INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 18-02-2024

Accepted: 28-03-2024

Published: 29-04-2024

Citation: Kasture P, Shirsath K (2024) Enhancing Stock Market Prediction: A Hybrid RNN-LSTM Framework with Sentiment Analysis. Indian Journal of Science and Technology 17(18): 1880-1888. https://doi.org/10.17485/IJST/v17i18.466

* **Corresponding author**.

pranjali.kasture@thakureducation.org

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Enhancing Stock Market Prediction: A Hybrid RNN-LSTM Framework with Sentiment Analysis

Pranjali Kasture[1]*, Kamini Shirsath[2]

1 Research Scholar, Department of Computer Engineering, K K Wagh Institute of Engineering Education and Research, Nashik, Savitribai Phule Pune University, Pune, Maharashtra, India
2 Professor, Department of Computer Engineering, Sandeep Institute of Engineering & Management, Nashik, Maharashtra, India

Abstract

Objectives: Predicting stock prices with accuracy is a difficult but crucial endeavor for market participants. To increase the precision of stock market forecasts, this study suggests a novel method that blends sophisticated neural network algorithms with sentiment assessment. **Methods:** News data is pre-processed. Each text document received a sentiment score reflecting overall sentiment. These scores were integrated into the feature set, combining textual sentiment information with historical stock price data from BSE Sensex. The proposed model of hybrid RNN-LSTM is applied and compared with the Random Forest Regressor (RFR), and Support Vector Regressor (SVR). The LSTM model is also applied and tested on data without sentiment analysis scores. **Findings:** The proposed model yields promising results in stock market prediction accuracy. It significantly gives a low value for mean absolute error (0.036), mean squared error (0.021), and root mean square error (0.046) when compared with the SVR and RFR models. The R2 value is also compared with literature methods, and it shows a 0.40% to 5.5% enhancement in the scores. The results prove that the incorporation of sentiment analysis enriches the predictive capabilities of the model. **Novelty:** Sentiment analysis combined with the hybrid RNN-LSTM framework provides a new technique to increase the accuracy of stock market forecasts. Using sophisticated knowledge of market dynamics and sentiments, the proposed approach gives important results to market participants, investors, and analysts of financial markets.

Keywords: Stock Market; Sentiment Analysis; Evaluation Metric; Prediction

1 Introduction

A vital economic arena that affects many different industries and people is the stock market. Accurately predicting stock price patterns is a difficult task that has repercussions on managing risks, market reliability, and sound financial decisions. Predictive modelling approaches have continuously advanced due to the complexity of the stock market, which is influenced by both internal and external forces. Stock prices are influenced by various factors, including the national and global economies. In the financial sector, traditional methods include time series analysis, technical analysis, and fundamental analysis[1]. In addition to evaluating intrinsic value, fundamental analysis takes into account inflation and interest rate variables. The volume, price trend, and investor mindset are the primary concerns of the extensive technical analysis[2]. However, traditional fundamental analysis faces challenges in terms of accuracy and dependence on analyst expertise. Financial time series, viewed as a random walk, pose additional challenges due to uncertainty and high noise[3]. Attempts to predict short-term stock prices using linear models like VAR, BVAR, ARIMA, and GARCH encounter skepticism due to the dynamic evolution of relationships over time[4].

Single models, whether linear or neural networks exhibit shortcomings in predicting stock prices effectively. The current trend emphasizes the synthesis of different techniques and efficient algorithms in deep learning for financial time series data[5]. To forecast stocks, some study presents a hybrid methodology using deep learning approaches. Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN) are the methods employed. To anticipate stock prices, the work expands on previous research by fusing deep learning and sentiment analysis. This framework offers a new direction for financial forecasting by integrating sentiment analysis and deep learning[6]. Extensive research has focused on achieving accurate stock market predictions encompassing various traditional and modern methodologies.

1.1 Traditional methods

Fundamental analysis examines financial statements, market patterns, and macroeconomic variables. However, it relies on accurate financial data and expert analysis, introducing subjectivity. Technical analysis uses tools like moving averages and candlestick charts to predict future moves based on historical price and volume trends. Time series analysis models stock prices as a time-dependent sequence, using methods like ARIMA and GARCH. While effective in capturing certain dynamics, they struggle with non-linear relationships and sudden shifts in market sentiment.

1.2 Modern approaches

Historical data is employed for training in machine learning techniques like linear regression and decision trees. However, their performance is hindered by the volatile and non-stationary nature of stock price data. Deep learning models, especially RNNs and LSTMs[7–9], have gained traction for capturing complex temporal dependencies. However, using these models alone may not fully leverage the rich information in stock market data. Sentiment analysis, a facet of natural language processing, enhances stock market predictions by gauging market sentiment from news articles, social media, and financial reports[10,11]. Hybrid models integrating sentiment analysis have shown promise in improving forecast accuracy. Table 1 lists different approaches used in stock prediction along with challenges. Our proposed model addresses these challenges by proposing a novel hybrid model, integrating sentiment analysis with an RNN-LSTM architecture to enhance stock price forecasting accuracy.

Table 1. Different approaches for stock prediction

Approaches	Methodology	Focus	Notable Models/ Parameters	Challenges	References
Traditional Approaches	Fundamental Analysis	Financial statements, industry trends, macroeconomic factors	Earnings per Share (EPS), Dividend Payout Ratio (DPR), Return on Equity (ROE)	Subjective evaluations, dependence on analyst expertise	[1]
	Technical Analysis	Past price and volume patterns	Moving averages, RSI, candlesticks	Reliability debate, self-fulfilling prophecies	[2]
	Time Series Analysis	Modeling stock prices as time-dependent sequences	ARIMA, GARCH	Non-linear relationship handling, sudden sentiment changes	[3], [4]
Modern Approaches	Machine Learning	Historical price and volume data	Linear regression, decision trees, support vector machines	Hindered by high volatility, non-stationary nature of data	[5] [6]
	Deep Learning	Capturing complex temporal dependencies	RNN, LSTM	Potential to overlook rich information in stock market data	[7], [8], [9]

Continued on next page

Table 1 continued

Sentiment Analysis in Finance	Measuring market sentiment from textual data	SVM, RNN, BERT, LSTM	Dependency on quality and diversity of textual data	(10), (11)

2 Methodology

2.1 Data Collection and Preprocessing

Historical stock price data of BSE Sensex is used, encompassing daily open, close, high, low, and trading volumes. Data is collected from Jan 2011 to Jan 2021 from reputable financial sources (finance.yahoo.com). News data is pre-processed which involves removing stop words, and special characters, and employing text normalization techniques to enhance data quality. A sentiment lexicon, comprising positive and negative words with assigned polarity scores, was utilized. Additionally, pre-trained sentiment analysis models like VADER were employed for automated sentiment scoring. Each text document received a sentiment score reflecting overall sentiment. These scores were integrated into the feature set, combining textual sentiment information with historical stock price data.

2.2 Proposed Model Architecture

The architecture for the proposed model is shown in Figure 1. Integrating sentiment analysis with a hybrid RNN-LSTM architecture holds promise in revolutionizing stock market prediction.

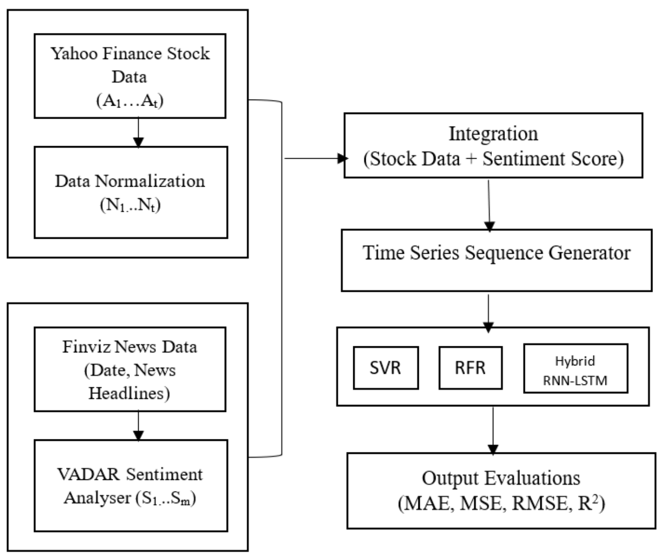

Fig 1. Proposed Stock Market Prediction Framework

RNN component, functioning as a feature extractor for text data, utilized a one-dimensional architecture with multiple recurrent layers of different filter sizes to capture patterns and relationships within sentiment data. The LSTM network is crafted to capture temporal dependencies within stock price data and process historical stock price and volume data. Its effectiveness in modelling sequential data was harnessed to learn long-term patterns and dependencies. Outputs from the RNN and LSTM components were concatenated and connected to a fully connected layer, followed by an output layer for stock price predictions. The training process tries to lessen the errors. The dataset underwent division into training, validation, and test sets. Training supported model learning, validation assisted in hyperparameter tuning, and the test set evaluated model performance. Hyperparameter tuning involved adjustments to the learning rate, number of convolutional filters, LSTM units, dropout rate, and experimentation with different sentiment lexicons and pre-trained sentiment analysis models. The evaluation matrix provides insights into the accuracy and reliability of predictions.

2.2.1 ***Proposed model step-by-step procedure***

Below is the step-by-step procedure for stock prediction using the proposed method.

Procedure: Stock-Prediction with hybrid RNN-LSTM
Input: Historical stock price data, Sentiment analysis-derived features
Output: Predicted stock prices
Step 1: Collect historical stock price data and sentiment analysis-derived features.
Step 2: Preprocessing the data involves data cleaning and normalization.
Step 3: Create 3 sets of data mainly training, validation, and test sets.
Step 4: Design the RNN layers to extract features from the sentiment analysis-derived data
Step 5: Construct the LSTM layers to capture temporal dependencies in the stock price data.
Step 6: Integrate the RNN and LSTM components into a hybrid model architecture.
Step 7: Compile the hybrid model with the appropriate loss function and optimization algorithm.
Step 8: With the training data, train the hybrid model, adjusting model parameters as necessary.
Step 9: Adjust hyperparameters and evaluate the model that was trained via the validation set.
Step 10: Assess the hybrid model's performance using the test data.
Step 11: To optimize the model, if needed, repeat steps 4–10.
Step 13: End Procedure

*2.2.2 **Mathematical Formulation***

With each row denoting a data point and each column an attribute, let X be the input data matrix. The actual stock prices are represented by the target variable vector, denoted by Y. An LSTM component's output is represented by the function $f_L(X)$. An RNN component's output is represented by the function $f_R(X)$.

$$Y' = f_H(X) \tag{1}$$

$$f_H(X) = \alpha . f_L(X) + (1-\alpha) . f_R(X) \tag{2}$$

Here, **Y '** represents the predicted stock prices, and α is a hyperparameter controlling the contribution of each component.

2.3 Other models for stock market prediction

2.3.1 Support Vector Regression (SVR)

The core idea behind SVR is to identify a hyperplane that effectively separates classes plotted in an n-dimensional space as shown in Figure 2.

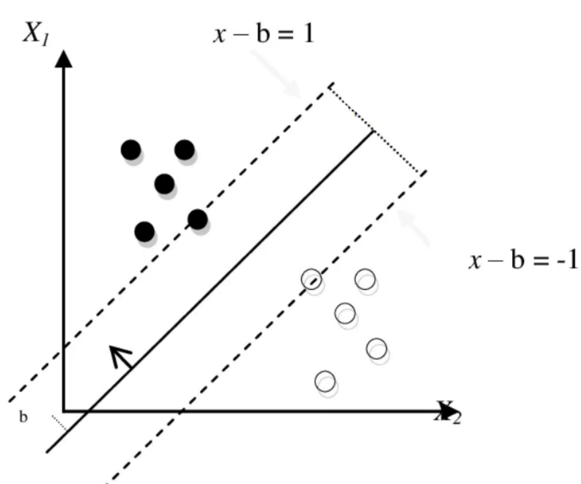

Fig 2. Hyperplane in SVR

We employed SVR as a powerful tool to capture intricate patterns in market sentiment and price dynamics. The SVR model was instantiated using the scikit-learn library, configured with a polynomial kernel of degree 2. This choice was made to accommodate potential nonlinear relationships inherent in market data[12].

The SVR model was trained on sentiment analysis-derived data, aiming to elucidate the interplay between sentiment signals and stock price movements. Our objective was to discern patterns and trends that could contribute to more accurate predictions. Following the training phase, the SVR model was employed to predict stock prices based on sentiment analysis inputs, providing valuable insights into the potential impact of market sentiment on future price trajectories. To assess the predictive prowess of our SVR-based approach, we generated predictions for unseen test data. Post-processing steps, including inverse scaling to revert predictions to their original scale, were meticulously applied. This facilitated a thorough evaluation of the model's performance against ground truth data, providing a comprehensive assessment of its effectiveness in real-world market scenarios.

2.3.2 ***Mathematical formulation of Support Vector Regression***

SVR with a linear kernel:

$$Y' = w^t \, . \, X + b \tag{3}$$

where **Y'** is the predicted stock prices, **w** represents the weight vector, **X** is the input feature vector, and **b** denotes the bias.

SVR with a non-linear kernel:

$$Y' = \sum_{i=1}^{n} \alpha i \, K(Xi, X) + b \tag{4}$$

where **Y'** represents the predicted stock prices, $\alpha_{\mathbf{i}}$ denotes Lagrange multipliers, **K (Xi, X)** represents kernel function computing the similarity between input feature vectors $\mathbf{X_i}$ and **X ,** and **b** is the bias term.

2.3.3 Random Forest Regression (RFR)

We configured 200 decision trees in a Random Forest Regressor through the scikit-learn module. This helps capture complex relationships between sentiment features and stock price changes. After putting together, the feature set and sentiment-driven predictions, we trained the Random Forest Regressor on our curated dataset. This allowed the model to learn and adjust to the intricate details of market sentiment and price changes. Then, we checked how well the model predicted unseen data, looking at its ability to understand and apply learned patterns to real-world market situations. Through careful checks and performance evaluations, our goal was to confirm the effectiveness and reliability of our Random Forest-based approach in predicting stock prices[(13)].

2.3.4 ***Mathematical Formulation of Random Forest Regression***

The Random Forest Regression model syndicates multiple decision trees to predict the output. The predicted stock price Y' is obtained by averaging the predictions of all decision trees in the forest.

$$Y' = \frac{1}{N} \sum_{i=1}^{n} y_i \tag{5}$$

where **N** is the number of decision trees in the forest, and y_i is the predicted stock price from the $\mathbf{i^{th}}$ decision tree.

3 Results and Discussion

The proposed hybrid model, enhanced by sentiment analysis, marks a noteworthy advancement in stock market prediction. The experimental results highlight its potential to reshape decision-making in the financial sector. Based on the analysis of the error comparison table, it becomes evident that the proposed model, augmented by sentiment analysis, emerges as the most accurate and reliable predictor of stock market trends. The model performance has been demonstrated through the use of evaluation metrics such as mean absolute error, mean square error, root mean square error, and R^2 score. Our hybrid model demonstrates better predictive abilities than other algorithms (Table 2), such as LSTM without sentiment analysis, Support Vector Regressor, and Random Forest Regressor, with significantly lower mean absolute error, mean squared error, root mean square error, and higher R^2 score. Figure 3 shows an analysis of evaluation metrics for existing models and the proposed model.

Table 2. Result analysis of different models

Continued on next page

Table 2 continued

Model	Features	MAE	RMSE	MSE
LSTM	O-H-L-C-V	0.163	0.184	0.337
Support Vector Regressor	O-H-L-C-V +News	0.077	0.109	0.118
Random Forest Regressor	O-H-L-C-V +News	0.039	0.069	0.047
Proposed	O-H-L-C-V +News	0.036	0.046	0.021

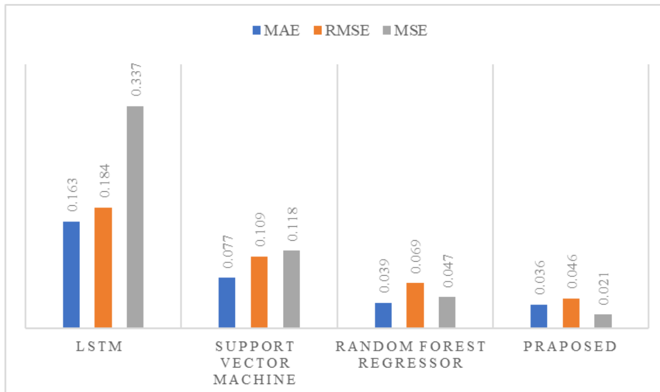

Fig 3. Result analysis of existing models and proposed model

Visual examination of prediction graphs as shown in Figures 4, 5, 6 and 7 further supported the efficacy of the proposed model, revealing a more consistent and coherent trend pattern compared to alternative models. The graph shows actual versus predicted closing prices for the last 8 months of periods.

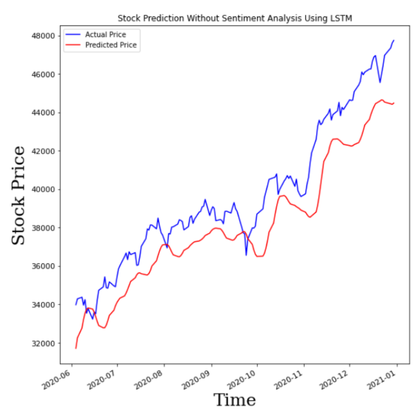

Fig 4. Stock price prediction without sentiment analysis using LSTM

Table 3. Comparative Analysis of existing and proposed models for R^2 score

Model	R^2 score
Ancy John et al. (14)	0.981
Srijiranon et al. (15)	0.965
Chen Y (16)	0.93
Proposed	0.985

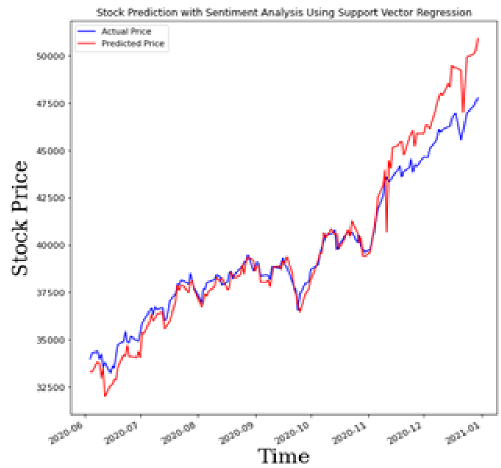

Fig 5. Stock price prediction with sentiment analysis using SVR

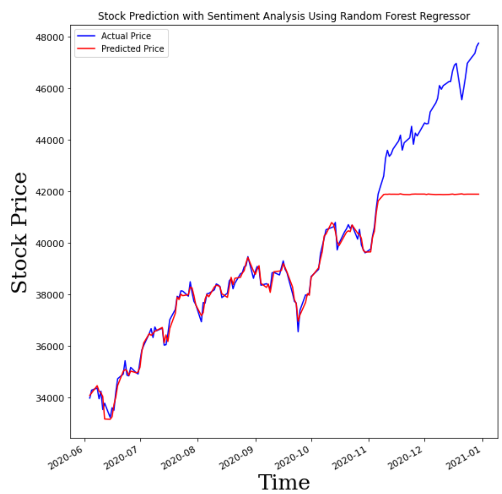

Fig 6. Stock price prediction with sentiment analysis using RFR

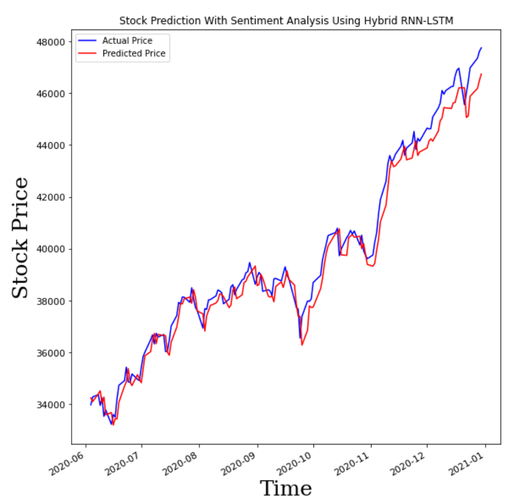

Fig 7. Stock price prediction using the proposed model

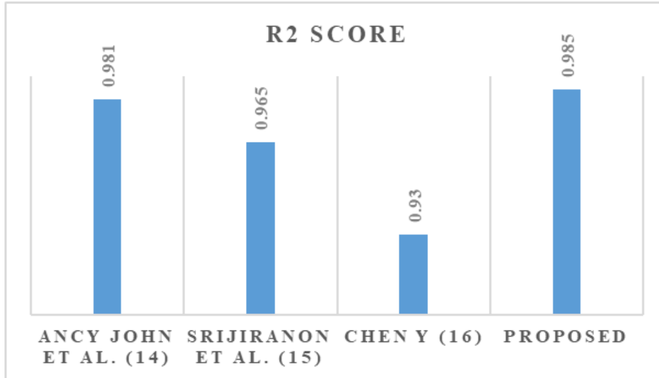

Fig 8. Comparative analysis of R^2 score for existing and the proposed model

4 Conclusion

Sentiment analysis functions as a key component in our predictive framework by assessing the emotional tone of news articles and headlines related to the stock market. This process involves cleaning and tokenizing textual data, assigning sentiment scores, and aggregating them to reflect the overall sentiment. Integrating this sentiment analysis into the stock market data of BSE Sensex enriches the dataset, allowing us to combine sentiment scores with traditional financial indicators. The proposed model introduced a groundbreaking approach to enhance stock market prediction by merging sentiment analysis and a hybrid RNN-LSTM architecture. We train the model to recognize patterns and relationships between the features and stock market movements. The observed values are compared with the Support Vector Regressor, Random Forest Regressor, and LSTM (without sentiment analysis) models. We get lower values of mean absolute error, mean squared error, and root mean square error. Results also show that the R^2 value for the proposed model is higher (0.40% to 5.5 %) when compared with literature methods. It is experiential that the proposed model enhances the prediction accuracy. In addition, this work signifies the continued evolution of predictive analytics in the financial realm, showcasing the potential of advanced techniques over traditional models. Future exploration could delve into refining sentiment analysis, integrating diverse data sources, and constructing interpretable models.

References

1) Agustin IN. The Integration of Fundamental and Technical Analysis in Predicting the Stock Price. *Jurnal Manajemen Maranatha*. 2019;18(2):93–102. Available from: https://doi.org/10.28932/jmm.v18i2.1611.
2) Agrawal M, Shukla PK, Nair R, Nayyar A, Masud M. Stock Prediction Based on Technical Indicators Using Deep Learning Model. *Computers, Materials & Continua*. 2022;70(1):287–304. Available from: https://dx.doi.org/10.32604/cmc.2022.014637.
3) Idrees SM, Alam MA, Agarwal P. A Prediction Approach for Stock Market Volatility Based on Time Series Data. *IEEE Access*. 2019;7:17287–17298. Available from: https://dx.doi.org/10.1109/access.2019.2895252.
4) Huang X, You P, Gao X, Cheng D. Stock Price Prediction Based on ARIMA-GARCH and LSTM. In: Proceedings of the 2nd International Academic Conference on Blockchain, Information Technology and Smart Finance (ICBIS 2023). Atlantis Press International BV. 2023;p. 438–448. Available from: https://doi.org/10.2991/978-94-6463-198-2_45.
5) Sen J, Mehtab S. A Time Series Analysis-Based Stock Price Prediction Using Machine Learning and Deep Learning Models. *International Journal of Business Forecasting and Marketing Intelligence*. 2020;6(4). Available from: https://dx.doi.org/10.1504/ijbfmi.2020.10038524.
6) Burgess M, Javed F, Okpara N, Robinson C. Stock Forecasts with LSTM and Web Sentiment. *Smu Scholar Data Science Review*. 2022;6(2):1–20. Available from: https://scholar.smu.edu/cgi/viewcontent.cgi?article=1219&context=datasciencereview.
7) Gaur Y. Stock Market Price Prediction Using LSTM. *International Journal for Research in Applied Science and Engineering Technology*. 2023;11(12):1881–1887. Available from: https://dx.doi.org/10.22214/ijraset.2023.57673.
8) Patel J, Patel M, Darji M. Stock Price Prediction Using RNN and LSTM. *Journal Of Emerging Technologies and Innovative Research (JETIR)*. 2018;5(11):1069–1079. Available from: https://www.jetir.org/papers/JETIRK006164.pdf.
9) Kasture P, Shirsath K. Recent Studies on Deep Learning Techniques for Stock Market Forecasting. In: 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS). IEEE. 2023. Available from: https://doi.org/10.1109/ICSCSS57650.2023.10169585.
10) Mitra M, Adhikary J, Nayek I, Baidya B, Ganguly K, Saha D. Stock Market Prediction Using Sentiment Analysis. *International Journal of Research Publication and Reviews*. 2023;4(6):2721–2729. Available from: https://ijrpr.com/uploads/V4ISSUE6/IJRPR14435.pdf.
11) Olawale IO, Iworiso J, Amaunam IA. Predicting Stock Investments Based on Sentiment and Historical Price Data. *International Journal of Data Envelopment Analysis and *Operations Research**. 2023;4(1):1–32. Available from: https://pubs.sciepub.com/ijdeaor/4/1/1/.
12) Ebrahimian K, Abbasi E, Tabriz AA, Mohammadzadeh A. Stock Price Forecasting With Support Vector Regression Based On. *International Journal Of Finance And Managerial Accounting*. 2023;8(29):53–64. Available from: http://www.ijfma.ir/article_21124.html.

13) Darapaneni N, Padur AR, Sharma H, Manjrekar M. Stock Price Prediction using Sentiment Analysis and Deep Learning for Indian Markets. 2022. Available from: https://doi.org/10.48550/arXiv.2204.05783.
14) John A, Latha T. Stock market prediction based on deep hybrid RNN model and sentiment analysis. *Automatika*. 2023;64(4):981–995. Available from: https://dx.doi.org/10.1080/00051144.2023.2217602.
15) Srijiranon K, Lertratanakham Y, Tanantong T. A Hybrid Framework Using PCA, EMD and LSTM Methods for Stock Market Price Prediction with Sentiment Analysis. *Applied Sciences*. 2022;12(21):1–23. Available from: https://dx.doi.org/10.3390/app122110823.
16) Chen Y. Stock Market Analysis and Prediction Using LSTM. *BCP Business & Management*. 2013;36:381–386. Available from: https://doi.org/10.54691/bcpbm.v36i.3489.