International Journal of Engineering Research and Science & Technology

IJERST

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(1) (2026)

ijerst.editor@gmail.com
editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

Purple Stock: "A Multi-Model Machine Learning Framework for Real-Time Stock Price Prediction with Adaptive Model Selection"

M. Bhargavi[1], P. Durga Prasad[2], Y. Abhiram Sai Manikantha[3], Y. Dileep[4], P. Sasidhar[5]

[1] Assistant Professor, Department of Computer Science and Engineering

[2],[3],[4],[5] Students, Department of Computer Science and Engineering

Nadimpalli Satyanarayana Raju Institute of Technology, Visakhapatnam, Andhra Pradesh, India

Corresponding Author: [1]bhargavi.cse@nsrit.edu.in ,[2] sweetyprasad1026@gmail.com ,

[3] abhiramwhy@mail.com ,[4] dileepdd2004@gmail.com ,[5] sasidharpalikala@gmail.com

Abstract

Stock market prediction is a complicated issue since it is volatile, non-linear, and external factors. The following paper describes Purple Stock, which is a real-time stock analysis and forecasting system whose architecture is founded on a multi-model machine learning structure of short-term price forecasting. The system combines various predictive models, such as Long Short-Term Memory (LSTM), Random Forest, and Extreme Gradient Boosting (XG Boost) as well as combinations between them, which allows the system to describe both the time-dependent and structural trends of financial time-series data.

Purple Stock, in contrast to the traditional methods based on one model, is a dynamical method that considers a variety of models, applies a well-organized validation procedure, and chooses the best model according to its results. Directional Accuracy is the most important parameter in the selection process so that the price movement trends can be predicted reliably, although the root mean square error (RMSE) and the mean absolute error (MAE) are utilized in order to preserve the numerical stability and reduce the error in prediction. This multi-metric analysis allows flexibility in different market situations, such as trending, volatile and sideways markets.

The system can access up-to-date financial information through the Yahoo Finance API and accommodate various exchanges in the world, such as the U.S., Indian, Japanese, and Korean ones. There is a time-consistent currency conversion system, which converts the stock values to both native currency and Indian Rupees (INR) to represent the stock values correctly across the markets. The platform also has an in-built dashboard where visualization of past trends, future projections, model confidence and technical indicators can be viewed, and the platform is transparent by showing the chosen model and its performance indicators.
Empirical evidence shows that there is no model that is always more effective than others. Alternatively, adaptive and hybrid models selection methods yield stronger and more effective forecasts, which points to the usefulness of ensemble learning in operational financial forecasting scenarios.

Keywords: Machine Learning, Stock Prediction, LSTM, XGBoost, Random Forest, Ensemble Learning, Time Series Forecasting, Financial Analytics, Real-Time Systems

Introduction

Financial markets are dynamic and complicated and they are affected by behaviour of investors, economic trends and external occurrences. This complicates short-term prediction of stock prices because of volatility, non-linearity and noise in financial time-series data. Conventional statistical tools can have difficulty in capturing such properties, and the use of machine learning methods is increasingly popular.
The LSTM, Random Forest, and XGBoost models have proven themselves to be effective in stock data modeling. The LSTM is good at capturing the temporal dependence and the tree-based model is good at capturing the non-linear relationships. Nevertheless, there is no one model that is excellent in every market environment, as there are trending, stable, and volatile markets.
Purple Stock solves this weakness by using a multi-model framework, which dynamically compares and picks the best model according to the performance. Directional Accuracy is of priority in the system when predicting a trend with the help of RMSE and MAE to remain stable. It is also capable of combining real-time worldwide market data with time-stable currency conversion, and offers a single dashboard to visualize and have transparency. It is aimed at obtaining more adaptive and reliable predictions than one-model methods.

Literature Survey

Machine learning has found its way into the stock prediction because it can operate complex and non-linear data. Time-series forecasting LSTM networks are frequently employed because these networks are able to capture time-dependent dependencies, whereas they are expensive to compute. Random Forest is more resistant and less prone to overfitting, whereas XGBoost is more accurate thanks to boosting methods; still, neither of them has a built-in time modeling. The recent studies are directed to ensemble and hybrid methods that utilize several models to exploit their benefits. Hybrid models, particularly those that combine LSTM with tree-based models have been demonstrated to be more effective under different market conditions.

Financial prediction depends on evaluation metrics. Although RMSE and MAE are used to gauge numerical error, they are not totally indicative of trading usefulness. Directional Accuracy has served as a significant indicator, since the direction of the correct movement is often more useful than precise price forecasting.

Irrespective of these developments, most of the systems are not transparent and use a fixed model selection. PurpleStock fills these gaps by dynamic model selection, multi-model integration and explicit exposure of model performance, enhancing the adaptability and practical usability.

Related Work

Numerous research studies have explored the use of machine learning techniques for stock market prediction. Random Forest and XGBoost have been widely used due to their efficiency in handling large datasets and their ability to model nonlinear relationships. These models have demonstrated strong performance in financial forecasting tasks. LSTM networks have also been extensively studied for their ability to process sequential data. They are particularly effective in capturing temporal dependencies in stock price movements. However, their performance can vary depending on the quality of data and market volatility.

Hybrid models have been proposed to improve prediction accuracy by combining multiple algorithms. These models aim to leverage the strengths of different techniques, such as combining the memory capabilities of LSTM with the structural learning of tree-based models. Despite these advancements, most existing systems rely on static models or predefined combinations. This limits their ability to adapt to different stocks or changing market conditions.
PurpleStock introduces a dynamic model selection mechanism that addresses this limitation. By evaluating multiple models for each ticker and selecting the best-performing one, the system ensures more accurate and reliable predictions. This adaptive approach distinguishes PurpleStock from traditional stock prediction systems.

System Architecture and Design

PurpleStock is a system that helps us work with stock market data in time. It does this by putting a few important steps like getting the data cleaning it up running models on it checking how well it works and showing us the results. This system is designed to be flexible and work well with the paced world of finance.

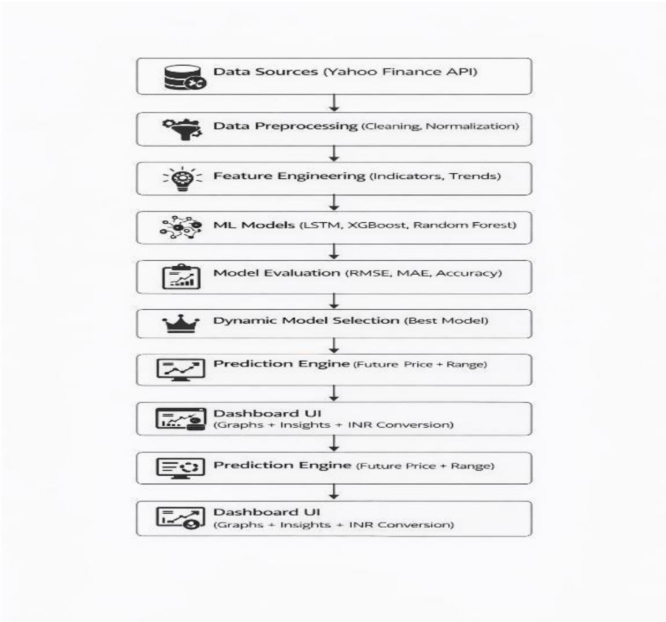
The system starts by getting data from the stock market. It uses the Yahoo Finance API to get both time and historical data on stocks. We can get data from markets around the world like the US, India, Japan and Korea which helps us compare what is happening in different places. The system also makes sure that when it looks at data it uses the exchange rates that were in place at that time and when it makes predictions it uses the current rates so the numbers are accurate.

After we get the data we need to clean it up. This means we fix any information make sure all the numbers are on the same scale and get rid of any noise that might be in the data. We also try to find patterns in the data like how prices have changed over time to help our models work better. We split the data into two parts: one for training and one for testing. We make sure to keep the data in the order it happened.

The system uses a few models to try to predict what will happen in the stock market. These models include things like LSTM, Random Forest and XGBoost. We also try combining them in different ways. Each model looks at the data and makes its own prediction so we can compare them fairly. We then check how well each model did using something called Directional Accuracy as our measure and also looking at RMSE and MAE to get a fuller picture. Based on this the system picks the model that did the best and uses it to make the prediction.

Finally the system shows us the results in a dashboard that's easy to understand. We can see what has happened to stock prices in the past what the model thinks will happen in the future how confident the model is and other important details, like support and resistance levels. The system is also transparent meaning it shows us which model it used and how well that model did so we can understand why it made the predictions it did. This helps us trust the system and its predictions.

System Architecture Diagram — End-to-End Pipeline of Purple Stock



Implementation Details

Methodology and Evaluation Metrics

The PurpleStock system uses a machine learning pipeline that combines models to make stock predictions. This pipeline is very reliable. Can adapt to changes in the stock market in real time. When someone asks for a stock prediction the system looks at data and processes it. Then it uses models at the same time to make predictions. These predictions are then checked using metrics to see which model is the best. The best model is chosen to give the prediction.

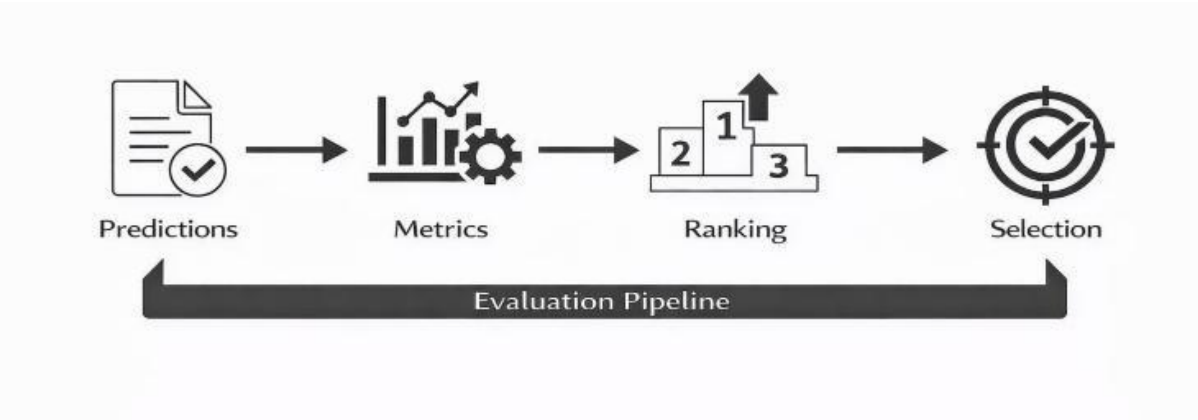
A big part of the PurpleStock methodology is using something called Directional Accuracy to evaluate the models. This is different from methods that just look at errors. Directional Accuracy checks if the predicted price movement is going in the direction as the actual trend. This is really important for trading because it is more important to know the direction of the trend than the value. So the system focuses on getting the trend rather than the exact number.

The PurpleStock system also uses something called Root Mean Square Error and Mean Absolute Error to check how stable the predictions are. Root Mean Square Error is good at finding predictions that're not stable and Mean Absolute Error gives a good idea of the average error without worrying too much about outliers. By using these metrics the system can look at both big mistakes and overall consistency.

When the system chooses a model it uses a step by step approach. First it ranks the models based on Directional Accuracy. If two models are tied it uses Root Mean Square Error to break the tie. Then it uses Mean Absolute Error to validate the choice. This approach helps the system adapt to market conditions while keeping the predictions accurate and stable.

The PurpleStock system does not just give a prediction. It also gives a confidence score based on how the model performed during validation. It gives a range of future values and a trend bias classification, which can be Bullish, Neutral or Bearish. These extra pieces of information help users understand the predictions better by combining numbers with insights about the trend. The PurpleStock system is really good, at giving users an idea of what is going on in the stock market.

Model Evaluation Flow — Multi-Metric Selection Process



Model Architecture

PurpleStock utilizes a range of machine learning algorithms and hybrid ensembling methods to efficiently deal with the complex features of financial time series data. The system utilizes a combination of three main machine learning algorithms: Long Short-Term Memory (LSTM), Random Forest, and Extreme Gradient Boosting (XGBoost). These algorithms are used for predicting stock prices based on their individual strengths.

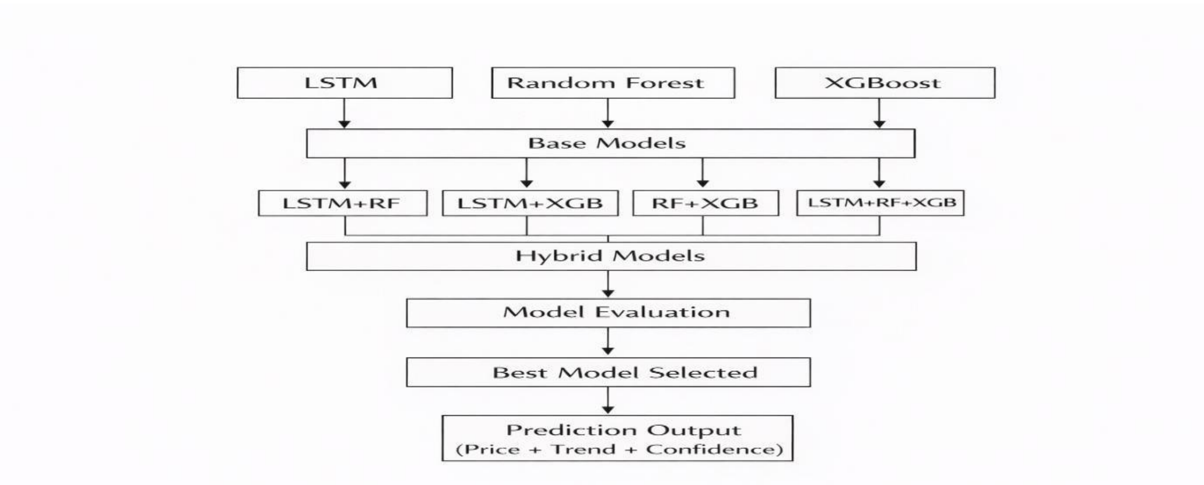
The first algorithm used is Long Short-Term Memory (LSTM). This algorithm is specifically used for time series data and can efficiently deal with long-term temporal dependencies. This allows for effective trend identification in stock price changes. The second algorithm used is Random Forest. This algorithm is based on decision trees and is used for improving the stability of predictions. This is achieved by reducing the probability of overfitting in machine learning models. The third algorithm used is Extreme Gradient Boosting (XGBoost). This algorithm is used for improving the efficiency of predictions based on gradient boosting methods.

Hybrid ensembling methods are used to combine the strengths of individual machine learning algorithms. These algorithms are used for dealing with complex features in financial data. These algorithms are based on pairwise combinations of individual algorithms as well as a hybrid combination of all algorithms. These algorithms are effective in dealing with market behavior characterized by mixed patterns.

All base and hybrid models are run in parallel for each stock request with the same input and preprocessing steps. This ensures a fair comparison and evaluation of the models. The output generated by each model is then passed on to the evaluation layer, where the model's performance is checked against various parameters before finally selecting the best model.

The flexibility of the architecture can also be seen in its response to changing market conditions. LSTM-based models perform better in trending markets because of the sequential learning capability of LSTM. Random Forest performs better in stable markets since it is able to give stable output in a stable environment. XGBoost performs better in handling short-term market fluctuations and unstable markets. This changing nature of model performance under different market conditions makes the dynamic model selection an essential component of the system.

Model Architecture — Base and Hybrid Model Structure



Data Pipeline and Feature Engineering

The performance of the PurpleStock system is heavily dependent on the quality, consistency, and structure of the data. A data pipeline is implemented in the system, which fetches, preprocesses, and transforms the financial data into a state suitable for machine learning algorithms. Data related to the stock exchange is fetched from the Yahoo Finance API, which includes historical stock prices, live stock prices, stock volumes, and other related financial information. The PurpleStock system supports data from different stock exchanges worldwide, including the US stock exchange, Indian stock exchange, Japanese stock exchange, Korean stock exchange, and so on.

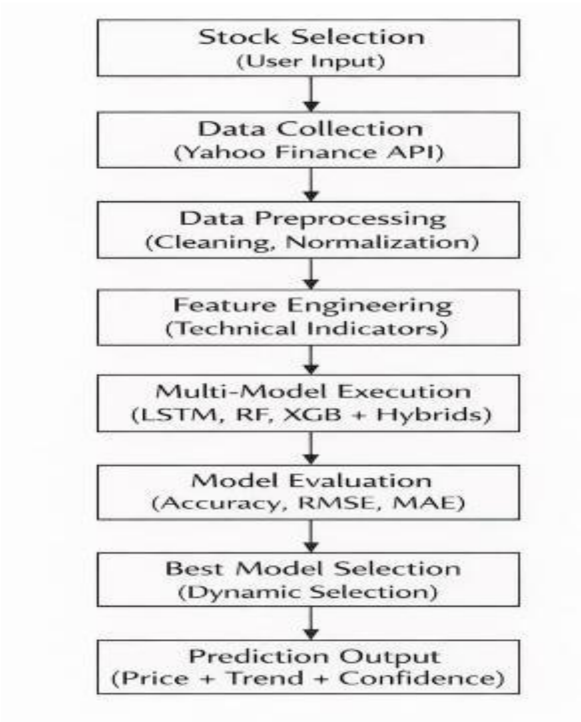
After the data is fetched, preprocessing is performed on the data to ensure better accuracy and consistency. This includes dealing with missing data, normalizing the stock prices, removing noise from the data, and so on. While dealing with time series data, the data must be in chronological order; hence, a temporal split is performed on the data, normally dividing the data into training sets and validation sets.

Feature engineering is performed on the data to improve the performance of the machine learning model by extracting relevant features from the data. Relevant features like differences between stock prices, moving averages, trends, and so on are generated on the data.
Another important aspect of the PurpleStock system is the implementation of time-consistent currency conversion. This is because the PurpleStock system supports different stock exchanges worldwide. Historical data is converted according to the exchange rates prevalent at the time period, while live data is converted according to the exchange rates prevalent at the time of prediction.

To maintain fairness, the data is fed into all the models, and the same preprocessing and feature engineering techniques are applied on each model. This is important because the performance of the model is heavily dependent on the data. All the models are updated dynamically by fetching data from the Yahoo Finance API; hence, the need for static data is avoided.

One of the properties of PurpleStock is its adaptive model selection system. Directional Accuracy is used to evaluate each model with RMSE and MAE being the secondary measures. The most appropriate model to use in the prevailing conditions is automatically chosen to produce the final prediction output.

Data Pipeline — From Data Source to Model Input



Experimental Setup

To assess the viability and effectiveness of PurpleStock, experiments were conducted using real-world stock market data. This ensures the diversity and realism in the sectoral behavior and market conditions. For instance, the system was tested on representative stocks such as AAPL and TSLA for the US market, INFY.NS and RELIANCE.NS for the Indian market, 7203.T for the Japanese market, and 005930.KS for the Korean market. These stocks were chosen to cover a wide range of steady, volatile, and cyclical market behaviors.
For the experiment, the stock market data was retrieved using the Yahoo Finance API and processed using the system pipeline. A temporal split was used to split the stock market data, where 80 percent was used for training and 20 percent for validation. For the experiment, all the basic and hybrid models were run in parallel on the same stock market data to maintain consistency in the experiment. Predictions were generated using the validation data, and the system was evaluated based on the accuracy using the Directional Accuracy, RMSE, and MAE metrics.

The experiment was conducted in different market conditions such as trending, highly volatile, and sideways markets to assess the adaptability and flexibility of the system. For the experiment, the best-performing model, accuracy, and confidence were recorded. This experiment was conducted using moderate computing conditions, where the system was able to generate predictions in a few seconds using the parallel processing capabilities. This ensures a balance in the system's computing capabilities and the accuracy of the predictions.
The main aim of the experiment was to assess the viability and effectiveness of the system in providing reliable predictions by using dynamic model selection and the viability and effectiveness of the system in providing consistent predictions using the hybrid models.

Latency Analysis

Stock	Latency (seconds)
Apple	3.0
Tesla	3.4
Infosys	2.8
Reliance	3.1
Toyota	3.2
Samsung	3.0

Results and Analysis

From the experimental results, it can be observed that the PurpleStock system performs well in adapting to changing conditions in the financial markets. The system uses the concept of dynamic model selection. Instead of depending on the performance of a single model in predicting the stock prices, the system evaluates the performance of several models and then uses the best model to predict the stock prices. This has resulted in more consistent and realistic predictions for the stock prices.

Performance Metrics of Best Performing Models for Individual Stocks via Dynamic Selection

Stock	Best Performing Model (Dynamic Selection)	Accuracy (%)	RMSE	MAE	Latency (sec)
Apple	LSTM + XGB	76.2%	2.65	2.05	3.0
Tesla	LSTM + RF + XGB	72.8%	3.42	2.68	3.4
Infosys	RF + XGB	75.1%	2.54	1.98	2.8
Reliance	LSTM + RF	74.3%	2.73	2.11	3.1
Toyota	XGB	73.6%	2.89	2.24	3.2
Samsung	LSTM	76.9%	2.99	2.24	3.2
Samsung	LSTM	76.9%	2.99	2.19	3.0

As observed from the performance of the models in predicting the stock prices, none of the models has performed better than the others in all conditions. The performance of the models has been consistent in the sense that the performance of the models depends on the conditions in the financial markets. The performance of the Random Forest model has been good in conditions of stability or moderate trending. The XG Boost model has performed well in conditions of short-term variation. The performance of the LSTM model has been good in conditions of trendiness. The performance of the hybrid model has been relatively consistent in all conditions.

As observed from the performance of the models in predicting the stock prices, the Directional Accuracy has varied depending on the conditions in the financial markets. The Directional Accuracy has been relatively high in conditions of trendiness. The Directional Accuracy has been relatively moderate in conditions of stability. The Directional Accuracy has been relatively low in conditions of high volatility. The low Directional Accuracy in conditions of high volatility has been realistic.
As observed from the performance of the models in predicting the stock prices, the low values of the Root Mean Squared Error and the Mean Absolute Error indicate the stability in the predictions. The high values of the Root Mean Squared Error indicate the conditions of high volatility in the financial markets. The use of both the Root Mean Squared Error and the Mean Absolute Error ensures the consideration of extreme conditions in the financial markets.

As observed from the performance of the models in predicting the stock prices, the confidence level of the system has been consistent with the performance of the models. The confidence level has been relatively high in conditions of stability. The confidence level has been relatively low in conditions of high volatility. The low confidence level in conditions of high volatility has been realistic.

Discussion

The study carried out in PurpleStock demonstrates the importance of considering multiple models in the prediction of financial time series. This study confirms the dependency of model performance on market conditions and asserts that there is no single model that performs optimally in all cases.

The application of multiple models in the system enables it to benefit from the different strengths of each model. For instance, LSTM is used for its capability in capturing temporal relationships, while tree-based models like Random Forest and XGBoost are used for capturing non-linear relationships. This is in addition to the hybrid model combinations, which improve model performance by minimizing model bias and improving model stability in mixed and uncertain market conditions.

The application of multiple evaluation metrics in model selection is also important in evaluating model performance. For instance, Directional Accuracy is used because it is relevant in decision-making in the stock market, as it is used in predicting trends. On the contrary, metrics like RMSE and MAE are used in evaluating numerical model stability.

The study also demonstrates the importance of model interpretability in model selection. For instance, PurpleStock enables users to understand not just the model's prediction but also its reliability. This is important in real-world applications like the stock market, where uncertainty is inherent.

Overall, it can be concluded that the discussion has highlighted the importance of integrating adaptability, multi-model integration, and transparent evaluation for developing more practical stock prediction systems. The results obtained are in line with the current research trends, and it is evident how effective the proposed approach is for handling more complex financial environments.

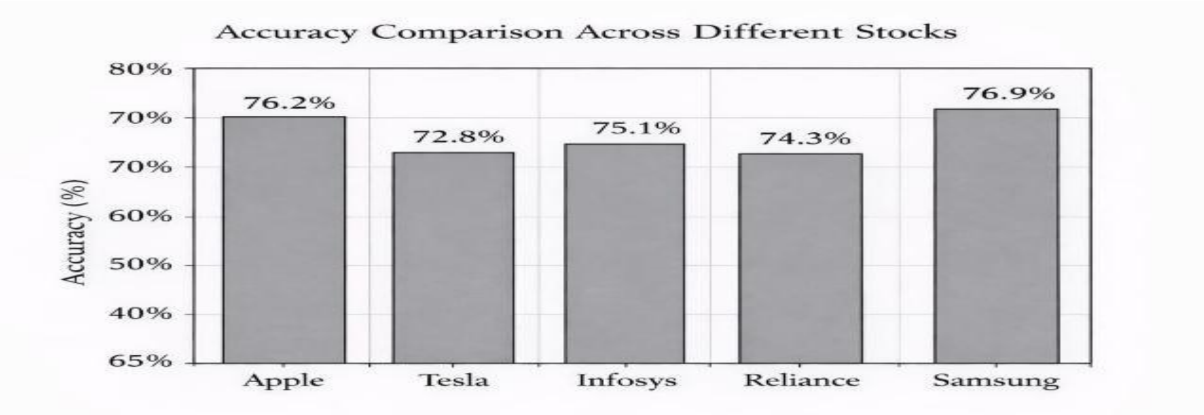


Table 4: Average Metrics per Model Type

Model Type	Avg Accuracy (%)	Avg RMSE	Avg MAE
LSTM	75.3	2.98	2.25
Random Forest (RF)	73.4	3.12	2.41
XGBoost (XGB)	74.6	2.85	2.20
Hybrid Models	75.9	2.74	2.12

Limitations

Despite its dynamic nature, PurpleStock has several inherent limitations, as is common with most stock prediction systems. The system is based on predicting stock prices based on historical price patterns. Although this is an efficient approach, it does not take into consideration other external factors, such as news, economic, and global events, which can greatly influence stock prices. However, it is to be noted that financial markets are highly unpredictable, as is evident from the results obtained. The system is most accurate when there is high volatility in the stock markets.

This is because, in such cases, prediction is more difficult, and it is more challenging to obtain high accuracy. The hybrid approach, which combines multiple models, can slow down the system to some extent. Although it is optimized for near real-time prediction, it is to be noted that there is an inherent trade-off between accuracy and efficiency. The system is more efficient for short-term prediction, and it is not advisable to use it for long-term investment purposes. The performance of the models can vary, as is evident from the results obtained. Although dynamic selection can handle this problem to a great extent, it is not completely possible.

Finally, the currency conversion process is based on the accuracy of exchange rate information. Although minor inaccuracies can occur due to dynamic rate changes, such information is readily available. Additionally, there is a trade-off between complexity and understandability. The use of multiple models and hybrid models is not entirely intuitive for non-technical users.

Table 1: Overall Performance Summary

Metric	Value (Average Across Stocks)
Directional Accuracy	74.8%
RMSE	2.87
MAE	2.21
Latency	3.1 sec

Future Work

PurpleStock is an excellent starting point for an adaptive stock prediction system, but there are several enhancements that can be made to improve its accuracy, efficiency, and usability. One such extension is to incorporate external information, such as news sentiment, social media, and economic information, which can offer more information than stock price history. This can improve the prediction system's accuracy.

Improvements can also be made to feature engineering. The system can be enhanced to incorporate technical stock features, such as RSI, MACD, and stock volatility. This can improve its ability to handle various operating conditions. The system's computational overhead can be reduced by optimizing model execution. This can be done by selectively activating models, pruning models, and hardware acceleration. Another feature can be to implement an auto hyperparameter tuner, which can optimize model parameters for various datasets.

To improve scalability, PurpleStock can be deployed on a cloud platform, which can handle more user requests. The system can be enhanced to include real-time data streaming.
Furthermore, the use of Explainable AI techniques would be useful in improving the interpretability of the model. The use of improved visualization techniques such as comparative stock analysis would further enhance the user experience. This would make the PurpleStock system more accurate, scalable, and user-friendly. Moreover, the system would be more applicable in the general financial analysis domain.

Conclusion

This paper has presented the PurpleStock system, which is a multi-model machine learning system for the prediction and analysis of stock prices in real-time. The system has integrated several models, both basic and hybrid. The system has the capability to dynamically choose the best model for use in the system based on the performance evaluation. The system has prioritized Directional Accuracy and has used RMSE and MAE as secondary parameters. The results obtained from the system show that no single model performs the best in all conditions.

The system has performed well in handling the diversity of the market behaviors. The system has utilized the advantages of the models to make the system more reliable and robust in generating the predictions. Moreover, the system has the advantage of using real-time data, the global markets, and the use of time-consistent currency conversion. The system has the advantage of being transparent in its operations. The results obtained from the system show that the use of the ensemble-based adaptive approach would be more realistic and would be more effective in solving the problem. The gap between the theoretical machine learning models and the actual world has been bridged.

References

[1] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[2] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[3] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. ACM SIGKDD*, 2016, pp. 785–794.

[4] T. Fischer and C. Krauss, "Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.

[5] J. Patel et al., "Predicting Stock Market Index Using Fusion of Machine Learning Techniques," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2162–2172, 2015.

[6] G. P. Zhang, "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[7] R. S. Tsay, *Analysis of Financial Time Series*, 2nd ed., Wiley, 2005.

[8] E. F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," *The Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[10] Yahoo Finance, "Financial Data API," Available: https://finance.yahoo.com/

[11] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *ICLR*, 2015.

[12] J. Brownlee, "Machine Learning Mastery for Time Series Forecasting," 2018.

[13] L. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *NeurIPS*, 2017.