

CAR RESALE VALUE PREDICTION

Name of the Team leader: PRANJAI V

Team ID: PNT2022TMID00833

Roll number: 2019PECCS347

Registration Number: 211419104195

Name of the Team Member 1: PRABHU SAI.D

Roll number: 2019PECCS344

Registration Number: 211419104192

Mobile Number: 9361283615

Mail ID: prabhusai660@gmail.com

Name of the Team Member 2: PRAKASH . R

Roll number: 2019PECCS345

Registration Number: 211419104193

Mobile Number: 9789801834

Mail ID: prakashstoneyn49@gmail.com

Name of the Team Member 3: PRANJIVAN V

Roll number: 2019PECCS348

Registration Number: 211419104196

Mobile Number: 8056249909

Mail ID: pranjivanvelmurugan24@gmail.com

Name of the Team Member 4: NAVEEN KUMAR

Roll number: 2019PECCS338

Registration Number: 211419104180

Mobile Number: 9499979041

Mail ID: naveenmema06@gmail.com

ABSTRACT

Predicting the price of used cars is one of the significant and interesting areas of analysis. As an increased demand in the second-hand car market, the business for both buyers and sellers has increased. For reliable and accurate prediction it requires expert knowledge about the field because of the price of the cars dependent on many important factors. This paper proposed a supervised machine learning model using KNN (K Nearest Neighbor) regression algorithm to analyze the price of used cars . Through this experiment, the data was examined with different trained and test ratios. As a result, the accuracy of the proposed model is around 85% and is fitted as the optimized model. The predictions are then evaluated and compared in order to find those which provide the best performances. A seemingly easy problem turned out to be indeed very difficult to resolve with high accuracy. All the four methods provided comparable performance. In the future, we intend to use more sophisticated algorithms to make the predictions . Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle's price on the market. The focus of this project is developing machine learning models that can accurately predict the price of a used car based on its features, in order to make informed purchases. We implement and evaluate various learning methods on a dataset consisting of the sale prices of different makes and models across cities in the United States. Our results show that Random Forest model and K-Means clustering with linear regression yield the best results, but are compute heavy. Conventional linear regression also yielded satisfactory results, with the advantage of a significantly lower training time in comparison to the aforementioned methods.

LITERATURE SURVEY

Several studies and related works have been done in previously predict used car prices around the world using different methodologies and approaches, with varying results of accuracy from 50% to 90%. In (Pudaruth, 2014) the researcher proposed to predict used car prices in Mauritius, where he applied different machine learning techniques to achieve his results like decision tree, K-nearest neighbours, Multiple Regression and Naïve Bayes algorithms to predict the used cars prices, based on historical data gathered from the newspaper. Achieved results ranged from accuracy of 60-70 percent, the author suggested using more sophisticated models and algorithms to make the evaluation, with the main weakness off the decision tree and naïve Bayes that it is required to discretize the price and classify it which accure to more inaccuracies. Moreover, he suggested a larger set of data of data to train the models hence the data gathered was not sufficient.

Sameerchand Pudaruth[1] proposed predicting the Price of Used Cars using Machine Learning Techniques. In this paper, they collected the historical data of used cars in Mauritius from the newspapers and applied different machine learning techniques like decision tree, K-nearest neighbours, Multiple Linear Regression and Naïve Bayes algorithms to predict the price. This model has the mean error about Rs.27000 for Nissan cars and about Rs45000 for Toyota cars using KNN and around Rs51000 using linear regression. The accuracy of decision trees and NaïveBayes algorithm dangled between 60 to 70 percentile with different parameters and the overall training accuracy of the model is 61%.

Nitis Monburinon et al. [2] proposed a prediction of Prices for Used Car by Using Regression Models. In this paper, the authors selected the data from the German ecommerce site. The main goal of this work is to find a suitable predictive model to predict the used cars price. They used different machine learning techniques for comparison and used the mean absolute error(MAE) as the metric. They proposed that their model with gradient boosted regression has a lower error with MAE value 0.28 and this gives the higher performance where linear regression has the MAE value 0.55, random forest with MAE value 0.35.

Enis Gegic et al. [3] proposed Car Price Prediction using Machine Learning Techniques. In this paper, they proposed an ensemble model by collecting different types of machine learning techniques like Support Vector Machine, Random Forest and Artificial neural network. They collected the data from the web portal www.autopijaca.ba and build this model to predict the price of used cars in Herzegovina and Bosnia. The accuracy of their model is 87%.

Kanwal Noor and Sadaqat Jan[4] proposed Vehicle Price Prediction System using Machine Learning Techniques. In this paper, they proposed a model to predict the price of the cars through multiple linear regression method. They selected the most influencing feature and removed the rest by performing feature selection technique. The Proposed model achieved the prediction precision of about 98%.

A machine learning model is proposed to estimate the cost of the used cars using the K-Nearest Neighbor algorithm. The model is trained with used cars data for different trained and test ratios. Then the proposed model is cross-validated using K fold method to examine the performance to avoid the over fit.

Method :

The topic such as this can be assessed with mathematical models derived from quantitative data. A multiple variable regression can analyze the data by assessing the role each independent variable plays in determining the dependent variable (in this case, resale value). Significance can also be assessed by observing the p-values for each variable. The use of a statistical model will aide in making a claim on this, and to identify some of the major contributors to resale value in automobiles.

Data Collection :

The data used for this regression will be quantitative in nature. The sources of data are what someone would expect for used car information. Four sources that are used include Kelly Blue Book, Edmunds, a government fuel economy resource, and Car and Driver. Kelly Blue Book and Edmunds will both serve as data sources, with each source providing different aspects of the independent variables used. With the cooperation of these sources, data regarding price of a car-including new and used-with the respective age, mileage, make, condition, miles per gallon, safety ratings, and hybrid technology information will be obtained. These variables will allow for a regression to be run and an equation to be estimated.

Expected Outcomes :

Before I can make predictions regarding the influence each variable will have on resale value, a review of prior research and literature is appropriate. This will allow me to make a more confident prediction as well as confirm which variables are needed to produce a strong equation that explains much of the variations in vehicle depreciation. An expected equation could look like this:

Resale Value (DV) = Intercept- B3(Age) - B4(Mileage) + B1(Make) + B2(MPG) + B5(Hybrid Tech)

REFERENCES

- [1] Pudaruth, Sameerchand. "Predicting the price of used cars using machine learning techniques." *Int. J. Inf. Comput. Technol* 4, no. 7 (2014): 753-764.
- [2] Monburinon, Nitis, Prajak Chertchom, Thongchai Kaewkiriya, Suwat Rungpheung, Sabir Buya, and Pitchayakit Boonpou. "Prediction of prices for used car by using regression models." In *2018 5th International Conference on Business and Industrial Research (ICBIR)*, pp. 115-119. IEEE, 2018.
- [3] Gegic, Enis, Becir Isakovic, Dino Keco, Zerina Masetic, and Jasmin Kevric. "Car price prediction using machine learning techniques." *TEM Journal* 8, no. 1 (2019): 113.
- [4] Noor, Kanwal, and Sadaqat Jan. "Vehicle price prediction system using machine learning techniques." *International Journal of Computer Applications* 167, no. 9 (2017): 27-31.
- [5] <https://ieeexplore.ieee.org/Xplore/home.jsp>
- [6] <https://www.analyticsvidhya.com/blog/2018/08/k-nearestneighbor-introduction-regression-python/>
- [7] <https://machinelearningmastery.com/k-fold-cross-validation/>