IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Implementation Of Machine Learning Algorithms Using Statistical Models For Predictive Analysis Of The Stock Market

Priyanka Gupta[1] Shradha Verma[2] Abhay Pratap Singh[3] Somya Rastogi[4]

School of Computer Science and Engineering, Lovely Professional University[123]

Abstract

Predicting stock prices accurately remains a challenging task, driving continuous research efforts in leveraging various machine learning, deep learning, and statistical analysis techniques. This paper presents a comparative study of four different models BiLSTM, GRU, CNN-LSTM, and ARIMA for stock price forecasting, using data from the S&P 500 index and five prominent technology companies: Meta, Apple, Google, Netflix, and Amazon. The study aims to identify the most effective model for maximising returns on investment in stock trading by forecasting future trends. Leveraging a dataset sourced from Yahoo Finance, comprising historical stock market data, we evaluate the performance of each model based on error metrics such as , Mean Squared Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The research findings highlight the strengths and weaknesses of each model in accurately predicting stock prices. Lower RMSE values indicate higher accuracy in predicting stock prices, guiding investors towards more informed decision-making. Through this comparative analysis, investors can identify the optimal model for stock price forecasting, thus enhancing their investment strategies and maximising returns.

Keywords: CNN, LSTM, GRU, ARIMA, MSE, RMSE, MAE, Forecasting, Stock Market

Introduction

The stock market serves as a dynamic platform where currencies, equities, and various financial instruments are bought and sold, enabling individuals to own shares in public companies. Traders seek to invest in stocks with potential for future growth while avoiding those anticipated to decline. Therefore, it is important to use statistical algorithms and machine learning and deep learning techniques for the correct and explanative prediction of the stock market. Stock market data presents challenges such as high levels of noise, non-stationarity, and non-linearity, which necessitate preprocessing techniques like differencing to achieve stationarity. Noise in the data is actually the changes in the trend in the prices of the stock in the market, the changes are due to many factors like national events, major economic decisions and trade wars, etc. Despite having all of the previous history and mapping the trends of the stock market the movement of the prices remains uncertain and vulnerable to different events across the world. Despite investors traditionally relying on historical analyses to guide their trading decisions, the effectiveness of this market hypothesis remains uncertain.

Machine learning and deep learning techniques have gained traction in financial markets due to their potential for forecasting market prices and predicting financial trends[1]. The stock market

plays a pivotal role in defining profit maximisation and risk minimization in any economy. Investing in promising stocks can yield significant returns, although the non-linear nature of stock prices poses challenges for accurate prediction. Work is being done to look into more accurate and explainable methods for stock prediction and determining the movement of the stock's price over a given time window which will incorporate more profits and better investing solutions for everyone[3].

This paper aims to contribute to the discourse by applying the ARIMA model, a statistical method, and deep neural networks to forecast stock prices. Additionally, we employ recurrent neural networks, specifically Long Short-Term Memory (LSTM) cells, to predict stock price trends. Our study also incorporates convolutional neural networks (CNNs), including preprocessing and full CNN architectures, to analyse time series trends in stock data. Deep learning models have been the go to algorithms for different tasks where normally lots of sequential data is involved but they have also shown almost reliable performance in deriving relations for time series tasks especially stock market forecasting.

Through comprehensive evaluation, we assess the performance of these models and analyse their comparative effectiveness. By examining mean accuracy across all models, we aim to ascertain the efficiency and productivity of each algorithm. This research endeavours to provide insights into the most effective techniques for stock price forecasting[4], thus empowering investors to make informed decisions and optimise their investment strategies.

Methodology Overview

In this study, we adopt a meticulous approach to construct an accurate predictive framework for stock price forecasting, leveraging various deep learning models. Recognizing the principle that simplicity often yields superior results in deep learning, we embrace a multifaceted strategy, employing multiple algorithms to analyse their efficacy in predicting time series data[5].

Our primary dataset comprises S&P 500 ETF data sourced from Yahoo Finance, encompassing crucial information such as date, open and close values, high and low prices, and trading volume. Focusing on the closing price, we delineate our study period from the starting dates of the individual stocks to July 2022, utilising almost all the data available on each stock to complete our extensive study.

To evaluate the performance of each model, we utilise Root Mean Square Error **(RMSE)** as a benchmark, seeking to minimise prediction errors.

Our methodology encompasses three primary deep learning models:

1. ARIMA: (AutoRegressive Integrated Moving Average):
ARIMA, a statistical analysis model widely employed in time series forecasting, forms the bedrock of our approach. By ensuring data stationarity and determining significant values such as AIC, BIC, and HQIC, we tailor parameters (p, d, q) for different stocks, optimising predictive accuracy. We use different plots like Standardised Residual plot, Normal Q-Q plot, and Correlogram which provide us with a deep insight into the model's performance over particular stocks.

2. LSTM: (Long-Short-Term Memory):
LSTM, a recurrent neural network **(RNN)** variant, is well known for its ability for capturing long-term dependencies in sequential data which can be incorporated in time series

tasks as well . Comprising forget, input, and output gates, LSTM nodes maintain connections to previous-data-streams, enhancing the model's ability to learn intricate temporal patterns.

3. CNN:(Convolutional Neural Network):
CNN, traditionally employed in image processing, offers promising results in time series prediction due to its ability to extract hierarchical features efficiently. By sharing weights and employing layers for convolution, pooling, and flattening, CNN enhances computational efficiency while preserving predictive accuracy.

Through experimentation with these models, our aim is to evaluate their relative effectiveness in stock price forecasting. Deep neural networks, while powerful, pose challenges in optimising network performance, influenced by factors such as neuron count, hidden layers, training techniques, activation functions, and feature sets[6].

By comparing the performance of ARIMA, LSTM, and CNN models, we endeavour to elucidate their respective strengths and weaknesses. The insights derived from our methodology will help in the development of novel methods to analyse the markets and predict them.

Dataset

The primary dataset utilised in this study consists of S&P 500 ETF data obtained from Yahoo Finance. This dataset encompasses crucial information including date, open and close values, high and low prices, and trading volume for each trading day. The study period spans from the starting dates of the individual stocks to July 2022, enabling a comprehensive analysis leveraging almost all available data for each stock. Focusing primarily on the closing price, the dataset provides a comprehensive view of the stock market dynamics over the specified period. This extensive dataset facilitates a thorough examination of the performance of various deep learning models in predicting stock prices.

Model Architecture

In this study, we employ four distinct deep learning models to predict time series data for five stocks: Meta, Apple, Amazon, Netflix, and Google. Each model offers unique architectures and functionalities, contributing to a comprehensive analysis of their efficacy in stock price forecasting.

1. ARIMA (Auto Regressive Integrated Moving Average):
ARIMA is a statistical analysis model widely used in time series forecasting. It involves ensuring data stationarity and determining significant values such as AIC, BIC, and HQIC. By employing parameters (p, d, q) tailored for different stocks, ARIMA optimises predictive accuracy. Diagnostic plots including Standardised Residual plots, Normal Q-Q plots, and Correlograms provide insights into model performance.

2. Bi-LSTM (Bidirectional Long Short-Term Memory):
Bi-LSTM is a variant of the recurrent neural network (RNN) that excels in capturing long-term dependencies in time series prediction[7]. Comprising multiple LSTM layers, including bidirectional layers, Bi-LSTM enhances the model's ability to learn intricate temporal patterns. The model architecture incorporates Convolutional 1D layers for feature extraction followed by dense layers for prediction.

3. GRU (Gated Recurrent Unit):
Similar to **Bi-LSTM**, **GRU** is another variant of the RNN architecture designed to capture temporal dependencies effectively[8]. The model consists of multiple GRU layers followed by dense layers for prediction. Convolutional 1D layers are used for feature extraction, enhancing the model's ability to learn complex patterns in time series data.

4. CNN-LSTM (Convolutional Neural

Network with Long Short-Term Memory): CNN-LSTM combines the hierarchical feature extraction capabilities of CNNs with the memory retention properties of LSTM networks. The model architecture incorporates Convolutional 1D layers for feature extraction, followed by LSTM layers for capturing temporal dependencies. Dense layers are utilised for prediction, enabling the model to effectively forecast stock prices.

Each model is trained using the Adam optimizer and Huber loss function to minimise prediction errors. Training is performed over multiple epochs, with validation data used to monitor model performance. The resulting models are evaluated based on metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) to assess their accuracy in predicting stock prices.

By examining the architectural intricacies of each model and analysing their performance on the dataset, we aim to gain insights into the strengths and weaknesses of different deep learning approaches in stock price forecasting[9]. These insights contribute to the advancement of predictive analytics in financial markets, enabling investors to make informed decisions and optimise their trading strategies.

Experimental Results

Analysis

Before training our deep learning models, several preprocessing steps were undertaken to ensure the data's suitability for training. These steps included visualising the stock prices of Meta, Apple, Amazon, Netflix, and Google using Plotly graphs. Additionally, moving averages for different time windows were calculated to

provide a smoother representation of stock price trends. Furthermore, the daily returns for each stock were computed to assess the volatility and performance over time[10].

To prepare the data for model training, the closing prices were scaled using Min-Max scaling to normalise the data within a specified range. Subsequently, the data was split into training and testing sets, with 95% of the data used for training and the remaining 5% for testing. The training data was further split into input features (X) and target labels (y), with a sequence length of 30 days chosen for each input sample.

The BiLSTM[11], GRU, CNN-LSTM, and ARIMA models were trained using the prepared data to forecast the stock prices of Meta, Apple, Amazon, Netflix, and Google. Evaluation of each model on all the particular stocks were done using the Root Mean Square Error (RMSE) metric, which helps us in determining the difference between the predicted and actual stock prices for all the stocks.

Results

The predictive performance of the BiLSTM, GRU, CNN-LSTM, and ARIMA models was assessed across five prominent stocks: Meta, Apple, Amazon, Netflix, and Google. The metric used for the evaluation of the performance of the models across different stocks was the Root Mean Square Error (RMSE). The results revealed notable variations in prediction accuracy across different models and stocks[12].

The BiLSTM model demonstrated relatively moderate performance across the stocks. For Meta, Apple, and Amazon, the RMSE values were 12.29, 16.50, and 12.24, respectively, indicating reasonably accurate predictions. However, the model's performance was less favourable for Netflix and Google, with RMSE values of 21.22 and 60.45, respectively, suggesting relatively higher prediction errors.

Fig 3.2[1]. BiLSTM Prediction of Amazon Stock

Fig 3.2[2]. BiLSTM Prediction of Google Stock

The loss graphs of the corresponding stocks for the BiLSTM model are as follows:

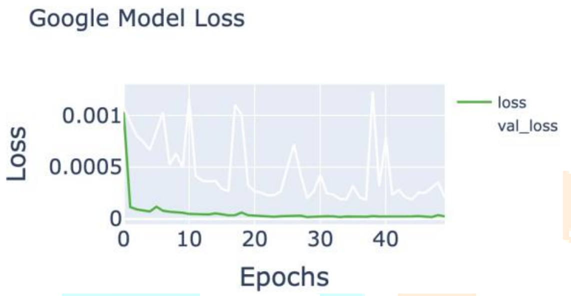


Fig 3.2[3]. BiLSTM Prediction Loss for Google

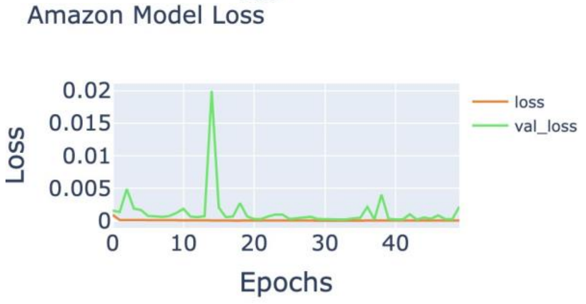


Fig 3.2[4]. BiLSTM Prediction Loss for Amazon

In contrast, the GRU model exhibited more consistent and improved performance across most stocks. It achieved lower RMSE values compared to BiLSTM for Meta (11.63), Apple (11.82), and Amazon (4.63), indicating more accurate predictions. However, its performance was similar to BiLSTM for Netflix (21.24) and Google (55.85).

Fig 3.2[5]. GRU Prediction of Amazon Stock

Fig 3.2[6]. GRU Prediction of Google Stock

The loss graphs of the corresponding stocks for the GRU model are as follows:

Fig 3.2[7]. GRU Prediction Loss for Amazon

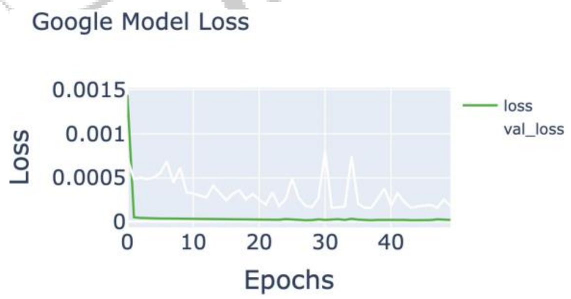


Fig 3.2[8]. GRU Prediction Loss for Google

The CNN-LSTM model, despite its potential, showcased notably higher RMSE values across all stocks, indicating less accurate predictions compared to BiLSTM and GRU[13]. The RMSE values were particularly elevated for Apple (129.22), Amazon (129.57), Netflix (373.90), and Google (2013.35), signifying significant prediction errors.

Fig 3.2[9]. CNN-LSTM Prediction for Meta

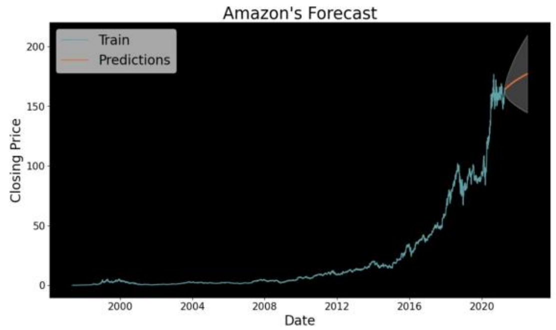
Fig 3.2[10]. CNN-LSTM Prediction for Google

The loss graphs of the corresponding stocks for the CNN-LSTM model are as follows:

Fig 3.2[11]. CNN-LSTM Prediction Loss for Meta

Fig 3.2[12]. CNN-LSTM Prediction Loss for Google

Additionally, the ARIMA model, a traditional time series forecasting approach, was evaluated for comparison. While it demonstrated competitive performance for some stocks, such as Amazon (RMSE = 28.10) and Apple (RMSE = 49.80), it yielded relatively higher RMSE values for others, including Meta (RMSE = 129.31), Netflix (RMSE = 203.82), and Google (RMSE = 324.77). The best parameters for each stock's ARIMA model were also identified, providing insights into the model's configuration for optimal performance.



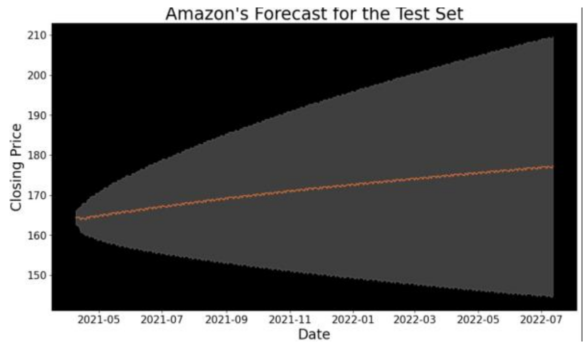
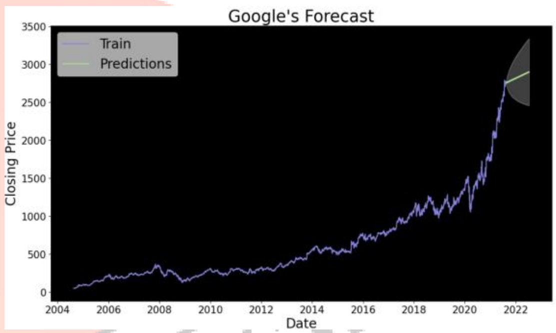


Fig 3.2[13,14]. ARIMA Prediction for Amazon



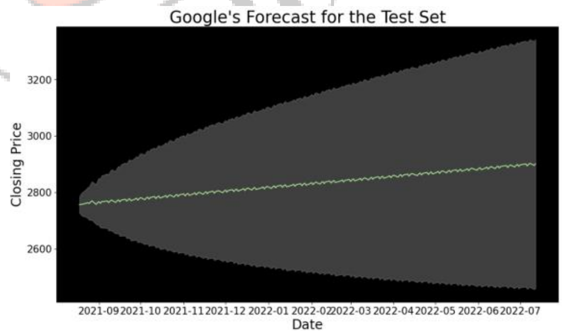


Fig 3.2[15,16]. ARIMA Prediction for Google

The following plots are the Normal Q-Q plots of the corresponding stocks mentioned above:

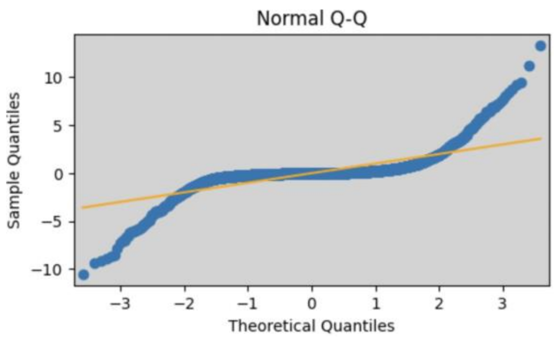


Fig 3.2[17]. ARIMA Normal Q-Q plot for Amazon

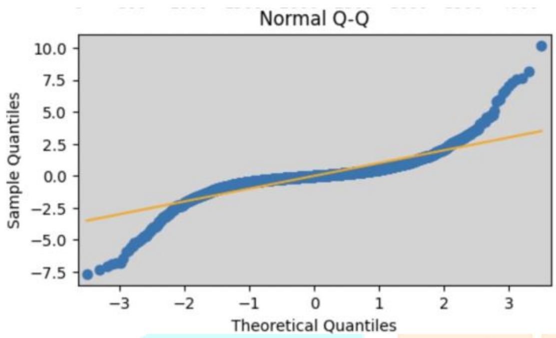


Fig 3.2[18]. ARIMA Normal Q-Q plot for Google

Overall, the experimental results underscore the importance of selecting appropriate models tailored to specific stocks for accurate stock price prediction. While newer deep learning models like BiLSTM and GRU show promise, traditional methods like ARIMA still hold relevance, emphasising the value of a diverse modelling approach in financial forecasting[14].

Overall, the evaluation of these models underscores the importance of selecting appropriate models and parameters for accurate stock price predictions. The results highlight the GRU model's superior performance compared to other deep learning models, while the ARIMA model's effectiveness varied across different stocks. These findings contribute to the understanding of predictive analytics in financial markets and provide insights for future research aimed at improving forecasting accuracy.

Conclusion

The present study investigated the efficacy of deep learning models, including BiLSTM, GRU, and CNN-LSTM, alongside the traditional ARIMA model, for stock price prediction across a diverse range of stocks including Meta, Apple, Amazon, Netflix, and Google. With the evaluation metric being the Root Mean Square Error (RMSE) which was the chosen performance metric, there were some key insights that were realised by analysing the numbers.

Firstly, the BiLSTM and GRU models demonstrated competitive performance, with varying degrees of accuracy across different stocks. While both models exhibited relatively accurate predictions for Meta, Apple, and Amazon, they encountered challenges in predicting the stock prices of Netflix and Google, where prediction errors were more pronounced.

In contrast, the CNN-LSTM model exhibited less favourable performance compared to BiLSTM and GRU, with significantly higher RMSE values observed across all stocks. This suggests limitations in leveraging convolutional neural networks for time series forecasting in the context of stock prices[15].

Furthermore, the traditional ARIMA model provided valuable insights, showcasing competitive performance for certain stocks such as Apple and Amazon, albeit with higher prediction errors for others like Meta, Netflix, and Google. The identification of optimal ARIMA parameters offers valuable guidance for future forecasting efforts.

Overall, the study highlights the importance of employing a diverse range of modelling techniques in stock price prediction, as no single model emerged as universally superior[16]. Instead, the selection of appropriate models should be tailored to the characteristics and dynamics of individual stocks.

Moving forward, future research could explore hybrid modelling approaches that combine the strengths of deep learning and traditional methods to further enhance predictive accuracy. Additionally, incorporating additional features and refining model architectures could offer avenues for improving forecasting performance in financial markets[17]. By advancing predictive analytics methodologies, researchers and practitioners can contribute to more informed decision-making and risk management in the dynamic domain of stock trading.

Improvements

Despite the advancements made in this study, there are several areas where further improvements can be made to enhance the effectiveness of stock price forecasting models. These include:

1. Feature Engineering: The selection and engineering of relevant features play a critical role in the performance of predictive models. Future research could explore advanced feature engineering techniques, including the extraction of non-linear relationships and interactions between variables, to improve model accuracy.

2. Hyperparameter Tuning: The performance of deep learning models is highly sensitive to the choice of hyperparameters such as learning rate, batch size, and network architecture,etc. More extensive hyperparameter tuning experiments could be conducted to identify optimal configurations for each model and dataset, thereby improving prediction accuracy.

3. Data Augmentation: Data augmentation techniques, commonly used in image processing, can be adapted for time series data to increase the diversity and size of the training dataset. Future research could explore the application of data augmentation methods such as time warping, temporal shifting and other techniques to improve the generalisation ability of predictive models.

4. Robustness to Market Dynamics: Stock markets are inherently volatile and subject to sudden shifts in sentiment and external factors[18]. Future research could focus on developing models that are robust to changes in market dynamics, including the ability to adapt to regime shifts and identify anomalous patterns in real-time.

By addressing these areas, we can build more accurate, reliable, and interpretable stock price forecasting models that meet the needs of investors and financial practitioners in dynamic market environments and also provide insights into the market in real-time.

Future Scope

While this study provides valuable insights into the efficacy of various deep learning and traditional models for stock price prediction, there are several avenues for future research that merit exploration. These include:

1. Hybrid Modeling Approaches: Future studies could investigate the potential of hybrid models that combine deep learning techniques with traditional statistical methods. By integrating the predictive power of models such as ARIMA with the feature extraction capabilities of deep learning architectures, hybrid models may offer improved forecasting accuracy.

2. Incorporation of External Factors: The inclusion of external factors such as macroeconomic indicators, news sentiment analysis, and geopolitical events could enhance the predictive capabilities of stock price forecasting models.

3. Ensemble Techniques: Ensemble learning techniques, which combine multiple models to make predictions. By aggregating the predictions of diverse models, ensemble techniques can mitigate the weaknesses of individual models and produce more robust forecasts.

4. Real-Time Prediction: Developing models capable of real-time stock price prediction is essential for enabling timely decision-making in the financial markets. Research could focus on the development of streaming data architectures and algorithms capable of processing and analysing data in real-time to provide forecasts on which people could take actions in real-time.

5. Interpretability and Explainability: Enhancing the interpretability and explainability of stock price forecasting models is crucial for gaining trust and acceptance from investors and financial practitioners. Future studies could explore techniques for interpreting the decision-making processes of complex deep learning models and providing transparent explanations for their predictions.

6. Evaluation Metrics: While this study primarily evaluated model performance using the Root Mean Square Error (RMSE) metric, future research could explore the use of alternative evaluation metrics that capture different aspects of prediction accuracy. Metrics such as Mean Absolute Percentage Error (MAPE) and Directional Accuracy (DA) may provide complementary insights into model performance.

Moving forward in the direction discussed above, a lot of new findings and developments can be made in the area of stock market prediction which can help retrieve valuable understanding of how the markets function.

References

[1] Forecasting Directional Movement of Stock Prices using Deep Learning. Ann Data Sci. 2022 Aug 1 : 1–18.doi: 10.1007/s40745-022-00432-6

[2] Choi, H.: Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model. arXiv:1808.01560 (2018)

[3] R. M. I. Kusuma, T.-T. Ho, W.-C. Kao, Y.-Y. Ou, and K.-L. Hua, "Using deep learning neural networks and candlestick chart representation to predict stock market," 2019, https://arxiv.org/abs/1903.12258.

[4] X. Hou, S. Zhu, L. Xia, and G. Wu, "Stock price prediction based on Grey Relational Analysis and support vector regression," in *Proceedings of the Thirtieth Chinese Control and Decision Conference (CCDC-2018)*, pp. 2509–2513, Shenyang, China, June 2018.

[5] N. Ritzmann and J. C. Nievola, "A generalised financial time series forecasting model based on automatic feature engineering using genetic algorithms and support vector machine," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, July 2018.

[6] R. Zhao, Y. Deng, M. Dredze, A. Verma, D. Rosenberg, and A. Stent, "Visual attention model for cross-sectional stock return prediction and end-to-end multimodal market representation learning," in *Proceedings of the Thirty-Second International Florida Artificial Intelligence Research Society Conference (FLAIRS-32)*, Melbourne, FL, USA, May 2018.

[7] J. Sen, "Stock price prediction using machine learning and deep learning frameworks," In Proceedings of the 6th International Conference on Business Analytics and Intelligence, Bangalore, India, 2018.

[8] F. Feng, H. Chen, X. He, Ji Ding, M. Sun, and T. Chua, "Enhancing stock movement prediction with adversarial training," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-19)*, pp. 5843–5849, Macao, China, August 2019.

[9] A. Mudinas, D. Zhang, and M. Levene, "Market trend prediction using sentiment analysis: lessons learned and paths forward," 2019, http://arxiv.org/abs/1903.05440.

[10] S. Mehtab, J. Sen and A.Dutta, "Stock price prediction using machine learning and LSTM-based deep learning models," In Proc. of the 2nd Symposium on Machine Learning and Metaheuristic Algorithms and Applications, Chennai, India, October 2020. (Accepted)

[11] H. Q. Nguyen, A. H. Rahimyar, and X. Wang, "Stock forecasting using M-band

wavelet-based SVR and RNN-LSTMs models," 2019, http://arxiv.org/abs/1904.08459.

[12] Li, X., Li, Y., Yang, H., Yang, L., Liu, X.: DP-LSTM: differential privacy-inspired LSTM for stock prediction using financial news. arXiv:1912.10806 (2019)

[13] Adusumilli, R.: Predicting stock prices using a keras LSTM model. NikolaNews (2019)

[14] Zhong, W., Gu, F.: Predicting local protein 3D structures using clustering deep recurrent neural network. IEEE/ACM Trans. Comput. Biol. Bioinform. (2020). https://doi.org/10.1109/tcbb.2020.3005972

[15] Sarode, Sumeet, et al. "Stock price prediction using machine learning techniques." 2019 International Conference on Intelligent Sustainable Systems (ICISS). IEEE, 2019.

[16] Lim, Yi Xuan, and Consilz Tan. 2021. Do negative events really have deteriorating effects on stock performance? A comparative study on Tesla (US) and Nio (China). *Journal of Asian Business and Economic Studies* 29: 105–19.

[17] Li, Yang, and Yi Pan. 2021. A novel ensemble deep learning model for stock prediction based on stock prices and news. *International Journal of Data Science and Analytics* 13: 139–49.

[18] Li, Audeliano Wolian, and Guilherme Sousa Bastos. 2020. Stock Market Forecasting Using Deep Learning and Technical Analysis: A Systematic Review. *IEEE Access* 8: 185232–242.