

CORPORATE PERFORMANCE: SMEs PERFORMANCE PREDICTION USING THE DECISION TREE AND RANDOM FOREST MODELS

Anjali Munde^{*}, Nandita Mishra^{**}

^{*} SOIL, School of Business Design, Gurugram, India

^{**} Corresponding author, Department of Management and Engineering, Linköping University, Linköping, Sweden

Contact details: Department of Management and Engineering, Linköping University, Campus Valla, A-huset, 581 83 Linköping, Sweden



Abstract

How to cite this paper: Munde, A., & Mishra, N. (2022). Corporate performance: SMEs performance prediction using the decision tree and random forest models. *Corporate Ownership & Control*, 20(1), 103–113.

<https://doi.org/10.22495/cocv20i1art10>

Copyright © 2022 The Authors

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

<https://creativecommons.org/licenses/by/4.0/>

ISSN Online: 1810-3057

ISSN Print: 1727-9232

Received: 25.08.2022

Accepted: 28.11.2022

JEL Classification: C52, C53, G41, G17

DOI: 10.22495/cocv20i1art10

Stock markets are volatile and continue to alter based on the functioning of the company, historical documents, market-rate, and news updates with the timings. Stock price prediction is the utmost stimulating assignment. In the present communication, a study with data on the stock prices of the top small and medium-sized enterprises (SMEs) in the National Stock Exchange of India (NSE) was utilized to estimate the functioning of the technique executed. The results of this study demonstrate the impact of COVID-19 on the financial distress of SMEs and also helps us in understanding how a better prediction model can help in predicting financial distress. Many studies have been conducted to estimate the bankruptcy of the SME sector using accounting-based financial. But in this study, the leading principle was to exemplify the means to utilize machine learning (ML) algorithms in the bankruptcy prediction of SMEs. The outcomes from the proposed a *decision tree* and a *random forest* prototype are observed to be effective with a high accuracy rate. The study has practical implications on the prediction accuracy and practical value for banks in supporting the financial decision and can be used to access the loan applications of SMEs.

Keywords: Stock Price Prediction, Relative Strength Index, Bollinger Bands, Machine Learning, Decision Tree, Random Forest

Authors' individual contribution: Conceptualization — A.M.; Methodology — A.M.; Formal Analysis — A.M.; Investigation — N.M.; Resources — N.M.; Data Curation — A.M.; Writing — Original Draft — A.M. and N.M.; Writing — Review & Editing — N.M.

Declaration of conflicting interests: The Authors declare that there is no conflict of interest.

1. INTRODUCTION

The stock market price trends are referred to as the upward and downward movement of the stock price. This is also referred to as bear and bull respectively. Understanding the stock movement and predicting the future movement has generated a range of methods and models (Malkiel, 1999). For many years, forecasting trends in stock market prices has been an area of interest for researchers because of its complicated and vigorous nature. Intrinsic volatility in the stock market makes

the prediction even more complex. Economic conditions, political stability, investors' sentiments towards a specific company, market psychology, traders' expectations, other uncertainties, and lots of variables make the prediction even more difficult. Stock markets are extremely unstable and create enormous data on daily basis. Stock price extrapolation is one of the crucial subjects in research and a huge task because of its complicated and unstable quality. The scheduling issue to purchase at low prices and trade at a high price is a non-trivial challenge. To elucidate this issue,

the estimation of the share prices trend is imperative to be studied.

SMEs and their growth are of prime importance for the global economy, especially for a country like India where the SME sector is the second largest employer. The Indian government has implemented various policies to support the growth of SMEs including better credit and tax policies. Nevertheless, access to credit and funds remain a challenge for SMEs. SMEs also lack reliable data leading to the poor credit rating and mistrust. A better prediction model for their financial distress can be the much-awaited answer to all these problems. This study tries to address these challenges by proposing a prediction model which is more accurate and is not dependent on the accounting data.

Stock markets are dynamic, nonlinear, noisy, chaotic, and non-parametric (Abu-Mostafa & Atiya, 1996; Khaidem et al., 2016). Therefore, the stock market price movement is deemed to be a random method. Technical analysis is a technique that aids in investigating share price trends through past price data which helps in envisaging the impending price of a stock. Technical analysis includes forecasting stock prices utilizing technical indicators such as simple moving average (SMA), exponential moving average (EMA), relative strength index (RSI), and moving average convergence/divergence (MACD). The primary methodologies used to predict stock market prices are: 1) technical analysis, 2) time-series forecasting, 3) machine learning (ML) data mining, and 4) modelling and predicting the volatility of stocks (Khaidem et al., 2016). The methodology that is discussed in this paper is machine learning and data mining applications in the stock market. The chief principle of this study was to exemplify in what manner to utilize ML Algorithms such as a *decision tree* and a *random forest*. Furthermore, the results of the comparison are done on different performance measures.

Accurate prediction of the stock market price can guide investors and traders in deciding their strategies and, therefore, increase the probability of gaining profits and reducing the chances of losses. Many researchers have focused on this area and studied it as a regression as well as a classification problem. It has been observed that taking this as a classification problem has given better and more accurate results. This is mainly because of the nature of the prediction where the movement may take two directions upward (bull) or downward (bear) and the classification model also, classifies the algorithms into binary. Classification algorithms used in machine learning have gained a lot of traction recently. In this paper, two prominent methods of machine learning have been used for the prediction of stock market prices. By using the decision tree and the random forest model we have tried to predict the stock price fluctuation which will be beneficial and useful for investors and traders. It will also give insight to policymakers about the future of the stock market. This paper has many practical and theoretical implications.

Small and medium-sized enterprises (SMEs) represent the strength of national economies in many countries. SMEs are the prevalent types of companies engaged in economic cooperation and typically account for two-thirds of

the total employment (Altman et al., 2010). They also contribute to the bank's profitability (Shin & Kolari, 2004). Especially during and after COVID-19 SMEs from all over the world have been impacted and many are in financial distress. This type of situation not only contributes to loss of employment but also impacts the bank's profitability and hence the country's economy. This study is an attempt to understand how better prediction model can help us in predicting financial distress.

The study's main contributions are as follows. First, this work extends the literature on SME default prediction models using stock price data. Second, this study compares and contrasts the prediction accuracy level of different models.

The remaining portion of this paper is organized as follows, Section 2 contains the literature review and Section 3 deals with the details of the data set and methodology. Section 4 highlights the empirical result from the data set and Section 5 discusses the results obtained by this study. Section 6 concludes and cites the recommendations and limitations.

2. LITERATURE REVIEW

Many studies have shown that supervised algorithms are more effective and accurate in forecasting stock market trends (Ballings et al., 2015; Kumar & Thenmozhi, 2006; Mishra et al., 2021). Several studies have concentrated on comparing different prediction algorithms to ascertain the superior algorithm. Kumar and Thenmozhi (2006) applied the random forest and support vector machine (SVM) learning method on the daily movements of the Nifty Market in India to compare the two models. Ballings et al. (2015) used the data of European companies to predict the stock price and used different machine learning models to do that. Furthermore, one more study by Ou and Wang (2009) investigated the predictive aptitude of ten machine learning algorithms in predicting the stock price of the Hang Seng Index (HSI) of the Hong Kong stock market. Similar work was done by Dai and Zhang (2013) where four machine learning algorithms have been compared in order to determine the most effective algorithm in predicting both, short and long-term stock price trends. Moghar and Hamiche (2020) built a model using recurrent neural networks (RNN) and a long-short-term model (LSTM) to predict future stock market values.

Subasi et al. (2021) presented a comparison of stock market prediction by inputting different classifiers and then the machine learning algorithm was tested against the NASDAQ, NYSE, Nikkei and FTSE.

Anbalagan and Maheswari (2015) suggested a Fuzzy Metagraph (FM) created on share market analysis, categorization and estimation for miniscule shareholders of the Indian share market and included Fuzzy Metagraph with SMA, MACD and RSI in the classification and prediction of share market investment. Wang and Kim (2018) built a successful technique for envisaging the share price movement and computed EMA by means of a varying weight established on the past instability and further evaluated the consistency of MACD-HVIX with MACD. Vijh et al. (2020) applied artificial neural

network (ANN) and the random forest methods for forecasting the subsequent closing price for the different companies and further assessed the prototypes by means of root-mean-square error (RMSE) and mean absolute percentage error (MAPE). Masoud (2013) examined the linking association between share market functioning and financial development and recommended a positive association between well-organized share markets and financial development. The backpropagation (BP) algorithm is utilised for training and the multilayer feed-forward network (MFFNN) is used for forecasting price. Chong and Ng (2008) investigated the MACD and the RSI to check whether these are beneficial and observed that both of them can make earnings better than the buy-and-hold approach applied frequently. Roman and Jameel (1996) suggested a novel method that supported proposing a set of investments across several share markets and examined the function of recurrent networks to the share market return forecast issue in comparison with backpropagation networks.

Mizuno et al. (1998) exhibited a neural network prototype for practical examination of the share market and its function to purchase and sell planning forecast technique for share index and suggested a cognitive process for refining forecast precision of new classes, monitoring the quantities of learning models utilising data regarding the significance of each class. Lahmiri (2018) introduced a method for examining share price performance built on unique classes of practical examination metrics and numerous extrapolative techniques. The extrapolative technique shown is created on an ensemble of neural networks linked to particle swarm intelligence for factor optimization. Moghaddam et al. (2016) examined the capability of artificial neural network in predicting the daily NASDAQ stock exchange rate by measuring the short-term past share prices as well as the day of the week as inputs. Guresen et al. (2011) assessed the efficiency of neural network prototypes which are identified are valuable in share-market estimations. The techniques studied are multi-layer perceptron, dynamic artificial neural network and hybrid neural networks that utilize generalized autoregressive conditional heteroscedasticity to obtain different input variables. Jiang (2021) examined the recent works on deep learning models for stock market prediction.

In this communication, a model is created that obtains the direction of the stock price trend based on several input variables. The machine learning algorithms comprising the decision tree and the random forest are employed and it is observed that utilising these two methods exhibits improvement in the accuracy of trend estimation. All these rules built predicting prototypes obtain main technical indicators as an input to envisage the movement of share price in increasing trend or decreasing trend.

The rationale of the research is to construct an application that can contribute to examining trends in share price progress and provide assistance to investors to acquire information and suggestion pertinent to these shares. The chief contribution of the study is to provide indicators to the trader about what is the correct time to purchase or sell applying the collective indicator variables

from SMA, EMA, RSI and MACD. This paper directs at determining the association of technical indicators with the decision tree and the random forest models, a robust and easily expressible algorithm, to envisage the trend of a stock price. Technical indicators such as SMA, EMA, RSI and MACD are considered.

3. RESEARCH METHODOLOGY

3.1. Data source

The study covers five SMEs listed on the National Stock Exchange of India (NSE). The selected five SMEs are leaders in their sector and have the maximum market share out of the total number of SMEs listed on the Indian stock market which is 394.

The stock data are acquired on a daily basis from April 11, 2011, to April 26, 2021, from the Yahoo Finance database, which contains high, low, close, adjacent close and volume for the five listed SMEs. The dataset is split into 70% as a training set and 30% as a testing set.

3.2. Method

Python applies machine learning methods essentially classification analysis, regression, recommendation, and data clustering (Sarker, 2021). After data pre-processing, a dataset is constructed by splitting it into two parts: 70% employed to train the models and 30% employed as a test dataset. The training dataset and test dataset are one-hot encoded, i.e., nominal variables are altered into numerical form to be postulated to distinctive machine learning algorithms for effectual extrapolation. K is chosen for K -fold cross-validation to approximate the precision of several models.

The decision tree is a renowned non-parametric supervised learning method. The decision tree techniques are utilised for the classification and regression assignments (Pedregosa et al., 2011).

ID3 algorithm (Quinlan, 1986), C4.5 (Quinlan, 1993), and classification and regression trees (CART) (Breiman et al., 1984) are recognised for the decision tree (DT) algorithms. Furthermore, lately recommended BehavDT and IntruDTree (Sarker, 2020) are operational in the pertinent application fields, for example, customer performance analytics and cybersecurity analytics, correspondingly.

By organising the tree from the root to leaf nodes, the decision tree arranges the illustrations. Examples are categorized by examining the characteristic outlined through that node, beginning at the root node of the tree, and later moving down the tree branch consequent to the characteristic value. For splitting, the widespread conditions are "Gini" for the Gini impurity and "entropy" for the information gain that can be stated mathematically (Pedregosa et al., 2011):

$$H(x) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (1)$$

$$Gini(E) = 1 - \sum_{i=1}^c p_i^2 \quad (2)$$

The decision tree is a supervised machine learning algorithm utilised for classification and regression problems. The decision tree is a string of subsequent decisions built to achieve a definite outcome.

The random forest is a tree-based machine learning algorithm that controls the potential of numerous decision trees for building decisions.

The decision tree utilises Gini impurity to determine which element moves to a decision node. Gini impurity is a degree of uniformity of the nodes. Therefore, by reducing the Gini impurity the decision tree locates the elements that split the data in the finest way.

One of the very frequently utilized approaches for data mining is the decision tree algorithm, which is extensively tapped in several disciplines. It has experienced an extended procedure of working from low to concentrated and from easy to difficult.

The decision tree is a hierarchical, iterative split that functions as a top-down, segregate-and-capture method, and its fundamental algorithm is principally greedy (Baldi, 2012). Preliminary from the root node, each non-leaf node is obtained to discover a characteristic in its consequent sample set to investigate the sample set, and the training sample set is segregated into numerous subsamples agreeing to the distinct outcomes of the test. Each subsample set comprises an original leaf node, and the procedure is replicated for the latest leaf node as a result the loop remains to spread an explicit closing situation (Balducci et al., 2018).

The principal gains of the decision tree-based learning algorithm are that it does not entail the consumer to attain a quantity of fundamental information through the learning procedure.

The procedure of fabricating the decision tree is distributed into two phases: tree building and pruning. The first phase is the tree building stage, which chooses a portion of the training data and constructs the decision tree through the breadth-first iterative algorithm till all leaf node pertains to the identical class. The second phase is the pruning phase, which consumes the outstanding data to verify the produced the decision tree and rectify the faults, and thus lastly prunes the decision tree and enhances nodes till an accurate the decision tree is constructed. The decision tree building algorithm is an iterative procedure that eventually develops into the decision tree, and pruning diminishes the influence of noisy data on classification accuracy. As a whole, the higher the information gain, the higher the "purity improvement" attained through attributes to segregate the dataset. Thus, information gain could be utilized to choose characteristics for the decision tree partitioning, which select the characteristics with the highest information gain.

The random forest classifier (Breiman, 1996) is an acclaimed ensemble classification technique that is employed in the area of machine learning and data science in several application fields. This technique aids "parallel ensembling" that corresponds to the numerous decision tree classifiers in parallel on distinct data set sub-samples and aids most voting or averages for the conclusion. It reduces the over-fitting issue and surges the extrapolation accuracy and restriction (Pedregosa et al., 2011). Thus, the random forest learning model with numerous decision trees is

characteristically more precise than a specific decision tree-based model (Sarker et al., 2019a, 2019b). To construct a sequence of decision trees with regulated alteration, it merges bootstrap aggregation (bagging) (Breiman, 1996) and random feature selection (Amit & Geman, 1997). It is conformable to both classification and regression difficulties and corresponds to both definite and continuous values.

The random forest is a sequence algorithm recommended by Breiman (2001), where if the envisaged outcome is a discrete value, it is the random forest classification, and if it is a continuous value, it is the random forest regression. Several experimental findings have established the principle that the random forest algorithm has high estimation accuracy with suitable acceptance for irregular value and noise.

The random forest classification algorithm is depleted into two stages. First, the random forest algorithm obtains subsamples from the fundamental samples utilising the bootstrap resampling technique and generates decision trees for each sample. Second, the algorithm categorizes the decision trees and applies a simple vote, with the chief vote of the classification as the conclusive outcome of the estimation. The random forest algorithm continually comprises three steps as follows:

1) *Choose the training set.* Employ the bootstrap random sampling method to recover K training sets from the innovative dataset (M properties), with the dimension of each training set the same as that of the unique training set.

2) *Construct the random forest model.* Establish a classification regression tree for every of the bootstrap training series to construct K -decision trees to develop a "forest", these trees are not pruned. Examining the progress of every tree, this method does not select the greatest characteristic as internal nodes for branches however the branching procedure is a random range of $m \ll M$ of all characteristics.

3) *Build simple voting.* Subsequently, the training procedure of every decision tree is independent, the training of the random forests can function in correspondence, which substantially corrects effectiveness. The random forest can be generated by uniting K -decision trees trained in the similar condition. When categorizing the input samples, the outcomes vary on the easy voting of the output of each decision tree. The random forest algorithm ascertains the samples by fabricating a sequence of independent and allocated decision trees and concludes the final group of the sample corresponding to every decision tree.

3.3. Technical indicators of the stock market

Technical indicators are applied to foster the characteristic of the data, enhance the effectiveness and ease the stock estimation method.

The technical indicators used in this study in predicting the direction of the predicting the direction are:

- simple moving average (SMA);
- exponential moving average (EMA);
- moving average convergence/divergence (MACD);
- relative strength index (RSI);
- Bollinger Bands.

The utmost prevalent kinds of moving averages are the SMA and the EMA.

SMA is computed by choosing the average value of the price within a definite time limit.

EMA is calculated to enhance the concept of a simple moving average by assigning additional weight to the highest and latest price data, which is believed to be more significant than previous data (Jelena et al., 2015).

MACD value attempts to predict share price trends by evaluating short and long-term trends. It is the difference between 26-day and 12-day EMA.

A nine-day EMA, termed as the "signal" line is outlined on the uppermost of the MACD to represent purchase/sell chances. If MACD is beyond the signal line, then purchase. If MACD is under the signal, then sell (Anghel, 2015).

The RSI studies that an asset is overpurchased or overvalued (McHugh et al., 2021).

$$RSI = 100 - (100 / (1 + AGL)) \quad (3)$$

where, $AGL = \text{Average profit} / \text{Average loss}$.

When RSI rises to beyond 70 then sell. If RSI is amid 30 and 70 then hold. If RSI declines to under 30 then purchase.

Bollinger Bands identified as volatility bands vary due to instability (Kocer, 2016; Lento et al., 2007). There are the lower Bollinger Band, the upper Bollinger Band and the middle Bollinger Band. The 20-day SMA computes the middle Bollinger Band value.

The upper Bollinger Band values are computed by totalling 20-day SMA and standard deviations. The upper band formula would be: $20_SMA + (20_Standard\ Deviation\ of\ Close * 2)$.

The lower Bollinger Band values are computed by subtracting standard deviations from 20-day SMA. The lower band formula would be: $20_SMA - (20_Standard\ Deviation\ of\ Close * 2)$.

The further sequence of actions involves:

1) The generation of technical indicators of the stock is carried out by implementing a moving average within: 10 days moving average; 20 days moving average; 10 days standard deviation; 20 days standard deviation.

2) Computing RSI indicator: 5 days RSI and 14 days RSI.

3) Calculating MACD: This indicator is the difference between two exponential moving averages: 12 days EMA; 26 days EMA.

4) Computing the Bollinger Bands using the 20-day Moving average.

We clean the output and generated a column called direction basis the following:

- if price > the upper Bollinger Band, and MACD value > MACD signal -> buy signal (1);
- if price < the lower Bollinger Band, and MACD value < MACD signal -> sell signal (-1);
- else, out of the market -> signal OOM (0).

4. RESEARCH RESULTS

In this section, the stock price trend of different companies has been illustrated (see Appendix).

Figures A.1–A.5 represent graphs showing the price trend over time from April 11, 2011, to April 26, 2021, for different companies.

Figures A.6–A.10 represent the MACD from April 11, 2011, to April 26, 2021, for ITC, HDFC Bank, Reliance, TCS and L&T.

Figures A.11–A.15 represent the signals from April 11, 2011, to April 26, 2021, for ITC, HDFC Bank, Reliance, TCS and L&T.

Figures A.16–A.20 represent the Bollinger Band plot from April 11, 2011, to April 26, 2021, for ITC, HDFC Bank, Reliance, TCS and L&T.

The daily stock data of 10 years with high, low, close, and adjacent close details were compared using the decision trees and the random forest model. The review of the literature supports these two models and, therefore, a detailed comparison is conducted to understand which method shows better performance. The comparative test has been performed on the training and test data and the results are shown in Table 1 and Table 2 respectively.

Table 1. Comparative analysis of accuracy values for the training datasets obtained using the decision tree and the random forest model

Company name	Decision tree (in %)	Random forest (in %)
ITC	92.7	91.8
HDFC Bank	93.4	92
Reliance	94.2	91.4
TCS	92.5	89.6
L&T	91.3	89.2

Table 2. Comparative analysis of accuracy values for the testing datasets obtained using the decision tree and the random forest model

Company name	Decision tree (in %)	Random forest (in %)
ITC	89.8	91.4
HDFC Bank	91.9	91.7
Reliance	91.7	90.9
TCS	89.7	89.4
L&T	88.3	89.1

5. DISCUSSION OF THE RESULTS

The comparative analysis of training datasets indicates, that for ITC, HDFC Bank, Reliance, TCS and L&T companies the decision tree confirms to be an effective procedure, providing improved accuracy values displayed in Table 1.

Whereas, the comparative analysis of testing datasets indicates, that for ITC and L&T company random forest demonstrates to be an effective method, providing improved accuracy values and for HDFC Bank, Reliance and TCS companies the decision tree demonstrates to be an effective method, providing improved accuracy values as displayed in Table 2. The comparative analysis constructed on accuracy values indicates that the decision tree model provides higher accuracy in comparison to the random forest model. Results exhibit the finest estimates attained through the decision tree model which provides an accuracy of 94.2% and the random forest model provides an accuracy of 92%.

The results of the prediction model and those discovered to be important factors also can be used as leading indicators to prevent SMEs bankruptcy. The result adds to the current literature on bankruptcy prediction by proposing the use of non-accounting data in the prediction model which also gives additional predictive power to the models.

The study proposes a prediction model where no accounting data is available or is biased. Especially post-COVID, where the importance of accuracy has reached a new level, relying on accounting data produced by the company may distort the accuracy of prediction. Banks can also take benefit from this unbiased prediction for their credit rating.

6. CONCLUSION

Envisaging the stock price trends is a difficult job because of constantly varying stock prices that are based on several factors that develop complicated models. The past dataset accessible on the company's website includes limited elements such as high, low, open, close, adjacent close value of stock prices, the volume of shares traded, etc., that are inadequate. This study explored the function of technical indicators in predicting share price trends and attaining higher accuracy with the decision tree and the random forest techniques. Functioning of the decision tree and the random forest are examined with extrapolative accuracy. Classification accuracy of the decision tree and random forest models shows that technical indicators have a vast impact on the estimation of stock price trend and attained about 94% of accuracy. The comparative

analysis constructed on accuracy values signifies that the decision tree model provides higher accuracy in comparison to the random forest model. Results exhibit the finest estimates attained through the decision tree model provides an accuracy of 94.2% and the random forest model provides an accuracy of 92%. Future scope includes deep learning models that study financial news articles and financial parameters such as a closing price, traded volume, profit and loss statements, etc., for reliable outcomes. A few limitations of the study include that bagging or bootstrap aggregating was not used which can improve the functioning of machine learning classification and regression. Bagging creates a novel training set constructed on a specified training set. It decreases variation and prevents overfitting. Bagging aids in outputting more precise results for unstable methods (Witten et al., 2016). We can employ AdaBoost (Adaptive Boosting) which is utilized by numerous techniques to better the accuracy. The output of all the weak learners is joined into a weighted sum that develops into the final output. AdaBoost exploits the boosting algorithm, but it creates it by joining all the outputs. Separate learners can be weak on their own but together they can prove to be strong learners (Hall et al., 2011).

REFERENCES

1. Abu-Mostafa, Y. S., & Atiya, A. F. (1996). Introduction to financial forecasting. *Applied Intelligence*, 6, 205–213. <https://doi.org/10.1007/BF00126626>
2. Altman, E. I., Sabato, G., & Wilson, N. (2010). The value of non-financial information in SME risk management. *Journal of Credit Risk*, 6(2), 95–127. <https://doi.org/10.21314/JCR.2010.110>
3. Amit, Y., & Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural Computation*, 9(7), 1545–1588. <https://doi.org/10.1162/neco.1997.9.7.1545>
4. Anbalagan, T. S., & Maheswari, S. U. (2015). Classification and prediction of stock market index based on Fuzzy Metagraph. *Procedia Computer Science*, 47, 214–221. <https://doi.org/10.1016/j.procs.2015.03.200>
5. Anghel, G. D. (2015). Stock market efficiency and the MACD. Evidence from countries around the world. *Procedia Economics and Finance*, 32, 1414–1431. [https://doi.org/10.1016/S2212-5671\(15\)01518-X](https://doi.org/10.1016/S2212-5671(15)01518-X)
6. Baldi, P. (2012). Autoencoders, unsupervised learning, and deep architectures. *Proceedings of Machine Learning Research*, 27, 37–49. <https://proceedings.mlr.press/v27/baldi12a.html>
7. Balducci, F., Impedovo, D., & Pirlo, G. (2018). Machine learning applications on agricultural datasets for smart farm enhancement. *Machines*, 6(3), 38. <https://doi.org/10.3390/machines6030038>
8. Ballings, M., Van den Poel, D., Hespels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, 42(20), 7046–7056. <https://doi.org/10.1016/j.eswa.2015.05.013>
9. Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. <https://doi.org/10.1023/A:1018054314350>
10. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
11. Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees* (1st ed.). CRC Press. <https://doi.org/10.1201/9781315139470>
12. Chong, T. T.-L., & Ng, W.-K. (2008). Technical analysis and the London stock exchange: Testing the MACD and RSI rules using the FT30. *Applied Economics Letters*, 15(14), 1111–1114. <https://doi.org/10.1080/13504850600993598>
13. Dai, Y., & Zhang, Y. (2013). *Machine learning in stock price trend forecasting*. Stanford University. <http://cs229.stanford.edu/proj2013/DaiZhang-MachineLearningInStockPriceTrendForecasting.pdf>
14. Dolgopolas, V., Dagiene, S., Minkevicius, S., & Sakalauskas, L. (2014). Python for scientific computing education: Modeling of queuing systems. *Scientific Programming*, 22, 164306 <https://doi.org/10.1155/2014/164306>
15. Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, 38(8), 10389–10397. <https://doi.org/10.1016/j.eswa.2011.02.068>
16. Hall, M., Witten, I., & Frank, E. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Morgan Kaufmann. <https://doi.org/10.1016/C2009-0-19715-5>
17. Jelena, S., Ivana, M., & Milos, S. (2015). Investment strategy optimization using technical analysis and predictive modelling in emerging markets. *Procedia, Economics and Finance*, 19, 51–62. [https://doi.org/10.1016/S2212-5671\(15\)00007-6](https://doi.org/10.1016/S2212-5671(15)00007-6)
18. Jiang, W. (2021). Applications of deep learning in stock market prediction: Recent progress. *Expert Systems with Applications*, 184, 115537 <https://doi.org/10.1016/j.eswa.2021.115537>
19. Khaidem, L., Saha, S., & Dey, S. R. (2016). *Predicting the direction of stock market prices using random forest*. arXiv. <https://doi.org/10.48550/arXiv.1605.00003>

20. Kocer, B. (2016). Bollinger bands approach on boosting ABC algorithm and its variants. *Applied Soft Computing*, 49, 292–312. <https://doi.org/10.1016/j.asoc.2016.08.023>
21. Kumar, A. (2010). *Python for education: Learning maths & science using Python and writing them in BTfiK*. Inter University Accelerator Centre. https://archive.org/stream/Ajith_Kumar_B.P._Python_for_Education/mapy_djvu.txt
22. Kumar, M., & Thenmozhi, M. (2006) Forecasting stock index movement: A comparison of support vector machines and random forest. Paper presented at the *Indian Institute of Capital Markets, 9th Capital Markets*. <https://doi.org/10.2139/ssrn.876544>
23. Lahmiri, S. (2018). A technical analysis information fusion approach for stock price analysis and modeling. *Fluctuation and Noise Letters*, 17(1), 1850007. <https://doi.org/10.1142/S0219477518500074>
24. Langtangen, H. P. (2011). *A primer on scientific programming with Python*. Springer-Verlag. <https://doi.org/10.1007/978-3-642-18366-9>
25. Lento, C., Gradojevic, N., & Wright, C. S. (2007). Investment information content in Bollinger bands. *Applied Financial Economics Letters*, 3(4), 263–267. <https://doi.org/10.1080/17446540701206576>
26. Malkiel, B. G. (1999). *A random walk down Wall Street: Including a life-cycle guide to personal investing*. W. W. Norton & Company.
27. Masoud, N. M. H. (2013). The impact of stock market performance upon economic growth. *International Journal of Economics and Financial Issues*, 3(4), 788–798. <https://www.econjournals.com/index.php/ijefi/article/view/557>
28. McHugh, C., Coleman, S., & Kerr, D. (2021). Technical indicators for energy market trading. *Machine Learning with Applications*, 6, 100182. <https://doi.org/10.1016/j.mlwa.2021.100182>
29. Mishraz, N, Ashok, S., & Tandon, D. (2021). Predicting financial distress in the Indian banking sector: A comparative study between the logistic regression, LDA and ANN models. *Global Business Review*. Advance online publication. <https://doi.org/10.1177/09721509211026785>
30. Mizuno, H., Kosaka, M., Yajima, H., & Komod, N. (1998). Application of neural network to technical analysis of stock market prediction. *Studies in Informatics and Control*, 7(2), 111–119.
31. Moghaddam, A. H., Moghaddam, M. H., & Esfandyari, M. (2016). Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science*, 21(41), 89–93. <https://doi.org/10.1016/j.jefas.2016.07.002>
32. Moghar, A., & Hamiche, M. (2020). Stock market prediction using LSTM recurrent neural network. *Procedia Computer Science*, 170, 1168–1173. <https://doi.org/10.1016/j.procs.2020.03.049>
33. Ou, P., & Wang, H. (2009). Prediction of stock market index movement by ten data mining techniques. *Mathematical Models and Methods in Applied Sciences*, 3(12), 28–42. <https://doi.org/10.5539/mas.v3n12p28>
34. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
35. Quinlan, J. (1986). Induction of decision trees. *Machine Learning*, 1, 81–106. <https://doi.org/10.1023/A:1022643204877>
36. Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann. <https://doi.org/10.1016/C2009-0-27846-9>
37. Roman, J., & Jameel, A. (1996). Backpropagation and recurrent neural networks in financial analysis of multiple stock market return. *Proceedings of HICSS-29: 29th Hawaii International Conference on System Sciences*, 2, 454–460. IEEE. <https://ieeexplore.ieee.org/document/495431>
38. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2, 160. <https://doi.org/10.1007/s42979-021-00592-x>
39. Sarker, I. H., Abushark, Y. B., Alsolami, F., & Khan, A. (2020). IntruDTree: A machine learning based cyber security intrusion detection model. *Symmetry*, 12(5), 754. <https://doi.org/10.3390/sym12050754>
40. Sarker, I. H., Colman, A., Hun, J., Khan, A. I., Abushark, Y. B., & Salah, K. (2019a). BehavDT: A behavioral decision tree learning to build user-centric context-aware predictive model. *Mobile Networks and Applications*, 25, 1151–1161. <https://doi.org/10.1007/s11036-019-01443-z>
41. Sarker, I. H., Kayes, A. S. M., & Watters, P. (2019b). Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage. *Journal of Big Data*, 6, 57. <https://doi.org/10.1186/s40537-019-0219-y>
42. Shin, G. H., & Kolari, J. W. (2004). Do some lenders have information advantages? Evidence from Japanese credit market data. *Journal of Banking & Finance*, 28(10), 2331–2351. <https://doi.org/10.1016/j.jbankfin.2003.09.006>
43. Subasi, A., Amir, F., Bagedo, K., Shams, A., & Sarirete, A. (2021). Stock market prediction using machine learning. *Procedia Computer Science*, 194, 173–179. <https://doi.org/10.1016/j.procs.2021.10.071>
44. Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia Computer Science*, 167, 599–606. <https://doi.org/10.1016/j.procs.2020.03.326>
45. Wang, J., & Kim, J. (2018). Predicting stock price trend using MACD optimized by historical volatility. *Mathematical Problems in Engineering*, 2018, 9280590. <https://doi.org/10.1155/2018/9280590>
46. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.
47. Zhang, H., Zimmerman, J., Nettleton, D., & Nordman, D. J. (2020) Random forest prediction intervals. *The American Statistician*, 74(4), 392–406. <https://doi.org/10.1080/00031305.2019.1585288>

APPENDIX

Figure A.1. Stock market price trend for ITC
April 11, 2011 – April 26, 2021

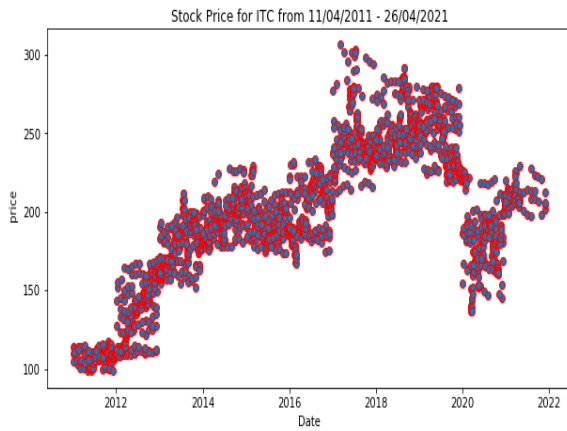


Figure A.2. Stock market price trend for HDFC Bank
April 11, 2011 – April 26, 2021

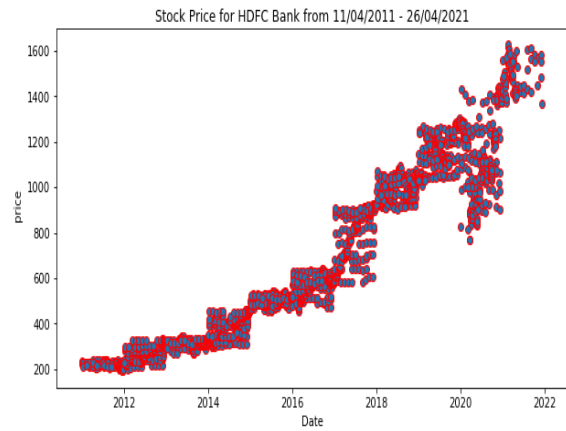


Figure A.3. Stock market price trend for Reliance
April 11, 2011 – April 26, 2021

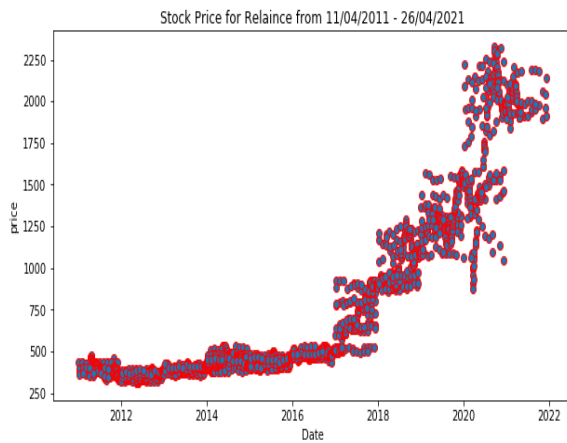


Figure A.4. Stock market price trend for TCS
April 11, 2011 – April 26, 2021

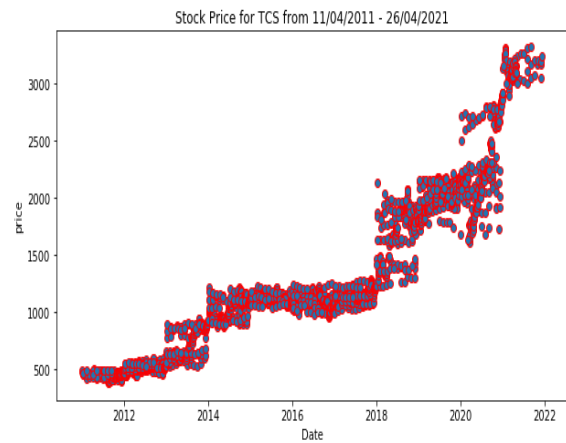


Figure A.5. Stock market price trend for L&T
April 11, 2011 – April 26, 2021

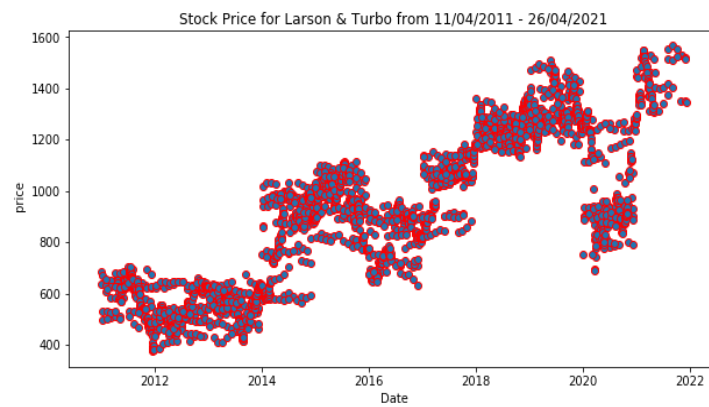


Figure A.6. MACD for ITC
April 11, 2011 - April 26, 2021

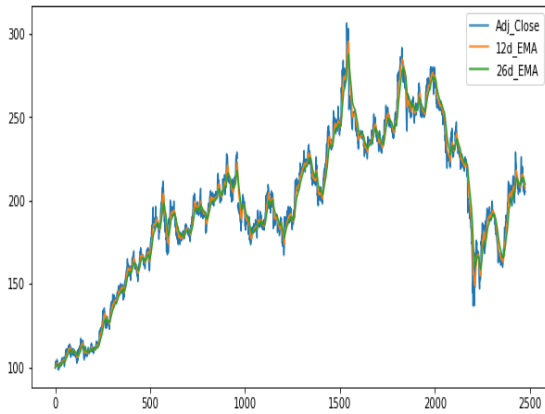


Figure A.7. MACD for HDFC Bank
April 11, 2011 - April 26, 2021

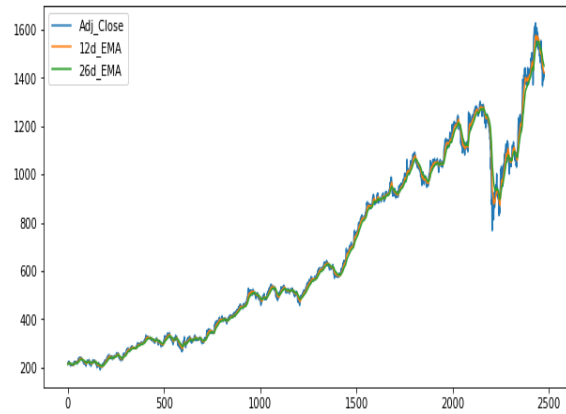


Figure A.8. MACD for Reliance April 11, 2011 - April 26, 2021

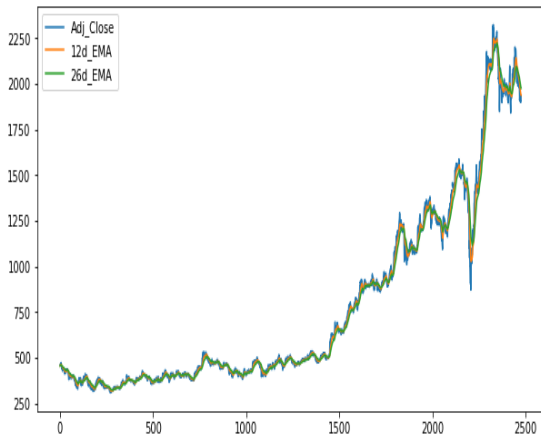


Figure A.9. MACD for TCS
April 11, 2011 - April 26, 2021

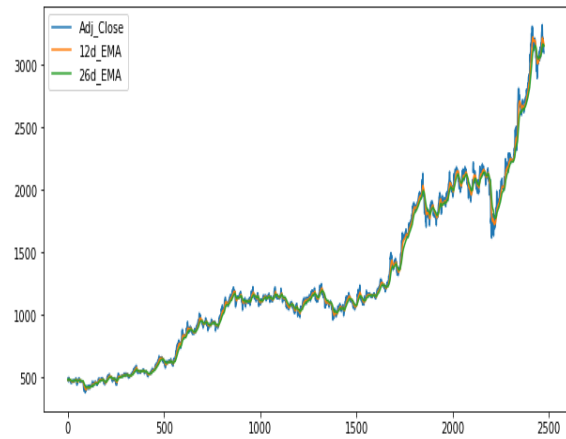


Figure A.10. MACD for L&T April 11, 2011 - April 26, 2021

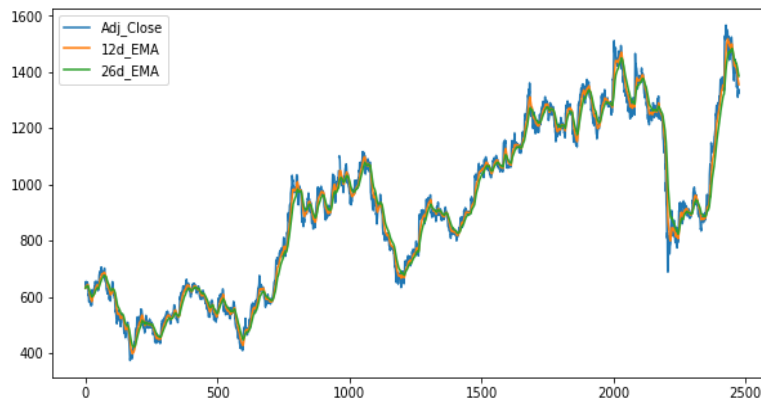


Figure A.11. Signal for ITC
April 11, 2011 - April 26, 2021

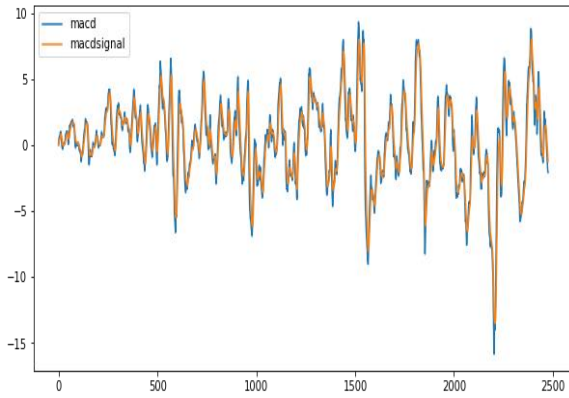


Figure A.12. Signal for HDFC Bank
April 11, 2011 - April 26, 2021

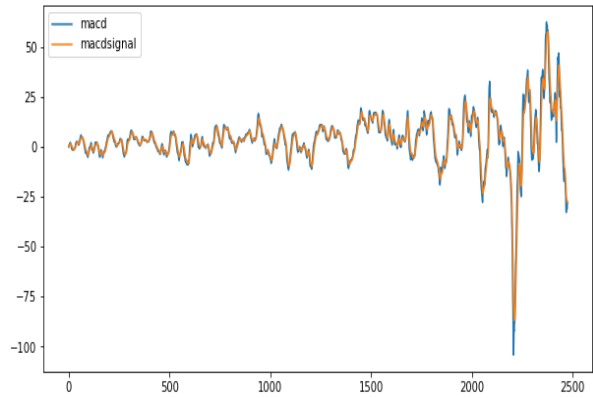


Figure A.13. Signal for Reliance
April 11, 2011 - April 26, 2021

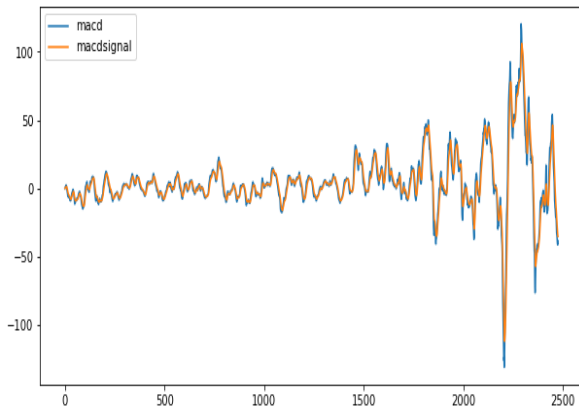


Figure A.14. Signal for TCS
April 11, 2011 - April 26, 2021

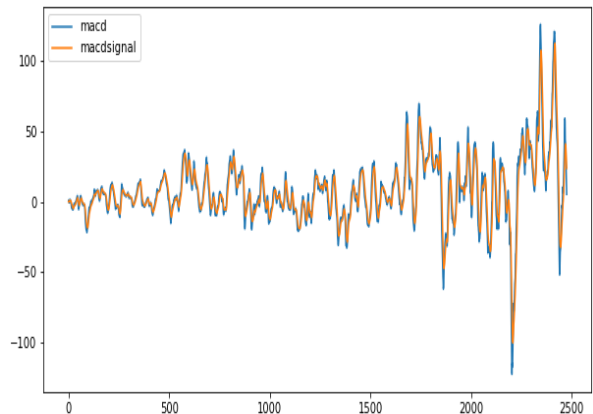


Figure A.15. Signal for L&T April 11, 2011 - April 26, 2021

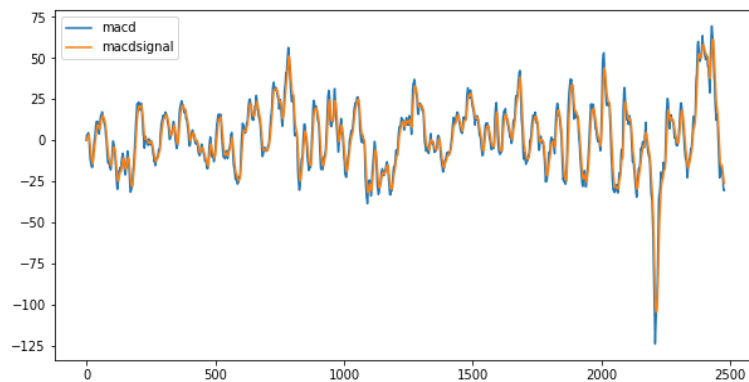


Figure A.16. Bollinger Band for ITC
April 11, 2011 - April 26, 2021

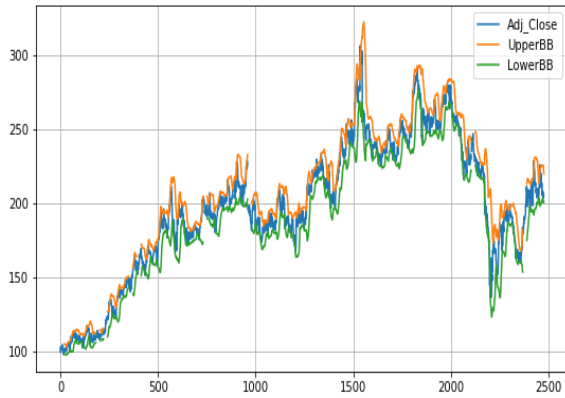


Figure A.17. Bollinger Band for HDFC Bank
April 11, 2011 - April 26, 2021

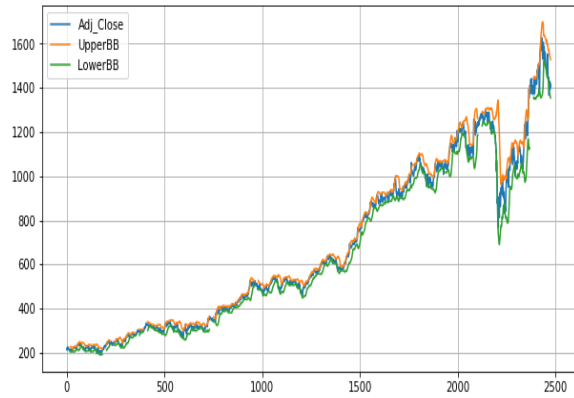


Figure A.18. Bollinger Band for Reliance
April 11, 2011 - April 26, 2021

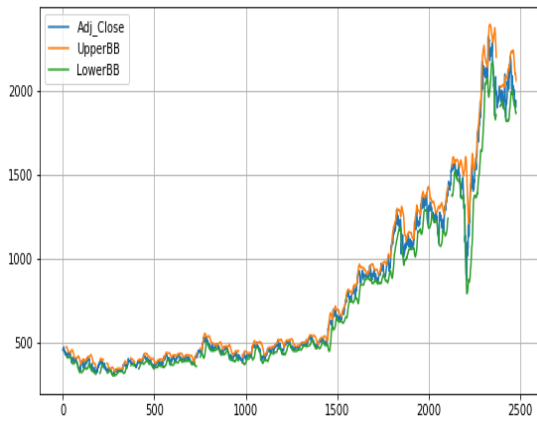


Figure A.19. Bollinger Band for TCS
April 11, 2011 - April 26, 2021

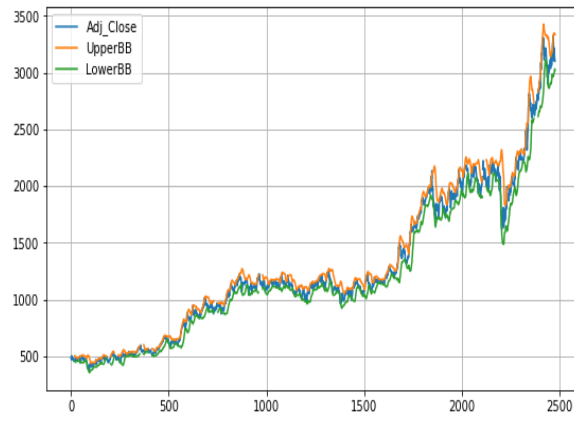


Figure A.20. Bollinger Band for L&T April 11, 2011 - April 26, 2021

