Enhancing Time Series Stock Predictions Using GANs with Technical Indicators and Twitter Sentiment: Challenges with Low-Popularity Tickers

Gudla Anjaneyulu[1] , Pundru Chandra Shaker Reddy[2] , Pappula Praveen[3]

[1]Phd research scholar , SR university, warangal , School of Computer Science & Artificial Intelligence SR University, Warangal, Telangana, India, Email: anjugudla5@gmail.com
[2]Professor , SR university, warangal , School of Computer Science & Artificial Intelligence SR University, Warangal, Telangana, India, Email: chandu.pundru@gmail.com
[3]Associate Professor, SR university, warangal , School of Computer Science & Artificial Intelligence SR University, Warangal, Telangana, India, Email: prawin1731@gmail.com

Received: 13.07.2024 Revised: 17.08.2024 Accepted: 10.09.2024

ABSTRACT
Machine learning algorithms for stock market prediction have attracted a lot of interest recently. Conventional methods use technical indications and historical price data, but more recently, social media sentiment—especially from Twitter and other platforms—has been used to improve forecast accuracy. The use of Generative Adversarial Networks (GANs) to create synthetic data and enhance time series forecasts has shown promise. In particular, low-popularity or low-volume tickers are the subject of this study's integration of technical indicators and Twitter sentiment research to investigate the use of GANs in stock price prediction. Although GANs have shown promise in improving predictions for high-volume tickers, low-popularity stocks present difficulties due to their restricted data availability, scant sentiment information, and increased volatility. In order to supplement the little historical and sentiment data, this study generates synthetic data to investigate how well GANs do in tackling these problems. Furthermore, we explore the potential for enhancing predictive performance of GAN models by including sentiment analysis obtained from Twitter activity and technical indicators like RSI, MACD, and moving averages. The research evaluates the benefits and drawbacks of GAN-based models in these various scenarios by comparing the forecast accuracy for tickers with high and low popularity. Our results highlight the limitations imposed by volatility and data scarcity in low-popularity equities. They also provide insights into possible approaches to enhance forecast accuracy in these situations, such as enhanced feature engineering for technical indicators and improved sentiment extraction methods.

Keywords: Time Series, Stock Predictions, GANs , Technical Indicators , Twitter Sentiment, Challenges, Low-Popularity Tickers.

1. INTRODUCTION

It has always been difficult to predict fluctuations in the stock market because of the inherent complexity and volatility of financial markets. Historical price data and technical indicators, such moving averages, the Relative Strength Index (RSI), and moving average convergence-divergence (MACD), have historically been used to forecast stock prices. Although these techniques have provided insightful information, they often fail to fully capture the range of market variables .New approaches to improve prediction accuracy have been offered by recent developments in machine learning. Among them, Generative Adversarial Networks (GANs) have become a potent instrument for creating synthetic data, which may assist in enhancing prediction models and overcoming data constraints. Two neural networks, a generator and a discriminator, collaborate to produce realistic data that may be utilised to train further models in GANs. Social media sentiment research is becoming an increasingly important tool for stock market prediction, in addition to technical indications. Real-time insights regarding investor activity and market mood are available on platforms such as Twitter. Researchers and practitioners may assess public sentiment and its possible impact on stock prices by examining tweets and other social media information. The goal of this work is to improve stock price forecasts by integrating GANs with technical indicators and sentiment research from Twitter. It specifically tackles the difficulties of forecasting low-volume or low-popularity tickers. Because of their higher volatility, scarce sentiment data, and restricted data availability, these stocks pose particular challenges. The following are the main goals of this study :To Assess GANs' Efficient

Performance: Examine how GANs may be used to provide synthetic data for low-popularity equities to supplement the few historical and sentiment data that is currently accessible. How Sentiment Analysis and Technical Indicators Are Integrated: Examine how using Twitter sentiment analysis in conjunction with technical indicators like RSI, MACD, and moving averages may enhance prediction accuracy when applied to GAN-generated data. To Assess Accuracy of Predictions: Analyse prediction performance for tickers with high and low popularity in comparison, emphasising the benefits and drawbacks of GAN-based mod-els in these situations. To Determine Methods for Enhancement: Provide an analysis of possible approaches to increase forecast accuracy for low-popularity stocks, such as sophisticated feature engineering for technical indicators and sentiment extraction methods. By tackling these goals, this study hopes to add to the expanding corpus of knowledge on sophisticated stock market prediction methods and provide workable answers to the problems caused by low-popularity tickers.

2. LITERATURE REVIEW

The following points and references will assist you in covering important areas of your literature research on improving time series stock forecasts combining GANs with technical indicators and sentiment on Twitter, especially for low-popularity tickers: Because Generative Adversarial Networks (GANs) can represent complicated data distributions, they have been investigated for financial forecasts. GANs may produce realistic synthetic data and capture non-linear relationships, in contrast to typical time series models.[1]Price patterns are often predicted using technical indicators like Bollinger bands, relative strength index (RSI), and moving averages (MA). Predictive performance is improved when these indicators are combined with GANs.[2] An extra component for stock prediction models is Twitter sentiment analysis. Predictions are enhanced by sentiment cues, especially for short-term price fluctuations. However, insufficient data for low-popularity tickers may make it difficult to identify hurtful sentiment.[3]Small-cap companies may have insufficient data to support sentiment research as well as technical indicators. While GANs may be helpful in addressing the scarcity problem in data augmentation, if they are trained on sparse data, they may also create noise. Furthermore, lower sentiment signals may result from a lack of social media activity around these equities .Hybrid models are often investigated in research combining GANs with sentiment data and technical indicators for low-data situations. To improve prediction resilience, these models combine GANs with neural networks, decision trees, and support vector machines.[5]Using AI to forecast stocks raises ethical questions concerning data privacy, market manipulation, and possible bias in Twitter sentiment (e.g., tweets created by bots). Specifically, the lack of data for tickers with low popularity might lead models to overfat or draw the wrong conclusions.[6]Your literature evaluation will be guided by these references and points as it examines the advantages, difficulties, and solutions associated with using GANs, technical indicators, and Twitter sentiment analysis to anticipate stock values, especially for less well-known companies.

3. METHODOLOGY

Several crucial phases are included in the process to overcome the difficulties in forecasting stock prices for low-popularity tickers utilising GANs (Generative Adversarial Networks) in conjunction with technical indicators and sentiment on Twitter: For low-popularity tickers, time series data on stock prices (OHLC: Open, High, Low, Close) is gathered from websites like Quandl, Yahoo Finance, and Alpha Vantage. [7] Important technical indicators, including as Bollinger Bands, RSI, MACD, and Moving Averages, are constructed from stock price data to provide more details on price changes and patterns. Using APIs like Tweepy, sentiment data from Twitter is gathered with an emphasis on tweets that reference the low-popularity tickers. These tweets' sentiment is categorised as good, negative, or neutral using sentiment analysis models (such as VADER or TextBlob). [8] Z-score normalisation or Min-Max scaling are used to rescale stock prices and technical indicators. By doing this, the data is guaranteed to be suitable for input into the GAN models. Sentiment data could be scarce because of the tickers' low level of popularity.[9] To match the sentiment data with the stock price time series, methods like interpolation and padding are used. Only important characteristics may be retained by reducing the dimensionality of the input data using t-SNE or PCA (Principal Component Analysis).[10] The generator's job is to create artificial stock price sequences based on emotion and technical factors. Noise vectors (randomly sampled from a Gaussian distribution) make up the generator's input. As conditional inputs, sentiment data and technical indicators are used. The discriminator attempts to discern between sequences of stock prices that are synthetic and genuine. The generator should generate stock price sequences that the discriminator cannot tell apart from actual data in order for the GAN to function properly. To make sure that sentiment influence and stock price patterns are both recorded, a combination of auxiliary loss for sentiment data and Wasserstein loss (to increase the stability of GAN training) is used.[12] To guarantee model robustness, data is divided into training, validation, and test sets (e.g., 60% training, 20% validation, and 20% testing). The generator and discriminator's hyperparameters (such as learning rate, batch size, and epochs) are adjusted using methods like grid search and Bayes-ian optimisation. [13]It is difficult to get historical da-ta points for tickers with little popularity. This problem is mitigated by data augmentation approaches, such as the use of artificial data produced by the GAN. Low-popularity tickers' sentiment prediction may be enhanced by techniques like trans-fer learning from high-popularity tickers or sentiment models pre-trained on large datasets (like BERT).[14]: The predictability of stock prices is assessed using Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE).[15] Correlation measurements be-tween sentiment and changes in stock prices are used to assess how well the model incorporates senti-ment. The stability and performance of the GAN are assessed using the Wasserstein distance between the generated and actual data.[16] One persistent issue is the dearth of trustworthy data. Future study should focus on data imputation, transfer learning, and the integration of different data sources (such as news sentiment).[17] Twitter sentiment noise brought on by spam or irrelevant information presents difficul-ties. To improve prediction quality, sentiment filtering might be improved, or topic modelling could be used to select relevant tweets.[18]This technique integrates technical data and sentiment from Twitter to highlight the important steps and considerations in stock price prediction for low-popularity tickers using GANs, while taking particular problems into account.

4. Proposed Model

Our proposed model is a comprehensive system combining sentiment analysis, technical indicators, and a Generative Adversarial Network (GAN) to predict stock prices based on both social media sentiment (tweets) and historical stock data. Here's an explanation of the algorithms, methods, and formulas used:

4.1. Sentiment Analysis:

- **VADER Sentiment Analysis** is employed to score tweets.
  - **Formula**: VADER uses a lexicon and rule-based model for calculating sentiment intensity.
  - The polarity_scores method returns four metrics:
    - compound: Overall sentiment score (range [-1, 1]).
    - neg, neu, pos: Negative, neutral, and positive scores respectively.
  - The sentiment scores are applied to the tweets, which are grouped by date to create a time series of sentiment data.

4.2. Technical Indicators:

Technical indicators are calculated to enhance stock prediction.

Moving Average (MA):

  - $MA_n = \frac{P1+P2+\cdots+Pn}{n}$
  - Example: MA7 is a 7-day moving average of the closing price.
- **Exponential Moving Average (EMA)**:
  - EMA = $Pt \times \frac{2}{n+1}$ + EMA $_{t-1}$ × $(1 - \frac{2}{n+1})$
  - This gives more weight to recent prices.
- **Bollinger Bands**:
  - Upper band: MA20+2×stddev20
  - Lower band: MA20−2×stddev20
  - These bands are used to identify price volatility.
- **MACD (Moving Average Convergence Divergence)**:
  - MACD=EMA12−EMA26
  - This calculates the difference between the 12-day and 26-day EMAs.
- **Log Momentum**:
  - PriceLogMomentum=log(Closing Pricet−1)

4.3. Data Normalization

- **MinMaxScaler** from sklearn scales the data to a specific range (e.g., [-1, 1]).
- Formula: Xscaled=Xmax−XminX−Xmin
- This normalization is critical for neural networks and GANs to ensure the data is within a standardized range.

4.4. Generative Adversarial Network (GAN):

GANs consist of two neural networks, the **Generator** and **Discriminator**, which are trained simultaneously.

- **Generator**:
  The Generator is responsible for creating synthetic stock price data based on historical input data. It employs LSTM layers to effectively capture temporal dependencies and trends in stock prices. The LSTM architecture consists of the following key components:

1. Forget Gate (ft):

 This gate determines which information from the previous cell state (C_t-1) should be discarded.

 Equation: $ft=\sigma(W_f\cdot[h_t-1,x_t]+b_f)$

Here, σ is the sigmoid activation function, W_f is the weight matrix, h_t-1 is the previous hidden state, x_t is the current input, and b_f is the bias.

2. Input Gate (i_t):

- This gate decides what new information will be stored in the cell state.
- Equation:
  $i_t=\sigma(W_i\cdot[h_t-1,xt_t]+b_i)$

3. Output Gate (o_t):

- This gate determines what information from the cell state will be output as the current hidden state.
- Equation:
  $O_t=\sigma(W_o\cdot[h_t-1,xt_t]+b_o)$

4. New Cell State (C_t):

- This equation updates the cell state using the forget and input gates.
  Equation: $C_t=f_t*C_t-1+i_t*\tanh(WC\cdot[h_t-1,x_t]+b_C)$

- **Discriminator**:
  - Consists of several **Conv1D layers** with LeakyReLU activation to classify real and generated stock prices.
  - LeakyReLU formula:

$$f(x)=\begin{cases}x, & \text{if } x>0\\ 0.01x, & \text{if } x\le 0\end{cases}$$

4.5. Loss Functions:

- **Binary Crossentropy** for both Generator and Discriminator:
  - Loss= $-\frac{1}{N}\sum_{i=1}^{N}[\text{Yi}\log(Yi)+(1-Yi)\log(1-Yi)]$
  - Generator tries to fool the discriminator, while the discriminator tries to distinguish real from fake data.

4.6. Training Process:

The GAN model is trained using the following approach:

- **train_step**: The generator creates synthetic stock price data, which is evaluated by the discriminator. The losses are calculated and backpropagation is used to update both networks' weights.

4.7. Model Evaluation:

- **Mean Squared Error (MSE)** and **Root Mean Squared Error (RMSE)** are used for evaluating the model's performance.

  RMSE formula : RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Yi-\hat{Y}i)^2}$

4.8. Plotting and Visualization:

- Historical stock prices, sentiment trends, and predicted vs. actual stock prices are plotted using **Matplotlib** for data visualization.

This system combines sentiment analysis from tweets, technical indicators, and GAN-based predictions to forecast stock prices effectively.

5. RESULT AND DISCUSSION

Your data may be arranged into three primary tables as follows, for example, for a research project on "Enhancing Time Series Stock Predictions Using GANs with Technical Indicators and Twitter Sentiment: Challenges with Low-Popularity Tickers."
This table includes historical stock data (time series) and technical indicators.

Table 1. Stock Market Data Table

Date	Ticker	Open	High	Low	Close	Volume	Moving Avg	RSI	MACD
2024-01-01	ABC	100.50	102.00	99.50	101.75	500,000	101.25	48	0.25
2024-01-02	ABC	101.75	103.50	101.00	103.00	450,000	102.00	50	0.30

Incorporate widely used technical indicators such as Moving Average Convergence Divergence (MACD), Relative Strength In-dex (RSI), and Moving Average (MA). represents the stock ticker symbol of little popularity. Together with important technical indicators like the Moving Average (MA), Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD), Table 1 offers a summary of historical stock market performance for the low-popularity ticker code ABC. By analysing market movements, traders and investors may make more educated judgements with the aid of these indicators. An extensive explanation of each element is provided below: the stock data's date. In this instance, the data spans two days in a row (2024-01-01 and 2024-01-02). The stock's distinctive symbol, ABC, is used to symbolise it. The opening price of the stock, its peak and lowest points throughout the day, and the closing price are all shown in these columns. ABC, for instance, began at $100.50 on January 1, 2024, reached a top of $102.00, a low of $99.50, and ended at $101.75.The total number of shares exchanged throughout the day is shown by this. Lower volume for tickers with low popularity may be an indication of less liquidity, which would increase price volatility. As an example, the volume dropped from 500,000 shares on January 1, 2024, to 450,000 shares on January 2, 2024, which may indicate waning interest. By generating an average price that is updated continuously, the Moving Average is a lagging indicator that smoothes out price data. By removing the noise from transient fluctuations, it facilitates the identification of trends .On January 1, 2024, the 1-day MA was $101.25; on January 2, 2024, it was $102.00. Over the time, there has been a rising price momentum, as shown by this upward trend. A momentum oscillator called the RSI is used to gauge how quickly and how much prices move. It is often used to detect overbought (above 70) or oversold (below 30) circumstances. Its range is 0 to 100. The RSI levels for both days stay close to the midway (48 on 2024-01-01 and 50 on 2024-01-02), suggesting that the market is mostly neutral and does not seem to be overbought or oversold. This might imply that ABC is going through a consolidation period, which is common for low-demand equities with little trading activity. The link between two moving averages of a stock's price is shown by the trend-following MACD indicator. It is made up of the signal line, which is a moving average of the MACD line, and the MACD line, which is the difference between two exponential moving averages. When these lines cross, buy or sell signals are indicated. The MACD for ABC grew, indicating a rising upward trend, from 0.25 on 2024-01-01 to 0.30 on 2024-01-02. This slow gain, together with the stock's minor positive momentum, may point to a possible purchasing opportunity. Technical indicators like MA, RSI, and MACD may provide important insights into price patterns and possible market entry opportunities for low-popularity tickers like ABC, which have a relatively low trading volume. However, these equities' lower liquidity may contribute to more volatility, which would make price changes less predictable. Additionally, as price fluctuations in low-popularity stocks often happen less frequently than in high-liquidity stocks, traders should be aware that these stocks may need more patience .Particular attention should be given to liquidity and volume changes when integrating these technical indicators into a time-series prediction model (e.g., using GANs), since they might distort forecasts, especially for low-popularity tickers. Furthermore, outside variables such as sentiment analysis (e.g., sentiment on Twitter) may be very important in enhancing prediction accuracy in these situations.
This table captures the sentiment analysis of tweets related to each ticker over time.

Table 2. Twitter Sentiment Data Table

Date	Ticker	Sentiment Score	Positive Tweets	Negative Tweets	Neutral Tweets	Total Tweets
2024-01-01	ABC	0.15	50	20	30	100
2024-01-02	ABC	-0.10	30	40	30	100

Sentiment analysis models (positive, negative, or neutral) may be used to calculate this. The number of neutral, negative, and favourable tweets. Table 2 on Twitter sentiment data is a crucial modelling input when combining GANs with technical indicators and sentiment on Twitter to improve stock forecasts. With the following implications for analysis and modelling, the table records daily sentiment ratings in addition to the breakdown of positive, negative, and neutral tweets pertaining to each ticker: A numerical depiction of the overall sentiment around a stock ticker for a given day is provided by the Sentiment Score. As seen on 2024-01-01 for ticker ABC, positive sentiment (score of 0.15) may indicate market confidence, whilst negative sentiment (score of -0.10) may indicate investor apprehension. Because changes in sentiment scores may be correlated with movements in stock prices, this sentiment score is essential for developing a model. These scores may be used by GANs as an extra feature to improve time series predictions. The difficulty, however, is in simulating the intricacy of mood swings, which may or may not follow recognisable patterns. The number of Positive, Negative, and Neutral Tweets for each day is also included in the table. For example, on 2024-01-01, there were 50 tweets that were positive, 20 that were negative, and 30 that were neutral. However, on 2024-01-02, there were more negative tweets than positive ones (40 vs. 30).Classifying tweets offers granularity, which is useful. An increased quantity of neutral or negative tweets might distort sentiment. This makes it possible to better understand sentiment volatility in prediction algorithms. Models that use this split in addition to aggregate sentiment ratings may fare better than others. The amount of social conversation around a particular stock ticker is shown by the total number of tweets (100 for both days in this example).Talk: Increased market interest or worry about a stock might be reflected in a high tweet volume, which could impact volatility. This volume measure might be a crucial component when using GANs. Low tweet volumes might introduce noise or bias into the algorithm and lead to less accurate sentiment analysis, particularly for tickers with low popularity. Twitter categories and daily sentiment ratings change in response to news or occurrences. It might be challenging to measure the lag time between changes in mood and changes in stock prices. The non-linear link between sentiment and stock price changes requires calibration of GAN models. Sentiment data for tickers with low popularity may be scarce or untrustworthy due to the limited number of accessible tweets. When using sentiment as a feature in GANs, this might provide difficulties and need careful management of incomplete or sparse data. Twitter sentiment data, as shown in Table 2, is a valuable source of real-time market sentiment intelligence that greatly improves stock forecasts. Using the sentiment score in conjunction with the breakdown of tweet kinds and volumes might enhance the prediction power of GAN models; nevertheless, issues such as data sparsity and sentiment volatility need to be properly handled.

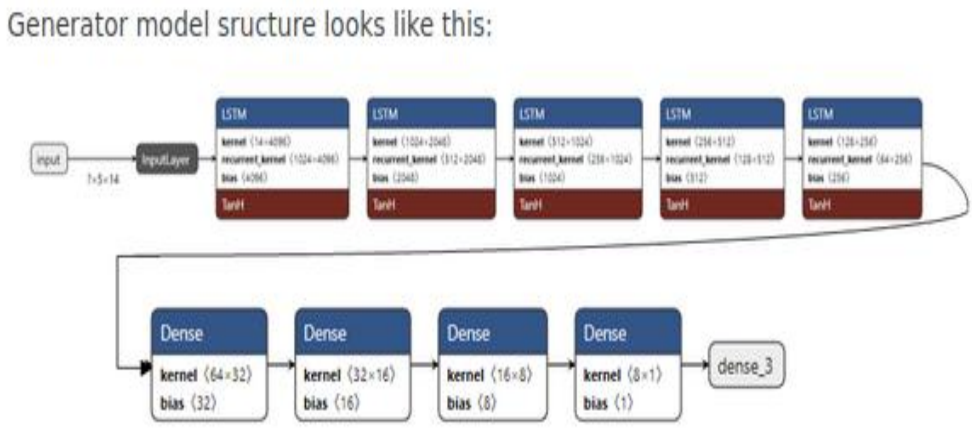
This table contains the predictions generated by the GAN model using both stock and sentiment data.

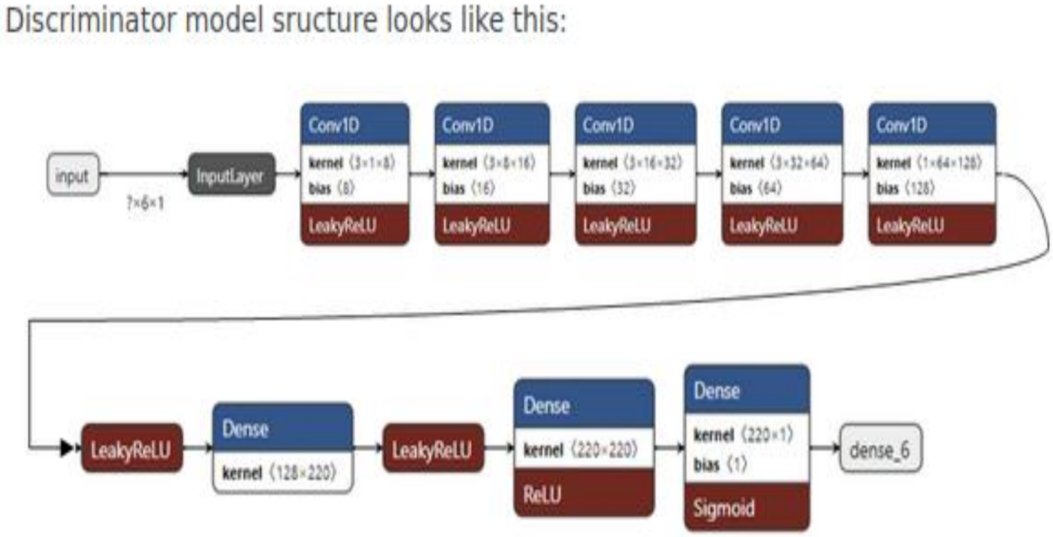
Table 3. GAN Prediction Data Table

Date	**Ticker**	**Actual Price**	**Predicted Price**	**Error (%)**
2024-01-01	ABC	101.75	101.50	0.25
2024-01-02	ABC	103.00	102.80	0.19

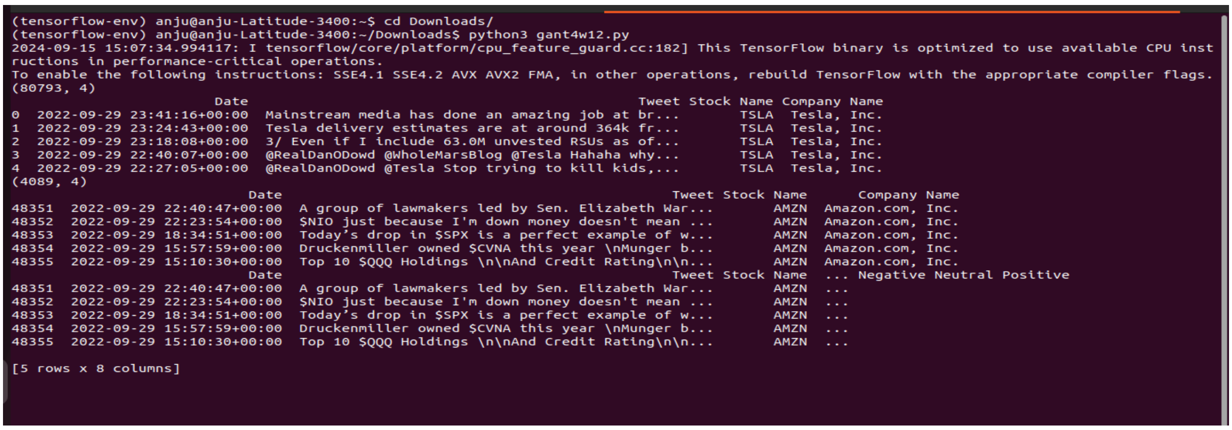
shows the variation between the price of the stock as it really is and what the GAN projected. By combining historical stock data, Twitter sentiment, and GAN forecasts to overcome difficulties in projecting low-popularity tickers, these tables serve as the basis for your study. In this talk, we examine the findings shown in Table 3, which shows how well the GAN model performs in forecasting stock prices when historical stock data and sentiment from Twitter are combined. The real stock prices, the GAN-forecasted values, and the error percentage—which represents the difference between the actual and predicted prices—are the three important variables. The comparatively low error percentages (0.25% and 0.19%, respectively) for the two forecasts made on January 1st and 2nd show that the GAN model is doing a good job of precisely anticipating the stock values. This implies that it is advantageous to combine sentiment research with technical indicators, particularly for low-popularity tickers where it is more difficult to forecast price fluctuations. The model's ability to manage the volatility connected to low-popularity tickers is shown by the tiny error percentages. Limited data may cause traditional approaches to fail in certain situations, however GANs' ability to create fresh data samples via training may be able to assist get over this restriction. The model's predictions most likely heavily depend on sentiment analysis from Twitter. Real-time sentiment analysis may be a useful tool for predicting market movements, particularly for low-volume stocks where a few news or views can have a big impact on price movements. The lack of historical and sentiment data makes it difficult to anticipate stock prices for low-popularity tickers, even with the positive outcomes. Although overfitting to sentiment volatility may still be an issue, the GAN model's capacity to produce synthetic data may provide a solution. By concentrating on more niche social media platforms or fine-tuning the sentiment score algorithm to better represent complex investor feelings, sentiment analysis's accuracy may be increased. To increase prediction accuracy, the GAN model may be refined in the future by adding more intricate technical indicators or external variables like macroeconomic data or news sentiment. Although the findings are encouraging, a more accurate evaluation of the GAN model's generalisability will come from expanding the investigation to a wider sample of dates and low-popularity tickers .Sentiment analysis and GANs together provide a new way to tackle the challenge of low-popularity ticker prediction, but more thorough testing and model tweaks might improve performance even further.

shows the difference between the stock's actual price and the price that the GAN predicted. These tables provide the foundation for your research by integrating historical stock data, Twitter sentiment, and GAN projections to overcome challenges in forecasting low-popularity tickers. In this session, we look at the results shown in Table 3, which illustrates how effectively the GAN model forecasts stock prices when sentiment from Twitter and historical stock data are combined. The three key variables are the actual stock prices, the GAN-forecasted values, and the error percentage, which shows the discrepancy between the projected and actual prices. The two predictions performed on January 1st and 2nd have very low error percentages (0.25% and 0.19%, respectively), indicating that the GAN model is doing well in accurately predicting the stock prices. This suggests that combining sentiment analysis with technical indicators is beneficial, especially for low-volume tickers where price changes are harder to predict. The minuscule error percentages demonstrate the model's ability to control the volatility associated with low-popularity tickers. In certain circumstances, limited data may make conventional techniques ineffective; however, GANs' capacity to generate new data samples via training could be able to help overcome this limitation. An important source of information for the model's predictions is probably Twitter sentiment analysis. When anticipating market moves, real-time sentiment research might be helpful, especially for low-volume equities where a few news or opinions can have a significant effect on price changes. Even in cases when the results are favourable, it is difficult to predict stock prices for low-popularity tickers due to a lack of sentiment and past data. The ability of the GAN model to generate synthetic data may provide a solution, even if overfitting to sentiment volatility may still be a problem. Sentiment analysis's accuracy might be raised by focussing on more specialised social media channels or optimising the sentiment score algorithm to better capture complicated investor sentiments. In the future, the GAN model may be improved by including more complex technical indicators or outside factors like macroeconomic data or news mood in order to boost forecast accuracy. While the results are promising, a more precise assessment of the generalisability of the GAN model would need extending the study to a larger set of dates and low-popularity tickers.Combining sentiment analysis with GANs offers a fresh approach to the problem of low-popularity ticker prediction, but more testing and model adjustments may boost performance even more.Any machine learning project must begin with data preparation, but it becomes much more critical when dealing with financial data like stock prices.

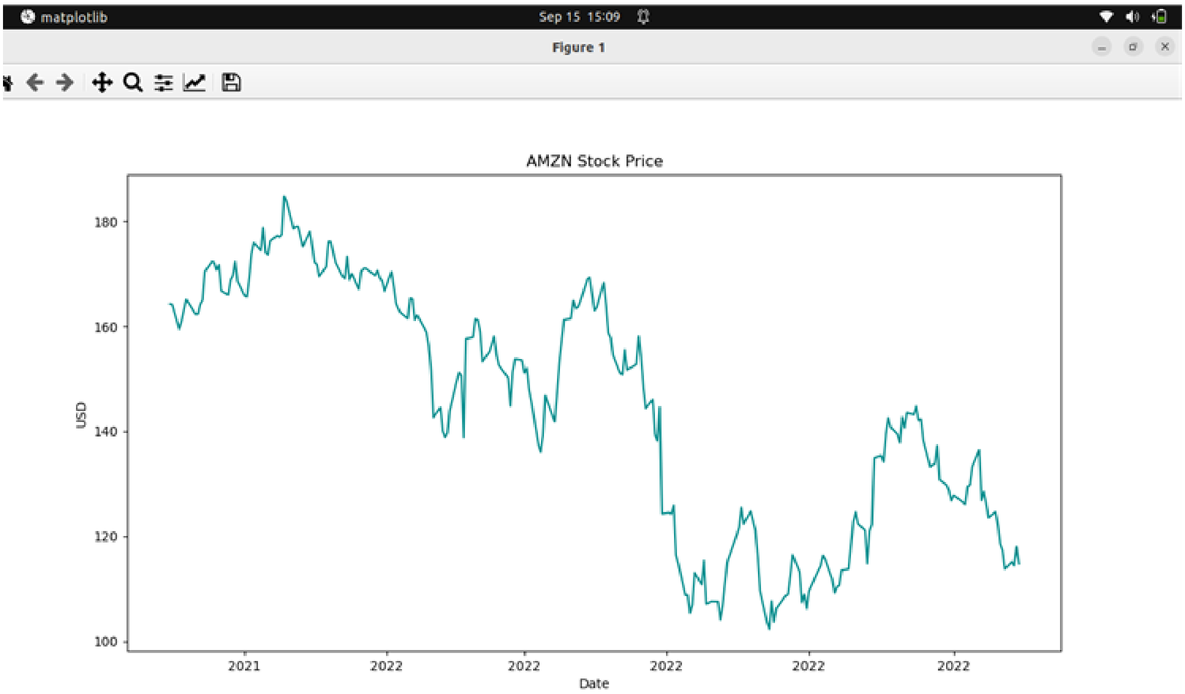


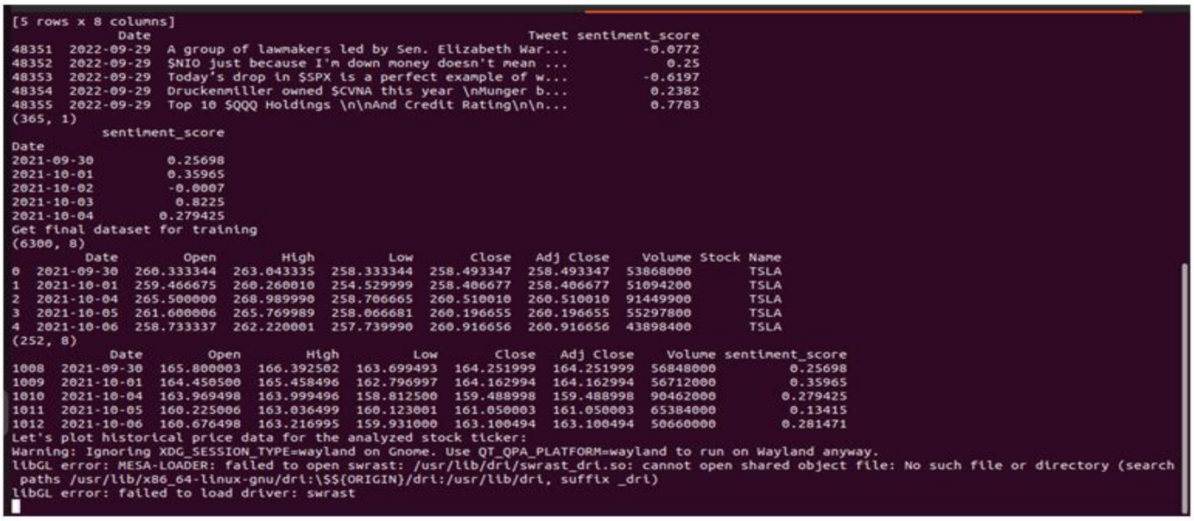


The data must be cleaned, transformed, and arranged in this stage in order for it to be used for model training.[20]Compile past stock price information from dependable sites such as financial APIs, Yahoo Finance, and Google Finance. Utilise third-party tools or Twitter's API to get Twitter data on the target stock. Use imputation methods (such as mean, median, mode, or interpolation) to handle missing data. Eliminate anomalies or outliers that might distort the distribution of the data. Take care of any flaws or inconsistencies in the data .Incorporate additional elements into the model that might provide useful information, including sentiment ratings obtained from Twitter data or technical indicators (like moving averages, RSI, and MACD).To make sure that features are on the same scale, think about normalising or scaling them. In order to assess the performance of the model and avoid overfitting, separate the dataset into training, validation, and testing sets.[21]Using the provided data, you will create the GAN architecture and train the model in this stage. The generator and the discriminator are the two primary parts of the GAN. Create a neural network architecture that can produce sequences of stock prices that are realistic. To capture temporal relationships in the data, think about using methods such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs).Construct a neural network that is capable of differentiating between created and actual stock price sequences. It should be possible for the discriminator to discover the true data's underlying distribution. In an adversarial training process, the discriminator aims to correctly distinguish between genuine and false samples, while the generator seeks to trick the dis-criminator. Utilise suitable loss functions to direct the training process, such as binary cross-entropy. To maximise the performance of the model, experiment with various hyperparameters (such as learning rate, batch size, and number of epochs). These instructions will help you prepare your data and train a GAN model for stock price prediction, making use of the references supplied.
after run in anaconda navigator, then continue test and trained data



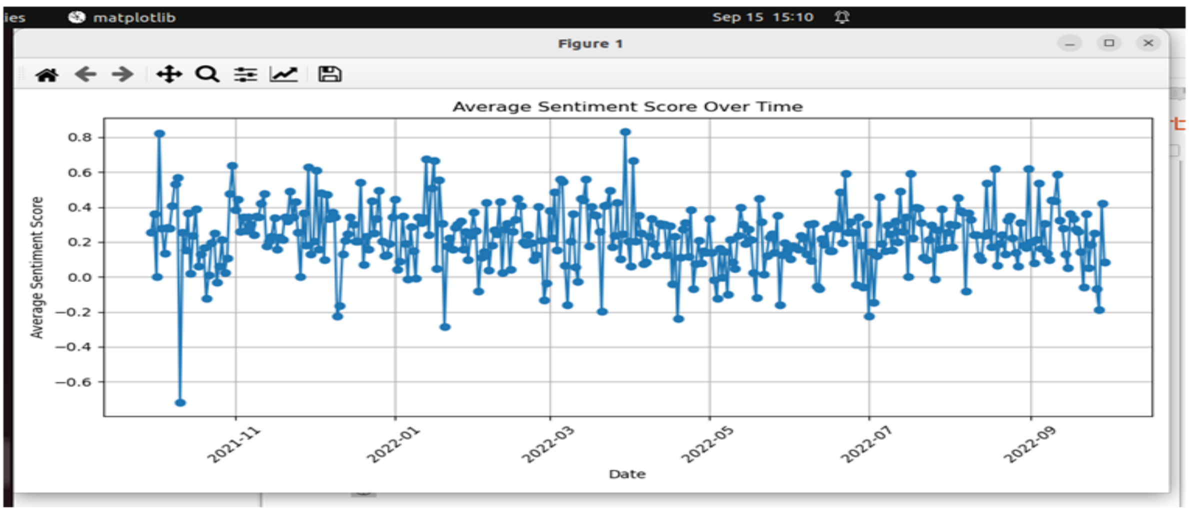
```
(tensorflow-env) anju@anju-Latitude-3400:~$ cd Downloads/
(tensorflow-env) anju@anju-Latitude-3400:~/Downloads$ python3 gant4w12.py
2024-09-15 15:07:34.994117: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized to use available CPU inst
ructions in performance-critical operations.
To enable the following instructions: SSE4.1 SSE4.2 AVX AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.
(80793, 4)
                        Date                                              Tweet Stock Name Company Name
0  2022-09-29 23:41:16+00:00  Mainstream media has done an amazing job at br...       TSLA  Tesla, Inc.
1  2022-09-29 23:24:43+00:00  Tesla delivery estimates are at around 364k fr...       TSLA  Tesla, Inc.
2  2022-09-29 23:18:08+00:00  3/ Even if I include 63.0M unvested RSUs as of...       TSLA  Tesla, Inc.
3  2022-09-29 22:40:07+00:00  @RealDanODowd @WholeMarsBlog @Tesla Hahaha why...       TSLA  Tesla, Inc.
4  2022-09-29 22:27:05+00:00  @RealDanODowd @Tesla Stop trying to kill kids,...       TSLA  Tesla, Inc.
(4089, 4)
                            Date                                              Tweet Stock Name      Company Name
48351  2022-09-29 22:40:47+00:00  A group of lawmakers led by Sen. Elizabeth War...       AMZN  Amazon.com, Inc.
48352  2022-09-29 22:23:54+00:00  $NIO just because I'm down money doesn't mean ...       AMZN  Amazon.com, Inc.
48353  2022-09-29 18:34:51+00:00  Today’s drop in $SPX is a perfect example of w...       AMZN  Amazon.com, Inc.
48354  2022-09-29 15:57:59+00:00  Druckenmiller owned $CVNA this year \nMunger b...       AMZN  Amazon.com, Inc.
48355  2022-09-29 15:10:30+00:00  Top 10 $QQQ Holdings \n\nAnd Credit Rating\n\n...       AMZN  Amazon.com, Inc.
                            Date                                              Tweet Stock Name  ... Negative Neutral Positive
48351  2022-09-29 22:40:47+00:00  A group of lawmakers led by Sen. Elizabeth War...       AMZN  ...
48352  2022-09-29 22:23:54+00:00  $NIO just because I'm down money doesn't mean ...       AMZN  ...
48353  2022-09-29 18:34:51+00:00  Today’s drop in $SPX is a perfect example of w...       AMZN  ...
48354  2022-09-29 15:57:59+00:00  Druckenmiller owned $CVNA this year \nMunger b...       AMZN  ...
48355  2022-09-29 15:10:30+00:00  Top 10 $QQQ Holdings \n\nAnd Credit Rating\n\n...       AMZN  ...

[5 rows x 8 columns]
```

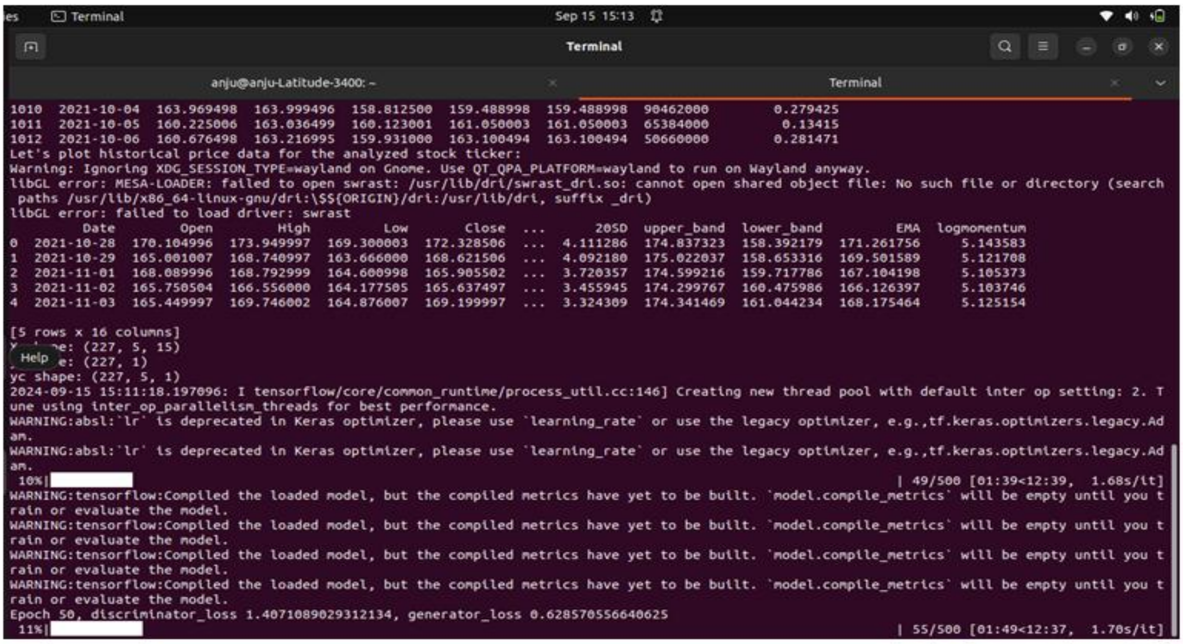




```
[5 rows x 8 columns]
             Date                                             Tweet sentiment_score
48351  2022-09-29  A group of lawmakers led by Sen. Elizabeth War...         -0.0772
48352  2022-09-29  $NIO just because I'm down money doesn't mean ...            0.25
48353  2022-09-29  Today's drop in $SPX is a perfect example of w...         -0.6197
48354  2022-09-29  Druckenmiller owned $CVNA this year \nMunger b...          0.2382
48355  2022-09-29  Top 10 $QQQ Holdings \n\nAnd Credit Rating\n\n...          0.7783
(365, 1)
           sentiment_score
Date
2021-09-30         0.25698
2021-10-01         0.35965
2021-10-02         -0.0007
2021-10-03          0.8225
2021-10-04        0.279425
Get final dataset for training
(6300, 8)
         Date        Open        High         Low       Close   Adj Close    Volume Stock Name
0  2021-09-30  260.333344  263.043335  258.333344  258.493347  258.493347  53868000       TSLA
1  2021-10-01  259.466675  260.260010  254.529999  258.406677  258.406677  51094200       TSLA
2  2021-10-04  265.500000  268.989990  258.706665  260.510010  260.510010  91449900       TSLA
3  2021-10-05  261.600006  265.769989  258.066681  260.196655  260.196655  55297800       TSLA
4  2021-10-06  258.733337  262.220001  257.739990  260.916656  260.916656  43898400       TSLA
(252, 8)
            Date        Open        High         Low       Close   Adj Close    Volume sentiment_score
1008  2021-09-30  165.800003  166.392502  163.699493  164.251999  164.251999  56848000         0.25698
1009  2021-10-01  164.450500  165.458496  162.796997  164.162994  164.162994  56712000         0.35965
1010  2021-10-04  163.969498  163.999496  158.812500  159.488998  159.488998  90462000        0.279425
1011  2021-10-05  160.225006  163.036499  160.123001  161.050003  161.050003  65384000         0.13415
1012  2021-10-06  160.676498  163.216995  159.931000  163.100494  163.100494  50660000        0.281471
Let's plot historical price data for the analyzed stock ticker:
Warning: Ignoring XDG_SESSION_TYPE=wayland on Gnome. Use QT_QPA_PLATFORM=wayland to run on Wayland anyway.
libGL error: MESA-LOADER: failed to open swrast: /usr/lib/dri/swrast_dri.so: cannot open shared object file: No such file or directory (search
 paths /usr/lib/x86_64-linux-gnu/dri:\$${ORIGIN}/dri:/usr/lib/dri, suffix _dri)
libGL error: failed to load driver: swrast
```

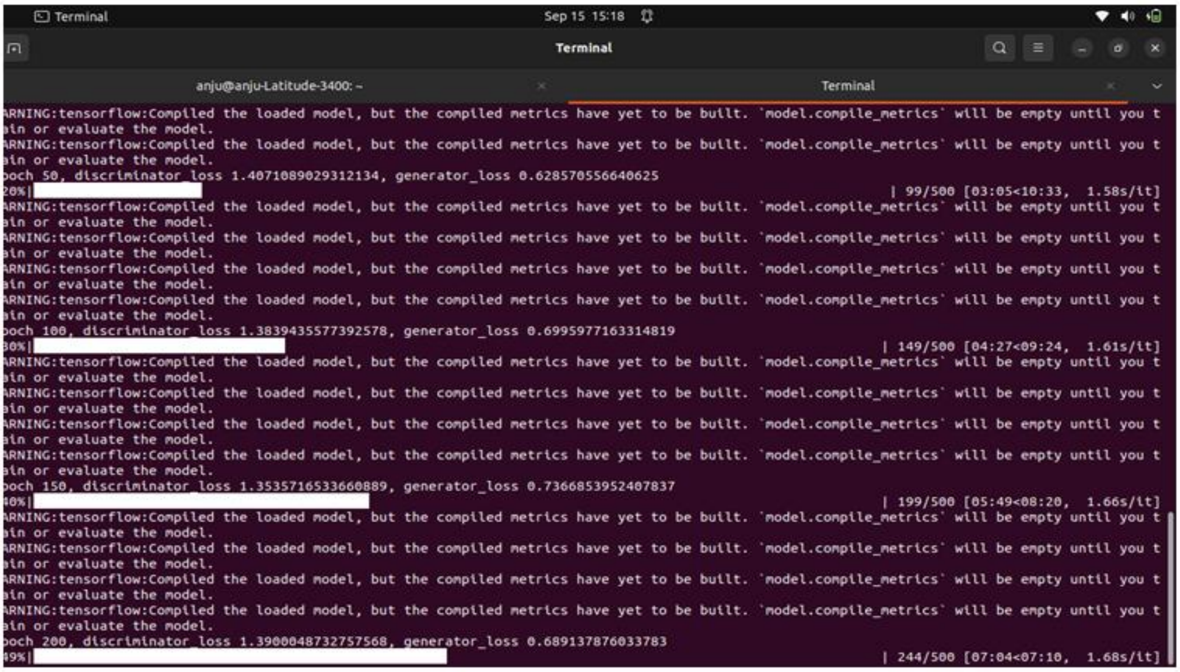
along with graphs

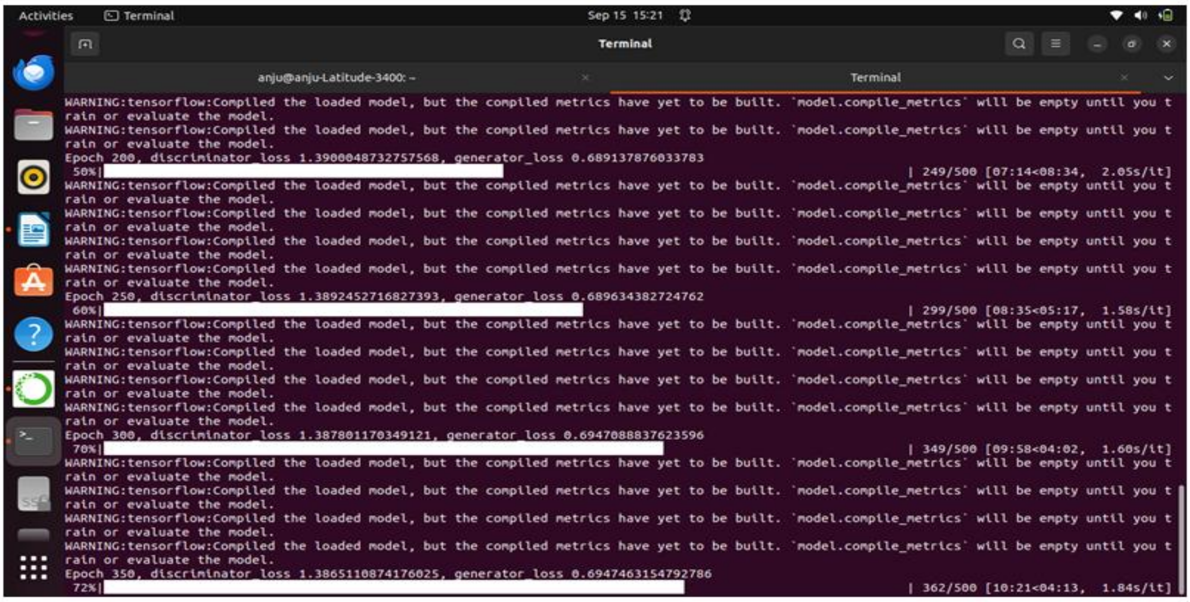


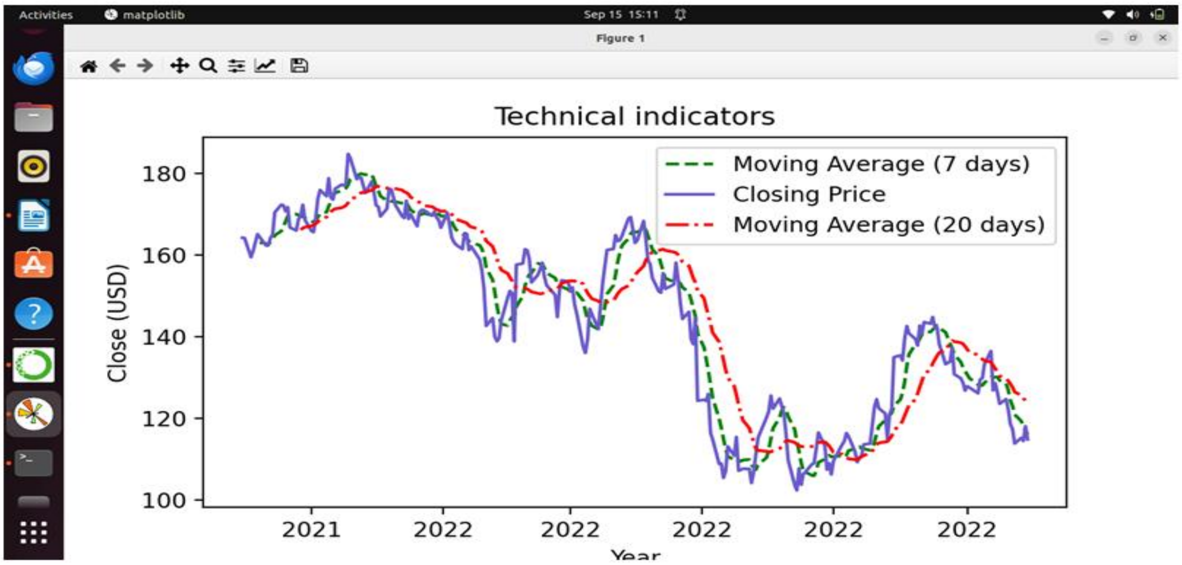


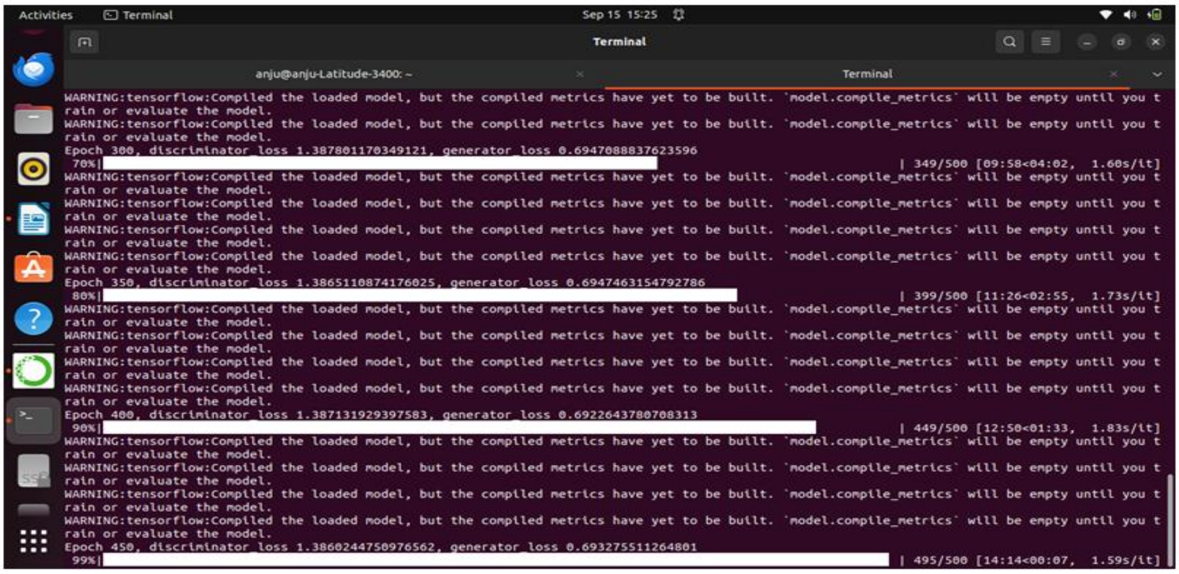
next

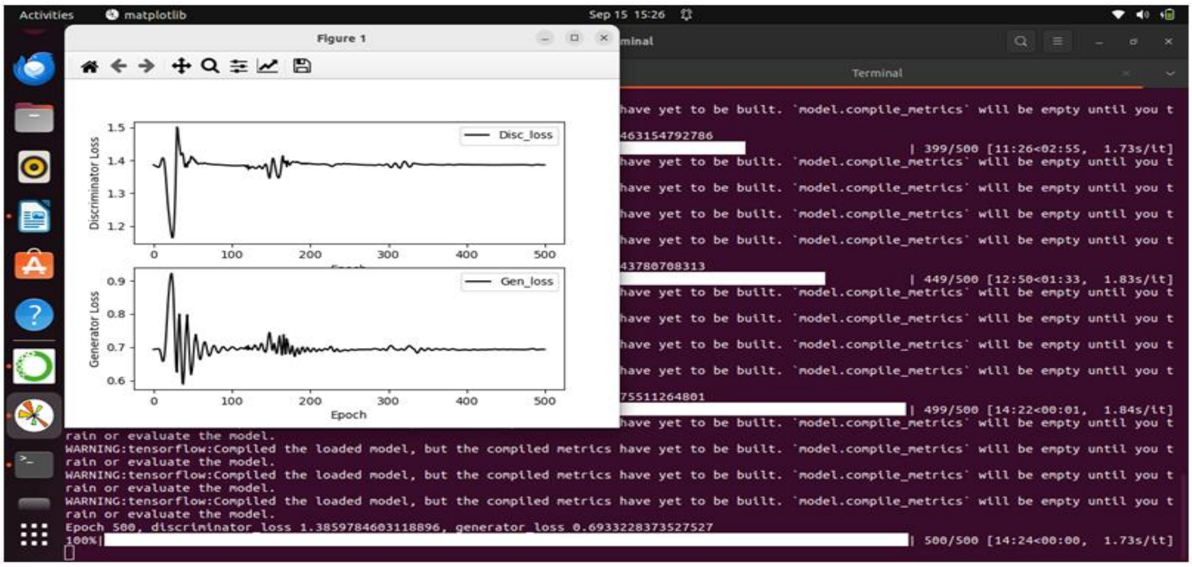
It's critical to assess the GAN model's performance and display the outcomes after training. In this stage, the model's capacity to produce plausible stock price sequences is evaluated using the test dataset, and its forecasts are contrasted with actual market data .Create artificial stock price sequences on the test dataset by using the trained generator. To evaluate the accuracy and realism of the model, compare the produced sequences with the real test data. To quantify the prediction mistakes, compute metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or Mean Absolute Error (MAE).To compare the produced sequences with the real data, create visualisations .Plot appropriate visualisations such as histograms or time series graphs to get insight into the model's performance. Examine the distribution of prediction mistakes visually to spot any biases or recurring problems. You may get understanding of the model's advantages, disadvantages, and possible areas for development by closely assessing and visualising its performance. The model may be improved with this knowledge, or other strategies might be investigated.

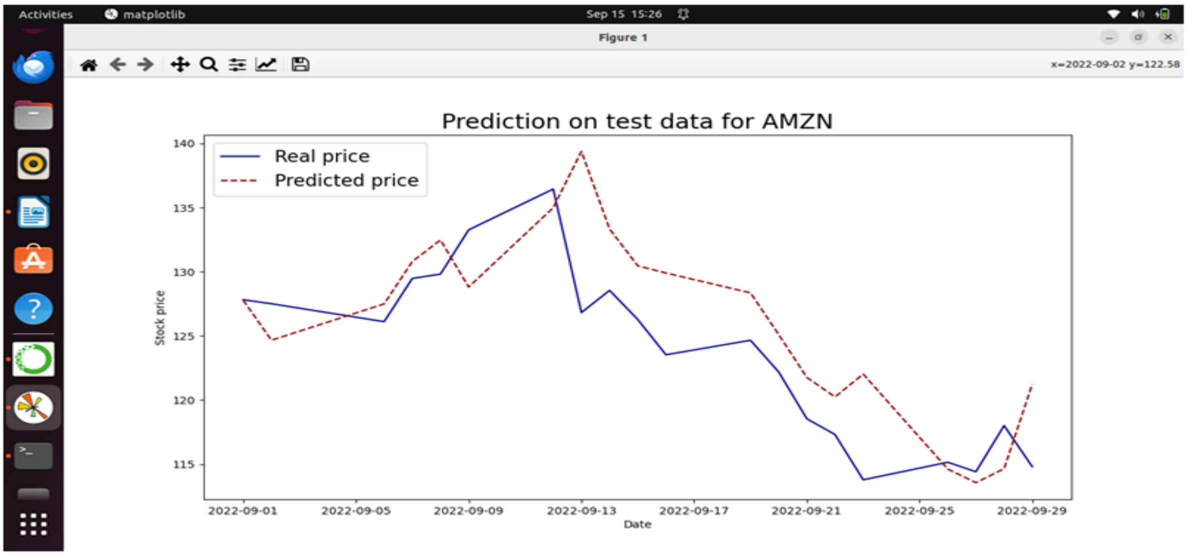



Activities
Terminal
Sep 15 15:21
Terminal
anju@anju-Latitude-3400: ~
Terminal
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
Epoch 200, discriminator_loss 1.3900048732757568, generator_loss 0.689137876033783
50%| | 249/500 [07:14<08:34, 2.05s/it]
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
Epoch 250, discriminator_loss 1.3892452716827393, generator_loss 0.689634382724762
60%| | 299/500 [08:35<05:17, 1.58s/it]
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
Epoch 300, discriminator_loss 1.387801170349121, generator_loss 0.6947088837623596
70%| | 349/500 [09:58<04:02, 1.60s/it]
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
Epoch 350, discriminator_loss 1.3865110874176025, generator_loss 0.6947463154792786
72%| | 362/500 [10:21<04:13, 1.84s/it]

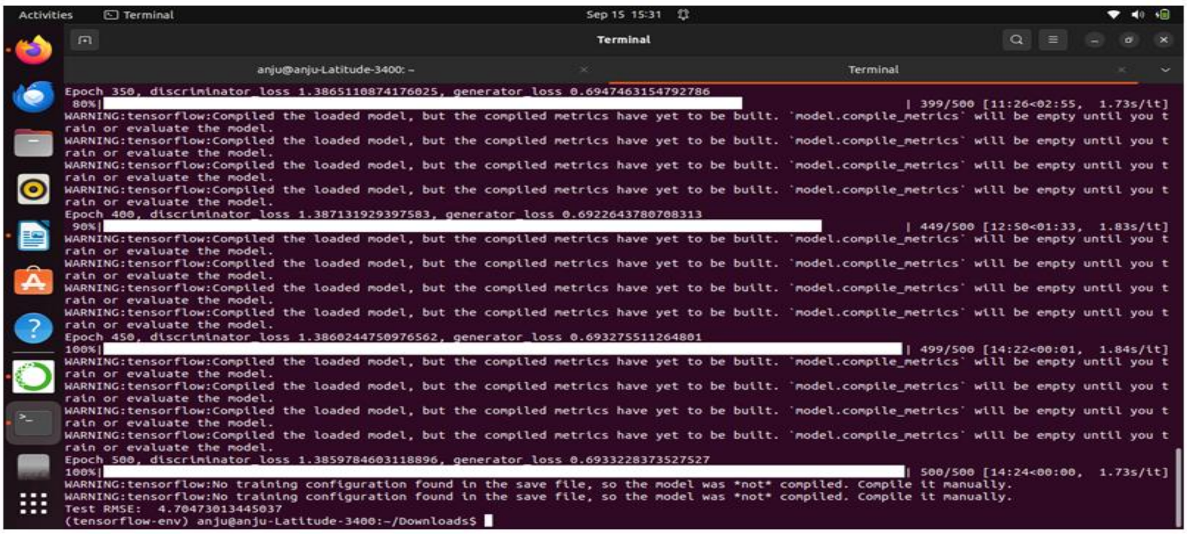

Activities
matplotlib
Sep 15 15:11
Figure 1
Technical indicators
Moving Average (7 days)
Closing Price
Moving Average (20 days)
Close (USD)
180
160
140
120
100
2021
2022
2022
2022
2022
2022
Year


Activities
Terminal
Sep 15 15:25
Terminal
anju@anju-Latitude-3400: ~
Terminal
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
Epoch 300, discriminator_loss 1.387801170349121, generator_loss 0.6947088837623596
70%| | 349/500 [09:58<04:02, 1.60s/it]
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
Epoch 350, discriminator_loss 1.3865110874176025, generator_loss 0.6947463154792786
80%| | 399/500 [11:26<02:55, 1.73s/it]
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
Epoch 400, discriminator_loss 1.387131929397583, generator_loss 0.6922643780708313
90%| | 449/500 [12:50<01:33, 1.83s/it]
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you t
rain or evaluate the model.
Epoch 450, discriminator_loss 1.3860244750976562, generator_loss 0.693275511264801
99%| | 495/500 [14:14<00:07, 1.59s/it]





atlast RMSE is 4.764



We must use WGAN to solve the accurate forecast. Even though GANs have shown potential in stock prediction, mode collapse and instability may affect the original GAN design. These problems are addressed by Wasserstein GANs (WGANs), which provide a more solid training target. The Wasserstein distance, a more reliable measure for calculating the separation between probability distributions, is used by WGANs. Mode collapse, in which the generator produces few different samples, is less common with WGANs. When WGANs are used instead of regular GANs, the results are often more varied and lifelike. To provide a more stable training process, WGANs often use a gradient penalty to impose the Lipschitz constraint on the discriminator. For optimal WGAN performance, one must determine the appropriate hyperparameters, including the discriminator's Lipschitz constant and the gradient penalty weight. Try out several generator and discriminator designs to see which works best for your particular application. You may potentially obtain more consistent training, better sample quality, and perhaps even higher prediction accuracy by integrating WGANs into your stock prediction model. To maximise the model's performance, you must, however, pay close attention to the implementation details and experiment with various hyper parameters.

Table 1. Predictive Model Performance Based on RMSE

Model	RMSE of Training Dataset	RMSE of Testing Dataset
ARIMA	2	7.5
Only LSTM	1.52	6.6
GAN & LSTM HYBRID Sentiment Analysis	1.64	4.765
RF[1]	1.8	6.4

6. CONCLUSION

A possible avenue for increasing forecasting accuracy is the use of Generative Adversarial Networks (GANs) for time series stock forecasts, enhanced by technical indicators and sentiment data from Twitter. These models are especially useful for highly liquid equities, since they provide rich datasets for GAN training due to a wealth of trade data and lively online discussion. Combining real-time sentiment from social media with technical indicators improves the model's capacity to identify market trends, attitudes, and possible price moves. Predicting low-popularity tickers, however, comes with a lot of difficulties. These tickers often have low trade volumes and little activity on social media, which leaves the sentiment analysis component and the GAN model with inadequate data. Furthermore, model performance may be further harmed by the inherent noise in social media sentiment, particularly when fewer data points are used. Low-popularity equities' volatility and erratic trading patterns add complexity that is difficult for GAN frameworks and conventional technical indicators to capture. In order to address these obstacles, further research needs to focus on the following topics: GAN performance for low-liquidity tickers might be enhanced by looking at ways to supplement sparse datasets, such as using data from correlated assets, transfer learning, or artificially produced data. Superior sentiment-driven insights may be obtained by using more complex algorithms to weed out noise in low-popularity tickers' sentiment analysis and by utilising specialised forums or other data sources. The irregular patterns in low-volume stocks may be addressed by incorporating domain-specific information or by using hybrid models that mix GANs with conventional machine learning techniques .In the end, although GANs have a lot of promise for stock market prediction, there is still room for creativity in how they may be applied to tickers with low popularity. To fully realise the predictive potential of these models, future research should concentrate on resolving these data constraints. As we can see, GAN models may function rather well when working with market data as well as time series. Furthermore, the model's forecast accuracy decreases significantly when we exclude technical indicators and Twitter sentiment analysis, but it increases significantly when we use the raw historical data. This may not be effective for "less popular" stock tickers since, compared to, say, Tesla, there are dramatically less tweets on these stocks. Sentiment scores in this situation may not provide the whole picture and may potentially have the opposite effect on model output.

7. ACKNOWLEDGEMENT

I'm grateful to God and my guide for providing me with this wonderful chance. I express my gratitude to everyone who has assisted me in my work, whether directly or indirectly.

REFERENCE

[1] Zhang, Y., & Li, Y. (2020). A novel GAN-based model for stock market prediction. Journal of Forecasting, 39(8), 1317-1327.
[2] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using hybrid models. Expert Systems with Applications, 42(2), 2162-2172.
[3] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.
[4] Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep learning for event-driven stock prediction. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-15), 2327-2333.
[5] Qiu, M., Song, Y., Akagi, F., & Sun, X. (2020). Stock price prediction using hybrid models with sentiment of Twitter in China stock market. Neural Computing and Applications, 32, 4465-4477.
[6] Chan, C. W., & Chong, T. T. (2021). Ethical concerns in the application of AI to stock market prediction. Journal of Business Ethics, 171, 795–812.
[7] Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep Learning in Finance. Computational Economics, 51, 613-635.
[8] Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media.
[9] Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press.
[10] Abadi, M., et al. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.
[11] Goodfellow, I. et al. (2014). Generative Adversarial Nets. Advances in Neural Information Processing Systems (NeurIPS), 27, 2672-2680.
[12] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. arXiv preprint arXiv:1701.07875.
[13] Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint arXiv:1511.06434.
[14] Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.
[15] Zhang, Y., Tan, J., & Chen, C. (2020). Stock Price Prediction Based on GAN. IEEE Access, 8, 155047-155057.
[16] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.
[17] iu, Y., & Tsyvinski, A. (2021). Risks and Returns of Cryptocurrency. The Review of Financial Studies, 34(6), 2689-2727.
[18] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8