International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

AI-Powered Stock Price Prediction System Using Machine Learning

Pavani K J[1], Akhila S Babu[2]

Abstract: ***The stock price prediction is difficult owing to the dynamic, volatile and non-linear nature of the financial markets. Proper forecasting will help users make good decisions. This study focuses on the AI-Powered Stock Price Prediction System Using Machine Learning, which uses Linear Regression and Random Forest Regression on past data of the stock market. The proposed system comprises data collection, data preprocessing, feature selection, model training, prediction, and model performance assessment. Linear Regression model is used as a baseline model to find the linear pattern of the data whereas Random Forest Regression finds the nonlinear patterns of the data. The performance of both models is assessed based on the following metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 Score. The findings of this experiment reveal that the prediction accuracy of the Random Forest Regression is higher than Linear Regression. The proposed system will help users to analyse past trends and make informed decisions.***

Keywords: Stock Price Prediction, Linear Regression, Random Forest Regression, Financial Forecasting, Historical Stock Data, Predictive Analytics, Data Preprocessing

1. Introduction

The stock market plays an important role in the economy of the world. It is difficult to predict the price of the stocks due to various reasons such as economy, performance of the companies, market trend, political influence, and behaviour of the investors.

Artificial Intelligence and Machine Learning offer powerful ways of analysis and pattern recognition using large amounts of historical financial data. These techniques can assist investors and financial analysts in analysing market trends and taking the right decisions.This paper considers Linear Regression and Random Forest Regression techniques in the stock price prediction. Linear regression will be used as an example of a basic and understandable algorithm, and the Random Forest regression technique consists of a set of decision trees.

The proposed system relies on historical data and uses the features such as Open, High, Low, Close, AdjClose, and Volume. The process includes data preprocessing, feature selection, model training and testing. The two considered models will be analysed by MAE, MSE, RMSE, and R^2 Score.As seen from the experiments, the Random Forest regression technique outperforms the Linear regression technique in terms of accuracy and prediction errors. In consequence, the proposed system can be considered as a decision-making tool to analyse historical trends and predict stock prices.

2. Literature Review

1) **Decision Fusion for Stock Market Forecasting:** A Systematic Review (2022) Zhang, Sjarif, and Ibrahim studied decision fusion approaches in stock market prediction. The article concluded that using multiple machine learning models is effective in improving the robustness and prediction results especially when there is high volatility [1].
2) **Intelligent Models to Predict Stock Prices:** A Comprehensive Review (2022) The authors Ansah, Denwar, and Appati reviewed algorithms including linear regression, random forest, SVM, decision tree, and LSTM. It was stressed that preprocessing and feature selection can enhance prediction results [2].
3) **Machine Learning Approaches and Data Sets for Stock Market Prediction:** Literature Review (2022) The article by Kumbure et al. reviewed over one hundred articles about machine learning for stock prediction. It was stated that it is crucial to use suitable financial indicators, features, and preprocessing [3].
4) Algorithmic Trading and Short-Term Financial Time Series Forecast Using Machine Learning Approaches (2022) The team of Joiner considered different regression models, ensembles, and deep learning in financial forecasting. Proper feature selection and noise reduction were pointed out as the key aspects of improved predictions [4].
5) Comparison of Predictions of Stock Prices via Linear Regression vs. Machine Learning Methods (2022) The study done by Jin and Yi made a comparison between Linear Regression and Random Forest, XGBoost, LSTM, and GRU. The study results showed that advanced machine learning techniques can model nonlinear correlations better than Linear Regression [5].

3. Methodology

The design process of AI-driven stock price forecasting system involves six key stages: data collection, data preprocessing, feature selection, model training, prediction and performance evaluation.

1) Data Collection

The historical stock price data are collected from trusted data sources like Yahoo Finance. The collected dataset consists of the following parameters: Open, High, Low, Close, Adjusted Close and Trading Volume.

2) Data Preprocessing

Missing, duplicate data entries and outliers in the collected data set are removed. Then, the cleaned dataset is partitioned into the training and test datasets.

3) Feature Selection
Important features such as Open, High, Low, Close, Adjusted Close, and Volume are chosen for prediction. The future value of closing price will be the target variable.

4) Model Training
Two types of supervised learning algorithms are implemented: Linear Regression and Random Forest Regression. Linear Regression detects linear correlations while Random Forest includes several decision trees to detect complicated nonlinear correlations.

5) Stock Price Prediction
Both trained models are used to forecast stock prices using test dataset. The actual and predicted values are compared numerically and graphically.

6) Performance Evaluation
The trained models are evaluated based on the metrics such as MAE, MSE, RMSE and R^2 Score. The lower values of errors and higher R^2 Score mean better performance.

4. Experimental Setup

The proposed model was developed through Python using Jupyter Notebook. Data was obtained for historical stocks on the Yahoo Finance platform, with variables including Open, High, Low, Close, Adjusted Close, and Trading Volume. Preprocessing of data and splitting of data into 80% train and 20% test sets took place. Models were built on the basis of Linear Regression and Random Forest Regression models with the chosen attributes. Pandas, NumPy, Scikit-learn, and Matplotlib were used for data preprocessing, model development, and visualization. Models were validated on the basis of MAE, MSE, RMSE, and R^2 Score.

Model Performance Analysis
The evaluation of Linear Regression and Random Forest Regression performance was done based on MAE, MSE, RMSE, and R^2 Score. Both regressions were tested on the same stock market data from the past, and the outcomes can be seen in Table I.

Table II: Model Performance Comparison of Linear Regression and Random Forest Regression

Algorithm	MAE	MSE	RMSE	R^2 Score
Linear Regression	2.85	15.72	3.96	0.91
Random Forest Regression	1.64	8.31	2.88	0.96

The findings indicate that Random Forest Regression model is more effective compared to Linear Regression in terms of all evaluation metrics used. The model generates lower values of MAE, MSE and RMSE, suggesting fewer errors in predictions. In addition, the model generates an improved value of R^2 Score of 0.96 compared to 0.91 for Linear Regression.

5. Conclusion

The current study introduced the design and implementation of an AI-Powered Stock Price Prediction System using Linear Regression and Random Forest Regression for predicting stock prices from market history data. The design of this proposed system consisted of several steps such as data acquisition, data pre-processing, feature selection, model training, prediction, and evaluation of model performances. The performance of both these models was assessed using MAE, MSE, RMSE, and R^2 Score. The experimental results revealed that Random Forest Regression outperforms the Linear Regression Model. Random Forest Regression gives an MAE value of 1.64, MSE value of 8.31, RMSE value of 2.88, and R^2 Score of 0.96, whereas the performance of the Linear Regression Model gives an MAE value of 2.85, MSE value of 15.72, RMSE value of 3.96, and R^2 Score of 0.91. Therefore, Random Forest can capture nonlinear relationships in stock market data. Hence, it becomes the preferred model for this experiment.

References

[1] Z. Cheng, N. N. A. Sjarif, and R. B. Ibrahim, "Decision Fusion for Stock Market Prediction: A Systematic Review," IEEE Access, vol. 10, pp. 82416–82438, 2022.
[2] M. M. Kumbure, C. Lohrmann, P. Luukka, and J. Porras, "Machine Learning Techniques and Data for Stock Market Forecasting: A Literature Review," Expert Systems with Applications, vol. 197, Art. no. 116659, 2022.
[3] X. Jin and C. Yi, "The Comparison of Stock Price Prediction Based on Linear Regression Model and Machine Learning Scenarios," Atlantis Press, 2022, pp. 506–512.
[4] D. Joiner, R. Sharma, and M. Patel, "Algorithmic Trading and Short-Term Forecast for Financial Time Series with Machine Learning Models: State of the Art and Perspectives," Mathematics, vol. 10, no. 23, Art. no. 4517, 2022.
[5] P. Soni, Y. Tewari, and D. Krishnan, "Machine Learning Approaches in Stock Price Prediction: A Systematic Review," Journal of Physics: Conference Series, vol. 2161, no. 1, Art. no. 012065, 2022.