INTRODUCTION

The need of transportation is a necessary demand. Everyone in the world rely on transportation to move from one point to another .A brand new car may not be affordable for everyone, so the alternative that everyone seeks is used cars. Cars are not a long term investment, therefore the market of used cars has drastic impact in commercial market. As compared to brand new cars, the price of used cars are much more flexible .Manual prediction of used cars depend on the interest of vendors .In this scenario, the commercial value of price prediction of used cars comes into picture. As the car’s value depend upon various factors like year of manufacture, number of seats, engine power,brand value ,type of fuel,braking system,number of previous owners, odometer range, vehicle type, transmission type etc..the prediction is not an easy factor. Based upon the minimal details available it is important find the feature importance and also predict better accuracy .All the features listed above were not taken in consideration as some of them were irrelevant to the target variable. Five C’s on ethical use of data was taken into consideration which includes consent, clarity, consistency, control and consequences.

The objective of this project is to predict the price of used cars using machine learning models. Since it is a regression problem various regression models were analyzed like Logistic regression,decision tree, k-nearest neighbour, naïve bayes, Support Vector Machine and linear regression. The dataset consists of 6020 rows and 12 columns. Analysis were done on a real time dataset from “AUTOMAX”, which is one among the retail car dealers. The performance of all the various models were analyzed to choose the best out of them which would give a better prediction. The project focused on deep learning techniques using Keras library.

RELATED WORKS

1.Price Prediction Model for 2nd Hand Cars. (2020). Retrieved 24 February 2020, from <https://medium.com/@elysekatrina.go/price-prediction-model-for-2nd-hand-cars-f1801d8c8d47>

Here the main heading of their project is “ We Made a Price Prediction Model for Used Cars in the Philippines “. In this project they were trying to predict the price of used cars by creating an online prototype application just like an online calculator. They train their model with the data that they have collected in Philippines. So in this article they haven’t mentioned much about their dataset, only thing which they said was that they were using 12 features which was different from other 2 –3 features from other online calculators. In their dataset they were having null values in the column “color family”, “fuel type” and “mileage”, in the “color family” and “fueltype” column they filled the null values with the value “ No color ” and “ No fuel type “ and they filled the null values in the “mileage” column by imputing the values using K-Nearest Neighbors Regressor. So in model building they used cross validation technique with 5 folds n they used Decision tree model (73.45%) , XGBoost (79.63%) and Random Forest (80.14%). So they got the highest accuracy for Random forest model.