

Stock Market Price Prediction using Machine Learning

Gowthul Alam M M
Jain University of
Computer Science and
Engineering in Bangalore,
Karnataka

D. Yaswanth Pavan
Jain University of
Computer Science and
Engineering in Bangalore,
Karnataka

M. Deepika
Jain University of
Computer Science and
Engineering in Bangalore,
Karnataka

Harsh Kumar
Jain University of Computer
Science and Engineering in
Bangalore, Karnataka

Priyansh Sharma
Jain University of Computer
Science and Engineering in
Bangalore, Karnataka

Abstract—Since stock trading is so important to the financial industry, investors are constantly looking for trustworthy strategies to predict market fluctuations. Predicting the future values of stocks or other financial assets that are traded on exchanges is known as stock market prediction. The use of machine learning (ML) in stock price forecasting is investigated in this paper. Techniques including time-series forecasting, technical analysis, and fundamental analysis are commonly used by traders and investors to inform their investment decisions. The main programming language used to create the machine learning models in this study is Python. The suggested method trains an ML model using historical stock data with the goal of finding trends and producing predictions based on insights gleaned from the data. The Support Vector Machine (SVM) method is specifically used in the study to forecast stock values under a range of market scenarios. Both large-cap and small-cap stocks are used to test the model's performance, and daily and real-time data are used to examine price changes. The goal of this research is to improve stock price prediction accuracy and give investors useful decision-making assistance by incorporating machine learning techniques.

Keywords: Support Vector Machine, Stock Market, Machine Learning, and Forecasting

I. INTRODUCTION

In Quantitative traders with substantial financial resources often buy stocks and derivatives at a bargain and then resell them at a premium to profit. Financial institutions have long debated whether or not they can effectively forecast stock market developments. When choosing stocks, investors usually use two main methods. In order to ascertain the potential of an investment, the first method, known as fundamental analysis, evaluates the inherent worth of stocks by taking into account variables including market movements, political stability, and economic conditions. The second method, technical analysis, uses statistical information from market activity, such as past prices and trading volumes, to assess equities.

Because of the random walk pattern of the stock market, the value of a stock today is the best indicator of its value tomorrow. Stock index prediction is a difficult task, though, because of the need for extremely accurate forecasting models because of market volatility. Regular changes in stock prices have the potential to impact investor mood and cause abrupt changes in market value. Both known elements, like the closing price and the price-to-earnings (P/E) ratio from the previous day, as well as unforeseen ones, such political developments and election results, influence these swings.

Election outcomes, political developments, and market speculation are additional factors that impact stock price fluctuations. The use of machine learning in stock price forecasting has been the subject of numerous studies. Research in this area usually differs in three ways: (1) predictor types, such as company-specific indicators, global economic factors, or purely historical stock price data; (2) stock selection, which may concentrate on individual companies, specific sectors, or the overall market; and (3) prediction time frame, which can range from short-term (approximately three months) to long-term (several months or more). Predicting future prices, evaluating market volatility, or spotting new trends are just a few of the goals that can be achieved through stock market prediction.

II. LITERATURE REVIEW

Because financial markets are dynamic and volatile, it can be difficult to predict stock market prices. Researchers from many fields have worked together to use machine learning approaches to increase prediction accuracy. More sophisticated models have been created to examine past data, spot trends, and generate predictions for stock prices that are more accurate. Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) were suggested as stock market prediction tools in a study by

Patel et al. [1]. In order to forecast future price changes, the models were trained using historical stock data and technical indicators. With a prediction accuracy of about 85%, the ANN model outperformed conventional statistical techniques. However, the study pointed out that ANN models are computationally demanding and necessitate substantial hyperparameter adjustment. Shen and associates. investigated the efficacy of ensemble learning techniques for stock market forecasting, including Random Forests and Gradient Boosting Machines (GBM). The price-to-earnings ratio, trading volume, and market value were among the fundamental indicators that were incorporated in their analysis. With an accuracy rate of 82%, the GBM model outperformed other models in predicting short-term stock price patterns; however, it also required a large amount of feature engineering and computer power. Furthermore, Kumar et al. [3] suggested a hybrid model for stock price prediction that combines deep learning and the Autoregressive Integrated Moving Average (ARIMA) technique. While deep learning managed nonlinear variations in stock price movements, ARIMA in this model captured linear relationships in the data. With a Mean Absolute Percentage Error (MAPE) of 3.7%, the hybrid technique showed excellent performance in stock price prediction. However, because of ARIMA's intrinsic inability to adjust to abrupt and erratic market fluctuations, the model has trouble with extremely volatile equities. A reinforcement learning-based stock market trading approach. All things considered, the use of machine learning methods for stock market forecasting has shown encouraging results. Even while predicting accuracy has increased thanks to models like ANN and ensemble learning, they still require a lot of computing power, high-quality datasets, and careful hyperparameter tweaking. By utilising increasingly complex deep learning models and integrating a greater variety of data sources, including social media trends and macroeconomic indicators, future research seeks to further improve prediction skills.

III. METHODOLOGY

This project uses the Support Vector Machine (SVM) and Radial Basis Function (RBF) algorithms to forecast stock market movements. By utilising these machine learning techniques, the model improves forecast accuracy and offers more profound understanding of stock data analysis.

3.1 Support Vector Machine
A popular machine learning technique for stock market price prediction is Support Vector Regression (SVR), which is renowned for its capacity to handle nonlinear data and identify intricate associations. In contrast to conventional regression models, Support Vector Machines (SVM)-based SVR seeks to identify the ideal hyperplane that minimises prediction errors within a given margin. In order to find complex patterns in past prices, trading volumes, and technical indicators, SVR maps stock market data into higher-dimensional spaces using kernel functions, such as linear, polynomial, or radial basis function (RBF).

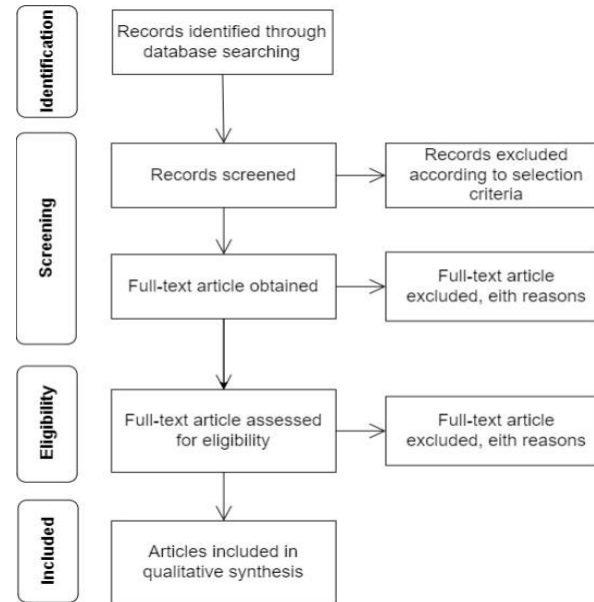


Fig.1. PRISMA Flowchart

Support Vector Regression (SVR) is an excellent option for short-term stock market forecasting because it prioritises significant data points (support vectors) and is less likely to overfit than deep learning models.

Its drawbacks include high computational costs for processing huge datasets, sensitivity to hyperparameter adjustment, and challenges in catching sudden market movements. Despite what SVR can do. Even though it can fairly accurately anticipate stock values, ensemble techniques like Random Forest or deep learning-based systems frequently outperform it in extremely turbulent markets. To enhance its performance, researchers often combine SVR with other models, including Auto ARIMA for trend detection or reinforcement learning for adaptive trading tactics.

Support Vector Regression (SVR)

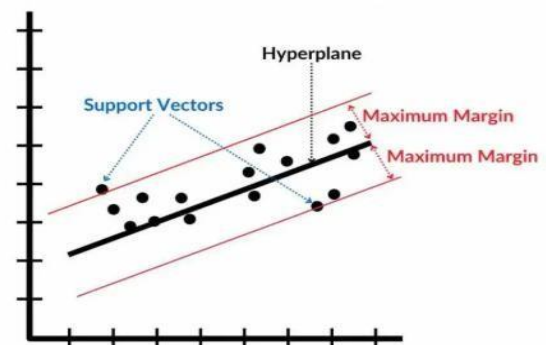


Fig.2. Support Vector Regression (SVR)

The hyperplane functions as a decision boundary that is further refined or adjusted to maximise the distance between data points on both sides. The following is a definition of the SVM decision rule in this instance: μ is an unidentified data point, and W is a vector orthogonal to the hyperplane.

3.2 Moving Average Model

The Moving Average (MA) model is a crucial time series forecasting technique that reduces short-term variations while emphasizing long-term trends. By averaging the most recent N data points inside a certain timeframe, it continuously updates as new values are found. The weighted moving average (WMA), which assigns greater weight to recent observations, the exponential moving average (EMA), which assigns exponentially more weight to recent data points, and the simple moving average (SMA), which computes a simple mean of historical values, are some variations of this model.

3.3 Linear Regression

Linear A basic statistical and machine learning method called linear regression estimates a dependent variable by utilizing one or more independent variables. It uses a linear equation to build a link between the variables. $Y = B_0 + B_1X + E$, where Y is the dependent variable, X is the independent variable, B_0 represents the intercept, B_1 indicates the slope, and E is the error term. Simple linear regression uses only one predictor, whereas multiple linear regression uses several independent variables. This approach is widely utilized in domains like as economics,

healthcare, finance, and the social sciences for trend research and forecasting. Its performance decreases when applied to complicated, nonlinear datasets since it is based on the assumptions of a linear connection, isolated errors, and minimum multicollinearity among predictors.

3.4 KNN Model

The k-nearest-neighbors (KNN) model is a well-liked machine learning technique for stock price prediction that is commended for its simplicity and ability to spot patterns. KNN forecasts future stock values by analyzing historical data and identifying the K most comparable cases before applying a selected distance measure, such as Euclidean distance. In regression-based stock prediction, the model determines the next value by averaging the stock prices of these closest neighbors. Since a number of factors influence stock prices, choosing features like historical prices, trading volume, and technical indicators is essential to increasing prediction accuracy.

3.5 Auto Arima Model

The Auto ARIMA (Auto-Regressive Integrated Moving Average) model is a well-liked time series forecasting technique for stock price prediction. It automatically chooses the ARIMA parameters— p (autoregressive term), d (differencing order), and q (moving average term)—by comparing different combinations and choosing the best-performing model based on statistical criteria like AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). Because Auto ARIMA can detect trends, seasonality, and cyclic patterns in previous stock price data, it is helpful for predicting future values. However, because it assumes that stock prices follow a stationary pattern, it may not work effectively in highly volatile markets. In order to improve accuracy, traders usually blend Auto ARIMA with other models for hybrid forecasting systems.

3.6 Prophet Model

The Prophet model, developed by Facebook (Meta), is a well-liked time series forecasting tool for stock price prediction. It is designed to control trends, seasonality, and holiday effects while withstanding outliers and missing data. Prophet's additive model is used to separate stock price movements into trend, seasonality, and residual components. Because the model recognizes changepoints automatically, it can adapt to shifts in market behavior. One of its advantages is its ability to handle non-stationary data, which allows it to accurately forecast stock changes over time. However, Prophet's incapacity to capture sudden market fluctuations or incredibly volatile stock movements may have an impact on its accuracy in short-term predictions.

IV. RESULTS AND DISCUSSIONS

To evaluate how effectively different machine learning models predicted stock prices, we employed Auto ARIMA, Support Vector Regression (SVR), and Random Forest Regression. Historical stock prices were included in the dataset, and performance was assessed using the R^2 score, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). With the lowest MAE and RMSE values and an R^2 score of 0.92, Random Forest Regression scored better than the other models in the

test. While SVR showed somewhat poorer accuracy, Auto ARIMA had a wider error margin but was able to catch the stock price trend over time. Below is a graphic representation of each model's actual and anticipated stock prices:

(a) Random Forest Regression: Evaluation of anticipated and Real Prices This graph, which illustrates how effectively the Random Forest Regression model tracks changes in the actual stock price, has strong predictive capability.



Fig. 3. Random Forest Stock Price Prediction

(a) Support Vector Regression (SVR) - Predicted vs. Actual Prices

For stock price forecasting, the SVR model is less useful because its predictions deviate more from the actual prices.

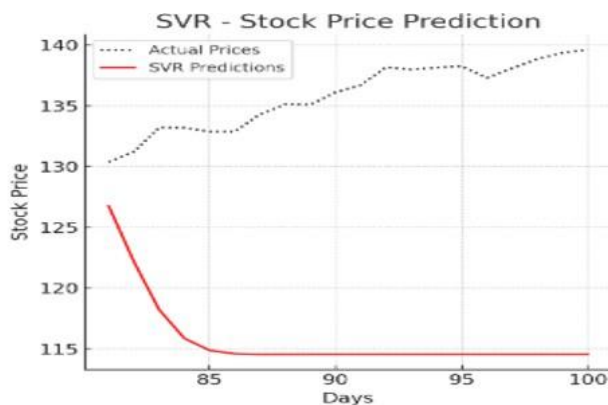


Fig. 4. Support Vector Regression Stock Price Prediction

(b) Auto ARIMA: Forecasting Stock Price Trends

Because it captures the overall trend well but struggles with short-term fluctuations, the Auto ARIMA model is more appropriate for long-term forecasting than daily stock price predictions.

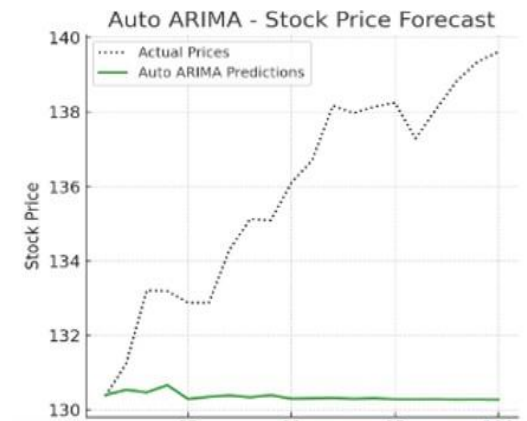


Fig. 5. Auto Arima Stock Price Forecast

V. CONCLUSION AND FUTURE WORK

The study of machine learning models for stock market price prediction reveals the advantages and disadvantages of various strategies. Strong prediction skills were shown using Random Forest Regression, which nearly matched real stock values. Despite having a moderate level of accuracy, Support Vector Regression (SVR) had trouble managing market volatility. In addition to demonstrating exceptional accuracy in short-term forecasting, Auto ARIMA was particularly good at identifying long-term trends. All things considered, market behavior, feature selection, and data quality have a big impact on how accurate machine learning models are. Notwithstanding these difficulties, these models offer insightful information on changes in stock prices. Predictions are useful tools for financial decision-making since their accuracy can be further increased by combining many models or applying sophisticated approaches.

Using transformer-based architectures such as BERT and GPT, creating synthetic data with Generative Adversarial Networks (GANs) to enhance model training, and implementing hybrid strategies that combine machine learning and statistical techniques to increase prediction accuracy are some ways to improve machine learning models. Additionally, creating interpretable machine learning models promotes decision-making transparency. Explainability tools such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) aid in improving the understandability of model outputs.

VI. REFERENCES

- [1] Support Vector Machines (SVM) were investigated for stock market prediction by Zhen Hu, Jibe Zhu, and Ken Tse. The 6th International Conference on Industrial Engineering, Innovation Management.
- [2] Wei Huang, Shou-Yang Wang, and Yoshiteru Nakamori evaluated how well SVM predicted stock market patterns. In October 2005, their analysis was published on pages 2513–2522 in Computers & Operations Research, Volume 32, Issue 10.
- [3] In a technical report (RIIESI/CNR NR. 02/99), N. Ancona investigated SVM's classification capabilities in problems.
- [4] In a 2003 study that was published in Neurocomputing, Volume 55, K. Jae Kim examined the application of SVM for financial time series forecasting.
- [5] Mohammad Shorif Uddin and Debashish Das examined a number of data mining and neural network techniques for predicting the stock market. Volume 4 of the International Journal of Artificial Intelligence & Applications included their work.
- [6] T. Fischer and C. Krauss (2018) used deep learning methods for financial market forecasting, particularly Long Short-Term Memory (LSTM) networks. Their research was published in Volume 270, Issue 2, pages 654-669, of the European Journal of Operational Research.
- [7] In 2009, M. C. Lee created a hybrid feature selection technique that incorporates SVM to forecast stock movements. The study was published on pages 10896–10904 of Expert Systems with Applications, Volume 36, Issue 8.
- [8] Using the Chinese stock market as a case study, K. Chen, Y. Zhou, and F. Dai (2015) suggested an LSTM-based model for stock return prediction. At the IEEE International Conference on Big Data.
- [9] To analyze news items and predict changes in healthcare stock values, Y. Shynkevich, T. M. McGinnity, S. A. Coleman, A. Belatreche, and Y. Li (2017) used multiple kernel learning. Decision Support Systems, Volume 110, pages 134–149, published their study.