Social Media Based Stock Market Analysis Using Big-Data Infrastructure

Vishakan Subramanian
Dept. of Computer Science and Engineering
SSN College of Engineering
Chennai, India
vishakan18196@cse.ssn.edu.in

Venkataraman Nagarajan
Dept. of Computer Science and Engineering
SSN College of Engineering
Chennai, India
venkataraman18192@cse.ssn.edu.in

Shashanka Venkatesh
Dept. of Computer Science and Engineering
SSN College of Engineering
Chennai, India
shashanka18145@cse.ssn.edu.in

Dr. Sujaudeen N
Dept. of Computer Science and Engineering
SSN College of Engineering
Chennai, India
sujaudeenn@ssn.edu.in

***Abstract*—Several factors influence the value of a stock apart from the typical quantitative and qualitative parameters seen in the fundamental analysis of stocks like balance sheets, income statements, cash flow statements etc. In recent years, one such factor that has gained prominence is social media trends. In this paper, we aim to study the correlation between social media trends and stock market movement with the usage of a big-data architecture. This work considers the sentiments expressed by Twitter users on relevant topics, and measures their correlation to stocks of companies within the relevant sector the Tweet appeals to. For efficient processing of such large-scale data, we use Apache Kafka for data ingestion and Apache Spark for Tweet data processing, i.e. - sentiment extraction and aggregation. Based on observed correlations, we discuss how to predict stock market movement based on the extracted Tweet sentiment data.**

***Index Terms*—Stock Market Analysis, Big-Data, Social Media, Sentiment Analysis, Machine Learning, Twitter, Apache Spark, Apache Kafka**

I. Introduction

Stock market analysis has been a topic of great interest ever since public stock exchanges came into existence. It is an avenue that offers great profit, which by itself is a stimulus for most researchers who work in this realm. The task is also considerably demanding due to these returns as well as high randomness and various external influences affecting the current valuation of a stock [4]. In recent years, social media has transformed in a revolutionary manner from a niche utility to a ubiquitous means of information exchange throughout the world. People have started utilizing social media as a tool not only for socialising, but also as a means to consume news, share their opinions, create awareness, follow the ideas of popular personalities and luminaries, among many others. The architecture of social media provides the avenue for interesting ideas to disseminate among society in near-instant times. In such an age, such sudden outbursts of information about a company or its product can make or break the valuation of that product or company - since information travels fast and people tend to react towards sensational articles. To highlight the severity of social media trends towards the stock market, we could consider the example of the Tweet posted by Elon Musk, the CEO of Tesla Motors, on 27 Jan 2021 towards a declining video-game retail company, GameStop. His Tweet about the company immediately raised the company's stock value by more than 60% within hours. Such is the influence of social media on an already volatile stock market. Hence, it is worthwhile to consider understanding the relationship between social media trends and the stock market.

Selvamuthu et. al [11] also express the opinion that sentiment analysis of the opinions of renowned personalities might help get an extra edge in stock price prediction, since their opinions are known to affect stock prices.

In this work, we analyse the statistical correlation between social media trends and its effect on the relevant industry's stock price valuations using a big-data architecture consisting of Apache Kafka and Apache Spark, since stock data and social media articles tend to be voluminous in nature.

II. Related Work

Lee et. al.'s [1] research delves into analyzing Twitter data, by classifying it into relevant industrial sector categories and performing sentiment analysis to predict the trend of stock prices, and comparing it to reality, to consequently find the correlation between the two. Their publication states that they used a sample of 1000 news articles to create a classifier. They were able to obtain a 77% accuracy on the classification of Tweets into different sectors. They have used 100 Tweets overall, with 20 Tweets per category in order to test out their hypothesis.

Mittal et al. [7] presented a research that utilized Twitter to capture and predict public mood, and use it along with previous stock data to predict the future movement of the stock. They used the DJIA (Dow Jones Industrial Average) value as their stock index under study. They have proposed an algorithm to approximate stock market data on days in which the market was closed due to public holidays and weekends. They carried out stock price prediction using several machine learning algorithms.

Nayak et. al [8] have done significant work on sentiment analysis of Twitter data for stock price prediction using trend analysis of stock data obtained from Yahoo Finance for three sectors: Banking, Mining & Oil and sentiment data extracted from relevantly collected Tweets. They have put forth algorithms to calculate closed price trends for each day, to check continuous days up/down for a stock, and an algorithm to combine sentiment with historical data. They also predict stock market movement based on available historical stock data and extracted social media sentiment data using 3 different supervised machine learning models, and obtain the best results using a Boosted Decision Tree model. Kalyanaraman et. al's [13] research tries to perform such an analysis purely using news articles. They considered 11 companies under the National Stock Exchange (NSE), and considered 100 news articles for each company, and manually pre-processed the articles to remove irrelevant ones. They used a custom sentiment dictionary to build a classifier of news articles using Linear Regression and Gradient Descent. Their model predicted the sentiment of articles with around 60% accuracy and the direction of stock price movement (positive or negative) with an accuracy of 81.92%. Peng uses a big-data architecture using Apache Spark [9] and Apache Hadoop to predict US Oil stock prices.

Kanavos et al. [4] have used the Stanford Core NLP library to perform text pre-processing along with Naive Bayes classifier to perform sentiment analysis on Twitter data for stock price predictions. Their classification was done on 6 dimensions, which were Alert, Calm, Happy, Kind, Sure & Vital. Pagolu et al. [10] performed correlation analysis between Tweets related to Microsoft and its stock price movements during the time period of August 2015 - August 2016. They transformed correlation analysis into a classification problems with the input features being the total negative, neutral and positive emotions in Tweets observed in a 3 day period. They utilized the methodologies of Word2Vec for sentiment analysis and Logistic Regression & SVM to perform the classification task, which yielded an accuracy of around 70% to predict the stock movement as an upward or downward trend.

Mehtab et al. [12] further builds upon the work done by Mittal et al. and developed eight regression and eight classification models for predicting the stock price movement of NIFTY 50, which is augmented using public mood data which was analysed and aggregated from relevant Tweets.

III. Design

Initially, Tweet and stock data are collected. They are then pre-processed before being streamed through Apache Kafka [20]. Following this, Apache Spark [19] is used to perform sentiment analysis on Tweets and aggregate the results date-wise. Finally, correlation analysis is performed on the aggregated Tweet data and stock data for various market sectors. Based on the observed correlation, machine learning models are then built to predict stock market movement based on observed Tweet sentiments.

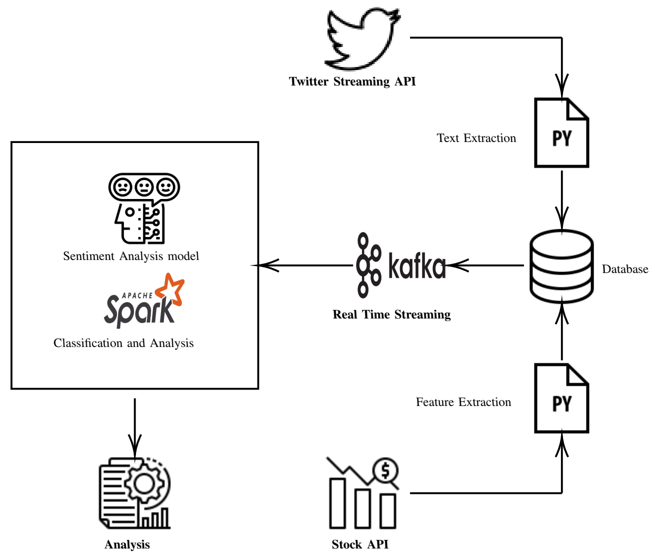


Fig. 1: Model Architecture

This architecture was built based upon the analysis of different tools available for various tasks involved in a big-data processing lifecycle, as researched by Gürcan et al [3].

An overview of the architecture is shown in Fig. 1.

IV. Data Collection and Preprocessing

A. Data Collection

The official Twitter API with the Academic Research access [17] is used to collect archival Tweet data. Polygon.io [18] is used to obtain 2 year historical stock market data. Tweets along with its retweet count are collected for certain market sectors based on appropriately chosen keyword(s) like #MSFT, #TSLA etc. and stock market data is collected for a representative company present in the corresponding industrial sector. The representative company of each sector was chosen based on it's market share in the sector and presence in the USA, since the majority of Twitter users are American (nearly 77M as of Jan 2022). Table I summarizes the volume of Tweet data collected and used in each sector.

TABLE I: Data Points Count

Sector	Data Points
Electric Vehicles (EVs)	116,355
Oil	53,253
Technology (Tech)	632,181
Pharmaceuticals (Pharma)	313,407

B. Data Preprocessing

The collected JSON data from the Twitter API is cleaned to remove emoticons and extraneous characters. Links and other irrelevant data are removed from the Tweet, if any. Duplicate Tweets are also removed. Upon analysing the distribution of the collected Tweets, a high left skew is observed since most

Tweets do not have high retweet counts. To filter out noise, Tweets with a very low retweet count are removed and Tweets with a high retweet count (outliers) are manually capped to a fixed retweet count to avoid bias and skew in the data.

Using Mittal et. al's [7] approximation algorithm, stock market data for missing days are calculated and imputed. Finally, all the preprocessed data is stored in a database for streaming through Apache Kafka.

C. Sentiment Analysis

A crucial component in the pursuit of our study to understand the relation between social-media (Tweets) and the stock market is analysing the sentiment (or mood) reflected by social media at any given date. To perform sentiment analysis on Twitter data, various NLP methods have been explored, as discussed in Section II. In the field of Natural Language, one of the recent developments has been the introduction of the BERT (Bidirectional Encoder Representations from Transformers) Model [2]. BERT differs from conventional NLP language models in the fact that it uses bidirectional training and attention mechanism in order to have a deeper sense of language context compared to other unidirectional models. Since its introduction, the BERT model has been extensively used for popular NLP tasks like Text Summarization, Question-Answering, Text Classification, Sentiment Analysis etc.

In our methodology, we chose to use Cardiff NLP research group's publicly available TimeLMs based sentiment analysis model [14]. This model builds upon the existing RoBERTa-base model, which in turn is a BERT-base model that has been pre-trained on a large corpus of English data using the training approach as discussed in its research paper [5]. This sentiment analysis model has been pre-trained on a dataset of approximately 124M Tweets collected between Jan 2018 - Dec 2021. To add further detail, their research aims at building language models (LMs) that specializes in understanding diachronic Twitter data [6]. Their model has been bench marked against the TweetEval task, and the results can be checked in their paper. We expect this model to perform optimally for our use-case, since it has been pre-trained on Tweet data that is temporally close to our collected Tweets, which translates to better understanding of the context of Twitter English language used during that period.

The model classifies a Tweet into three sentiments - Negative, Neutral and Positive, and returns an array of three confidence values which correspond to the probabilities of the Tweet being classified into the corresponding sentiments.

Thus, utilizing this transformer model, all Tweets available for a given day and a given sector are scored and aggregated by the Spark Processing Engine. With the use of Apache Spark and its fundamental RDD (Resilient Distributed Datasets) data structure, we are able to take advantage of parallelization to process Tweets efficiently. The overall mood observed on Twitter on a given day for each market sector is thus captured and quantified.

V. Correlation Analysis

1) Methodology: After processing the entire stream of Tweet data and obtaining intermediate results based on the sentiment scores, correlation analysis is performed.

The aggregated sentiment scores obtained from the Spark Engine is grouped sector-wise and merged with its relevant stock market ticker data, according to a date-wise order (ref. Table II). For example, the aggregated Tweet results obtained under the EVs (Electric Vehicles) category are merged with the stock market data of Tesla Motors (NASDAQ: TSLA). In essence, we now have a combined dataset that describes the stock market performance of a given stock and the social media mood captured on Twitter on that day for the stock's relevant market sector.

TABLE II: Description of dataset schema

Attribute	Description
Category	Market sector under consideration
Date	UTC date that the Tweet belongs to
Open	Opening price of the stock on the given date in ($)
Close	Closing price of the stock on the given day in USD ($)
Wted_Neg	Daily aggregated negative sentiment score weighted on retweet count
Wted_Neu	Daily aggregated neutral sentiment score weighted on retweet count
Wted_Pos	Daily aggregated positive sentiment score weighted on retweet count

We then perform feature scaling, i.e. normalization of the data using min-max scaling, shown in (2). This is done since different features present in the dataset have different ranges, and it is necessary to bring them all to the same scale, to avoid bias and to perform further analysis.

$$y' = \frac{y - x_{min}}{x_{max} - x_{min}}(x'_{max} - x'_{min}) + x'_{min} \tag{1}$$

Typically $\langle x'_{min}, x'_{max} \rangle = \langle 0, 1 \rangle$.

$$y' = \frac{y - x_{min}}{x_{max} - x_{min}} \tag{2}$$

where:
$\langle x_{min}, x_{max} \rangle$ is the old range.
$\langle x'_{min}, x'_{max} \rangle$ is the new range.
$y \in \langle x_{min}, x_{max} \rangle$ is the value to be normalized.
$y' \in \langle x'_{min}, x'_{max} \rangle$ is the min-max normalized value.

The next step is to perform correlation analysis with the Spearman's Correlation Coefficient, which is a non-parametric measure of rank correlation between two variables. It is equal to the Pearson's Correlation Coefficient between the rank values of those two variables. While the Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (irrespective of linearity). The confidence levels of each correlation score are also calculated, to accept/reject the Null Hypothesis (whether the correlation is significant or not). It is a better correlation metric than the Pearson's correlation metric for this specific scenario since the stock market may not be linearly related to the sentiments observed on social media.

A statistical test of significance is also performed for the observed Spearman's correlation coefficient values. The correlation value is considered statistically significant (i.e. the alternate hypothesis is accepted) if the p-value is lower than 0.01 (for a 99% confidence interval). Otherwise, the correlation is statistically insignificant, and can be ignored for all practical purposes.

$$\rho = \frac{cov(R(X), R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}} \tag{3}$$

where:

ρ is Spearman's coefficient, $\rho \in \langle -1, 1 \rangle$
$cov(R(X), R(Y))$ is the covariance of the rank variables.
$\sigma_{R(X)}$ and $\sigma_{R(Y)}$ are the standard deviations of the rank variables.

If all n ranks are all distinct, it can be computed using the formula:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{4}$$

where:

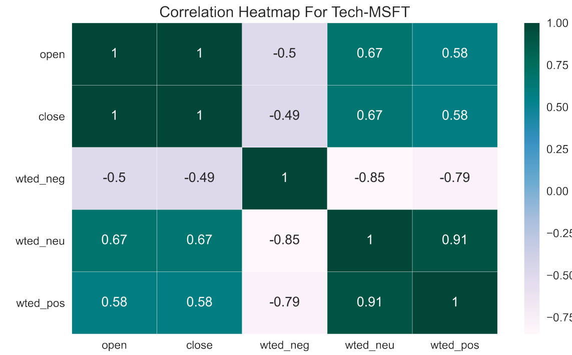
ρ is Spearman's coefficient, $\rho \in \langle -1, 1 \rangle$
n is sample size.
d_i is the difference between the ranks of the data being analyzed.

This procedure is carried out for each aggregated sector data separately to observe and tabulate the correlation between that sector's stock market trend and its' respective social media mood.

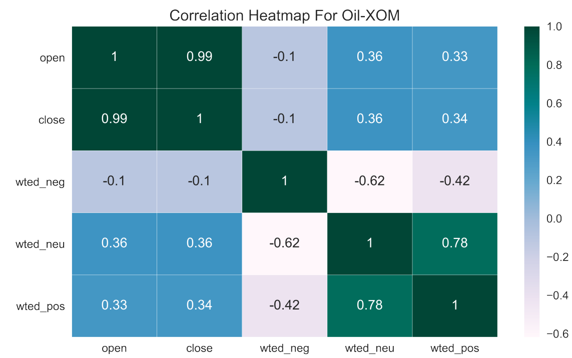
2) Results Obtained: For our analysis, we consider only the daily close value to understand the correlation between a sector's market performance and the corresponding day's social media trend.

Upon category-wise examination of the obtained results shown in Fig. 2, we see that there is a strong correlation between the weighted neutral sentiment score (Wted_Neu) and Close (0.66), as well as between the weighted positive sentiment score (Wted_Pos) and Close(0.59) for the Tesla (EVs) dataset (Fig 2c). There is a moderate negative correlation between the weighted negative sentiment score (Wted_Neg) and Close (-0.31). These results can be interpreted as follows: The neutral sentiment's result indicates that the Tesla stock performs well with a rise in the amount of discussion about electric vehicles in general. An overall positive mood about the EV market correlates to the appreciation of Tesla Inc.'s stock market value. An overall negative sentiment, however, does not translate as much to depreciation of the market value, due to its moderate correlation.

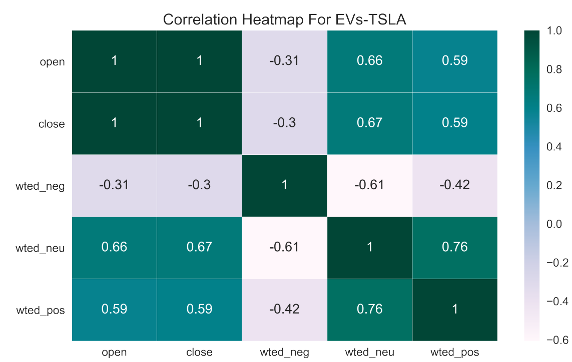
Upon examining the Oil sector tweets with the Exxon Mobil. Corp stock performance (Fig 2b), there is a moderate positive correlation between the Wted_Neu (0.36) and Wted_Pos (0.34) to the closing price. The weak correlation between the Wted_Neg and the closing price (-0.1) indicates that the oil market does not fluctuate if there are negative sentiments being expressed about it on Twitter - making it a lowrisk stock in that aspect. But the market closing price shows considerable appreciation with positive/neutral discussion about it on the social media platform, although not being very susceptible to social media trends.



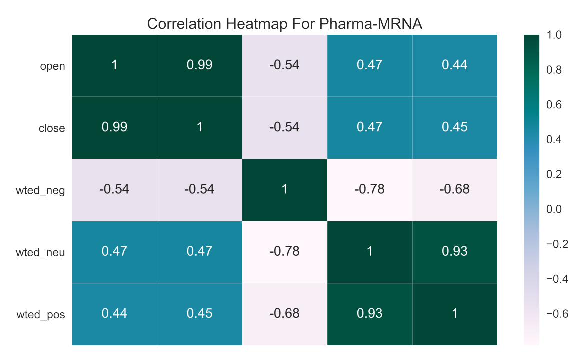
(a) MSFT



(b) XOM



(c) TSLA



(d) MRNA

Fig. 2: Correlation heatmap between Tweets and Stock Market for different industrial sectors

For the Technology sector, and its reference stock market entity - Microsoft Corp., a strong positive correlation was obtained between the closing price and Wted_Neu (0.67) and Wted_Pos (0.58), respectively (Fig 2a). A medium negative correlation was obtained between the closing price and Wted_Neg (-0.49) attribute. These results indicate that there is a good correlation between general social media discussion of the Tech sector and the stock market in both ways, indicating that the technology market is volatile to the mood of social media.

In the Pharmaceutical sector, with its stock market representative - Moderna Inc., it was observed (Fig 2d) that there is a medium positive correlation between the three weighted attributes - Wted_Neg (-0.54), Wted_Neu (0.47) & Wted_Pos (0.45). This indicates that the Pharma sector's market behaviour is also volatile to the sentiments expressed on social media, especially on the negative side - a negative mood correlates more to the pharmaceutical sector's market price going down.

VI. Prediction

1) Methodology: Following our primary objective of performing analysis, we decided to additionally perform prediction of the next-day's stock market closing value using machine learning algorithms, since we observed a good correlation between the weighted Tweet sentiment scores and the close price of the market. From Nayak et. al's [8] research on stock market prediction augmented by sentiment analysis of Twitter and news, it can be seen that Boosted Decision Trees perform the best in this scenario when compared to other algorithms like Logistic Regression and Support Vector Machines. Thus, we decided to compare and contrast two widely-used decision tree algorithms that employ the concept of boosting to predict the close value of the above used stock listings, augmented by the weighted Tweet sentiment scores.

The first boosting algorithm under consideration is DMLC's XGBoost (eXtreme Gradient Boosting) [15] which was presented in 2014 as an optimized gradient-boosting library that is highly efficient. From its inception, it has gained universal acclaim and usage in the data science community. The second algorithm under consideration is Yandex's CatBoost [16], also a gradient-boosting algorithm introduced in 2017 with native support of categorial features and GPU usage for faster training times.

Optimal hyper parameters for the models were found with the usage of a grid search algorithm. Separate regressor models were constructed for each market sector for analysis. The datasets were split in an $85 : 15$ ratio for training and testing purposes. Their results are highlighted in Tables IV & III.

A. Results

It was observed that both models were able to capture the trend of the closing price movement (increase/decrease), and were able to predict the next day's closing price with high accuracy. The XGBoost model performs best $(83.81\%$ accurate) on the Tech-MSFT dataset, which follows from

TABLE III: CatBoost Regressor Metrics

(a) (NASDAQ: MSFT)

RMSE	0.0480
MSE	0.0023
MAE	0.0383
R2 Score	0.8319
Explained Variance Score	0.8427
Max Error	0.1288

(b) (NYSE: XOM)

RMSE	0.1962
MSE	0.0385
MAE	0.1457
R2 Score	-0.7114
Explained Variance Score	0.1810
Max Error	0.4340

(c) (NASDAQ: TSLA)

RMSE	0.0711
MSE	0.0050
MAE	0.0562
R2 Score	0.5328
Explained Variance Score	0.6047
Max Error	0.2063

(d) (NASDAQ: MRNA)

RMSE	0.0337
MSE	0.0011
MAE	0.0251
R2 Score	0.8461
Explained Variance Score	0.8642
Max Error	0.1491

TABLE IV: XGBoost Regressor Metrics

(a) (NASDAQ: MSFT)

RMSE	0.0471
MSE	0.0022
MAE	0.0377
R2 Score	0.8381
Explained Variance Score	0.8427
Max Error	0.1403

(b) (NYSE: XOM)

RMSE	0.1914
MSE	0.0366
MAE	0.1420
R2 Score	-0.6285
Explained Variance Score	0.2131
Max Error	0.4236

(c) (NASDAQ: TSLA)

RMSE	0.0562
MSE	0.0031
MAE	0.0463
R2 Score	0.7086
Explained Variance Score	0.7555
Max Error	0.1630

(d) (NASDAQ: MRNA)

RMSE	0.0366
MSE	0.0013
MAE	0.0255
R2 Score	0.8185
Explained Variance Score	0.8370
Max Error	0.1885

the observed high correlation between its weighted sentiment scores and its closing price. The observed metrics are better in all cases for the XGBoost model, except for the Pharma-Moderna dataset, for which the CatBoost model showed the best performance (84.61% accuracy). Thus, we believe that using either model for this task would be prudent, although there is a slight advantage to using XGBoost in terms of reliability, as evidenced by our results.

VII. Conclusion

In our study, we have undertaken four stock market segments - Electric Vehicles, Oil & Gas, Technology and Pharmaceuticals - and four representative stock market entities - Tesla Inc., Exxon Mobil Corp., Microsoft Corp., Moderna Inc. to analyse the correlation between the sentiments observed in social media to the respective entity's stock market valuation.

In this paper, we have shown conclusive proof that there is a strong correlation between the mood expressed on social media to stock market behaviour, although the extent of the correlation varies for each market sector. We were also able to identify two candidate machine learning algorithms - CatBoost and XGBoost, that can be utilized for efficient stock market prediction using the sentiment values. Our main contribution in this area of research has been the incorporation of state-of-the-art NLP models for sentiment analysis, and a more fine-grained approach to correlate stock market listings to related social media data, instead of a generalized approach that performs mood analysis to stock market indices like the DJIA, as done by related authors such as Lee et al. [1] and Mittal et al. [7]. We also used a significantly larger dataset to broaden the sample space under consideration, and the results we obtained show great promise in terms of practical use and further research.

Finally, we have also shown that stock market prediction based on sentiment-analysis of social media data is a worthwhile methodology that can be used to bolster existing stock prediction algorithms that make use of other features.

VIII. Future Work

With regard to future work, the inclusion of data from other social media websites like Facebook, Reddit etc. can present an even broader picture of the sentiment expressed by users online. The inclusion of the sentiments reflected in news articles could also prove to be more effective in trend analysis of the stock market. Thus, data from multiple sources can be pooled together and aggregated with big-data architecture. Another consideration is the fact that the social media data that we work with are primarily produced by the English-speaking audience. Similar NLP techniques that work with other languages would be required to perform similar analyses in non-English speaking countries. With the inclusion of non-English Tweets (and other social media posts), it is possible to obtain a higher correlation score. The stock market data prediction task can further be improved and made reliable with the addition of other features that affect the stock market, apart from social media. To broaden the scope of this research, many other market sectors can also be considered to perform such sentiment-based analysis and prediction.

References

[1] Chungho Lee, Incheon Paik, *Stock Market Analysis from Twitter and News Based on Streaming Big Data Infrastructure*

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

[3] Gürcan, Fatih & Berigel, Muhammet, Real-Time Processing of Big Data Streams: Lifecycle, Tools, Tasks, and Challenges, 2018 2^{nd} International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)

[4] Kanavos, Andreas & Vonitsanos, Gerasimos & Mohasseb, Alaa & Mylonas, Phivos. (2020). *An Entropy-based Evaluation for Sentiment Analysis of Stock Market Prices using Twitter Data*

[5] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov, *RoBERTa: A Robustly Optimized BERT Pretraining Approach*

[6] Loureiro D., Barbieri F., Neves L., Anke L. & Camacho-Collados J. (2022). *TimeLMs: Diachronic Language Models from Twitter.*

[7] Mittal, A. (2011). Stock Prediction Using Twitter Sentiment Analysis.

[8] Aparna Nayak, M. M. Manohara Pai and Radhika M. Pa, *Prediction Models for Indian Stock Market*

[9] Zhihao Peng, Stocks Analysis and Prediction Using Big Data Analytics, 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)

[10] Pagolu, Sasank & Reddy, Kamal & Panda, Ganapati & Majhi, Babita. (2016). *Sentiment analysis of Twitter data for predicting stock market movements.*

[11] Dharmaraja Selvamuthu, Vineet Kumar and Abhishek Mishra, *Indian stock market prediction using artificial neural networks on tick data*

[12] Mehtab, Sidra & Sen, Jaydip. (2019). A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing. SSRN Electronic Journal. 10.2139/ssrn.3502624.

[13] Vaanchitha Kalyanaraman, Sarah Kazi, Rohan Tondulkar, Sangeeta Oswal, Sentiment Analysis on News Articles for Stocks.

[14] *Twitter-Roberta,*https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest

[15] Tianqi Chen and Carlos Guestrin. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, (785–794).

[16] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. (2018). CatBoost: unbiased boosting with categorical features. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18). Curran Associates Inc., Red Hook, NY, USA, (6639–6649).

[17] *Twitter-API,* https://developer.twitter.com/en/products/twitter-api

[18] *Polygon-API,* https://api.polygon.io/v1/open-close/

[19] *Apache Spark,*https://www.apache.org/dyn/closer.lua/spark/spark-3.2.1

[20] *Apache Kafka,* https://kafka.apache.org/downloads