International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Predicting the Price of Used Cars using Machine Learning Techniques

Tanmay Jain
Student, Department of Electronics, MITS, Gwalior, India

ABSTRACT

With the growing automobile industry, the market for second hand car has increased exponentially and it becomes important to figure out what should be the value of a certain car with certain features be. As customers and sellers both need to be better informed about the exact price in the market with the data available so that the user doesn't encounters the loss. In this paper, we investigate the applications of Machine Learning techniques to predict the price of car. Using Machine Learning Algorithms such as Linear regression, Decision tree, XG Boost regression, I tried to develop a model which help us to find the more accurate value of the used car based on the features and the data available. In this we will compare the accuracy of these models.

Keywords: Machine Learning, XG Boost regression, Linear regression, Decision tree

I. INTRODUCTION

At present the ratio of second-hand car sold to the new cars in India is 1:1, while in developed countries it is almost 3:1 [1]. When it comes to expensive cars, most of the users buy second hand car rather than buying it, because they can't afford it. For this they need to find the exact price of the car, or to buy or sell the car, both buyers and sellers need to know the exact price of the used cars. Here, I developed a model which help user to find the more accurate price of the car with the given data. In this model I used supervised Machine Learning algorithm.

Machine Learning commonly known as ML is a branch of Artificial Intelligence. It is a study of computer algorithm which focuses on the use of data that can improve automatically through experience and by the use of this data. In this model we used Supervised Machine Learning algorithm because it trains and teach the machine using data that is well labelled, when we talk about Unsupervised Machine Learning Algorithm, data is neither classified nor labelled and act on the information without guidance.

Here, we use linear regression, decision tree and XG Boost regression but XG Boost regression gives us the most accurate price of the car because of its processing speed and its good performance. XG Boost or Extreme Gradient Boosting is an extreme level boosting algorithm which comes under ensemble learning in which a sequence of model is trained where every new model tries to correct the error of its previous models. Errors came out to be very low in XG Boost and R^2 score which tells how good the regression model is came out to be very high, standing at 90% on train data and 87% on test data.

II. METHODOLOGY

Selling price of the second-hand car can be estimated by the features of the car, what type of car and which car are we talking about. In our dataset we have 'n' numbers of car and lot of other features mainly name of the car, selling price, year of release, number of seats in the car, fuel used, kilometres driven, seller type, mileage, engine type, torque, max-power and transmission using these features we crate the model and trained it and depict the price. So in order to create the model we have to create the dataset which included encoding the categorical values which were not in numerical values. To get the prediction of Machine Learning, we need numerical values.
To select the decision variable, we need to check the correlation coefficient between decision variable and target variable and then we decided to train the Machine Learning model by using decision variable. Now to select which Machine Learning algorithm to use, we studied about a lot of Machine Learning algorithm and then decided to move on with XG Boost regression.

STEPS:
1. Understanding the dataset.
2. Pre-processing the data.
3. Analysing target variable and selecting decision variable.
4. Understand which Machine Learning algorithm to used.
5. Selecting XG Boost regression.
6. Training the model.
7. Getting out the prediction.

III. RESULTS

The R^2 score of the model with XG Boost regression algorithm came out to be 0.9 for the training dataset and 0.87 for the test dataset, when we tried to solve the problem using linear regression algorithm, the model was underfitting and when we tried Decision tree to solve the problem, the model was overfitting. It means that XG Boost regression algorithm gives us the more accurate prediction of the car.

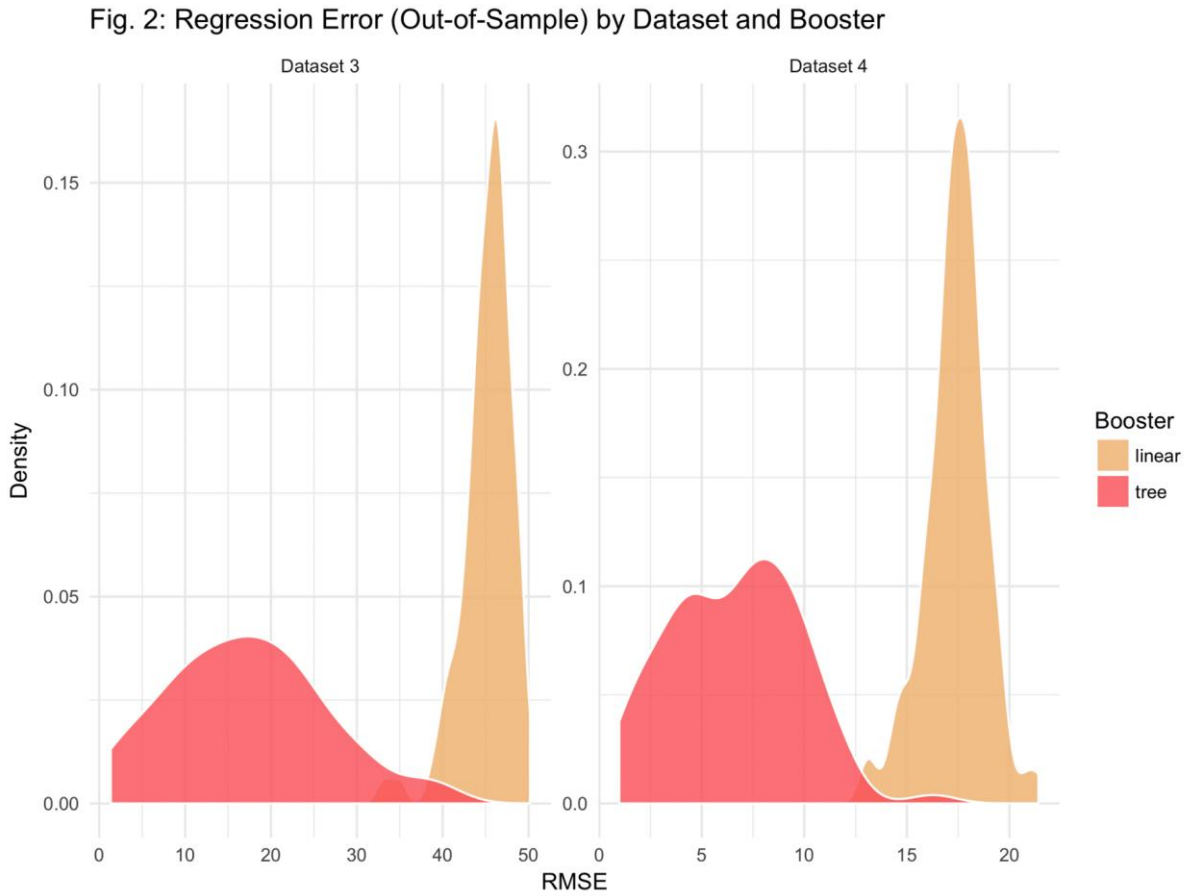

Figure 1. Error comparison of XG Boost tree and linear regression [2]

IV. CONCLUSION

This paper has proposed a model which predict the second-hand car price using data available. It can play a very crucial role in future to depict the price so that no user encounters the loss. This model has further scope of improvement, if we are able to gain more accurate data of second-hand car available in the market and we can also apply Hyperparameter tuning to our model when we have the access to more powerful systems to train the model.

V. REFERENCE

[1] Team-BHP.com. 2022. *The ratio of used cars to new cars sold in India is 1:1 | Team-BHP*. [online] Available at: <https://www.google.com/amp/s/www.team-bhp.com/news/ratio-used-cars-new-cars-sold-india-11%3famp> [Accessed 27 April 2022].

[2] statworx®. 2022. *XGBoost Tree vs. Linear*. [online] Available at: <https://www.statworx.com/en/content-hub/blog/xgboost-tree-vs-linear/> [Accessed 29 April 2022].