

Tata Motors Equity Forecasting System using Machine Learning

R. Kavitha¹, D. Nishanth², V. Sutharsan³, S. Barathbalaji⁴

^{1,2,3,4}Department of Information technology, SRM Institute of Science & Technology, Ramapuram campus, Chennai, India.

E-mail: ¹kavithar8@srmist.edu.in, ²dn6832@srmist.edu.in, ³vs3687@srmist.edu.in, ⁴bs7272@srmist.edu.in

Abstract

The TATA Motors stock prediction system is a machine learning tool that uses regression algorithms to analyze historical data on TATA Motors stock prices, market trends, economic indicators, and industry news. This system aims to help TATA Motors build a business model for the next few months by providing accurate predictions of future stock prices. By mitigating risks associated with investing in the stock market, TATA Motors can make better investment decisions and avoid losses. The system is designed to be user-friendly, with an intuitive interface that makes it easy for TATA Motors employees to input data and access predictions. Furthermore, the system can inform longer-term strategic planning and investment decisions. However, the accuracy of the predictions is reliant on the quality and relevance of the input data. Regular updating and refining of both the input data and the algorithms used to analyze, may be necessary to maintain the system's accuracy.

Keywords: Stock prediction, machine learning, predictive analysis, Regression.

1. Introduction

Using machine learning algorithms for stock prediction offers many advantages. These algorithms can quickly analyze vast amounts of data, allowing investors to make faster, more informed decisions based on the latest market trends and news. Machine learning algorithms can also be used for other purposes, such as identifying potential risks and opportunities in the market, optimizing investment portfolios, and identifying fraudulent activity.

One of the benefits of using machine learning algorithms for stock prediction is that they can identify patterns in the data that might not be visible to the human eye. This can

lead to more accurate predictions and better-informed investment decisions. Machine learning algorithms can also be used to analyze unstructured data, such as news articles and social media posts, which can provide valuable insights into market sentiment and trends. However, it's important to keep in mind that machine learning algorithms are not infallible. They are based on statistical models and are subject to error, particularly when the data being used is noisy or incomplete. Additionally, machine learning algorithms are not a substitute for human judgment and expertise.

2. Problem Statement

The objective of this research is to develop a robust stock price prediction system for Tata Motors, which will help investors, traders, and other stakeholders to make informed decisions regarding buying, selling, or holding Tata Motors' stock. The prediction system should be designed to handle a variety of scenarios, including changes in market conditions, company-specific news, and unexpected events that may impact the stock price of Tata Motors. It should be able to process large volumes of data quickly and efficiently, using advanced machine learning and statistical techniques to generate accurate predictions. The system should leverage historical stock prices, financial data, and other relevant information to predict future stock prices accurately. Additionally, it should provide users with visualizations and analytics that can help them understand the trends and patterns in the data and make better decisions.

To achieve these goals, the system will need to be developed using a combination of domain expertise, data science, and software engineering skills. The ultimate goal is to provide investors with a valuable tool that can help them achieve higher returns on their investment in Tata Motors. The success of this research will depend on the quality of the data, the accuracy of the predictions, and the usability and scalability of the system.

3. System Modules

A. Gradient Boost Regressor Module

The gradient boost regressor module is an essential component of the Tata Motors stock price prediction system. It is a machine learning algorithm that is used to build a predictive model based on historical stock prices and other relevant data. The gradient boost regressor module works by iteratively adding new models to the ensemble, with each model correcting the errors made by the previous models. This process continues until the desired level of

accuracy is achieved, or until a specified number of models have been added. The gradient boost regressor module is particularly useful in situations where there is a large amount of data and a high degree of complexity in the relationships between the input variables and the target variable. It can handle non-linear relationships and interactions between variables, making it a powerful tool for predicting stock prices. To use the gradient boost regressor module in the Tata Motors stock price prediction system, historical data on Tata Motors stock prices and other relevant variables such as financial reports, news articles, and economic indicators are fed into the algorithm. The algorithm then generates a predictive model that can be used to forecast future stock prices. To ensure that the gradient boost regressor module produces accurate predictions, the algorithm must be tuned carefully, and the hyperparameters must be adjusted to optimize performance. Additionally, the model's performance should be regularly evaluated and validated to ensure that it is still accurate and relevant to the current market conditions.

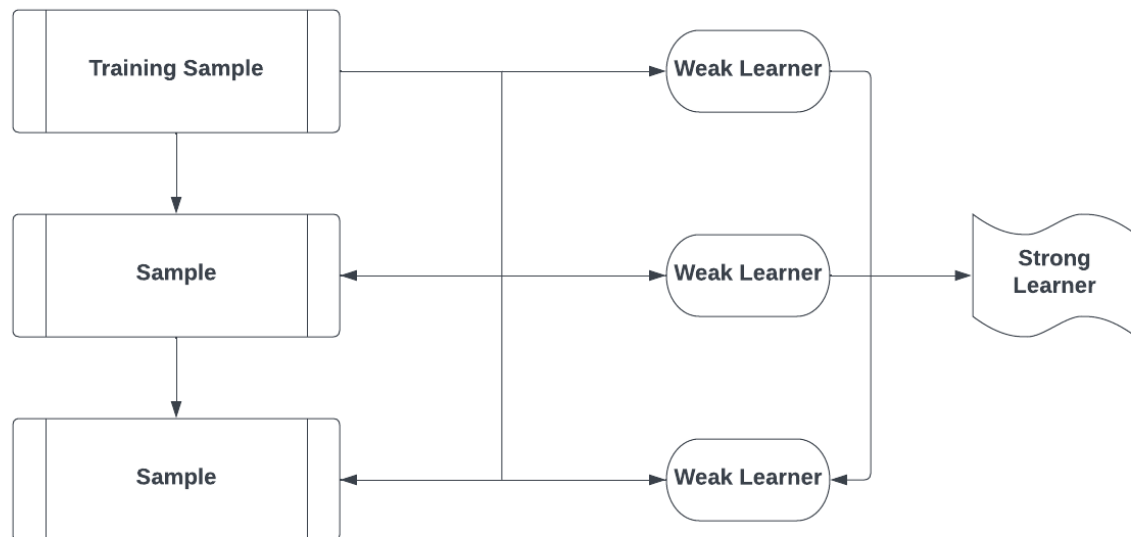


Figure 3.1. Gradient boost regressor module

B. Xg Boost Regressor Module

The XGBoost regressor module is another important machine learning algorithm used in the Tata Motors stock price prediction system. XGBoost stands for Extreme Gradient Boosting, and it is a popular algorithm for building predictive models due to its speed, accuracy, and flexibility. Like the gradient boost regressor module, the XGBoost regressor module uses a boosting technique to iteratively add new models to the ensemble, with each model correcting the errors made by the previous models. However, XGBoost incorporates additional features such as regularization and parallel processing, which help to improve its performance and scalability. To use the XGBoost regressor module in the Tata Motors stock

price prediction system, historical data on Tata Motors stock prices and other relevant variables such as financial reports, news articles, and economic indicators are fed into the algorithm. The algorithm then generates a predictive model that can be used to forecast future stock prices. One of the key advantages of the XGBoost regressor module is its ability to handle missing data and outliers, which can be common in financial datasets. Additionally, it is highly customizable, with many hyperparameters that can be adjusted to optimize performance for a particular dataset. To ensure that the XGBoost regressor module produces accurate predictions, the algorithm must be tuned carefully, and the hyperparameters must be adjusted to optimize performance. Additionally, the model's performance should be regularly evaluated and validated to ensure that it is still accurate and relevant to the current market conditions.

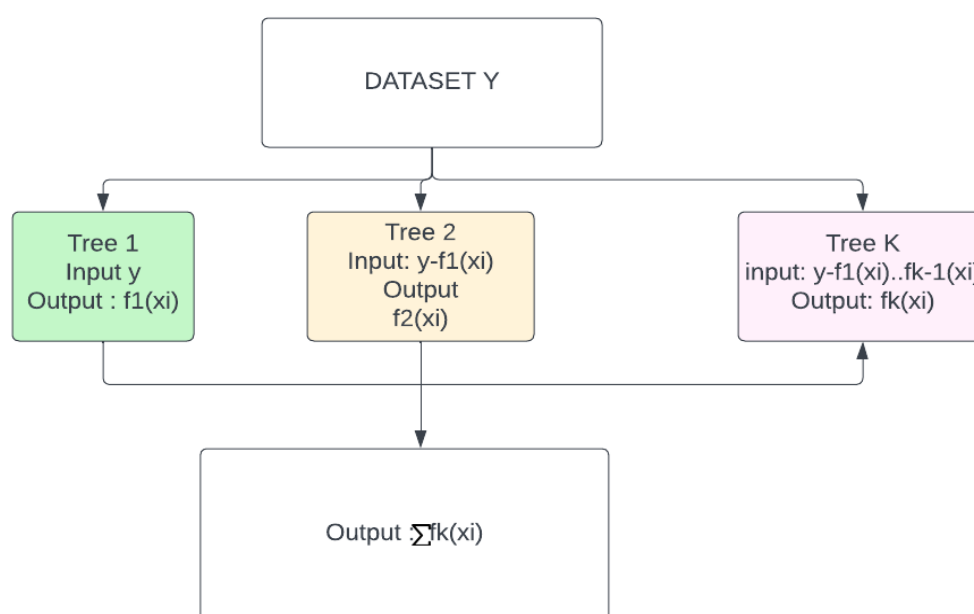


Figure 3.2. XGBoost Regressor Module

C. Catboost Regressor Module

The CatBoost regressor module is another machine learning algorithm used in the Tata Motors stock price prediction system. CatBoost stands for Categorical Boosting, and it is a gradient boosting algorithm that is particularly well-suited for datasets with categorical features. Like the other boosting algorithms, the CatBoost regressor module iteratively adds new models to the ensemble, with each model correcting the errors made by the previous models. However, CatBoost incorporates additional features such as gradient-based feature importance, which helps to identify the most important features for predicting stock prices. To use the CatBoost regressor module in the Tata Motors stock price prediction system, historical

data on Tata Motors stock prices and other relevant variables such as financial reports, news articles, and economic indicators are fed into the algorithm. The algorithm then generates a predictive model that can be used to forecast future stock prices. One of the key advantages of the CatBoost regressor module is its ability to handle categorical features, which are common in financial datasets. Additionally, it has built-in mechanisms for handling missing data and outliers, which can be important for maintaining accuracy in real-world applications. To ensure that the CatBoost regressor module produces accurate predictions, the algorithm must be tuned carefully, and the hyperparameters must be adjusted to optimize performance. Additionally, the model's performance should be regularly evaluated and validated to ensure that it is still accurate and relevant to the current market conditions.

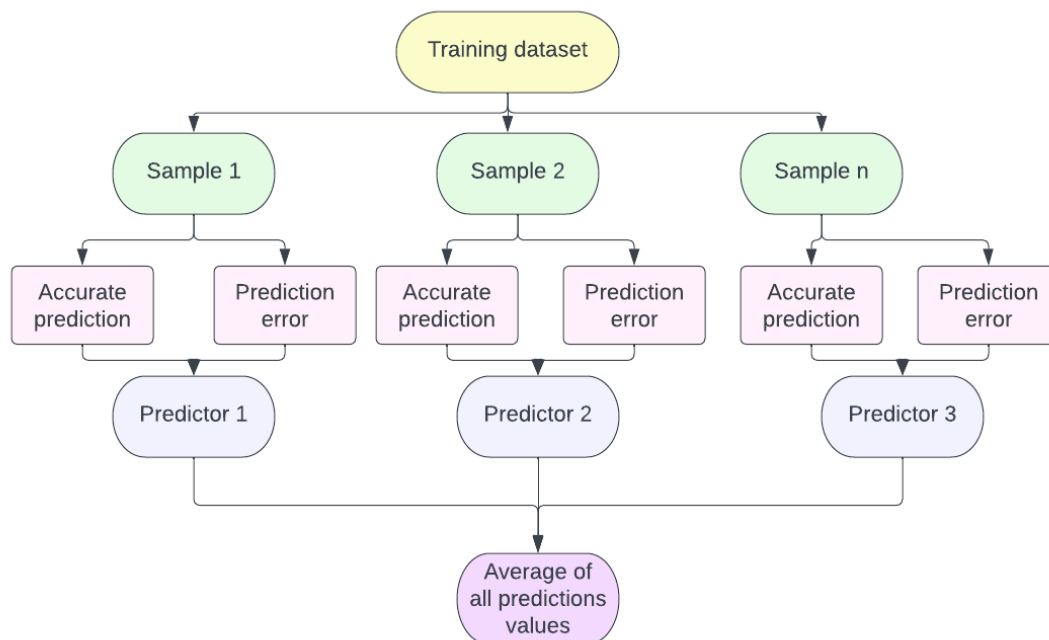


Figure 3.3. Catboost Regressor model

The above diagram shows the flow of the working of Catboost regressor.

4. Results & Discussion

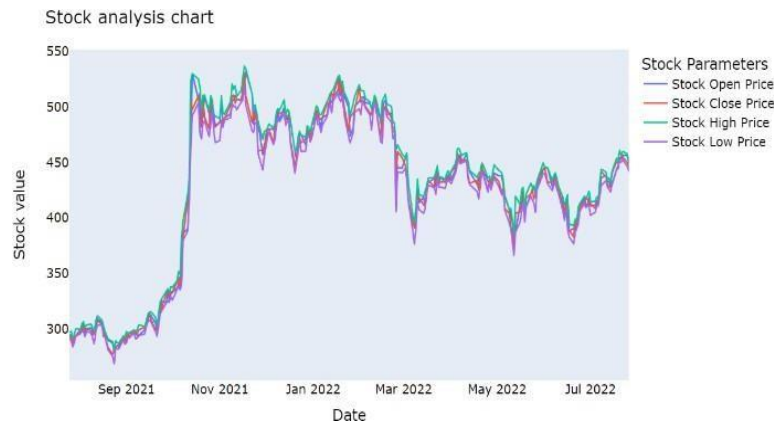


Figure 4.1. Stock analysis chart

The dataset for this research is sourced from Yahoo finance historical dataset sources. The dataset is available for open-source testing purposes. The sample dataset contains 7 columns (Date, Open, High, Low, Close, Adj Close, Volume) and 252 rows of data from the date 2021-07-26 to the date 2022-07-26. The feature extraction is implemented using the Standard Scaler method from the library sci-kit learn. The splitting of the dataset is performed by generalized method of splitting. The chart in Fig.4.1 describes about the analysis of the open price, closing price, high price, and low price of the stocks. Tata motors open and closing price are in a constant move from July 2021 to December 2021. Tata motors open price started from the value 295. The open price started to increase in September 2021 with a game changer value of 299.

The value for the open price was at the peak for the date Nov 18, 2021, at the value of 531.45. At last, the open price ended in the date December 30, 2021, at the value of 472.55. The open price is a major resource for predicting the closing price too at the end. The closing price started at the date July 26, 2021, with the value of 293.15. The closing price also witnessed a drastic increment in the following months. The closing price analysis ended in the date December 29, 2021, with the value of 470.4.

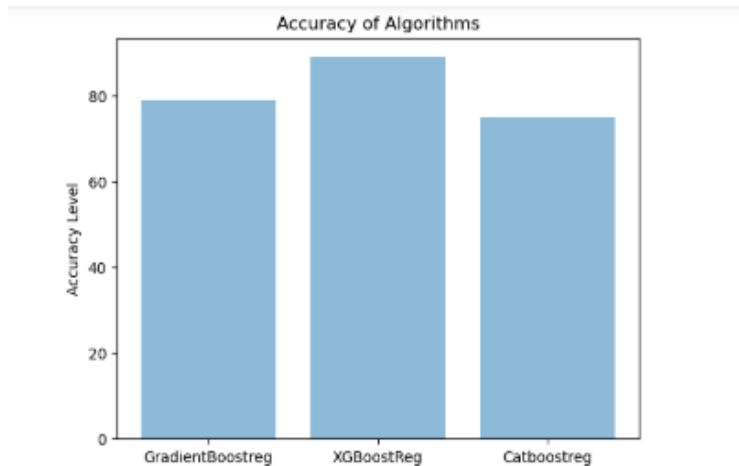


Figure 4.2. Comparison of algorithm's performance

From the analysis, it can clearly be visualized that the algorithms Gradient boost regressor and Catboost regressor attained less accuracy than XGBoost regressor algorithm. This analysis is made up of generating the algorithms' performance values based on the dataset. Xgboost has an accuracy of 89%, Gradient boost regression has an accuracy of 78% and Catboost regression has an accuracy of 75%. As per the analysis, the suitable algorithm is XGboost for the dataset which was used to predict.

```
In [63]: s=xgb.predict(X_test)
          print(s)

[477.99478 448.76797 415.61734 429.6075 478.9719 376.0042 491.64572
 413.43744 429.6075 295.78696 435.0402 428.4127 504.54166 435.0402
 491.82605 509.1776 434.36188 350.1265 294.38354 506.7778 400.93793
 306.79596 498.7045 428.47885 513.9206 508.08786 451.0998 428.6682
 302.09964 434.86005 399.9693 335.98117 485.08572 428.9654 422.72675
 412.91016 298.9652 296.59857 284.89618 415.61734 419.48007 281.61935
 494.2812 458.8627 298.9652 494.01968 310.50546 296.0605 296.59857
 294.38354 412.0457 453.494 489.1089 449.26578 433.91153 505.6819
 503.89563 293.33847 298.9652 502.58304 425.91272 450.54825 448.46988
 434.36188 335.98117 376.0042 306.79596 434.86005 501.28775 488.57782
 476.82425 414.1516 451.0998 428.9654 433.71976 487.0334 ]
```

Figure 4.3. Output – Predicted closing values

The above screenshot provides the predicted data as the predicted closing values for a set of days. These values are generated using extreme gradient boost algorithm. The predicted values contain lots of variations. Hence the users can use the predicted values to sell or buy share in the right time. This gives the user the ease of trading without losing money and provides a stress-free life to the traders. The upgradation for this research will be connecting the frontend and a database to store the predicted values. The user will also be able to select

the period to identify the predicted closing values in the frontend. The frontend will be eye catchy and it provides a powerful ease of access to the users. This includes account creation, sign up and login form to provide the users the data and their complete trade history. This research is mainly designed as a universal tool to predict the stocks for any dataset corresponding to stocks and shares.

5. Conclusion

Grounded on the analysis and prediction performed, it can be concluded that the TATA Motors equity prediction system has the implicit to give precious perceptivity for investors and stakeholders in the automotive industry. The system utilizes a variety of fiscal and request data to predict unborn equity prices and identify implicit investment openings. Likewise, the TATA Motors equity prediction system uses advanced algorithms and machine learning ways to continually upgrade its predictions and adapt to changing request conditions. By using this system, investors and stakeholders can make informed opinions about when to buy, vend, or hold TATA Motors equity.

6. Future Enhancements

The statement highlights the significance of incorporating all applicable variables that could impact stock price movements into a neural network model to achieve more accurate predictions. When creating a predictive model, it's pivotal to consider numerous factors as possible that could affect the stock price, including but not limited to fiscal rates, company performance criteria, macroeconomic pointers, and news sentiment analysis. By doing so, the model can learn and identify the underpinning patterns and connections between different variables, allowing it to make further accurate predictions.

7. References

- [1] Abhishek Bajpai, Anjali Singh-“An Efficient Method to Predict the Tata- Motors Stock Price using Hybrid Machine Learning Methods”, 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN), Al-Khobar, Saudi Arabia, 2022

- [2] Pranita Hosmani, Dr Supriya Shanbhag – “Tata Motors Stock Market Price Prediction using Artificial Neural Networks with MATLAB”, International Journal of Advances in Engineering and Management (IJAEM), Volume 4, Issue 9 Sep. 2022
- [3] Aloysius Edward, JyothiManoj – “Forecast Model Using Arima For Stock Prices Of Automobile Sector”, Volume 6, Issue 4 (April, 2016) (ISSN 2231- 5985), International Journal of Research in Finance and Marketing (IMPACT FACTOR– 5.861)
- [4] Bollen, J., Mao, H., and Zeng, X-“Twitter mood predicts the stock market”, Journal of Computational Science, 2(1),1–8 2010.
- [5] Chan, S. W. K., and Franklin, J-“A text- based decision support system for financial sequence prediction. Decision Support Systems”, 52(1),189-198,2011.
- [6] Chatrath, A., Miao, H., Ramchander, S., Villupuram, S., 2014. Currency jumps, cojumps and the role of macro news. Journal of Money and Finance, 40, 42-62,2014.
- [7] Li, X., Xie, H., Wang, R., Cai, Yi., Cao, J., Wang, F., Min, H., and Deng, X- “Empirical analysis: stock market prediction via extreme learning machine.” Neural Computing and Applications, 27(1), 67- 78,2014.
- [8] Ling, W., and Mei, Fu. D-“Estimation of missing values using a weighted k-Nearest Neighbors algorithm”. International Conference of Environmental Science and Information Application Technology (ESIAT), 2009.
- [9] Arjun Singh Sauda, Subarna Shakya- “Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE”, International Conference on Computational Intelligence and Data Science (ICCIDS 2019).
- [10] Htet Htet Htun, Michael Biehl – “Survey of feature selection and extraction techniques for stock market prediction”, Financ Innov 9, 26 (2023)
- [11] Shruti Mittal, C.K. Nagpal - “Predicting a reliable stock for mid and long term investment”, Journal of King Saud University- Computer and Information Sciences, Volume 34, Issue 10, Part A, November 2022.