CAR RESALE VALUE PREDICTION USING DATA SCIENCE

LITERATURE SURVEY

TITLE : Used Cars Price Prediction using Supervised
Learning Techniques

AUTHOR : Pattabiraman Venkatasubbu, Mukkesh Ganesh

YEAR : 2019

ABSTRACT :

Overfitting and underfitting come into picture when we create our statistical models. The models might be too biased to the training data and might not perform well on the test data set. This is called overfitting. Likewise, the models might not take into consideration all the variance present in the population and perform poorly on a test data set. This is called underfitting. A perfect balance needs to be achieved between these two, which leads to the concept of Bias-Variance tradeoff. Pierre Geurts has introduced and explained how bias-variance tradeoff is achieved in both regression and classification. The selection of variables/attribute plays a vital role in influencing both the bias and variance of the statistical model. Robert Tibshirani proposed a new method called Lasso, which minimizes the residual sum of squares. This returns a subset of attributes which need to be included in multiple regression to get the minimal error rate. Similarly, decision trees suffer from overfitting if they are not pruned/shrunk. Trevor Hastie and Daryl Pregibon have explained the concept of pruning in their research paper. Moreover, hypothesis testing using ANOVA is needed to verify whether the different groups of errors really differ from each other. This is explained by TK Kim and Tae Kyun in their paper. A Post-Hoc test needs to be performed along with ANOVA if the number of groups exceeds two.

TITLE : Price Prediction of Used Cars Using Machine Learning

AUTHOR : Chuyang Jin

YEAR : 2021

ABSTRACT :

This paper aims to build a model to predict used cars' reasonable prices 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 users 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.

TITLE : Used Cars Price Prediction and Valuation using Data Mining Techniques

AUTHOR : Abdulla AlShared

YEAR : 2021

ABSTRACT :

Due to the unprecedented number of cars being purchased and sold, used car price prediction is a topic of high interest. Because of the affordability of used cars in developing countries, people tend more purchase used cars. A primary objective of this project is to estimate used car prices by using attributes that are highly correlated with a label (Price). To accomplish this, data mining technology has been employed. Null, redundant, and missing values were removed from the dataset during pre-processing. In this supervised learning study, three regressors (Random Forest Regressor, Linear Regression, and Bagging Regressor) have been trained, tested, and compared against a benchmark dataset. Among all the experiments, the Random Forest Regressor had the highest score at 95%, followed by 0.025 MSE, 0.0008 MAE, and 0.0378 RMSE respectively. In addition to Random Forest Regression, Bagging Regression performed well with an 88% score, followed by Linear Regression having an 85% mark. A train-test split of 80/20 with 40 random states was used in all experiments. The researchers of this project anticipate that in the near future, the most sophisticated algorithm is used for making predictions, and then the model will be integrated into a mobile app or web page for the general public to use.

TITLE : Used car prediction using Machine Learning with optimal features

AUTHORS : Muhammad Asghar, Khalid Mehmood, Samina Yasin, Zimal Mehboob Khan

YEAR : 2021

ABSTRACT :

We all are needed the personal vehicle that could help us to travel from home to office and travel to vocations means we need the personal vehicle for traveling for this we purchase the new vehicle or used vehicle this is some time take so much to take decision for purchasing the new one and most difficult decision is to take how to sale the old one that is already we have keep using if we sale and what is best price we can get or gives us more benefits. More over the purchasing power of the customers is low due to the prices of the new cars. There are different methods to predict the price of the car according to market value. Our proposed method helps the both the purchase and seller for to purchase and sale their vehicle and they can predict the best for their vehicle and make their decision good for personal and business. Our proposed model performance shows that the proposed study is productive and efficient. In the proposed study the machine learning algorithm Regression helps in the outperform. Here we use the Statistical test to get the design value of P and get the optimal features and using the linear regression. First, we find the RFE and then apply the statistical test for VIF for the OLS Regression. Prediction results shows the study is efficient and effective.

TITLE : Car's Selling Price Prediction using Random Forest Machine Learning Algorithm

AUTHOR : Abhishek Pandey, Vanshika Rastogi, Sanika Singh

YEAR : 2020

ABSTRACT :

India has one of the biggest automobile markets all over the globe every day many buyers usually sell their cars after using for the time to another buyer, we call them as 2nd /3rd owner etc. Many platforms such as cars24.com, cardekho.com and OLX.com provides these buyers with a platform where they can sell their used cars, but what should be the price of the car, this is the toughest question ever. Machine Learning algorithms can bring a solution to this problem. Using a history of previously used cars selling data and using machine learning techniques such as Supervised Learning can predict a fair price of the car, here I also used machine learning algorithms such as Random Forest and Extra Tree Regression along with powerful python library Scikit-Learn to predict the selling price of the used car. The result has shown that these both algorithms are highly accurate in prediction even the dataset is large or small, irrespective of the size of the dataset they give a precise result.