

LSTM GAN ARCHITECTURE FOR STOCK PRICE PREDICTION

1

INTRODUCTION

- Stock market prediction is a challenging and popular problem in the field of finance.
- Deep learning models can be used to derive patterns in historical data and natural language processing can be used to understand market sentiment.
- Using extensive feature engineering techniques and exploring novel architectures to do so can greatly improve the accuracy of the prediction of price movements.
- Such models can be used by investors and economists to beat the market and understand the dynamics of the system.

PROBLEM STATEMENT

System for an accurate prediction of stock price can be built with:

- effective pre-processing and feature engineering of relevant stock data
- performing time series and textual analysis to build a deep learning architecture
- using RMSE (root mean squared error) as performance metric to compare the predictions made by the deep learning models

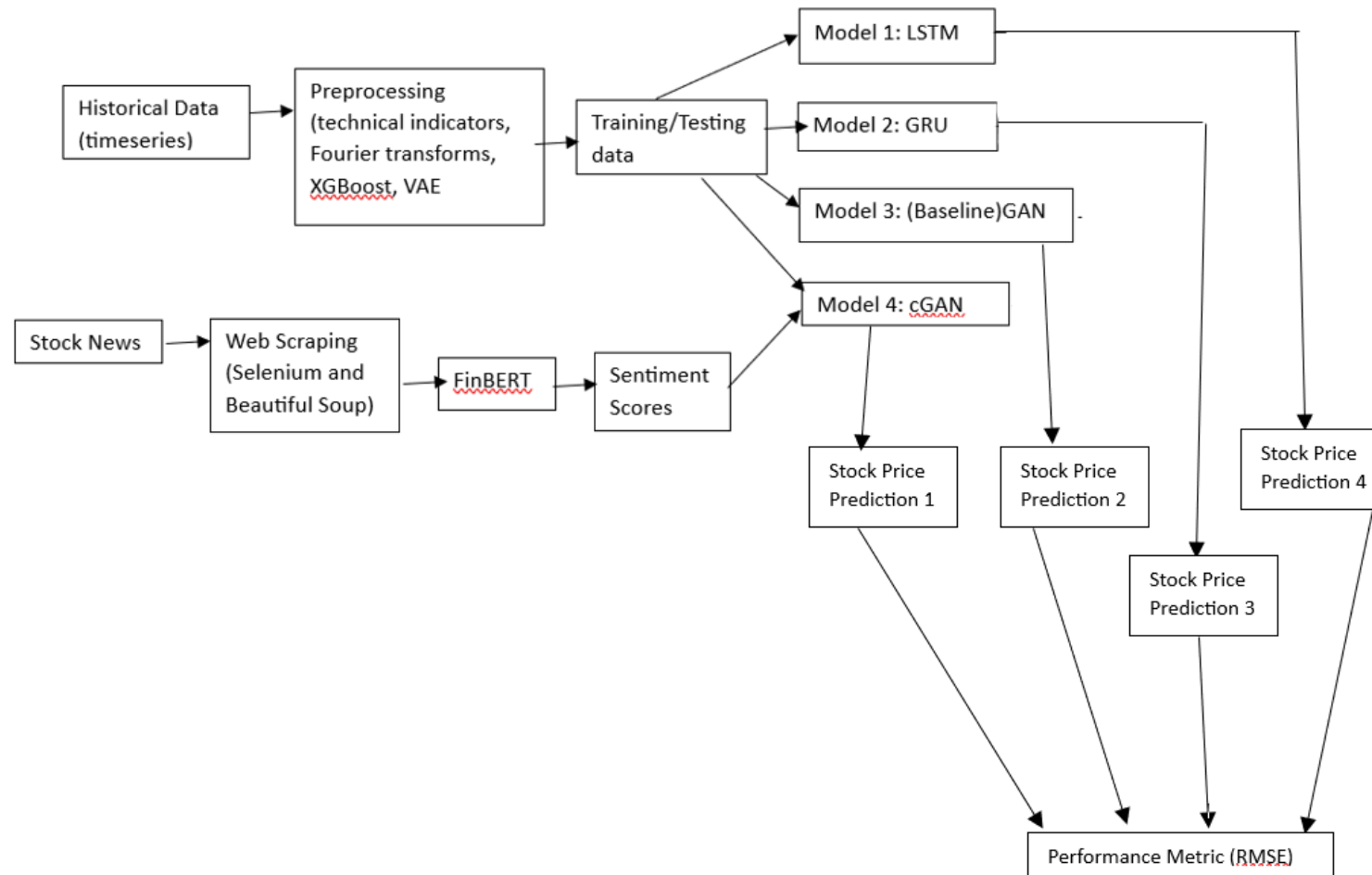
RESEARCH OBJECTIVE

The research aims to achieve accurate stock market prediction by:

- Feature engineering to extract latent space features using variational autoencoders
- Build a cGAN architecture to accurately predict stock prices
- Generating stock specific news sentiment scores to condition the GAN architecture
- Compare the RMSE of the cGAN model with baseline deep learning models namely RNN, LSTM and GAN

As a whole, this project aims to build a cGAN architecture that analyses historical data and news sentiment for stock price prediction and comparing its performance with baseline deep learning models.

PROPOSED SYSTEM: DIAGRAM



PROPOSED SYSTEM: MODULES

- Data (collection, pre-processing and feature engineering)
- Deep learning models (LSTM, GRU, GAN)
- Designing and developing CGAN model
- Web Scraping and Sentiment Analysis

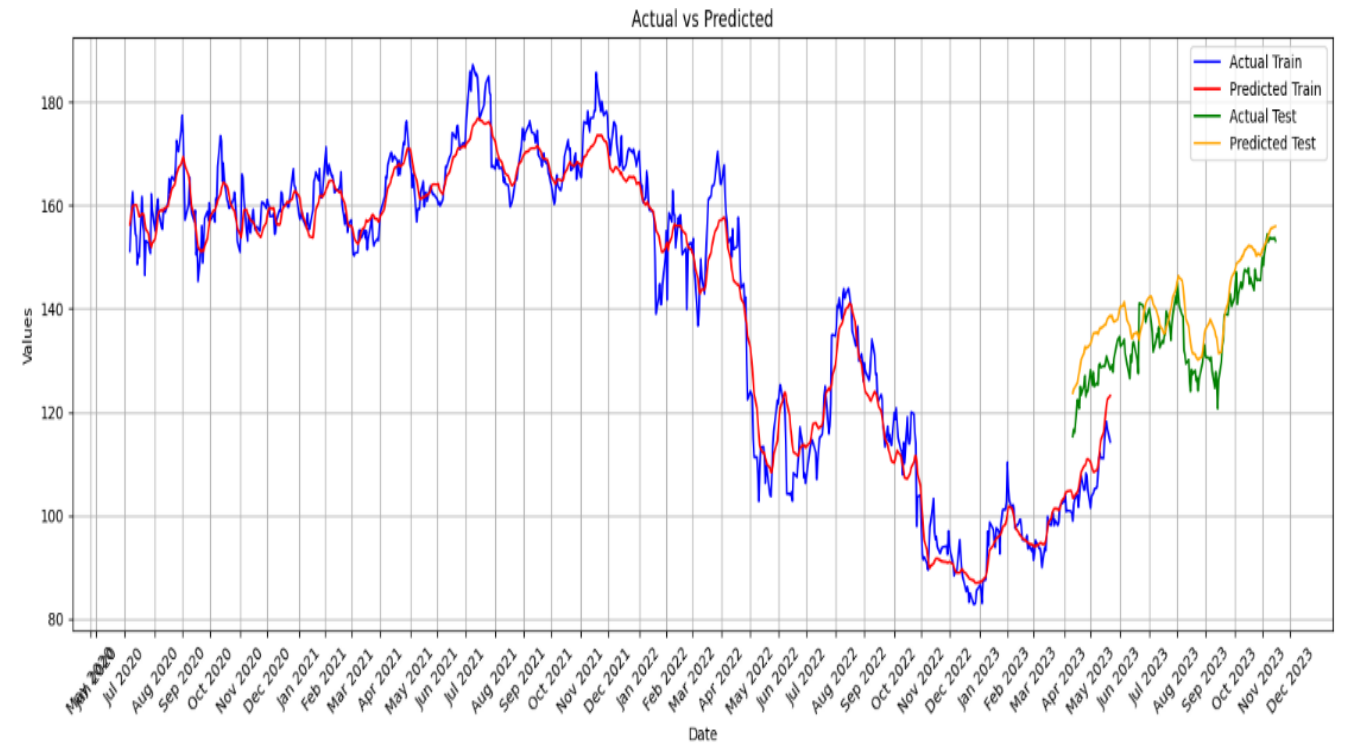
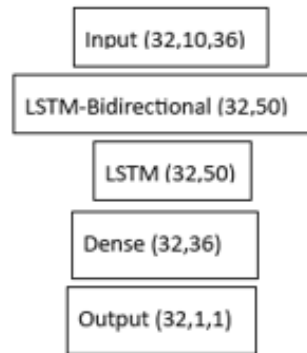
PROPOSED SYSTEM: EXPLANATION

- Stock price time-series data is collected from Python library yfinance. The historical data contains Open, Low, High, Close, Adjusted Close and Volume attributes.
- Amazon.com Inc. stock is used and data from 01-01-2021 to 01-01-2024 is collected.
- Technical indicators like SMA, MACD, RSI, ATR, Bollinger Bands and RSV, and Fourier transform values of closing price of stock are added as additional parameters to the dataset.
- XGBoost is performed for feature analysis.
- Variational encoders are used to extract latent space features.

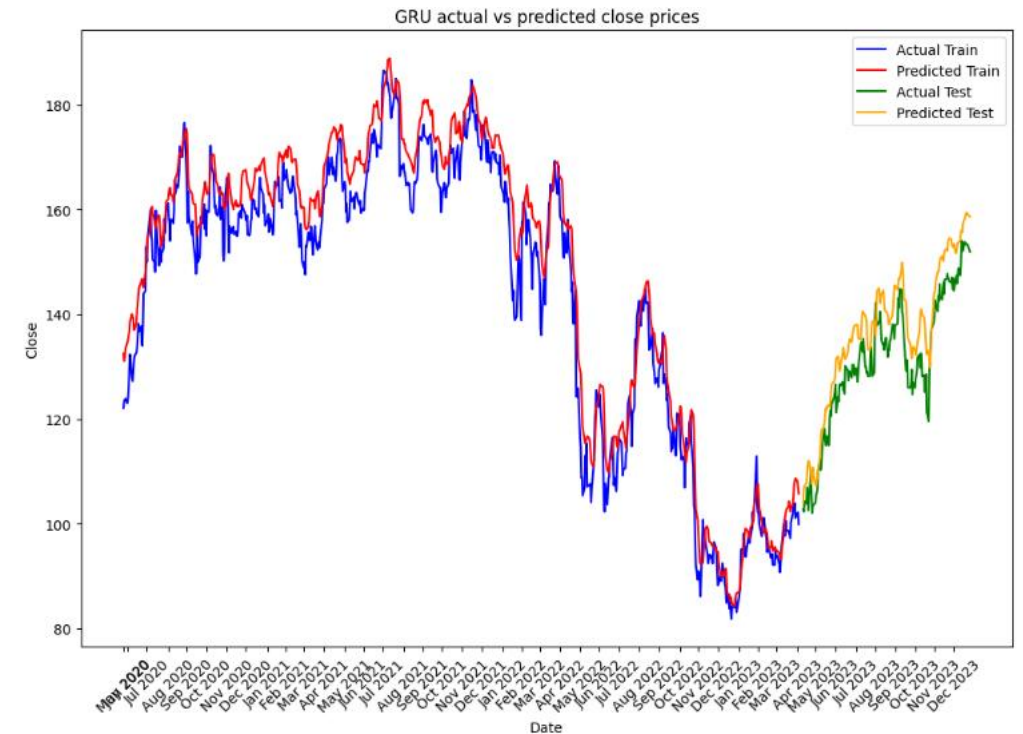
PROPOSED SYSTEM: EXPLANATION

- Designing and developing the deep learning models namely, LSTM, GRU and GAN which will later be used to compare prediction results with proposed CGAN architecture. A time step of 10 is used to predict next 1 day stock price.
- The financial news is collected from SeekingAlpha, a leading finance news outlet.
- Amazon.com Inc. stock is used and news from 01-01-2021 to 01-01-2024 is collected.
- Web scraping is done using Selenium and BeautifulSoup.
- FinBERT is used to get sentiment scores.

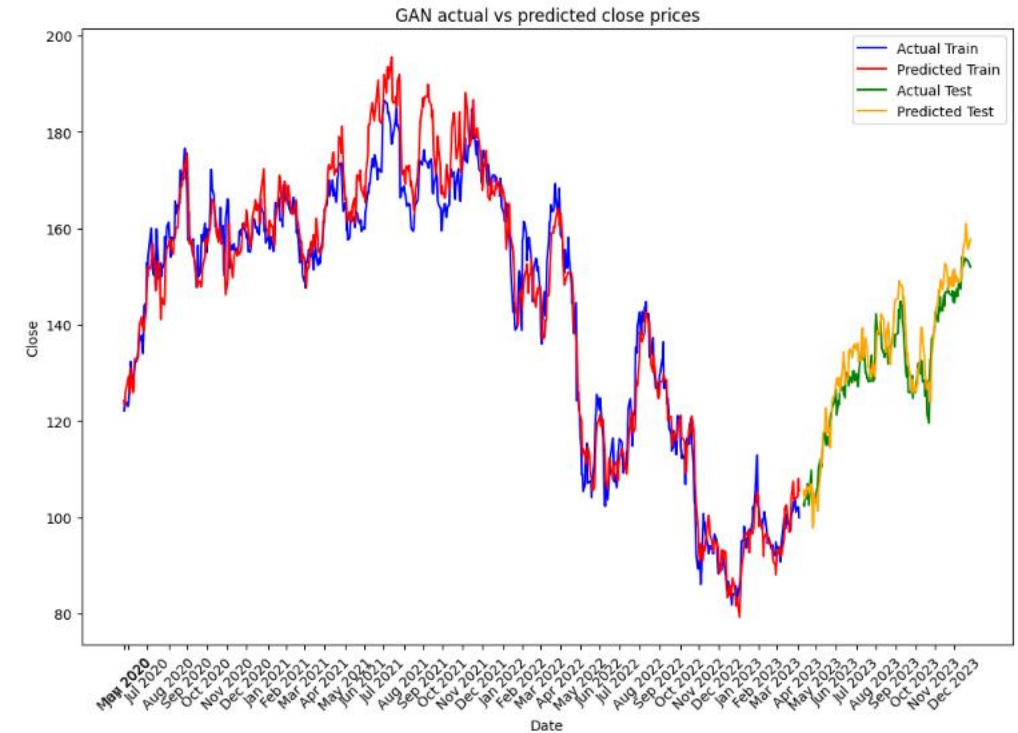
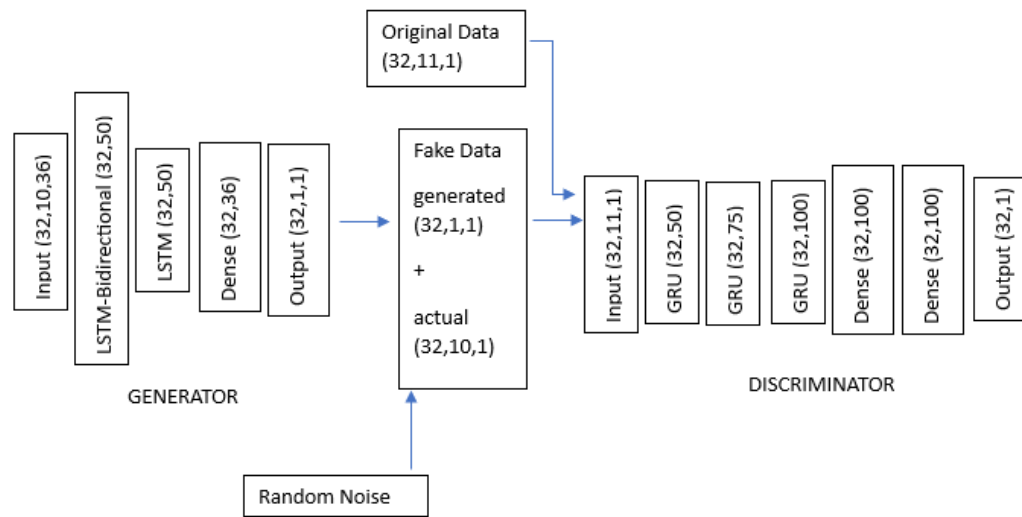
IMPLEMENTATION AND RESULTS (LSTM)



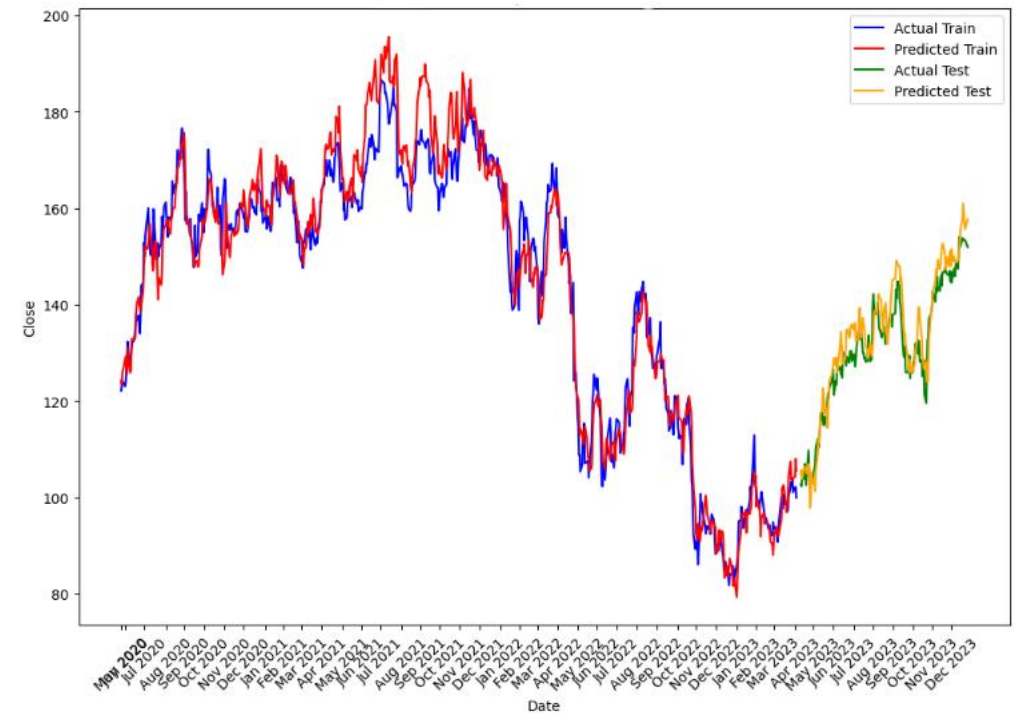
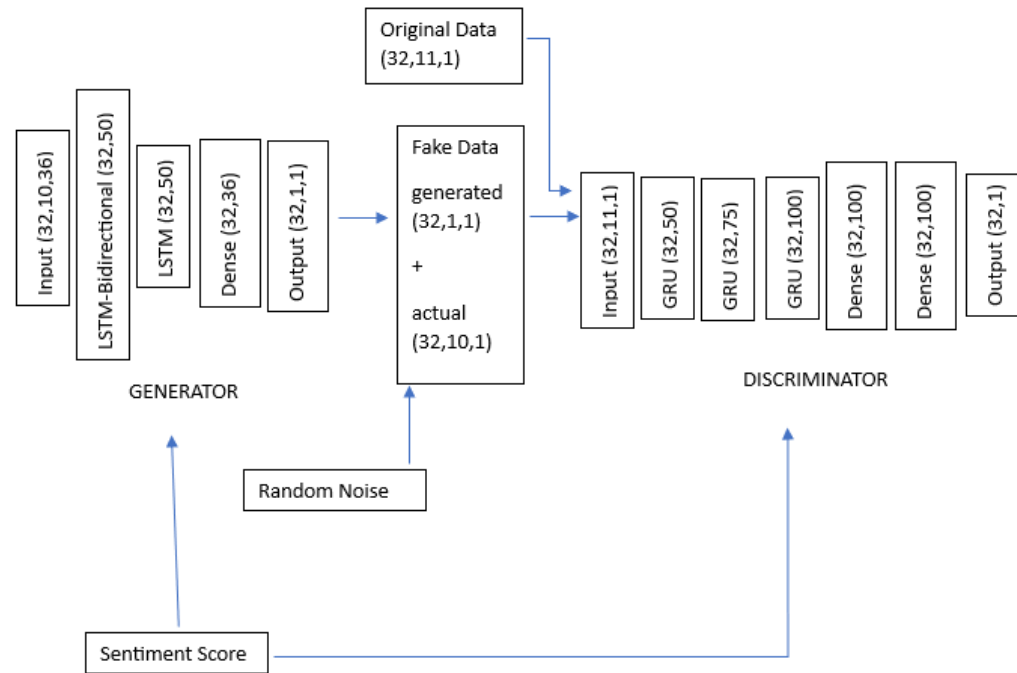
IMPLEMENTATION AND RESULTS (GRU)



IMPLEMENTATION AND RESULTS (GAN)



IMPLEMENTATION AND RESULTS (CGAN)



IMPLEMENTATION AND RESULTS (RMSE)

Model	LSTM	GRU	Basic GAN	<u>cGAN</u>
Training Data	4.79	6.30	5.07	4.90
Testing Data	5.93	6.37	4.44	4.48

CONCLUSION

This work combines time-series data and twitter sentiment analysis model to predict the price of a stock for a given day. A CGAN with stacked bi-directional LSTM as generator and GRU as discriminator along with conditional parameter of sentiment scores provides best results according to the chosen performance metrics.

FUTURE WORK

- Scraping a variety of news outlets can be done to get a fuller picture and avoid any inherent bias.
- Real-time prediction by accessing live tweets through Twitter's paid developer interface can actively make suggestions to investors in real-time.
- Studying a longer period of time can help in establishing long term trends that can aid investors in long term investments.
- Comparing different types of stock as research suggests that they have different trends and hence require different trading strategies.

REFERENCES

1. Çelik, T. B., İcan, Ö., & Bulut, E. (2023). Extending machine learning prediction capabilities by explainable AI in financial time series prediction. *Applied Soft Computing*, 132, 109876.
2. R. Patel, V. Choudhary, D. Saxena and A. K. Singh, "LSTM and NLP Based Forecasting Model for Stock Market Analysis," 2021 First International Conference on Advances in Computing and Future Communication Technologies (ICACFCT), Meerut, India, 2021, pp. 52-57, doi: 10.1109/ICACFCT53978.2021.9837384.
3. Ganesh, P., & Rakheja, P. (2018). Vlstm: Very long short-term memory networks for high-frequency trading. arXiv preprint arXiv:1809.01506.
4. Lin, C., & Lobo Marques, J. A. (2023, July). An Exploratory Comparison of Stock Prices Prediction using Multiple Machine Learning Approaches based on Hong Kong Share Market. In *Proceedings of the 2023 14th International Conference on E-business, Management and Economics* (pp. 311-316).
5. Zhang, K., Zhong, G., Dong, J., Wang, S., & Wang, Y. (2019). Stock market prediction based on generative adversarial network. *Procedia computer science*, 147, 400-406.
6. Brophy, E., Wang, Z., She, Q., & Ward, T. (2023). Generative adversarial networks in time series: A systematic literature review. *ACM Computing Surveys*, 55(10), 1-31.
7. Tiano, D., Bonifati, A., & Ng, R. (2021, June). FeatTS: Feature-based time series clustering. In *Proceedings of the 2021 International Conference on Management of Data* (pp. 2784-2788).
8. Saha, S., Gao, J. & Gerlach, R. A survey of the application of graph-based approaches in stock market analysis and prediction. *Int J Data Sci Anal* 14, 1–15 (2022). <https://doi.org/10.1007/s41060-021-00306-9>

REFERENCES

9. Fataliyev, K., Chivukula, A., Prasad, M., & Liu, W. (2021). Stock market analysis with text data: A review. arXiv preprint arXiv:2106.12985.
10. Ho, T. T., & Huang, Y. (2021). Stock price movement prediction using sentiment analysis and CandleStick chart representation. *Sensors*, 21(23), 7957.
11. Masi, G., Prata, M., Conti, M., Bartolini, N., & Vyetrenko, S. (2023, November). On Correlated Stock Market Time Series Generation. In *Proceedings of the Fourth ACM International Conference on AI in Finance* (pp. 524-532).
12. S. M. Carta, S. Consoli, L. Piras, A. S. Podda and D. R. Recupero, "Explainable Machine Learning Exploiting News and Domain-Specific Lexicon for Stock Market Forecasting," in *IEEE Access*, vol. 9, pp. 30193-30205, 2021, doi: 10.1109/ACCESS.2021.3059960.
13. Gite, S., Khatavkar, H., Kotecha, K., Srivastava, S., Maheshwari, P., & Pandey, N. (2021). Explainable stock prices prediction from financial news articles using sentiment analysis. *PeerJ Computer Science*, 7, e340.
14. Ghahramani, M., & Najafabadi, H. E. (2022). Compatible deep neural network framework with financial time series data, including data preprocessor, neural network model and trading strategy. arXiv preprint arXiv:2205.08382.
15. Borovkova, S., & Tsiamas, I. (2019). An ensemble of LSTM neural networks for high-frequency stock market classification. *Journal of Forecasting*, 38(6), 600-619.