

Data Analysis of Indian Stock Market Prediction Using Machine Learning Algorithms

¹D.RAM MOHAN REDDY, ² UDDANDAM TRIVENI

¹Associate Professor, Mtech, Newton's Institute of Engineering, Macherla.

².Mtech Student, Newton's Institute of Engineering, Macherla.

Abstract: *The complexity and variety of the stock market has always attracted researchers at the place where it takes a long time to find a way to prepare for the future. However, international changes in the retail sector make this project more difficult. Statistical techniques and modeling are effective but cannot solve the major problems of economic forecasting. Traditional strategies are not able to address and advise on the complex issues that exist in the market. Learning Gadget and intellectual tools are used to solve complex problems and big data. In this article, the authors recommend the use of six best algorithms, which are general models, deep learning, decision trees, random forests, gradient boosted trees and boosting vector machine, and to identify models that are expected to be close to reality. These algorithms were implemented in BSE data analysis from April 2015 to March 31, 2020 and the model with the least error was confirmed. Of all the models used, Gradient Boosted Trees is chosen as the best, because it has the least impact and variance. Another gradient boosted tree is used to estimate the result.*

Keywords—BSE SENSEX, Big Data, Predictive algorithms, Artificial Intelligence, Random Forest, SVM, Gradient Boosted Trees, Machine Learning

I INTRODUCTION

Today, predicting price movements on the stock market has become a challenge due to the unpredictability of the business and its volatile nature. The stock market is so called because the measure of the financial system that causes the business boom is random, noisy, unpredictable, non-linear

and deterministic. [1] Many factors that make up the market, along with politics, culture and money. From a business perspective, foreign trade organizations choose which makes it difficult to predict the cost of business time. Conventionally, there are methods for anticipating trading

behavior such as the Chartism approach (technical analysis) and the intrinsic value theory (fundamental analysis). In general, these thoughts reflect the Grand Market that follows a random walk. The logic of Random Walk is mainly based on the idea that the continuous price variation does not depend on yesterday's price and the market [8]. Along the same lines, the EMH concept relies entirely on the idea that stock prices reflect all available and current information and even if buyers do not know how to obtain a cheap rate of return compared to financial experts [13]. Recently, researchers and experts around the world are using learning algorithms in various fields. Machine learning is often used to predict product prices in the market because of the process it uses to identify patterns. This item is used for important purposes use the knowledge algorithms acquired by the system to anticipate additional values of the BSE SENSEX.

A. Support vector machine

Support Vector Machine (SVM) is a classic and widely used tool to acquire knowledge of special rules by controlling decision ability, kernel functionality utilization, and response sparsity [5]. It can be used effectively for classwork and regression. The support vector statistics are close to the hyperplane and can leave the

impact of the web page and the direction of the hyperplane.

B. SVM algorithm

In SVM algorithm, we try to extend the edges of data points and hyperplane. There is a drop function also known as the hinge function which is used to maximize results.

$$L(w) = \sum_{i=1} \underbrace{\max(0, 1 - y_i[w^T x_i + b])}_{\text{Loss function}} + \underbrace{\lambda \|w\|_2^2}_{\text{regularization}}$$

Failure to do so will penalize the classifications and time constants will avoid the solution vector.

C. Random forest

Random forests are a group of tree estimation, in which the value of male or female tree depends on the value of the error random vector model [3]. The random forest algorithm works in the plethora of "a group of uncorrelated woods (patterns) working in a combination outperforms the constituent version of an individual".

D. Decision trees

In the selection tree analysis of automatic learning, selection is made from the tree structure, its branches and its nodes. Decision trees can be based on rank or regression. The classification bush is used to analyze the problem and find a solution, while the real regression tree requires a variable such as the stock price. Each node has a purpose and each branch provides

options. After the decision of the decision tree is established, the error rate is calculated to cut the decision tree [16].

E. Gradient-enhanced wood

Gradient Boosted is one of the most popular gadget learning sites. Using unemployment, classification problems can also be solved, as well as recovery problems [2].

II REVIEW OF LITERATURE

Compares four prediction models, artificial neural network (ann), support vector machine (svm), random forested area (rf) and naive bayes for reward distribution and conclude that random forested area (rf) surpasses the three different ones. predictive modeling of normal performance [15]. however, pay attention to the use of special power tools such as mlp, cnn and rnn to anticipate the rise of the product market in the s&p500 index and conclude that the connection of neural networks (cnn) is higher. . than other architectures in the prediction of primes [7] estimated and correct at 60% at the same time as estimated. price forecast for next week. additionally, logistic regression and support vector machines (svm) were confirmed to achieve slightly better results [6]. analysis which used 2 specific ans i.e. feed forward neural networks and recurrent neural networks to predict the ephemeral price of 10 stocks on

nyse showed that feed forward worked better than short term (lstm) to predict the cost percentage [10]. estimation model in predicting the delivered product. furthermore, this version is compared to the random walk model derived from efficient market (emh) speculation. their analysis concluded that the support vector model has predictive power[9]. there is also a study in which the authors investigated the value of ensuring the accuracy of money transfers through four models of gadget learning algorithms. they concluded that deep learning is more predictive than other techniques and provided a vector regression approach in 2d level neural networks and random wooded area techniques at level 1/3 [14]. this view proposed a hybrid method for estimating the value of stocks. in this hybrid approach, they incorporated the support vector regression model and the HOD RICK-PRESCOTT filter to optimize the query value estimation. they concluded that the proposed model is more accurate in predicting stock prices [12]. there are some successful reviews of the proposed multi-filter neural network (mfnn) for extracting features from financial model data and stock price forecasts. this model has been converted to the Chinese stock index csi300. they have managed to make their company perform higher in terms of cost effectiveness, accuracy and stability compared to conventional systems of knowledge models, statistical modes and

convolution networks and lstm [11]. in this article, the authors argue that in-depth research allows investors to provide good results in the case of additional products that are difficult to arbitrage. however, the performance of the control tool decreases very well in the presence of buying and selling prices due to increased trading. in addition to the in-depth research, there are additional benefits for investors in the end [4]. gadget learning knows that machine learning models outperform statistical models and business models. furthermore, it is said that the machine knows fashion better than the single subject. in this regard, the performance of the tree-based ensemble machine mastery model random forest area (rf), xgboost classifier (xg), bagging classifier (bc), ada boost classifier (ada), extra trees classifier (et) and voting classifier (vc) is compared to predict stock prices. the kendall w concordance test is used to evaluate the overall performance of tree-based ml algorithms. the extra trees classifier (et) is considered to be the first-class version in the estimation of the award-winning product [2]. however, there are other ways to examine the importance of time complexity and the authors, in their analysis, calculate the time complexity of the algorithms and supporting data.

III RESEARCH PROBLEM

This paper attempts to demonstrate and verify the forecast of market index (Sen sex) direction by applying various machine learning algorithms and secondly, to contrast the output of SVM and random forest to foresee the direction of the stock market motion and deploying the selected model to find out the predicted values.

IV DATA & RESEARCH METHODOLOGY

Research Data in this paper pertains to every day closing charges of BSE SENSEX from 1st April 2015 to thirty first March 2020. Data is retrieved from official website of BSE. Data Analysis will be in following stages (Fig1):

Loading the statistics

General Pte-processing of Data

Handling of Text Columns

Preparing facts for correlation calculation

Encoding the data

Removing Columns with consistent

Sample data down primarily based on attributes

Creating real correlation matrix

Define a result call.

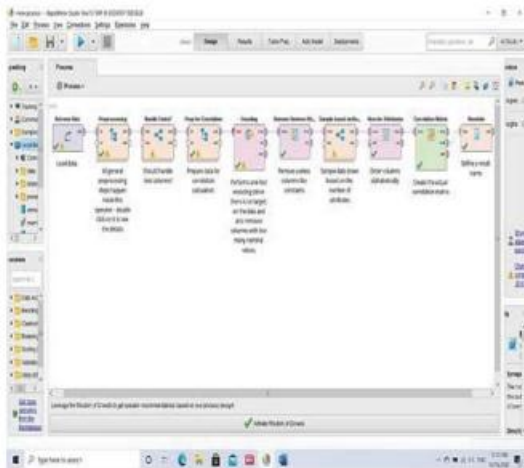


Fig.1. Process of Application of Machine Learning Algorithms

Data Visualization

Fig2 depicts the stock price movement for 5 years (2015-2020), BSE Sen sex closing prices seems to be rising with some volatility during the said duration, but a dip in prices is visible in January & February 2020 due to Covid19 and weakened global markets.

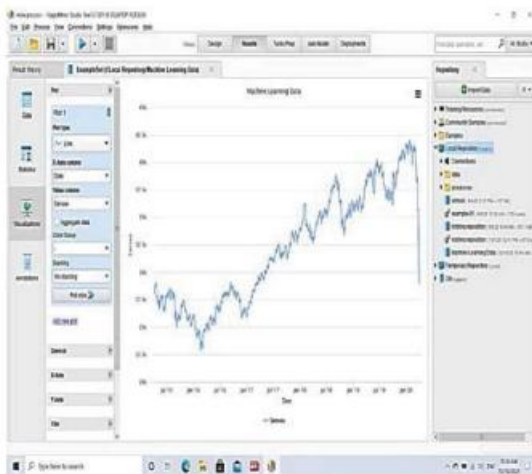


Fig.2. Data Visualization

V DEPLOYMENT OF GRADIENT BOOSTED TREES:

Gradient Boosted Trees is deployed as shown in the

Fig14. The machine is deploying this prediction model and it is clear that the values are estimated by making tree like models.

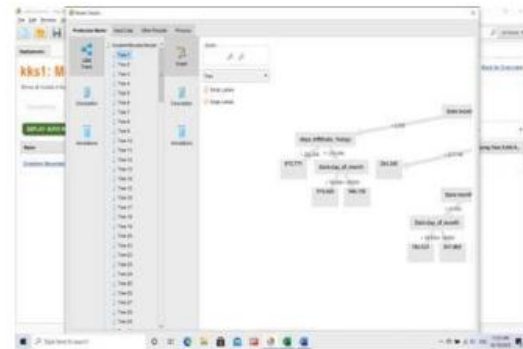


Fig.3. Deployment of Gradient Boosted Trees

After the Gradient Boosted Trees model deployed its version summary Fig15 has given the mean residual deviance 43422.95, mean absolute error 124.873276, root suggest squared log error 0. 00641. It also indicates the overall variety of timber are a 150.

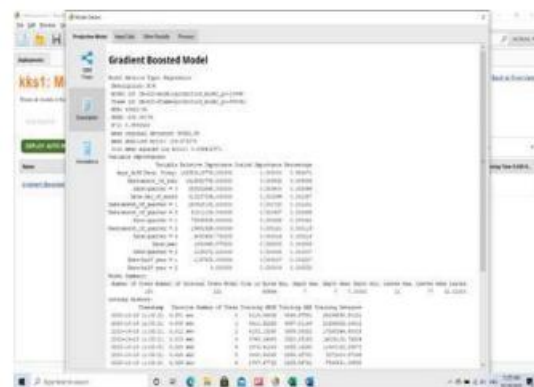


Fig.4. Model summary- Gradient Boosted Trees

The following Fig16 shows the final result of the prediction given by way of Gradient Boosted Trees version. In the figure,

Column I denote the date and Column II denotes Sen sex values and the 0.33 column denotes the prediction price. The BSE Sen sex values from tenth April 2015 to 14th April 2015 were 28516.Fifty nine, 28707.Seventy five, 28885.21, 28879.38 respectively and the expected price for those dates have been 28259.885, 28238.Sixty nine, 28247.723,28312.072 respectively.

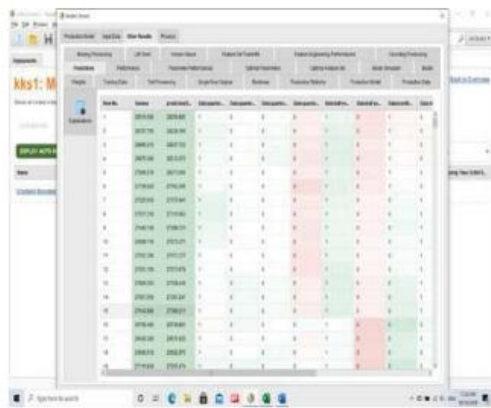


Fig.5. Predicted Values as per Gradient Boosted Trees

VI RESULT ANALYSIS

In this study, all 6 models were used to predict BSE Sen sex, namely Generalized Linear Model, Deep Learning, Order tree, random forest, gradient enhanced tree and support vector machine.

As it has already been explained, since the relative error and variance of the gradient boosted tree is the lowest, for example, zero, eight percent, and ± 0.0 percent compared to the five opposing models, so it was sent to hope for

the BSE Sen sex. As the gradient boosted model has the lowest residual deviance, 43422.Ninety-five (Fig. 16), and the lowest reported error, 124.873276, this version is submitted. After applying the gradient boosted model, it was found that the closest approximation was given by this model only. Other standards deviated much higher.

VII CONCLUSION

Finding a stable financial forecasting model for the stock market is the need of the hour. This is difficult but important when creating a financial marketing plan. Many statistical tools, understudy programming tools and gear knowledge are used in this course. In this look, 6 models namely Generalized Linear Model, Deep Learning, Decision Tree, Random Forest, Gradient Boosted Trees and Support Vector Machine have been used by the closing price of BSE Sen sex from April 2015 to March 31, 2020 to get rewards.

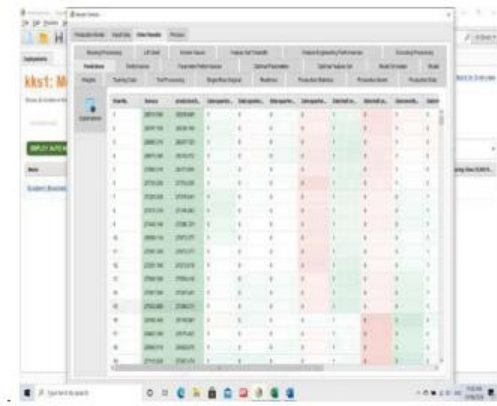


Fig. 6. Predicted values as per Gradient Boosted Trees

The estimate given by the gradient boosted model on day one is 28,259.885, which is behind the actual 28,516.590 by 256.705 bps. However, the estimates for the same day provided by deep learning models and machine learning models are 26,598.2 bp and 24,643.4 bp, respectively. The deep learning model gives an offset of 1918.39 bp while the machine learning model has an offset of 3873.19 bp. Thus, it is concluded that the gradient support model has achieved all the variables of predicting the BSE Sen sex price.

REFERENCES

1. Chen, Y., & Hao, Y. (2017). A feature weighted support vector machine and K-nearest neighbour algorithm for stock market indices prediction. *Expert Systems with Applications*, 80, 340-355.

2. Dingli, A., & Fournier, K. S. (2017). *Financial Time Series Forecasting – A Deep Learning Approach*. International

Journal of Machine Learning and Computing, 7(5), 118–122. <https://doi.org/10.18178/ijmlc.2017.7.5.632>

3. Di Persio, Luca & Honchar, Oleksandr. (2016). Artificial neural networks architectures for stock price prediction: Comparisons and applications. 10. 403-413.

4. Fama, E. F. (1995). Random walks in stock market prices. *Financial analysts journal*, 51(1), 75-80.

5. Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science*, 4(3), 183–201. <https://doi.org/10.1016/j.jfds.2018.04.003>

6. K. Khare, O. Darekar, P. Gupta and V. Z. Attar, (2017)"Short term stock price prediction using deep learning," 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, 2017, pp. 482- 486, doi: 10.1109/RTEICT.2017.8256643.

7. Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering

for stock price movement prediction. Knowledge-Based Systems, 164, 163–173.

<https://doi.org/10.1016/j.knosys.2018.10.034>

8. M. Ouahilal, M. El Mohajir, M. Chahhou and B. E. El Mohajir,(2016) "Optimizing stock market price prediction using a hybrid approach based on HP filter and support vector regression," 4th IEEE International Colloquium on Information Science and Technology (CiSt), Tangier, 2016, pp. 290-294, doi: 10.1109/CIST.2016.7805059

9. Malkiel, B. G. (1989). Efficient market hypothesis. In Finance (pp. 127-134). Palgrave Macmillan, London.

10. Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. Intelligent Systems in Accounting, <https://doi.org/10.1002/isaf.1459>

11. Prasadi Peddi and Dr. Akash Saxena (2014), "EXPLORING THE IMPACT OF DATA MINING AND MACHINE LEARNING ON STUDENT PERFORMANCE", International Journal

of Emerging Technologies and Innovative Research, ISSN:2349-5162, Vol.1, Issue 6, page no.314-318.

12. Prasadu Peddi, & Dr. Akash Saxena. (2016). STUDYING DATA MINING TOOLS AND TECHNIQUES FOR PREDICTING STUDENT PERFORMANCE. International Journal Of Advance Research And Innovative Ideas In Education, 2(2), 1959-1967.