OLD CARS PRICE PREDICTION

GROUP MEMBERS

SHREYA CHOUDHARY - 9

MANSI DEVRUKHKAR - 11

ABSTRACT :

Predicting the price of used cars in both an important and interesting problem. For this research, we conducted a comparative study on performance of regression based on supervised machine learning models. Each model is trained using data of used car market collected from [www.kaggle.com](http://www.kaggle.com) used cars price prediction dataset.csv.

We are going to use multiple linear regression, random forest regression.

LITERATURE SURVEY :

Several related works have been done previously on the subject of used car price prediction. Pudaruth Predicted the price of used cars in Mauritius using multiple linear regression, k-nearest neighbors, naive Bayes and decision trees. Although their results was not good for prediction due to a less number of car observation. Pudaruth concluded in his paper that the decision tree and naive Bayes are unable to use for variable with a continuous value. Noor and Jan used multiple linear regression to predict vehicle car price. They performed variable selection technique to find the most influencing variables then eliminate the rest. The data contain only selected variable that used to form the linear regression model.

Peerun et al. did a research to evaluate the performance of the neural network in used car price prediction. The predicted value, however, are not very close to the actual price, especially on cars with a higher price. They concluded that support vector machine regression

slightly outperform neural network and linear regression in predicting used car price. Sun et al. proposed the application of online used car price evaluation model using the optimized BP neural network algorithm. They introduced a new optimization method called Like Block-Monte Carlo Method (LB-MCM) to optimize hidden neurons. The result shown that the optimized model yielded higher accuracy when it compared to the non-optimized model Based on the previous related works, we realized that none of them had implemented random forest technique in the prediction of used car price yet. Thus, we decided to build a used car price evaluation model using random foresr regression trees.

ALGORITHMS AND TECHNIQUES :

1.Linear Regression : In statistics, **linear regression** is a **linear** approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple **linear regression**.

Here we are going to multiple Linear Regression model.

**Multiple linear regression** (MLR), also known simply as **multiple regression**, is a statistical technique that uses **several** explanatory variables to predict the outcome of a response variable. **Multiple regression** is an extension of **linear** (OLS) **regression** that uses just one explanatory variable.

2. Random forest : Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees.