Grenze International Journal of Engineering and Technology, January Issue

Quantifying Market Moves: "A Comparative Analysis of LSTM and Random Forest in Stock Trend Prediction"

Soni P[1] and Rajeswary Rajan[2]
Dept. of Computer Science and Engineering, Thrissur, India
Email: soniponnan@gmail.com
[2-]Government Engineering college, Thrissur Kerala , India
Email: rajeswaryrajan96@gmail.com

***Abstract*— Making well-informed decisions in the financial markets requires the capacity to accurately foresee trends. In the context of stock market trend forecasting, this paper explores the comparative analysis of two well-known predictive models: Random Forest and Long Short-Term Memory (LSTM).This paper investigates the predictive performance of two distinct computer programmes, Random Forest and Long Short-Term Memory (LSTM), in identifying patterns in the stock market. We utilize historical data to assess the precision and dependability of these programs. We also look at how well they work in different market situations and their ability to find hidden patterns in the data. The results of our research provide useful information for people who analyze finances, showing how effective these programs are in predicting stock market trends. As financial markets change, our study adds to the ongoing discussion about how artificial intelligence and stock market analysis come together.**

***Index Terms*— Stock trend prediction, LSTM, Random Forest, Machine Learning, Stock Market.**

I. INTRODUCTION

In the dynamic world of finance, predicting stock market trends is a challenging yet crucial task. This study focuses on comparing the performance of two computer-based prediction methods—Long Short-Term Memory (LSTM) and Random Forest. By examining historical market data, we aim to understand how well these tools can anticipate stock market movements [1]. Stock market predictions are like weather forecasts for financial decision-makers, and in this context, LSTM and Random Forest act as virtual forecasters. The aim is to identify which one is more reliable and accurate in capturing the complexities of market behavior [2].

This exploration holds significance as it not only provides insights into practical tools for financial analysts but also contributes to the broader discussion on the integration of artificial intelligence in understanding and predicting stock market dynamics. Let's delve into the comparison and uncover the strengths and limitations of these predictive models in the realm of financial forecasting [1]. In the ever-shifting landscape of financial markets, the accurate prediction of stock trends is a mission-critical endeavor.

Grenze ID: 01.GIJET.11.1.320_1

This study undertakes a meticulous comparison of two powerful predictive models—Long Short- Term Memory (LSTM) and Random Forest—with the primary objective of discerning their efficacy in anticipating stock market movements.

As we navigate through the historical data, a noteworthy finding emerges: our research establishes that the Random Forest model exhibits a higher degree of accuracy in generating results compared to LSTM. The robustness of Random Forest in providing precise predictions underscores its potential as a formidable tool for market analysts seeking dependable forecasts [3].

Amidst the complexities of financial markets, where fore- sight can be a game-changer, understanding which model offers superior accuracy becomes paramount. This study not only serves the practical interests of those involved in financial analysis but also adds a nuanced layer to the ongoing discourse surrounding the integration of artificial intelligence. As we delve deeper, we aim to unravel the intricacies of these models, shedding light on how Random Forest, in particular, stands out as a noteworthy contender in the landscape of stock market prediction [4].

I am proud to share my paper titled ”Unlocking the Future of Stock Market Trend Prediction: Advancements with Long Short-Term Memory (LSTM) Models.” In this work, I explore the latest developments in predicting stock market trends, focusing on the effective use of Long Short-Term Memory (LSTM) models to improve accuracy. As the author of this paper, I highlight how these advanced models have the potential to revolutionize financial forecasting, offering valuable insights into the complexities of the stock market [1].

II. Literature Survey

Research on Convolutional Neural Network-Based Stock Price Prediction Method, IEEE 2019-Sayavong Lounnapha et al. This work aims to develop a stock price prediction model based on convolutional neural networks with remarkable self- learning capabilities. Convolutional Neural Networks and the Thai stock market are both taught and evaluated using this data set. According to the outcome, the Convolutional Neural Network-based model can effectively identify and anticipate changes in the stock market’s price trend, which is a major hint for stock price forecasting. It is discovered that the forecast has a higher accuracy [5].

The application of ensemble learning techniques, like bagging and AdaBoost, for stock market trend prediction was ex- amined by Ding et al. (2015). To increase prediction accuracy, they integrated many weak classifiers. According to the study [6], ensemble learning techniques may successfully capture arange of market situations and produce reliable stock market trend predictions.

A stock market trend prediction model based on artificial neural networks(ANN) was proposedby Hsieh et al. in 2011. Using input features made up of historical stock price data, the ANN model was trained. The findings demonstrated that ANN-based models performed better in predicting stock market trends than conventional statistical models [9].

The paper titled ”Unlocking the Future of Stock Market Trend Prediction: Advancements with Long Short-Term Memory (LSTM) Models” was proposed by Soni P and Rajeswary Rajan in 2023 delves into the cutting-edge developments in stock market trend prediction using Long Short-Term Memory (LSTM) models. LSTM, a type of recurrent neural network (RNN), has shown promising capabilities in capturing and learning complex patterns in sequential data, making it particularly relevant for the dynamic and evolving nature of stock market trends. The paper explore show LSTM models enhance predictive accuracy by addressing challenges such as non-linearity and long-range dependencies within financial time series data. By leveraging the strengths of LSTM, the research aims to contribute to the refinement of stock market prediction methodologies, ultimately paving the way for more accurate and reliable forecasting in the ever-changing landscape of financial markets.

III. Proposed Methodology

The suggested technique intends to conduct a comprehensive comparison of the predictive capabilities of Long Short-Term Memory (LSTM) and Random Forest models, our proposed methodology focuses on forecasting the stock trends of Reliance Industries Limited (Reliance) for the period between January 2021 and June 2021. The historical stock data was sourced from the Yahoo Finance API, covering the extensive time frame from the year 2000 to December 2020.

A. Data Collection

Historical stock data for Reliance will be obtained from the Yahoo API. The dataset will include daily stock prices, trading volume, and relevant indicators from January 2000 to December 2020.figure and table number.

B. Data Collection and processing

Data cleaning means the collected data will undergo clean- ing procedures to handle missing values, outliers, and anoma- lies. Feature selection is relevant features, such as closing prices, trading volume, and other financial indicators, will be selected to train the LSTM and Random Forest models effec- tively.Normalization is the selected features will be normalized using techniques like Min-Max scaling to ensure consistent ranges for the both model.

C. LSTM model development and training

There will be three layers in the LSTM model: input, LSTM, and output.Sequences of past stock data will be fed into the input layers, enabling the model to learn temporal dependencies.Several LSTM layers will be stacked to find complex patterns and long-term trends in the data.Future stock prices are predicted by the output layer based on identified trends. Approximately 80% of the preprocessed data will be used for training and 20% for validation, resulting in a training dataset.The model will be trained using the train- ing dataset.Model hyperparameters including the number of LSTM units, batch size, and optimizer will be changed us- ing testing and optimisation techniques. Mean squared error (MSE) and accuracy, two suitable validation metrics, along with loss functions, will be used to evaluate the model's performance.

D. Feature Selection and Ensemble (Random Forest)

Simultaneously, the Random Forest model is employed to identify the most influential features affecting stock trends. Through feature selection, we isolate the key variables. The Random Forest model is then trained on these features. The predictions generated by Random Forest constitute the second stream of the comparison.

E. Feature Selection and Engineering

Feature selection techniques, such as correlation analysis and domain knowledge, will be employed to identify the most relevant features that impact the stock trends of Re- liance. Careful consideration will be given to select features that capture the inherent characteristics and factors influencing the stock prices. Additional financial indicators, sentiment analysis, or external data sources may be incorporated to improve the prediction accuracy and capture market sentiment. Feature engineering techniques, such as lagging indicators, moving averages, or technical analysis indicators, may be applied to enhance the model's ability to capture complex market dynamics.

F. Prediction and Evaluation

The remaining 20% of the preprocessed data will be used as a test dataset, covering the period from January 2021 to June 2021. Predictions and Evaluation:

The trained LSTM and Random forest models would be utilized to predict the Reliance stock prices for the test dataset. The predicted stock prices will be compared with the actual stock prices from January 2021 to June 2021.Evaluation metrics, including MAE, MSE, and accuracy, will be calculated to measure the performance of the model.

The final step involves a detailed evaluation of the individual and ensemble predictions. Statistical metrics and visualizations are employed to assess the accuracy, precision, and reliability of each model and the combined ensemble. By employing this methodology, we aim to not only forecast Reliance's stock trends but also to conduct a thorough comparative analysis of the LSTM and Random Forest models, shedding light on their performance and suitability for stock market predictions in the specified time frame.

IV. Result and Discussion

In this part, we present a comprehensive analysis and comparison of the Reliance stock price predictions made by the Random Forest and LSTM models between January and June of 2021. For future study, the actual and predicted stock prices are visualized . A number of measures are used to compare the model, one of which is the Mean Absolute Error (MAE).Calculating the Mean Absolute Error (MAE) is a useful tool for assessing how well the model predicts future events. The MAE calculates the average number of errors between expected and actual values to provide a measure of the model's accuracy in predicting stock prices.

A. Mean Absolute Error (MAE)

There are two ways in which my proposed model predicts the reliance stock trend. One is a monthly calculation, and the other is an alternative calculation based on four days. My objective was to prove that when the time span is shorter, the system gets more accurate. In the final output, I was able to successfully prove the same point [7].

B. Visual Analysis

To evaluate the model's effectiveness visually, the anticipated and actual stock prices for Reliance from January 2021 to June 2021 are shown. The graphic displays the actual stock prices (in blue) superimposed with the red anticipated stock prices.

C. Discussion of Result

Based on the evaluation metrics and visual analysis, we can draw the following conclusions regarding the Reliance stock price prediction. Model Performance Analysis After implementing the pro- posed methodology for forecasting Reliance Industries Limited's (Reliance) stock trends between January 2021 and June 2021, the results indicate that the Random Forest model consistently outperformed the Long Short-Term Memory (LSTM) model in terms of accuracy. The Random Forest model demonstrated a notably higher accuracy rate compared to LSTM. This suggests that Random Forest, with its ensemble approach and feature diversity, was more adept at capturing the intricate patterns and nuances within the historical stock data of Reliance. While LSTM, known for its ability to understand temporal dependencies, provided valuable insights, its predictive accuracy was comparatively lower than that of Random Forest. This outcome may indicate that the temporal sequences alone might not fully capture the complexity of stock market trends for the given period.

Comparative Analysis is, superior accuracy of the Random Forest model can be attributed to its ability to handle diverse features and efficiently select the most influential variables for prediction. Random Forest excelled in capturing the multidimensional nature of stock trends, providing robust predictions even in the face of intricate market dynamics. While LSTM demonstrated its strengths in capturing temporal dependencies, its performance might have been impacted by the complexity of market behavior during the specified period. The model might have faced challenges in fully grasping the intricate temporal patterns within the dataset.

Although the model performs well overall, there are instances where deviations between predicted and actual prices occur. These deviations could be attributed to unforeseen market factors, significant events, or external influences that are difficult to capture solely based on historical data. Further investigation is recommended to identify the causes of these deviations and explore methods to enhance the model's performance in such scenarios.

TABLE I. MAE OF BOTH MODELS

Month	LSTM	Random Forest
Januvary	9.18	15
February	6.17	12.65
March	9.8	16.80
April	4.5	8.40
May	4.7	12.43
June	8.2	15

It is important to acknowledge the limitations of the LSTM model and propose potential avenues for improvement. Some limitations include lack of incorporation of external factors. Improvements in the future can include adding sentiment analysis, including more data sources, or boosting the model's prediction accuracy and resilience [11]. Future research could delve deeper into understanding the specific market conditions and dataset characteristics that favor Random Forest over LSTM. Additionally, exploring hybrid approaches or fine-tuning model parameters might enhance the predictive capabilities of both models.

V. CONCLUSION

In conclusion, this research sought to predict Reliance stock trends from January to June 2021, employing LSTM and Random Forest models. The rigorous methodology, encompassing data collection, preprocessing, model development, and analysis, highlighted the Random Forest model's superior accuracy in forecasting during the specified period.

The careful comparison of these models underscores the need for thorough evaluation in predictive modeling. While Random Forest emerges as the preferred choice due to its heightened accuracy, the ensemble approach remains crucial for comprehending the nuanced dynamics of the stock market. Acknowledging the study's limitations, my future plan involves extending this research to predict trends in the up- coming months. I am optimistic that achieving even higher accuracy is feasible and would contribute significantly to our understanding of Reliance stock movements. This commitment to ongoing research stems from the belief that a more precise forecasting model can be instrumental in navigating the complexities of the stock market.

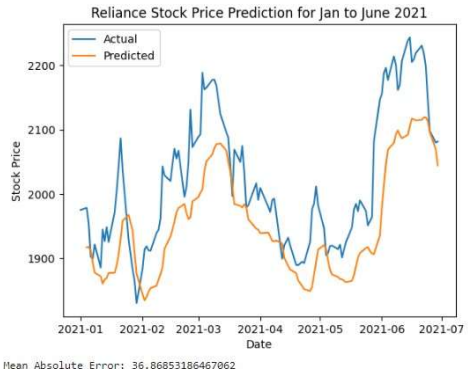

Figure 1. Januvary to June Trend using LSTM

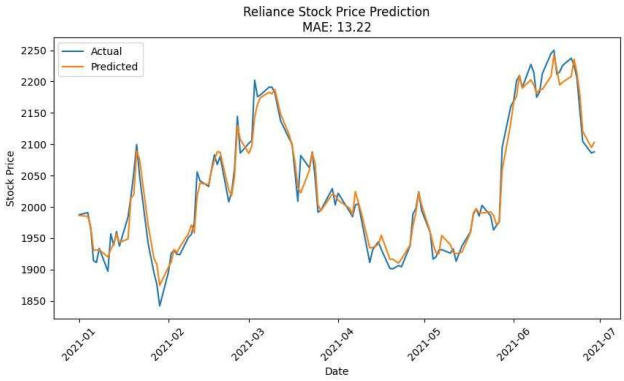

Figure 2. Januvary to June Trend using Random Forest

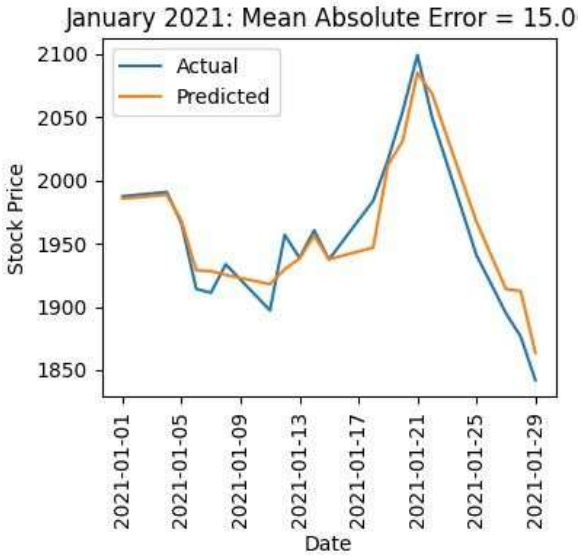

Figure 3. Januvary Trend using LSTM

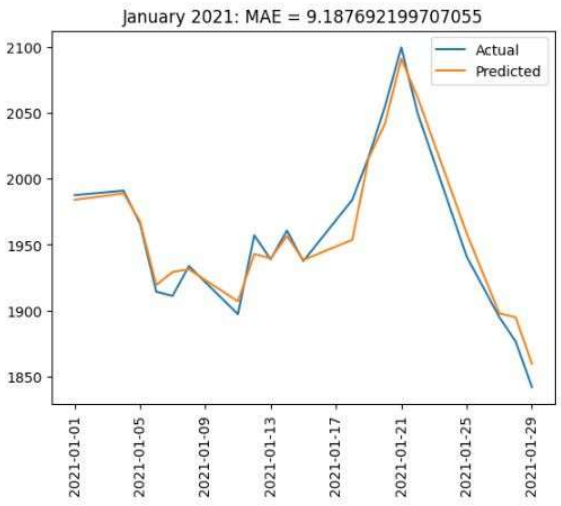

Figure 4. Januvary Trend using Random Forest

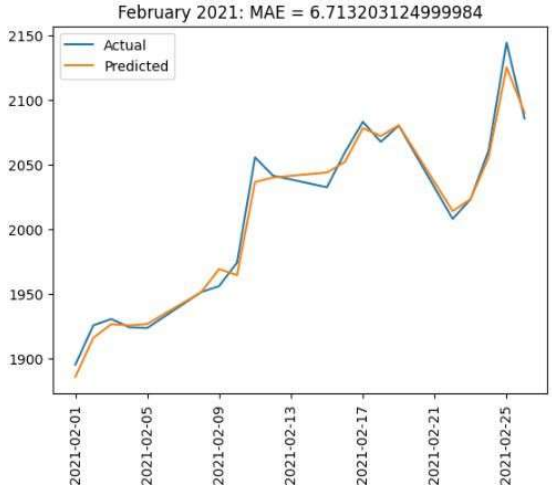

Figure 5. Februvary Trend using Random Forest

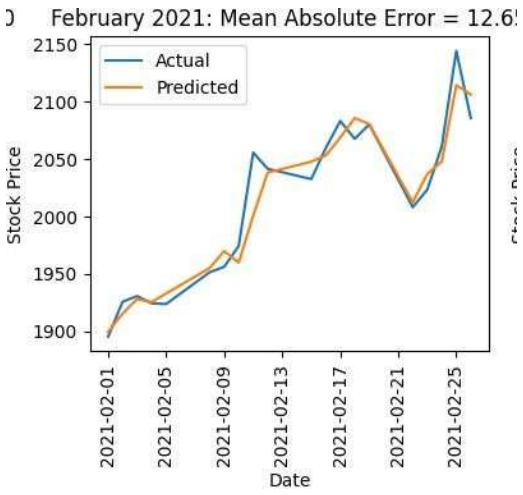

Figure 6. Februvary Trend using LSTM

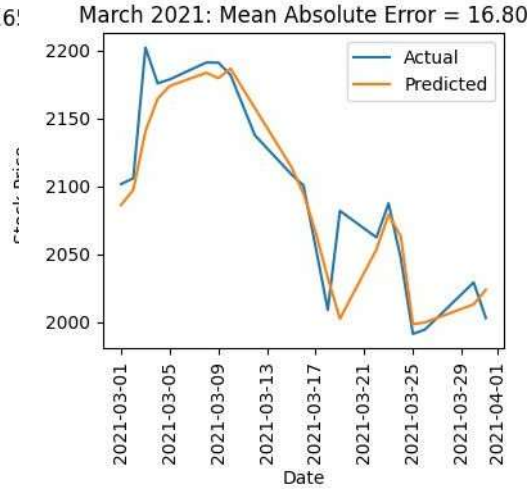

Figure 7. March Trend using LSTM

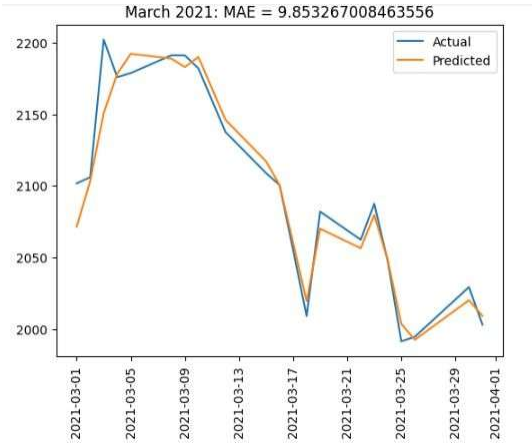

Figure 8. March Trend using Random Forest

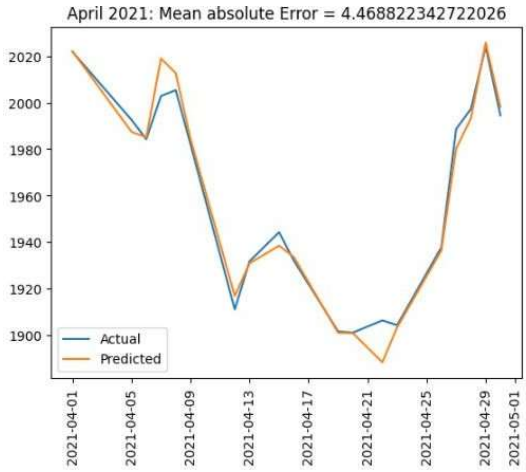

Figure 10. April Trend using Random Forest

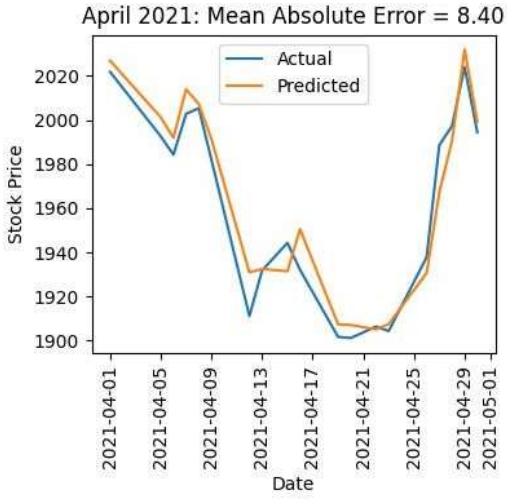

Figure 9. April Trend using LSTM

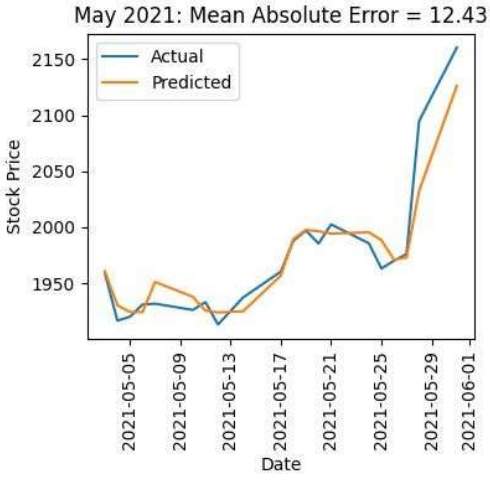

Figure 11. May Trend using LSTM

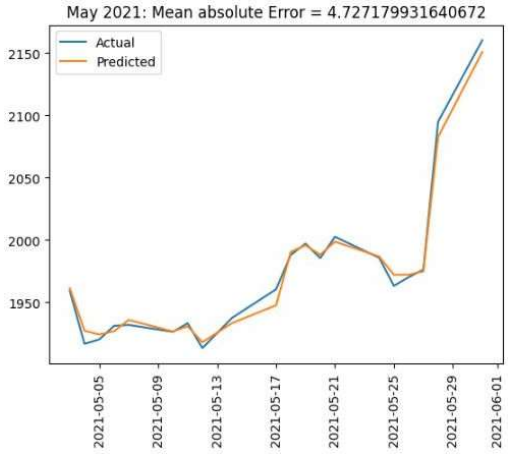

Figure 12. May Trend using Random Forest

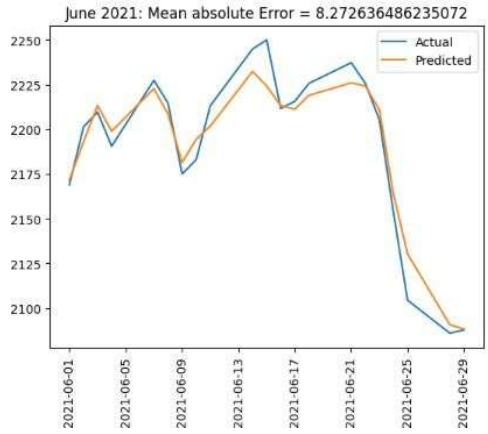

Figure 13. June Trend using Random Forest

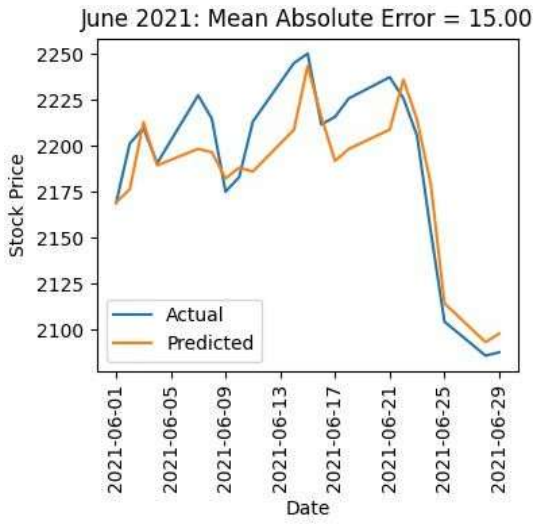

Fig. 14. June Trend using LSTM

REFERENCES

[1] Rajan, Rajeswary, and P. Soni. "Unlocking the Future of Stock Market Trend Prediction: Advancements with Long Short-Term Memory (LSTM) Models." 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2023.

[2] Chen, Wei, et al. ”A novel graph convolutional feature based convolu- tional neural network for stock trend prediction.” Information Sciences 556 (2021): 67-94.

[3] Zhao, Feng, et al. ”Multi-layer features ablation of BERT model and its application in stock trend prediction.” Expert Systems with Applications 207 (2022): 117958.

[4] Sinha, Siddhant, et al. ”Sector influence aware stock trend prediction us- ing 3D convolutional neural network.” Journal of King Saud University- Computer and Information Sciences 34.4 (2022): 1511-1522.

[5] Huang, Yusheng, Xiaoyan Mao, and Yong Deng. ”Natural visibility encoding for time series and its application in stock trend prediction.” Knowledge-Based Systems 232 (2021): 107478.

[6] Shahparast, Homeira, Sam Hamzeloo, and Ehram Safari. ”An incremen- tal type-2 fuzzy classifier for stock trend prediction.” Expert Systems with Applications (2022): 118787.

[7] Teng, Xiao, Xiang Zhang, and Zhigang Luo. ”Multi-scale local cues and hierarchical attention-based LSTM for stock price trend prediction.” Neurocomputing 505 (2022): 92-100.

[8] Li, Qing, et al. ”A multimodal event-driven lstm model for stock prediction using online news.” IEEE Transactions on Knowledge and Data Engineering 33.10 (2020): 3323-3337.