SENTIMENT SCORE COMPUTATION USING NEWS SENTIMENT ANALYSIS AND STOCK PRICE PREDICTION USING GAN + TWITTER SENTIMENT ANALYSIS

Ananya Jaikumar [1], Kunal Sharma [2]

[1] New Jersey Institute of Technology, Newark, NJ 07102 USA
[2] New Jersey Institute of Technology, Newark, NJ 07102 USA

ABSTRACT In the dynamic realm of stock market analysis, this report delves into the fusion of an innovative approach to stock market analysis by offering users two distinctive options. The first option provides real-time and week-long sentiment scores of a stock, derived from an analysis of current and historical news headlines. Leveraging VADER sentiment analysis, this approach delves into the intricate sentiments embedded in textual data from financial news articles. The second option introduces an advanced predictive model that analyzes historical tweet data alongside stock price data using a Generative Adversarial Network (GAN). This model is meticulously built and trained to forecast stock prices, providing users with a forward-looking perspective. The integration of sentiment analysis from diverse sources, including news and tweets, coupled with predictive modeling, enriches our understanding of market dynamics. This comprehensive approach empowers users to make more informed decisions in the face of market volatility, contributing to a resilient and adaptive financial ecosystem.

INDEX TERMS Beautiful Soup, Bollinger Bands, Finviz, Flask, Generative Adversarial Network (GAN), Machine Learning, matplotlib, News Headlines, numpy, Relative Strength Index, Sentiment Analysis, Stock Prediction, VADER.

1. INTRODUCTION

In the rapidly evolving landscape of financial markets, the stock market has emerged as a focal point of interest, particularly among millennials seeking to capitalize on monetary gains. The accessibility of stock trading platforms and the ease of entry into the market have drawn a diverse range of participants. As individuals venture into this dynamic realm, two pivotal questions inevitably arise: What factors drive changes in stock prices, and at what price points should one consider buying or selling stocks?

This research endeavors to delve into the intricate interplay between technical analysis, which traditionally evaluates structured financial and fundamental data, and the burgeoning field of sentimental analysis, which explores the impact of market participants' sentiments primarily derived from unstructured sources such as financial news, social media, blogs, and web-based forums. By combining these analytical approaches, we aim to provide a comprehensive understanding of the forces shaping stock prices.

In this research endeavor, we embark on a journey to redefine the landscape of stock market analysis, presenting users with a novel and comprehensive approach to augment their decision-making capabilities. Our exploration unfolds through two distinctive options, each tailored to provide users with nuanced insights and forward-looking perspectives.

The first option is a dynamic real-time and week-long sentiment scoring system for stocks, achieved through a meticulous analysis of both current and historical news headlines. Utilizing the cutting-edge VADER sentiment analysis, our approach unravels the intricate sentiments embedded in the textual fabric of financial news articles. By offering users a detailed and timely understanding of the sentiment landscape, this option serves as a valuable resource for navigating the complexities of the stock market.

The second option introduces an advanced predictive model that extends beyond conventional methodologies. This model harnesses the power of Generative Adversarial Networks (GANs) to analyze historical tweet data in conjunction with stock price information. Meticulously crafted and rigorously trained, this predictive model provides users with forecasts for stock prices, presenting a forward-looking perspective. By seamlessly integrating sentiment analysis from diverse sources, encompassing both news headlines and tweets, with state-of-the-art predictive modeling, this option enriches our understanding of the intricate dynamics governing market behavior.

The convergence of sentiment analysis and predictive modeling forms the cornerstone of our research, offering users a holistic and multi-faceted toolkit for navigating the dynamic and often volatile terrain of financial markets. This comprehensive approach is designed to empower users, be

they seasoned investors or newcomers, with valuable insights that foster more informed decision-making. In doing so, we contribute to the development of a financial ecosystem characterized by resilience, adaptability, and a heightened capacity for navigating the complexities of the contemporary stock market.

2. RELATED WORK

In the pursuit of refining stock price prediction models, researchers have explored diverse methodologies, integrating innovative techniques to capture the dynamic relationship between financial news, sentiments, and market movements. Several noteworthy studies contribute valuable insights to this evolving field.

[1] employed a distinct strategy, employing a Support Vector Machine (SVM) to label news articles. The resulting model achieved an impressive 83% accuracy compared to a random news labeling accuracy of 51%. Falinouess demonstrated the efficacy of SVM in discerning sentiment and highlighted the potential of machine learning algorithms in extracting meaningful signals from financial news.[2] adopted a comprehensive approach by categorizing news articles into five distinct subsets based on their relevance to specific stocks, sub-industries, industries, groups of industries, and sectors. Employing various kernels for learning, the Math Kernel Library (MKL) demonstrated promising results, emphasizing the significance of nuanced categorization in capturing diverse market influences.Building on this foundation,[3] concentrated on predicting intraday stock trends using financial articles, achieving a 63.58% accuracy on average. The study highlighted the importance of feature selection based on sentiment scores and utilized linear regression, random forest, and Gradient Boosting Machine Algorithm to enhance predictive performance. [4] explored the relationship between news-driven information and implied volatility. Focusing on predicting changes in volatility rather than close prices, their model showcased the potential of news data in forecasting market dynamics. Implied volatility, derived from option pricing formulas, served as a valuable indicator for assessing the impact of news on stock behavior.

The forecasting of stock market trends necessitates a comprehensive analysis, integrating both fundamental financial indicators and real-world opinions. While fundamental analysis covers aspects like financial reports, quarterly balances, dividends, and audit reports, the hybrid approach acknowledges the significance of external factors such as public opinions and rumors, which can significantly impact stock prices. Recognizing the value of each forecasting model,[5] a Proposed Hybrid Model (PHM) has been introduced, combining the strengths of the Exponential Smoothing Method (ESM), Autoregressive Integrated Moving Average (ARIMA), and Backpropagation Neural Network (BPNN) models. Through empirical testing on the Shenzhen Integrated Index and Dow Jones Industrial Average (DJIA), the hybrid model has demonstrated superior performance with a directional accuracy of 70.16%, outperforming individual sub-models and conventional forecasting methods. In the pursuit of anticipating future stock prices, the model introduced in [6] evaluates its accuracy by benchmarking against comparable models. This approach involves combining the Adaline Neural Network (ANN) with modified Particle Swarm Optimization (PSO). Additionally, [7] presents a hybrid intelligent model utilizing an Adaptive Network-based Fuzzy Inference System (ANFIS) in conjunction with quantum-behaved particle swarm optimization, showcasing the integration of diverse methodologies for enhanced forecasting capabilities in the stock market.The findings from [8] highlight the efficacy of employing deep learning techniques, particularly LSTM models, in combination with hybrid multilingual sentiment data for stock market forecasting. The study suggests that this approach, incorporating translated data from non-native English-speaking countries into English, outperforms other models utilizing different data types.[9] introduced a hybrid stock prediction model, encompassing a noise-filtering approach, distinctive features, and machine learning-based predictions. The noise-filtering technique is employed to refine historical stock price data by removing the cyclic component of the time series. The newly extracted features play a pivotal role in predicting future stock prices. The study explores both traditional and deep machine learning techniques for stock price prediction, employing a machine learning-centric approach.The model presented in [10] features an LSTM-GRU network coupled with 25 distinct features. Performance indicators highlight that the proposed model outperforms competing models, showcasing its enhanced accuracy in stock market forecasting.The hybrid approach introduced in [11] for predicting financial time series integrates LSTM, Polynomial Regression (PR), and Chaos Theory. The model assesses the presence of chaos, models it, and utilizes LSTM for initial forecasts. The sequence of errors obtained from LSTM predictions is then employed by PR for error forecasting. The ultimate hybrid model generates forecasts by combining error forecasts with the original model projections. In [12], the researcher makes a dual contribution. Firstly, they introduce a distinctive and robust deep convolutional GAN architecture, serving in both generative and discriminative capacities for stock price forecasting. Secondly, the researcher recommends enhancing the generator's loss function by incorporating additional terms to improve prediction accuracy.

Sentiment analysis, a method for examining research literature to identify underlying attitudes and emotions, plays a crucial role in predicting the stock market. This technique involves evaluating sentences to discern whether they convey optimistic or pessimistic sentiments, assigning

a value of 1 to positive news and 0 to negative news to gauge overall sentiment. The integration of deep learning models, leveraging existing data to identify patterns, and lexicon-based approaches, examining word prevalence to infer emotions, has significantly contributed to stock market forecasts. Additionally, text mining and natural language processing have bridged the gap between market fluctuations and news sentiments. The ongoing development and exploration of these techniques aim to enhance stock price forecasting, as highlighted in previous research (Table 1).

Collectively, these studies showcase the diversity of approaches in the quest for accurate stock price prediction, underscoring the significance of sentiment analysis, machine learning algorithms, and multi-source data integration in forecasting market dynamics.

3. METHODOLOGY

This comprehensive methodology integrates web scraping, sentiment analysis, and machine learning techniques, showcasing the role of Python, BeautifulSoup, and VADER in deriving actionable insights for stock price prediction based on financial news sentiments.

3.1. SENTIMENT SCORE FROM NEWS HEADLINES

This section details the methodology for computing real-time sentiment scores for news headlines related to a specific stock.

3.1.1. DATA COLLECTION

In our research endeavors, we relied on the FinViz platform, a highly regarded stock screener and financial visualization tool. As a browser-based resource, FinViz offered us valuable insights into financial news on a diverse range of actively traded stocks. Renowned for its user-friendly interface and powerful screening capabilities, FinViz swiftly provided a macro view of the market landscape, facilitating efficient analysis and informed decision-making. This platform's robust features make it a popular choice among investors and traders, underscoring its significance in our research for gaining comprehensive financial information and insights.

3.1.2. WEB SCRAPING

Python, renowned as the second-most popular programming language in 2020, following closely behind C, has evolved since its introduction in 1991, with a pivotal upgrade in 2008 marked by the release of Python 3.0. In our research, Python played a central role, offering an extensive standard library equipped with invaluable tools. Notably, we employed BeautifulSoup, a widely acclaimed package for scraping and parsing website data.

After selecting FinViz as our data source, we executed a Python script utilizing BeautifulSoup to extract article headlines. This web scraper, crafted with tools from the BeautifulSoup library, showcases the language's versatility. BeautifulSoup, celebrated for its simplicity and Pythonic approach, efficiently constructs navigable and searchable parse trees. A key advantage lies in its ability to interpret any input, translating parsed data into the widely used UTF-8 format on the internet, enhancing the efficacy of our data collection process.

3.1.3. PRE-PROCESSING, CLEANING AND LABELING FOR PREDICTIVE ANALYSIS

In our data preparation phase, we implemented a rigorous pre-processing and cleaning regimen for the financial news extracted from FinViz, ensuring the data's quality and reliability. Leveraging the Scikit-learn library for data analysis and the NLTK library for language processing, each headline was meticulously classified as positive, negative, or neutral. The subsequent step involved extracting news headlines for stocks under consideration using web scraping techniques. Labels were then assigned to predict the directional movement of stock values for the following day. To introduce dynamism to our predictive model, we adopted a moving average approach for label calculation. Addressing instances where selected stocks lacked coverage in major finance news publications on specific dates, sentiment scores were conservatively recorded as 0, demonstrating our approach to managing missing data. The data preprocessing stage involved essential steps such as tokenization, eliminating extraneous elements like numbers and punctuation, implementing stop words to filter out common terms, and applying stemming to reduce word redundancy. This meticulous preparation was crucial, considering that text data is inherently unstructured, necessitating thoughtful handling for effective input into the classifier.

3.1.4. SENTIMENT ANALYSIS TOOL

In the pursuit of sentiment analysis for financial news related to stock market dynamics, we employed the VADER model (Valence Aware Dictionary for Sentiment Reasoning), a rule-based system embedded within the Natural Language Toolkit (NLTK) package in Python. VADER, designed by Hutto and Gilbert, surpasses traditional sentiment lexicons like LIWC, offering improved generalization across diverse domains and heightened responsiveness to sentiment expressions in social media environments. Empirically validated against eleven prominent sentiment analysis tools, VADER demonstrated robust performance. However, it's crucial to note that on various dates, selected stocks lacked coverage in major finance news publications utilized by FinViz for data collection, resulting in a sentiment score of 0 for those instances. Subsequently, in our modeling phase, linear autoregressions were executed, both with and without the exogenous factor of sentiment scores, showcasing their impact on stock price changes. VADER is used to calculate the mean sentiment scores of the news headlines.

After the sentiment score was obtained, in the modeling phase two linear autoregressions were performed; one without an exogenous factor and one with one (the

exogenous factor being the sentiment score) as Equation (1) and Equation (2), respectively, show:

$$y=\beta_1+\beta_2 y_{t-1}+\beta_3 y_{t-2}+\beta_4 y_{t-3}+\beta_5 y_{t-4}+\varepsilon. \quad (1)$$
$$y=\beta_1+\beta_2 y_{t-1}+\beta_3 y_{t-2}+\beta_4 y_{t-3}+\beta_5 y_{t-4}+\beta_6 x+\beta_7 x_{t-1}+\beta_8 x_{t-2}+\varepsilon \quad (2)$$

where $\beta 1$ is the intercept, $\beta 2$is the slope, $y_{t-1/2/3/4}$ are the stock-opening-price change variables in the previous days, x and x_{t-1} are the sentiment score exogenous variables in the present and previous days and ε is the error term. y is the stock price change in time t.

Additionally, linear, quadratic, and cubic regressions were employed to analyze the nuanced relationship between sentiment scores and stock-opening-price changes.
The relation for the linear, quadratic and cubic autoregressions are presented in Equation (3), Equation (4) and Equation (5), respectively:

$$y=\beta_1+\beta_2 x+\varepsilon. \quad (3)$$
$$y=\beta_1+\beta_2 x+\beta_3 x^2+\varepsilon. \quad (4)$$
$$y=\beta_1+\beta_2 x+\beta_3 x^2+\beta_4 x^3+\varepsilon \quad (5)$$

where y is the stock opening price change, β_1, β_2, β_3 and β_4 are the unknown parameters, x is the sentiment score variable and ε is the error term. Y is the stock price change in time t.

After running the linear, quadratic and cubic regressions, we ran a nonlinear autoregression with an exogenous factor (NARX) as can be seen in Equation (6):

$$y=\beta_1 y_{t-1}+\beta_2 y_{t-2}+\beta_3 y_{t-3}+\beta_4 y_{t-4}+\beta_5 x+\beta_6 x^2+\varepsilon. \quad (6)$$

where y is the opening price change in time t, x is the sentiment score and $y_{t-1/2/3/4}$ are the opening price changes in time t−1, t−2, t−3 and t−4.

To enhance our regression analysis, we further employed nonlinear autoregressions with an exogenous factor (NARX), utilizing ordinary least squares in SPSS for the regression computations. The use of aggregated data, with sentiment scores weighted by market capitalization, provided insights into the performance of different models, ultimately leading to the adoption of equal weight for all regressions in our study.

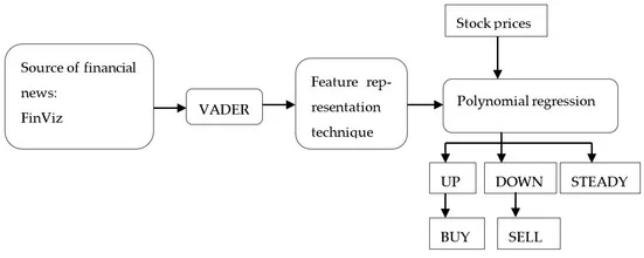

FIGURE 1. Research Design

3.2. STOCK PRICE PREDICTION USING TWEETS

This section outlines the methodology employed for predicting stock prices using historical tweets and stock data.

3.2.1. DATA COLLECTION

In our research, we utilize a comprehensive dataset that integrates historical tweets (September 2021 to September 2022, ~81k rows) from Kaggle and stock-related information, including opening/closing prices and trading volumes, sourced from Yahoo Finance. This diverse dataset combines qualitative insights from social media (tweets) with quantitative measures such as stock prices and trading volumes, providing a holistic foundation for our analysis. The tweet dataset captures user-generated content related to 50 curated stocks, featuring attributes like "Source," "Tweet Text," and "Date." The historical price data from Yahoo Finance is presented in CSV format, contributing to a multifaceted and robust dataset for our research.

In the predictive model, historical stock data undergoes technical analysis, while social media data and financial news undergo sentiment analysis using NLP. The combined insights from sentiment analysis and technical analysis are utilized in a unified model for stock market prediction, as illustrated in Fig. 1. This holistic approach aims to leverage both textual and numerical data for a more nuanced understanding of market dynamics and improved predictive accuracy.

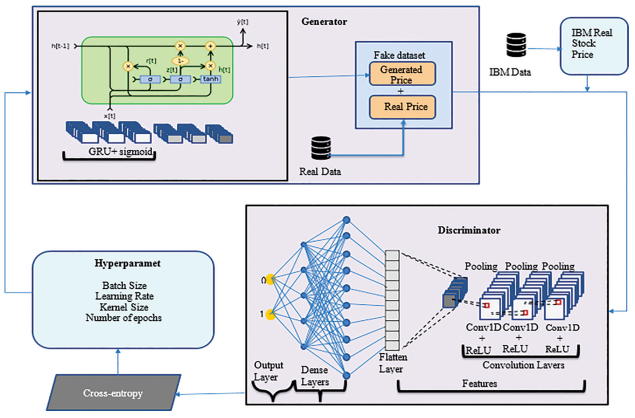

FIGURE 2. Proposed a details model of IBM stock market prediction

3.2.2. TECHNICAL INDICATORS

Our experimental setup incorporates a selection of key technical indicators aimed at capturing diverse aspects of stock price behavior.

- **Moving Averages (MA):** Utilizing both 7-day and 21-day simple moving averages (SMA) to represent historical price trends.

- **Bollinger Bands (BB):** Employing three lines - the SMA, upper (+20 standard deviations), and lower (-20 standard deviations) - to gauge price volatility.

- **Moving Average Convergence Divergence (MACD):** Offering insights into overbought or oversold market conditions through the calculation of 26-day and 12-day exponential moving averages.

- **LogMomentum:** Capturing the overall stock trend by aggregating the logarithm of daily closing prices.

- **Relative Strength Index (RSI):** Generating signals based on thresholds, identifying overbought conditions (RSI > 70%) and oversold conditions (RSI < 30%).

3.2.3. GANs WORK

Generative Adversarial Networks (GANs) constitute a revolutionary paradigm in artificial intelligence, featuring two neural networks engaged in a competitive duet. The Discriminator and Generator operate collaboratively, with the Generator striving to craft synthetic data indistinguishable from the training set. Concurrently, the Discriminator endeavors to accurately discern genuine from generated data, avoiding deception. This competitive dynamic proves particularly potent for intricate data types such as audio, video, or images. In each training cycle, the Generator refines its approach, amplifying the likelihood of fooling the Discriminator, while the latter enhances its acumen in distinguishing between real and synthetic samples. This iterative process, governed by a loss function, propels both networks toward greater sophistication. Mathematically, the model's competitiveness is expressed through the loss function, guiding the iterative refinement of parameters through backpropagation. GANs have demonstrated prowess in diverse applications, from image synthesis to stock price prediction, harnessing the power of adversarial learning to generate data distributions of remarkable realism. It can be mathematically described by the formula below:

$$V(D,G)=E_{x\sim Pdata(x)}[\log D(x)]+E_{z\sim Pz(z)}[\log(1-D(z))]$$

where G = Generator, D = Discriminator, Pdata(x) = distribution of real data, P(z) = distribution of generator, x = sample from Pdata(x), z = sample from P(z), D(x) = Discriminator network, G(z) = Generator network.

3.2.4. GAN GENERATOR

In the construction of our GAN model, the authors opted for the Gated Recurrent Unit (GRU) as the generator, attributing its selection to its inherent stability. The dataset under consideration encompasses stock price history, featuring 36 distinct features ranging from Open, Low, High, and Close to various market indices and indicators. To facilitate multi-step forward prediction, the study meticulously defines the input and output phases for the generator. The generator is designed to take input data parameters, including batch size, input step, and features, producing corresponding output data parameters, specifically batch size and output step. This configuration is underpinned by a three-layer GRU architecture with neuron counts of 1024, 512, and 256, supplemented by two layers of dense networks. The final dense layer's neuron count aligns with the desired output step for prediction. The term "multi-steps-ahead prediction" elucidates the task of forecasting a sequence of values within a time series. The study employs a multi-stage prediction approach, incrementally applying the predictive model and leveraging the estimated value of the current time step to project its value in the subsequent time step.

3.2.4. THE DISCRIMINATOR

Within our GAN model, the discriminator takes on the role of a convolutional neural network (CNN), entrusted with discerning the authenticity of input data. Its primary task is to distinguish between genuine and artificially generated data, whether received in its original form or newly created by the generator. The discriminator architecture is composed of six layers, incorporating three 1D Convolution layers with 32, 64, and 128 neurons, three Dense layers featuring 220, 220, and 1 neuron respectively at the final layer, and three additional Dense layers. Leaky Rectified Linear Unit activation functions are applied across all layers, except for the output layer, where the Sigmoid activation function is employed for GAN, and the linear activation function for Wasserstein GAN-Gradient Penalty (WGAN-GP). The output layer, using Rectified Linear activation function (ReLU), produces a single-scalar output that signifies either 0 or 1, corresponding to true and false values, respectively, in the context of GAN.

4. RESULTS

4.1. SENTIMENT SCORE FROM NEWS

In the sentiment score calculation process for news headlines, we employ the VADER sentiment analysis tool to assess the positive, negative, and neutral tones conveyed in each headline. VADER provides scores indicating the intensity of each sentiment category. Subsequently, based on these sentiment scores, we generate visual representations in the form of hourly and daily graphs for the respective stock. These graphs offer a dynamic visualization of how the sentiment in news headlines evolves over time and its potential correlation with the stock's performance. By observing these graphical trends, we can draw more nuanced and detailed conclusions about the interplay between sentiment analysis and stock movements, enhancing our understanding of the market dynamics.

FIGURE 3. **Hourly Sentiment Score Graph of TSLA Stock.**

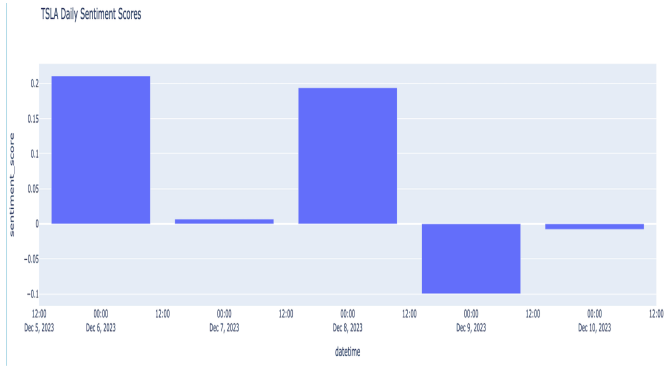

FIGURE 4. **Daily Sentiment Score Graph of TSLA Stock.**

4.2. STOCK PRICE PREDICTION USING TWEETS

In the stock price prediction methodology utilizing historical tweets and stock data, we adopt a sophisticated GAN model architecture. The generator comprises 5 LSTM blocks, while the discriminator is equipped with 5 convolutional layers and 3 dense layers, employing a sigmoid activation function. Following the model's construction, we conduct comprehensive computations, including the generation of technical indicators for the stock. Subsequently, during the training and testing phases, we visualize the stock based on its price. As part of our analysis, we plot key technical indicators, generator loss, and discriminator loss. Finally, the culmination of our efforts is reflected in graphical representations juxtaposing the real stock price against the predicted price. These visualizations offer insights into the model's performance and its ability to predict stock movements based on historical data and sentiments derived from tweets.

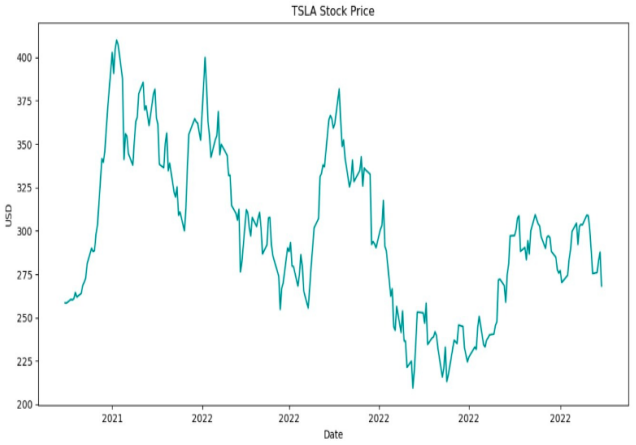

FIGURE 5. **Stock Price of TSLA Stock.**

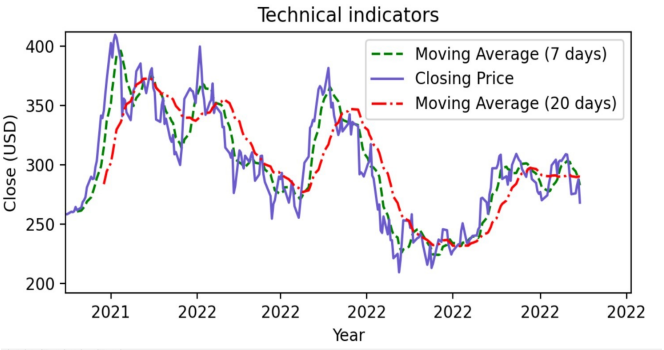

FIGURE 6. **Technical Indicators on the TSLA Stock.**

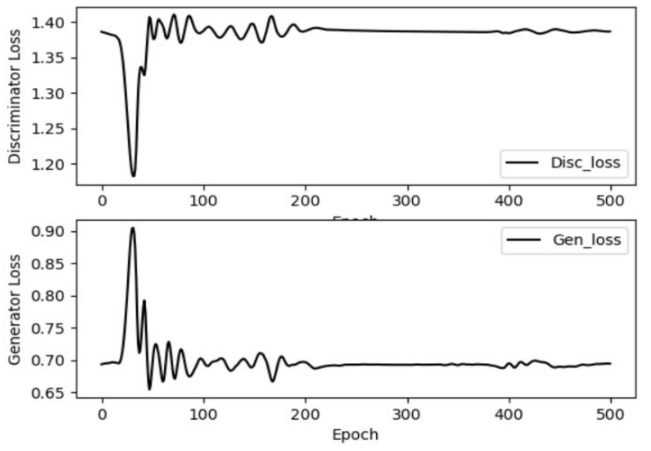

FIGURE 7. **Result of discriminator and generator's loss in GAN basic**

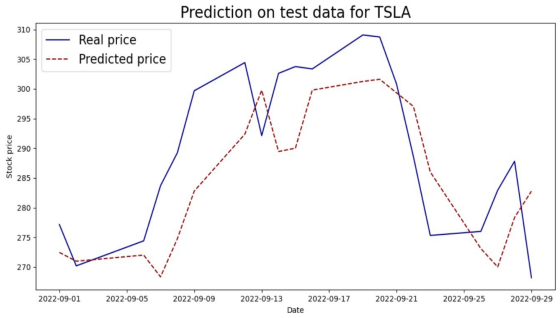

FIGURE 8. Output of testing model of GAN basic

5. CONCLUSION

In this study, we harnessed the power of the VADER model to generate sentiment scores daily, derived from financial news headlines provided by FinViz for selected stocks. Our research delves into the intricate interplay of artificial intelligence techniques in the stock market, examining pivotal factors influencing stock price fluctuations across 36 variables, encompassing both quantitative and qualitative aspects. These variables span daily market value, currency rates, competitors' performance, and global market indicators, while qualitative inputs are sourced from Twitter and news sites discussing the target stock.

Our exploration builds upon a thorough review of previous financial stock studies, forecasting techniques, and contemporary artificial intelligence methodologies. The study introduces a Generative Adversarial Network (GAN) featuring Convolutional Neural Networks (CNN) as the discriminator and Gated Recurrent Unit (GRU) as the generator. Notably, the GAN's output serves as hyperparameters for further model tuning. Experimental findings underscore the GAN model's potential to enhance time series forecasting for the stock market, with the GRU model in the generator exhibiting robust performance in sequence memorization and data generation.

Investors can leverage these findings for strategic decision-making regarding stock transactions, such as buying, holding, or selling. The research not only contributes valuable insights into the effectiveness of different approaches but also serves as a reference for future work in the field. The study stands out for its comprehensive exploration of diverse techniques, incorporating statistical, machine learning (supervised and unsupervised), and hybrid models. It innovatively integrates quantitative and qualitative data, introducing GANs for processing time series data in stock market prediction. The model seamlessly combines two robust networks, RNN-GRU and DL-CNN, employing multiple cost functions to enhance predictive capabilities.

Future research directions may involve exploring additional data sources, experimenting with diverse processing methods, and employing advanced evaluation techniques. Researchers could consider incorporating both numerical and nominal data for a holistic stock market analysis, merging quantitative and qualitative insights. Importantly, future studies can aim for generalizability beyond specific stocks, offering universally applicable predictive models for various stocks in the market.

REFERENCES

[1] P. Falinouss, “Stock trend prediction using news articles: a text mining approach,” Dissertation, 2007, pp. 83-84.

[2] Y. Shynkevich, T. M. McGinnity, S. Coleman and A. Belatreche, "Predicting Stock Price Movements Based on Different Categories of News Articles," 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, 2015, pp. 703-710.

[3] P. Kaviani and S. Dhotre, “Short Survey on Naive Bayes Algorithm,” International Journal of Advance Research in Computer Science and Management, vol. 4, 2017.

[4] A. Atkins, M. Niranjan and E. Gerding, “Financial news predicts stock market volatility better than close price,” The Journal of Finance and Data Science, 2nd ed., vol. 4, 2018, pp. 120-137.

[5] J. J. Wang, J. Z. Wang, Z. G. Zhang and S. P. Guo, “Stock index forecasting based on a hybrid model,” Omega, vol. 40, no. 6, pp. 758–766, 2012.

[6] M. Sedighi, H. Jahangirnia, M. Gharakhani and S. Farahani Fard, “A novel hybrid model for stock price forecasting based on metaheuristics and support vector machine,” Data, vol. 4, no. 2, pp. 75, 2019.

[7] A. Bagheri, H. M. Peyhani and M. Akbari, “Financial forecasting using ANFIS networks with quantum-behaved particle swarm optimization,” Expert Systems with Applications, vol. 41, no. 14, pp. 6235–6250, 2014.

[8] Y. L. Lin, C. J. Lai and P. F. Pai, “Using deep learning techniques in forecasting stock markets by hybrid data with multilingual sentiment analysis,” Electronics, vol. 11, no. 21, pp. 3513, 2022.

[9] Q. M. Ilyas, K. Iqbal, S. Ijaz, A. Mehmood and S. Bhatia, “A hybrid model to predict stock closing price using novel features and a fully modified hodrick–Prescott filter,” Electronics, vol. 11, no. 21, pp. 3588, 2022.

[10] G. R. Patra and M. N. Mohanty, “An LSTM-GRU based hybrid framework for secured stock price prediction,” Journal of Statistics and Management Systems, vol. 25, no. 6, pp. 1491–1499, 2022.

[11] M. Durairaj and K. M. BH, “Statistical evaluation and prediction of financial time series using hybrid regression prediction models,” International Journal of Intelligent Systems and Applications in Engineering, vol. 9, no. 4, pp. 245–255, 2021.

[12] A. Staffini, “Stock price forecasting by a deep convolutional generative adversarial network,” Frontiers in Artificial Intelligence, vol. 5, pp. 1–16, 2022.

WORKLOAD

Weeks 1: Data Collection and Preprocessing

Ananya: Collected and preprocessed historical stock price data.
Kunal: Collected and preprocessed news headlines and tweets.

Weeks 2-3: Sentiment Analysis Implementation and Optimization

Ananya: Developed and optimized the sentiment analysis algorithm for news headlines.
Kunal: Develop and optimize the sentiment analysis model for tweets.

Weeks 4: Feature Engineering

Ananya: Engineered features based on sentiment scores from news headlines and explored additional relevant features.
Kunal: Engineered features based on sentiment scores from tweets and explored additional relevant features.

Weeks 5: Model Development for User Option 2 (Stock Price Prediction)

Ananya and Kunal: Collaborated on building the predictive model using historical tweets and stock price data, experimented with various algorithms, optimized hyperparameters, and validated the model's performance.

Week 6: Integration and Testing

Ananya and Kunal: Integrated sentiment analysis models with the overall stock price prediction model and conduct thorough testing.

Week 7: Documentation, Presentation, and Reporting

Ananya: Documented the methodology for sentiment analysis on news headlines and prepared relevant documentation.
Kunal: Documented the approach and challenges in sentiment analysis of tweets and prepared relevant documentation.

User Option 1: Get Sentiment Score of the Stock from Latest News Headlines

Ananya: Implemented the functionality to provide sentiment scores based on the latest news headlines.
Kunal: Assisted in integrating the sentiment analysis for news headlines into the user interface.

User Option 2: Predict Stock Price Using Historical Tweets and Stock Price Data

Kunal: Lead the development of the functionality to predict stock prices using historical tweets and stock price data.
Ananya: Assist in integrating the sentiment analysis for tweets into the user interface.

EXECUTION OF CODE

1. Download the historical tweet dataset from stock_tweets.csv and stock_yfinance_data.csv.
2. Libraries to be imported/ Modules to be installed: tensorflow, flask, beautiful soup(bs4), matplotlib, plotly and numpy, pandas, urllib3, nltk and gunicron.
3. Download the files from the provided github link.
4. Change the path of the file in the code (app.py line 177, line 217.
5. The link for the executable code is given below: https://github.com/ananya-jaikumar/DM-Project