**Car Price Prediction**

**Abstract**

Used car price prediction using AI/Machine Learning techniques has picked researchers interest since it takes a significant amount of work and expertise on the part of the field expert. For a dependable and accurate forecast a large number of unique attributes are analyzed. We employed 6 different machine learning approaches to develop a model for forecasting the price of used automobiles.

**Problem Statement**

With the coronavirus away on the lookout, we have seen a lot of changes in the vehicle market. Presently some vehicles are sought after subsequently making them exorbitant and some are not popular and consequently less expensive. With the adjustment of the market due to the coronavirus 19 effect, people/sellers are facing issues with their past car price valuation AI/Machine learning models.

The primary point of this used car price prediction AI/machine learning project is to create a dataset with the help of web scraping and anticipate the cost of a trade-in vehicle given different elements.

**The Objective of the Project**

Data Collection: To scrape the date of atleast 5000 used cars from various websites like OLX, cardeho, cars24 etc

Model Building: To build a supervised machine learning model for forecasting the value of a vehicle based on multiple attributes.

**Motivation Behind the Project**

There are a few major worldwide multinational participants in the automobile sector, as well as several merchants. By trade, international companies are mostly manufacturers, although the retail industry includes both new and used automobile dealers. The used automobile market has seen a huge increase in value, resulting in a bigger percentage of the entire market. In India, about 3.4 million automobiles are sold each year on the secondhand car market.

**Collecting the Data**

We have scraped the data for over 5000 cars using Selenium script from 4 different websites from different locations around the country. The websites are as followed:  
1. OLX  
2. Cars24  
3. CarDekho

1.’Brand & Model’: It gives us the brand of the car along with its model name and      manufacturing year

2.’Varient’: It gives us a variety of particular car model

3.’Fuel Type’: It gives us the type of fuel used by the car

4.’Driven Kilometers’: It gives us the total distance in km covered by car

5.’Transmission’: It tells us whether the gear transmission is Manual or Automatic

6.’Owner’: It tells us the total number of owners cars had previously

7.’Location’: It gives us the location of the car

8.’Date of Posting Ad’: It tells us when the advertisement for selling that car was posted online

9.’Price (in ₹)’: It gives us the price of the car.

**Data Cleaning**

We check for null values and find that there are few in column ‘Variant’ and we will treat them with Mode.  
Since all the features are categorical hence we need not check for outliers and skewness.  
Exploratory data analysis  
Firstly, we will plot the boxplot and distribution plot for the target variable. And find that few outliers need not be treated and the data is tightly distributed with an almost normalized distribution.

**Bar Graph**

Since Brands, Varients, Driven Kilometers & locations have a wide range of values in them, we will not perform bivariate analysis for them as they will not give us any specific details. Now by plotting the graph of Fuel Type, Transmission, and Owner against Price, we conclude that a Car that uses Diesel has automatic Transmission, and Has only 1 owner is more likely to have a high price.

**Model Building**

Linear Regression

RandomForestRegressor

**Conclusion**

Here, we can see that all the predicted prices are either equal or nearly equal to the original prices of the car. Hence we conclude that our model ‘price car’ is working very well. And we shall save it for further use.