

Stock Market Prediction Using Machine Learning (ML) Algorithms

Prof. Mayur Tembhurney¹, Akshad Sontakke², Aarti Chourawar³, Sakshi Pise⁴,
Swity Kosarkar⁵

^{1,2,3,4,5} Electronic and Telecommunication Engineering, S.B Jain Institute of Technology Management and Research,
Nagpur- 441501, Maharashtra, India

ABSTRACT

In the world, stock marketing is one of the most important activity. The main objective of this paper is to predict the value of the stock market index Nifty 50 and compare the Algorithms which is best for Stock Market Prediction by comparing the graph of the four Algorithms. In this Programing Language used is Python Programing Language. In this paper we used a Machine Learning (ML) approach for training module from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this we going to usefour machine learning techniques called Support Vector Machine (SVM),Random Forest, Linear Regression and LongShort-Term Memory (LSTM) Algorithmsto predict stock market index Nifty 50 (Close).

Keywords: Python Programing Language, Machine Learning (ML), Support Vector Machine (SVM), Random Forest(RF), Linear Regression(LR), Long Short-Term Memory (LSTM) and Stock market index Nifty 50 (Close).

1. INTRODUCTION

A share market is a market in which shares of publicly held companies or people are traded to raise money. The prices of shares fluctuate and depends upon the demand and supplies of shares. Only Companies are allowed to carry out trading are registered companies. Stock market is the process of trying to predict the future stock value of a company. Stock market prediction is one of the most widely studied and challenging problems, attracting researchers from many fields including economics, history, finance, mathematics, and computer science. This is going to help the individuals or organization to invest in the stock market. This is going to predictvalue of the Nifty 50 by taking dataset form nse.com and Kaggle.

This paper will help develop a financial data predictor program in which there will be a dataset storing all historical stock prices and data will be treated as training sets for the program. The main purpose of the prediction is to reduce uncertainty associated to investment decision making.

In this paper we are using a Machine Learning Techniques i.e., Support Vector Machine (SVM) and Random Forest, Linear Regression and Long Short-Term Memory (LSTM) in order to predict the stock market and we are using Python language for programming, API key of nse.com is used to take the real time value, we are giving Raw Data from Kaggle. The Data given is from January 2000 to July 2021 i.e., data of 21 Years, 252months and 5535 days.

2. CHRONICLE OF STOCKMARKET

Stock Market Ground Work

The stock market brings together investors and buyers to sell and buy the shares in companies at an agreed price. Prices are determined by demand and supply. The primary market deals directly with the company's new securities issues. A share is a company's issued document entitling its holder to be one of the company's owners. One can get a dividend by owning a share which in turn get capital gain by selling the shares. Stock Exchanges act as a clearing house for each transaction which guarantees the payment of the security to the seller. The smooth operation of all these activities

facilitates the expansion of businesses, economic growth, employment and the production of goods and services. It must be listed there to be able to trade a security on a certain stock exchange. The listing requirements are a set of conditions imposed on companies wishing to be listed by a given stock exchange. Traders buy and sell financial instruments like stocks, bonds, and derivatives. Traders can be either professional from financial institutions or a corporation, or individual investors. Stock market basically serves as (1) Primary market and (2) Secondary market.

Importance of Stock Market

Indian stock market stood at third rank in the world. The Stock is essentially a share in a company's ownership. Stocks are partial ownership of businesses instead of stock tickers piece of paper, which can be traded in the stock market. If company ownership is divided into 100 parts, the investor purchase one part which is equal to one share then we can own 1 percent of that company. Stock exchange uses an automated matching system driven by order. Stock prices are defined as any time how many buyers and sellers available for the same stock in the market. If the number of buyers is more than sellers then stock price becomes high and if the number of sellers higher than buyers then stock price becomes low. SEBI (Security and Exchange Board of India) regulates the stock market. In stock markets customers preferences and requirements are different. The estimated world stock market was at \$36.6 trillion in early October 2008. The total world market for derivatives was estimated at approximately \$791 trillion in face value or nominal value, 11 times the size of the world economy.

3. PROBLEM STATEMENT

Everyone want to be rich in life with low efforts and great advantages. Similarly, we want to look in our future with inner most desire as we do not want to take risks or we want to decrease risk factor. Stock market is a place where selling and purchasing can provide future aims of life (Kang Zhang et al, 2019). Now the question is that how we can get advantages from stock market? Or what are the steps that can give us stocks market predictions before taking yourself in risk zoon (Yue-gang Song et al, 2018).

4. STOCK MARKET PREDICTION (SMP)

If stock market trend predicted then we can avoid wastage of money. SMP is a process of predicting future on the base of past data. Prediction decreases the risk level to investors and increases the confidence level for investment. If they predicted goals before reach then they can avoid loss of money. All these consideration work as SMP. On the basis of historical data trends, we guess future trend that is called SMP.

5. METHODOLOGY

In this project the prediction of stock market index Nifty 50 is done by the Support Vector Machine (SVM) and Random Forest, Linear Regression and Long Short-Term Memory (LSTM). The details of each step will be discussed in this section.

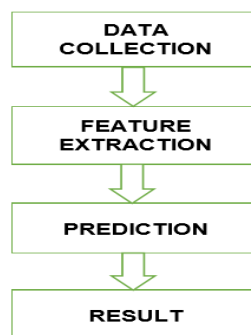


Fig: Proposed Methodology

Data Collection and Analysis Stage

In this stage, we shall look at the raw data available to us and study it in-order to identify suitable attributes for the prediction of our selected label. Now the data that we're going to use for our program is taken from, a premier dataset providing platform. The dataset taken can be extracted from Kaggle and we are going to use the API key of nse.com to take the real time value. We are going to use data of last 21 years.

Feature Extraction

The raw data is converted into processed data by using feature extraction. The raw data contains unrequired data therefore required data should be extracted for training and testing. The attributes of the dataset include: Open (Opening price of Stock), High (Highest price possible at an instance of time), Low (Lowest price possible at an instance of time), Close (Closing price of stock).

Prediction

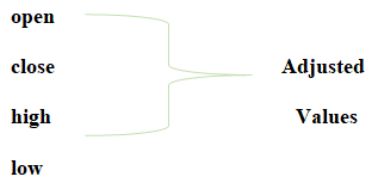
The attributes of the dataset include: Open (Opening price of Stock) High (Highest price possible at an instance of time) Low (Lowest price possible at an instance of time) Close (Closing price of stock).

Training & Testing-

Training and Testing of the model is very straightforward and the Data should be proper. The 80 percent of total data will be used for Training and 20 percent of the total data is will be used for the Testing. According to this the graph will be plotted.

Support Vector Machine

A Support Vector Machine (SVM) is a linear model which is used for classification as well as regression problems. It can solve linear and non-linear problems and work well for many practical problems. In other words, the given labeled training data (supervised learning), the algorithm outputs the optimal hyperplane which categorizes new examples. In the two-dimensional space this hyperplane is a line dividing a plane into two parts where in each class lay in either side. Support Vector Machine (SVM) is considered to be as one of the most suitable algorithms available for the time series prediction. It is a training algorithm for classification and regression, which works on a larger dataset. There are many algorithms in the market but SVM provides with better efficiency and accuracy. The correlation analysis between SVM and stock market indicates strong interconnection between the stock prices and the market index. The supervised algorithm can be used in both, regression and classification. The SVM involves in plotting of data as point in the space of n dimensions.



$$HL \% = \frac{high - low}{A_{close}} \times 100$$

$$Percent\ change = \frac{close - open}{open} \times 100$$

Random Forest

The use of machine learning techniques to predict the prices of the stock is an increasing trend. More and more researchers invest their time every day in coming up with ways to arrive at techniques that can further improve the accuracy of the stock prediction model. Due to the vast number of options available, there can be n number of ways on how to predict the price of the stock, but all methods don't work the same way.

The output varies for each technique even if the same data set is being applied. In the cited paper the stock price prediction has been carried out by using the random forest algorithm is being used to predict the price of the stock using financial ratios form the previous quarter. This is just one way of looking at the problem by approaching it using a predictive model, using the random forest to predict the future price of the stock from historical data. However, there are always other factors that influence the price of the stock, such as sentiments of the investor, public opinion about the company, news from various outlets, and even events that cause the entire stock market to fluctuate.

By using the financial ratio along with a model that can effectively analyze sentiments the accuracy of the stock price prediction model can be increased.

Linear Regression

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things:

- (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable?
- (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they—indicated by the magnitude and sign of the beta estimates—impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables.

The simplest form of the regression equation with one dependent and one independent variable is defined by the formula,

$$y = c + b \cdot x$$

where, y = estimated dependent variable score,

c = constant,

b = regression coefficient, and

x = score on the independent variable.

Longshort-Term Memory (LSTM)

It is special kind of recurrent neural network which is capable of learning long term dependencies in data. This is achieved because the recurring module.

Long Short-Term Memory Network is an type of RNN, a sequential network, that allows information to persist. It is capable of handling the vanishing gradient problem faced by RNN

LSTMs are a complex area of deep learning. It can be hard to get your hands around what LSTMs are, and how terms like bidirectional and sequence-to-sequence relate to the field.

There are few that are better at clearly and precisely articulating both the promise of LSTMs and how they work than the experts that developed them.

We will explore key questions in the field of LSTM using quotes from the experts, and if you're interested, you will be able to dive into the original papers from which the quotes were taken.

6. COMPARATIVE ANALYSIS OF DIFFERENT APPROACHES

Sr.no	Algorithm used	Evaluation metrics used	Advantages	Disadvantages
1	Linear Regression	Mean Absolute Error- 17.0039, Mean Squared Error- 940.30664, Root Mean Squared Error- 30.6644, (R ²) Score- 0.9999.	Easier to implement, interpret & efficient to train.	It is often quite prone to noise and overfitting.
2	Support Vector Machine (SVM)	Mean Absolute Error- 6014.7, Mean Squared Error- 53419696.4, Root Mean Squared Error- 7308.8779, (R ²) Score—2.0966.	Support Vector Machine is comparably memory systematic.	Exaggerate to minor fluctuations in the training data which decrease the predictive ability.
3	LongShort-Term Memory LSTM)	Mean Absolute Error- 463.5558, Mean Squared Error- 434545.7882, Root Mean Squared-Error- 659.200, (R ²) Score- 0.9999.	Speech and Handwriting Recognition.	Comparative analysis is not extensive
4	Random Forest	Mean Absolute Error- 3.6471, Mean Squared Error- 92.3353, Root Mean Squared Error- 9.6091, (R ²) Score- 1.0.	It is used in radar for object detection.	A forest is less interpretable than a single decision tree.

Here, we calculate a Mean Absolute Error(MAE), Mean Squared Error(MSE), Root Mean Squared Error(RMSE).

Mean Absolute Error:- MAE are used to measure the error produced by a predictive model.

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

Mean Squared Error:-There are no acceptable limits for MSE the higher the accuracy of prediction as there would be excellent match between the actual and predicted data set.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Root Mean Squared Error:-Root Mean Square Error (RMSE) is a standard way to measure the error of a model in predicting quantitative data. The lower the RMSE, the better a given model is able to “fit” dataset.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

7. RESULT

The model will obtain the desired results in any form we want. In this, we will be plotting a graph of our results as per our requirements which we have discussed earlier in the paper. The accuracy of the predicted result will be high. The graph will look like

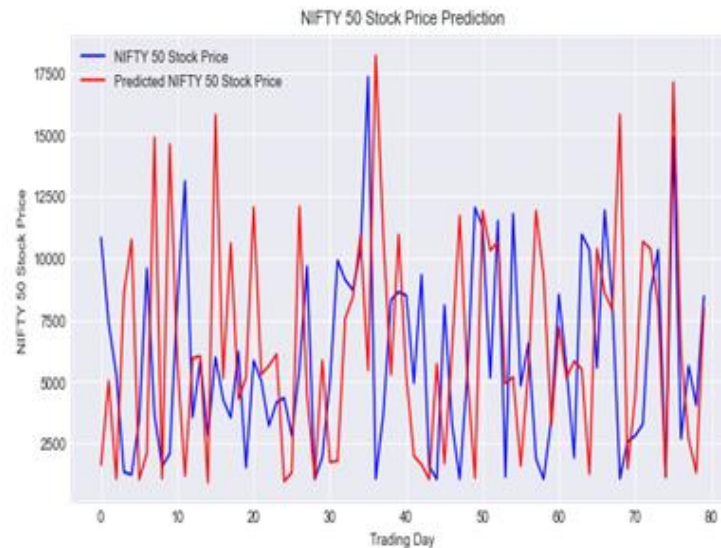


Fig. 1

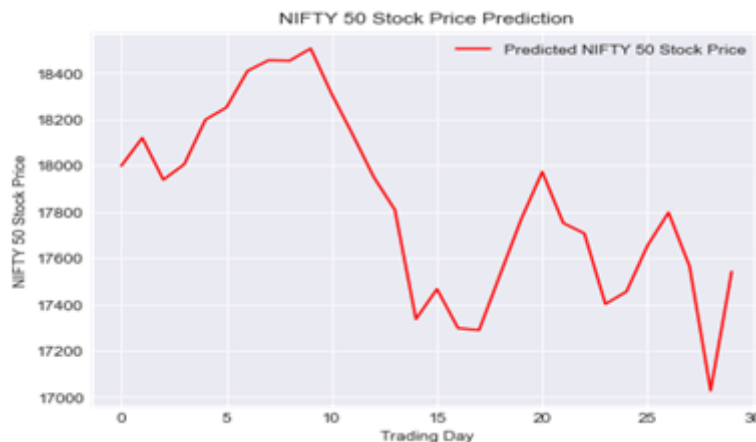


Fig. 2

As fig. 1 and fig. 2 show, the prediction result by using linear regression model from scikit-learn library in python. The first 80% data is the real data used for training set and the last 20% data for testing the predicted data set. Here, the prediction of Linear Regression is accurate as compared to SVM but as compared with the Random Forest and LSTM model predictions it is slightly inaccurate,

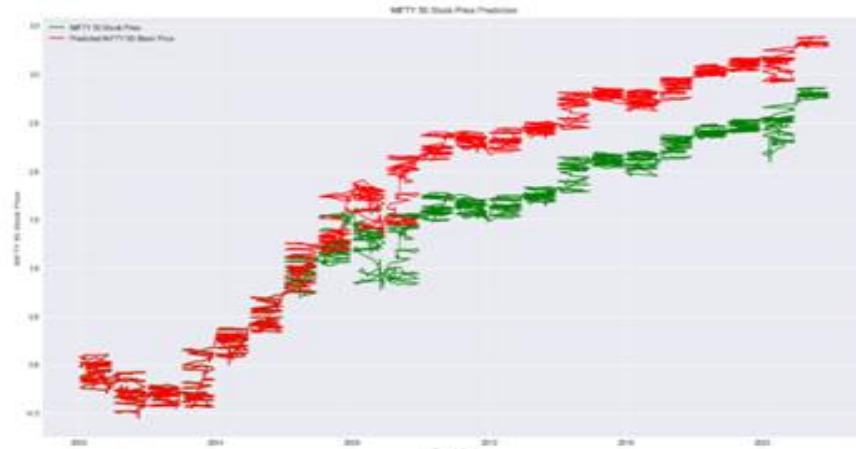


Fig. 3

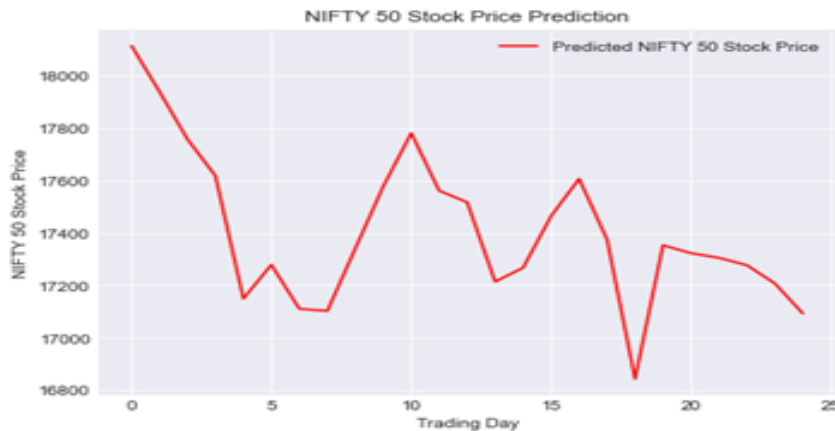


Fig. 4

As fig. 3 and fig. 4 show, the prediction result by using SVM model. In this same, we are using first 80% data as training set and the remaining 20% dataset as testing dataset. Here, the prediction of SVM model's prediction varies much, as compared to Linear Regression, Random Forest and LSTM.

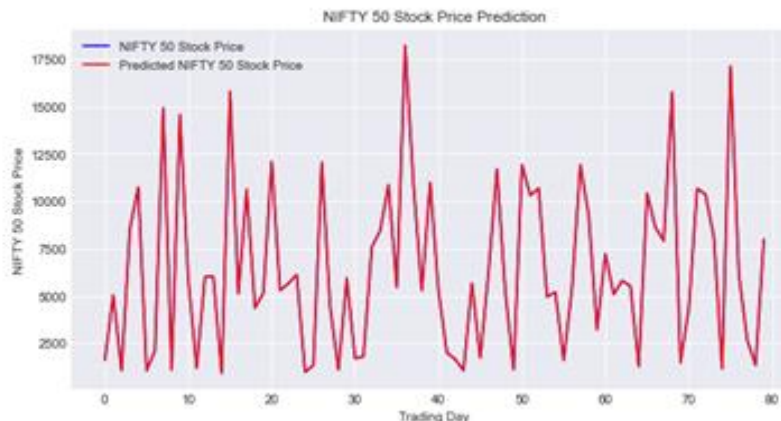


Fig. 5

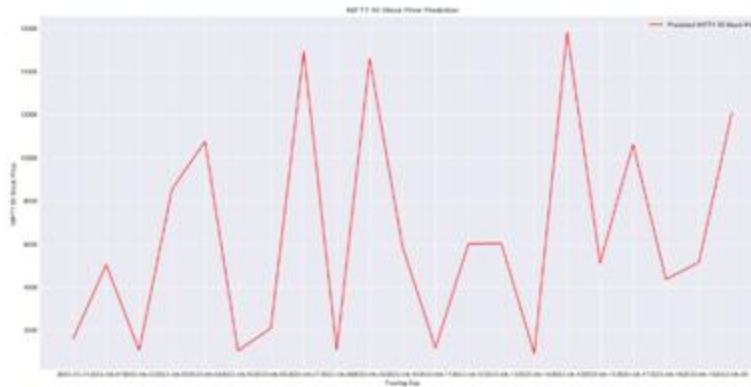


Fig. 6

As fig. 5 and fig. 6 show, the prediction result by using Random Forest model. In this same, we are using first 80% data as training set and the remaining 20% dataset as testing dataset. Here, the prediction of Random Forest are better as compared to the SVM and Linear Regression but as compared to LSTM predictions are low.

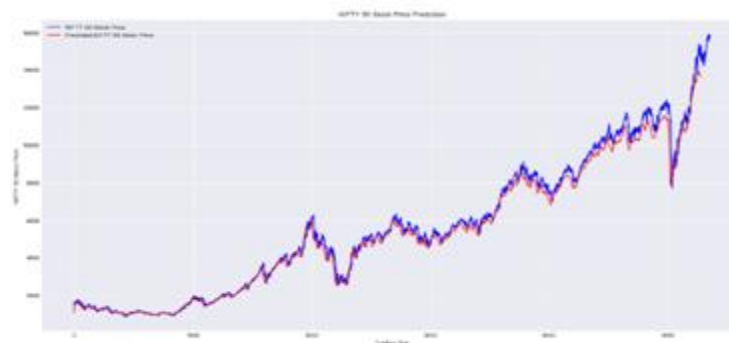


Fig. 7

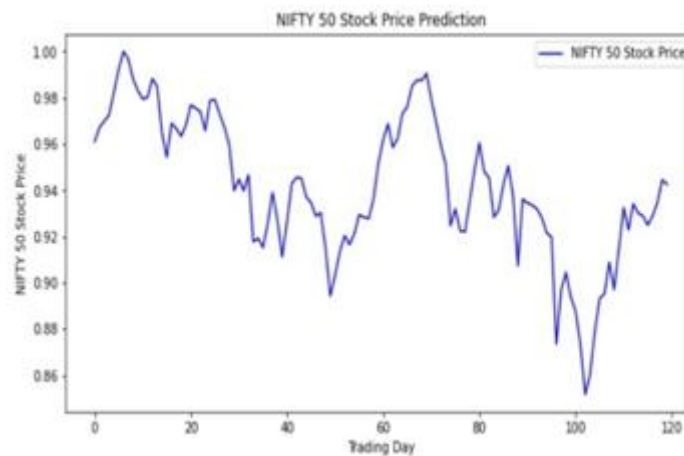


Fig. 8

As fig. 7 and fig. 8 show, the prediction by using LSTM model. For LSTM we install the packages i.e. Keras, Tensorflow. For the twenty one year dataset, there are 5535 days in total. The 80% of the total dataset used as training set and remaining 20% is used for the testing of the predicted data set. From the graph we conclude that LSTM is giving much better and accurate graph as compared to the other three algorithms.

CONCLUSION AND FUTURE SCOPE

Financial markets provide an excellent platform for investors and traders, who can trade from any gadget that connects to the internet. Over the last few years, people have become more attracted to stock trading. Like any other walk of life, the stock market has also changed due to the advent of technology. Now, people can make their investments grow. Online trading has only changed the way individuals purchase and sell stocks. The budgetary markets have advanced rapidly, and have formed an interconnected global marketplace. These advancements pave the way to new opportunities. This research based on several results and we used machine learning algorithm (ML) as Linear Regression (LR), Support Vector Machine(SVM), Long Short-Term Memory(LSTM), Random forest(RF) with respect relations to business priority.

REFERENCES

- [1]. Stock Market Prediction Using Machine Learning by V Kranthi Sai Reddy www.irjet.net.
- [2]. Stock Market Prediction Using Machine Learning (ML) Algorithms by M Umer Ghania, M Awaisa and Muhammad Muzammala ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal Regular Issue, Vol. 8 N. 4 (2019), 97-116 eISSN: 2255-2863 DOI: <http://dx.doi.org/10.14201/ADCAIJ20198497116>.
- [3]. <https://www.researchgate.net/publication/305329827> Stock Market Index Forecasting of Nifty 50 Using Machine Learning Techniques with ANN Approach.
- [4]. Kang Zhang et al. "Stock Market Prediction Based on Generative Adversarial Network", Procardia Computer Science, 147(2019):400-406.
- [5]. Stock Market Prediction using Supervised Machine Learning Techniques: An Overview 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering CSDE 978-1-665419741/20/\$31.00 ©2020 IEEE | DOI: 10.1109/CSDE50874.2020.9411609.
- [6]. Stock Price prediction using LSTM and SVR 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC) | 978-1-7281-7132-6/20/\$31.00 ©2020 IEEE | DOI: 10.1109/PDGC50313.2020.9315800.
- [7]. STOCK PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES International Conference on Intelligent Sustainable Systems (ICISS 2019) IEEE Xplore Part Number: CFP19M19-ART; ISBN: 978-1-5386-7799-5.
- [8]. Combining of Random Forest Estimates using LS boost for Stock Market Index Prediction by Nonita Sharma. 2017 2nd International Conference for Convergence in Technology (I2CT).
- [9]. Yue-gang Song et al. Corrigendum to "Towards a new approach to predict business performance using machine learning" [Cogn. Syst. Res. 52 (2018): 1004-1012].