CAR RESALE VALUE PREDICTION

<by:Bendram Sravanthi>

CONTENTS:

INTRODUCTION:

1.1 OVERVIEW

With difficult economic conditions, it is likely that sales of second-hand imported cars and used cars will increase. In many developed countries, it is common to lease a car rather than buying it outright. 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. Thus, it is of commercial interest to sellers/financers to be able to predict the residual value of cars with accuracy.

1.2 PURPOSE

Deciding whether a used car is worth the posted price when you see listings online can be difficult. Several factors, including mileage,power, fuel type,year, etc. can influence the actual worth of a car. Considering the main factors which would affect the resale value of a vehicle a regression model is to be built that would give the nearest resale value of the vehicle. We will be using various regression algorithms and algorithm with the best accuracy will be taken as a solution, then it will be integrated to the web-based application where the user is notified with the status of his product.

LITERATURE SURVEY:

2.1 EXISTING PROBLEM

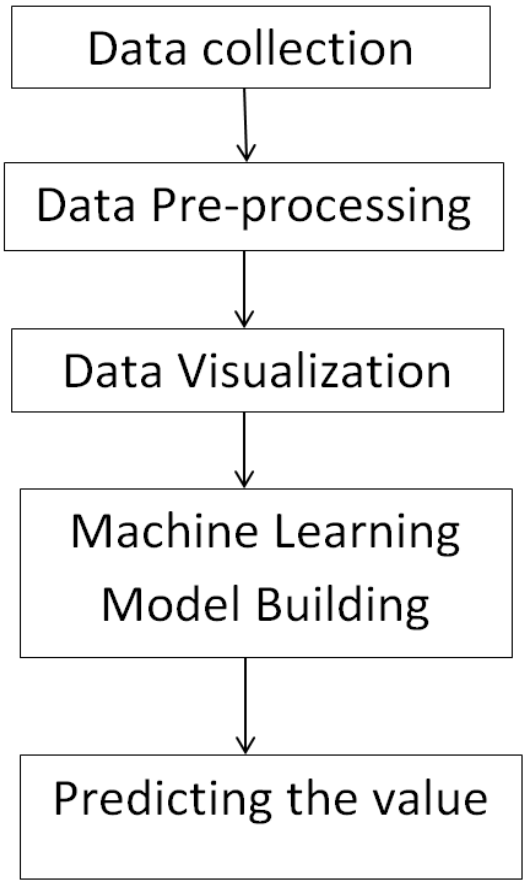
Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle's price on the market. The focus of this project is developing machine learning models that can accurately predict the price of a used car based on its features, in order to make 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 India.

2.2 PROPOSED SOLUTION

In order to predict the resale value of the car, we proposed an intelligent, flexible, and effective system that is based on using regression algorithms.Conventional linear regression yields satisfactory results, with the advantage of a significantly lower training time in comparison to the aforementioned methods.

THEORITICAL ANALYSIS:

3.1 BLOCK DIAGRAM



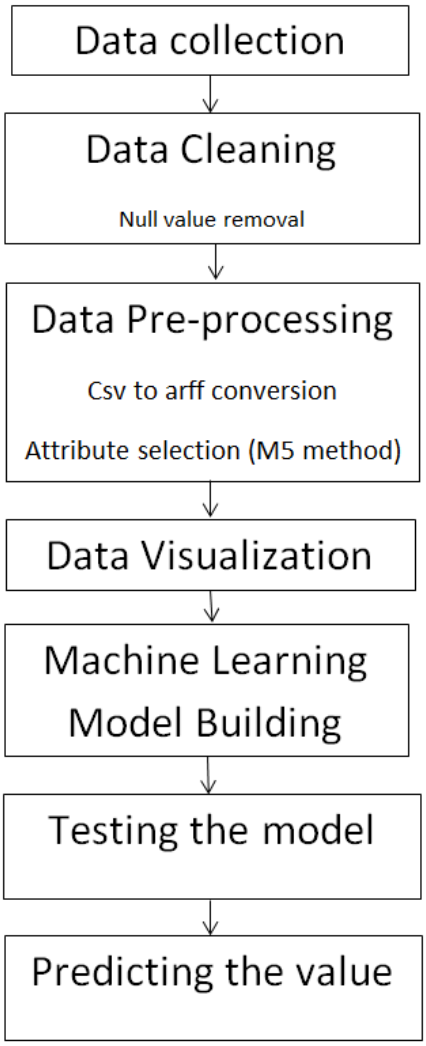
3.2 HARDWARE AND SOFTWARE DESIGNING

- A computer with basic support system is required as a hardware requirement.
- ECLIPSE IDE
- WEKA
- EXCEL

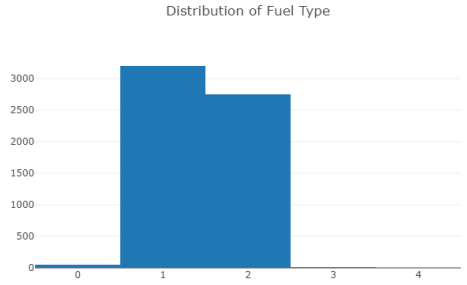
EXPERIMENTAL INVESTIGATION:

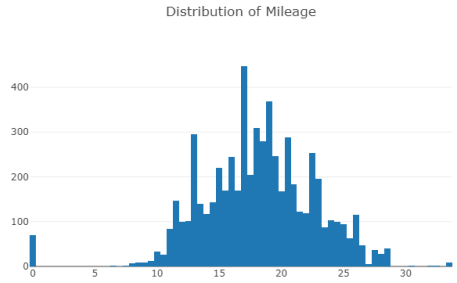
To analyze the degree to which our features are linearly related to price, we plotted the Price against Mileage and Year for a particular Make and Model. There seemed to be a fair degree of linearity for these two features. We utilized several classic and state-of-the-art methods, including ensemble learning techniques, with a 90% - 10% split for the training and test data. For most of the model implementations, Linear Regression as our baseline methods. Linear Regression was chosen as the first model due to its simplicity and comparatively small training time. The features, without any feature mapping, were used directly as the feature vectors. No regularization was used since the results clearly showed low variance

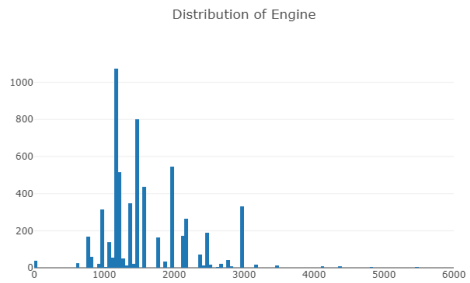
FLOWCHART:

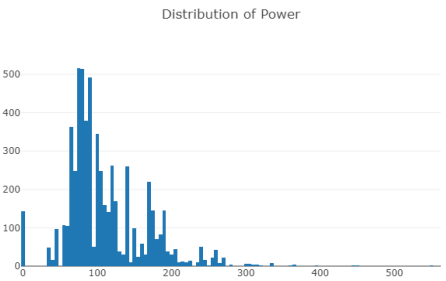

Data collection
Data Cleaning
Null value removal
Data Pre-processing
Csv to arff conversion
Attribute selection (M5 method)
Data Visualization
Machine Learning
Model Building
Testing the model
Predicting the value

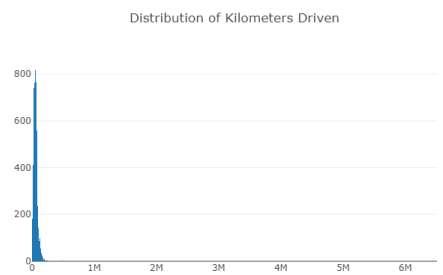
RESULT:

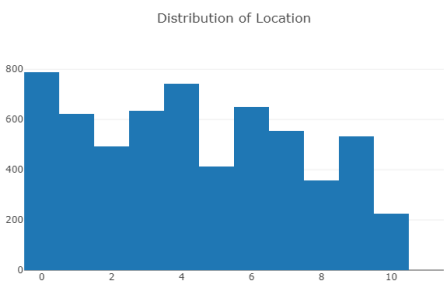

Distribution of Fuel Type
3000
2500
2000
1500
1000
500
0
0
1
2
3
4

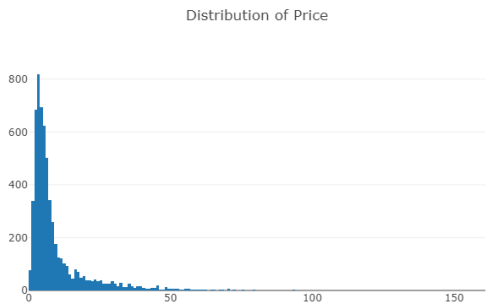

Distribution of Mileage
400
300
200
100
0
0
5
10
15
20
25
30

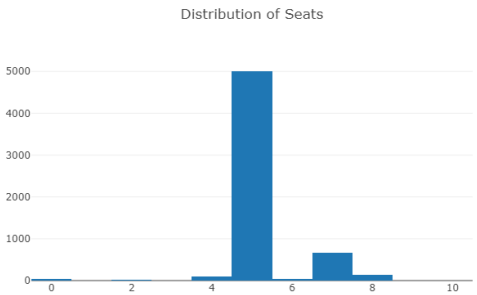

Distribution of Engine
1000
800
600
400
200
0
0
1000
2000
3000
4000
5000
6000

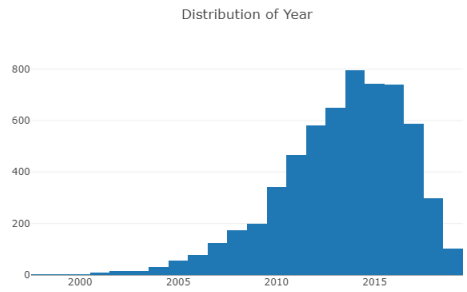

Distribution of Power
500
400
300
200
100
0
0
100
200
300
400
500


Distribution of Kilometers Driven
800
600
400
200
0
0
1M
2M
3M
4M
5M
6M


Distribution of Location
800
600
400
200
0
0
2
4
6
8
10


Distribution of Price
800
600
400
200
0
0
50
100
150


Distribution of Seats
5000
4000
3000
2000
1000
0
0
2
4
6
8
10



The results of our tests were quantified in terms of the R 2 score of our prediction. R 2 score is a statistical measure of how close the data are to the fitted regression line.

```
Correlation coefficient                  0.9819
Mean absolute error                      0.0743
Root mean squared error                  0.1817
Relative absolute error                 13.712  %
Root relative squared error             18.9232 %
Total Number of Instances            1234

-------------------
Predicted label:
7.4336052009869995
```

ADVANTAGES:

- Can be used in making good investments.
- Effectively determine the worthiness of the car using a variety of features.
- The algorithm used is the best because of it's less complexity to compared to other algorithms.
- Gives information about the relevance of features.
- Overfitting is a situation that arises when a machine learning model fits a dataset very closely and hence captures the noisy data as well
- Regularization is a technique that can be.easily implemented and is capable of effectively reducing the complexity of a function so as to reduce the risk of overfitting.

DISADVANTAGES:

- Prone to underfitting,A sitiuation that arises when a machine learning model fails to capture the data properly.This typically occurs when the hypothesis function cannot fit the data well.
- Sensitive to outliers.Data outliers can damage the performance of a machine learning model drastically and can often lead to models with low accuracy.
- Outliers can have a very big impact on linear regression's performance and hence they must be dealt with appropriately before linear regression is applied on the dataset.

APPLICATIONS:

- The Client,to be able to predict used cars market value can help both buyers and sellers.
- Used car sellers (dealers)
- Online pricing services
- Individuals: There are lots of individuals who are interested in the used car market at some points in their life because they wanted to sell their car or buy a used car. In this process, it's a big corner to pay too much or sell less then it's market value

CONCLUSION:

Building up a relatively good performance of Linear Regression, ensembles Learning Method (with K = 3) producing the best R score on test data without high variance as 2 it fits linear relationships categorically. The deep neural network was converging to local minima due to small batch-sizes. By performing linear regression model, it was aimed to get different perspectives. With this study, it purpose was to predict prices of used cars by using a dataset that has 13 predictors and 6019 observations. With the help of the data visualizations and data analysis, the dataset was uncovered and features were explored deeply. The relation between features were examined. At the last stage, predictive models were applied to predict price of cars in an order.

FUTURE SCOPE:

For better performance, we plan to judiciously design deep learning network structures, use adaptive learning rates and train on clusters of data rather than the whole dataset. To correct for overfitting in Random Forest, different selections of features and number of trees will be tested to check

for change in performance

BIBILOGRAPHY:

1. Reffered To: https://github.com/kshitijkumbar/CS229-Project.
2. For Dataset: https://www.kaggle.com/c/1056lab-used-cars-price-prediction
3. ML | Linear Regression - GeeksforGeeks
4. N. Monburinon, P. Chertchom, T. Kaewkiriya, S. Rungpheung, S. Buya and P. Boonpou, "Prediction of prices for used car by using regression models," 2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, 2018, pp. 115-119

APPENDIX:

SOURCE CODE:

https://github.com/smartinternz02/SPS-10601-Car-Resale-Value-Prediction