LITERATURE SURVEY

|  |  |
| --- | --- |
| Date | 19 September 2022 |
| Team Id | PNT2022TMID30244 |
| Project Name | Car Resale Value Prediction |

**1.USED CAR PRICE PREDICTION Praful Rane1, Deep Pandya2, Dhawal Kotak3[4 APR 2022]**

The first paper is Predicting the price of Used Car Using Machine Learning Techniques.In this paper, they investigate the application of 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.

**2.Used Cars Price Prediction using Supervised Learning Techniques[Pattabiraman Venkatasubbu, Mukkesh Ganesh ,Dec 2019]**

Overfitting and underfitting come into picture when wecreate our statistical models. The models might be too biasedto the training data and might not perform well on the test dataset. This is called overfitting. Likewise, the models might nottake into consideration all the variance present in thepopulation and perform poorly on a test data set. This is calledunderfitting. A perfect balance needs to be achieved between these two, which leads to the concept of Bias-Variancetradeoff. Pierre Geurts [2] has introduced and explained howbias-variance tradeoff is achieved in both regression andclassification. The selection of variables/attribute plays a vital role in influencing both the bias and variance of the statistical model. Robert Tibshirani [3] 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 [4] 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 [5]. A Post-Hoc test needs to be performed along with ANOVA if the number of groups exceeds two

**3. Used Cars Price Prediction and Valuation using Data Mining Techniques[Abdulla AlShared,12-2021]**

Today, the transportation industry is considered to be one of the backbones of the economy. Automobiles are referred to as the "Industry of Industries" in developed nations. According to industry professionals, the UAE's automotive industry has seen remarkable growth. Besides being the fastest-growing nation in the automobile industry, it represents its global presence. In Dubai, like most other countries, cars are gaining a great deal of popularity among the local population and the ex-pat community who work in the country. There are used cars for sale in the UAE of all makes and models, even cars from well-known brands (Rizvi, 2019). UAE's auto industry is experiencing constant growth, registered at 27%, with a total industry volume (TIV) of 310,403 cars. Approximately 1.49 million units were sold within the Gulf Cooperation Council (GCC). Compared to the global market, the Gulf Cooperation Council countries are growing at 10% in 2021 (Research, 2020). So far, the market in the UAE has grown by 19%. It is thus the world's largest market in terms of growth rate.

**4.Enis Gegic et al 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%.

**5.Nitis Monburinon et al 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.

**PROJECT DESCRIPTION**

Car resale value prediction system is made with the purpose of predicting the correct valuation of used cars that helps users to sell the car remotely with perfect valuation and without human intervention in the process to eliminate biased valuation.

