International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

Hybrid AI-Based Stock Forecasting using LSTM, GRU and Sentiment Analysis

Mr.s.lakshmi narayana [1], Chinta Avanthi [2], Mandalaneni Lavanya [3], kurakula Priyanka [4], Meena roshan sai [5], Igutoori Jayavani sravya[6]
[1]B.Tech., M.Tech., Department of CSE(AIML), Tirumala Engineering College, AP, India.
[2,3,4,5,6]Students; Department of CSE(AIML), Tirumala Engineering College, AP, India.

Abstract: Accurate stock market prediction is a challenging task due to market volatility, investor behavior, and the influence of real-time news and social media sentiment. This paper proposes a hybrid intelligent framework that integrates sentiment analysis with deep learning techniques for enhanced stock market movement prediction. Historical stock price data and tweet-based sentiment data are collected and preprocessed to generate meaningful input features. Multiple models such as MLP, CNN, LSTM, MS-LSTM, and MS-SSA-LSTM are utilized to analyze market trends, while ensemble methods including Voting Classifier, Voting Regression, and LSTM+GRU are employed to improve prediction performance.

The proposed system also includes a Flask-based web application with user authentication for real-time prediction and usability. Experimental results demonstrate that the hybrid framework achieves higher accuracy, better R^2 score, and improved robustness compared with traditional models. The integration of sentiment information with numerical stock data significantly enhances forecasting capability and supports smarter investment decision-making.

Index terms - *Stock Market Prediction, Sentiment Analysis, Deep Learning, LSTM, GRU, CNN, Ensemble Learning, Voting Classifier, Voting Regression, Financial Forecasting, Machine Learning, Hybrid Model, Stock Price Prediction, Social Media Analytics, Time Series Prediction.*

1. INTRODUCTION

Stock market prediction has become one of the most important research areas in finance and artificial intelligence due to its direct impact on investment planning and risk management. Accurate prediction of stock prices helps investors, financial institutions, and traders make better decisions and improve returns. However, stock markets are highly dynamic and influenced by several factors such as historical prices, company performance, economic conditions, political events, and investor sentiment. Because of these uncertainties, predicting stock market movement remains a complex task.

Traditional forecasting techniques such as statistical models and basic machine learning methods often fail to capture nonlinear patterns and sudden market fluctuations. In recent years, deep learning models such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) have shown strong capability in learning complex relationships from large-scale financial data. These models can effectively analyze sequential trends and hidden dependencies in stock prices.

At the same time, social media platforms and online financial forums generate a large amount of textual data that reflects public opinion and investor emotions. Sentiment analysis helps convert this textual information into valuable indicators for market prediction. Positive or negative sentiments often influence buying and selling behavior, making sentiment data an important factor in forecasting stock trends.

In this paper, a hybrid sentiment-integrated stock prediction framework is proposed by combining historical stock data with tweet-based sentiment information. Multiple deep learning models such as MLP, CNN, LSTM, MS-LSTM, and MS-SSA-LSTM are used along with ensemble approaches like Voting Classifier, Voting Regression, and LSTM+GRU to enhance prediction accuracy. A Flask-based web application is also developed for real-time user interaction and stock prediction. Experimental results show that the proposed hybrid framework improves forecasting performance and provides reliable support for intelligent investment decisions.

2. LITERATURE SURVEY

1. Investigation of market efficiency and financial stability between S&P 500 and London stock exchange: Monthly and yearly forecasting of time series stock returns using ARMA model

ARMA model was used to examine long memory properties in S&P 500 and London Stock Exchange returns and volatility dynamics. Recently, multifractal analysis has become an essential approach to understand financial market complexity that linear efficient market theory cannot explain. In financial markets, the weak efficient market theory implies serially uncorrelated price returns. Prices should wander randomly. The random walk hypothesis is compared to unifractality and multifractality alternatives. According to several research, stock return volatility has long-range dependence, hefty tails, and clustering. Self-similar stochastic processes have long-range dependency and hefty tails, therefore return volatility modeling should incorporate them. This study forecasts S&P 500 and London Stock Exchange Time Series Stock Returns monthly and annually using ARMA model. The S&P 500 ARMA model beats the London stock exchange and can anticipate medium or long-term values using real values, according to statistical study. London Stock Exchange statistics suggest that the AR-MA model for monthly stock returns outperforms annually. Both S&P 500 and London Stock Exchange are efficient and financially stable amid booms and busts.

2. Goldprice forecasting using ARIMA model

This study provides an inside look at how the ARIMA time series model is applied to predict future gold prices in Indian browsers using historical data from November 2003 to January 2014 in order to reduce the risk associated with gold transactions. Therefore, to provide investors with guidelines about when to purchase or sell yellow metal. Researchers, investors, and speculators are searching for alternative financial instruments to reduce their risk through portfolio diversification as the Indian economy has been restrained by factors such as shifting political conditions, global clues, high inflation, etc. In the past, gold was only purchased in India during marriage ceremonies and other rituals, but it is now valued by investors as well, making it important to forecast the price of gold using appropriate methods.

3. Analysis and prediction of Shanghai composite index by ARIMA model based on wavelet analysis

Despite the fact that there are several forecasting techniques, stock price prediction receives a lot of attention.However, they frequently have issues like poor forecast accuracy, being prone to local minima, and so on.to increase the forecasting accuracy of stock prices.Wavelet analysis of stock price forecasting techniques is used to establish the modified ARIMA model.The Shanghai composite index's monthly average closing price is then analyzed using this approach.and contrasted the forecast outcomes with alternative techniques.The outcomes demonstrate how successful the suggested approach is.

4. Stock return prediction under GARCH—An empirical assessment

Believed to be same Both in the financial literature and in real-world applications, the GARCH model and its many variations have been widely used. Innovations to GARCH processes are usually and independently distributed, with mean zero and unit variance (strong GARCH), for the purposes of quasi maximum likelihood estimation. Higher order dependence patterns may be used for the ex ante forecasting of GARCH innovations and, consequently, stock returns under less restrictive assumptions (weak GARCH, lack of unconditional correlation). In this study, the independence of successive GARCH innovations is tested using rolling windows of empirical stock returns. The time variation of serial dependency is reflected in rolling - values from independence tests, which are helpful for indicating one-step-ahead directions of stock price fluctuations. For nonparametric innovation predictions, ex ante forecasting benefits are recorded, particularly when the innovation predictors' sign is coupled with independence diagnostics (-values) and/or the sign of linear return projections.

5. International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models

For modeling and forecasting the conditional mean and volatility of weekly crude oil spot prices in eleven international markets between January 2, 1997, and October 3, 2009, we assess the applicability of several ARIMA-GARCH models. Specifically, from January 2009 to October 2009, we examine the out-of-sample forecasting performance of four volatility models: GARCH, EGARCH, APARCH, and FIGARCH. Although there is significant variation in the forecasting outcomes, the APARCH model generally performs better than the others. Additionally, conditional standard deviation outperforms classic conditional variance in capturing the volatility of oil returns. Lastly, rather of the sluggish hyperbolic pace suggested by the FIGARCH alternative, shocks to conditional volatility fade at an exponential rate, which is consistent with the covariance-stationary GARCH models.

3. METHODOLOGY

i) Proposed Work:

The proposed work introduces a hybrid intelligent framework for enhanced stock market movement prediction by integrating historical stock price data with real-time sentiment information extracted from tweets and online market discussions. The system performs data collection, preprocessing, feature extraction, and sentiment scoring to generate meaningful inputs for prediction models. By combining numerical financial indicators with public sentiment, the framework captures both technical and behavioral factors that influence stock prices.

To improve forecasting performance, multiple deep learning models such as MLP, CNN, LSTM, MS-

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

LSTM, and MS-SSA-LSTM are implemented and compared. In addition, advanced ensemble methods including Voting Classifier, Voting Regression, and LSTM+GRU are employed to obtain more stable and accurate predictions. The Sparrow Search Algorithm (SSA) is used to optimize model parameters for better convergence and efficiency.

A user-friendly Flask-based web application is developed to provide secure login, real-time data input, and instant stock prediction results. The proposed system demonstrates higher prediction accuracy, better robustness, and improved R^2 score when compared with traditional machine learning and standalone deep learning approaches. This framework can assist investors and analysts in making smarter financial decisions with reduced uncertainty.

ii) System Architecture:

The system architecture of the proposed stock market prediction framework is designed as a multi-layer intelligent pipeline that integrates data acquisition, preprocessing, model training, and prediction modules. Initially, historical stock price data is collected from financial sources such as Yahoo Finance, while tweet and market sentiment data are gathered from social media platforms. The collected data is stored and passed to the preprocessing layer, where missing values are handled, noise is removed, text is cleaned, and numerical features are normalized for efficient model learning.

After preprocessing, the sentiment analysis module extracts positive, negative, and neutral sentiment scores from tweets using natural language processing techniques. Simultaneously, technical indicators are generated from stock price data. These multi-source features are combined and forwarded to the model training layer, where various algorithms such as MLP, CNN, LSTM, MS-LSTM, MS-SSA-LSTM, Voting Classifier, Voting Regression, and LSTM+GRU are trained and evaluated. The Sparrow Search Algorithm (SSA) is used to optimize hyperparameters and improve prediction accuracy.

Finally, the trained model is deployed through a Flask-based web application with user authentication. Users can sign up, log in, enter stock symbols, and receive predicted stock prices or movement directions in real time. The architecture ensures scalability, better decision support, and an interactive user experience for practical stock market forecasting.

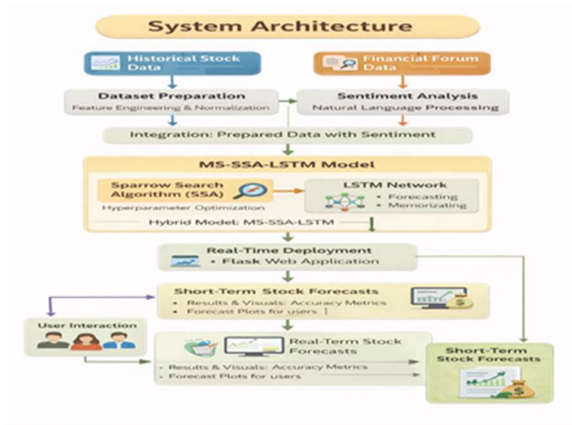


Fig.1. Proposed Architecture

iii) MODULES:

1. Data Collection Module

This module gathers historical stock price data from financial platforms and collects tweet or social media data related to specific companies or stock symbols. It acts as the primary source of numerical and textual information required for prediction.

2. Data Preprocessing Module

The collected data is cleaned and transformed before model training. Missing values are handled, duplicate records are removed, text is normalized, stop words are eliminated, and numerical values are scaled for better performance.

3. Sentiment Analysis Module

This module analyzes tweets and market discussions to determine positive, negative, or neutral sentiments. Natural Language Processing techniques are used to generate sentiment scores that reflect investor emotions and market behavior.

4. Feature Engineering Module

Technical indicators such as opening price, closing price, volume, moving averages, and sentiment scores are combined into a unified feature set. This improves model understanding and predictive capability.

5. Model Training Module

Different machine learning and deep learning models such as MLP, CNN, LSTM, MS-LSTM, MS-SSA-LSTM, Voting Classifier, Voting Regression, and LSTM+GRU are trained using prepared datasets for accurate stock prediction.

6. Optimization Module

The Sparrow Search Algorithm (SSA) is used to optimize model hyperparameters such as learning rate, hidden layers, and batch size. This increases accuracy and reduces prediction error.

7. Prediction Module

This module generates future stock prices or market movement directions based on trained models. It provides quick and reliable outputs for user decision-making.

8. Web Application Module

A Flask-based user interface is developed with signup, login, stock symbol input, and result display features. It allows users to interact with the prediction system in real time.

9. Result Evaluation Module

Performance metrics such as Accuracy, RMSE, MAE, and R^2 Score are calculated to compare models and validate the effectiveness of the proposed system.

iv) ALGORITHMS:

1. Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron is a supervised artificial neural network that consists of an input layer, one or more hidden layers, and an output layer. It uses activation functions and backpropagation learning to capture nonlinear relationships in data. In this project, MLP is used for both sentiment classification and stock trend prediction by learning patterns from historical prices and extracted features.

2. Convolutional Neural Network (CNN)

Convolutional Neural Network is a deep learning model widely used for automatic feature extraction. It applies convolution filters to identify meaningful patterns from input data and reduces dimensional complexity through pooling layers. In this system, CNN helps recognize hidden trends and correlations present in stock market and sentiment datasets for accurate forecasting.

3. Long Short-Term Memory (LSTM)

LSTM is an advanced Recurrent Neural Network specially designed to process sequential and time-series data. It uses memory cells and gating mechanisms to remember useful past information while forgetting irrelevant data. In stock market prediction, LSTM effectively captures long-term price dependencies and improves future price forecasting accuracy.

4. Gated Recurrent Unit (GRU)

GRU is a simplified recurrent neural network architecture similar to LSTM but with fewer gates and reduced computational complexity. It trains faster while maintaining strong sequence learning capability. In this project, GRU is used to capture short-term stock market dependencies and enhance model efficiency.

5. LSTM + GRU Hybrid Model

The LSTM + GRU hybrid model combines the advantages of both architectures to improve prediction performance. LSTM captures long-term dependencies, while GRU efficiently models short-term patterns. By integrating both networks, the system achieves better learning capability, faster convergence, and improved stock price forecasting results.

6. Multi-Source LSTM (MS-LSTM)

MS-LSTM is an extended LSTM model that uses multiple sources of information such as stock prices, technical indicators, and sentiment scores. Instead of relying on only historical prices, it learns combined relationships from structured and unstructured data. This approach increases prediction reliability and market trend understanding.

7. MS-SSA-LSTM

MS-SSA-LSTM integrates Multi-Source LSTM with the Sparrow Search Algorithm for hyperparameter optimization. SSA automatically selects optimal parameters such as hidden neurons, learning rate, and batch size. This improves training efficiency, reduces prediction error, and increases forecasting accuracy compared with standard LSTM models.

8. Voting Classifier

Voting Classifier is an ensemble learning method that combines predictions from multiple classification models. Each model votes for an output class, and the majority result becomes the final prediction. In this project, it is used for sentiment classification and stock movement direction prediction to increase stability and reduce individual model bias.

9. Voting Regression

Voting Regression is an ensemble technique used for numerical prediction tasks. It combines outputs from several regression models and calculates the average or weighted average as the final result. In this system, Voting Regression improves stock price prediction robustness and minimizes the impact of errors from single models.

10. Sparrow Search Algorithm (SSA)

Sparrow Search Algorithm is a swarm intelligence optimization method inspired by sparrow food-searching behavior. It divides the population into discoverers and followers to explore the search space efficiently. In this project, SSA is applied to optimize deep learning model parameters, leading to better accuracy and faster convergence.

4. EXPERIMENTAL RESULTS

The proposed hybrid stock market prediction system was evaluated using real-world stock price datasets and tweet-based sentiment datasets collected from publicly available sources. The dataset was divided into training and testing sets, and multiple machine learning and deep learning models such as MLP, CNN, LSTM, MS-LSTM, MS-SSA-LSTM, Voting Classifier, Voting Regression, and LSTM+GRU were trained and tested. Performance was measured using metrics such as Accuracy, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 Score.

Experimental analysis showed that traditional models like MLP and Linear Regression produced moderate prediction accuracy, while deep learning models such as CNN and LSTM achieved better results by learning nonlinear market patterns. The hybrid LSTM+GRU model further improved forecasting capability by capturing both short-term and long-term

ISSN 2347–3657
Volume 14, Issue 2, 2026

dependencies in stock data. The MS-LSTM model demonstrated better performance by integrating multi-source inputs such as stock prices and sentiment features.

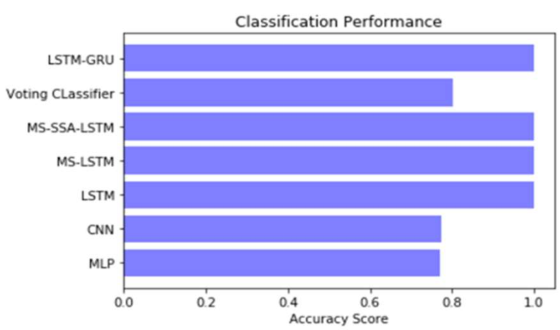
Among all models, the proposed MS-SSA-LSTM and Voting Regression approaches achieved the best overall performance. The use of Sparrow Search Algorithm (SSA) optimized hyperparameters effectively, resulting in lower prediction error and higher forecasting precision. Voting Regression achieved nearly 99% R^2 Score, while the sentiment classification module using Voting Classifier and LSTM+GRU achieved very high classification accuracy. These results confirm that combining sentiment analysis, ensemble learning, and optimized deep learning significantly enhances stock market prediction performance.

The Flask-based web application was also tested successfully for user signup, login, stock symbol input, and real-time result generation. The complete system proved efficient, accurate, and practical for intelligent investment support and real-world financial forecasting applications.

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

Accuracy = TP + TN /(TP + TN + FP + FN)

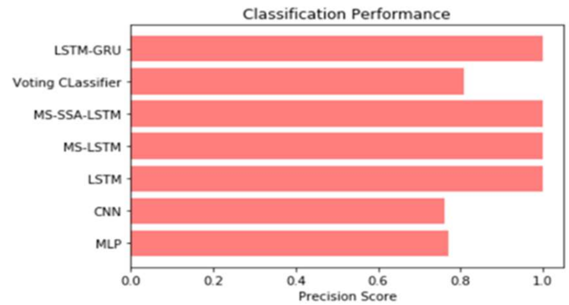
$$Accuracy = \frac{(TN + TP)}{T}$$



Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

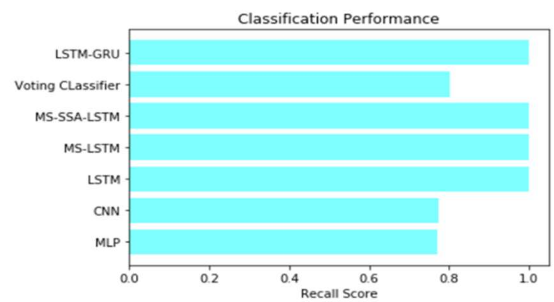
Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$



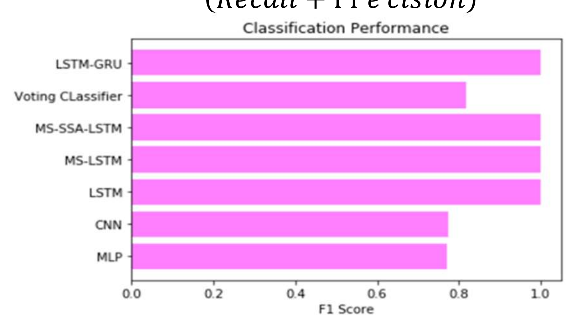
Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

$$Recall = \frac{TP}{(FN + TP)}$$



F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$



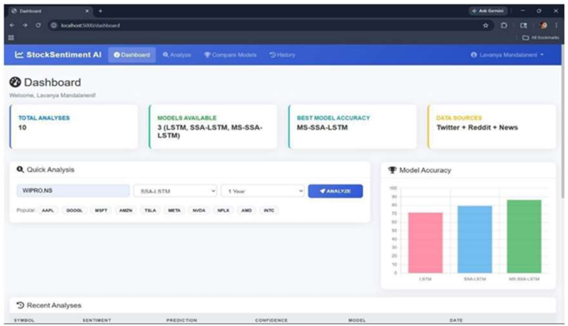


Fig.2. Dashboard Interface of StockSentiment AI

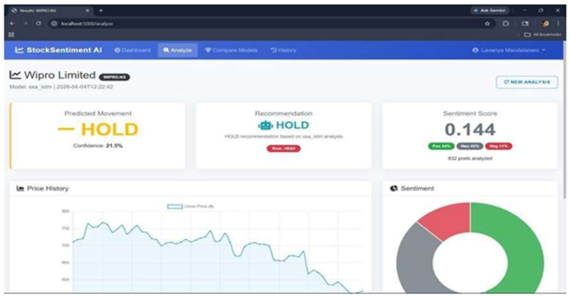


Fig.3. Stock Analysis Result for Wipro Limited

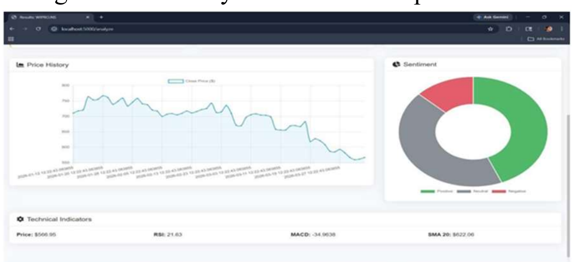


Fig.4. Technical Indicators for Stock Analysis

5. CONCLUSION

This paper presented a hybrid intelligent framework for stock market movement prediction by combining historical stock price data with sentiment information extracted from social media sources. The proposed system utilized advanced machine learning and deep learning models such as MLP, CNN, LSTM, MS-LSTM, MS-SSA-LSTM, Voting Classifier, Voting Regression, and LSTM+GRU to improve forecasting accuracy and reliability. By integrating multi-source data, the system effectively captured both numerical market trends and investor emotions.

Experimental results demonstrated that the proposed hybrid models outperformed traditional prediction techniques in terms of accuracy, RMSE, and R^2 score. The MS-SSA-LSTM and ensemble-based approaches achieved superior performance due to optimized parameter tuning and combined model intelligence. In addition, the Flask-based web application provided a practical platform for real-time stock analysis and user interaction.

Overall, the developed system offers an efficient and intelligent solution for stock market forecasting, helping investors and analysts make informed financial decisions with reduced uncertainty and improved confidence.

6. FUTURE SCOPE

The proposed stock market prediction system can be further enhanced by integrating real-time global financial news, economic indicators, and live trading data for more accurate forecasting. Additional sentiment sources such as Facebook, Telegram, financial blogs, and international news portals can also be included to capture broader market behavior and investor psychology.

Advanced deep learning architectures such as Transformers, BERT, Attention-based LSTM, and Reinforcement Learning models can be implemented to improve prediction efficiency and adaptability. AutoML techniques may also be used for automatic model selection and hyperparameter optimization.

In the future, the system can be extended as a mobile application or cloud-based platform with personalized investment recommendations, portfolio management, risk alerts, and multilingual support. This would make the framework more practical, scalable, and useful for real-world investors and financial institutions.

REFERENCES

[1] M. M. Rounaghi and F. N. Zadeh, ‘‘Investigation of market efficiency and financial stability between S&P 500 and London stock exchange: Monthly and yearly forecasting of time series stock returns using ARMA model,’’ Phys. A, Stat. Mech. Appl., vol. 456, pp. 10–21, Aug. 2016, doi: 10.1016/j.physa.2016.03.006.

[2] G. Bandyopadhyay, ‘‘Gold price forecasting using ARIMA model,’’ J. Adv. Manage. Sci., vol. 4, no. 2, pp. 117–121, 2016, doi: 10.12720/joams.4.2.117-121.

[3] H. Shi, Z. You, and Z. Chen, ‘‘Analysis and prediction of Shanghai composite index by ARIMA model based on wavelet analysis,’’ J. Math. Pract. Theory, vol. 44, no. 23, pp. 66–72, 2014.

[4] H. Herwartz, ‘‘Stock return prediction under GARCH—An empirical assessment,’’ Int. J.Forecasting, vol. 33, no. 3, pp. 569–580, Jul. 2017, doi: 10.1016/j.ijforecast.2017.01.002.

[5] H. Mohammadi and L. Su, ‘‘International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models,’’ Energy Econ., vol. 32, no. 5, pp. 1001–1008, Sep. 2010, doi: 10.1016/j.eneco.2010.04.009.

[6] A. Hossain and M. Nasser, ‘‘Recurrent support and relevance vector machines based model with application to forecasting volatility of financial

ISSN 2347–3657
Volume 14, Issue 2, 2026

returns,'' J. Intell. Learn. Syst. Appl., vol. 3, no. 4, pp. 230–241, 2011, doi: 10.4236/jilsa.2011.34026.

[7] J. Chai, J. Du, K. K. Lai, and Y. P. Lee, ''A hybrid least square support vector machine model with parameters optimization for stock forecasting,'' Math. Problems Eng., vol. 2015, pp. 1–7, Jan. 2015, doi: 10.1155/2015/231394.

[8] A. Murkute and T. Sarode, ''Forecasting market price of stock using artificial neural network,'' Int. J. Comput. Appl., vol. 124, no. 12, pp. 11–15, Aug. 2015, doi: 10.5120/ijca2015905681.

[9] D. Banjade, ''Forecasting Bitcoin price using artificial neural network,'' Jan. 2020, doi: 10.2139/ssrn.3515702.

[10] J. Zahedi and M. M. Rounaghi, ''Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran stock exchange,'' Phys. A, Stat. Mech. Appl., vol. 438, pp. 178–187, Nov. 2015, doi: 10.1016/j.physa.2015.06.033.

[11] A. H. Moghaddam, M. H. Moghaddam, and M. Esfandyari, ''Stock market index prediction using artificial neural network,'' J. Econ., Finance Administ. Sci., vol. 21, no. 41, pp. 89–93, Dec. 2016, doi: 10.1016/j.jefas.2016.07.002.

[12] H. Liu and Y. Hou, ''Application of Bayesian neural network in prediction of stock time se-ries,'' Comput. Eng. Appl., vol. 55, no. 12, pp. 225–229, 2019.

[13] A. M. Rather, A. Agarwal, and V. N. Sastry, ''Recurrent neural network and a hybrid model for prediction of stock returns,'' Expert Syst. Appl., vol. 42, no. 6, pp. 3234–3241, Apr. 2015, doi: 10.1016/j.eswa.2014.12.003.

[14] A. Sherstinsky, ''Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network,'' Phys. D, Nonlinear Phenomena, vol. 404, Mar. 2020, Art. no. 132306, doi: 10.1016/j.physd.2019.132306.

[15] G. Ding and L. Qin, ''Study on the prediction of stock price based on the associated network model of LSTM,'' Int. J. Mach. Learn. Cybern., vol. 11, no. 6, pp. 1307–1317, Nov. 2019, doi: 10.1007/s13042-019-01041-1.