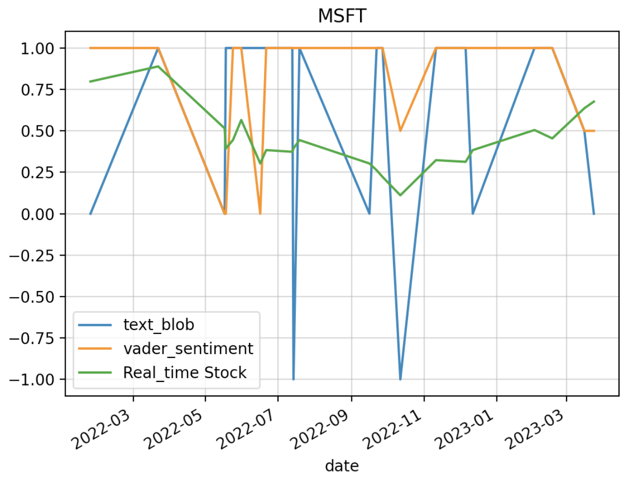

Figure 8: *MSFT-Successful people Sentiment Analysis*

2. Analyzing the Correlation between Random People's Sentiment and Microsoft's Stock Trends

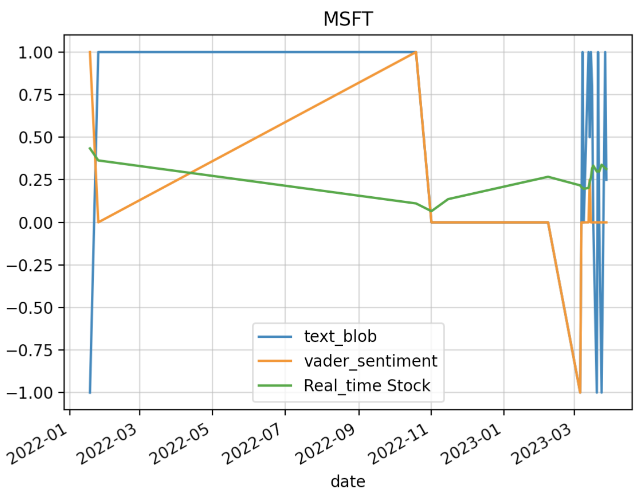

Figure 9: *MSFT-Random people Sentiment Analysis*

The trend seem to be better for successful people when compared to random people as the graph does not follow any pattern between sentiment and stock price for Microsoft.

3. Analyzing the Correlation between Successful People's Sentiment and Amazon's Stock Trends

With respective to amazon, for the successful people category, if we look at the graph at time period 2021-23, the text blob sentiment has a negative sentiment and stock price is also decreasing, seems like the vader sentiment did not do well for this case as its not aligned at in the beginning but after May 2022, the vader sentiment and stock both are decreasing.

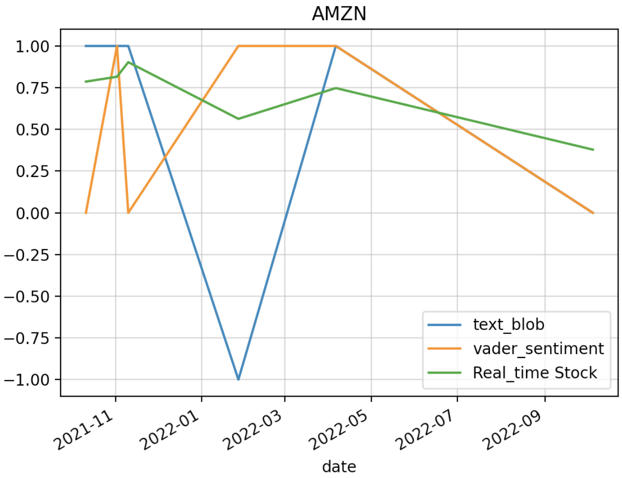

Figure 10: *Amazon-Successful people Vader and Text Blob Sentiment Analysis*

4. Analyzing the Correlation between Random People's Sentiment and Amazon's Stock Trends

The co relation doesn't seam to be relevant it appears random and the co relation values are also negative.

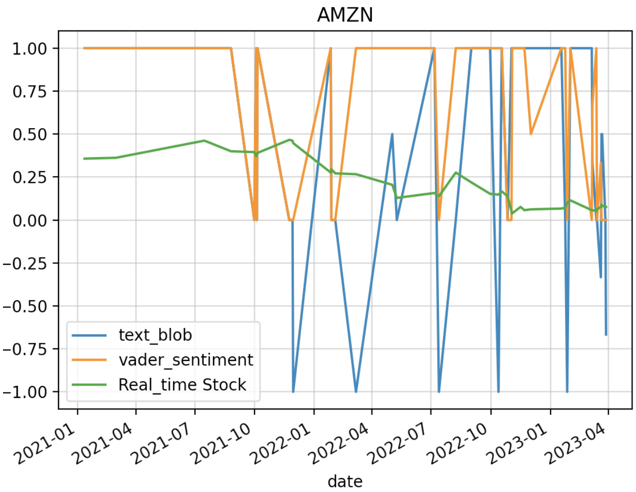

Figure 11: *Amazon Random people-Sentiment and stock anaylsis*

V. LIMITATIONS AND FUTURE WORK

There are limitations to this study that needs to be acknowledged. Firstly, our project only focused on two companies, it may not generalize well for all the companies. Second, we only analyzed tweets from specific time period, and only took around 30 user handles from forbes list. The topics discussed, the sentiment of it may change over time.

Future research could include, other factors like the news articles or financial reports that could provide a better understanding of the stock market trends. Additionally, a more in depth analysis of the sentiment expressed on twitter could be done using more advanced NLP techniques. To expand the study, we could include a larger sample of companies, individuals and also use other social media platforms like reddit or stocktwits which might provide a more comprehensive understanding of the corelation.

VI. CONCLUSION

We conducted an analysis on tweets related to companies Microsoft and Amazon from two categories of people over 2021-2023. We examined the corelation of market sentiment and the stock market between two groups successful and random and found that successful people sentiment had a better corelation when compared to random. To further examine the properties/factors, that is affecting the corelation we categorized tweets by topics and found that successful people had more tweet count related to company performance and financial outlook where as random people tweet count was comparatively less for these topics and their tweets revolved around daily market updates.

REFERENCES

1. Pagolu, Sasank Reddy, Kamal Panda, Ganapati Majhi, Babita. (2016). Sentiment analysis of Twitter data for predicting stock market movements. 1345-1350. 10.1109/SCOPES.2016.7955659.
2. Pak, A., Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No, pp. 1320-1326).
3.Deng, S., Mitsubuchi, T., Shioda, K., Shimada, T., Sakurai, A. Combining technical analysis with sentiment analysis for stock price prediction.