Old Vehicle Price Prediction Using Machine Learning

By,

Debjyoti Basak (18700120158)

Rabi Majumder (18700120141)

Tushar Singh (18700120146)

Shirsha Mullick (18700117032)

Introduction-Old Vehicle Price Prediction

- Vehicle price prediction especially when the vehicle is used and not coming directly from the factory, is equally critical and important task.
- Increasing demand of used cars in the market, more and more vehicle buyers are finding alternatives of buying new cars.
- The aim is to develop machine learning models that accurately prognosticate a used car's price based on its features, empowering customers to make informed choices.

Problem Definition

Problem Statement:

- The prices of new cars in the market is generally standardised by the manufacturer company with some additional/hidden costs incurred by the Government. But due to the increased price of new cars and the incapability of customers to buy new cars due to the lack of capital, used cars sales are on a global increase.

- We tried to create a model that can reliably justify the price of a used car using historical data on old automobiles, including parameters like brand, model, manufacturing year, fuel type, distance driven and other important criterias.

Problem Definition-(Contd.)

Input:

- A collection of informations about used cars, including company, model, year of manufacture, fuel type, mileage, and other pertinent details.

Output:

- Predicted price as per the dataset model of the used vehicle.

Challenges:

- Predicting the price of used cars can be difficult for a number of reasons, such as dealing with incomplete or missing data, handling categorical variables like make and model, managing outliers, dealing with potential multicollinearity among features, and choosing a machine learning algorithm that best fits the data.

Tools Used

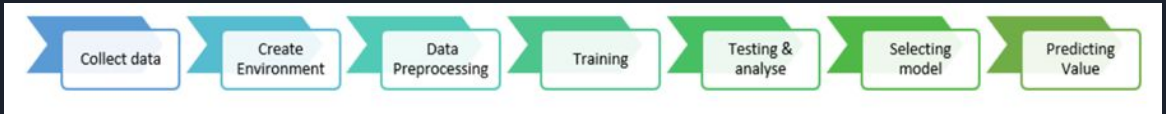
To make this project possible, some data extraction techniques are used, which helped to extract the most important and desired data from the raw dataset, by cutting out lots of unnecessary data present in the same, and it was removed using sorting technique. For this “panda” library of python is used.

This project is incomplete without Graphs and Arrays, so to represent graphs and doing modification with arrays, “numpy”, ”matplotlib” & “matplotlib.pyplot” libraries are used.

For the data visualization, Seaborn visualization Library is used, since it provides high level interface for drawing attractive and informative statistical graphics

Since the base of the project is “Machine Learning”, so to train the models, “Sklearn” or “SciKit-Learning” data analysis library is used.

Methodology and Architecture



- Data Collection: Picking up raw and unprocessed dataset from internet and making an environment for filtering the dataset.

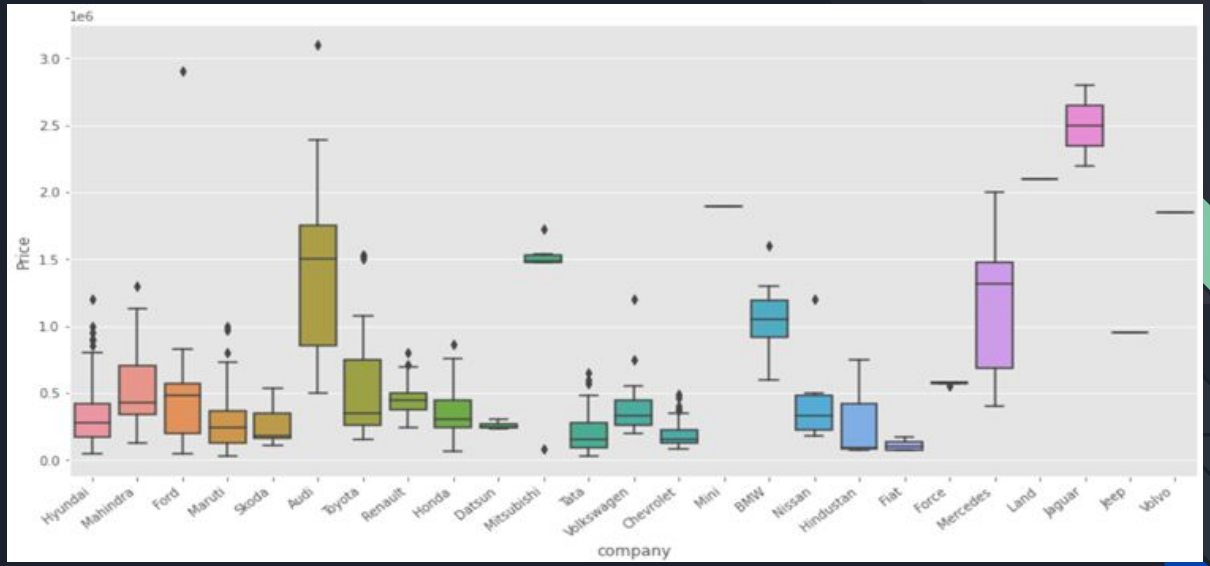
- Preparing Dataset: Identifying unimportant data and picking out the values of wrong attributes.

Methodology and Architecture-(Contd.)

- Analysing Dataset: Spotting all the the inconsistencies ib the dataset and noting them down.

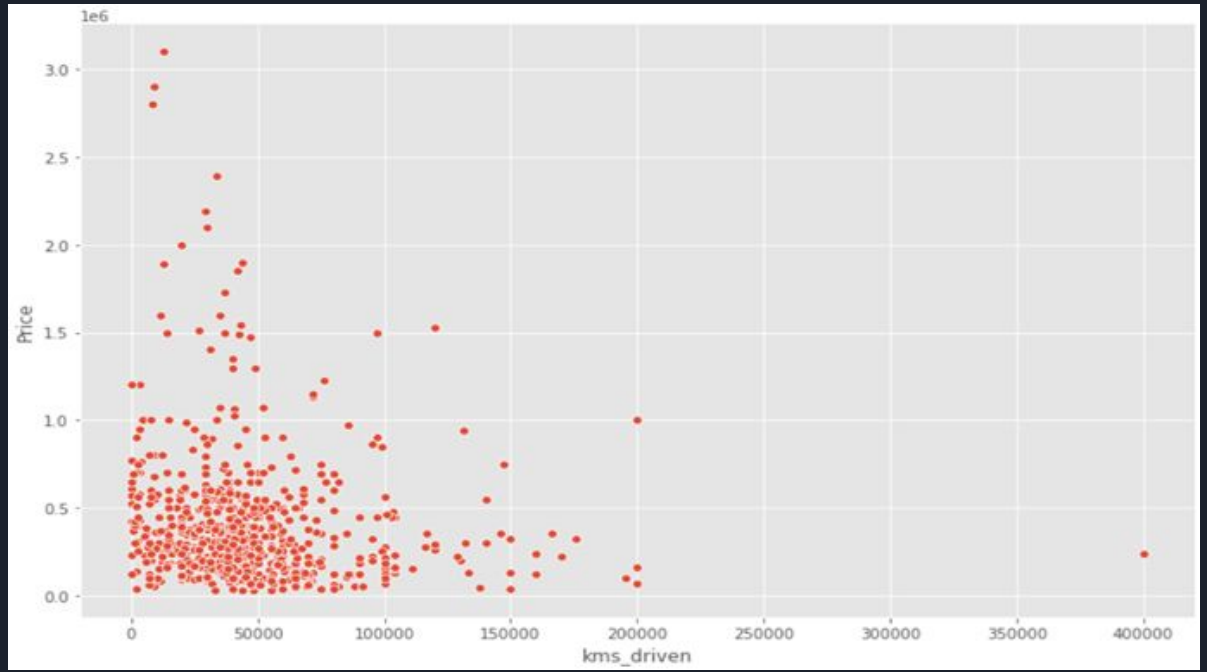
- Cleaning Data: After analysis, the data should be cleaned. In this case, name spam, kilometer values to object, null valued fuel-types etc. are fixed and thus is cleaned the data and saved it.

Relationships of other factors with price



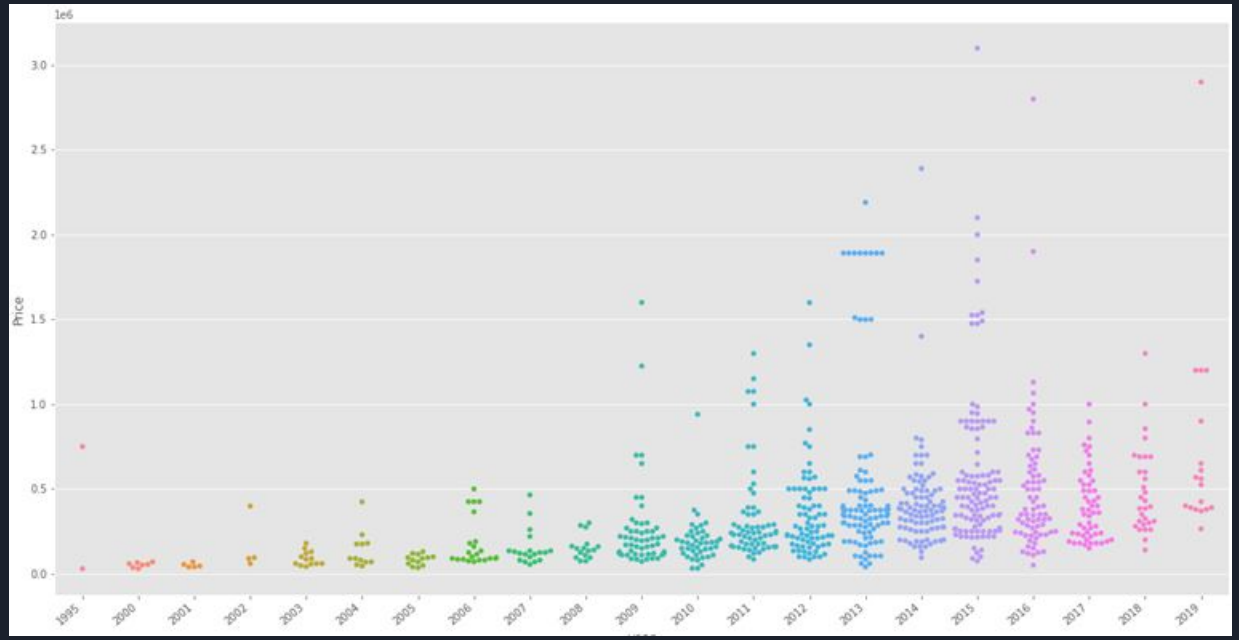
This graph represents the price range of the cars of every individual company, given in dataset.

Relationships of other factors with price



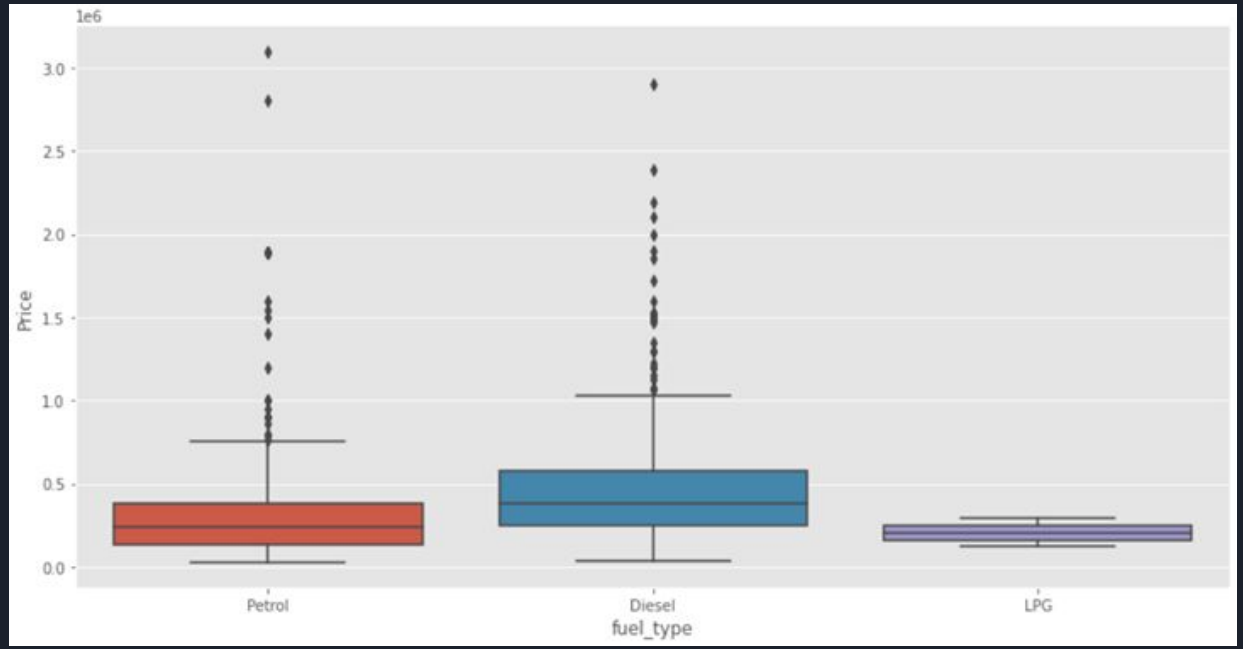
This graph plotting represents the kilometers_driven by the cars present in the cleaned data compared to the price.

Relationships of other factors with price



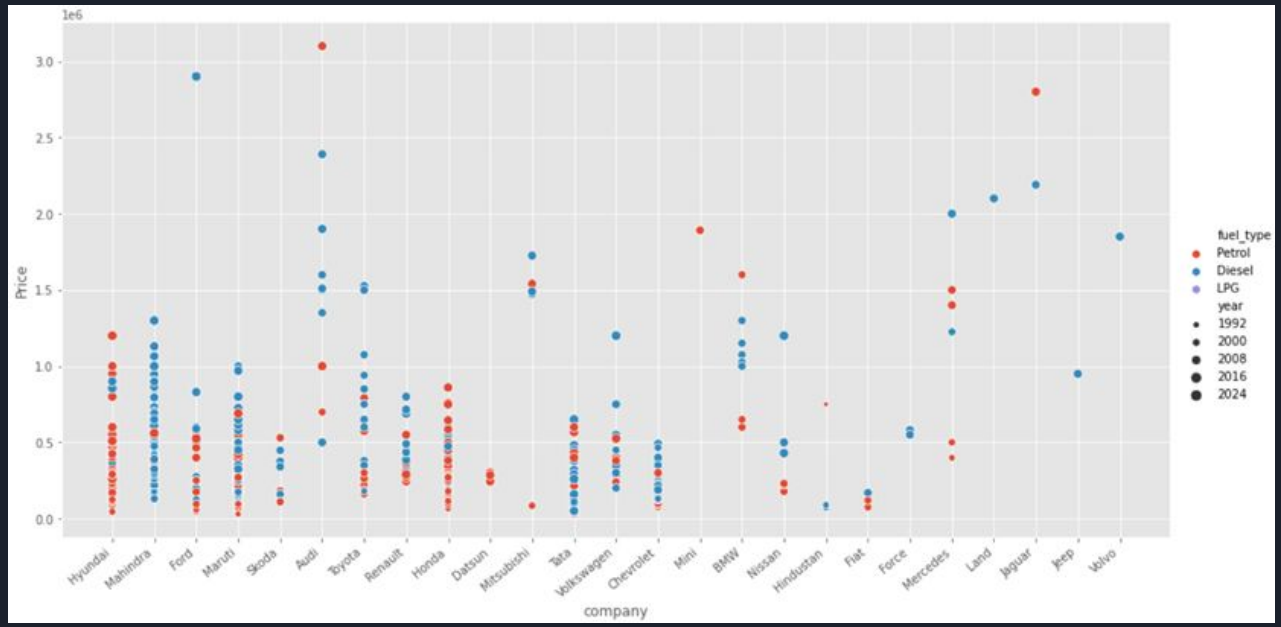
This graph represents the manufacturing year of the cars present in the cleaned data compared to the price.

Relationships of other factors with price



This candle representation portrays the fuel_type of the cars present in the cleaned dataset compared to the price

Relationships of other factors with price



This dotted pattern representation portrays the summary of the fuel_type, manufacturing year and company of the cars present in the cleaned dataset compared to the price

Initiating Machine Learning with the dataset

In this step, '**sklearn**' library is imported and from sklearn, there are various modules are later imported, like train_test_split, LinearRegresion, r2_score, OneHotEncoder, make_column_transformer and make_pipeline.

Train_test_split is a model validation procedure, allowing user to simulate how the model would perform on new or unseen data.

LinearRegression is a data analysis technique to predict the value of unknown data value by using another known data value. There are another 2 regression, but for this case LinearRegression is best.

R2_score, which is also known as the coefficient of determination is used to evaluate the the performance of a regression-based machine learning model.

OneHotEncoder is a common method for dealing with categorical data, which is transformed into such a format that may be fed into machine learning algorithms to improve prediction accuracy.

Make_column_transformer helps to applying OneHotEncoder to the input data and **make_pipeline** helps to pass the input to OneHotEncoder and later to LinearRegression, so that machine can be fed. And also from the other end of the pipeline the output can be obtained.

Future Scope

Used car market:

- As the demand of used cars continues to grow, accurately predicting used car prices can provide valuable insight to buyers and sellers in the used car market.

Auto insurance:

- Accurate price predictions also help detecting fraud by identifying exaggerated claims related to vehicle value.

Fleet management:

- Fleet managers often need to assess the value of vehicles in order to make decisions about buying, selling, or retaining them in their fleet.

Future Scope-(Contd.)

Financial operations:

- Banks, credit unions, and other financial institutions often need to value older vehicles for loan or lease purposes.
- Machine learning models help provide accurate vehicle valuations based on historical data and current market trends, helping financial institutions make informed decisions about credit terms, interest rates, and collateral valuations.

Personalized vehicle evaluation service:

- Advances in digital technology may lead to the proliferation of personalized vehicle valuation services.
- Users can enter specific details of their older vehicle into the online platform, and machine learning models can generate estimates based on historical data and market trends.

Conclusion

- In summary, using machine learning to predict used car prices is promising for a variety of applications. By leveraging historical data, market trends, and vehicle attributes, machine learning models provide valuable insights to buyers, sellers, insurance companies, fleet managers, financial institutions, and individual vehicle owners.

- Accurate price forecasts promote fair trading, streamline fleet operations and financial planning, help detect fraud, and enable users to make data-driven decisions.

- As technology continues to advance and more data becomes available, the future scope of used car price prediction may expand, offering new opportunities for the automotive and related sectors.

Thank You!