TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 4227 - 4233

Technical analysis of Stocks listed on NSE using Machine Learning

Abhishek Rai, Manas Shukla, Ankur Chokhani, Kiran Gawande
Department of Computer Engineering
Sardar Patel Institute of Technology, Mumbai, India - 400058
Email: abhishek.rai@spit.ac.in, manas.shukla@spit.ac.in, ankur.chokhani@spit.ac.in, kirangawande@spit.ac.in

Article Info
Volume 82
Page Number: 4227 - 4233
Publication Issue:
January-February 2020

Article History
Article Received: *18 May 2019*
Revised: *14 July 2019*
Accepted: *22 December 2019*
Publication: *21 January 2020*

Abstract
The authors of this paper have proposedthe technical analysis of various stock prices as listed on National Stock Exchange (NSE) of India. Various prediction models have been tried and built using different analytical techniques. The openly available historical stock data has been usedand various models have been trained with it. Various parameters were explored, which affect the prediction models. This paper also tries to explore various machine learning techniques available and compares their results to understand which one is better and to what extent. In totality, the authors have analysed the stock prices of Tata Steel, Bank of Baroda and Tata Consultancy Services (TCS) using 4 different models, which are the Moving Average model, Linear Regression model, K-nearest neighbours (KNN) model, and the Long Short Term Memory (LSTM) model. The performances of these prediction models have been compared by calculating the Root Mean Square Error (RMSE) values of each model, which gives us an idea about how accurate are the predictions made by eachmodel.

Keywords*—Stock prediction, Machine Learning, Artificial Intelligence, Moving Average, Linear Regression, KNN, LSTM.*

I. INTRODUCTION

Stock market prediction has been a topic of great interest and research since a very long time, due to its ability of generating huge monetary profits in a much smaller period of time. The volume of trading that takes place throughout the globe daily indicates that this is a field of study that would be area of research for many more years to come. There have been numerous attempts to find different and better ways to predict the stock prices with higher accuracy. In general, there are broadly 3 ways through which we can analyse and predict the future stock values, that are - Fundamental, Sentimental and TechnicalAnalysis.

Fundamental Analysis refers to analyzing the fundamentals of a business in order to predict, primarily, the tangible estimates like revenue and earning potential. The factors considered while analysing a business fundamentally include studyingtheexistingmarketscenarioforthebusiness,business model, competitors analysis, industry P/E ratio, scope of business expansion etc. It mainly deals with how viablea business will be, and the scope of being able to generate appropriate return on the investment being made.Fundamental Analysis helps an investor to understand whether it is a good idea to invest in a particular business or not in the long term, and also speaks about the pros and cons of the business, and the possible headwinds that the business might face in the future.Sentiment Analysis refers to gauging the latest trend about the business, its industry, and the continual reaction of the people to the business. Sentiment Analysis is majorly done by identifying words that can be classified as either positive, negative, or neutral. It will depict the general perception or outlook towards the business, giving weightage to the social aspect of the opinion about a business. Sentiment Analysis helps in analysing the short-term outlook of the business, and can be triggered by any adverse happening affecting the business, provided it attracts peoples attention. It canhelp gauge a business from a short-term as well as a long-term perspective.

Technical Analysis refers to analysing the various charts of stock price movement in order to predict how the future stock price variation would be. There are different types of graphs that can be used in technical

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 4227 - 4233

analysis, including chart patterns as well as statistical indicators. However, technical analysis is based entirely upon the historical data of the stock price, and does not factor in the external influences on the business, like the effects of market sentiment and fundamentals of the business. Technical analysis is mainly about identifying trends in the price movement, and the estimation of the trigger point for these trends to replicate themselves and also the intensity with which the trends shall be replicated. Trends can be both short term or long term. Technical Analysis isextensively usedbytraders,majorlytheoneswhotradeinshort-terms.

The authors have chosen to do Technical analysis among these 3 techniques, so as to explore the features and parameters that affect the stock prices, which is possible by analysing the historical data that is openly available on the internet. The authors also wanted to understand how different machine learning tools and techniques can affect the accuracy of prediction, and thus help in obtaining better throughput for the stock market.

II. Implementation

This paper discussed the implementation of 4 different models to predict the stock prices, which are - Moving Average model, Linear Regression model, KNN model, and the LSTM model. The moving average model uses the Simple Moving Average (SMA) of a given window size to calculate the future stock price. The other 3 models are basically machine learning techniques/algorithms, that use thehistoricaldatasettotrainthemodelandthenpredictthe stock prices. These 4 models have been chosen since we wanted to try out different types of algorithms and understand the advantages and limitations of each one of them, so that we can find which one of them is better andwhy.

The datasets of Tata Steel, Bank of Baroda and TCS have been considered, that were openly available on Kaggle. The prices of the stocks are the same as listed on National Stock Exchange (NSE) of India. The dataset consists of attributes such as opening and closing values of a stock for the day, highest and lowest values in the day, as well as the total trading volume throughout the day. But our main focus was on the closing price of the stock at the end of day, which is one of the most important factors that decides the fluctuation in the stock price on the next day. The dataset ranges from mid 2008 to early 2018. We chose these 3 organizations as they are among the leading firms in India and have been in the stock market game for a long time now. Also, all 3 of them represent different sectors, which gives us a versatile dataset with different stock price ranges. All the 4 models have been implemented using Python. Machine learning and neural network packages such as sklearn and keras have been used to implement these models. The training and testing datahavebeendividedintheratioof75:25.

III. Moving AverageModel

Simple moving average is one of the various time series analysis techniques that exist. Time series analysis is a way through which data that is timely structured is processed, and it helps in finding valuable results or important characteristics for multiple reasons [1]. The formula for calculating the Simple Moving Average is given in Figure1.

$$\bar{p}_{\mathrm{SM}} = \frac{p_M + p_{M-1} + \cdots + p_{M-(n-1)}}{n} = \frac{1}{n}\sum_{i=0}^{n-1} p_{M-i}$$

Fig. 1. Formula for Simple Moving Average

The ‘n’ in the above formula represents the window size that was chosen, i.e. it will calculate the average stock price of the past ’n’ days from any given day. This Moving Average model calculates the average value of the stock prices for a specific time period, which can be set according to the use case. For stock prediction purposes, the window size has been chosen to be the same as that of the testing data. Hence if the stock price has to be predicted many days into the future, this model would also consider the already predicted average stock prices, which is the reason why it is known as the “moving”average.Figures 2,3 and 4 show the performance of this model for various stocks, where the blue curve represents the training data, curve in green represents the predicted values and the orange curve represents the actual values.

January - February 2020
ISSN: 0193 - 4120 Page No. 4227 - 4233

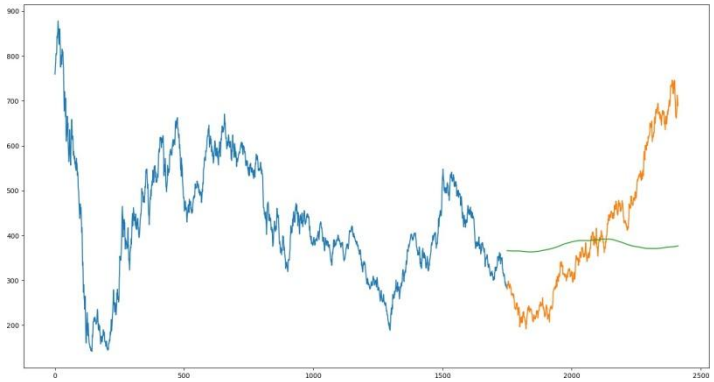

Fig. 2. Simple Moving Average for Tata Steel

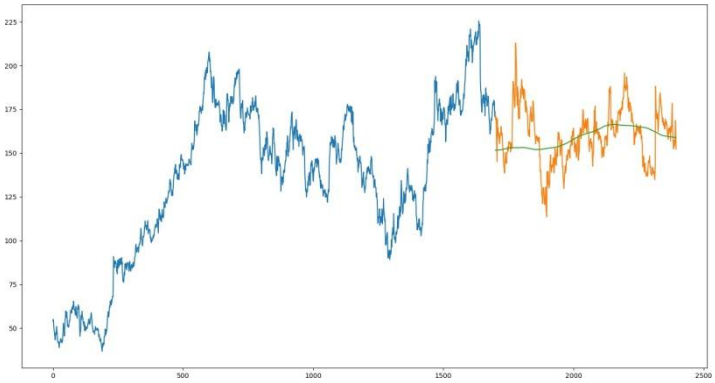

Fig. 3. Simple Moving Average for Bank of Baroda

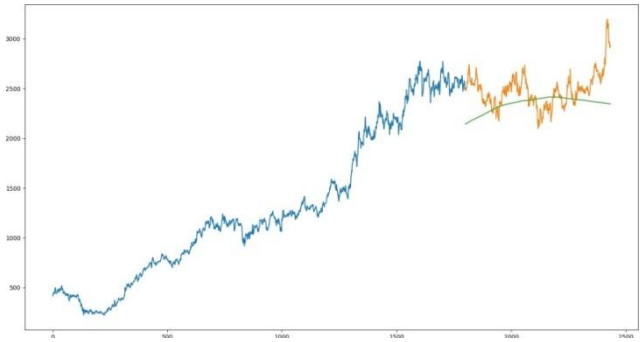

Fig. 4. Simple Moving Average for TCS

As visible in these above figures, Simple Moving Average model output is often a smooth curve that tends to intersectthe actual curve a few times, but isn't able to catch the variations exactly.

IV. Linear RegressionModel

Regression techniques have always been a part of the machine learning domain. In 1805, Legendre published the method of least squares, which was one of the earliest forms of regression, while Gauss published further developments to the theory of least squares in 1821 that also includes the Gauss-Markov theorem[2].

Linear regression is one of the most basic and well known machine learning algorithms used for prediction. There are different dependent and independent variables in a problem definition and this algorithm returns the relationship between them. For this particular case, there are no independent variables, there are just dates. The dependant variable is the stock price. The linear regression algorithm was applied for these variables and stock prices were predicted.

January - February 2020
ISSN: 0193 - 4120 Page No. 4227 - 4233

Figures 5,6 and 7 show the performance of this model, where the blue curve represents the training data, green curve represents the predicted values and the orange curverepresents the actual values. This algorithm did not perform very well for stock price prediction, the main reason being that linear regression can more often than not overfit to the datecolumns.

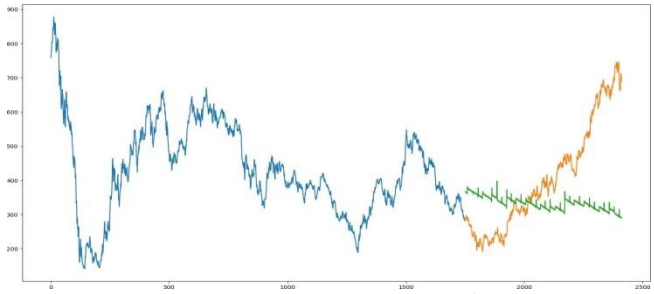

Fig. 5. Linear Regression for Tata Steel

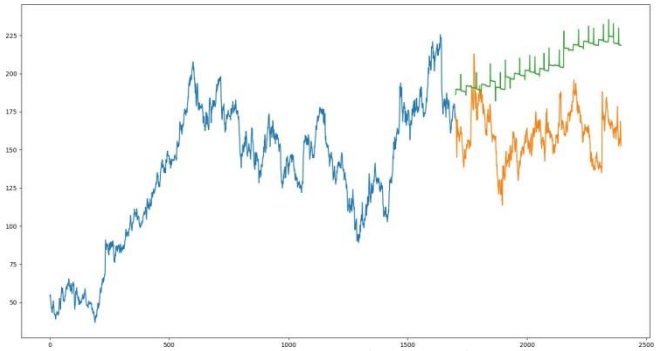

Fig. 6. Linear Regression for Bank of Baroda

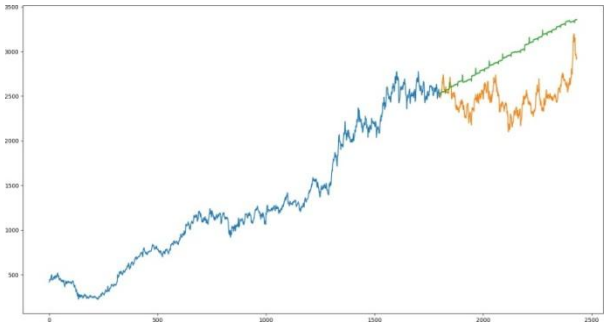

Fig. 7. Linear Regression for TCS

As it is visible in the above figures, Linear Regression model tries to best fit a line through the past data which is available, and extrapolates it to predict the future values. But since the stock market is a highly volatile market, it does not always follow a linear path, which is the primary reason why Linear Regressionisunabletogetgoodresultsforsuchusecases.

V. KNNMODEL

KNN or the K-Nearest algorithm is an algorithm that can either be used for regression or classification. For the purpose of this paper, KNN has been used for regression. KNN algorithm uses the concept of feature similarity and predicts the value corresponding to the point in question. Basically, it assigns a value to the point based on how similar the point is with the existing points in the training data. In this case, the closing prices of a stock for a certain period of time were available and the new closing price had to be predicted. The number of nearest neighbors was considered as 5. Hence, for the new closing price, the algorithm computes the value of the average of 5 nearest closing prices and that value is the predicted closingprice.

January - February 2020
ISSN: 0193 - 4120 Page No. 4227 - 4233

Different types of distance functions can be utilized to calculate the distance between two data points, like Manhattan and Euclidean distances. But there are limitations and drawbacks of using the distance function as well. The independent variables would more often than not have varying measurement scales. Thus normalization is needed to nullify such effects[3].

Figure 8,9 and 10 show the performance of this model, where the green curve represents the training data, orange curve represents the predicted values and the blue curve represents the actual values. This algorithm too did not perform very well for stock price prediction, but was able to detect some trends in the later stages.

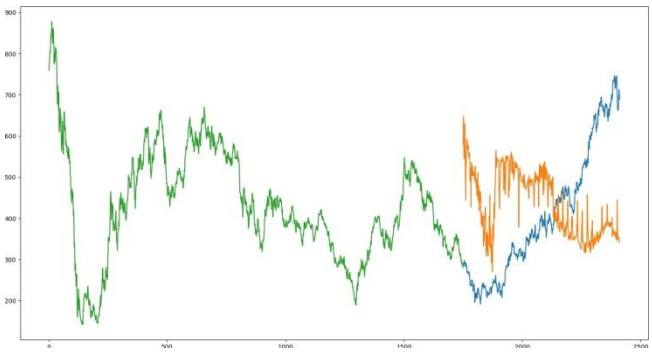

Fig. 8. KNN for Tata Steel

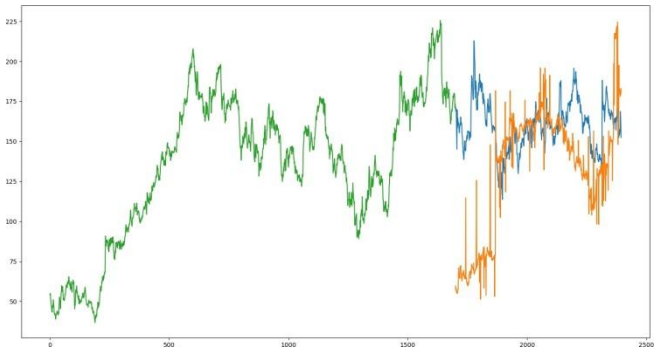

Fig. 9. KNN for Bank of Baroda

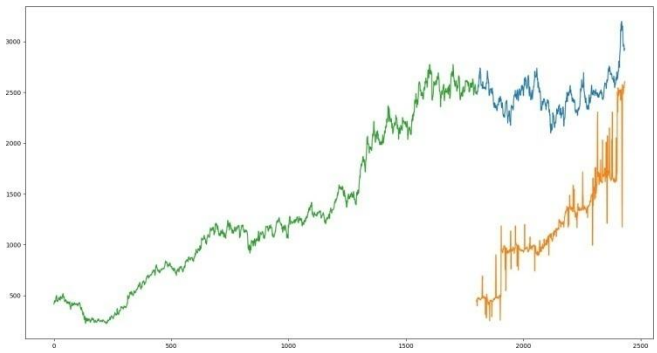

Fig. 10. KNN for TCS

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 4227 - 4233

VI. LSTMMODEL

LSTM Networks are a part of Deep learning architecture. They use artificial neural networks to train the model with the help of multiple neurons and hidden layers. Neural networks help in prediction problems by adjusting the weights between the hidden layers in the network through back-propagation, by learning from the supervised dataset, so that better predictions can be made for future inputs. LSTM (Long-Short Term Memory) is a type of Time Recurrent Neural Network (RNN), that is aptly suitable for processing and predicting theimportanteventsofintervalandlongdelayintimeserieslike the stock market [4].

Recurrent Neural Networks were designed to overcome the limitations of a normal Feed-forward neural network, which was that they were unable to capture the sequences/time series information, like that in stock market data. But RNNs themselves were unable to capture the important trends in the stock price fluctuations, due to the problem of vanishing gradient [5]. Thus LSTM Network was introduced, which is a kind of modified RNN, that helped to overcome this drawback. LSTM Networks try to remember the important information from the past and discard/forget the information that is not so important. They achieve this with the help of 3 gates, namely the Input, Forget and Output gates. It includes extra interactions between the hidden layers which help in recognizing which information can be discarded and which havetorememberedforalongerduration.

Figures 11, 12 and 13 show the performance of this model, where the blue curve represents the training data, green curve represents the predicted values and the orange curve represents the actual values. This model clearly performs very well and predicts almost all the trends as visible from thefigure.

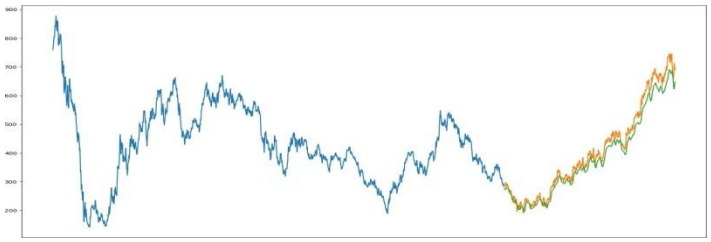

Fig. 11. LSTM for Tata Steel

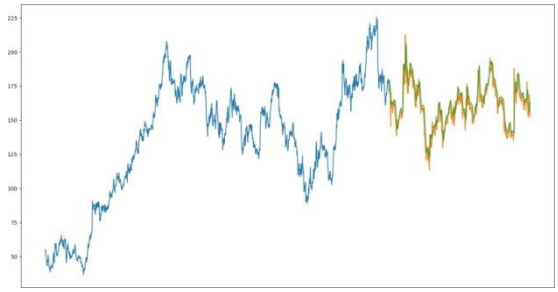

Fig. 12. LSTM for Bank of Baroda

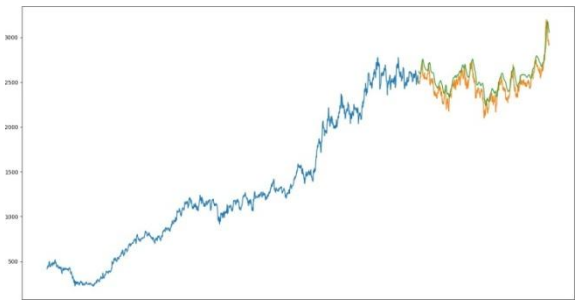

Fig. 13. LSTM for TCS

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 4227 - 4233

VII. Results

The RMSE (Root Mean Square Error) value for each model has been calculated to understand which of them is relatively better. The formula for RMSE is given in Figure 6.

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

Fig. 14. Root Mean Square Error

In the above formula, 'N' represents the number of days in the testing dataset. Hence the RMSE value of each model has been calculated for the duration of the testing dataset. The comparitive results of all models for each of the 3 organizationsaresummarizedinTables1,2and3respectively.

TABLE I
RMSE Values for Tata Steel

Sr No	Model	RMSE
1	Moving Average	157.138
2	Linear Regression	185.226
3	KNN	212.762
4	LSTM	28.676

TABLE II
RMSE Values for Bank of Baroda

Sr No	Model	RMSE
1	Moving Average	15.285
2	Linear Regression	48.375
3	KNN	52.306
4	LSTM	7.244

TABLE III
RMSE Values for TCS

Sr No	Model	RMSE
1	Moving Average	236.387
2	Linear Regression	537.321
3	KNN	1389.886
4	LSTM	103.349

RMSE is sensitive to the magnitude of stock prices, i.e. bigger the value of a stock, bigger will be its RMSE value. Hence, only the RMSE values of various models within the same stock have been compared, and not inter-stock values, as every stock has a different price range, and hence the magnitude of RMSE will also vary from stock to stock. For example, in case of TCS, where the stock prices are in the range of thousands, the RMSE value is expected to be high, evenifttheerrorisactuallylow.

VIII. Conclusion

Since the RMSE value of LSTM model is minimum in all the cases, it is the best technique among all the 4 models tested, and the most efficient as well. This highlights the advantages of artificial neural networks over other models. We can also conclude that even though algorithms like linear regression and KNN are very good in general, they aren't able to perform well for time series applications like that of stock marketdata.

Further work can include exploring other external features like crude oil prices, currency exchange rates, etc. that could potentially affect the stock prices of a firm like Tata Steel. The LSTM model could also be extended and the hyperparameters could be tuned to produce better results.

References

1. Stefan Lauren, S. Dra. Harlili, "Stock trend prediction using simplemoving average supported by news classification", *International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 2014, pp135-139.
2. Han Lock Siew, Md Jan Nordin, "Regression techniques for the prediction of stock price trend", International Conference on Statistics in Science, Business and Engineering (ICSSBE),2012,pp1-5.
3. AthitPhongmekin, PisitJarumaneeroj, "Classification Models for Stock's Performance Prediction: A Case Study of Finance Sector in the Stock Exchange of Thailand", International Conference on Engineering, Applied Sciences, and Technology (ICEAST), 2018, pp 1-4.
4. Siyuan Liu, Guangzhong Liao, Yifan Ding, "Stock transaction prediction modeling and analysis based on LSTM", 13th IEEE Conference on
5. Industrial Electronics and Applications (ICIEA), 2018, pp 2787-2790.
6. Shuai Li, Wanqing Li, Chris Cook, Ce Zhu, YanboGao, "Independently Re current Neural Network (IndRNN): Building A Longer and Deeper RNN",
7. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018,pp 5457-5466.