Stochastic Modelling and Computational Sciences

COMPARATIVE ANALYSIS FOR INDIAN RAILWAY FINANCE CORPORATION STOCK PRICE USING MACHINE LEARNING TECHNIQUES

Krishna Veni.M[1], Arumugam. P[2] and Pradeep Nijanthan. P[3]
Research Scholar[1], Professor[2] and Research Scholar[3], Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu
krish20393@gmail.com[1], sixfacemsu@msuniv.ac.in[2] and pradeepnijanthanp@gmail.com[3]

ABSTRACT
The goal of stock market prediction is to forecast the future worth of a company's financial stocks. The most recent development in stock market forecasting technology is the application of machine learning, the historical values of stock market indexes to generate forecasts based on their present values. Various models are used in machine learning itself to facilitate and authenticate prediction. This study focuses on comparing the machine learning algorithms Multiple Linear Regression, Random Forest Regression and K- Nearest Neighbor Regression to forecast stock values. Open, close, low, high, and volume are the factors are considered. The models are evaluated using two popular strategic metrics, the R^2 score and Root Mean Square Error (RMSE). Lower values of these two metrics indicate that the trained models are more effective. This proposed study examines the practical applications of the prediction system and provides a machine-learning model to estimate the duration of stocks market price existence, while also addressing concerns regarding the accuracy of the total value. In addition to providing useful solutions for the issues faced by stock capitalists, accurate stock forecasting would surely be extremely advantageous for stock exchange governments.

***Keywords**: Multiple Linear Regression, Random Forest Regression, K-Nearest Neighbor, Root Mean Square Error.*

1. INTRODUCTION
The Indian government created the Indian Railway Finance Corporation (IRFC), a government-owned financial organization, to collect money for the Indian Railways. The main responsibility of IRFC is to finance rolling equipment (wagons, locomotives, etc.) and other infrastructure projects for Indian Railways. This includes leasing assets and taking out loans from the debt and stock markets. The Ministry of Railways is in charge of IRFC, which was established in 1986. The company is a major participant in the Indian banking and transportation sectors because of its vital role in the development and upgrading of the country's rail network. IRFC made its debut on the Indian stock market in 2021 with its Initial Public Offering (IPO), giving individual investors the opportunity to engage in a dependable, government-backed business. The IPO marked a significant turning point in the company's history by announcing its transition from a government-only organization to one that is more openly traded and market-driven. The stock of IRFC (symbol: IRFC on the NSE/BSE) is highly attractive to investors due to the backing of Indian Railways, which ensures a certain level of stability and security. Nonetheless, the success of the stock is influenced by several factors, including the status of the Indian economy, laws, interest rates, and the overall efficiency of the nation's rail network. Institutional and retail investors have expressed interest in IRFC's stock performance since it is frequently thought of as a comparatively low-risk investment in the infrastructure industry. However, the larger macroeconomic variables that affect the Indian economy, such inflation, interest rates, and government infrastructure spending, may have an effect on its stock price and financial stability.

1.1 Objective of the Study
The aim of this study is to use a variety of predictive models, including both contemporary machine learning techniques and conventional statistical approaches, to examine the stock performance of IRFC. This study will shed light on the variables influencing IRFC stock prices and assess how well machine learning models- Random Forest, and Linear Regression, KNN Regression predict future stock prices. The study will also investigate how several characteristics, like volume, past stock prices, and general economic factors, affect the prediction

accuracy. This study intends to advance knowledge of how the market perceives government-backed financial institutions and the main factors influencing their stock movements by analyzing IRFC's stock performance. Investors looking for useful information will find it in this study.

1.2 Related Research

Machine learning techinque for stock price prediction using Random Forest Regression and compare support vector regression, decision tree.choose more effective algorithm to estimate stock price prediction discussed by R Raksha Prabhu , Saritha Shetty[1]. . R. Harikrishnan , Shivali Amit Wagle have compared the prediction of stock price using Sentimant analysis, Twitter sentiments and Support vector machine,Linear regression, Random Forest Regression and Predective analysis.The result provide sentiment analysis into the prediction model[3].Improve the supervised learning algorithms to predict the stock price discussed by . R S Abirami , K.Varalakshmi, Maddika Jaswanth Reddy[4]. Compare the analysis of machine learning algorithm Linear,LSTM,KNN for stock prediction finally choose the effective algorthim for stock discussed by A. Sravani, Chintam Anusha, N.V.S. Shankar[7]. The importance of feature selection and data pre-processing in improving prediction model accuracy was emphasized by the authors. The linear regression machine learning algorithm used to predict stock prices was the subject of this work by Nusrat Rouf, Saurabh Singh[11].

2. METHODOLOGY

2.1 Multiple Linear Regression

To predict the value of a response variable, a statistical technique known as multiple linear regression (MLR), or simply multiple regression, uses a number of explanatory factors. The goal of MLR is to model the linear relationship between the response (dependent) and explanatory (independent) variables. many regression is basically an extension of ordinary least-squares (OLS) regression since it takes into account many explanatory variables.

$$y_{i=}\beta_0+\beta_1 x_{i1+}\beta_2 x_{i2+\_+}\beta_p\ x_{ip+\varepsilon}$$

where, for i=n observations:

y_i= dependent variable

x_i = explanatory variable

β_0= y-intercept (constant term)

β_p=slope coefficients for each explanatory variable

ε=the model's error term (also known as the residuals)

2.2 Random forest Regression:

One effective tree learning method in machine learning is the Random Forest algorithm. During the training stage, it generates a number of Decision Trees. To measure a random subset of characteristics in each partition, a random subset of the data set is used to build each tree. By introducing variety across individual trees, this randomness lowers the possibility of overfitting and enhances prediction performance overall.

Steps for Random Forest Regression Algorithm:

Step 1: Choose arbitrary samples from a training or data collection.

Step 2: For each training set of data, this algorithm will build a decision tree.

Step 3: The decision tree will be averaged for voting.

Step 4: Finally, decide which forecast outcome received the most votes to be the final one.

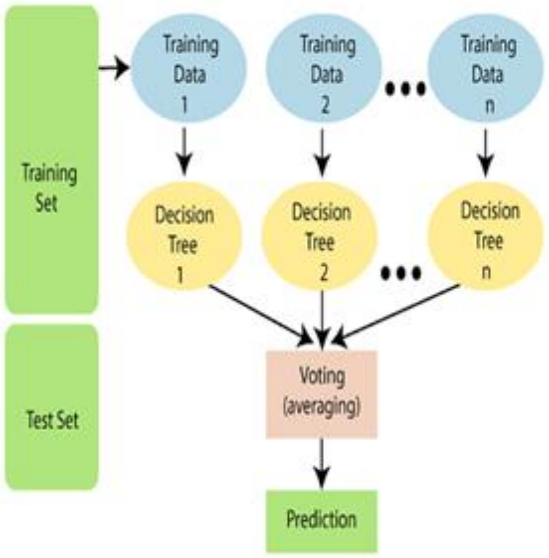

Fig 1: Working of the Random Forest Algorithm

2.3 KNN Regression: KNN regression is a non-parametric method used for predicting continuous values. The core idea is to predict the target value for a new data point by averaging the target values of the K nearest neighbors in the feature space. The distance between data points is typically measured using Euclidean distance, although other distance metrics can be used.

Steps for KNN Regression Algorithm:

1. **Choosing the number of Neighbors (K):** Choosing the number of neighbors, K, is the first stage. The model's performance is significantly impacted by this decision. While a bigger value of K produces smoother predictions, a lesser value of K makes the model more susceptible to noise.
2. **Determining Distances:** For a new data point, calculate the distance between this point and all points in the training set.
3. **Determining K Nearest neighbors:** Determine which K training set points are closest to the newly added a data point
4. **Target value Prediction:** Determine the expected value for the new data point by averaging the target values of the K nearest neighbors.

2.4 Evalution Metrices:

1. **R^2 error:** It measures the proportion of variance in the dependent variable that can be explained by the independent variable, with a larger value indicating a better model.
2. RMSE is the standard deviation of prediction errors, giving a high weight to large errors, while a smaller value indicates a better model.
3. **MAE:** The Mean absolute error measures the average difference between actual and predicted values in a dataset, encompassing the average residuals.4.MSE: Mean Squared Error measures the variance of residuals by calculating the average difference between the original and predicted values in the data set.
4. **MSE:** Mean Squared Error measures the variance of residuals by calculating the average difference between the original and predicted values in the data set.

3. DATA SOURCES:

This research on stock market forecasting in India outlines the data collection and pre-processing procedures used for ensuring the reliability and accuracy of the results.

3.1 Data Collection

The Indian Railway Finance Corporation's data, collected from NSE.com, is a widely used indicator of the Indian stock market's performance, covering 249 records from 1 January 2022 to 31 December 2022.

3.2 Data Pre-processing

The raw data was pre-processed to remove discrepancies, missing numbers, and outliers, replacing missing values with interpolated ones and normalizing the data for consistency. The data was divided into training and testing sets using an 80:20 split, and the machine learning model's performance was assessed on the testing set.

3.3 Visualize the Data

The dataset includes attributes such as date, open, high, low, close and volume. *Figure: 1* displays the closing price over a year.

Fig 1: Closing price over a year

This graph shows the "Stock Close Price Trend" over time. The horizontal axis (X-axis) represents the timeline, spanning from January to December 2022, while the vertical axis (Y-axis) shows the closing price, ranging from 20 to 35. Blue dots mark the individual daily closing prices, and a connecting blue line indicates the trend over time.

4. RESULTS

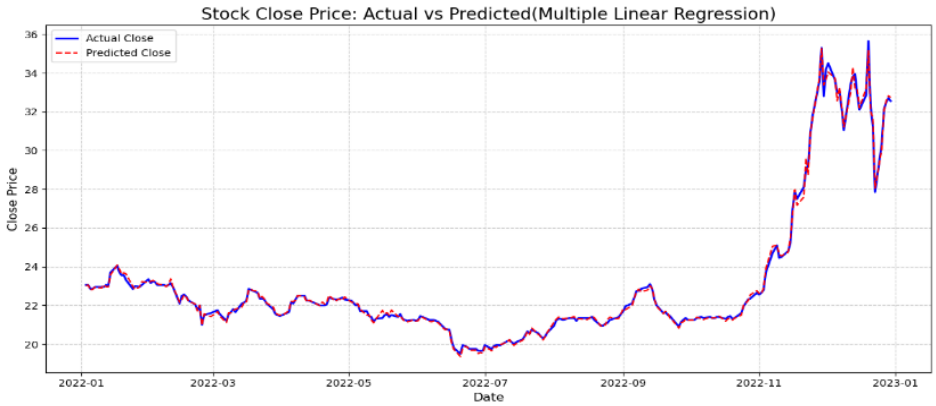

Fig 2: Actual vs Predicted Values for Stock Close price- Multiple Linear Regression

The graph (Fig 2) showing the comparison of actual and predicted stock closing prices over a specific time using a Multiple Linear Regression model. The blue line denotes the actual closing prices noted in the market, while the red dashed line displays the prices predicted by a forecasting model. The model's performance in predicting stock prices is indicated by the close alignment between these lines.

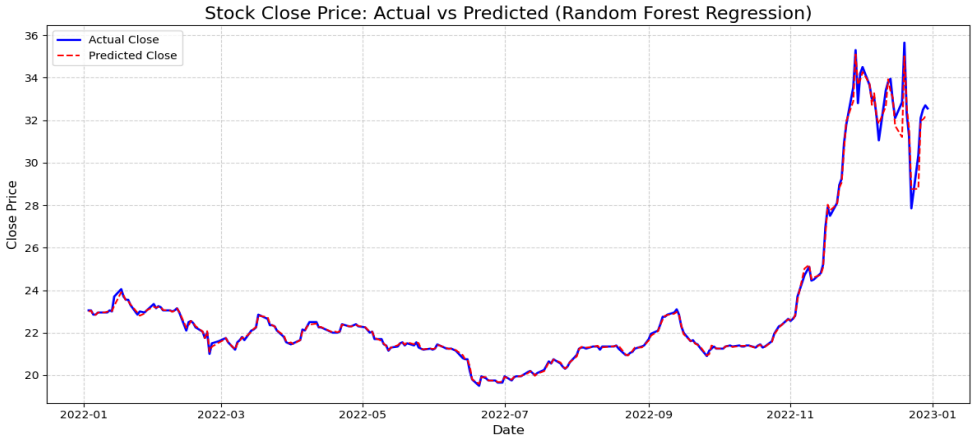

Fig 3 Actual vs Predicted Values for Stock Close Price- Random Forest Regression

The graph (Fig 3) compares actual and predicted stock closing prices from January 2022 to December 2022, using a Random Forest model, with blue lines representing market observations and red dashed lines representing predicted prices. The Random Forest model accurately predicts stock price trends, with minimal deviations between actual and predicted prices. This high prediction accuracy makes it a valuable tool for market analysis, despite minor deviations.

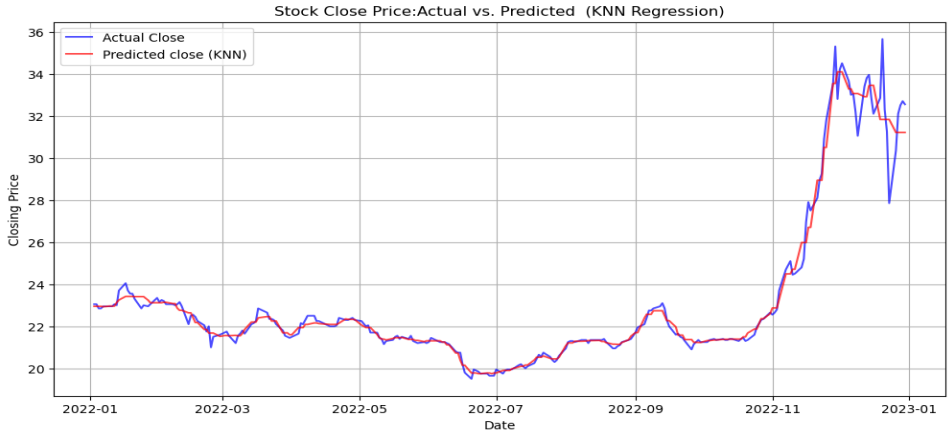

Fig 4: Actual vs Predicted Values for Stock Close-Knn Regression

This graph (Fig 4) compares the **actual stock price trends** (blue line) with the **predicted trends** (red line) using the **K-Nearest Neighbors (KNN) regression model**. The x-axis represents the time period from January 2022 to November 2022, while the y-axis shows the stock's closing prices. The KNN regression model effectively forecasts stock prices, accurately capturing significant market movements and predicting trends with minor discrepancies, despite minor discrepancies.

This research on stock market forecasting in India utilizes machine learning techniques, comparing the evaluation metrics for all three algorithms. Table 1 displays the MAE,MSE,RMSE, R^2 of all three models.

Table 1: Result of Accuracy

Regression	MAE	MSE	RMSE	R^2
Multiple Linear Regression	0.11	0.03	0.17	0.99
Random Forest Regression	021	0.16	0.40	0.99
KNN Regression	0.33	0.30	0.55	0.98

4.1 Multiple Linear Regression (MLR)

MLR achieved the best performance with the lowest MAE (0.11), MSE (0.03), and RMSE (0.17). The R^2 score of 0.99 indicates that the model explains 99% of the variance in the target variable. These results demonstrate MLR's ability to accurately predict stock closing prices with minimal error.

4.2 Random Forest Regression (RFR)

RFR performed comparably well, with an R^2 of 0.99, showing it explains the same variance as MLR. However, it had higher errors (MAE: 0.21, MSE: 0.16,

RMSE: 0.40), suggesting that while RFR is robust, it is slightly less precise than MLR for this dataset.

4.3 KNN Regression (KNN)

KNN Regression underperformed relative to the other models, with the highest MAE (0.33), MSE (0.30), and RMSE (0.55). Although its R^2 value of 0.98 is still high, the errors indicate that KNN struggles to predict precise stock prices, likely due to challenges in identifying accurate neighbors in this dataset.

5. CONCLUSION

This study compared the performance of three regression models for stock price prediction: Multiple Linear Regression, Random Forest Regression, and KNN Regression. The results show that Multiple Linear Regression outperforms the other models, achieving the lowest error metrics and the highest predictive accuracy. Random Forest Regression is also highly effective but less precise than MLR, while KNN Regression performs the least effectively due to higher error metrics. These findings highlight the importance of selecting appropriate models based on the dataset's characteristics. Future research could explore additional models, such as Support Vector Regression or neural networks, to further enhance predictive performance.

REFERENCES

1. R Raksha Prabhu , Saritha Shetty et al," STOCK PRICE PREDICTION USING RANDOM FOREST REGRESSION", Volume:5,Issue 05 May 2023.
2. Subba Rao Polamuri, K. Srinivas, A. Krishna Mohan et al,"Stock Market Prices Prediction using Random Forest and Extra Tree Regression", Volume:8,Issue 3,September 2019.
3. R. Harikrishnan , and Shivali Amit Wagle et al," A Comparative Study on Different
4. Machine Learning Algorithms for Predictive Analysis of Stock Prices", July 2023, https://doi.org/ 10.1007/978-981-99-1479-1_44.
5. R S Abirami , K.Varalakshmi, Maddika Jaswanth Reddy et al," Stock Market Price Prediction Using Random Forest And Support Vector Machine", Volume: 10,Issue 6,June 2022.
6. Ghanashyam Vagale, Matur Rohith Kumar, Bhanuprakash Darbha, Durga Shankar Dalayi et al," Stock Price Prediction Using Machine Learning [Random Forest Regerssion Model]", Volume: 10,Issue 7,July 2023.
7. Shreya Pawaskar et al," Stock Price Prediction Using Machine Learning Algorithms", Volume10,Issue 1, January 2022.
8. A. Sravani, Chintam Anusha, N.V.S. Shankar et al," A Comparative Analysis of Machine Learning Algorithms in Stock Prediction ", February 2022.

9. Amit Gupta, T.J.Nagalakshmi et al," Stock Price Prediction using Linear Regression in Machine Learning",ISSN: 2278-3075,Volume: 8,Issue-12,October 2019.
10. Anderson, B. and Ledolder, J.: Statistical Methods for Forecasting, John Wiley & Sons, NewYork, 1983.
11. Tae Kyun Lee et al, "Global Stock Market Investment Strategies Based On Financial Network Indicators Using Machine Learning Techniques", Volume: 117, Issue: 1, Date: 2019.
12. S. Shukla and B.Shah's "Stock Price Prediction Using Multiple Linear Regression Models"International Journal of Engineering Science Invention (IJESI). Published in October 2018.