Project Description:-

The following project aims to predict mileage per gallon(mps) using various technical specifications (features) as input to the regression algorithms.

Abstract:-

In this paper, we 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. 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.

Introduction:-

In many developed countries, it is common to lease a car rather than buying it outright. A lease is a binding contract between a buyer and a seller (or a third party – usually a bank, insurance firm or other financial institutions) in which the buyer must pay fixed instalments for a pre-defined number of months/years to the seller/financer. After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e. its expected resale value.

Database Description:

The data is technical spec of cars. This dataset is a slightly modified version of the dataset provided in the StatLib library. In line with the use by Ross Quinlan (1993) in predicting the attribute "mpg", 8 of the original instances were removed because they had unknown values for the "mpg" attribute. The original dataset is available in the file "auto-mpg.data-original".

The data concerns city-cycle fuel consumption in miles per gallon, to be predicted in terms of 3 multivalued discrete and 5 continuous attributes." (Quinlan, 1993)

Number of Instances: 398

Number of Attributes: 9 including the class attribute

Attribute Information:-

mpg: continuous cylinders: multi-valued discrete displacement: continuous horsepower: continuous weight: continuous acceleration: continuous model year: multi-valued discrete origin: multi-valued discrete car name: string (unique for each instance) Missing Attribute Values: horsepower has 6 missing values.

Machine Learning Steps Involved:-

1. Linear Regression:-

The performance of the linear regression is not much good compared to random forest. The difference between actual values and predicted values is worthy. That’s why, we need different models that may give better predictions results.

Chart, bar chart

Description automatically generated

1. **Ridge Regression**:-

Ordinary least square (OLS) gives unbiased regression coefficients (maximum likelihood estimates “as observed in the data-set”). Ridge regression and lasso allow to regularize (“shrink”) coefficients.

Chart

Description automatically generated

**Pre-processing the Data Flowchart:-**

* **Label Encoding**
* **Train the data**
* **Scaling the Data**

Diagram

Description automatically generated

Libraries Involved:-

1. pandas

2. Numpy

3. Seaborn

4. Matplotlib

Steps Involved:

1. Importing the libraries

2. Loading the dataset

3. Data Preprocessing

4. Study Correlation

5. Univariate Analysis

6. Bivariate Analysis

7. train and test data split

8. Building the model

**Conclusion:-**

By performing different models, it was aimed to get different perspectives and eventually compared their performance. With this study, it purpose was to predict prices of used cars by using a dataset that has 13 predictors and 380962 observations. With the help of the data visualizations and exploratory data analysis, the dataset was uncovered and features were explored deeply.

References:-

* IBM Knowledge Center. (n.d). Use of KNN. Retrieved from: <https://www.ibm.com/support/knowledgecenter/SSHRBY/com.ibm.swg.im.dashdb.analytics.doc/doc/r_knn_usage.html>
* Gareth, J., Daniela, W., Trevor, H., & Tibshirani, R. (2013). *An Introduction to StatisticalLearning*(Vol. 8). <https://doi.org/10.1016/j.peva.2007.06.006>
* Hurwitz, E., & Marwala, T. (2012). Common mistakes when applying computational intelligence and machine learning to stock market modelling. *arXiv preprint arXiv:1208.4429*.