

# Evaluation of Stocks to find Multibaggers using Decision Tree Algorithm in Python

<sup>[1]</sup>Mirza Mohammed Baig, <sup>[2]</sup> Syed Fouzaan Ahmed, <sup>[3]</sup> Syed Muzammil Hussain, <sup>[4]</sup> Dr. Syed Asadullah Hussaini

<sup>[1]</sup>Student, Department of Computer Science and Engineering, ISL Engineering College

<sup>[2]</sup>Student, Department of Computer Science and Engineering, ISL Engineering College

<sup>[3]</sup>Student, Department of Computer Science and Engineering, ISL Engineering College

<sup>[4]</sup>Associate Professor, Department of Computer Science and Engineering, ISL Engineering College

## Article Info

**Page Number:** 1284-1291

**Publication Issue:**

**Vol. 72 No. 1 (2023)**

## Abstract

The idea is to identify a multibagger trade by researching past Multibagger Shares and their characteristics. The decision tree approach is utilized for carrying out statistical operations like data discovery, text extraction, and erroneous information detection in a class. In this article, researchers sought to analyze information on the stock market over the last 20 years. Samples containing basic and technological data are taken into account. The techniques of machine learning and data science methods are used in this investigation. Machine learning is highly efficient in many industries for automating activities that used to require human labor. One such application of ML involves predicting whether or not a certain deal will be profitable. In this study, the stock analysis of Tata Motors, Indian Oil, and Tata Steel are utilized as demonstrations. The findings from this study will assist investors in making faster and better judgments so that they can make money from the investments they make.

**Keywords:** Analysis of Stocks, Decision Tree Algorithm, Multibagger, Fundamental analysis, Yfinance.

## Article History

**Article Received:** 15 October 2022

**Revised:** 24 November 2022

**Accepted:** 18 December 2022

---

## 1. Introduction

A network of securities brokers, financiers, and shareholders who purchase, sell, or transact in shares is known as the stock market. Because numerous companies keep their share lists available to the public, investors find their stocks to be appealing. Investors are experimenting with various methods to learn more about various firms in order to increase the return on their investments. It is crucial to improve the financial standing of emerging nations like India. All are aware that the enormous earnings have kept the share market in the public spotlight for a long time. The stock exchange represents one of the largest and most lucrative financial markets since it sees everyday transactions of enormous capital. At present, among the indicators of a nation's economy is its share market. Many individuals make big investments in the stock market, but occasionally they tend to suffer extremely large losses as they rely on stockbrokers, who provide the shareholders' advice according to fundamental, technical, and time frame series. Shareholders have long been looking for a clever solution to these issues. This is a scenario wherein predicting stocks is useful.

Machine learning approaches may be used in a variety of sectors, including risk analytics, consumer analytics, identifying fraud, and forecasting the stock market. Because of market turbulence, predicting return on investment is extremely challenging. The primary aspect of share market prediction is a high degree of preciseness and accuracy. The emergence of artificial intelligence and excellent computational capacity has enhanced performance. The highly conceptual and hypothetical character of the stock exchange has been investigated during the past several years by collecting and employing repeated patterns. Different machine learning techniques, such as Multiple Linear Regression and Polynomial Regression, are applied in this case. The prototypes are assessed with standard analytical metrics the RMSE & R2 ranking. Smaller readings of these two variables indicate the increased performance of the developed programs. Multiple companies utilize different sorts of statistical techniques for predicting and the major objective is for precision to earn the most profit. [10]

Estimating profits in the shares industry is often framed as a forecasting issue wherein prices are anticipated. Inherent volatility in the global share market renders forecasting difficult. As a result, prediction and diffusion analysis address a wide variety of issues faced in projecting equity market developments. Reducing prediction error would reduce risk associated with investments. [8]

## 2. Literature Survey

Many stockers have been attempting to forecast the value of shares by employing multiple algorithms based on ML that include Support Vector Analysis, Linear Regression, Support Vector Machine, Neural Networks, Genetic Algorithms, and a number of others on shares from multiple businesses. For example, in the study [1] researchers' trained four models: Multi-Layer Perceptron, Recurrent Neural Network, Convolutional Neural Network, and Long Short-Term Memory, where CNN outperformed each of the other three networks. Some authors, on the contrary, feel that Support Vector Regression, which is recognized to tackle regression & forecasting issues, provides higher results, as demonstrated in the article [2]. Within the scope of the study [3] developed three models: Support Vector Machine, CBR algorithm, and Neural Networks (NN), providing the highest accurate forecasting. [4] Developed a method for combining two separate areas for exchange trading analysis. It combines price prediction using historical and current information with news assessment. The informational databases are compiled from big collections of business news that contain important and current data information. The outcomes of both studies are then merged to generate an interpretation that aids in visualizing suggestions for future improvements.

The procedure of choosing stocks for investment decisions in the share market [5] is a massive effort due to the vast number of equities accessible on the stock exchange. Technical evaluation and fundamental analysis are the two most commonly used instruments in this procedure. Fundamental research of the stock under consideration includes an examination of the economic situation, the service sector, and the firm. The inherent /actual worth of a share is then assessed utilizing financial & economic research, which can then be contrasted with present trading prices to figure out if the shares are overpriced or underpriced.

Decision tree classifiers [6] are widely recognized as one of the most well-known approaches for representing classifiers of data. The usage of Decision tree classifiers has been suggested in a variety of domains, including medical illness evaluation, text recognition, user categorization, photographs, and several more. This study takes a deep look at decision trees. Additionally, article details like algorithms/approaches employed, information sets, and results attained are thoroughly examined and discussed.

In [7], researchers examined the classification efficiency of two computational models, Random Forest, & J48, for categorizing 20 diverse datasets. Authors examined the classification statistics derived from Random Forest and Decision Tree (J48) approaches. Correctly categorized occurrences, erroneously classed scenarios, F-Measure, Quality, Precision, and Recall are the classification variables. We addressed the advantages and disadvantages of utilizing these mathematical frameworks for big and small sets of information. The classification findings demonstrate that Random Forest performs well for a comparable range of characteristics and big data sets, i.e. with a larger number of examples, but J48 does well with small data sets (fewer instances).

### 3. Proposed Methodology And Procedure

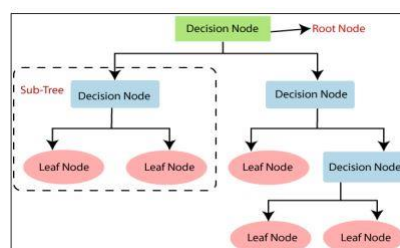
Supervised Machine Learning is employed for both classification and regression analysis. CART-Classification and Regression Trees are two more names for it. In this method, there are two kinds of nodes: the Decision Node, which generates choices and may be split into several subdivisions, and the Leaf Node, which provides decision results and is unable to be split into numerous nodes. The mathematical equation for Leaf Node is as follows:

$$\text{InformationGain} = \text{ClassEntropy} - \text{EntropyAttribute} \quad (1)$$

Branches-Decision constraints are established here to develop separate nodes. Prediction begins at the root node, checks the outcomes of the actual value to that of the root value, and then continues the branching process and leaps to the next node depending on that comparison's outcome. This procedure repeats till it gets to the tree's leaf node. Entropy is a statistic that aids in the measurement of error in a particular characteristic. The formula for calculating entropy is [5]: -

$$\text{Entropy} (I) = -P(\text{yes}) \log_2 P(\text{yes}) - P(\text{no}) \log_2 P(\text{no}) \quad (2)$$

Here (I) denotes the entire number of instances. P (yes) denotes the Probability of S, while P (no) denotes the Probability of no.



**Fig 1: Decision tree classifier process**

### 3.1 Process for Stock Market Prediction

The database used in the experiment sets is obtained from genuine SEBI-authorized domains such as the Money Control website as well as verified sites such as "Yahoo Finance, Screener, Ticker Tape, Moneycontrol.com", and so on. In this study, the developers are exploring with statistics employing Python and R. Researchers utilized libraries written in Python like "numpy, pandas, BeautifulSoup4, Yfinance, and FbProfit" (changed to profit) for fundamental analysis.

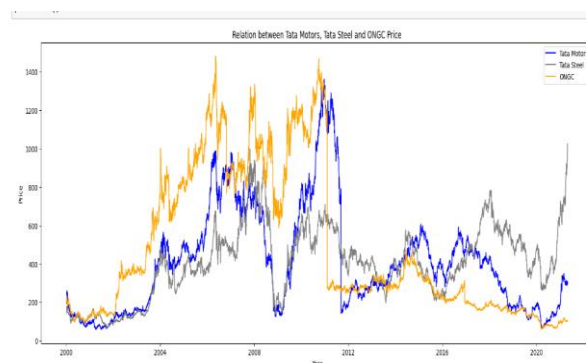
### 3.2 Fundamental Analysis

Fundamental analysis is a way of appraising securities that involves calculating an asset's intrinsic value. Fundamental analysis examines the entire process from the broader business and market circumstances to a company's liquidity and administration.

The goal of this study is to determine an equity price that takes into account all of the key fundamental characteristics. Because the technique does not incorporate short-term price and trade movements, it is regarded as a long-term investing strategy. Because these variables influence progressive demands, this process is used to develop an appraisal based on both historical and progressive data.

## 4. Result Analysis

In this study, we examined the share markets of Tata Motor, Tata Steel, and ONGC. We determined that a firm with excellent positive funding, revenue and earnings, development, and fewer debts, which pays dividends consistently, has a greater long-term gain.

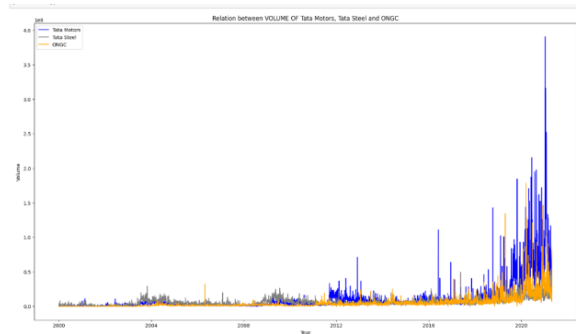


**Fig 2: Price comparison of different companies**

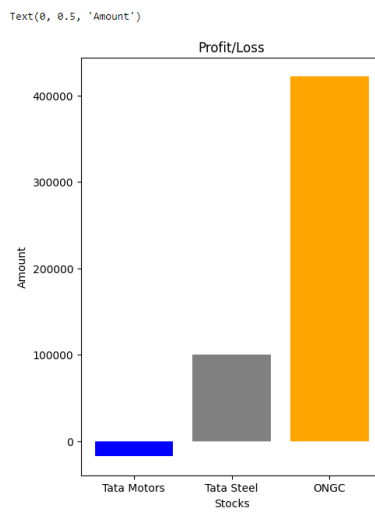
Referring to Figure 2, the market value of ONGC has risen much more than the rates of Tata Steel & Tata Motors. Since its inception, ONGC's revenue curve has moved largely upward, whilst Tata Steel and Tata Motors remain more focused on a consolidation trajectory.

Despite the fact that the market value of ONGC is up considerably more than that of Tata Steel as well as Tata Motors, it can be seen from Figure 3 that ONGC possesses the lowest volume, indicating the fact that the Python analysis of the share market has been dealt in comparison less than that of Tata Steel as well as Tata Motors and is, therefore, less profit.

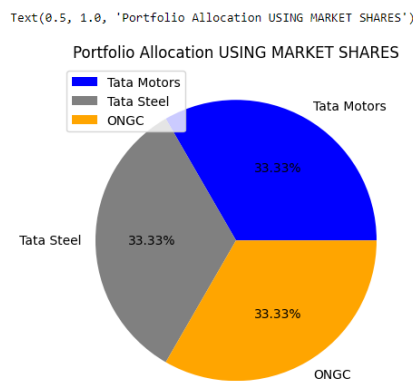
Tata Motors, on the contrary, has been the most traded, indicating increased liquidity and better execution of order.



**Fig 3: Volume Comparison of Different Companies**



**Fig 4: Profit/loss amount**



**Fig 5: Portfolio Allocation**

This research reveals a stock's long-term trajectory and demonstrates SIP's potential for the future.

## 5. Conclusion and Future Works

To anticipate greater profits, companies would require multiple AI ML algorithms. In mathematical terms, it is feasible to decrease risk, enhance revenues, and saves the time required to make decisions and process information. According to our findings, a firm's stock may do extremely well if its revenues are consistently increasing, if it possesses a dominance in the sector or significant market share, or if the business operates in profitable industries (Power, Solar, Electrical, Technological, and Pharmaceuticals).

In the future, numerous upgrades can be explored to improve a program meant to locate multi-bagger shares (stocks having the capacity to provide very large profits). Below are several recommendations:

**Real-time Data and Market Monitoring:** Improve the program so that it can accept feeds of real-time information and continually track market circumstances. Accessibility to accurate prices for shares, the latest news, financial statements, and other important details that may affect the performance of stocks are included.

**Risk Assessment and Portfolio Management:** Incorporate risk evaluation and asset management tools to analyze the risk of possible multi-baggers and help consumers effectively handle their investments. Risk measurements, diversity evaluation, and asset optimization instruments are examples of such tools.

**Back-testing and Performance Evaluation:** Develop a back testing capability that enables consumers to compare the efficacy of specific multi-bagger equities over time. This validates the efficacy of the application's suggestions and gives an understanding of prior performance.

**Integration with Brokerage Platforms:** Investigate possibilities for integrating with prominent brokerage systems to enable efficient trade management depending on the found multi-bagger sector suggestions. This improves consumer comfort while also streamlining the investing procedure.

## 7. References

1. Hiransha, M., et al. "NSE stock market prediction using deep-learning models." *Procedia computer science* 132 (2018): 1351-1362.
2. Yang, Haiqin, Laiwan Chan, and Irwin King. "Support vector machine regression for volatile stock market prediction." *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, Berlin, Heidelberg, 2002.
3. Yoo, Paul D., Maria H. Kim, and Tony Jan. "Machine learning techniques and use of event information for stock market prediction: A survey and evaluation." *International Conference on Computational Intelligence for Modeling, Control and Automation*.
4. Sarode, Sumeet, et al. "Stock price prediction using machine learning techniques." *2019 International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE, 2019.
5. Agrawal, S. A., Umbarkar, A. M., Sherie, N. P., Dharme, A. M., & Dhabliya, D. (2021). Statistical study of mechanical properties for corn fiber with reinforced of polypropylene fiber matrix composite. *Materials Today: Proceedings*, doi:10.1016/j.matpr.2020.12.1072

6. Anupong, W., Azhagumurugan, R., Sahay, K. B., Dhablya, D., Kumar, R., & Vijendra Babu, D. (2022). Towards a high precision in AMI-based smart meters and new technologies in the smart grid. *Sustainable Computing: Informatics and Systems*, 35 doi:10.1016/j.suscom.2022.100690
7. Anupong, W., Yi-Chia, L., Jagdish, M., Kumar, R., Selvam, P. D., Saravanakumar, R., & Dhablya, D. (2022). Hybrid distributed energy sources providing climate security to the agriculture environment and enhancing the yield. *Sustainable Energy Technologies and Assessments*, 52 doi:10.1016/j.seta.2022.102142
8. Aoudni, Y., Donald, C., Farouk, A., Sahay, K. B., Babu, D. V., Tripathi, V., & Dhablya, D. (2022). Cloud security based attack detection using transductive learning integrated with hidden markov model. *Pattern Recognition Letters*, 157, 16-26. doi:10.1016/j.patrec.2022.02.012
9. Dhablya, D. (2021). Delay-tolerant sensor network (DTN) implementation in cloud computing. Paper presented at the *Journal of Physics: Conference Series*, , 1979(1) doi:10.1088/1742-6596/1979/1/012031 Retrieved from [www.scopus.com](http://www.scopus.com)
10. K. Anusha, A. Murugan. "Evaluate Intrinsic Value of Stock Using Decision Tree Algorithm." *Alochana Chakra Journal*, ISSN NO:2231-3990, Volume IX, Issue VI, June/2020.
11. Bahzad Taha Jijo, Adnan Mohsin Abdulazeez. "Classification Based on Decision Tree Algorithm for Machine Learning." *Journal of Applied Science and Technology Trends*, Vol. 02, No. 01, Pp. 20 – 28 (2021).
12. Jehad Ali, Rehanullah Khan, Nasir Ahmad, Imran Maqsood. "Random Forests and Decision Trees, *IJCSI International Journal of Computer Science Issues*, Vol. 9, Issue 5, No 3, September 2012.
13. Suryoday Basak, Saibal Kar, SnehanshuSaha, LuckysonKhaidem. "Predicting the direction of stock market prices using tree-based classifiers." *The North American Journal of Economics and Finance* , DOI:10.1016/j.najef.2018.06.013.
14. Jagruti Hota, Sujata Chakravarty, Bijay K. Paikaray and Harshvardhan Bhojar. "Stock Market Prediction Using Machine Learning Techniques." *ACI'22: Workshop on Advances in Computation Intelligence, its Concepts & Applications at ISIC 2022*, May 17-19, Savannah, United States.
15. Shreya Pawaskar. "Stock Price Prediction using Machine Learning Algorithms." *Ijrasnet Journal For Research in Applied Science and Engineering Technology*
16. Syed Shehriyar Ali, Mohammed Sarfaraz Shaikh, Syed Safi Uddin, Dr. Mohammed Abdul Bari, "Saas Product Comparison and Reviews Using Nlp", *Journal of Engineering Science (JES)*, ISSN NO:0377-9254, Vol 13, Issue 05, MAY/2022
17. Hafsa Fatima, Shayesta Nazneen, Maryam Banu, Dr. Mohammed Abdul Bar," *Tensorflow-Based Automatic Personality Recognition Used in Asynchronous Video Interviews*", *Journal of Engineering Science (JES)*, ISSN NO:0377-9254, Vol 13, Issue 05, MAY/2022
18. MohammedShoeb, Mohammed Akram Ali, Mohammed Shadeel, Dr. Mohammed Abdul Bari, "Self-Driving Car: Using Opencv2 and Machine Learning", *The International*

journal of analytical and experimental modal analysis (IJAEMA), ISSN NO: 0886-9367,  
Volume XIV, Issue V, May/2022

19. Mr. Pathan Ahmed Khan, Dr. M.A Bari,: Impact Of Emergence With Robotics At Educational Institution And Emerging Challenges”, International Journal of Multidisciplinary Engineering in Current Research(IJMEC), ISSN: 2456-4265, Volume 6, Issue 12, December 2021,Page 43-46
20. Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali,” Smartphone Security and Protection Practices”, International Journal of Engineering and Applied Computer Science (IJEACS) ; ISBN: 9798799755577 Volume: 03, Issue: 01, December 2021
21. Shahanawaj Ahamad, Mohammed Abdul Bari, Big Data Processing Model for Smart City Design: A Systematic Review “, VOL 2021: ISSUE 08 IS SN : 0011-9342 ;Design Engineering (Toronto) Elsevier SCI Oct
22. Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali,” Smartphone Security and Protection Practices”, International Journal of Engineering and Applied Computer Science (IJEACS) ; ISBN: 9798799755577 Volume: 03, Issue: 01, December 2021 (International Journal,U K) Pages 1-6
23. Dr.M.A.Bari, ”EffectiveIDS To Mitigate The Packet Dropping Nodes From Manet “, JACE, Vol -6,Issue -6,June 2019
24. M.A.Bari&Shahanawaj Ahamad,” Process of Reverse Engineering of Enterprise InformationSystem Architecture” in International Journal of Computer Science Issues (IJCSI), Vol 8, Issue 5, ISSN: 1694-0814, pp:359-365,Mahebourg ,Republic of Mauritius , September 2011.