Proceedings of ICFTBA 2024 Workshop: Finance in the Age of Environmental Risks and Sustainability
DOI: 10.54254/2754-1169/99/2024OX0209

A Comparative Study of NIFTY-50 Prediction Using Linear Regression and Regression Tree Models

Yunshu Tang[1,a,*]

[1]*Faculty of Arts and Sciences, Beijing Normal University, Zhuhai, Zhuhai, 519085, China*
a. 202011059272@mail.bnu.edu.cn
**corresponding author*

Abstract: It is important for investors to predict the tendency of stock. However, the stock market fluctuation is too hard to predict. A single stock may be inflected by many elements, so it would be much more feasible to study a stock index. A stock index may be decided by many stocks, which means that it would not be effected by some only one company. In this passage, NIFTY-50 would be analyzed to make a prediction. The linear regression and regression tree are used to predict the close price of NIFTY-50. To make sure the data size is big enough, the data includes the price from 2000 to 2021. By comparing two models, a proper model to predict stock index would be found. After the comparison of the MSE and R square index, it is easy to find that the regression tree is the model much more reliable. However, there is still long way to go before researchers find way to predict stock price precisely.

Keywords: Stock Index Prediction, NIFTY-50, Linear Regression Model, Regression Tree Model.

1. Introduction

For investors, it would be a great help to get a model to predict the tendency of stock. A suitable model can help investors reduce the risk and increase interest [1]. There have been many methods used to predict the price of stocks. However, most of them only focus on one single stock, which is not a good choice to predict. One single stock may be effected by many elements, such as the decision of the leader in company. These factors may make the price fluctuate wildly [2]. Many investors cannot get information about these factors on time. Therefore, it is impractical to predict the price of one single stock. Because of the stability of the stock index, it is much more practical to predict the price of stock index [3]. The NIFTY-50 is chosen as an example for the huge data size. To analyze the price of NIFTY-50, two models are used in the article. By comparing the linear regression and regression tree, the method which is suitable for prediction would be found. The study may make an example for other researches on stock index and help investors make wise decisions when they want to make an investment even if investors cannot get some special information. The close price is easy to find on the Internet, which would be good material for the models show in this essay. Once the investors get the close price seven days before the target date, a prediction could be made by these models.

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/99/2024OX0209

2. Literature Review

There has been many researches about NIFTY-50. Some of them use OLS regression and Granger Causality for finding the contact between the price of NIFTY-50 and some financial sector in India [4]. In some researches, the Toda-Yamamoto methodology is used for studying. By analyzing, the result about the price of the index and the prices of some stocks of some companies is gotten [5]. A research used Deep Learning based on Long Short-Term Memory shows a feasible way for the prediction of the price of NIFTY-50. The result shows a relative high degree of exactitude [6]. Even the result gotten in this research is so accurate that some investors can use that model to predict the price and make decisions for investing, there is still something could be done to make the decision easier. If one investor need to do a decision without a high-performance computer, the investor may need another model for prediction. Even some old models have performs good in prediction of the price of INFTY-50, some researchers made new models to study and predict it.

After developing the standardization method and the automated feature interactions learner, a new model is made for prediction. The high accuracy of the result shows that it performs well in predicting the price of NIFTY-50 [7]. However, whether this model would perform well in predicting other stock index is still a mystery. So there may still be something to study and improve. In some study, the LSTM strategy performs well in investing [8]. By comparing the normal invest strategy and the LSTM strategy, the result shows that it is much excellent to use the LSTM strategy to invest. However, it needs accuracy information for prediction in time, which is not easy for everyone. In another research, the long short-term network architecture is used for studying. Three variables were input into the model for comparing to find out which one is suitable to be added for prediction [9]. Long short-term memory networks as well as densely connected neural networks are both used in another research. The result shows the excellent of these model in predicting the price of stock index especially the NIFTY-50 [10]. Time series forecasting is a good choice for prediction, especially for something stock prices. A research has shown the result of it and by comparing, it is find that Exponential Smoothing Model performs best among them [11]. There is many examples of the machine learning models in researches, so it would be feasible to use machine learning to predict the price of index. The next step is to try more models for a better one or improve the model for higher accuracy.

3. Methodology

At first, getting the data from Kaggle to make sure that worth to believe. The data would be dealt by two methods and compare the result. Before building the model, the data is split into two parts. One of them contains 70% data and is used for training, the other one is used for testing.

3.1. Regression Tree

By making code, the regression tree model is built by using training data. Then the predicted price is gotten by input testing data. After comparing the difference between the predicted price and real price, the MES and R^2 for regression tree are calculated.

3.2. Linear Regression

Using the same training data, the linear regression model is built. After rejecting the factors whose P is larger than 0.05, an equation for prediction is finished. Using the equation, the predicted price is gotten. After comparing the difference between the predicted price and real price, the MES and R square for regression tree are calculated.

DOI: 10.54254/2754-1169/99/2024OX0209

3.3. Others

Figure 1 about the close price of NIFTY-50 is made. The lowest price and the largest difference are also gotten by code for analyzing.

4. Results

The result showed in the code tells that the MES of regression tree model is about 62.58 and the R Square of it is about 0.9959, while the MES of linear regression is about 230.96 and the R square of it is about 0.9847. The equation as the result of the linear regression shows that 0.9666*close7+0.0797*close1, can predict the price. The equation means that 0.0797 multiply the close price in the first day plus 0.9666 multiply the close price in the seventh day would get the predicted close price in the eighth day. It can also find in the code that the lowest close price is 108, and the highest price difference in two days is 93.3. The line chart of the close price is shown in the Figure 1.

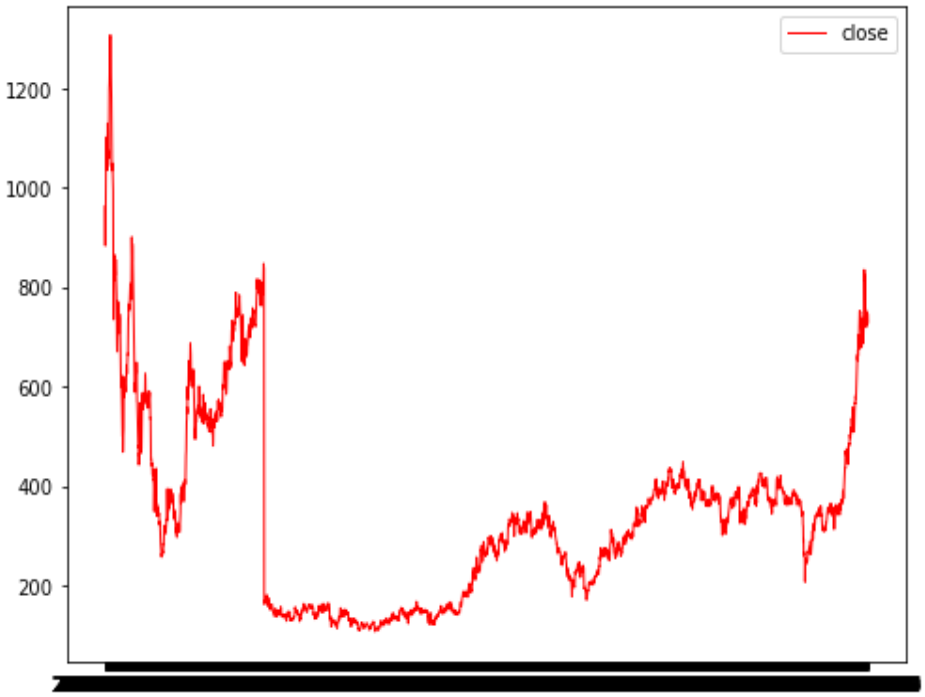

Figure 1: NIFTY-50 (Photo/Picture credit: Original).

By analyzing the results, it is easy to find that there is still a long way to go before using the model into practice, especially the line regression model. Even the R square indexes of them seem to be excellent, the large numbers of MES still show the problem of these model. The highest price difference is much lower than The MES of linear regression model, which means the difference between the predicted price and the real price might be large. By comparing the MES of them, it is easy to find that regression tree is much more suitable to predict the close price.

From the coefficient gotten from the linear regression model, it could be found that the predicted close price is hugely decided by the close price in the day before the target day. It means that the model is not as useful as the wish. Every investor knows that the difference between the close price in two days would not be very large, so the weight of independent variable in the equation shows that in this model, most other independent variable is useless when the prediction is made. It means that

DOI: 10.54254/2754-1169/99/2024OX0209

the model is relatively failed because even without this model the investors can predict that the price in the target day is close to the price in the day before that.

However, there is still something can be learned from the equation. The other coefficient, the 0.0797, some conclusion could be gotten. Even the number is not large, the exist of it shows that the price is influenced not only by the price in the day before the target day. The price in the day a week before the target day may affect the price. The result seems to mean that the long term data is not totally useless, it can also effect the close price in some way. It is not sure that whether the period of one week is important. Would this time is an important time period for price prediction? Or it is just a special case in this data or in this model? This question is worth to be study in the future. If the result means that it is an important period for prediction, it would be a help for investors to predict the close price.

By the coefficient in the equation, some tendency of India finance could be learned. Even in the Figure 1, some sharp decrease seems to be fearsome, the coefficient still tells that the India finance is not in a difficult situation. 0.0797+0.9666=1.0463, which is larger than 1. This trait means that even facing some sharp decreasing, the Indian finance is still growing slowly, so it is in a good situation. The regression tree performs much better than linear regression in the prediction. The difference of MES has proven it. However, there is still a lot of work to do in the future study. It would be better to make the regression tree visualize by code so that the investors can use the model even they cannot reach their computer. The other thing to do is to figure out which element effect the prediction more in this model.

Even the result gotten from regression tree seems to be good, there is still some problems. 62.58 is still too large for a price prediction. It is indeed that there is some great difference between the price in two days. However, only little differences is large like that. So 62.85 means that it can hardly use in prediction for the great error. It is necessary to find a better model for prediction.

5. Conclusion

After analyzing the whole study, there is still something could be done to make an improvement. Firstly, it would be better if more data can be analyzed. If the models are feasible, more data would make the degree of accuracy higher. If the models are not feasible, putting more data into it may be a great help to find out what makes the model infeasible, which would be great help to improve models or finding other models for further research in the future. So it is necessary to get more data for further study.

Selecting other models would also be an important work for further research. The values of R square show that there might be some problems to use the two models to predict the price of stock indexes. It would be a good choice to change the model or even build a model for price prediction. The using of new models may be helpful to increase the degree of accuracy and find out what factors made the values of R square high in the model used above.

The variables would also be a topic for discussion in the further research. It could not be find that all variables perform significant roles in prediction. Which variables should be given up in the future study and which variables should be added into study would be a good topic for discussion.

References

[1] *Dijiu, E. (2024). Research and implementation of stock trend prediction model based on deep neural network.*

[2] *Lv, L. (2011). Empirical study of mutual relationship between stock index future and stock market based on S&P CNX Nifty 50 Index Future of India. Hunan University.*

[3] *Shao, Z. (2018). The international comparative research on the volatility characteristics and causes of stock index futures. Jilin University.*

DOI: 10.54254/2754-1169/99/2024OX0209

[4] *Bhuvaneshwari, D. (2021, December). An analytical study of Nifty 50 and financial sector indices. In Proceedings of the First International Conference on Combinatorial and Optimization (ICCAP 2021), December 7-8, 2021, Chennai, India.*

[5] *Abinaya, P., Kumar, V. S., Balasubramanian, P., & Menon, V. K. (2016, September). Measuring stock price and trading volume causality among Nifty50 stocks: The Toda Yamamoto method. In 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1886-1890). IEEE.*

[6] *Sisodia, P. S., Gupta, A., Kumar, Y., & Ameta, G. K. (2022, February). Stock market analysis and prediction for NIFTY50 using LSTM deep learning approach. In 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM) (Vol. 2, pp. 156-161). IEEE.*

[7] *Nagula, P. K., & Alexakis, C. (2022). A novel machine learning approach for predicting the NIFTY50 index in India. International Advances in Economic Research, 28(3), 155-170.*

[8] *Seshu, V., Shanbhag, H., Rao, S. R., Venkatesh, D., Agarwal, P., & Arya, A. (2022, January). Performance analysis of Bollinger Bands and Long Short-Term Memory (LSTM) models-based strategies on NIFTY50 companies. In 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 184-190). IEEE.*

[9] *Jain, V., Jain, A., Garg, V., & Gandhi, C. (2021, July). HybridLSTM for NIFTY50 prediction using global indices and technical indicators. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.*

[10] *Garg, D., Pandey, R. K., & Tiwari, P. (2023, December). A systematic deep learning approach to forecast Nifty50 index trend. In 2023 11th International Conference on Intelligent Systems and Embedded Design (ISED) (pp. 1-5). IEEE.*

[11] *Thapar, H., & Shashvat, K. (2018, December). Predictive analysis and forecasting of S&P CNX NIFTY50 using stochastic models. In 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC) (pp. 122-127). IEEE.*