

Stock Data Analysis and Prediction in Machine Learning

Ankit Kumar^{1*}, Jasbir Singh Saini²

^{1,2}Guru Nanak Dev Engineering College, Ludhiana (Punjab), India

*Corresponding Author: ankitkumargill95@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v8i9.7078> | Available online at: www.ijcseonline.org

Received: 06/Sept/2020, Accepted: 12/Sept/2020, Published: 30/Sept/2020

Abstract- In the world of stock market Machine Learning has a very unique role to play when it comes on to the stock prediction. Machine learning library which is also known as MLIB helps in determining the future values of the stocks. This Research finds out the future ups and downs of stock market by providing you a signal for the same, whether the stock will be closed up or down. This has done by analysing the historical data. In this study stock data of NSE (National Stock Exchange of India) from 2000 to 2019 have been analysed which includes top forty eight companies of various sectors from all over India. With the help of machine learning libraries six technical indicators known as Bollinger Band, Relative Strength Index(RSI), Stochastic Oscillator, Williams %R, Moving Average Convergence Divergence (MACD), Rate of Change have been applied on to the nineteen years of stock data and finally, Random Forest algorithm and Artificial Neural Network Model have been applied on it to predict the stock movement, at last a comparison between Random forest and ANN model has also been done to check the better prediction.

Keywords- Stock data, Nifty-50, Stock Indicators, Random Forest, Artificial Neural Network.

I. INTRODUCTION

There are many technical indicators which are available in stock market that aims to predict the market trends. These technical indicators are totally based on mathematical calculations that uses past stock data or historical data to forecast the market direction. In the world of stock market many indicators are still under development by traders so that they can get the better results because market is keep on changing day by day and it demands upgradation. In this work we tried to maximize the prediction with two most popular machine learning algorithm i.e Artificial Neural Network model and Random forest model. These are the most commonly used algorithm and they have different way of learning.

I.I ANN (Artificial Neural Network)

ANN is basically a connectionist systems inspired by the biological neural networks that represent animal brain. Without being programmed with any kind of task or specific rule these systems "learn" to perform tasks by considering examples. The connections between artificial neurons are known as "edges". Artificial neurons and edges generally have a weight that adjusts as learning issue. The load will increase or decreases the strength of the signal at a association. Artificial neurons could have a threshold specified the signal is merely sent if the combination signal crosses that threshold. The original task of the ANN technique was to solve the issues within the same manner that an individual's brain would, so instead of taking the entire knowledge set ANN only takes knowledge samples to make solutions, that saves time as well as cash. ANNs square measure thought of fairly straight-forward mathematical models to boost existing knowledge analysis technologies [2] [10] [14].

I.II Random Forest

This approach comes under the predictive modeling approach of machine learning. It's a kind of decision tree learning where observations are represented by its branches and conclusion are represented by the leaves of the tree. Instead of depending on a single decision this method depends on multiple decisions and then it will give you the final output or result. Random forest divides the data into two sets, training set and testing set. In training set it creates the decision trees by picking up the random samples from your data and in the second phase (testing set) it will give you an output by considering decisions from all the trees that has made in the training phase [1][7][16].

I.III National Stock Exchange of India

NSE (National Stock Exchange of India Limited) is a ruling stock exchange of the country (India). It provides fully automatic electronic trading system which makes trading easy for the investors and it also provides different facilities to the investors. Its indices are used in India which involves NIFTY 50, NIFTY NEXT 50, NIFTY 500. Indices like NIFTY 50 provides a portfolio for the investment to the indian market and covers around fourteen sectors like Nifty Bank, Nifty IT, Nifty Pharma, Nifty Serv Sector etc of the indian economy. Earlier people like J.Patel et al [1](2014), P.Aithal et al [2](2019), A.Giri et al [4](2017), M.Kumar et al [16](2006) have been worked on Nifty-50 stock data.

II. RELATED WORK

Jigar Patel et al [1](2014), Studied four prediction models Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest and Naïve-Bayes. To

demonstrate the methodology ten years of stock data of Reliance Industries and Infosys Ltd from 2003 to 2012 was taken into consideration and two stock indices CNX Nifty and BSE (Sensex) were considered for the evaluation. Ten technical indicators have also been applied in this work and it was found that Random Forest improves the overall performance of other three prediction models.

Giri A. K., Joshi Pooja [4](2017), Studied the impact of economic indicator on Indian stock market. It is an empirical analysis which examined the impact of GDP, inflation, exchange rate and crude oil on Indian stock market. Their study shows that inflation, GDP and exchange rate have a positive response on the Indian stock market whereas crude oil price has a negative impact on Indian stock market.

Nonita Sharma, Akanksha Juneja [8](2017), In this study historical data of ten years of data indices, namely, CNX Nifty and S&P Bombay Stock Exchange (BSE) Sensex from Indian stock market was carried out to predict the future stock market index. This research combines the prediction of the ensemble of trees in a random forest using LS boost and then it compares the performance with support vector regression. Finally, the results show that the scheme outperforms support vector regression and can be applicable for building predictive models for stock price prediction.

Zhihao Peng [13](2019), In this paper, it is proposed to analyze the robust Kloder-Hadoop-based data pipeline's approach for any scale and data to estimate the daily returns of selected US stocks based on real-time data from Yahoo Finance Analysis. Apache Hadoop Big-Data Framework is provided to manage large data sets through

distributed storage and processing, US stock market shares are selected, and their daily profit data is used to estimate stocks with high profit margins. For this, the training and testing data is divided into sets to predict the stocks with high daily gains module of Spark.

R. Yamini Nivetha, C. Dhaya [17](2017), In this research, various prediction algorithms are analyzed to create a prediction model. The estimation model is based on a monthly estimate and a daily estimate for the next day's market value. This model predicts the next day's open price on the market. Sentiment analysis requires identifying and eliminating emotions from each person on social media. The relationship between sentiment and stock prices must be determined. A comparative study of these three algorithms: Multiple Linear Regression, Support Vector Machine and Artificial Neural Network. The stock price is estimated with the best algorithm.

III. METHODOLOGY

The stock data frame for trading has been obtained for analysis offered through www1.nseindia.com. In this work, stock data of nineteen years i.e. from 2000 to 2019 has been taken into consideration. The indicators applied in this work are as follows:- Relative Strength Index (RSI), Stochastic Oscillator, Williams% R, Moving Average Convergence Divergence (MACD), Rate of Change (ROC), Bollinger Band. To improve the result of ANN and Random Forest model all the NAN values from the data frame were neglected and finally a prediction column has been created at last in which all the results have been stored and that will give you an indication for the future. The process flow has been shown in the enumerated Figure 1.

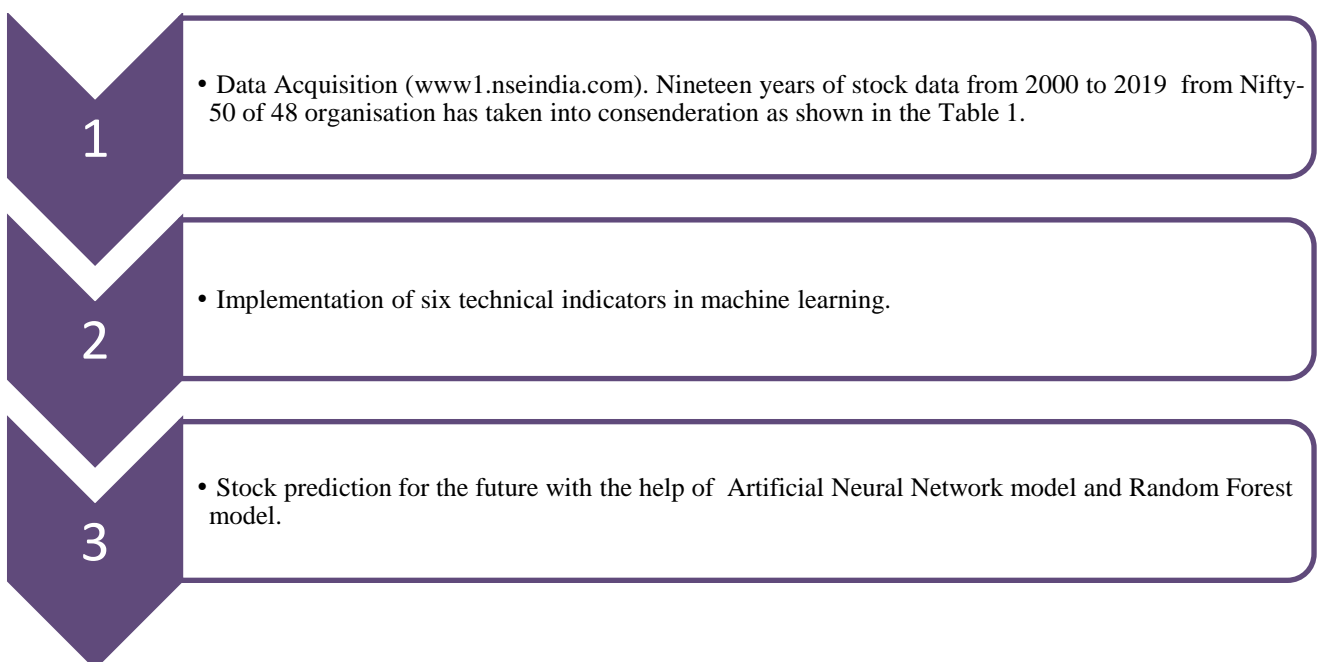


Figure 1: Workflow of proposed approach.

Figure 2 shows the working of six technical indicators (that has applied in this work) with the models (ANN and Random Forest). It also shows the relationship between the indicators and the two models that has applied in this work. It's basically a representation of how indicators give their opinion on stock movement and then how models use

that opinion to predict the stock. Each technical indicator (Relative Strength Index (RSI), Stochastic Oscillator, Williams% R, Moving Average Convergence Divergence (MACD), Rate of Change (ROC), Bollinger Band) gives an opinion on the stock by UP or DOWN and then that opinion has been used in predicting the stock.

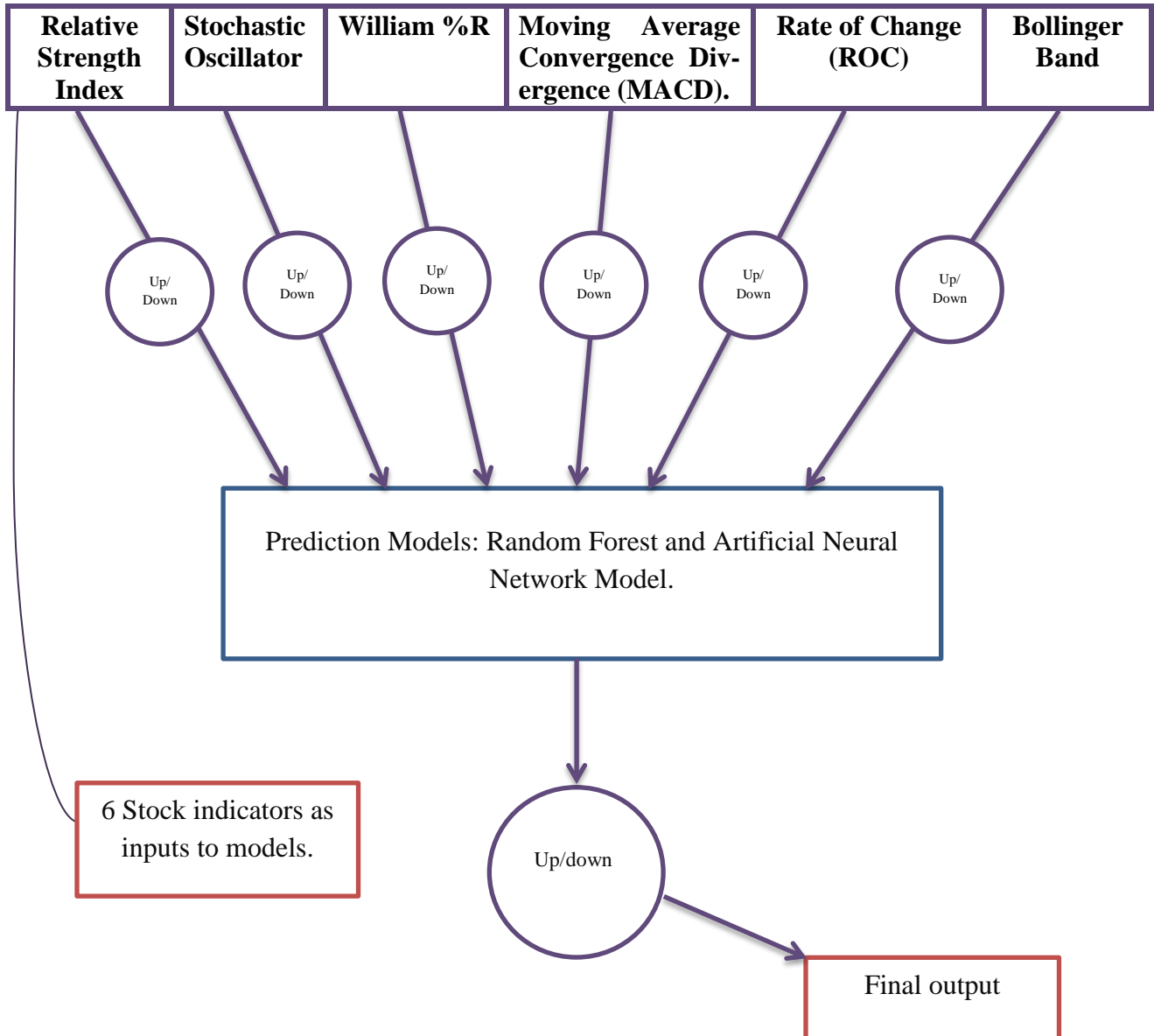


Figure 2: Showing the relationship of six stock indicators with models.

IV. RESULT AND DISCUSSION

Forty-eight organizations stock data that has been taken into the research are listed in the table given below. It

involves different – different organizations from various sectors and each organisation has nineteen years of stock data.

Table 1: Stock data of forty-eight organizations.

S. No.	Name of the Organization	S. No.	Name of the Organization
1.	WIPRO	25.	ICICIBANK
2.	HDFC	26.	INDUSINDBK
3.	ADANI PORTS	27.	INFRA TEL
4.	ASIAN PAINT	28.	JSW STEEL
5.	AXIS BANK	29.	INFY

6.	BAJAJFINSV	30.	BAJAJ-AUTO
7.	BAJFINANCE	31.	KOTAKBANK
8.	BHARTIARTL	32.	LT
9.	BPCL	33.	MARUTI
10.	BRITANNIA	34.	MM
11.	CIPLA	35.	NESTLEIND
12.	COALINDIA	36.	NTPC
13.	DRREDDY	37.	ONGC
14.	EICHERMOT	38.	POWERGRID
15.	GAIL	39.	RELIANCE
16.	GRASIM	40.	SBIN
17.	HCLTECH	41.	SUNPHARMA
18.	HDFCBANK	42.	TATAMOTORS
19.	HEROMOTOCO	43.	TATASTEEL
20.	HINDALCO	44.	TECHM
21.	IOC	45.	TITAN
22.	ITC	46.	UPL
23.	TCS	47.	VEDL
24.	HINDUNILVR	48.	ZEEL

In this work all the technical indicators that has been applied on the stock data of forty-eight organisation are as follows:

Relative Strength Index (RSI): RSI is one of the best indicator if you want to analyze the financial market. In order to calculate the RSI you need to calculate the EWMA (Exponential Weighted Moving Average) first for both Up and Down which means older values are given less weight as compared to newer values and it also known as

Relative Strength in general language and then with the use of Relative Strength you will be able to calculate the RSI. Formula to calculate the Relative Strength is given in the equation-1 and equation-2 respectively.

$$\text{Relative Strength} = \frac{\text{ewma up}}{\text{ewma down}} \quad (1)$$

$$\text{Relative Strength Index} = 100 - \left(\frac{(1 + \text{Relative Strength})}{100} \right) \quad (2)$$

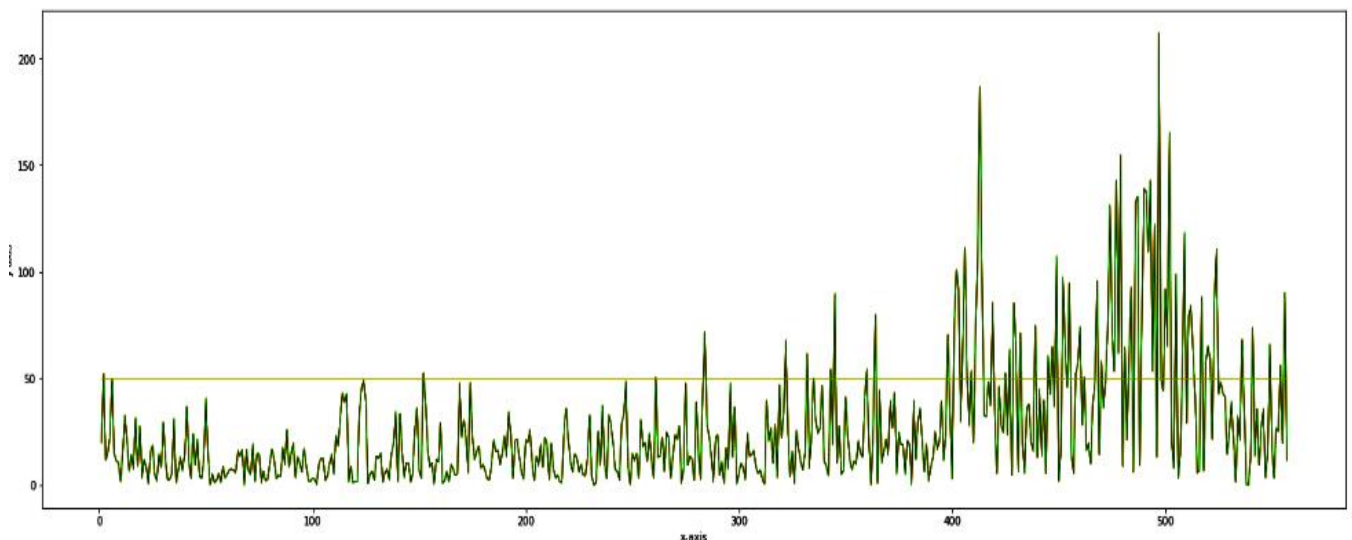


Figure 3: Representing the applied RSI indicator on the stock data.

Stochastic Oscillator: This indicator basically indicates the movements of the market. In Stochastic Oscillator sensitivity of the oscillator is reducible, all you need to adjust the time period. Stochastic Oscillator can be obtained with the help of mentioned formula given below.

$$\text{Stochastic Oscillator} = 100 \times \frac{\text{Close price} - \text{Low}}{\text{High} - \text{Low}},$$

$$D\% = \frac{K1 + K2 + K3}{3} \quad (3)$$

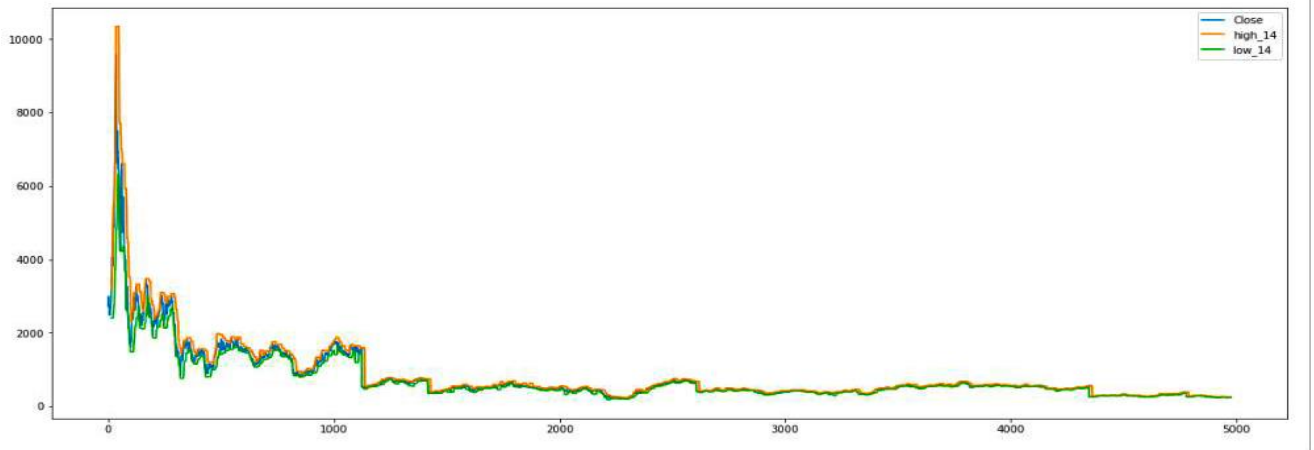


Figure 4: Representation of Stochastic Oscillator. Blue line in the middle showing the last close whereas orange and green lines are representing the high and low values of the stock.

Williams %R: It works on the current closing price and tells the relation of high and low of the past N days with that closing price. Formula to calculate the Williams %R is shown in equation-4.

$$\text{Williams \%R} = \frac{((\text{HighNdays}) - (\text{CloseToday}))}{((\text{HighNdays}) - (\text{LowNdays}))} \times -100 \quad (4)$$

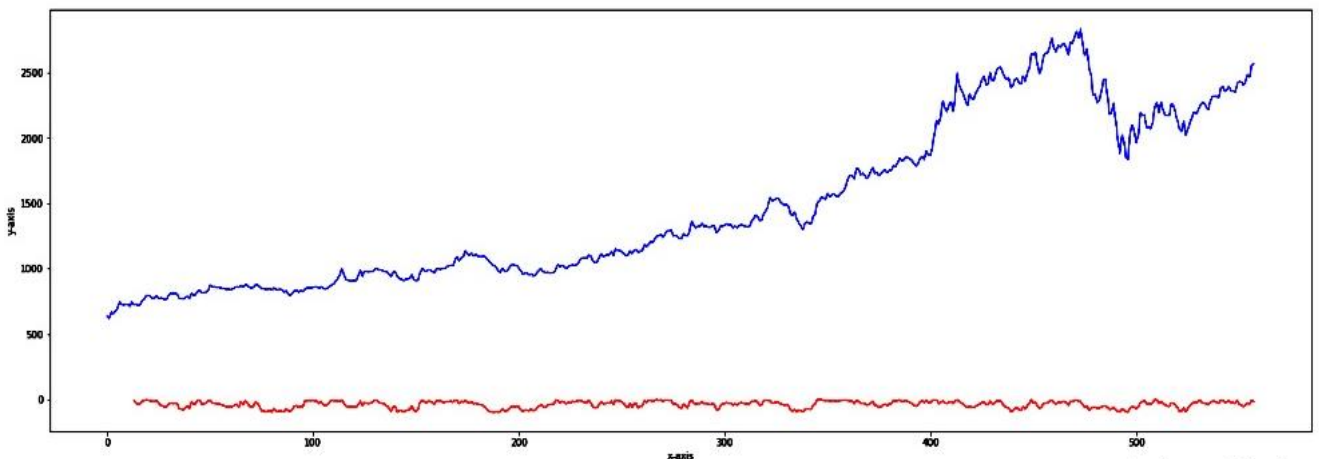


Figure 5: Graphical representation of Williams %R.

MACD (Moving Average Convergence Divergence): Moving Average Convergence Divergence can be obtained by just subtracting the 26-periods of EMA

(exponential Moving Average) from 12-periods of EMA (exponential Moving Average).

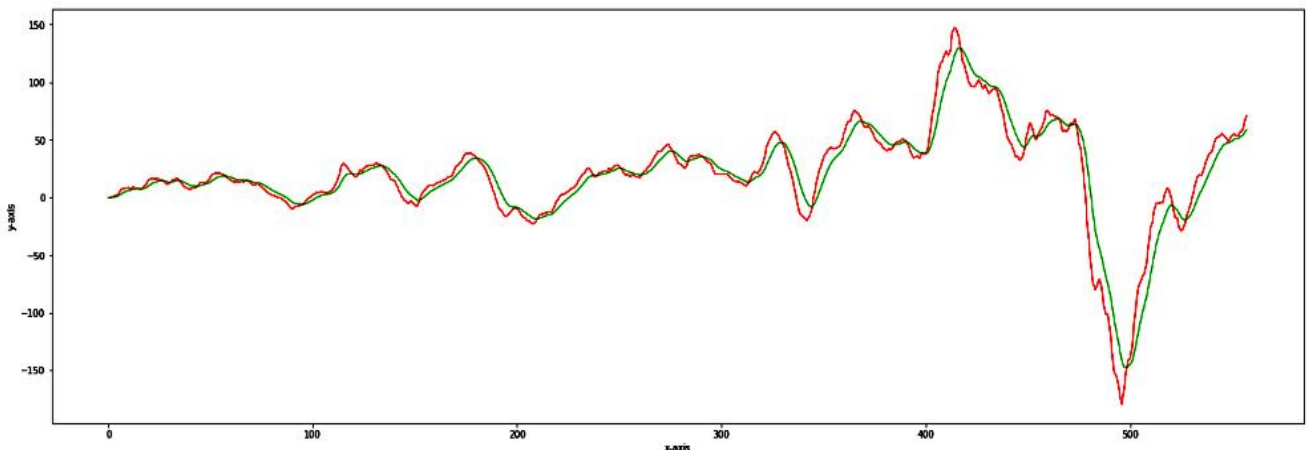


Figure 6: Graphical representation of MACD indicator.

ROC (Rate of Change): Figure 7 shows that how ROC works as it is plotted against zero. If price changes is more inclined towards positive direction it means that price changes will be upward (positive) from the zero margin and if its moving towards negative direction then price changes

will be downward (negative) from the zero margin. You can also spot the divergences and crossovers with the help of ROC. Formula for ROC is given below.

$$ROC = \left(\left(\frac{\text{Current Value}}{\text{Previous Value}} \right) - 1 \right) \times 100 \tag{5}$$

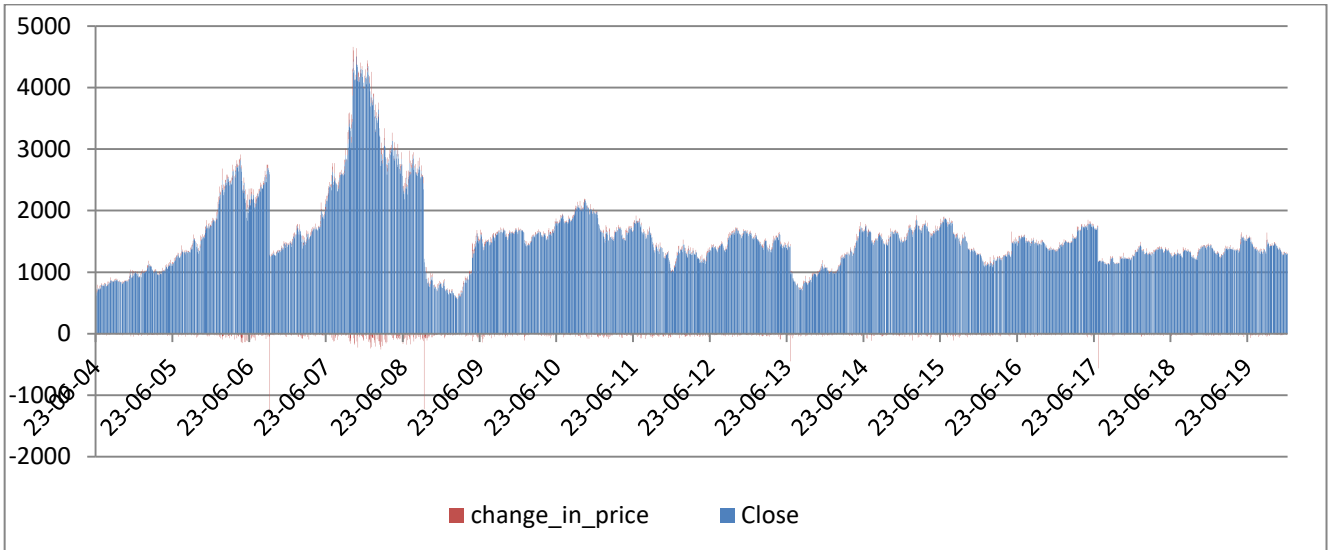


Figure 7: Graphical representation of ROC indicator.

Bollinger Band it is a chart of statistic which gives the information about the price and volatility over a period of time. Graphical representation of bollinger band is given in the Figure 1.4. It basically consists of two bands upper and lower. It gives you an indication whenever the closing price touches the upper band that it is going to be a time

slot where the stock will rise and will fall down in the case lower band. Formula is given below.

$$\%B = \frac{\text{Last} - \text{LowerBB}}{\text{UpperBB} - \text{LowerBB}} \tag{6}$$

$$\text{BandWidth} = \frac{\text{UpperBB} - \text{LowerBB}}{\text{MiddleBB}} \tag{7}$$

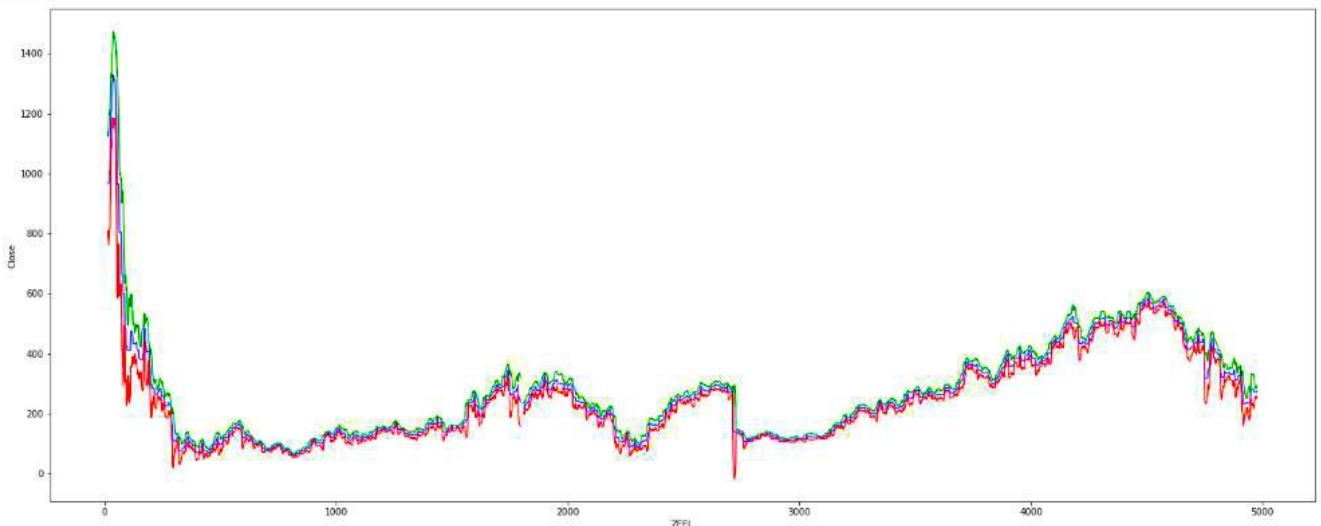


Figure 8: Representation of Bollinger Bands, which consist of three bands i.e last close (blue colour line in the middle), upper band (shown by green colour) and lower band (shown by red colour).

After the implementation of all the six indicators (ROC, BB, Stochastic Oscillator, Williams %R, MACD, RSI) result has been obtained with the help of Random Forest algorithm as shown in the Figure 9. Here a prediction column has also been created which gives you day wise prediction. It gives you an indication for the stock market

by +1(Positive) and -1 (Negative). +1 indicates that stock will rise so this is the right time for the investment where as -1 indicates that stock will fall down so be careful with the investment as shown in the prediction column of the Figure 9.

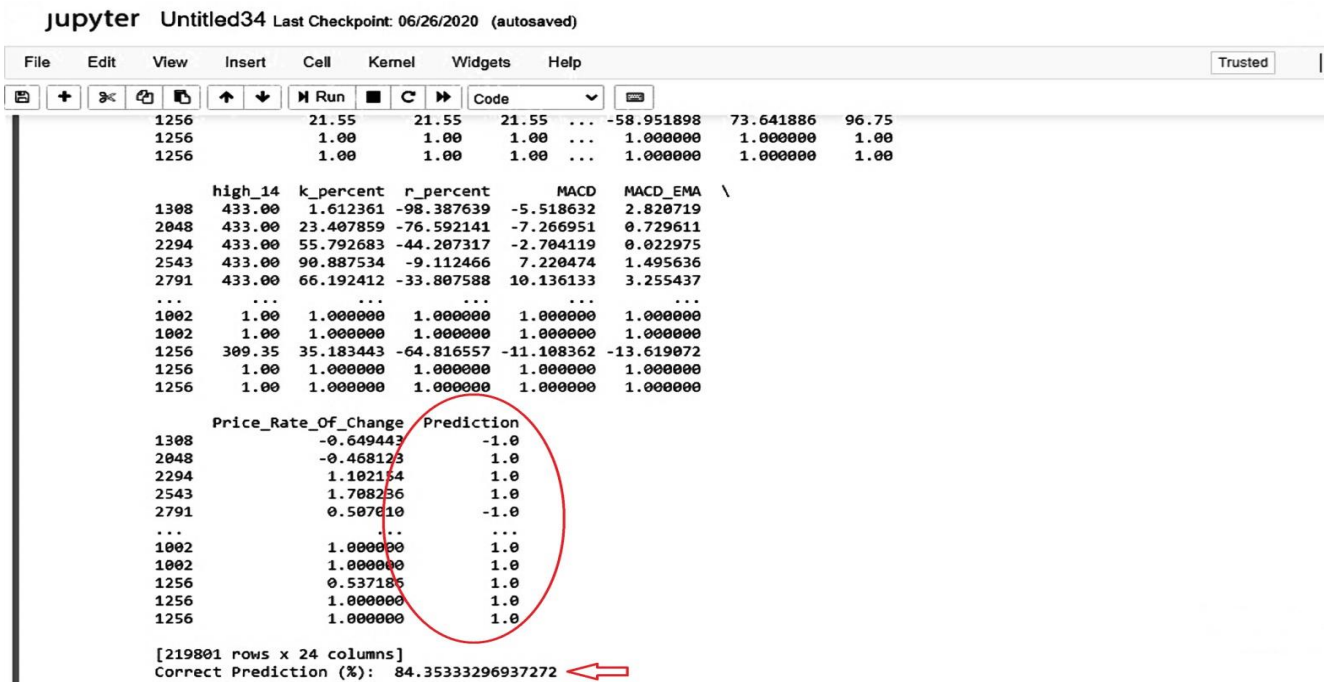


Figure 9: Showing the prediction column and the accuracy of random forest model.

Accuracy has also been calculated for this work in the output as shown in the Figure 9. In total 219801 rows and 24 column were there at the end. And it consist of nineteen years of stock data of forty eight organization from Nifty-50.

Figure 10 and 11 are showing the prediction and accuracy of Artificial Neural Network Model. All the three layers of ANN model has also been shown in the Figure 10 which takes the four parameters i.e Low, High, Open and Close. On the basis of these parameters it will give you the prediction for the stock by +1 (Positive) and -1 (Negative) which works same as explained in the case of random forest model.

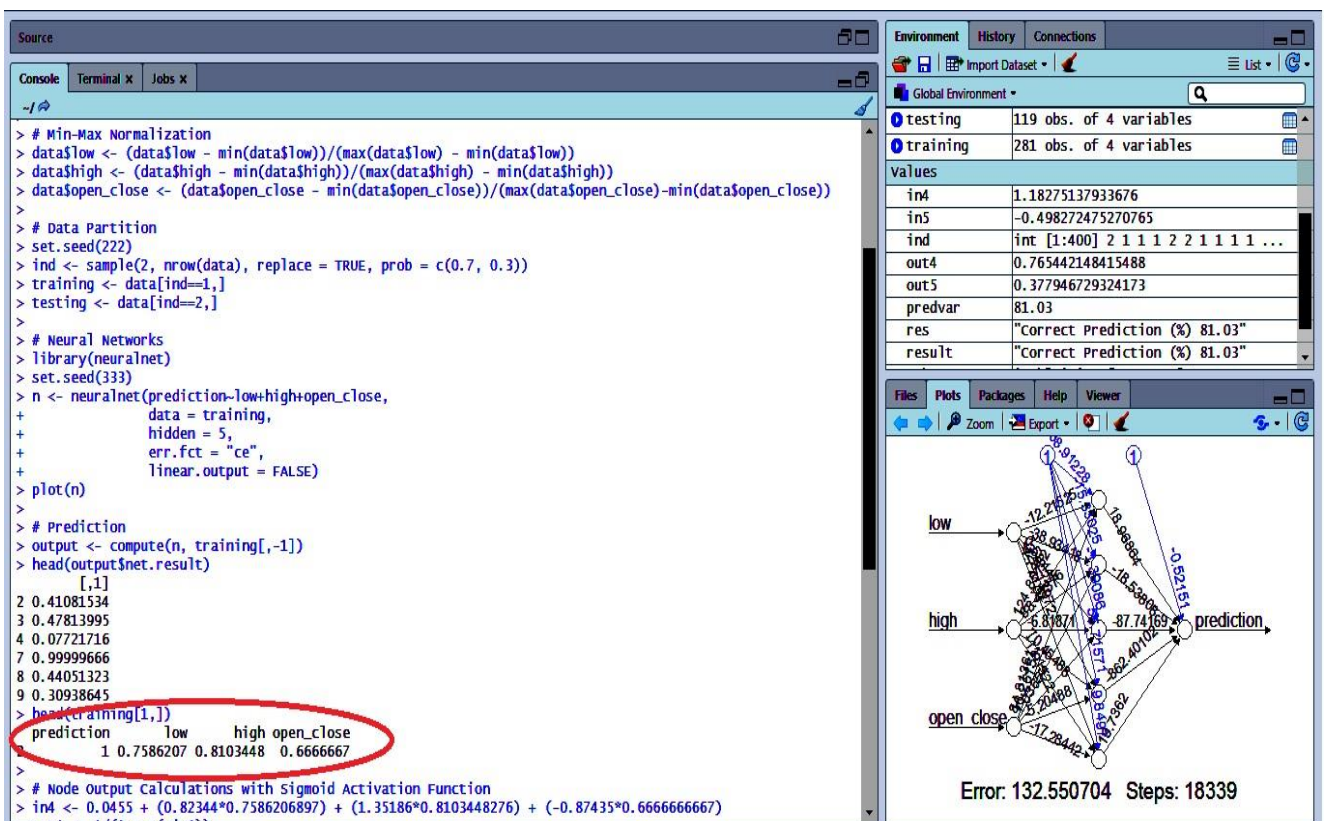


Figure 10: Shows the prediction column of Artificial neural network model.

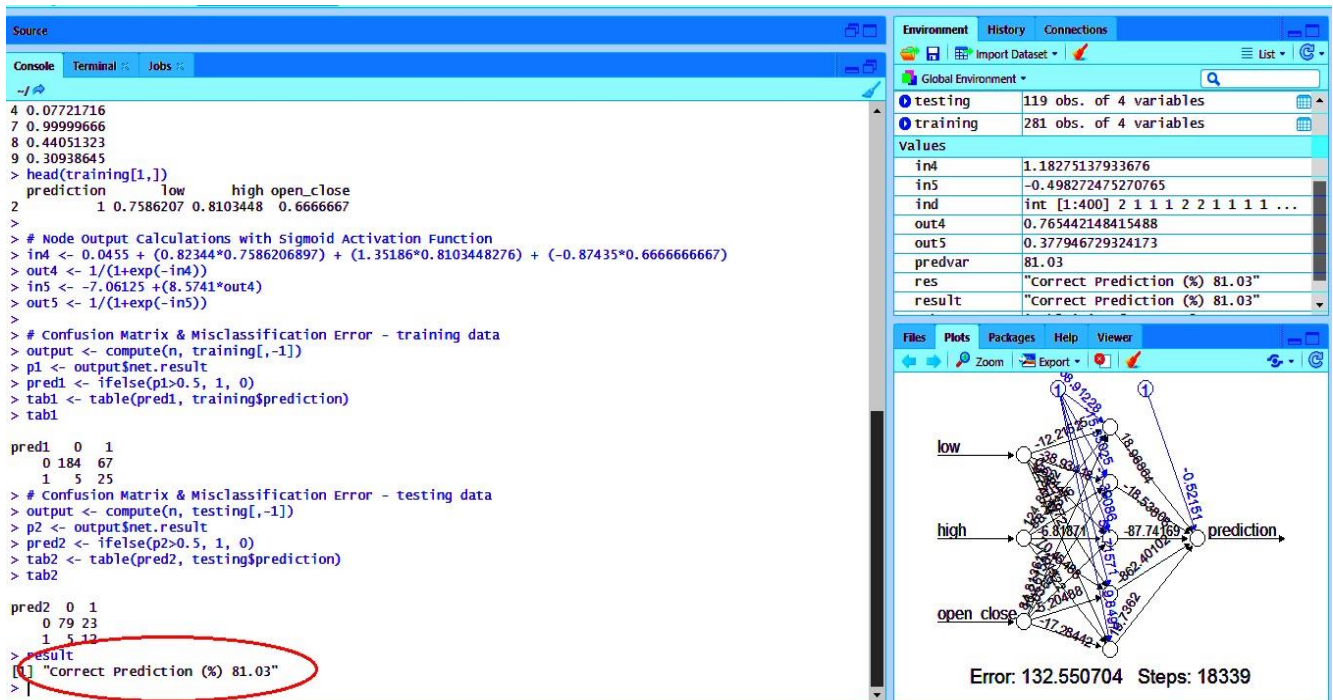


Figure 11: Showing the accuracy of Artificial neural network model.

Accuracy: Accuracy considers all the testing samples. In this work accuracy is calculated with the help of accuracy_score function from Sklearn package of machine learning and it depends on the number of test sets that a model has made. Formula to obtain the accuracy is given in the equation number 8.

$$\text{Accuracy} = \frac{(tp+tn)}{(tp+tn)+(fp+fn)} \quad (8)$$

Where,

Tp = True Positive.

Fp = False Positive.

Tn = True Negative.

Fn = False Negative.

V. CONCLUSION AND FUTURE SCOPE

In this research we have compared the two machine learning techniques (Artificial Neural Network model and Random Forest model) on the stock data, which provides you the prediction of stock by giving you an indication for that. Nineteen years of stock data from Nifty-50 was taken as sample and with the help of six technical indicators (RSI, Stochastic Oscillator, Williams %R, ROC, Bollinger Band) an accuracy of 84.3% in the case of Random Forest and 81% in the case of Artificial Neural Network model have been obtained and it has found that Random Forest model gives the better prediction and results in the case of large data set. It does not lose the accuracy with the increment of the data. This method can be applied on to the stock market of any nation to test the working of these six technical indicators and the results can also be improved with the use of some other stock indicators like Accumulation Distribution, ATR, and On Balance Volume which are available in the world of stock market.

ACKNOWLEDGEMENT

The Authors would like to thank their families for providing all kind of support financially and socially

throughout the research. We want to thank Guru Nanak Dev Engineering College also for providing the good environment for research.

REFERENCES

- [1] J. Patel, S. Shah, P. Thakkar and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques", *Expert Systems with Applications*, vol. **42**, no. **1**, pp. 259-268, **2015**.
- [2] P. Aithal, A. Dinesh and M. Geetha, "Identifying Significant Macroeconomic Indicators for Indian Stock Markets", *IEEE Access*, vol. **7**, pp. 143829-143840, **2019**.
- [3] Y. Alsubaie, K. Hindi and H. Alsalman, "Cost-Sensitive Prediction of Stock Price Direction: Selection of Technical Indicators", *IEEE Access*, vol. **7**, pp. 146876-146892, **2019**.
- [4] A. Giri and P. Joshi, "The Impact of Macroeconomic Indicators on Indian Stock Prices: An Empirical Analysis", *Studies in Business and Economics*, vol. **12**, no. **1**, pp. 61-78, **2017**.
- [5] P. Kanade, "Machine Learning Model for Stock Market Prediction", *International Journal for Research in Applied Science and Engineering Technology*, vol. **8**, no. **6**, pp. 209-216, **2020**.
- [6] H. M, G. E.A., V. Menon and S. K.P., "NSE Stock Market Prediction Using Deep-Learning Models", *Procedia Computer Science*, vol. **132**, pp. 1351-1362, **2018**.
- [7] M. Vijh, D. Chandola, V. Tikkiwal and A. Kumar, "Stock Closing Price Prediction using Machine Learning Techniques", *Procedia Computer Science*, vol. **167**, pp. 599-606, **2020**.
- [8] Sharma, N., & Juneja, A., "Combining of random forest estimates using LSboost for stock market index prediction". *2017 2Nd International Conference For Convergence In Technology (I2CT)*.
- [9] Y. Snezhko, "The use of technical analysis indicators in the Russian stock market", *Russian Journal of Entrepreneurship*, vol. **16**, no. **16**, p. 2681, **2015**.
- [10] M. Paluch and L. Jackowska-Strumiłło, "Prediction of Closing Prices on the Stock Exchange with the Use of Artificial Neural Networks", *Image Processing & Communications*, vol. **17**, no. **4**, pp. 275-282, **2012**.

- [11] P. Pai and C. Liu, "Predicting Vehicle Sales by Sentiment Analysis of Twitter Data and Stock Market Values", *IEEE Access*, vol. 6, pp. 57655-57662, 2018.
- [12] T. Gao and Y. Chai, "Improving Stock Closing Price Prediction Using Recurrent Neural Network and Technical Indicators", *Neural Computation*, vol. 30, no. 10, pp. 2833-2854, 2018.
- [13] Z. Peng, "Stocks Analysis and Prediction Using Big Data Analytics", 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS).
- [14] A. Moghaddam, M. Moghaddam and M. Esfandyari, "Stock market index prediction using artificial neural network", *Journal of Economics, Finance and Administrative Science*, vol. 21, no. 41, pp. 89-93, 2016.
- [15] M. Firdaus, S. Pratiwi, D. Kowanda and A. Kowanda, "Literature review on Artificial Neural Networks Techniques Application for Stock Market Prediction and as Decision Support Tools", 2018 Third International Conference on Informatics and Computing (ICIC).
- [16] M. Kumar and T. M., "Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest", *SSRN Electronic Journal*, 2006.
- [17] R. Nivetha and C. Dhaya, "Developing a Prediction Model for Stock Analysis", International Conference on Technical Advancements in Computers and Communications (ICTACC), 2017.

AUTHORS PROFILE

ANKIT KUMAR is currently pursuing the M.Tech degree in Computer Science & Engineering Department of Guru Nanak Dev Engineering College, Ludhiana (Punjab), India. He has completed his B.Tech degree in Computer Science & Engineering in 2018.



He has also published one Research paper in IIC-STEMPH International Conference held at Nanyang Technological University, Singapore in 2017.

PF. JASBIR SINGH SAINI is working as Associate Professor cum System Analyst at Guru Nanak Dev Engineering College, Ludhiana, Punjab, India. Presently he is the Head of Central Computer Center of the college and also serving as a Head of Computer Applications



Department. His areas of interest are Computer Networks, Cyber Security, DBMS, Big Data and Cloud Computing.