[1] Kushal Kumar Potta
[2] Koushik S
[3] Khushi Agrawal
[4] Dr. Shubha Rao
[5] Dr. Sindhu K

NeuroTrader: Redefining Portfolio Strategies through Advanced Machine Learning

Journal of Electrical Systems

Abstract: - The stock market is cooperative trading network involving company shares and their derivatives. This market provides dominant amount of contribution on contemporary economies as it's a place where companies raise huge amount of money to speed up start-ups, to extend existing business, consolidate actions and pay off debt. Stock price prophecy is important for value investments in the stock market. In particular, short-term prediction that exploits financial news articles is promising in recent years. In this paper, an innovative approach to enhance stock market prediction, addressing key limitations in existing forecasting models is introduced. This research explores how advanced technologies like Artificial Intelligence and Machine Learning can improve stock market predictions and sentiment analysis from web-scraped news articles. By using unstructured data for models finBERT, Vader, and LSTM, the work aims to build a model that aids to make financial analysis more accurate and accessible to everyone.

I. Introduction

In financial market analysis, there's been a shift from traditional methods to adapting advanced technologies, mirroring the broader evolution of the finance industry. In the past, humans were the main players in financial analysis, manually sifting through loads of data for insights. However, these traditional methods were not very quick in responding to fast-changing market dynamics. Recently, the finance world has undergone a significant transformation due to digitalization and the rise of big data. The surge in diverse and large-scale data has made traditional methods less effective, creating a demand for more advanced, automated, and accurate analytical tools. This is where machine learning comes into play, revolutionizing various industries, including finance. Initially, machine learning in finance was limited, but its potential to transform the sector became clear over time. Three standout models in financial analysis are finBERT, VADER and LSTM.

Through this research, we aim to contribute to the field of stock market and provide people with a more accurate analysis and prediction of the future market. Three prediction models using LSTM, VADER and finBERT have been built and proposed in the paper.The rest of this research paper is organized as follows: Section II provides a detailed review of related literature and existing approaches to stock market prediction. Section III outlines the architecture of the proposed model, Section IV lays insight into the methodology used, Section V presents the flow of the developed project. Finally, Section VI concludes the paper with a summary of the findings, contributions, and avenues for future research.

II. Detailed Survey

Alhujaili, et.al [1] explores sentiment analysis for Arabic educational YouTube videos using machine learning and deep learning methods. The research focuses on understanding viewer emotions through comments and employs various techniques like oversampling and SMOTE for data balancing. The study demonstrates that machine learning classifiers and deep learning models, especially SVC, RF, and DL, can achieve high accuracy up to 96% in sentiment classification. This approach opens new possibilities for evaluating educational content quality and has broader applications in areas like market trend prediction

Saraswat, S., et. al [2] presents sentiment analysis of audio files. The study converts spoken language into text and then uses machine learning algorithms to analyze these texts for emotional sentiment. The practical applications of this study are significant, especially in customer service and AI development, offering a new dimension to understanding public perception and sentiment in financial markets.

Boukabous et. al [3] emphasizes the role of sentiment analysis in security intelligence and crime detection. The paper discusses the limitations of text-based sentiment analysis and the potential of audio sentiment analysis. The authors combine NLP and machine learning to analyze text data sentiment and leverage speech features and machine learning algorithms to process audio data. Using the Sphinx Library and the XD-Violence video dataset, the paper benchmarks the dataset on audio using Convolutional Neural Network (CNN) and Recurrent Neural Network

[1, 2, 3, 4, 5] Department of ISE, BMS College of Engineering, Bangalore, India

(RNN), then applies BERT for text analysis. The combination of CNN and BERT yields the best results, with an accuracy of 85.63%, a loss of 30.47%, and an F1-score of 85.16%. The study highlights the potential of multimodal sentiment analysis in crime detection and the need for further research to enhance the performance of such systems.

Juyal, et.al [4] explores multimodal sentiment analysis (MSA) using audiovisual data. The research introduces a method that combines feature extraction and emotion recognition from both text and visual modalities using CNN. Human emotions are understood coining audiovisual streams in sentiment analysis. The datasets used were RAVDESS and COVAREP.

Saumya et. al [5] studies on isolating sexist remarks, hateful speeches and sarcasm detection using DNN and CNN. The model is trained on audio and text based inputs to predict speaker sentiment. The proposed work achieves a 61.89% accuracy rate and minimal divergence between testing and training loss and accuracy.

S. Roy et. al [6] proposes a model for answering multiple choice questions (MCQs) by using BERT with CNN. The datasets used where TQA and SciQ. The model achieves a 22.7% improvement over LSTM based systems.

Xiaojun, et.al [7] proposed an improved model for named entity recognition by combining BERT with BiLSTM-CNN, especially for Chinese railway construction texts. This work is important both in terms of quick speed, accuracy, and specificity of complex information. The scope of research goes beyond railways into potential applications regarding stock market predictions.

Pranav, et. al [8] deals with the issue of hate speech and toxic language detection in text. The different approaches used comprise traditional machine learning models, BERT, fastText embeddings, and DNN. The authors compare different algorithms and techniques, and concluded that the deep learning models using embeddings of BERT and fastText outperform basic methods in terms of accuracy and performance. The highest accuracy of 98.75% was achieved using CNN with BERT embeddings on the merged dataset (ALONE-HASOC-Mixed). This study underlines a need for developing new techniques in the improvement of toxic speech detection systems and, therefore, warrants further research in this direction.

Matheus et. al [9] uses BERT for sentiment analysis of news articles in the stock market. The authors fine-tune a BERT model on a dataset consisting of 582 stock news articles, having achieved a 72.5% F-score. They also show how the model can be used for predicting the movements of the Dow Jones Industrial (DJI) Index. This paper explores the challenge of breaking financial news analysis and proposes BERT for rapid sentiment analysis to support investors. The paper explains in detail the architecture of BERT and its applications in sentiment analysis.

Wei, et.al [10] proposed "News2Trend," a deep learning framework for forecasting stock trends from news. This framework uses NLP and deep learning to extract relevant information from news to predict stock trends. The authors prove that their approach outperforms the existing methods on different stock markets' datasets. The paper demonstrates the importance of textual analysis for predicting stocks and the leverage of deep learning and NLP in the exploitation of news data towards better decision-making for investors. The new model employed the BERT language model in predicting future stock movement, a correlation matrix enhancing predictive outcomes. Experimental results have provided adequate confirmation of the method as it has greatly improved the accuracy of the stock prediction compared to baseline systems.

Sidogi, et. al [11] discussed how financial news sentiment impacts a firm's stock prices. The research paper utilizes LSTM networks and FinBERT to conduct the sentiment analysis and shows that including news headline sentiment improves the accuracy of predicting stock price drastically. The findings provide useful knowledge toward integrating sentiment analysis with complex models of financial forecasting.

The authors, Ayyappa et. al [12], analyse stock price prediction using LSTM along with BERT on sentiment analysis of tweets. The use of LSTM for sequential patterns and BERT for contextual understanding of sentiment in tweets better interprets the robust and accurate predictions. The authors recognize the importance of quantitative factors along with qualitative aspects and then put focus light on the implications that sentiment of the masses may play in the stock market. It is one of the models that boasts of offering high accuracy for financial forecasting.

H. Ma et.al [13] present a forecasting model using BERT, BiGRU, and Attention mechanisms to investigate how investor sentiments affect stock prices. The model combines sentiment analysis of online investor comments and traditional trading data for stocks, with higher accuracy compared to traditional LSTM models. The present study highlights the importance of combining both sentiment and trading data to make accurate predictions of stock prices; many scopes are available for improvement in the future.

J. Choi, et. al [14] suggests a technique of stock market movement prediction through the fusion of numerical stock price data with textual information from news. This hybrid approach, merging time-series data with semantic features, could prove to be an all-around analysis tool for stock market prediction. The study, therefore, gives an indication that in the future, stock prediction is no longer a matter of number-crunching but also deciphering a story.

Sarangpure et al [15] use PyCaret with Streamlit to streamline the ML processes. It focuses on AutoML by providing an automation of tasks such as data pre-processing and model selection. The integration of PyCaret for rapid model deployment and Streamlit for user-friendly interfaces is emphasized. The paper thus represents a significant milestone in making ML more accessible and efficient in many industries.

III. Proposed Architecture

The user interface is built and hosted with Streamlit, an open source python library. It serves as the central component, interacting with three models: VADER for sentiment analysis, an LSTM model for price prediction trained on past data, and FinBERT for advanced sentiment analysis. The user interface collects stock ticker(s) of interest and sends requests to the models. The VADER model fetches news data from Finviz, parses headlines, calculates sentiment scores, and plots sentiment analysis. The LSTM model fetches historical stock price data, preprocesses it, sequences the data for LSTM input, creates and trains the LSTM model, and saves the trained model and scalar parameters. The FinBERT model fetches news data from Finviz, parses headlines, calculates sentiment scores using a pre-trained model, and plots sentiment analysis. The UI displays the predictions from all models, providing a comprehensive analysis for the specified stock ticker.

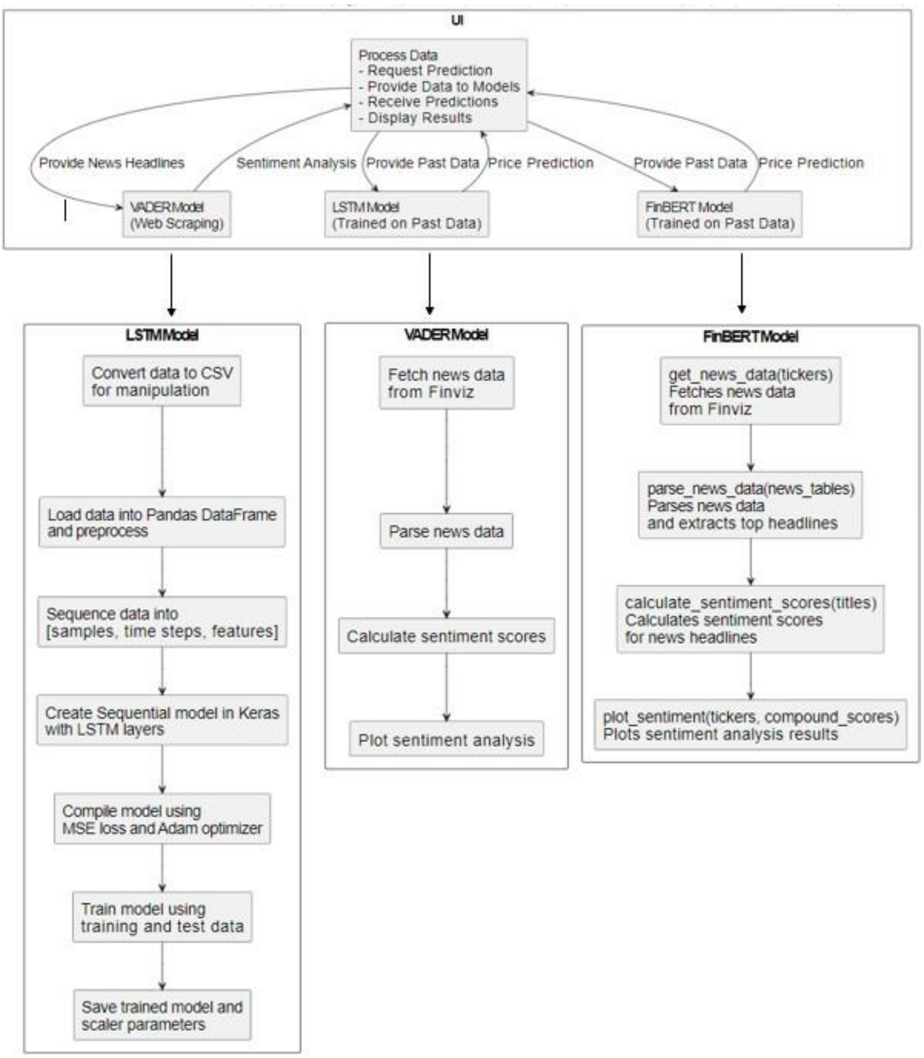

Fig. 1 Architecture

The user can select the company name and the date for which he wants to know the predicted closing price. The date selected by the user should be within 30 days from the current date. On hitting the predict button, the LSTM model runs in the background to display the predicted closing price for that particular stock. A graph is also generated which plots the closing price of that stock over the past three months and its predicted closing price for the next 30 days. The model fetches historical stock price data from Alpha Vantage API and preprocesses it. It then sequences data for LSTM model input and creates LSTM model architecture in Keras. It compiles model using MSE loss and Adam optimizer, trains model with training data and validates with test data and saves trained model and scalar parameters.

**# Recent Data: Taken From Quandl **

	index	Date	open	high	low	close	adjusted close	volume	dividend amount	split coefficient	Scaled High
90	6,147	2024-04-09	190.54	191.25	186.66	189.31	None	None	None	None	0.8174
91	6,148	2024-04-10	187.42	187.915	185.52	186.04	None	None	None	None	0.7406
92	6,149	2024-04-11	186.04	186.795	184.58	185.9	None	None	None	None	0.7149
93	6,150	2024-04-12	184	185.1699	181.685	182.27	None	None	None	None	0.6774
94	6,151	2024-04-15	185.57	187.48	180.88	181.25	None	None	None	None	0.7306

Fig. 2. LSTM model prediction

A. VADER Model

The VADER model fetches news data from Finviz for given stock tickers and organizes it into tables. It extracts top 10 news headlines for each ticker, calculates compound sentiment scores for each headline, plots sentiment analysis graphs with sentiment categories and mean sentiment score. The main function defines stock ticker(s) and calls other functions in sequence to fetch news data, parse headlines, calculate sentiment scores, and plot analysis.

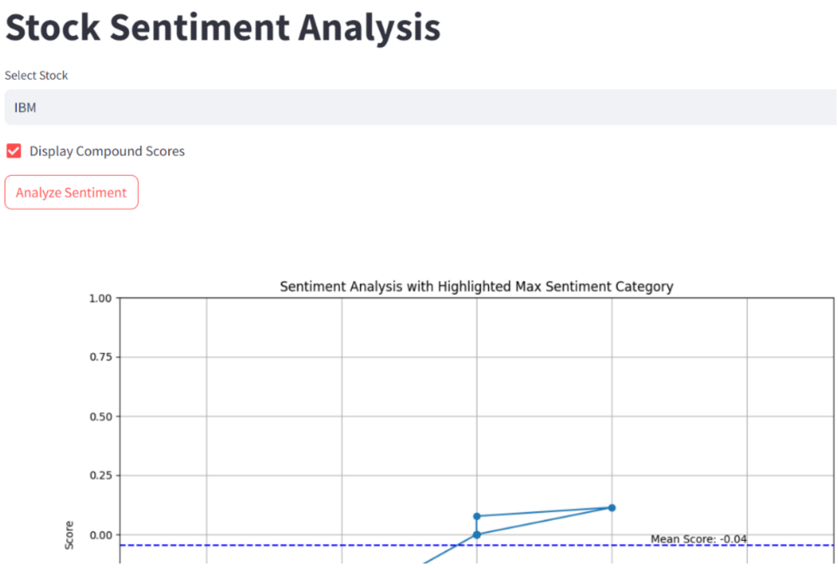

Fig. 3. Sentiment analysis using Vader

Sentiment Analysis Plot

Titles with Compound Scores:

	Title	Compound Score
1	Unlocking IBM (IBM) International Revenues: Trends, Surprises, and Prospects	0.4767
2	Is Trending Stock International Business Machines Corporation (IBM) a Buy Now?	0.0000
3	3 Dividend Stocks to Buy on the Dip: April 2024	0.0000
4	What is an Annuity for Retirement? 15 Dividend Stocks to Buy Instead	0.0000
5	Why I have hated this earnings season and you might too	-0.6369
6	American Politicians are Buying These 10 AI Stocks	0.0000
7	3 Quantum Computing Stocks That Could Be Multibaggers in the Making: April Edition	0.0000
8	Slingshot Stocks: 3 Picks Primed to Catapult Higher in the Next Market Surge	0.0000
9	Buy the Dip on IBM Stock	0.0000
10	Microsoft is 'the highest quality company one can own': Analyst	0.0000

Fig. 4. News titles with their predicted compound score

B. *finBERT Model*

The finBERT model fetches news data from Finviz for given stock tickers, parses news data and extracts top 10 news titles for each ticker, calculates sentiment scores for each news title using a pre-trained sentiment analysis model. It plots sentiment analysis results, categorizes scores, and displays them on a bar chart. It also plots sentiment scores for each news title, highlights maximum sentiment category, and shows mean sentiment score.

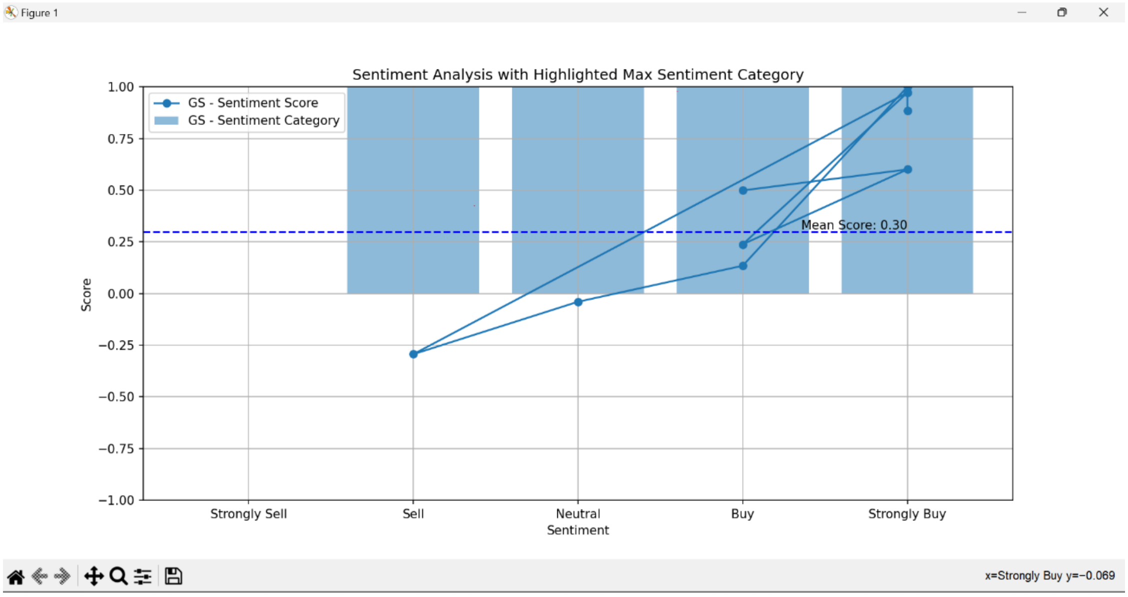

Fig. 5. Sentiment analysis using finBERT model

IV. Flow Of Proposed Service

The proposed system is designed to provide accurate and timely predictions of stock prices through a series of well-defined steps. It begins with the collection of historical stock price data from the Alpha Vantage API in JSON format. This data is then exported in the form of a CSV file for easier manipulation and analysis. Then, the data will be loaded to a Pandas DataFrame and preprocessed-that is, extracting 'high' prices and scaling with MinMaxScaler to a range between 0 and 1-and then sequenced into the required format for input into the LSTM model. The two additional models are also integrated to enhance the process of stock price prediction. The VADER model will be used for sentiment analysis by fetching news data from Finviz for the specified stock tickers, then parse the headlines to calculate compound sentiment scores. These scores give additional insight into the market's sentiment, influencing stock price. In contrast, the use of the FinBERT model encompasses a pre-trained sentiment analysis model, yiyanghkust/finbert-tone. This further examines news headlines and computes normalized sentiment scores between -1 and 1. The sentiment scores and predictions from the LSTM model provide a full analysis of stock price trends to help investors make informed decisions.

V. Results

The project demonstrates the predictive capabilities of the LSTM model in predicting the closing price of a stock for a certain date in the future. To achieve this, the user selects a stock of interest, upon which the LSTM model uses historical stock price data to train and make the predictions. This model is based on multiple factors, including previous price trends and patterns, to make the closing price prediction of the equity for that date in the future. Users can really use this prediction insight to make decisions on investment.

In addition, both VADER and FinBERT are applied to identify news that will be related to your stock and further to analyze sentiment behind them. VADER and FinBERT have different approaches to sentiment analysis, with the former relying on lexicon-based analysis, and the latter on a pre-trained model tailored for financial news. After this extraction, both models give sentiment scores to each of the articles, showing the overall sentiment expressed. The compound scores given by VADER and FinBERT give us insight into the market sentiment encompassing the stock chosen.

One important result obtained is a comparison of compound sentiment scores taken from FinBERT and VADER in 10 common news titles. This interesting piece of insight indicates the performance of the two models in sentiment analysis.

Title	Compound score: finBERT	Compound score:VADER
Why Goldman Sachs (GS) is a Great Dividend Stock Right Now	0.88	0.62
Goldman Sachs doesn't want its bankers to play hooky at the Paris Olympics	1.00	0.29
Goldman Sachs cracks down on Paris trips	0.13	0.00
Top 12 Alternatives to Edward Jones	-0.04	0.20
Dividend Stock Portfolio For Income	-0.29	0.20
20 States with Highest Hispanic Population Growth Rates	0.97	0.38
25 Richest Billionaires in Real Estate Industry	0.24	0.53
Republicans are Buying These 10 Oil and Gas Stocks	0.60	0.00
30 Most Miserable Countries in the World	0.50	-0.54
Goldman Sachs' Top 15 Stock Picks for 2024	-1.00	0.20

Fig. 5. Comparison of compound score between finBERT and vader

VI. Conclusion

The proposed work demonstrated the integration of advanced AI and machine learning techniques, such as finBERT and LSTM models, in the realm of stock market prediction and sentiment analysis. Using these technologies, the proposed work displays a remarkable prediction accuracy regarding stock prices and market sentiment, valuable information for investors and traders. The user-friendly interface and real-time processing, in summary, make the system highly accessible and usable. Therefore, the future for this project is immense in terms of further development and applications of the financial analysis field, which has the potential to revolutionize market data interpretation and utilization in decision-making processes.

References

[1] Alhujaili, Rawan Fahad, and Wael MS Yafooz. "Sentiment analysis for YouTube educational videos using machine and deep learning approaches." 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI). IEEE, 2022.

[2] Saraswat, S., Bhardwaj, S., Vashistha, S. and Kumar, R., 2023, March. Sentiment Analysis of Audio Files Using Machine Learning and Textual Classification of Audio Data. In 2023 6th International Conference on Information Systems and Computer Networks (ISCON) (pp. 1-5). IEEE.

[3] Boukabous, Mohammed, and Mostafa Azizi. "Multimodal sentiment analysis using audio and text for crime detection." 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET). IEEE, 2022.

[4] Juyal, Prachi. "Multi-modal Sentiment Analysis of Audio and Visual Context of the Data using Machine Learning." 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC). IEEE, 2022.

[5] Roy, S., Ghoshal, S., Basak, R., Basu, P. and Roy, N., 2022, February. Multimodal sentiment analysis of human speech using deep learning. In 2022 Interdisciplinary Research in Technology and Management (IRTM) (pp. 1-4). IEEE.

[6] S. Roy, N. Ehtesham, M. S. Islam and Marium-E-Jannat, "Augmenting BERT with CNN for Multiple Choice Question Answering," 2021 24th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2021, pp. 1-5, doi: 10.1109/ICCIT54785.2021.9689877.

[7] X. Wu, T. Zhang, S. Yuan and Y. Yan, "One Improved Model of Named Entity Recognition by Combining BERT and BiLSTM-CNN for Domain of Chinese Railway Construction," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022, pp. 728-732, doi: 10.1109/ICSP54964.2022.9778794.

[8] P. Malik, A. Aggrawal and D. K. Vishwakarma, "Toxic Speech Detection using Traditional Machine Learning Models and BERT and fastText Embedding with Deep Neural Networks," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1254- 1259, doi: 10.1109/ICCMC51019.2021.9418395.

[9] M. G. Sousa, K. Sakiyama, L. d. S. Rodrigues, P. H. Moraes, E. R. Fernandes and E. T. Matsubara, "BERT for Stock Market Sentiment Analysis," 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), Portland, OR, USA, 2019, pp. 1597-1601, doi: 10.1109/ICTAI.2019.00231.

[10] Wei, Feng, and Uyen Trang Nguyen. "Stock trend prediction using financial market news and BERT." Wall Street Journal (2018).

[11] T. Sidogi, R. Mbuvha and T. Marwala, "Stock Price Prediction Using Sentiment Analysis," 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia, 2021, pp. 46- 51.

[12] Y. Ayyappa, B. V. Kumar, S. P. Priya, S. Akhila, T. P. Vardhan Reddy and S. M. Goush, "Forecasting Equity Prices using LSTM and BERT with Sentiment Analysis," 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 643-648, doi: 10.1109/ICICT57646.2023.10134443.

[13] H. Ma, J. Ma, S. Liang and W. Du, "A Model of Integrating Bert and BiGRU+ Attention Dual-channel Mechanism for Investor Sentiment Analysis of Stock Price Forecast," 2022 IEEE/ACIS 23rd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Taichung, Taiwan, 2022, pp. 126-131, doi: 10.1109/SNPD54884.2022.10051779.

[14] J. Choi, S. Yoo, X. Zhou and Y. Kim, "Hybrid Information Mixing Module for Stock Movement Prediction," in IEEE Access, vol. 11, pp. 28781-28790, 2023, doi: 10.1109/ACCESS.2023.3258695.

[15] Sarangpure, N., Dhamde, V., Roge, A., Doye, J., Patle, S. and Tamboli, S., 2023, March. Automating the Machine Learning Process using PyCaret and Streamlit. In 2023 2nd International Conference for Innovation in Technology (INOCON) (pp. 1-5). IEEE.