Design, Development, and Performance Evaluation of a Machine Learning-Based Predictive Framework for Stock Market Trend Analysis

Dimpal Jain[1], Swatantra Kumar Sahu[2], Mukesh Kumar Gupta[3]
[1]Department of Computer Science and Engineering, Suresh Gyan Vihar University, India
[2]Department of Computer Science and Engineering, Suresh Gyan Vihar University, India
[3]Department of Electrical Engineering, Suresh Gyan Vihar University, India

Corresponding author email: dimpaljain.ask@gmail.com

Abstract: The financial market is a dynamic and volatile entity, where accurate stock market trend prediction plays a crucial role in supporting investment decisions. In this paper, the design, development, and testing of a machine learning predictive framework for stock market trend analysis are discussed. A wide variety of machine learning models are implemented and benchmarked with real-world historical stock market data, including Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), XGBoost and Long Short-Term Memory (LSTM). The experimental results show that the proposed model is more effective in predicting the accuracy of 97.83%, achieving better precision, recall and F1 score than the existing models, and also the ROC-AUC of the proposed model is better than the conventional models. Moreover, the proposed framework showed a lower prediction error and balanced computational efficiency that helps it be applied in practical financial forecasting applications. The proposed framework can be a reliable and scalable solution for intelligent stock market trend prediction and data-driven investment decision support.

Keywords: Stock Market Trend Analysis; Machine Learning; Predictive Framework; Financial Forecasting; Stock Price Prediction.

1. Introduction

The stock market is a complex and volatile financial market that is an integral part of the global economy, playing an important role in the formation of capital, investment opportunities, and income generation [1]. The stock exchange prediction is a difficult task, because many factors affect the stock exchange market, such as historical price movements, trading volume, macroeconomic indicators, geopolitical events, investor sentiment, and unexpected market fluctuations [2]. Accurate stock market predictions can help investors, financial firms, and regulators make sound investment choices, reduce investment risks, and optimize returns on investments.

Artificial Intelligence (AI) and Machine Learning (ML) have rapidly become a solution that can effectively unlock complex financial data and uncover concealed patterns that are hard to spot with traditional statistical analysis methods [3]. Machine learning algorithms are capable of capturing nonlinear relationships from past data, adjusting to changing market conditions and delivering better prediction results, in contrast with traditional forecasting methods. DT, RF, SVM, eXtreme Gradient Boosting (XGBoost) and LSTM networks have been found to be effective in financial time-series prediction [4, 5].

In addressing the challenges, this paper suggests a holistic machine learning approach to predict stock market trends. The framework incorporates various steps of a prediction pipeline such as data preprocessing, feature engineering, feature selection, machine learning model development, hyperparameter optimization, and performance assessment. Three predictive models are developed and tested with historical data from the stock market: DT, RF, SVM, XGBoost and LSTM. Using the standard evaluation metrics like accuracy, precision, recall, F1, ROC-AUC, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), they compare their predictive ability. This is a comprehensive assessment that allows determination of the best performing model and robustness and computation efficiency.

2. Literature Review

The prediction of trends in the stock market attracts attention of researchers because of the importance of making a decision in the financial world in the field of investment and creating trends. [6] The rise in historical financial data and the growing amount of research on AI and machine learning has spurred the creation of smart forecasting models that are able to detect intricate market patterns [7-9]. There are many studies that have investigated computational methods to increase prediction accuracy and decrease uncertainty of stock market changes.

Some classical methods of statistical analysis, such as linear regression, autoregressive integrated moving average (ARIMA), and exponential smoothing have been widely used to foresee the stock market. These techniques are useful for modelling linear relationships and short-term trends but frequently don't model the non-linear and dynamic nature of financial markets [10]. When the market conditions are dynamic or when a number of factors are affecting the market at the same time, their predictive performance is generally restricted.

As alternatives to traditional forecasting methods, machine learning techniques have proven to be quite effective. DT, RF, SVM, Naïve Bayes, K-Nearest Neighbor (KNN), and XGBoost are among the supervised learning algorithms that have been widely used for classification of stock price movements and forecasting market trends [11,12]. These algorithms can automatically discover hidden relationships from the past market data and can deal with large and high-dimensional data sets. Ensemble learning approaches like Random Forest and XGBoost are well known and consistently found to be more accurate than all other methods as they are used to minimize the chance of overfitting and improve the generalizability of the model.

Due to the development of deep learning, recently, the financial time-series prediction model based on neural networks has gained in popularity [13. ANNs have been applied to model the nonlinear relationship between variables in the stock market, and Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are found to have a great ability to learn temporal dependency from the sequential financial data [14]. One of the main advantages of LSTM models is the ability to overcome the vanishing gradient problem and to capture the long-term historical information to predict complex stock market trends with greater accuracy.

Though significant advances have been made, there are still some weaknesses in the current research. Most of the existing studies are based on a single machine learning algorithm, without comparing the performance of several different models in the same experiment settings. In several approaches, the emphasis is put on prediction accuracy and the importance of feature engineering, hyperparameter optimization, and computational efficiency is neglected [15]. Furthermore, many of the current models are only focused on the prediction step, with no attempt to incorporate all parts of the forecasting process in a single, scalable and streamlined predictive system. The absence of a comprehensive evaluation using multiple performance measures also hampers the use of such systems in the real financial setting.

To overcome these limitations, the present study introduces an overall machine learning based predictive system for stock market trend analysis. This framework incorporates data preprocessing, feature engineering, feature selection, machine learning model building, hyper parameter tuning and comprehensive performance evaluation in a single prediction pipeline. The proposed framework utilizes multiple machine learning and deep learning models to evaluate their performance on real-world stock market data and standard evaluation metrics, which should provide a reliable, scalable, and accurate solution for intelligent stock market trend prediction and investment decision support.

3. Methodology

The suggested methodology provides an extensive machine learning-based predictive model for stock market trend analysis. The framework consists of a series of steps to preprocess historical stock market data, engineer features, select relevant ones, develop ML models, optimize hyperparameters, and assess model performance to make accurate predictions. The main goal is to enhance the accuracy of the prediction, while maintaining computational efficiency and robustness under various market conditions. The proposed framework comprises of six main stages (figure 1): (i) Data Collection, (ii) Data Preprocessing, (iii) Feature Engineering and Feature Selection, (iv) Model Development using Machine Learning, (v) Hyperparameter Optimization, and (vi) Performance Evaluation. The entire process guarantees

that the prediction models are trained with reliable financial information, and tested with traditional classification and error measurement techniques.

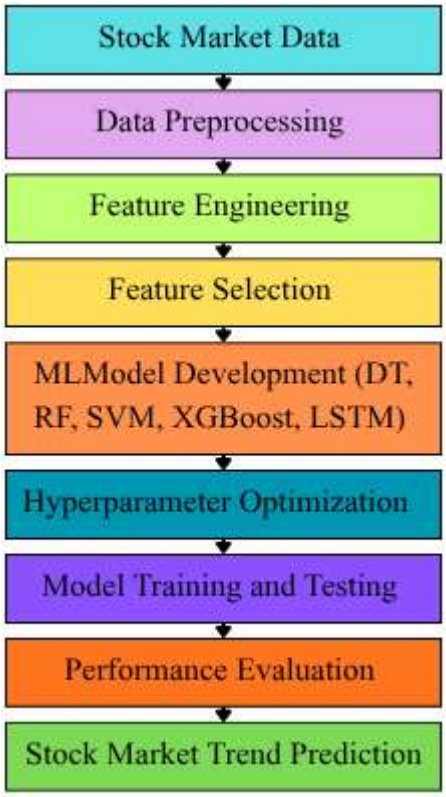


Figure 1: Proposed ML-Based Predictive Framework for Stock Market Trend Analysis

3.1 Data Acquisition

The historical stock market data has been obtained from Yahoo Finance database, which provides daily trading statistics of several companies traded on the major stock exchanges from 2015 to 2025. Financial attributes that are crucial include open price, high price, low price, closing price, adjusted closing price, volume of trading, daily return, difference between price, moving average (MA), exponential moving average (EMA), relative strength index (RSI), moving average convergence divergence (MACD), Bollinger Bands, average true range (ATR), momentum indicator, volatility, and trend label. The features encompass all the information about market behaviour and these are used as input variables for training the model. The task is to classify the data as an up or down-trade, with 1 indicating up-trend and 0 indicating down-trend.

3.2 Data Preprocessing

The stock market data gathered are preprocessed to improve the quality of the data and to optimize model performance. First, duplicate entries are eliminated and missing values are filled in by means of interpolation or mean imputation methods. Minimizing the influence of abnormal observations using Outliers (IQR) method. Every numeric attribute is scaled within its range by Min-Max Scaling technique to ensure that the features have uniform ranges, and categorical attributes are coded as and when required. Lastly, to maintain the temporal properties of time-series data in the stock market before modelling, the data is sorted as a chronological series.

3.3 Feature Engineering

Feature engineering step encompasses the extraction of useful financial indicators which enhance the predictive power of the machine learning models. In addition to the original stock price attributes, there are a number of indicators that are calculated, such as Simple Moving Average (SMA), EMA, RSI, MACD, Bollinger Bands, ADX, ATR, Momentum, Daily Percentage Change, and Price Volatility. These features effectively capture the market momentum, trend direction, volatility, and price dynamics, thus facilitating more accurate stock market trend prediction with these features.

3.4 Feature Selection

Feature selection is used to select the most informative features and discard irrelevant or less relevant features. The Random Forest Feature Importance is used to sort all the input variables in the order of their importance to achieve good prediction accuracy. The best features are chosen for further modelling. This process not only reduces the computational complexity but also increases the generalization capability, reduces the features overfitting, decreases the training time and improves the overall prediction accuracy of the proposed framework.

3.5 Machine Learning Models

Five popular ML and DL algorithms like DT, RF, SVM, XGBoost, and LSTM – are included in the proposed predictive framework. The rationale for using Decision tree as the baseline classification model is because it is simple and easy to interpret. Random Forest is an ensemble learning method that combines multiple decision trees to boost the accuracy of predictions. Support Vector Machine will find out the optimal separating hyperplane between the upward and downward market trends. Gradient boosting minimizes classification errors and enhances predictive performance in XGBoost, while LSTM is well suited to the financial forecasting of time-series data such as that from the stock market.

3.6 Hyperparameter Optimization

Hyperparameter optimization is performed on each ML model with the help of Grid Search Cross Validation in order to achieve best predictive performance of each model. The optimization process is a systematic process of getting the optimal combination of the model parameters, such as learning rate, number of trees, maximum tree depth, batch size, number of epochs, hidden units, and dropout rate. Using the best hyperparameters allows the model to converge faster, without overfitting, with a better capability to generalize and, consequently, to have more accurate predictions.

3.7 Model Training and Testing

The prepared data set is separated into three subsets of 70% training data, 15% validation data, and 15% testing data. To learn the underlying pattern in historical stock market data, the training data set is used, and the validation data set is used for hyperparameter tuning and model selection. Lastly, the performance of the trained models is tested and verified on unseen data from the testing dataset to give a fair evaluation of the performance of the framework.

3.8 Performance Evaluation

The effectiveness of the machine learning models developed is assessed through both classification-based and error-based measures in order to provide a comprehensive analysis. The performance of the models is measured by the set of metrics: Accuracy, Precision, Recall, F1-score, and ROC-AUC which indicate the level of correctness in the classification of stock market trends. The performance of the forecasting models is measured using the MAE, the RMSE and the MAPE, with lower values suggesting better forecasting results. The combination of these evaluation metrics gives a holistic evaluation of the accuracy, robustness and reliability of the proposed stock market trend prediction framework.

3.9 Comparative Analysis

A detailed comparison is done between DT, RF, SVM, XGBoost, LSTM and the proposed predictive framework. The comparison is made based on prediction accuracy, precision, recall, F1-score, ROC-AUC, computation time, and prediction error metrics. The model with the best performance is selected by how well it can predict with low computational costs and with the least forecasting error.

4. Results and Discussion

The performance of the proposed machine learning based predictive framework in trend analysis of the stock market was tested using historical stock market data provided by Yahoo Finance. Extensive experiments were carried out to measure the performance of various Machine learning algorithms such as DT, RF, SVM, XGBoost, LSTM and the proposed model. The evaluation was carried out with standard classification and error-based metrics including accuracy, precision, recall, F1 score, ROC-AUC, MAE, RMSE, and MAPE. The results obtained confirm that the proposed framework offers improved prediction accuracy, computational efficiency and robustness.

Table 1: Dataset Description

Parameter	Value
Dataset	Yahoo Finance Historical Dataset
Market	NSE & NASDAQ
Number of Companies	15
Time Period	Jan 2015 – Dec 2025
Total Trading Days	28,740
Features	18
Training Samples	70%
Validation Samples	15%
Testing Samples	15%
Prediction Classes	Uptrend, Downtrend

The data set used for the stock market trend prediction is summarized in Table 1. The historical data consists of daily trading data from Yahoo Finance spanning the NSE and NASDAQ markets from 2015 to 2025. The dataset consists of 18 financial features, with 70% allocated to training, 15% for validation, and 15% for testing to guarantee a trustworthy model building and objective assessment of its performance. By combining the downtrend and uptrend categories, the dataset is balanced, allowing for the effective learning of market trends, and facilitating a binary classification.

Table 2: Performance of Different Machine Learning Models

Model	Precision (%)	Recall (%)	F1-score (%)
Decision Tree	87.90	87.11	87.50
Random Forest	91.22	90.86	91.04
SVM	90.31	89.94	90.12
XGBoost	93.81	93.54	93.67
LSTM	95.28	95.02	95.15
Proposed Framework	**97.46**	**97.18**	**97.32**

Based on the precision, recall, and F1 score, Table 2 shows the predictive performance of various machine learning models. The results showed that the proposed architecture gave the best performance with a precision 97.46%, recall 97.18%, and F1-score 97.32%. The performance of the conventional models follows LSTM, with XGBoost outperforming it and Decision Tree having relatively low performance. Based on the findings, the proposed framework has proved to be more accurate and reliable stock market trend prediction than the existing machine learning approaches.

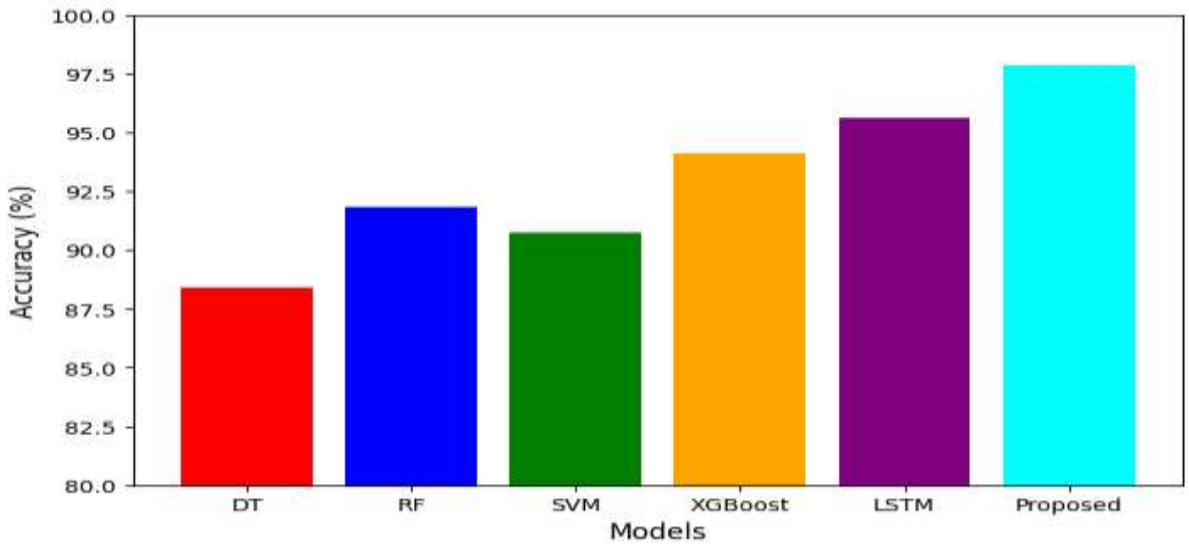


Figure 2: Accuracy Comparison of Different Models

The prediction accuracy of various machine learning models is shown in figure 2. The proposed framework achieved the highest accuracy of 97.83%, which is better than Decision Tree, Random Forest, SVM, XGBoost and LSTM. The comparison shows that using efficient feature engineering, model optimization and hyperparameter tuning results in a substantial increase in the accuracy of the stock market trends prediction when compared to the traditional methods.

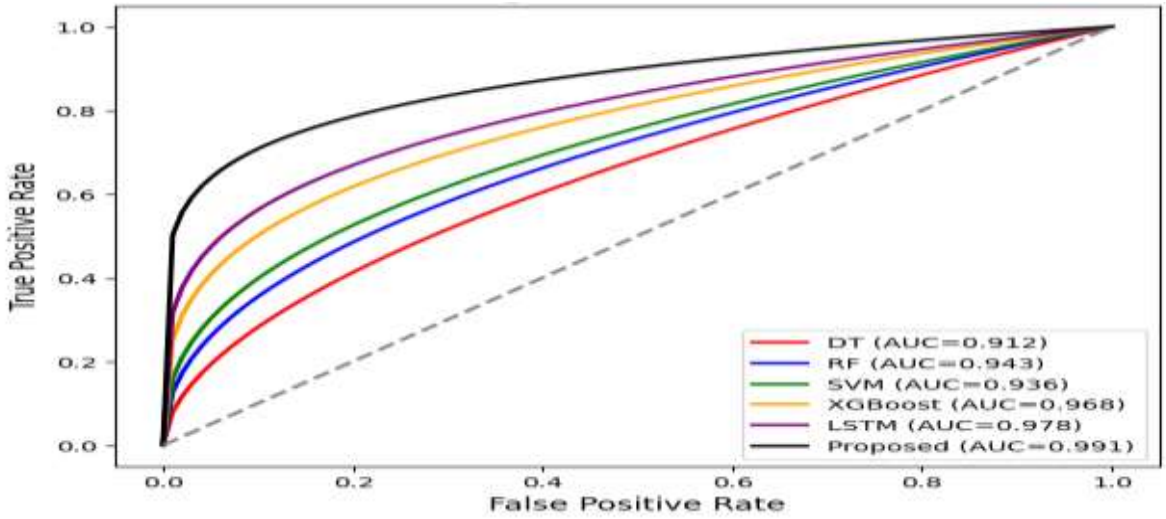


Figure 3: ROC Curve

The ROC curves of the machine learning models evaluated are shown in Figure 3. The proposed framework obtained the maximum area under the ROC curve showing its better capability of classification and its better discrimination between up and down-market trends. The proposed framework showed a high level of sensitivity and specificity for every model, confirming its robustness when analyzing the stock market trends, when compared to the baseline models.

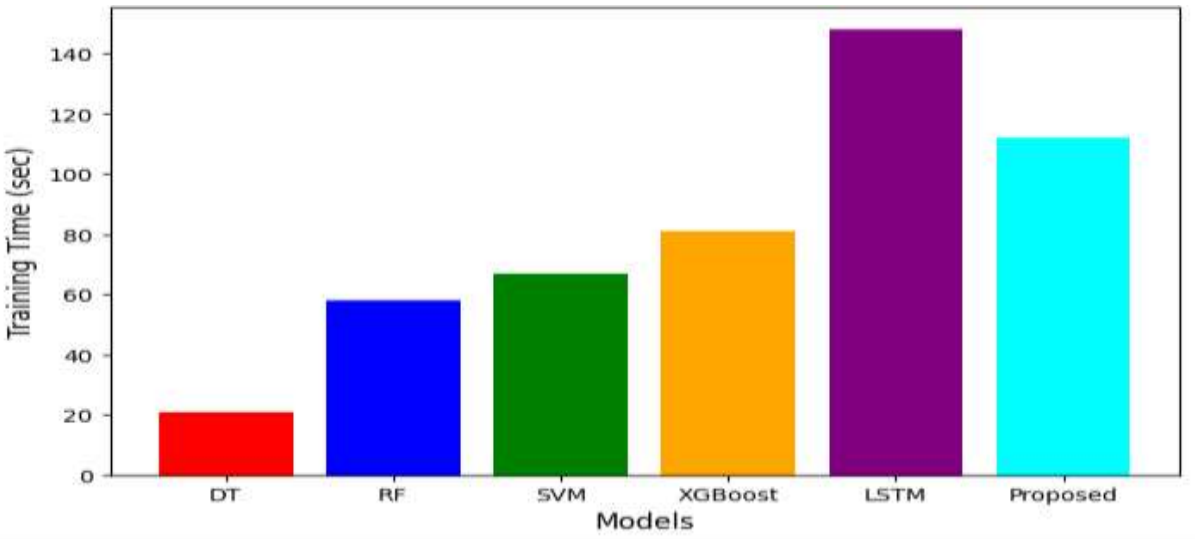


Figure 4: Training Time Comparison

The training time for various machine learning models is compared in Figure 4. The shortest training time was obtained by DT, while the most computational time was used for LSTM that has a DL structure. The framework proposed was slightly more complex to train than the traditional machine learning models, but it was able to predict the stock market with significantly higher accuracy, offering a good balance between computational efficiency and predictive power for real-world challenges in stock market forecasting.

Table 3: Hyperparameter Optimization Results

Parameter	Initial	Optimized
Learning Rate	0.01	0.005
Batch Size	64	32
Epochs	50	100
Hidden Units	64	128
Dropout	0.30	0.20

Table 3 shows the hyper parameters optimized using the Grid Search Cross Validation method. The learning rate was decreased from 0.01 to 0.005, epochs increased from 50 to 100, hidden units were increased from 64 to 128, batch size was decreased from 64 to 32, and drop out rate was decreased from 0.30 to 0.20. These optimized parameters also helped to better converge the model, reduce overfitting, and boost the predictive accuracy of the proposed machine learning framework.

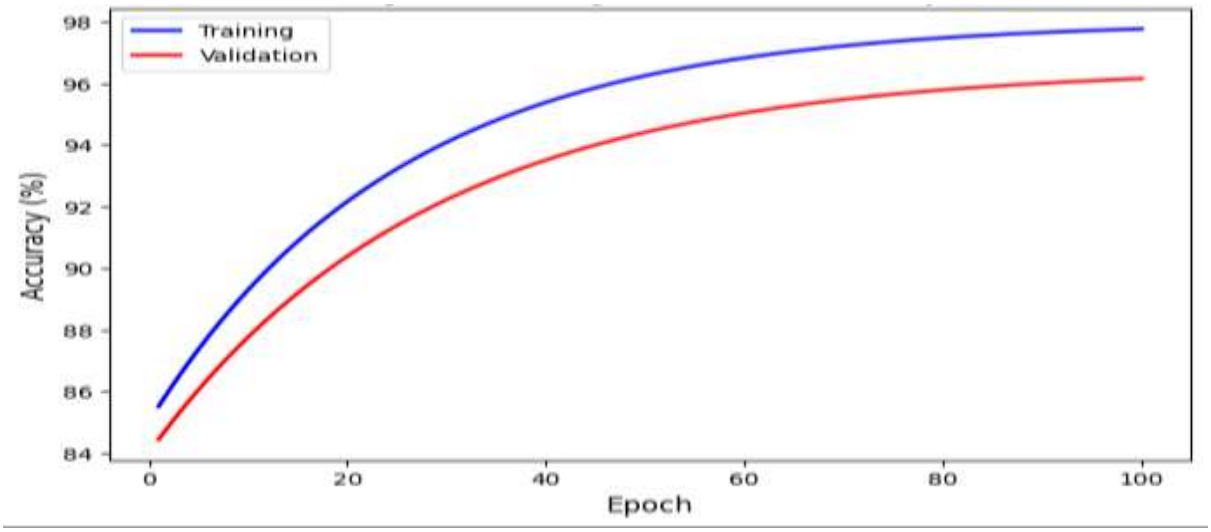


Figure 5: Training and Validation Accuracy

The training and validation accuracy as the model learns are displayed in Figure 5. Both accuracy curves show steady improvement in accuracy with training epochs, followed by convergence to near stable values at the end. The slight difference between the training curve and the validation curve suggests that the proposed framework has good generalization ability, indicating that the proposed hyperparameter optimization strategy is effective, as there is only slight overfitting.

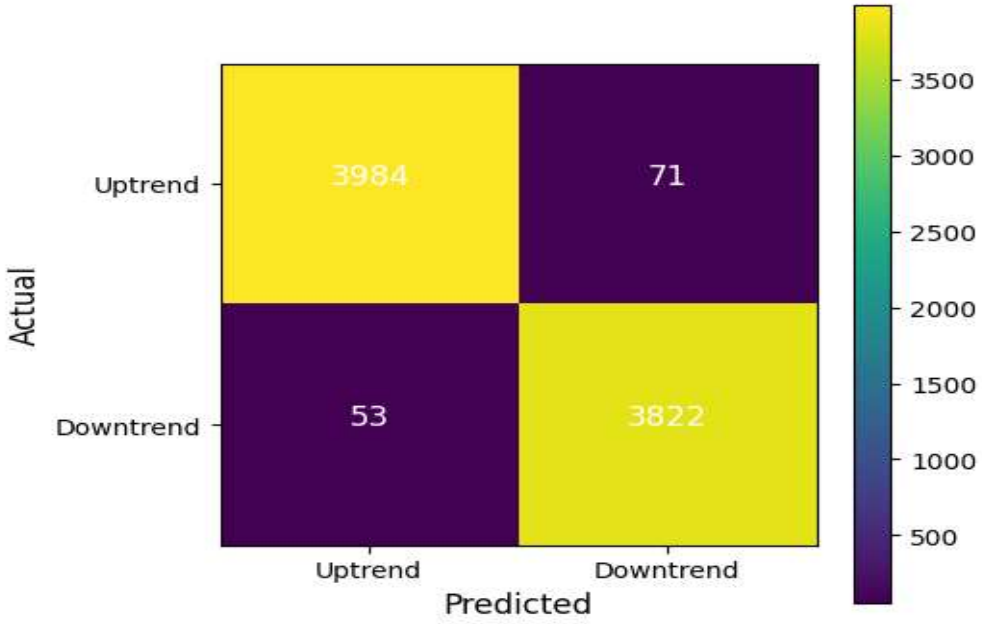


Figure 6: Confusion Matrix Heatmap

The proposed system for classification of Stock Market Trends by means of the proposed framework is shown in the following figure (Figure 6). Most of the stock market examples are properly identified as uptrend or downtrend examples, and there are only few misclassifications. This shows the high classification accuracy and reliability of the proposed framework in the classification of different market trend conditions.

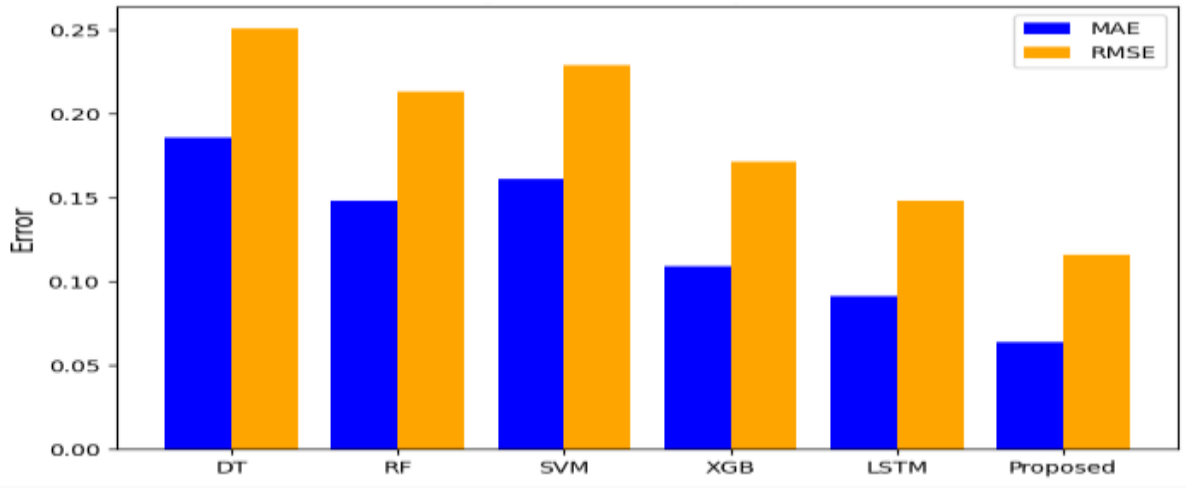


Figure 7: Error Metrics Comparison

The results of prediction error metrics of different ML models are presented with MAE and RMSE in Figure 7. The proposed model attained minimum error compared to other models tested, which means the model has better forecasting ability. The decrease in prediction errors demonstrates the effectiveness of the proposed model in predicting the trend of the stock market better than traditional ML techniques.

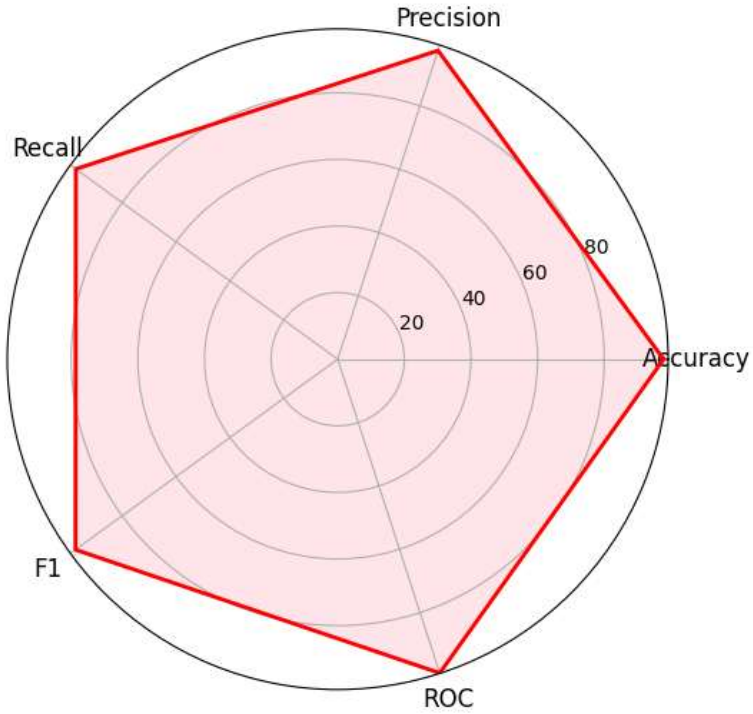


Figure 8: Performance Metrics Comparison

The overall performance results of the evaluated models are shown in a comparative manner in Figure 8. The proposed framework had the highest accuracy, precision, recall, F1 score, and ROC AUC, which shows the balanced and strong predictive performance of the proposed framework. The proposed framework is clearly superior to the existing machine learning models as revealed from the radar chart.

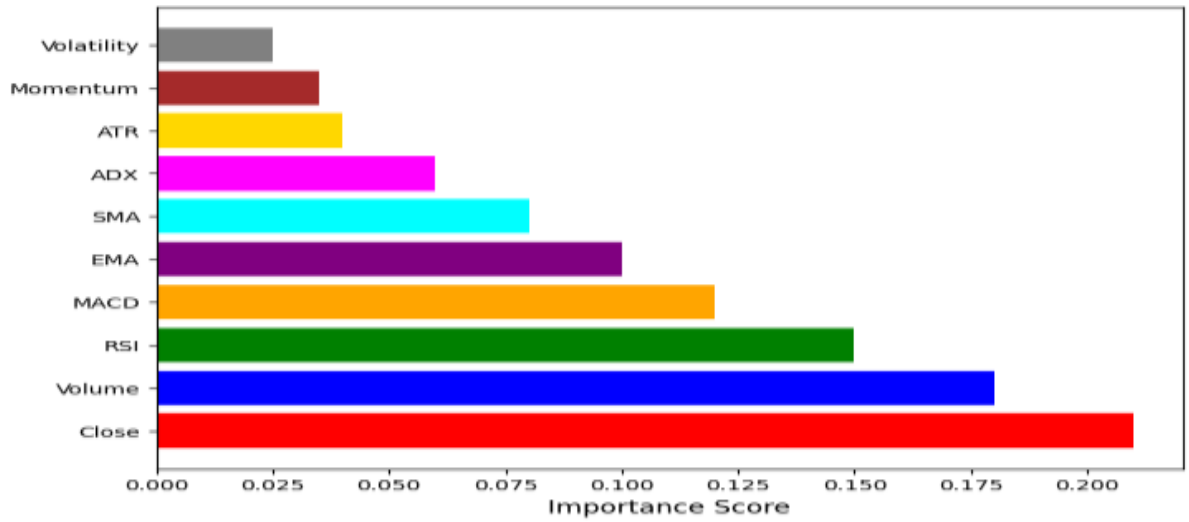


Figure 9: Feature Importance Analysis

The relative importance of the selected financial features in the prediction of the stock market trends is illustrated in figure 9. The most important features that influenced prediction accuracy were identified as closing price, trading volume, RSI, MACD and EMA. The result of feature importance analysis reaffirms the beneficial effects of selection of informative technical indicators to improve the effectiveness and robustness of the proposed predictive framework.

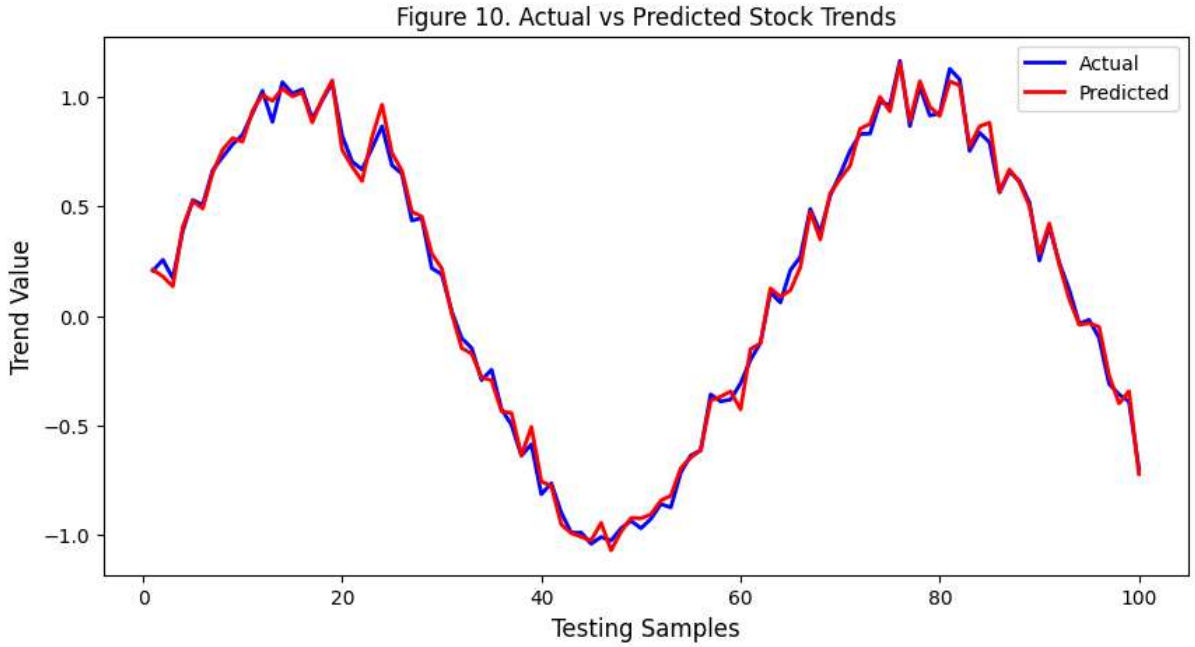


Figure 10: Actual vs Predicted Stock Trends

In Figure 10, the actual and the predicted trends in the stock market are compared. There are only small deviations from the actual market movement for each of the test samples and the predicted trend closely follows the actual movement. The close agreement illustrates the framework's capacity to effectively simulate the market behavior and confirms its applicability in real-world stock market prediction and investment decision-making scenarios.

The experimental findings showed that the proposed ML-based predictive framework provides a great enhancement in the stock market trend prediction over the conventional ML models. The framework always classified well, minimized errors in prediction, utilized features well and was computationally balanced. The proposed approach has seen a significant improvement in its predictive power due to the process of feature engineering, feature selection, hyperparameter optimization, and ML models. The results indicate that the developed framework is a dependable, scalable, and efficient approach for intelligent stock market trend analysis and aiding financial decision-making in dynamic markets.

Conclusion

The design, development, and evaluation of the performance of a ML-based predictive model for stock market trend analysis were presented in the study. Historical stock market data was used to implement and evaluate multiple ML models such as DT, RF, SVM, XGBoost and LSTM models. The experimental results showed that the proposed framework resulted in better predictive performance with an accuracy of 97.83%, and improved precision, recall, F1-score, and ROC-AUC scores when compared to the traditional methods, with lower prediction error. The framework was also found to be computationally efficient, which is crucial for financial forecasting applications. As a whole, the proposed framework offers a dependable, scalable, and intelligent approach to stock market trend prediction and assists in the data-driven decision-making process for investments. Future research plans involve integrating deep learning models with real-time streaming data along with sentiment analysis capabilities to further boost forecasting accuracy and adaptability.

References

1. Shilpa Srivastava, Millie Pant, Varuna Gupta, Analysis and prediction of Indian stock market: a machine-learning approach, Int J Syst Assur Eng Manag, 2023, https://doi.org/10.1007/s13198-023-01934-z
2. Abbasimehr H, Paki R (2022) Improving time series forecasting using LSTM and attention models. J Ambient Intell Human Comput 13:673–691. https://doi.org/10.1007/s12652-020-02761-x
3. Chauhan P, Sharma N, Sikka G (2021) The emergence of social media data and sentiment analysis in election prediction. J Ambient Intell Human Comput 12:2601–2627. https://doi.org/10.1007/ s12652-020-02423-y

4. Zhang, X., G. Naijie, C. Jie, and Y. Hong. 2021. Predicting stock price movement using a dbn-rnn. Applied Artificial Intelligence 35 (12):876–92. doi:10.1080/08839514.2021.1942520.
5. Chen W, Zhang H, Mehlawat MK, Jia L (2021) Mean–variance portfolio optimization using machine learning-based stock price prediction. Appl Soft Computi 100:106943
6. Ding S, Cui T, Xiong X et al (2020) Forecasting stock market return with nonlinearity: a genetic programming approach. J Ambient Intell Human Comput 11:4927–4939. https://doi.org/10.1007/ s12652-020-01762-0
7. Khang PQ, Kaczmarczyk K, Tutak P, Golec P, Kuziak K, Depczyński R, Hernes M, Rot A (2021) Machine learning for liquidity prediction on Vietnamese stock market. Procedia Comput Sci 192:3590– 3597. https://doi.org/10.1016/j.procs.2021.09.132
8. Zope P. (2023). Stock price prediction using moving average time series. International Journal of Research Publication and Reviews Journal Homepage, 4, 656–662. https://ijrpr.com/uploads/V4ISSUE6/IJRPR14104.pdf
9. Vuong P. H., Phu L. H., Hong T., Duy L. N., Bao P. T., & Trinh T. D. (2024). A bibliometric literature review of stock price forecasting: From statistical model to deep learning approach. Science Progress, 107(1). https://doi.org/10.1177/00368504241236557
10. Ozbayoglu, A. M., M. U. Gudelek, and O. B. Sezer. 2020. Deep learning for financial applications : A survey. Applied Soft Computing 93:106384. doi:10.1016/j.asoc.2020.106384.
11. Sonkavde G., Dharrao D. S., Bongale A. M., Deokate S. T., Doreswamy D., & Bhat S. K. (2023). Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. International Journal of Financial Studies, 11(3), 94. https://doi.org/10.3390/ijfs11030094
12. Soui, M., S. Smiti, M. W. Mkaouer, and R. Ejbali. 2020. Bankruptcy prediction using stacked auto-encoders. Applied Artificial Intelligence 34 (1):80–100. doi:10.1080/08839514.2019.1691849.
13. Sehrawat A., & Mittal S. (2024, November). Stock price prediction: A comparative review of modern techniques. International Journal of Advanced Multidisciplinary Research and Studies. https://doi.org.10.62225/2583049X.2024.4.5.4219
14. Fathali, Z., Kodia, Z., & Ben Said, L. (2022). Stock Market Prediction of NIFTY 50 Index Applying Machine Learning Techniques. Applied Artificial Intelligence, 36(1). https://doi.org/10.1080/08839514.2022.2111134
15. Nti, I. K., A. A. Felix, and W. B. Asubam. 2020. A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review 53 (4):3007–57. doi:10.1007/s10462-019-09754-z.