IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

FORECASTING OF STOCK PRICES USING NEURAL NETWORK AND RANDOM FOREST REGRESSION

[1] **Ms. Y. Navrin Jumana**,
[1] STUDENT, DEPARTMENT OF MBA,
[1]CARE COLLEGE OF ENGINEERING,
Tiruchirappalli-09, TAMILNADU, INDIA.

[2]**Dr.S. Manochandar,***
[2] HEAD OF THE DEPARTMENT
[2] DEPARTMENT OF MBA,
[2] CARE COLLEGE OF ENGINEERING,
Tiruchirappalli-09, TAMILNADU, INDIA.

Abstract: This paper presents a study on the use of Machine Learning Models for predicting stock prices. In this study, we used Neural Network and Random Forest Regression to predict stock prices based on historical data. The data used in this study is the daily stock prices of top five NIFTY50 companies for the past one year. The models were trained and tested on the pre-processed data. The study demonstrates the effectiveness of machine learning for predicting stock prices. The findings of this study could be beneficial for investors and traders who are interested in predicting stock prices and making informed investment decisions.
KeyWords: Prediction; Neural Network; Random Forest; NIFTY; Time Series

I. INTRODUCTION

The use of machine learning in finance has gained significant attention due to its potential to predict stock prices with high accuracy. The objective of this project is to identify and compare the performance of various machine learning models for predicting the stock prices of the top five companies in the Nifty 50 index, and to determine the most accurate and reliable model for future predictions.

Stock market predictions are complex, and traditional forecasting methods are not always accurate. Therefore, there is a need for more precise and reliable methods for predicting stock prices to help investors in making better decisions.(Bhatti, U. & Hanif. M. (2010))

The scope of this project includes selecting the top 5 companies in the Nifty 50 index based on their market capitalization on each sector and analyzing their stock price data over a last year. Machine learning models such as neural networks and random forest will be trained and compared based on their accuracy, reliability, and efficiency.

The deliverables of this project include a comparative analysis of various machine learning models for predicting stock prices, a recommendation of the most accurate and reliable model, and a future prediction of the stock prices of the selected companies based on the recommended model. This project has the potential to aid investors in making informed decisions and to contribute to the development of more accurate and reliable methods for predicting stock prices.

II. LIERATURE REVIEW

Cheng, Huang, and Wu (2018) proposed an innovative model; an attention-based Long and Short Term Memory model which is part of Machine Learning technique to predict stock price movement and to enhance the investment decision making process.

Chen and Ge (2019) investigated the attention mechanism in Long and short term memory network specifically adopted to predict stock price movement in Hong Kong stock trading.

Polamuri, S. R., Srinivas, K., & Mohan, A. K. (2019) predicts the stock market prices using random Forest and Extra Tree Regression. The results obtained from this study clearly states the decision tree and random forest regressor provides 80% of accuracy in the predicting stock market price.

Wang, C., Chen, Y., Zhang, S., & Zhang, Q. (2022) Deep Learning plays a vital role in predicting stock market prices specifically, convolutional neural networks, and recurrent neural networks. The experimental analysis was conducted on main stock market indices, including CSI 300, S&P 500, Hang Seng Index, and Nikkei 225. The results clearing showed that the Deep transformer model provides better earnings from the investors compare to the classical methods.

III. RESEARCH METHODOLOGY

3.1Population and Sample

The top five companies of each sector are chosen based on the Nifty 50 index and their market capitalization. The selected constituents by weightage NIFTY50 are:-

Table 3.1 Summary of the Stock price Data

Company's Name	Weight (%)	Sector
Reliance Industries Ltd.	10.41	OIL & GAS
ICICI Bank Ltd.	7.44	BANK
Infosys Ltd.	7.2	IT
Housing Development Finance Corporation	6.06	FINANCIAL SERVICES
Tata Consultancy Services Ltd.	4.41	IT

3.2 Data and Sources of Data

For this study secondary data has been collected. From the website of moneycontrol.com. The stock price list of the selected companies had been collected for the time period of 01 FEB 2022 to 22 FEB 2023.

3.3 Data Preprocessing

STATIONARITY CHECK

Stationarity means that the statistical properties such as mean, variance etc. of a time series do not change over time. Many analytical tools and statistical tests and models rely on the stationarity. As such, the ability to determine whether a time series is stationary state is important. Corellogram is used to check the stationary of the data.

CORELLOGRAM

A correlogram is a graphical representation of the relationship between variables in a dataset. It is also known as a correlation matrix or a correlation plot. It is a matrix of pairwise correlations between variables, where each cell in the matrix represents the correlation coefficient between two variables. The correlation coefficient is a measure of the strength and direction of the linear relationship between two variables, ranging from -1 to +1. A value of -1 indicates a perfect negative correlation, a value of +1 indicates a perfect positive correlation, and a value of 0 indicates no correlation. A correlogram is typically visualized using a heat map or a color-coded matrix, where the color of each cell indicates the strength and direction of the correlation between two variables. Correlograms are commonly used in exploratory data analysis and can provide insights into the relationships between variables in a dataset. They can also be used to identify potential multicollinearity issues in regression analysis, where two or more independent variables are highly correlated with each other. Overall, correlograms are a useful tool for understanding the structure and relationships within a dataset.

NEURAL NETWORK

A machine learning model built based on the model of human nervous system with a series of mathematical and physical processes. Neural networks have proven to be highly effective in various domains, empowering IT professionals since their inception in the 1950s to accomplish diverse tasks, including decision-making.The set of algorithms in neural network can recognise the patterns and make predictions according to those patterns. It has three layers namely input, output and hidden layer.

Input Layer: - The values fed into the network

Output Layer: - The final layer that produces predicted results.

Hidden Layer: - It has a set of nodes which are interconnected, it represents mathematical operations that process the input data. These are called neurons. These hidden layers are so powerful they make complicated calculations and determine the output

Neural networks has the ability to learn from past events and makes them ideal for predicting. To improve the accuracy of prediction these neural networks will adjust the strength of connection between each nodes during training.

RANDOM FOREST REGRESSION

Random forest regression is a powerful machine learning technique that offers numerous benefits across various applications. This approach, which has gained significant popularity in recent years, involves constructing an ensemble of decision trees to make accurate predictions and handle complex regression tasks. By combining the predictions of multiple decision trees, random forest regression overcomes some of the limitations of individual trees, such as over fitting and sensitivity to outliers. Each decision tree in the ensemble is trained on a different subset of the data, using a random selection of features. This diversity in training allows the ensemble to capture a broader range of patterns and relationships present in the data.

Random forest regression is known for its robustness and ability to handle high-dimensional datasets with numerous features. It can effectively handle missing data and is less prone to over fitting compared to other regression techniques. Additionally, random forests provide valuable insights into feature importance, allowing practitioners to understand the relative contributions of different variables in the prediction process. The versatility of random forest regression extends to various domains, including finance, healthcare, and ecology. In finance, it can be employed for tasks such as stock market prediction and credit risk assessment. In healthcare, random forest regression can aid in disease diagnosis and personalized treatment recommendations. In ecology, it can help analyze environmental data and predict species distributions.

IV. RESULTS AND DISCUSSION

4.1 Results of Descriptive Statics of Study Variables

RELIANCE INDUSTRIES LTD.

Table 4.1: Results obtained from Neural Network

Mean absolute error	33.1844
Root mean squared error	44.5604
Relative absolute error	107.697 %
Root relative squared error	108.208 %
Total Number of Instances	263

Table 4.2: Results obtained from Random Forest Regression

Mean absolute error	38.1534
Root mean squared error	49.3714
Relative absolute error	123.8234 %
Root relative squared error	119.8907 %
Total Number of Instances	263

ICICI BANK LTD.

Table 4.3: Results obtained from Neural Network

Mean absolute error	9.341
Root mean squared error	11.8519
Relative absolute error	106.9096 %
Root relative squared error	104.4101 %
Total Number of Instances	263

Table 4.4: Results obtained from Random Forest Regression

Mean absolute error	10.2872
Root mean squared error	13.0292
Relative absolute error	117.7402 %
Root relative squared error	114.7815 %
Total Number of Instances	263

INFOSYS LTD.

Table 4.5: Results obtained from Neural Network

Mean absolute error	21.9369
Root mean squared error	28.2862
Relative absolute error	106.4251 %
Root relative squared error	106.3855

Table 4.6: Results obtained from Random Forest Regression

Mean absolute error	26.1244
Root mean squared error	33.1561
Relative absolute error	126.7403 %
Root relative squared error	124.7013 %

HOUSING DEVLOPMENT FINANCE CORPORATION

Table 4.7: Results obtained from Neural Network

Mean absolute error	32.9005
Root mean squared error	43.4895
Relative absolute error	107.1768 %
Root relative squared error	104.6444 %

Table 4.8: Results obtained from Random Forest Regression

Mean absolute error	37.4165
Root mean squared error	50.9408
Relative absolute error	121.8883 %
Root relative squared error	122.5739 %

TCS

Table 4.9: Results obtained from Neural Network

Mean absolute error	36.7858
Root mean squared error	50.139
Relative absolute error	103.6829 %
Root relative squared error	103.7452

Table 4.10: Results obtained from Random Forest Regression

Mean absolute error	44.887
Root mean squared error	58.137
Relative absolute error	126.5168 %
Root relative squared error	I.2943

V. CONCLUSION

In evaluation of machine learning regression models for stock price prediction, lower values for Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) indicate better performance, as they represent the average magnitude of the prediction errors. Similarly, the relative errors show the percentage difference between the predicted and actual values, with lower values indicating better accuracy.

Comparing the metrics of the two models, we can observe the following:

- The Random Forest Regression (RFR) model has the highest MAE and RMSE values among the models, indicating the highest prediction errors.
- Neural Network has the lowest MAE and RMSE values among the models, indicating better predictive performance in terms of accuracy and goodness of fit.

Therefore, Neural Network provides better performance in predicting of stock prices compare to Random Forest Regression.

REFERENCES

[1] https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

[2] https://www.nseindia.com/

[3] Cheng, L. C., Huang, Y. H., & Wu, M. E. (2018, December). Applied attention-based LSTM neural networks in stock prediction. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 4716-4718). IEEE.

[4] Chen, S., & Ge, L. (2019). Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction. Quantitative Finance, 19(9), 1507-1515.

[5] Polamuri, S. R., Srinivas, K., & Mohan, A. K. (2019). Stock market prices prediction using random forest and extra tree regression. Int. J. Recent Technol. Eng, 8(1), 1224-1228.

[6] Wang, C., Chen, Y., Zhang, S., & Zhang, Q. (2022). Stock market index prediction using deep Transformer model. Expert Systems with Applications, 208, 118128.

[7] Bhatti, U. and Hanif. M. 2010. Validity of Capital Assets Pricing Model. Evidence from KSE-Pakistan "European Journal of Economics, Finance and Administrative Science", 3 (20).