

Applied Data Science

Car Resale Value Prediction

IBM Project – Team ID: PNT2022TMID12178

Team Lead

Thaniyeal P -73151921056

Team Members

Thavanish P.S -73151921057

Natheem J -73151921032

Rajkumar S -73151921041

Of

BACHELOR OF TECHNOLOGY

In

INFORMATION TECHNOLOGY

K.S.R COLLEGE OF ENGINEERING

ANNA UNIVERSITY: CHENNAI 600025

ABSTRACT

This paper aims to build a model to predict used second hand cars reasonable price; based on multiple aspects, including vehicle mileage, year of manufacturing, fuel consumption, transmission, road tax, fuel type, and engine size. This model can benefit sellers, buyers, and car manufacturers in the used cars market. Upon completion, it can output a relatively accurate price prediction based on the information that user's input. The model building process involves machine learning and data science. The dataset used was scraped from listings of used cars. Various regression methods, including linear regression, polynomial regression, support vector regression, decision tree regression, and random forest regression, were applied in the research to achieve the highest accuracy. Before the actual start of model-building, this project visualized the data to understand the dataset better. The dataset was divided and modified to fit the regression, thus ensure the performance of the regression. To evaluate the performance of each regression, R-square was calculated. Among all regressions in this project, random forest achieved the highest R-square of 0.90416. Compared to previous research, the resulting model includes more aspects of used cars while also having a higher prediction accuracy.

LITERATURE REVIEW

SURVEY-1

Stefan Lessmann (2017)

"Car resale value prediction using regression method"

This paper study statistical models for forecasting the resale prices of used cars. An empirical study is performed to explore the contributions of different degrees of freedom in the modelling process to the forecast accuracy. First, a comparative analysis of alternative prediction methods provides evidence that random forest regression is particularly effective for resale price forecasting. Second, the empirical results demonstrate the presence of heterogeneity in resale price forecasting and identify methods that can automatically overcome its detrimental effect on the forecast accuracy. Finally, the study confirms that the sellers of used cars possess informational advantages over market research agencies, which enable them to forecast resale prices more accurately. This implies that sellers have an incentive to invest in in-house forecasting solutions, instead of basing their pricing decisions on externally generated residual value estimates.

SURVEY-2

Doan Van Thai (2019)

"Car resale value prediction using quantify qualitative data and knowledge-based system"

Car pricing using machine learning has a strong relationship with the process of knowledge acquisition for expert systems. This model, the primary technique for knowledge acquisition has been the time-consuming process of recommendation, posting for car buying or selling on internet market websites. After discovering the data, It is divided into two types: structured and unstructured that require knowledge-based analysis. This paper will involve the techniques for extraction of meaning, data inference, and rules for qualitative data. The main purpose of the current research is to explore different data types of car data and the objective is to create an automated technique to predict car prices.

SURVEY-3

Feng Wang(2021)

"Car resale value prediction based on supervised learning techniques"

In this approach feng wang used machine learning algorithms to predict the price of used cars with less human intervention to make the results more objective. This method is used to preprocess the dataset through Python's Pycaret package and compare the performance of each algorithm through the algorithm comparison function. Finally, the algorithm was optimized by using the hyperparameter function. The results show that $R^2 = 0.9807$ obtained from extreme random numbers is the best performance. When new used car data flows into the used car system, used car prices will be automatically generated by this algorithm, which will make the workflow of the used car market faster and more competitive for that used car market.

SURVEY-4

Jaideep A Miley (2017)

"Prediction of used cars prices by using SAS EM"

The aim of jai deep is to analyse the market trend of used car industry and find out what are the factors that are important decide the price of a used car and finally predict the price of a used car. With the help of SAS Enterprise mine he has used statistical methods such as Transformations, Decision Trees, and Regression to identify the target variable.

SURVEY-5

Sameerchand Pudaruth (2014)

"Car resale value prediction using machine learning"

Sameerchand pudaruth is used supervised machine learning techniques to predict the price of used cars in Mauritius. The predictions are based on historical data collected from daily newspapers. Different techniques like multiple linear regression analysis, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions. 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.

SURVEY-6

Praful Rane, Deep Pandya, Dhawal Kotak (2021)

“USED CAR PRICE PREDICTION”

Praful rane and there team members are used regression algorithms because they provided them with continuous value as an output and not a categorized value. Because of which it will be possible to predict the actual price a car rather than the price range of a car. User Interface has also been developed which acquires input from any user and displays the Price of a car according to user's inputs.

REFERENCES

1. Stefan Lessmann (2017) "Car resale value prediction using regression method".
2. Doan Van thai(2019) "Car resale value prediction using quantify qualitative data and knowledge-based system".
3. Feng Wang (2021) "Car resale value prediction based on supervised learning techniques".
4. Jaideep A Miley (2017) "Prediction of used cars prices by using SAS EM".
5. Sameerchand Pudaruth(2014) "Car resale value prediction using machine learning".
6. Praful Rane, Deep Pandya, Dhawal Kotak(2021)"Used Car Price Prediction".