



NATIONAL UNIVERSITY OF COMPUTER AND EMERGING SCIENCES

MAIN CAMPUS KARACHI

Department of Computer Science

Spring 2021

COURSE:

DATA SCIENCE CS-481

COURSE INSTRUCTOR:

DR. MUHAMMAD NOUMAN DURRANI

REPORT ON:

PROJECT SAMUSH

USED VEHICLES' PRICE PREDICTION MODEL

TEAM MEMBERS:

[18K-0366]

SAYED MUSHAHID RAZA QUADRI

[18K-1147]

ABDUL SAMI SHAIKH

STUDENTS OF SECTION 6-C

ABSTRACT:

Used vehicles are one of the most widely sold products in online market also one of the most neglected and forsaken products by data scientists of the platform providers. Unlike new vehicles they do not have a set price or any method of validation to make sure that their price is according to the market rate. We try to solve this issue for car sellers and buyers by making an AI based Supervised Learning model which can accurately predict the prices of the used vehicles given some parameters about the car.

INTRODUCTION:

This project is developed in python 3.0, on windows platform. It uses many supervised learning models to accurately predict the price of used cars. The dataset is Dataset obtained from Keggles and is authorized and collected by "Pak Wheels" for learning purposes, the dataset contained record of more than 56,000 different vehicles and it had 16 different features. Our goal was to develop an algorithm which will use the above dataset to train a model which can give us a line of regression on which we can predict price of used vehicles of different types. This resultant model can be utilized for multiple purposes by car selling and buying platforms and will provide and set a standard for it.

METHODOLOGY:

- Load Data in to Pandas Dataframe
- Printed it's features and head records
- Dropped useless features
- Column Renaming, Type Casting & Attributes Division
- Deleted the records with outliers and NAN values
- Separation of Target Variable 'Price'
- Label Encoding to handle Categorical Data
- Data Scaling to avoid unwanted Biasness
- Test/Train Split (80% Train, 20% Test)
- Data Visualization to see the relationship among features
- Converted Non Linear Related Data To Linear Taking Natural Log
- Stepwise Regression Technique Applied
- Hold Out Method used for Validation
- Model Training and Fitting
- Accuracy Testing
- Visualization of Line of Regression

ACCURACY RESULTS:

- Random Forest Tree Regressor (Accuracy = 94.23%)
- KNN Regressor (Accuracy = 93.85%)
- Decision Tree Regressor (Accuracy = 91.45%)
- SVM Regressor (Accuracy = 90.49%)
- Neural Networking Regressor (Accuracy = 90.33%)
- Linear Regressor (Accuracy = 78.74%)

CONCLUSION:

This projected had been a lot better if Covid-19 Lockdown would not have happened. More features and hyper parameters would have been added and more details would have been taken into account. In future image recognition can also be included to improve the price predictions. Overall all algorithms except Linear Regression give an accuracy of above 90% but KNN and Random Forest outperforms every other algorithm. Since KNN is lazy learner and slow compared to Random Forest, Random Forest is the best choice for this MODEL.

This report is under your kind consideration and perusal.

THE END