International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Stock Prediction using GAN and Sentiment Analysis

Shubham Shah, Prachi Patel, Riya Bihani, Dhruv Gada, Dr. Kranti Ghag
Department of Computer Engineering
Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
shahshubham024@gmail.com, prachipatelp09@gmail.com, riya.bihani7492@gmail.com, drg020502@gmail.com, kranti.ghag@djsce.ac.in

Abstract— Tweet sentiment analysis can be used to forecast stock market movements. Examining the attitude expressed in tweets about a company or industry might reveal more about how the public perceives and thinks about it. This information can then be utilised to forecast potential changes in the stock price of the firm or industry. This paper proposes a method for forecasting the stock market based on sentiment analysis of Twitter-sourced messages. To create the dataset, information about various companies via the Twitter API is mapped and linked with their stock prices. This paper employs a Generative Adversarial Networks-based framework that integrates three GAN models with Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). To evaluate and compare the efficacy of the proposed model, the Root Mean Square Error (RMSE) is employed as a metric. The stock price and sentiment score of each company were individually assessed to determine which of the three GAN models generated the most accurate outcomes.

Keywords- Twitter, Stock market, GAN, Tweets, Sentiment analysis, LSTM

I. Introduction

Social networking and content sharing platforms like Twitter and Facebook have permeated every aspect of the internet. These websites generate an enormous volume of messages on a daily basis, providing researchers an unprecedented chance to utilize the public opinions expressed within them for a wide variety of applications. One such application is analyzing the stock market based on tweets. The widespread adoption of Twitter and user-friendly open Application Programming Interfaces (APIs) have spawned numerous studies that utilize the platform for analyzing and predicting stock market movements. Stock market predictions has been considered as one of the highly challenging and essential tasks due to its nonlinear or dynamic behavior as mentioned in [1] by Albahli et. al. Numerous factors influence the rise and decline of stock prices, causing them to fluctuate constantly. This can include eco- nomic indicators, company performance, international events, and even social media trends. Currently, there is an ongoing discussion regarding the reliability of social media sentiments in predicting stock market fluctuations. The Efficient Market Hypothesis (EMH) states that stock market prices are largely driven by new information and follow a random walk pattern mentioned in [2] by Anshul Mittal and Arpit Goel. Addition- ally, a number of studies suggest that these sentiments can influence market movements and serve as useful predictors of trade-off outcomes. For a variety of reasons, sentiment analysis of tweets about stocks can be extremely important. The market's impression of a stock can first be determined by examining the mood of tweets regarding that stock. Positive sentiment might indicate an upbeat view while negative feeling can indicate one that is pessimistic. Second, sentiment analysis could help in spotting new stock market tendencies. Investors may be able to gain a deeper grasp of market mood and make better investing decisions by analyzing tweets in real-time. Third, sentiment analysis can be used to spot possible cases of deception. Unusual patterns or dubious behavior that may be a sign of insider trading or market manipulation can be found by examining tweets about a specific stock.

Traditional sentiment analysis methods may find constraints when assessing tweets due to sarcasm, irony, ambiguity, and context. Incorrect sentiment classification occurs frequently as a result of the difficult-to-understand language used in tweets. It does not take into account other key factors that can affect the stock market, such as economic indicators or international events. This paper employs Generative Adversarial Networks (GANs) to overcome these limitations by generating synthetic data that encompasses a broader spectrum of factors. More- over, GANs can learn to capture the nuances of language and context, enabling them to make more accurate sentiment predictions.

The objective of this paper is to predict stock market movements by analyzing recent tweets. The initial dataset is comprised of tweets retrieved from the Twitter API between January 1, 2021 and September 30, 2022. VADER, a pre- built machine learning tool for natural language processing in Python specifically intended to analyze social media text, is used to analyze the sentiment of each tweet. VADER con- siders the distinctive characteristics of social media language, such as the use of vernacular, emoticons, and abbreviations. Furthermore, a second dataset, known as the Stock Market Twitter Sentiment Analysis (SMTSA) dataset, is created which includes the stock prices for each company gathered from the previously mentioned dataset.

The stock price was mapped to the previous dataset, and the two datasets were merged to form the final SMSTA dataset used in this paper.

The rest of the paper is organized as follows. The second section provides a brief overview of previous approaches adopted to solve the problem. In Section 3, the methodology is discussed and it describes the dataset used for this paper and data preprocessing measures adopted. It also includes the sentiment analysis technique, Generative Adversarial Networks, developed for the purpose of this paper. Section 4 includes the results obtained from the implementation of the proposed model. Finally, in Section 5, the conclusions drawn from the paper are mentioned and highlight potential areas for future research.

II. PREVIOUS WORK

The role of investor sentiment in defining stock market dynamics has been the topic of significant attention and debate in finance, as it has the capacity to influence market participants' behavior and, ultimately, stock prices. Related work to this field for using public sentiment for predicting stock prices and forecasting future dynamics include a few notable mentions. Most of these works have not maximized the outputs and a lot of improvement can be made for the accuracy as well as the methodology.

In [3], Shuaijie Deng and Changjiang Zhang use dual attention networks and sentiment to predict stock prices and use different sentiment parameters to improve the approach. Somenath Mukherjee et. al. in [4] performed a comparative study of the for deep learning models like ANN and CNN.

In [5], Harshil Prajapati et. al. provided three algorithms using LSTM, ARIMA and SARIMA models to reduce the errors in prediction. Jaydip Sen and Sidra Mehtab in [6], per- formed stock prediction on the NSE (national stock Exchange) of India using deep learning approach. E Naresh et. al. in [7] proposed an approach which was used sentiment and use the random forest classifier. In [8], Marcus Jun Rong Foo et. al. have used the likes and dislikes as weights and predicted stock prices using naive bayes approach.

In [9], Yash Thesia et. al. used DSdT, a new method for enhancing stock price prediction and trading strategy. This approach is dynamic and scenario-driven, meaning it incorporates the current market situation into the forecasting process. To achieve this, a scenario recognition and integration module, was developed which identifies the present scenario and integrates it into the forecasting pipeline. This technique used a shallow neural network with a gating mechanism and a wide range of technical indicators to capture and incorporate the current market scenario in the prediction process. Experiments were carried out on eleven Indian Stock Market stocks to validate the mentioned approach.

J.L.Wu et. al. in [10] give a overview of work which takes the actual stock market data and uses it as a measure for predicting the highs, lows and what action has to be taken. GAN are used for this purpose wherein the generator function is used for noticing and predicting daily trading actions, and the discriminator function is used for detecting the fake trading actions. Sanjam Singh and Amandeep Kaur in [11] describes work on the Sentiment 140 dataset and work on finding the stocks predictions related to the differences in Apple Inc(AAPL) and DJIA and use boosted Regression Trees and Multi-layer Perceptron(MLP) Neural networks. The neural network method mentioned outperforms the traditional methods mentioned. [12] by M. P. Cristescu et. al. uses polynomial regression on various stocks on DJIA and provides a detailed analysis on volatility and uses VADER model for sentiment analysis.

In [13], Zhongxia Zhang and Meng Wu proposed a method for forecasting real-time locational marginal prices (RTLMPs) using a generative adversarial network (GAN) approach. This technique learns the spatiotemporal correlations from historical market data tensors and applies them to predict RTLMPs in real-time. In [14], Priyank Sonkiya et. al. have used sentiment analysis and GAN to predict stock prices for Apple Inc. In [15], Nadeem Malibari et. al. used transformers to predict the closing prices of the next seven days for the Saudi exchange.

N. Jing et. al. in [16] talks about a hybrid approach for combining deep learning with sentiment analysis for prediction of stock movement.

Works like [17] by T. Sun et. al. suggests a novel design ESAN for providing better learning ability for the dataset and also focuses on return rate, risk and sharpen ratios for A-share stocks. H. Lin et. al. in [18] show work related to using deep learning and GAN models together to increase the accuracy of results. Additionally , FinBERT has been used to generate news sentiment index for Apple Inc and comparison is made with baseline models. Zhang et. al. in [19] describes the use of CGAN which is conditional GAN and uses LSTM for generator network and MLP in the discriminator. These models were able to outperform traditional LSTM models on few datasets. In the work by Karlemstrand et. al. in [20] a model is created for time-series forecasting using LSTM. VADER is used for sentiment analysis of the dataset.

In [21], S. Boonmatham and P. Meesad performed a com- parative study of the machine learning models like SVM, MLP, GRU and LSTM. In [22] by Yingzhe Dong et. al., twitter sentiment has been analyzed using BERT-LSTM(BELT) approach and they were able to improve accuracy than the StockNet which is a state of the art for news-based stock prediction.

[23] by R. Gupta and M. Chen, relates to work done mainly on providing a outlook on how sentiment from tweets can help to improve the learning of models and thus provide for better accuracy.

In [24], Kang Zhang et. al. used only GAN to try and improve the accuracy of stock price prediction. In [25], Saloni Mohan et. al. have tried to increase the dataset so that

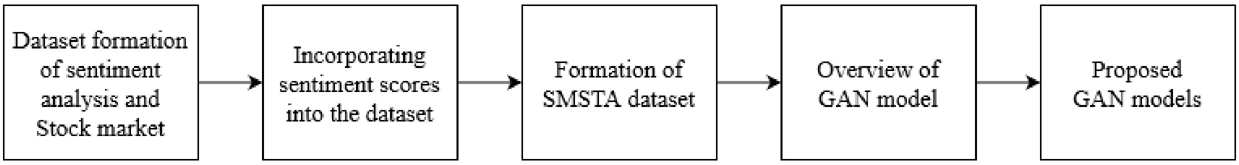


Fig. 2. System Architecture Flowchart

the models can learn more and improve accuracy. In [26], Yassine Al Amrani et. al. used sentiment analysis and classified it for the amazon product reviews. Xi Zhang et. al. in [27] proposed a novel model named multiple-source multiple instance which processed historical as well as news simultaneously. C. Wang et.al. in [28] focuses on using traditional models and doing a comparative analysis of them.

In [29] by Michael Jermann, the tweets of only the company executives were taken into consideration to improve the accu racy. It was also mentioned that a larger dataset would have helped in better training of the model. In [30], M.D. Devika et. al. provides a comparative study on machine learning based, rule based and lexicon-based approach for stock prediction. Pagolu Venkata Sasank et. al. [31] includes work for the proper prediction for the rise and the fall of price and their movements. They incorporate a sentiment analyzer consisting of random forest, SMO, and logistic regression and then aggregating scores for sentiments. After this a correlation analyzer is incorporated consisting of LibSVM, Logistic regression and SMO.

Significant improvements have been made to reduce errors. M. Arias et. al. in [32] present work related to datasets that include not only stocks but other fields as well as a graphical network creation and use of SVM for analysis and prediction. This work does take into account a lot many stocks and accounts for volatility in the index as a whole. Anshul Mittal and Arpit Goel in [2] work on finding the correlation between people's and market's sentiment on the Dow Jones Industrial Average (DJIA) using a Self-Organizing Fuzzy Neural Network (SOFNN). Works like [33] by Hu et. al. have used transformers from Hugging Face BERT for sentiment classification but have found difficulties in predicting proper movements for the dataset.

Thus we noticed various gaps in not only the accuracy but also few models only used the historical stock data to train the models for predicting the daily trading actions and sentiment plays an important role in helping improve decision; making for sellers, buyers as well as short-term holders. Sentiment Analysis was applied in a few papers but the approaches were traditional algorithms so improvement was made for accuracy as well as the prediction process.

III. METHODOLOGY

The idea indicated is made up of three fundamental components. The first is a sentiment analysis of tweets from 15 companies - Tesla, Google, Microsoft, Amazon, Meta, Netflix, Apple, TSM, Xpeng Inc - ADR, ZS, NIO, PG, and others. The sentiment score was computed through the use of the VADER Python tool for Natural Language Processing, which was applied to stock tweets from various companies.

The second phase involves developing a novel dataset - the SMSTA dataset, which not only includes daily stock price fluctuations for each company stock but was also mapped with the sentiment scores assessed in the prior dataset formation. The concept is to combine the benefits of sentiment analysis with various stock price parameters in order to improve the accuracy of daily stock tweet price predictions.

Lastly, the GAN model is applied to anticipate stock values more accurately. LSTM and GRU layers make up the Generator function of the GAN model. Different neuron values used in the CNN layers make up the Discriminator function of the GAN model.

By taking into account both historical patterns and public sentiment, adding sentiment analysis to SMTSA dataset used to train a GAN model increases the precision of the stock price data that is produced. This improves how accurately the data supplied reflects actual market circumstances and offer a useful tool for evaluating investment strategies and risk management. Consequently, the paper's system architecture flow is depicted in Figure 1

A. Dataset formation for Sentiment analysis of Stock market Tweets

From January 1, 2021 to September 30, 2022, a total of 90,564 tweets about stock market were retrieved using the Twitter API. Twitter4J is a java program that aids in tweet extraction from Twitter. The tweets were gathered using the Twitter API and contained tweets about shares in companies including Tesla, Google, Microsoft, Amazon, Meta, Netflix, Apple, TSM, Xpeng Inc - ADR, ZS, NIO, PG and others. The companies in the dataset are divided as follows: Tesla 37%, TSM 21%. Among the remaining companies, Amazon makes up 20%, Microsoft 9%, Google 9%, and others 4%. When pre- dicting stock prices for different companies, the dataset created considers not only historical data and technical indicators but also external market influences like trader sentiment and brand reputation as seen in social media posts. Fields in the dataset include: *Date*, *Tweet*, *Stock name, and Company name*.

B. Incorporating sentiment scores into the dataset

The creation of stock prices is significantly more complex than simply looking at previous data. Over the course of a single day, a single online post could change the course of events and cause the market to crash. The evidence for this may be found in tweets from Elon Musk, the coronavirus, and the beginning of the Russian invasion of Ukraine.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Consequently, considering yet another crucial external indicator, such as stock market participants' mood, sentiment scores were incorporated into the dataset.

Tasks involving sentiment analysis are very specialized. Many sentiment analyzers are open source and many have been developed through extensive research on the subject of reading the emotions in movie reviews and news articles. This analyzers' primary flaw is that they were trained on a different corpus, which is a problem. Movie corpus and stock corpus, for instance, are not interchangeable.

VADER, a pre-built Python machine learning tool for natural language processing, was used to analyze the sentiment of each tweet. VADER is accessible via vader-sentiment. Different words and punctuation can be given sentiment scores by VADER, ranging from extremely negative (-1) to extremely positive (+1), with zero representing neutral. VADER may then combine the results of several token analyzes and grammatical frame analyzes to provide a total sentiment score for the sentence. VADER emphasizes the ability to recognize capital letters, slang, exclamation points, and the most popular emojis. Since tweets are not written in an official or academic manner, VADER is appropriate for social media analysis. However, some new words were added to VADER's original dictionary because VADER mispronounced or overlooked some crucial words in the financial industry, which led to inaccuracy. For instance, the phrases "bully" and "bullish" are derogatory in the VADER language but favorable terms in the financial market. The Loughran-McDonald Financial Sentiment Word Lists, a financial dictionary that includes several terms used in the stock market, were utilized to update the VADER vocabulary. There are seven categories in this dictionary. The positive, negative, and neutral lists were updated in the current model. All relevant tweets were examined using VADER to determine the scores of each tweet in order to obtain the overall sentiment score for per stock for one day. This average's daily percentage change was taken into account as a general factor for 1 day.

Fields in the dataset after incorporating sentiment analysis include: *Date, Tweet, Stock name, Company name, Sentiment score, Positive, Negative and Neutral*.

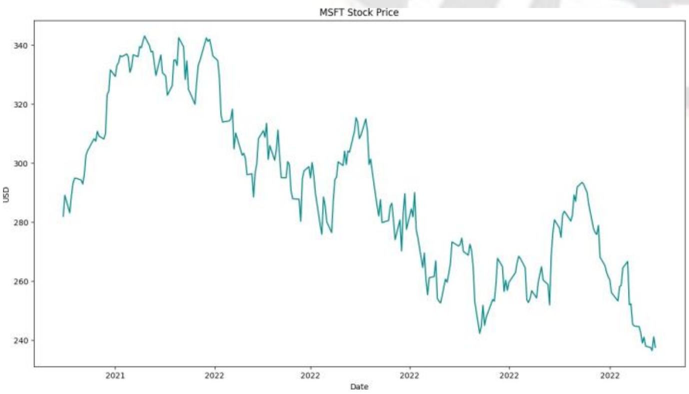
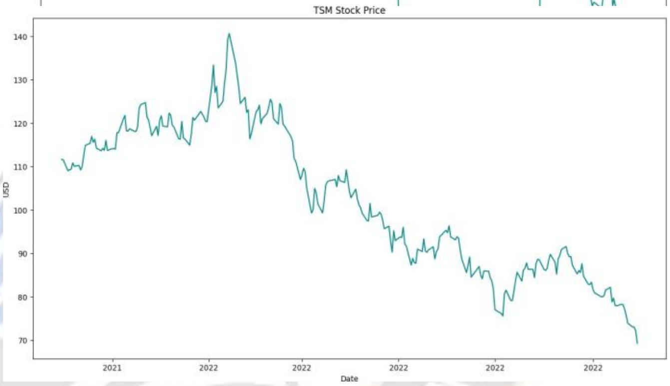


Fig. 2. SMTSA Dataset - Microsoft Stock price

C. *Formation of SMTSA Dataset*

From January 1, 2021, to September 30, 2022, stock prices for each firm in the prior dataset were gathered as shown in Figure 1, 3, 4. This new dataset - SMTSA Dataset includes stock prices for each company, and it comprises five columns of time-series price data: Open, High, Low, Close, and Adjclose. Based on field values of the date and stock name, the sentiment score values from the old dataset were mapped to SMTSA Dataset. Fields in SMTSA Dataset include: *Date, Open, High, Low, Sentiment score and Stock Name*.

Fig. 3. SMTSA Dataset - Amazon Stock price



Different technical indicators, such as moving averages, bollinger bands as middleline in Equation 2, upperbound in Equation 3 and lowerbound in Equation 4 etc., are added to the training data as shown in Figure 5 to help the network understand the bigger picture of the market. These indicators describe the development of stock price not only for the current day but for the previous week or more.

Fig. 4. SMTSA Dataset - TSM Stock price

Moving Average for the last 7 days is designated as **MA(7)**, while Moving Average for the last 20 days is designated as **MA(20)**.The **EMA**, or exponential moving average, can be calculated as follows in Equation 1:

$$EMA_t = Pclose + (EMA_t - 1(100 - P))* \quad (1)$$

Bollinger Bands are calculated as follows in Equations 2, 3, 4:

$$middleline : stdev(MA(20)) \quad (2)$$

$$upperbound : MA(20) + 2stdev(MA(20)) \quad (3)$$

$$lowerbound : MA(20) - 2stdev(MA(20)) \quad (4)$$

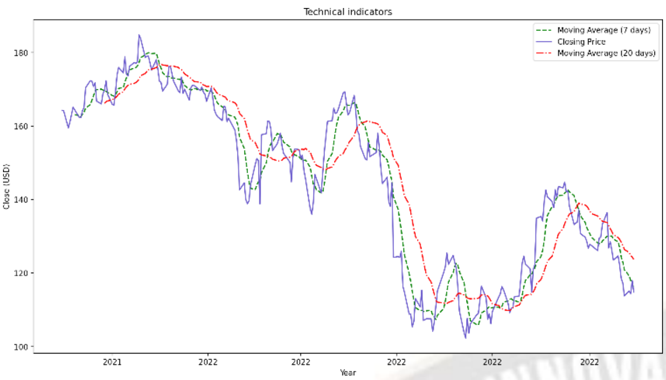


Fig. 5. SMTSA Dataset with Technical Indicators

To make the sequential data from the training datasets into an acceptable input for the model, data cleaning, transformation, and approximation normalization were all completed.

D. Overview of the GAN Model

GAN model consists of multiple-input single-output system that takes as inputs n previous days' worth of data together with m (16) stock market parameters each day and generates one of four types of future price data as the prediction: the open, high, low, or close. A LSTM network in the generative model is partly used since the stock data is typical time- series data. The stock prices on day t+1 was predicted using LSTM, which had been trained to extract the characteristics of price data from the previous n days. The embedding layer and LSTM layer make up the generator model. Between two LSTM layers in a generator, a dropout layer is added to improve generalization and reduce the likelihood of overfitting. The discriminator's job is to categories incoming data as authentic or fraudulent. When presented with accurate input data, this classification model produces 0, and when presented with erroneous data, it produces 1. The embedding layers and multilayer convolution neural network layers (CNN) make up the discriminator's two components. The embedding layers convert the data and labels into CNN input. The CNN model successfully completes the classification assignment by projecting the data into a higher dimensional space. The output layer of the CNN uses the sigmoid function, whereas the hidden layers employ the Leaky ReLU activation function.

1) The Generator: According to its stability, the LSTM layer serves as the generator for the majority of the GAN model. The SMSTA dataset has 16 characteristics and also contains the stock price history for the last two years. The model forecast many steps in advance. As a result, the input step and the output step of the generator need to be speci- fied. The generator will take as input three-dimensional data, including batch size, input-step, and features; it will generate two-dimensional data, including batch size and output-step. The model employs four or five layers of LSTM or GRU, followed by three or four layers of dense layer, in order to construct a generator with high performance. The neuron counts used in the model are 1024, 512, 256, 128 and 64.

2) The Discriminator: Convolution neural networks are used as the discriminator in the GAN model with the goal of determining if the input data was authentic or fraudulent. The generator's generated data or the original data is used as the discriminator's input. Four or five 1D Convolution layers with 512, 256, 128, 64, and 32 neurons each are included in this discriminator, coupled with two or three Dense layers organised inside the layers that contain 220, 220, and 1 neuron each. All of the layers' activation functions have been set to the Leaky Rectified Linear Unit (ReLU), except for the output layer, which has its activation function set to Sigmoid activation for four-layer GANs and linear activation for five- layer GANs.

3) Loss functions: Binary cross-entropy is used in the GAN model structure to determine the loss for the generator and discriminator. For the discriminator in particular, the produced stock price was coupled with the previous stock price of the input stages as our input. This step lengthens the data and improves the discriminator's accuracy as it learns the categorization.

E. Proposed GAN Model

In order to provide trading actions with high performance, the proposed model is a GAN-based framework that integrates CNN, LSTM, and three GAN models.

The first GAN architecture, SMAM-1, Stock Market Analysis Model-1 used has a five-layer LSTM network as a generator and a five-layer CNN as the discriminator. The discriminator was implemented with four convolutional layers, followed by a flattened layer, one dense layer which preceded one last convolutional layer, followed by two dense layer with sigmoid activation.

The second GAN architecture, SMAM-2, Stock Market Analysis Model-2 used has a four-layer LSTM network as a generator and a four-layer CNN as the discriminator. The discriminator was implemented with four convolutional layers, followed by one flattened layer and two dense layer with sigmoid activation.

The third GAN architecture, SMAM-3, Stock Market Analysis Model-3 used has a four-layer GRU network as a generator and a four-layer CNN as the discriminator. The discriminator was implemented with two convolutional layers, followed by one flattened layer, one dense layer which preceded next two convolution layers, followed by one flattened layer and two dense layers with sigmoid activation.

IV. RESULTS

Along with the proposed model, the Root Mean Square Error (RMSE) metric was chosen for evaluation and com- parison. Finally, the models were compared based on stock price as well as sentiment analysis on tweets for three of the SMTSA dataset's most widely distributed companies: Microsoft, TSM, and Amazon. To determine which model produced the best results, the stock price and sentiment score of each company were individually assessed using the three defined GAN models as shown in Table 1, 2 and 3.

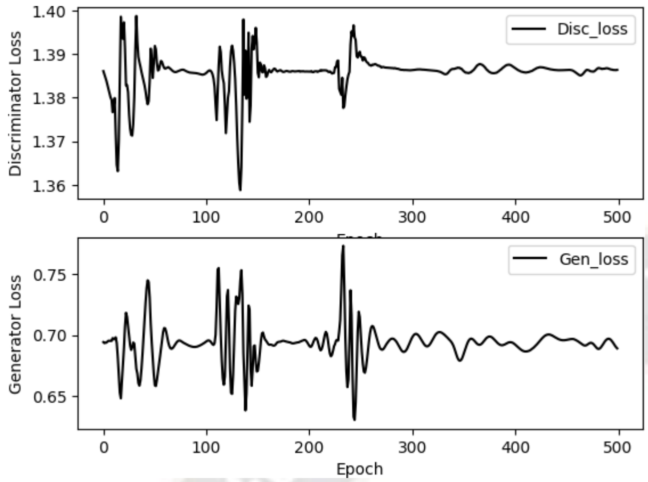


Fig. 6. SMAM-1 (Amazon Company) Discriminator and Generative loss

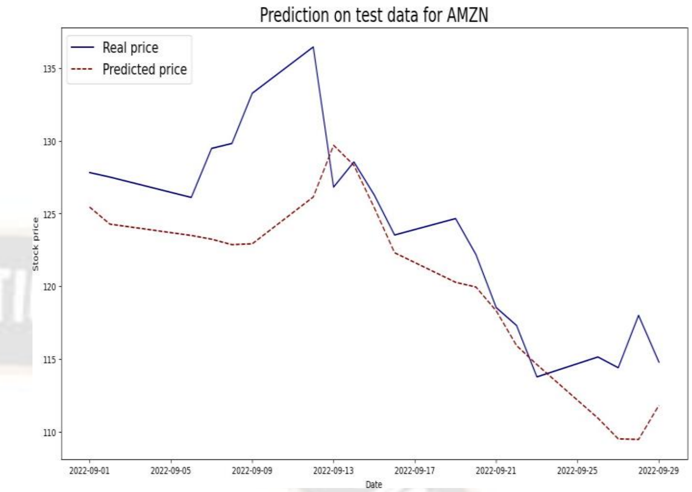


Fig. 7. SMAM-1 (Amazon Company) Real and Predicted Data graph

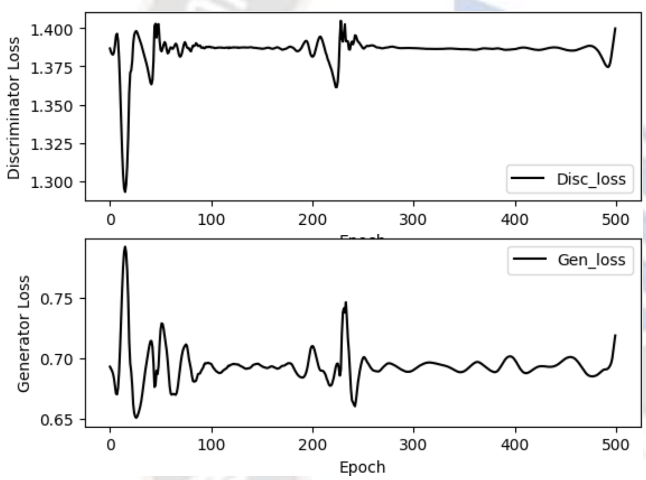


Fig. 8. SMAM-2 (Amazon Company) Discriminator and Generative loss

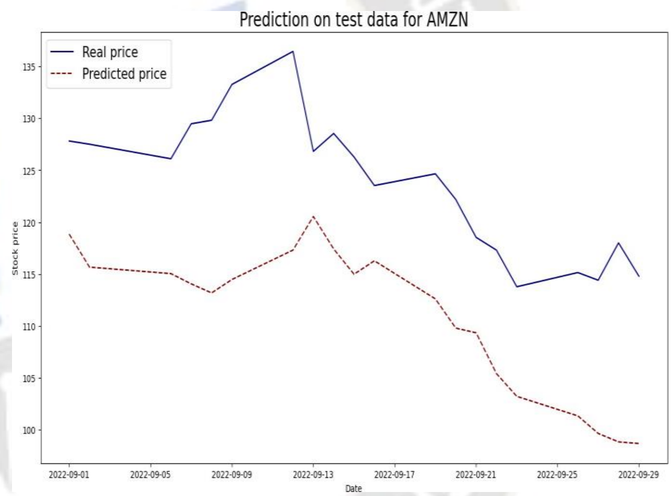


Fig. 9. SMAM-2 (Amazon Company) Real and Predicted Data graph

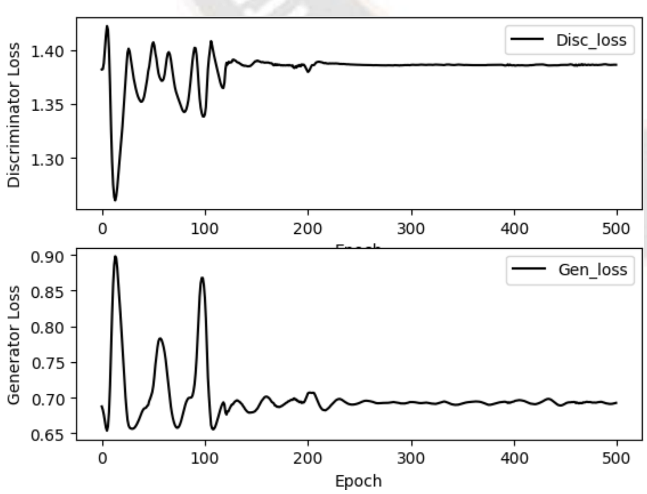


Fig. 10. SMAM-3 (Amazon Company) Discriminator and Generative loss

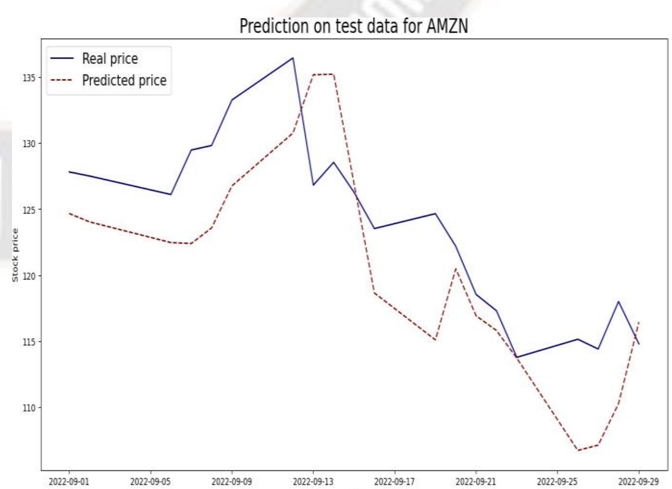


Fig. 11. SMAM-3 (Amazon Company) Real and Predicted Data graph

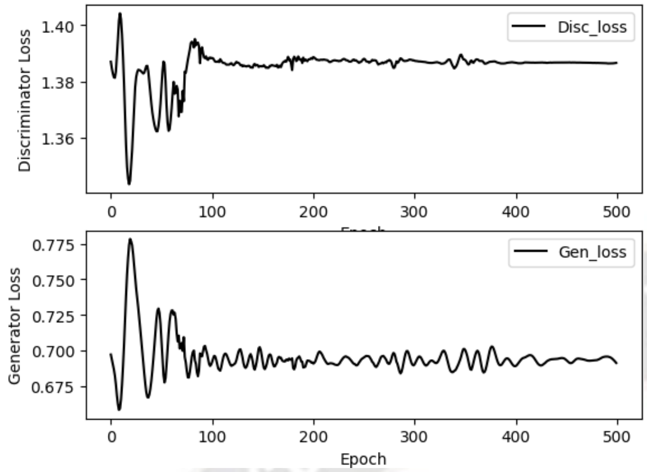


Fig. 12. SMAM-1 (TSM Company) Discriminator and Generative loss

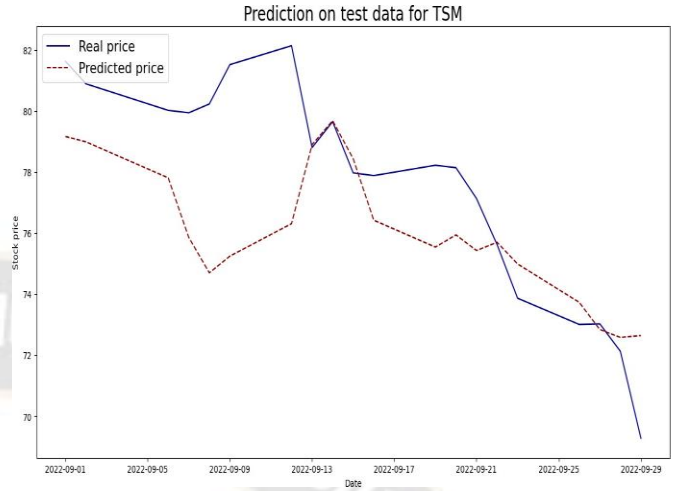


Fig. 13. SMAM-1 (TSM Company) Real and Predicted Data graph

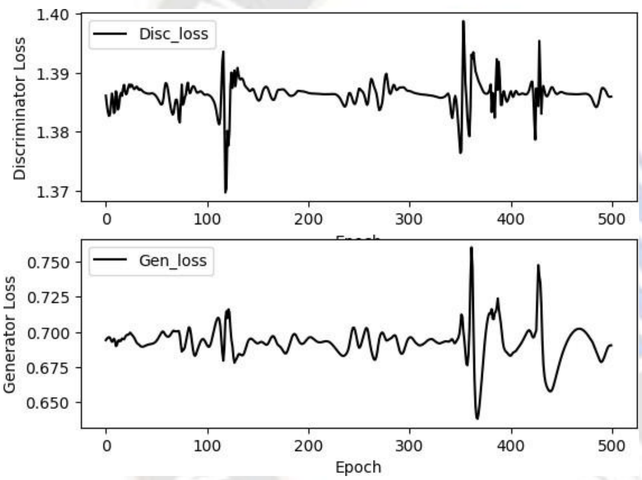


Fig. 14. SMAM-2 (TSM Company) Discriminator and Generative loss

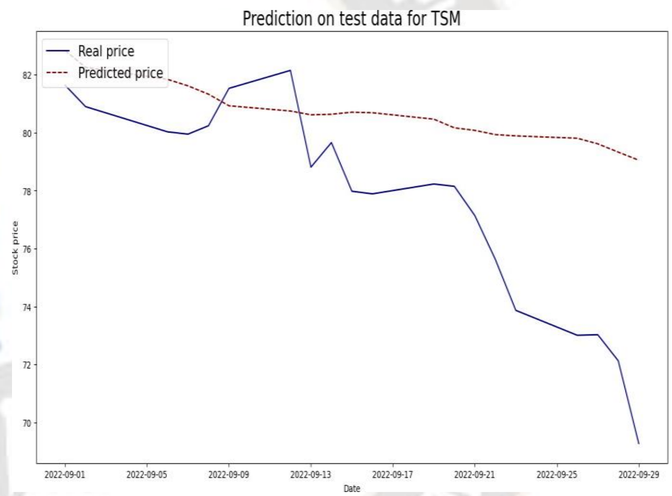


Fig. 15. SMAM-2 (TSM Company) Real and Predicted Data graph

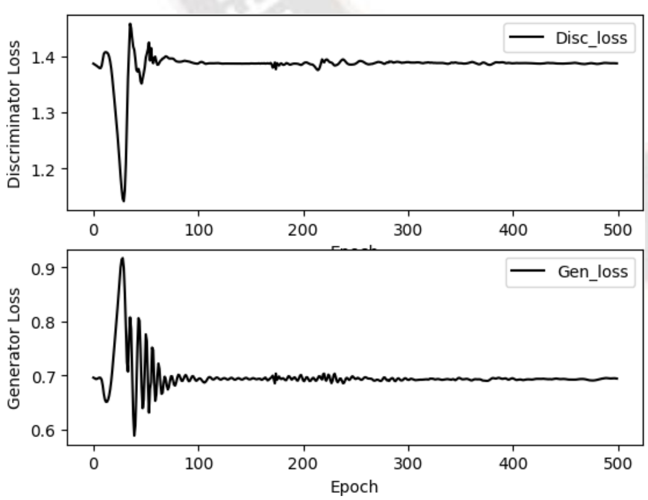


Fig. 16. SMAM-3 (TSM Company) Discriminator and Generative loss

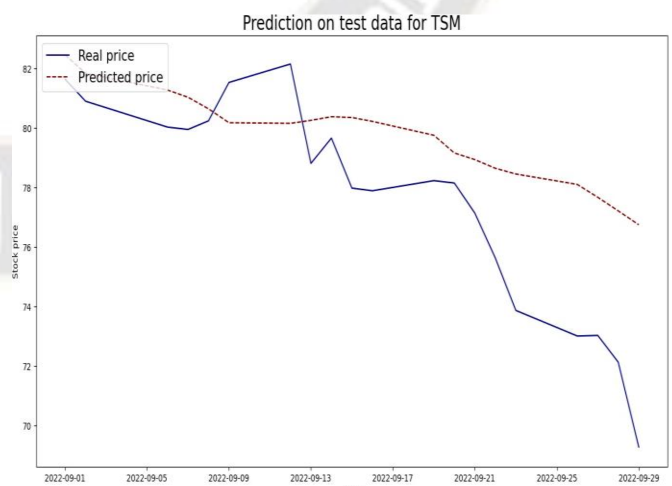


Fig. 17. SMAM-3 (TSM Company) Real and Predicted Data graph

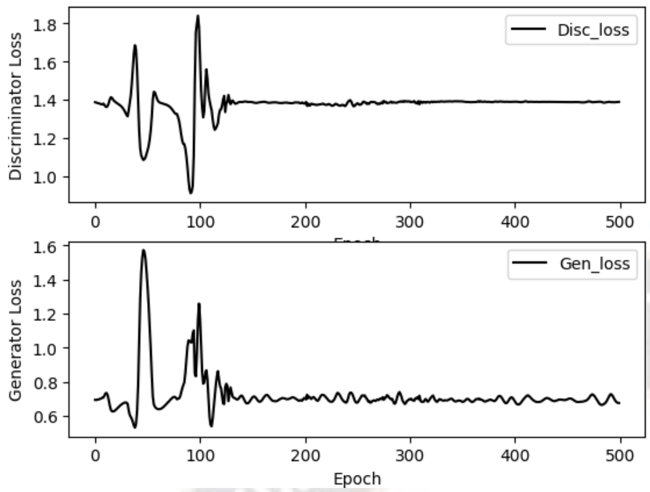


Fig. 18. SMAM-1 (Microsoft Company) Discriminator and Generative loss

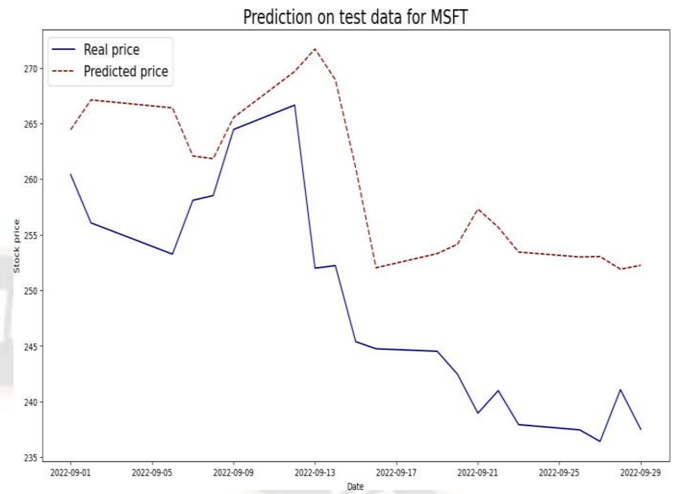


Fig. 19. SMAM-1 (Microsoft Company) Real and Predicted Data graph

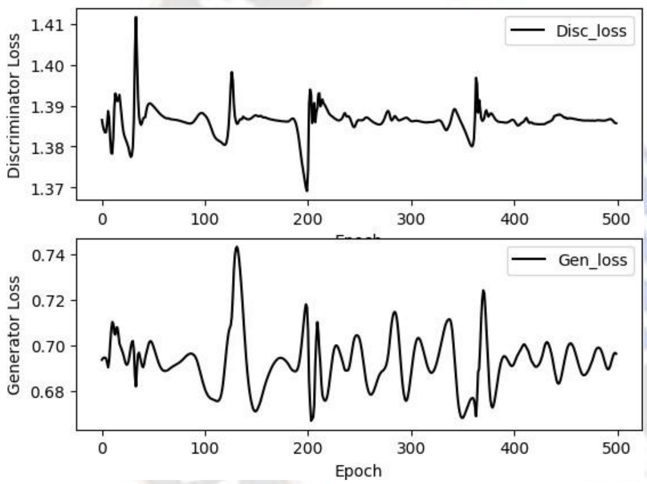


Fig. 20. SMAM-2 (Microsoft Company) Discriminator and Generative loss

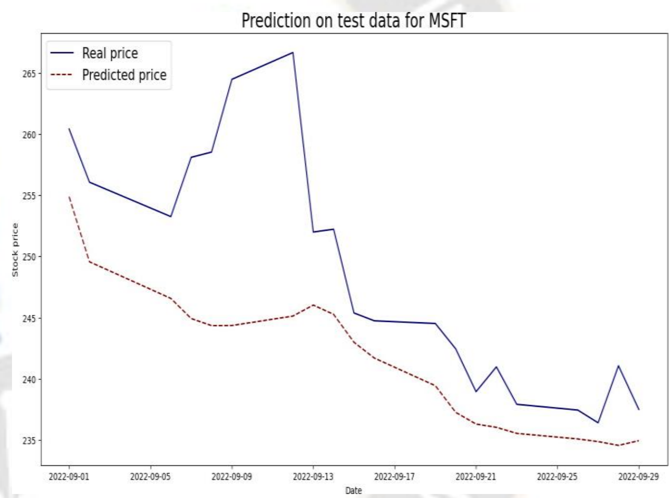


Fig. 21. SMAM-2 (Microsoft Company) Real and Predicted Data graph

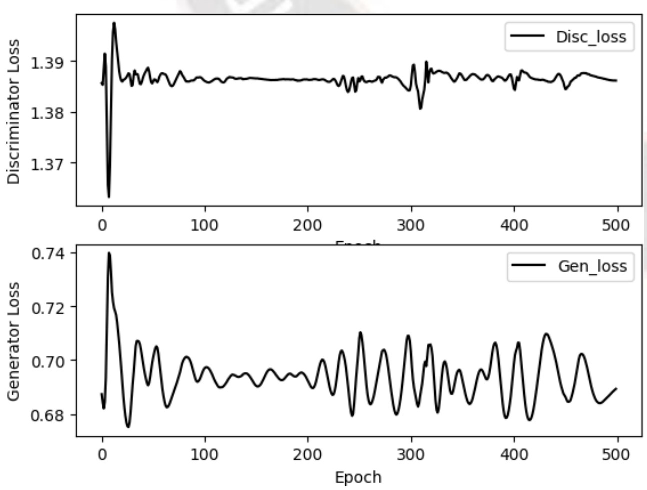


Fig. 22. SMAM-3 (Microsoft Company) Discriminator and Generative loss

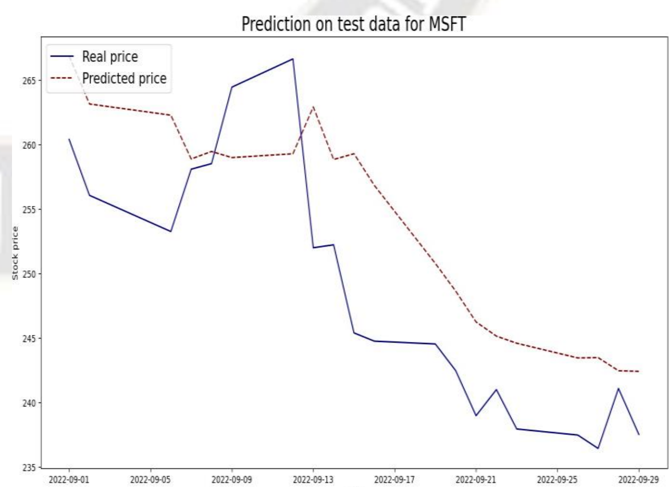


Fig. 23. SMAM-3 (Microsoft Company) Real and Predicted Data graph

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

TABLE 1
VALUES FOR THE GENERATOR AND DISCRIMINATOR LOSSES WITH THE RSME SCORE FOR THE AMAZON COMPANY DATASET

Amazon	SMAM-1	SMAM-2	SMAM-3
Discriminator Loss	1.3864	1.3864	1.3997
Generator Loss	0.6889	0.6928	0.7189
RSME	4.9185	5.5817	13.4010

The model performed best with the SMTSA Dataset from the Amazon company out of the three datasets examined. It provides precise results with minimal Generator and Discriminator loss as well as RSME score. For each of the three proposed GAN models, the SMTSA Dataset from the Amazon Company performs well as illustrated in Figure 6 to Figure 7. As shown in Table 1 and Figure 13 to Figure 18 of the three GAN models, SMAM-1 displays the best results with RSME score of 4.9185.

TABLE 2
VALUES FOR THE GENERATOR AND DISCRIMINATOR LOSSES WITH THE RSME SCORE FOR THE TSM COMPANY DATASET

TSM	SMAM-1	SMAM-2	SMAM-3
Discriminator Loss	1.3865	1.3869	1.3859
Generator Loss	0.6911	0.6938	0.6905
RSME	2.8823	3.07992	4.1335

For stock forecasts from January 1, 2021 to September 30, 2022, the SMSTA Dataset for TSM and Microsoft company likewise produced strong, accurate results as illustrated in Figure 12 to Figure 23. Although the Generator and Discriminator losses are very low, the predicted and real values graphs indicate that there is room for improvement in the proposed GAN models. SMAM-1 exhibits the best outcomes of the other two models for both the TSM and Microsoft companies with RSME score of 2.8823 and 7.3005 respectively as shown in Table 2 and 3.

TABLE 3
VALUES FOR THE GENERATOR AND DISCRIMINATOR LOSSES WITH THE RSME SCORE FOR THE MICROSOFT COMPANY DATASET

Microsoft	SMAM-1	SMAM-2	SMAM-3
Discriminator Loss	1.3861	1.3857	1.3881
Generator Loss	0.6893	0.6962	0.6749
RSME	7.3005	8.9569	12.5966

V. CONCLUSION AND FUTURE SCOPE

Stock prediction is of great significance as it helps investors and financial analysts to make informed decisions regarding investments in the stock market. Accurately forecasting the fluctuations in stock prices can enable investors to maximize their returns while minimizing their risks. Developing an effective and reliable method for predicting stock market fluctuations using tweets sourced from Twitter is a complex task, and researchers are still exploring various approaches to enhance the accuracy and effectiveness of such techniques. This paper proposes a model that predicts accurate trends, though the values may be not close enough to the original ones. Using these trends, investors can make the choice to buy, sell or hold the particular stocks. This paper utilized three different GAN models that integrated LSTM along with CNN. With the Amazon company dataset, the first model, SMAM-1, had the best results with RSME score of 4.9185. Similarly, the TSM and Microsoft company datasets produced the best results with SMAM-1 with RSME scores of 2.8823 and 7.3005 respectively. Hence, out of the three models, the first model, SMAM-1, produced better outcomes. Additionally, out of the three datasets analyzed, the models yielded the most accurate results when tested on the SMTSA dataset sourced from Amazon.

Despite the advancements in machine learning and sentiment analysis, there is still no foolproof method for detecting and predicting stock market changes based on Twitter data. Stock prediction requires a large dataset, which can be challenging to collect and prepare. Using a large dataset can introduce complexities such as data noise and quality issues, which can reduce the accuracy and effectiveness of the model. Additionally, ensuring the reliability and consistency of the dataset can be a concern, as market conditions and sentiments can vary significantly over time. Additionally, the accuracy in detecting neutral comments can be improved further. Neutral comments may not have a significant impact on stock market fluctuations, but their exclusion from the analysis can potentially skew the results. The analysis presented does not account for these factors that may impact stock market fluctuations, and these factors provide potential areas for future research.

REFERENCES

[1] S. Albahli, A. Irtaza, T. Nazir, A. Mehmood, A. Alkhalifah, and W. Albattah, "A machine learning method for prediction of stock market using real-time twitter data," *Electronics*, vol. 11, no. 20, Oct 2022.

[2] A. Mittal and A. Goel, "Stock prediction using twit- ter sentiment analysis," *Stanford University, CS229 (2011(http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf))*, vol. 15, 2012.

[3] S. Deng and C. Zhang, "Stock price prediction with technical analysis and market sentiment via dual-attention networks," *Journal of Forecasting*, 2023.

[4] S. Mukherjee, B. Sadhukhan, N. Sarkar, D. Roy, and S. De, "Stock market prediction using deep learning algorithms," *CAAI Transactions on Intelligence Technology*, vol. 8, no. 1, 2023.

[5] P. Jishtu, H. Prajapati, and J. Fiaidhi, "Prediction of the stock market based on machine learning and sentiment analysis," 2022.

[6] J. Sen, "Stock price prediction using deep learning and natural language processing jaydip sen and sidra mehtab," 2022.

[7] E. Naresh, B. J. Ananda, K. Keerthi, and M. Tejonidhi, "Predicting the stock price using natural language processing and random forest regressor," in *2022 IEEE International Conference on Data Science and Information System*

(ICDSIS). IEEE, 2022.

[8] M. J. R. Foo and C. S. Pun, "Stock movement prediction with social sentiments and interactional data: Integrating nlp and Bayesian frameworks," in 2022 *4th International Conference on Natural Language Processing (ICNLP)*. IEEE, 2022

[9] Y. Thesia, V. Oza, and P. Thakkar, "A dynamic scenario-driven technique for stock price prediction and trading," Journal of Forecasting, vol. 41, no. 3, 2022.

[10] J.-L. Wu, X.-R. Tang, and C.-H. Hsu, "A prediction model of stock market trading actions using generative adversarial network and piecewise linear representation approaches," Soft Computing, 2022.

[11] S. Singh and A. Kaur, "Twitter sentiment analysis for stock prediction," Available at SSRN 4157658, 2022.

[12] M. P. Cristescu, R. A. Nerisanu, D. A. Mara, and S.-V. Oprea, "Using market news sentiment analysis for stock market prediction," Mathematics, vol. 10, no. 22, 2022.

[13] Z. Zhang and M. Wu, "Predicting real-time locational marginal prices: A gan-based approach," IEEE Transactions on Power Systems, vol. 37, no. 2, 2021.

[14] P. Sonkiya, V. Bajpai, and A. Bansal, "Stock price prediction using bert and gan," 2021.

[15] N. Malibari, I. Katib, and R. Mehmood, "Predicting stock closing prices in emerging markets with transformer neural networks: The saudi stock exchange case," International Journal of Advanced Computer Science and Applications, vol. 12, no. 12, 2021.

[16] N. Jing, Z. Wu, and H. Wang, "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," Expert Systems with Applications, vol. 178, 2021.

[17] T. Sun, W. Zheng, S. Yu, M. Li, and J. Ou, "Esan: Efficient sentiment analysis network of a-shares research reports for stock price prediction," 2021.

[18] H. Lin, C. Chen, G. Huang, and A. Jafari, "Stock price prediction using generative adversarial networks," J. Comp. Sci, 2021.

[19] Y. Zhang, J. Li, H. Wang, and S.-C. T. Choi, "Sentiment-guided adversarial learning for stock price prediction," Frontiers in Applied Mathematics and Statistics, vol. 7, 2021.

[20] R. Karlemstrand and E. Leckstr"¨om, "Using twitter attribute information to predict stock prices," 2021.

[21] S. Boonmatham and P. Meesad, "Stock price analysis with natural language processing and machine learning," in Proceedings of the 11th International Conference on Advances in Information Technology, 2020.

[22] Y. Dong, D. Yan, A. I. Almudaifer, S. Yan, Z. Jiang, and Y. Zhou, "Belt: A pipeline for stock price prediction using news," in 2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020.

[23] R. Gupta and M. Chen, "Sentiment analysis for stock price prediction," in 2020 IEEE conference on multimedia information processing and retrieval (MIPR). IEEE, 2020.

[24] K. Zhang, G. Zhong, J. Dong, S. Wang, and Y. Wang, "Stock market prediction based on generative adversarial network," Procedia computer science, vol. 147, 2019.

[25] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," in 2019 IEEE fifth international conference on big data computing service and applications (BigDataService). IEEE, 2019.

[26] Y. Al Amrani, M. Lazaar, and K. E. El Kadiri, "Random forest and support vector machine based hybrid approach to sentiment analysis," Procedia Computer Science, vol. 127, 2018.

[27] X. Zhang, S. Qu, J. Huang, B. Fang, and P. Yu, "Stock market prediction via multi-source multiple instance learning," IEEE Access, vol. 6, 2018.

[28] C. Wang, Y. Xu, and Q. Wang, "Novel approaches to sentiment analysis for stock prediction," Stanford University, 2018.

[29] M. Jermann, "Predicting stock movement through executive tweets," 2017.

[30] M. Devika, C. Sunitha, and A. Ganesh, "Sentiment analysis: A comparative study on different approaches," Procedia Computer Science, vol. 87, 2016.

[31] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of twitter data for predicting stock market movements," in 2016 international conference on signal processing, communication, power and embedded system (SCOPES). IEEE, 2016.

[32] M. Arias, A. Arratia, and R. Xuriguera, "Forecasting with twitter data," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 5, no. 1, 2014.

[33] K. Hu, D. Grimberg, and E. Durdyev, "Twitter sentiment analysis for predicting stock price movements," 2021.