Stock Market Forecasting

Rahul Kumar Yadav
Technology and Development
Indian Institute of Technology, Bombay
22m0212@iitb.ac.in

Abhishek Raghuwanshi
Technology and Development
Indian Institute of Technology, Bombay
22m0204@iitb.ac.in

Sarita Kumari
Technology and Development
Indian Institute of Technology, Bombay
22m0227@iitb.ac.in

***Abstract*—This research paper explores the effectiveness of various machine learning algorithms in predicting the stock prices of NIFTY, one of the leading stock market indices in India. The paper uses NIFTY data collected from Y Finance dataset and applies data preparation techniques, feature selection methods, and model selection techniques to compare the performance of four different algorithms: LSTM, Support Vector Regression (SVR), Ridge regression, and Random Forest. The research findings indicate that the LSTM algorithm provides the most accurate predictions, followed by the Support Vector Regression algorithm. This research paper concludes that LSTM can be an effective tool for stock market prediction and decision making.**

Keywords—Y Finance, NIFTY, LSTM, SVR, Ridge Regression, Random Forest

I. Introduction

The stock market along with its tendencies are among the most erratic in the financial world. Researchers are drawn to it in an effort to measure the volatility and forecast its future course (Kalyani, J., Bharathi, P., & Jyothi, P., 2016). The possibility for profit exists for stock traders who properly forecast price trends. For stock traders' decision-making, forecasting stock price patterns is crucial. The stock market, however, exhibits extremely intricate tendencies (Srijiranon, K., Lertratanakham, Y. and Tanantong, T., 2022). According to (Han, X., & Li, Y., 2017), the stock market is a complex system, which is also very dynamic in nature, that is influenced by various economic factors such as market capitalization, overall economic conditions, social media sentiment indices, and financial news. However, forecasting stock prices is considered one of the most challenging tasks in time series prediction due to the market's inherent volatility and noise (Wang, B., Huang, H., & Wang, X., 2012). If the investors can pick the appropriate stocks to invest in, they will get good returns; otherwise, they risk losing their money, which would be bad for both them and their nation. Therefore, it is necessary to create prediction models that could aid in more accurate and efficient stock price predictions. (Gupta, I., Madan, T. K., Singh, S., & Singh, A. K., 2022).

The financial sector has witnessed significant advancements with the use of machine learning owing to its ability to process large volumes of data and identify trends from historical patterns. This technology enables the analysis of years' worth of data in a more efficient manner to identify patterns and provide actionable insights that can guide traders towards making informed decisions (Ramprasath, E., Manojkumar, P., & Veena, P., 2015). To analyze the stock market and forecast the future behavior of a stock, both technical and fundamental analysis methods can be employed. The fundamental analysis approach evaluates stocks by estimating their intrinsic value, which considers several factors apart from assets and liabilities, such as the economic conditions, financial stability, management strategies, and other relevant factors. Contrarily, technical analysis makes predictions about a stock by observing statistical trends related to it, including its closing and opening prices, high and low prices, and volume traded. The price of the stock is assumed to have already taken into account all fundamental variables (Harikrishnan, R., Gupta, A., Tadanki, N., Berry, N., & Bardae, R., 2021).

Various machine learning techniques, such as Support Vector Machine (SVM), Linear Regression, Long Short-Term Memory (LSTM), Random Forest, Artificial Neural Network (ANN), and many more have been utilized for the development of prediction models in the stock market. These prediction models were developed using a variety of stock price influencing elements (Gupta, I., Sharma, V., Kaur, S., & Singh, A. K., 2022). Out of all the different factors, the prediction based on historical data is the one that is most important. The model must be more effective by taking into account sentiment data (financial news, user tweets), along with historical data. It is because the sentiment data takes current events into account, such as any natural disaster, a new government policy, new foreign investments, etc. (Gupta, I., Sharma, V., Kaur, S., & Singh, A. K., 2022).

II. Literature Review

In a recent study (Bansal, M., Goyal, A. and Choudhary, A., 2022) the application of machine learning and deep learning algorithms for stock price prediction in the Indian stock market was explored. The research compared the effectiveness of several algorithms, including K-Nearest Neighbors, Linear Regression, Support Vector Regression, Decision Tree

Regression, and Long Short-Term Memory. The research used a dataset of stock prices for 12 companies over the past 7 years. The study concludes that the deep learning algorithm outperforms all other algorithms, providing the most accurate results for stock price or time series prediction. The paper also highlights more efficient and robust techniques for forecasting trends in the stock market**.**

In a study conducted (Shen, S., Jiang, H. and Zhang, T., 2012), a novel prediction algorithm was introduced that leverages the temporal correlation among global stock markets and financial products to forecast the next-day trend in the stock market. The algorithm employed Support Vector Machine (SVM) and achieved high prediction accuracy of 74.4% in NASDAQ, 76% in S&P500, and 77.6% in DJIA. Additionally, the study employed various regression algorithms to track the actual increment in the markets and established a simple trading model to assess the performance of the proposed algorithm against other benchmarks.

In the paper (Singh, S., Madan, T.K., Kumar, J. and Singh, A.K., 2019), it is noted that forecasting stock prices is a difficult task due to the volatile nature of the stock market and external factors such as news events and public sentiments. Machine learning algorithms have been employed to create prediction models that can accurately predict stock prices and market trends. The study reviews several algorithms, including support vector machine, deep learning, random forest, boosted decision trees, and ensemble methods, and their applications in building prediction models for various stock exchanges. The paper also discusses the challenges encountered while constructing such models.

The paper (Gite et.al.,2021) shows stock market is unpredictable and influenced by positive/negative sentiments from media releases. Accurately predicting stock prices is a challenging task. The research paper suggests the integration of advanced machine learning methods with the deep learning approach of Long Short-Term Memory (LSTM) to achieve precise forecasts of stock prices. The sentiment derived from news headlines influences traders, so the proposed model also incorporates sentiment analysis to improve prediction accuracy. By incorporating historical stock data and news sentiments, the LSTM model is capable of generating a more accurate predictive model. The LSTM algorithm is particularly beneficial for predicting temporal data that exhibits long-term dependencies.

The study (Srijiranon, K., Lertratanakham, Y., and Tanantong, T., 2022) presented a research article that proposed a hybrid prediction model for forecasting the closing price of the stock market in Thailand. The model combined Long Short-Term Memory (LSTM), Empirical Mode Decomposition (EMD), & Principal Component Analysis (PCA) methods. The study also incorporated news sentiment analysis using FinBERT to enhance the model's performance. The study analyzed data from 2018-2022 and evaluated the model's performance using various statistical indicators. The results showed that the proposed framework outperformed baseline methods in predicting stock market prices, and the incorporation of news sentiment analysis improved the LSTM model's performance.

The research, (Indika, A., Warusamana, N., Welika, E. and Deegalla, S., 2021) focused on comparing the effectiveness of various heterogeneous ensembles in financial forecasting for predicting the next day's closing stock price and categorizing it as either HIGH or LOW. They employed regression to predict the stock price and used classification to label it as either HIGH or LOW. To evaluate the performance of the ensembles, they utilized individual models such as LSTM, LR, and SVM. The researchers collected data from the top 20 most active companies in the NASDAQ stock exchange for ten years. The study revealed that the blending ensembles approach outperformed other methods, while SVM underperformed, and LSTM produced satisfactory results, with linear regression showing promising outputs.

III. Models Used

A. Lineer Regression

Ridge regression, a regularization technique used in linear regression, has been applied to stock market prediction. The linear relationship between dependent and independent variables in stock market data can be established using the Ridge regression algorithm, which can handle multicollinearity and improve model accuracy. This technique adds a penalty term to the cost function of linear regression, which reduces the influence of features with low predictive power and prevents overfitting. Ridge regression has been used in combination with other techniques such as principal component analysis and autoregressive integrated moving average to improve stock market prediction performance. However, the algorithm's performance can be affected by outliers in the dataset.

B. Support Vector Regression

Support Vector Regression (SVR) is a popular machine learning algorithm used in predicting stock market prices. It is a regression-based variant of the Support Vector Machine (SVM) algorithm that is well-suited for regression problems. SVR identifies a hyperplane with the maximum margin such that most data points are within those margins. The hyperplane with the maximum number of points is chosen as the best fit line. Support vectors, which are extreme vector points, are used to create an appropriate hyperplane. SVR is effective in predicting stock market prices because it can handle non-linear relationships between dependent and independent variables. Studies have shown that it performs well in predicting both short-term and long-term stock prices.

C. Long Short-Term Memory (LSTM) Algorithm

LSTM is a popular deep learning algorithm that has been extensively used for predicting stock prices in the financial markets. Its ability to capture long-term dependencies and handle noisy and non-linear data makes it an ideal choice for this task. LSTM-based models can analyse historical stock price trends and patterns, identifying potential buying or selling opportunities. The algorithm can also process vast amounts of data in real-time, allowing for more accurate and timely predictions. LSTM has been applied to a wide range of stock market prediction tasks, including portfolio management, price forecasting, and risk analysis, and has shown promising results

in improving investment strategies and decision-making processes.

D. Random Forest

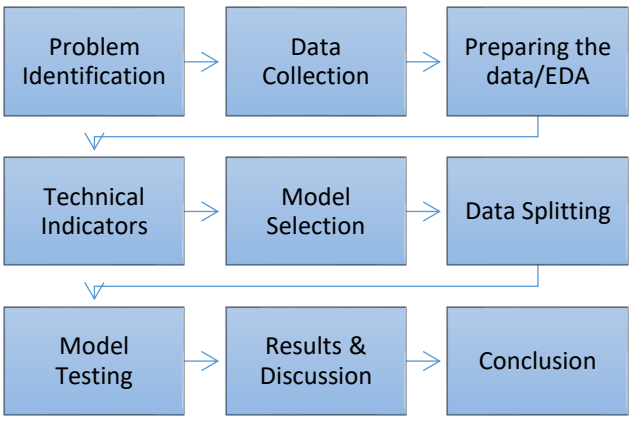
Random Forest is a popular supervised learning algorithm that uses a collection of decision trees to make predictions. It is well-suited for handling high-dimensional data and can capture complex interactions between features. In stock market prediction, Random Forest can be used to identify relevant features that impact stock prices and to make predictions on future prices. By using multiple decision trees, the algorithm can reduce the risk of overfitting and improve the accuracy of predictions. Random Forest has been shown to be effective in predicting stock prices and outperforms other traditional machine learning algorithms. It is a widely used algorithm in finance and has been used for risk management, portfolio optimization, and trading strategy development.

IV. Dataset

We have used Y Finance dataset for training and testing of models. The Y Finance dataset for NIFTY is a popular financial dataset that provides historical data for various stocks and indices traded on the National Stock Exchange of India (NSE). This dataset contains data for the NIFTY 50 index, which represents the performance of the top 50 companies listed on the NSE. The data includes daily stock prices, trading volume, and other market indicators such as open price, close price, high price, low price, and adjusted close price. This dataset is widely used in financial research and machine learning projects for stock market prediction and analysis. It provides a rich source of data that can be used to build predictive models and make informed investment decisions. The data can be easily downloaded from the Y Finance API and processed using Python libraries such as Pandas and NumPy.

Technical indicators such as moving average, volatility, returns are included in the nifty50 data, which are obtained using TA-Lib library which is a very well-known library for technical analysis.

V. Methodology



Problem Identification: The research problem is identified as the prediction of stock market trends using machine learning algorithms.

Data Collection: Data is collected from NIFTY, which is an index of the National Stock Exchange of India. The data includes historical stock market prices and other relevant variables.

Data Preparation: The collected data is pre-processed, cleaned, and transformed into a format suitable for analysis.

Model Selection: Four machine learning algorithms (LSTM, Support Vector Regression, Ridge Regression, and Random Forest) are selected for analysis.

Data Splitting: The dataset is divided into training and testing sets to evaluate the performance of the selected models.

Feature Selection: Relevant features are selected using techniques such as correlation analysis and feature importance scores.

Model Testing: The selected models are trained on the training set and tested on the testing set to evaluate their performance.

Result and Discussion: The results of the analysis are presented and discussed in detail, highlighting the strengths and weaknesses of each model.

Conclusion: Based on the analysis, conclusions are drawn regarding the effectiveness of each machine learning algorithm in predicting stock market trends.

VI. Results and Discussion

Before hyper tuning and cross-validation the Ridge Regression, SVR and RF showed MAE, RMSE, NRMSE and R^2 values as shown in the table 1.

The LSTM model metrices based on 50 number of neurons and 3 layers scores were better than other models.

Model	Mean Absolute Error	RMSE	NRMSE	R^2
Ridge Regression	0.183	0.236	0.242	0.165
SVR	0.113	0.164	0.167	0.598
Random Forest Regression	0.100	0.140	0.140	0.750
LSTM	0.081	0.105	0.108	0.840

After hyper tuning and cross-validation the Ridge Regression, SVR and RF showed MAE, RMSE, NRMSE and R^2 values as shown in the table 2.

Model	Mean Absolute Error	RMSE	NRMSE	R^2
Ridge Regression	0.183	0.238	0.251	0.162
SVR	0.107	0.143	0.146	0.702
Random Forest Regression	0.090	0.130	0.140	0.760
LSTM	0.081	0.105	0.108	0.840

The results suggest that the LSTM model outperforms other models in predicting stock prices, with an MAE of 0.081 and an R2 score of 0.84. The random forest model also performed well with an MAE of 0.1 and an R2 score of 0.75. However, after hyper-tuning and cross-validation, the SVR model showed a significant improvement in its performance, with an MAE of 0.107 and an R2 score of 0.702, which was comparable to the random forest model. The Ridge Regression model performed the worst among all models. These results suggest that LSTM and random forest models are strong candidates for predicting stock prices, and SVR can also be a viable option with proper hyper-tuning and cross-validation. However, it is essential to note that the performance of these models may vary with different datasets and market conditions.

VII. Conclusion

In conclusion, this project aimed to predict stock market prices using various machine learning techniques, including Ridge Regression, Support Vector Regression, Random Forest Regression, and Long Short-Term Memory (LSTM). The results showed that the LSTM model with 50 neurons and three layers along with one output layer outperformed the other models, with the lowest MAE, RMSE, and NRMSE values and the highest R2 value. However, after hyperparameter tuning and cross-validation, the performance of Ridge Regression, SVR, and Random Forest Regression models improved. The project's findings demonstrate the potential of machine learning algorithms, particularly deep learning techniques such as LSTM, in predicting stock prices. This could have significant implications for investors, traders, and financial analysts looking to make informed decisions.

Despite the promising results, this study has some limitations. First, it only used data from one stock market, which limits its generalizability to other stock markets. Second, the study considered a small number of features, and including more features could improve the performance of the models. Finally, the study only considered technical analysis and it did not consider the potential impact of external factors such as geopolitical events, economic indicators, and natural disasters on the stock market, which could also affect the models' predictive power.

References

Bansal, M., Goyal, A. and Choudhary, A., 2022. Stock Market Prediction with High Accuracy using Machine Learning Techniques. Procedia Computer Science, 215, pp.247-265.

Gite, S., Khatavkar, H., Kotecha, K., Srivastava, S., Maheshwari, P. and Pandey, N., 2021. Explainable stock prices prediction from financial news articles using sentiment analysis. PeerJ Computer Science, 7, p.e340.

Gupta, I., Madan, T. K., Singh, S., & Singh, A. K., 2022. HiSA-SMFM: historical and sentiment analysis-based stock market forecasting model. arXiv preprint arXiv:2203.08143.

Gupta, I., Sharma, V., Kaur, S., & Singh, A. K., 2022. PCA-RF: An Efficient Parkinson's Disease Prediction Model based on Random Forest Classification. arXiv preprint arXiv:2203.11287.

Han, X., & Li, Y., 2017. Can investor sentiment be a momentum time-series predictor? Evidence from China. Journal of Empirical Finance, 42, 212-239.

Harikrishnan, R., Gupta, A., Tadanki, N., Berry, N., & Bardae, R., 2021. Machine Learning Based Model to Predict Stock Prices: A Survey. In IOP Conference Series: Materials Science and Engineering (Vol. 1084, No. 1, p. 012019). IOP Publishing.

Indika, A., Warusamana, N., Welika, E. and Deegalla, S., 2021. Ensemble Stock Market Prediction using SVM, LSTM, and Linear Regression.

Kalyani, J., Bharathi, P., & Jyothi, P., 2016. Stock trend prediction using news sentiment analysis. arXiv preprint arXiv:1607.01958.

Ramprasath, E., Manojkumar, P., & Veena, P., 2015. Analysis of Direct Current Motor in LabVIEW. International Journal of Electrical and Computer Engineering, 9(5), 1146-1149.

Shen, S., Jiang, H. and Zhang, T., 2012. Stock market forecasting using machine learning algorithms. Department of Electrical Engineering, Stanford University, Stanford, CA, pp.1-5.

Singh, S., Madan, T.K., Kumar, J. and Singh, A.K., 2019, July. Stock market forecasting using machine learning: Today and tomorrow. In 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (Vol. 1, pp. 738-745). IEEE.

Srijiranon, K., Lertratanakham, Y. and Tanantong, T., 2022. A Hybrid Framework Using PCA, EMD and LSTM Methods for Stock Market Price Prediction with Sentiment Analysis. Applied Sciences, 12(21), p.10823.

Wang, B., Huang, H., & Wang, X., 2012. A novel text mining approach to financial time series forecasting. Neurocomputing, 83, 136-145.

Project concept: Rahul Kumar Yadav
Data Collection: Rahul Kumar Yadav, Abhishek Raghuwanshi & Sarita Kumari
Coding: Rahul Kumar Yadav
Experimentation: Rahul Kumar Yadav & Abhishek Raghuwanshi
Report Writing: Abhishek Raghuwanshi & Sarita Kumari
Video Making: Rahul Kumar Yadav