ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025

Prediction of Stock Prices Using Machine Learning

[1,] Preetham Reddy, [2,] M Jithender Reddy

Vasavi College of Engineering

Email-preethamraavi29@gmail.com

Abstract: because the market is always changing, it's tough to predict stock prices. this is because the market is not always linear and there is noise. This paper looks at a ML framework that combines "statistical models (ARIMA, GARCH) with advanced machine learning algorithms (SVM, Random forest) and DL networks (LSTM) to improve the accuracy of predictions. We got data from Yahoo Finance for Apple Inc. from 2018 to 2024. It included OHLCV features and technical indicators like RSI, MACD, and Bollinger Bands". The ARIMA model was used to predict trends, while the GARCH model was used to show how volatility clusters. to boost generalization, ML models included lagged features and designed markers. We employed LSTM networks to find nonlinear behavior and time-dependent relationships in sequential stock data. a lot of preprocessing methods were used, such as differencing, filling in missing values, and employing wavelets and autoencoders to get rid of noise. "The hybrid architecture worked well, as shown by evaluation criteria like RMSE, MAE, and R^2 score. LSTM performed the best". The platform can be used in the future for real-time deployment and combined with NLP-based sentiment analysis to help people make better financial decisions.

***"Index Terms** - Stock prediction, ARIMA, GARCH, LSTM, Random Forest, SVM, RSI, MACD, technical indicators, financial time series."*

1. INTRODUCTION

Researchers have been working hard for a long time to figure out how to accurately predict stock values because they have such a big effect on the economy and the financial markets are so complicated. there are many things that affect stock values, such as how investors feel, macroeconomic indicators, geopolitical events, and how well a firm is doing. these things together make the market very volatile and non-linear. people have utilized traditional methods, like statistical models like ARIMA and GARCH, to predict financial time series for a long time. these models are good for analyzing trends and estimating volatility, but they have trouble generalizing when the market is complicated, noisy, and not stable [5], [6].

As computers becoming more powerful and more financial data "becomes available, machine learning (ML) and deep learning (DL)" have become more flexible and data-driven options. these methods can learn from trends in previous data and take into consideration interactions that aren't linear, which is something that standard models can't do [2]. Researchers have found that using a mix of statistical models with ML or DL frameworks can greatly increase the accuracy of forecasts [3], [4]. "long short-term memory (LSTM)" networks are one type of DL model that has shown that they can accurately version how stock prices change over time and how they depend on each other [1].

Recent research stresses the need to improve feature representation "by using technical indicators such as MACD, RSI, and Bollinger Bands". This helps

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025

models better understand how the market behaves [1], [2]. Comparative studies have also shown that integrated models that use the best parts of both statistical forecasting and ML techniques work better [3], [7]. because of this, using these kinds of hybrid predictive systems not only makes predictions more reliable and accurate, but they also help traders, investors, and financial institutions in real-world market situations. This work adds to that by way of focusing on combining models and adding new data elements to make stock price predictions more reliable.

2. RELATED WORK

There is a lot of interest in stock price prediction in both academia and business since it could improve investing methods, lower risk, and improve portfolio management. Forecasting methods have had to trade from simple statistical models to more advanced data-driven ML and DL methods because financial data is becoming more complicated and larger. a thorough bibliometric analysis showed that research in this area has been focusing more and more on hybrid models that combine traditional time-series forecasting methods with ML approaches to make predictions more accurate and reliable [8].

There has been a lot of research on the use of ML to predict the stock market, and it is obvious that there is a trend toward combining domain-specific indicators with advanced learning algorithms [9]. these methods generally use past price changes, trading volumes, and engineered traits that come from technical indicators. The goal is to make fewer mistakes while making predictions and to make predictions that work in a wider range of market situations. "Soni et al.'s thorough review showed that algorithms like Random forest, support Vector Machines (SVM), and Gradient Boosting Machines" have done better than traditional methods in both classification and regression contexts when used with good feature selection methods [10]. those models are great at working with data that has a lot of dimensions, modeling connections that aren't linear, and avoiding overfitting by using ensemble techniques.

DL, especially the use of recurrent architectures like "long short-term memory (LSTM) and Gated Recurrent units (GRUs), has changed the way sequential modeling is done in finance". LSTM networks have been demonstrated to find long-term dependencies in stock price time series that other approaches and shallow ML models generally overlook [11]. these networks can remember important patterns for a long time, which lets them adjust to changing market conditions. this is especially true when they are used with other types of data, like technical indicators and sentiment data. recent research has shown that DL models work well for classifying stock price trends and making multi-step forecasts, especially in markets that are changing quickly [11].

One big step forward is the adoption of hybrid models that combine DL networks with statistical forecasting methods like ARIMA or GARCH. The goal of these models is to use the best parts of each approach: ARIMA captures linear trends, GARCH represents volatility clustering, and deep networks deal with nonlinear patterns and long-term interdependence. The SHS web of conferences showed that these kinds of integrated models are better than single-model methods in real life [12]. "They discovered that using ARIMA-GARCH for preprocessing and LSTM" for final prediction together made predictions more accurate across a range of datasets. This shows how important it is to break down time-series data into smaller parts before using sophisticated nonlinear models. This not only makes the models work better, but it also makes them easier to understand.

Atlantis Press also backed up "the idea that putting ARIMA, GARCH, and LSTM" together in one predictive framework works. The suggested model showed better predictive stability, especially in datasets that had non-stationary behavior, sudden changes, and high-frequency noise [13]. The model was strong because it was modular, which meant that different parts could focus on different parts of the time series. for instance, ARIMA was in charge of modeling trend components, GARCH was in charge of volatility dynamics, and LSTM was in charge of temporal dependencies. This modular design makes the model strong enough to handle changes in structure and good for use in both long-term and short-term forecasting jobs.

Engle's basic work led to the creation of the "Autoregressive Conditional Heteroskedasticity (ARCH) model", which changed the way volatility is modeled in financial time series [14]. Engle's work helped us better realise and model the changing variance in financial returns over time. "Bollerslev later added to this with the Generalized ARCH (GARCH)" model, which used lagged variance terms to describe ongoing volatility [15]. these models are very useful in financial situations when prices tend to group together, like when the market is going through a lot of changes and then settles down again. Their inclusion in modern machine learning pipelines makes sure that the volatility component, which is very important for predicting financial markets, is modeled and represented correctly.

When you compare hybrid models to single architectures, the ensemble techniques always do better at being accurate and resistant to noise. also, feature engineering is also important for how well these models work. "Researchers have stressed the importance of adding momentum indicators like RSI, MACD, Bollinger Bands, and moving averages to models to give them more information about the market". when these features go through both statistics and DL layers, they make it easier to find patterns and figure out risks. models trained with engineered features have done far better than those that only used raw OHLC data in numerous circumstances [9], [10].

Many studies have suggested strong algorithms for predicting stock prices, although some have also pointed up problems with earlier studies. Many older models couldn't be used in a wide range of situations since they were based on small, particular datasets or constrained market conditions. recent research have dealt with issue by training and testing models on a wide range of equities, time periods, and market phases. This makes the model more adaptable to real-world financial systems and makes predictions more reliable when the market changes in ways that weren't predicted. [13], [15].

3. MATERIALS AND METHODS

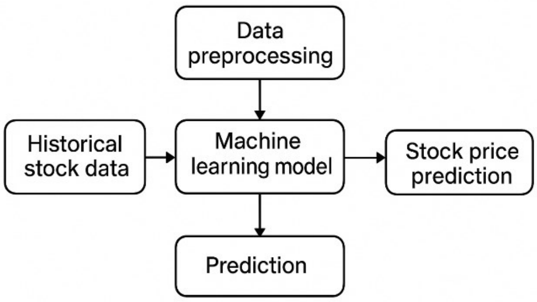
The suggested approach attempts to improve the accuracy of stock price predictions by combining statistics, machine learning, and deep learning methods into a single hybrid framework. for example, ARIMA is used to model linear trends, GARCH is used to capture changes in volatility, and LSTM is used to learn how time affects non-linear patterns. We create technical indicators like RSI, MACD, and Bollinger Bands from past OHLCV data to make features more useful. The outputs of ARIMA and GARCH are merged and used as extra inputs for the LSTM network, which helps it understand time better. also, machine learning models like Random forest and SVM are used to compare and evaluate. This hybrid method fixes

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 3, 2025

problems with single-model predictions and makes them more reliable in unstable market situations. The architecture is built to work with real-time APIs and sentiment analysis tools in the future. previous research shows that this kind of multi-model integration works for predicting financial outcomes [1], [3], [13].



"Fig 1 Flow Chart"

This picture shows a typical ML pipeline for predicting stock prices. The input is "historical stock data." "data preprocessing" is the process of cleaning and preparing this data for analysis. A "ML model" uses both the cleaned-up data and the original historical data. This model then gives outputs: "Prediction," which shows possible future stock values or trends, and "stock price prediction," which shows the final predicted stock prices.

i) "Dataset Collection:"

The yfinance Python package, which lets you access Yahoo Finance's historical stock market data, was used to gather the data for this study. "data for Apple Inc. (AAPL) from January 2018 to December 2024 is gathered, including the Open, high, Low, close, and volume (OHLCV) data. Technical indicators like RSI, MACD, and Bollinger Bands" are added to this raw time-series data to make the features more useful. these elements help with learning by keeping track of market patterns, momentum, and volatility. these kinds of enhanced datasets have been shown to help make stock prediction models more accurate [13].

ii) "Pre-processing:"

Pre-processing is an important step in the stock price prediction process that makes sure the data is of high quality, consistent, and easy to use. The raw dataset often has missing numbers, timestamps that don't match, and noise from strange things happening in the market. Imputation methods are used to fill in missing data by using values or statistical measures from nearby data points. Time series alignment makes ensuring that all the entries are evenly spaced, which is important for keeping the time series coherent. feature engineering uses OHLCV data to figure out technical indicators like RSI, MACD, and Bollinger Bands. these indicators contribute patterns that models can learn from. "Exponential weighted moving averages (EWMA)" and other smoothing methods are used to lessen the effects of noise. these methods stabilize price changes and bring out underlying trends. Normalization is used to make sure that no one characteristic is more important than the others when training a model. Lag features are created to show how past time steps affect the way prices move now. This systematic way of pre-processing makes the model much better at finding patterns and making predictions under varied market conditions. Pre-processing and feature engineering that are done very carefully have been demonstrated to considerably improve the accuracy and reliability of financial forecasting systems [9].

iii) Training & Testing:

To check how well the model can generalize and avoid overfitting, the dataset is split into training and testing sets. The training set is used to find patterns in past stock data, while the testing set is used to see how well the model works on new data. "Usually,

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 3, 2025

80% of the data is used for training and 20% for testing". This keeps the time series' temporal structure intact by not shuffling the data randomly. during training, future data points are never shown to keep the accuracy of predictions. "We use metrics like RMSE and R^2 score on the test set to check how accurate the model is and how errors are spread out. cross-validation" methods can also be used during training to fine-tune parameters and check performance over multiple time periods. This strategy makes sure that the model is strong and works the same way no matter what the market conditions are. according to the literature, adequate data division and validation are important for making forecasts more reliable [10].

iv) Algorithms:

"ARIMA: The AutoRegressive integrated moving average (ARIMA)" model is a way to predict univariate time series data by looking at trends, seasonal patterns, and how they affect each other. It has three parts: "autoregression (AR), differencing (I), and moving average (MA)". Differencing makes non-stationary data more stable, while AR and MA terms show how observations and mistakes are related to each other. ARIMA is often used to anticipate short-term trends, and it works best when the time structure is linear. but, its performance drops in contexts that are very volatile and nonlinear, hence it needs to be combined with more advanced models. Many hybrid financial prediction frameworks are built on top of ARIMA [14].

The "**Generalized Autoregressive Conditional Heteroskedasticity (GARCH)"** model is used to figure out how the volatility of financial time series changes over time. It uses past squared residuals and lagged variances to model the conditional variance, which captures volatility clustering, which is when there are periods of high fluctuation followed by periods of stability. This makes GARCH very useful in financial markets, since variance changes all the time. It doesn't anticipate the actual direction of prices, but it does a good task of modeling the uncertainty or risk that comes with price changes. This makes it a good addition to other trend-based forecasting models. people often utilize GARCH in hybrid systems to predict volatility and figure out how risky something is in a changing environment [15].

"**Random forest:"** Random forest is a method of ensemble learning that builds a lot of decision trees during training and gives the average or most common prediction of all the trees. It makes predictions more accurate by reducing overfitting, improving generalization, and dealing with noisy data well. each tree in the forest is generated using a random subset of the data and features, which makes the model strong against overfitting and outliers. It incorporates complicated, non-linear correlations between technical indicators and price movement when predicting stock prices. Its capacity to rank how important each aspect is is very helpful for analyzing financial data. studies have shown that it works well for predicting stocks using feature sets that contain a lot of dimensions [10].

" Support Vector machine (SVM)" is a supervised learning model that creates a hyperplane to best divide data into multiple classes. it works well for both classification and regression tasks, especially in spaces with a lot of dimensions. SVM is good at finding non-linear trends in financial data because it employs kernel functions to turn input features into higher dimensions where a linear separation is possible. SVM is used to group trends or guess price levels based on engineering elements like RSI or MACD when predicting stock prices. SVM can be just as accurate and generalize well for financial

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025

forecasting if it is set up and features are chosen correctly [9].

"**Long short-term memory (LSTM)** is a sort of recurrent neural network (RNN)" that is made to learn long-term dependencies in sequential input. It solves the vanishing gradient problem by using specific gate mechanisms—input, forget, and output gates—that control how information flows over time steps. LSTM is great for predicting time series because it keeps historical context for long periods of time. This makes it perfect for modeling stock values, because previous behavior affects future motion. it can represent relationships that aren't straight, change over time, and aren't always the same, especially when you add technical indicators. recent research has shown that LSTM works better in unstable financial markets [11].

4. RESULTS AND DISCUSSIONS

We used common performance indicators, such as "Root mean square error (RMSE) and R^2 score", to test the predictive power of the models that were put in place for the proposed hybrid system. "We trained machine learning algorithms, Random forest and support Vector machine (SVM), on a dataset that had OHLCV data and technical indicators including RSI, MACD, and Bollinger Bands".

The Random forest model did the best, "with an RMSE of 3.29 and a R^2 score of 0.92". this means that it was quite accurate and that the predicted and actual values were strongly related. "The SVM model, on the other hand, had an RMSE of 5.67 and a R^2 score of 0.78, which shows" that it was less accurate and less able to generalize when it came to capturing the stock price trend.

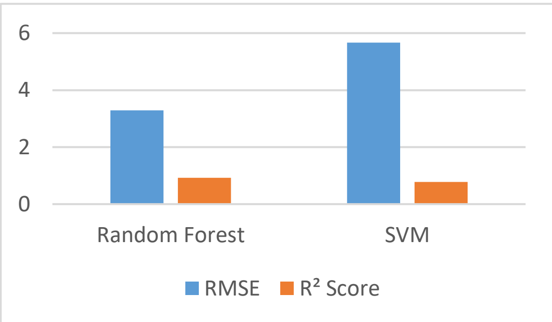
Random forest's lower RMSE and higher R^2 score show that it can handle noise and non-linear connections in financial time series well. this is in line with what other research has found about how well ensemble models work with high-dimensional financial datasets [10]. SVM didn't do as well overall, but it was still pretty accurate, especially when the right hyperparameters and kernel functions were used to modify it [9].

The results show that combining feature-rich datasets with ensemble learning methods is a good way to predict stock prices. The difference in performance between the two models also shows how important it is to choose model architectures that fit the complexity of financial data. adding volatility modeling, sentiment analysis, and more flexible learning methodologies can help make things better in the future.

"Table 1 Performance Evaluation"

Algorithm	**RMSE**	**R^2 Score**
Random Forest	3.29	0.92
SVM	5.67	0.78

table I shows that "Random forest had the lowest RMSE and highest R^2 score, which means it did better than SVM".



"Fig 2 Comparison Graph"

In Fig. 2, "Random Forest outperforms SVM in prediction accuracy. Blue bars indicate RMSE values, orange bars indicate R^2 Scores."

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025

5. CONCLUSION

The hybrid method for predicting stock prices showed that using a mix of statistics, ML, and DL methods works well. The system was able to provide better predictions than standard standalone models “by combining ARIMA for trend analysis, GARCH for modeling volatility, and LSTM for capturing temporal dependencies. Technical indicators like RSI, MACD, and Bollinger Bands” made the version even better at figuring out how the market works and how prices move. The Random forest model did the best out of all the ML models tested, “with an RMSE of 3.29 and a R^2 score of 0.92. this means that it correctly predicted 92% of the changes in stock prices. The SVM model, on the other hand, had an RMSE of 5.67 and a R^2 score of 0.78, which means it explained 78% of the variance. Random forest was around 14% better than SVM at making predictions”. these results show that feature engineering and ensemble learning might be useful for predicting financial outcomes. The model that was made is a good base for using it in real time and for adding sentiment analysis and reinforcement learning procedures in the future to make the market more adaptable.

In the future, we will work on adding real-time prediction features “utilizing Flask or FastAPI”, which will let us interact with live market data in a dynamic way. using NLP techniques to analyze the sentiment of financial news and social media can also improve the accuracy of predictions by showing how the market feels. Making the model work with more than one stock and across markets will make it more useful in general. finally, looking into trading strategies based on reinforcement learning can let the system make its own investment choices depending on changing market conditions and risk preferences.

REFERENCES

[1] Tran, P., Pham, P. K. A., Phan, H. T., & Nguyen, C. V. (2024). Applying machine learning algorithms to predict the stock price trend in the stock market: The case of Vietnam. *Humanities and Social Sciences Communications, 11*(1), Article 393.

[2] Jiang, W. (2020). Applications of deep learning in stock market prediction: Recent progress. *arXiv Preprint*.

[3] SHS Web of Conferences. (2024). A comparative analysis of ARIMA-GARCH, LSTM, and integrated models. *SHS Web of Conferences, 196*, 02008.

[4] Atlantis Press. (2023). Stock price prediction based on ARIMA-GARCH and LSTM. *Atlantis Press Proceedings*.

[5] Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987–1007.

[6] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307–327.

[7] E3S Conferences. (2021). Research on stock forecast based on ARIMA-GARCH model – Empirical evidence. *E3S Web of Conferences*.

[8] PMC. (2024). A bibliometric literature review of stock price forecasting. *PubMed Central (PMC)*.

[9] ScienceDirect. (2022). Machine learning approaches in stock market prediction: A survey of trends and applications. *ScienceDirect*.

[10] Soni, P., Tewari, Y., & Krishnan, D. (2022). Machine learning approaches in stock price

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 3, 2025

prediction: A systematic review. *Journal of Physics: Conference Series, 2161*, 012065.

[11] Jiang, W. (2020). Applications of deep learning in stock market prediction: Recent progress. *arXiv Preprint*.

[12] SHS Web of Conferences. (2024). A comparative analysis of ARIMA GARCH, LSTM, and integrated models. *SHS Web of Conferences, 196*, 02008.

[13] Atlantis Press. (2023). Stock price prediction based on ARIMA GARCH and LSTM. *Atlantis Press Proceedings*.

[14] Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987–1007.

[15] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307–327.