MATHEMATICS FOR INTELLIGENT SYSTEM

GROUP-8

TOPIC: AUTOMOBILE MPG PREDICTION

MEMBERS:

ABHISHEK ASHOK KUMAR - AM.EN.U4AIE21002

C GURU ASWINI DATH - AM.EN.U4AIE21023

ERIC OOMMEN - AM.EN.U4AIE21027

MANAV MANIPRASAD - AM.EN.U4AIE21043

ANUPA SAJIKUMAR - AM.EN.U4AIE21071

VISHNUJITH MANU - AM.EN.U4AIE21082

1. INTRODUCTION:

The fuel consumption of motor vehicles powered by internal combustion engines unchangeably constitutes an essential issue in forming not only the development trends of the automotive industry, but is also an important factor in the economics of transport . This problem is certainly a complex issue and must be considered as a many-faceted one. In logistics and business economics, fuel consumption generates costs relating to the transport of materials and products, which are an important component of a company's expenditure . From the viewpoint of enterprise economics, an underlying direction of an optimisation is to minimise the costs linked to fuel consumption . Climate change and the degradation of the natural environment, related to the human and industrial activities, are an aetiology of an ecological approach to the processes of the consumption of petroleum-based fuels in vehicles powered by internal combustion engines . The manufacturers from the automotive industry are obliged to fulfil the more and more strict EURO emission standards, which are applicable within the European Union.

MPG, or **miles per gallon**, is the distance, measured in miles, that a car can travel per gallon of fuel. MPG is also the primary measurement of a car's fuel efficiency: The higher a car's MPG, the more fuel efficient it is.

Dataset:

The Automobile MPG prediction data set contains the list of cars and their miles distances. We analysis the data into several sub class.

Dataset is 398*9 dimensions and their properties.

#	Column	Non-Null Count	Dtype
0	mpg	398 non-null	float64
1	cylinders	398 non-null	int64
2	displacement	398 non-null	float64
3	horsepower	398 non-null	object
4	weight	398 non-null	int64
5	acceleration	398 non-null	float64
6	model year	398 non-null	int64
7	origin	398 non-null	int64
8	car name	398 non-null	object

Methodology:

This notebook contains the following columns: **mpg, cylinders, horsepower, weight, acceleration,** etc., which should all be self-explanatory.

Displacement is the volume of the car's engine, usually expressed in liters or cubic centimeters.

Origin is a discrete value from 1 to 3. This dataset does not describe it beyond that, but for this notebook we assumed 1 to be American-origin vehicle, 2 is European-origin, 3 is Asia/elsewhere.

Model year is given as a decimal number representing the last two digits of the 4-digit year

The model in this dataset will be trained on many different cars, and it should give us a good estimate for our unknown car's mpg. Note that some of the values in the dataset are incorrect, so we will be fixing those values as we preprocess the data.

2. Data ingestion

According to others using this dataset, some of the mpg values for the cars are incorrect, meaning that some of our predictions will be off by a large amount, but we shouldn't always trust the listed mpg value.

There are also unknown mpg values in the dataset, marked with a '?'. We will need to manually replace these with the correct mpg value.

While our model is the end result of this notebook, the data analysis section will be incredibly important in visualizing trends without having to use any machine learning techniques.

3. Data preprocessing

The purpose of the data preprocessing stage is to minimize potential errors in the model as much as possible.

Generally, a model is only as good as the data passed into it, and the data preprocessing we do ensures that the model has as accurate a dataset as possible.

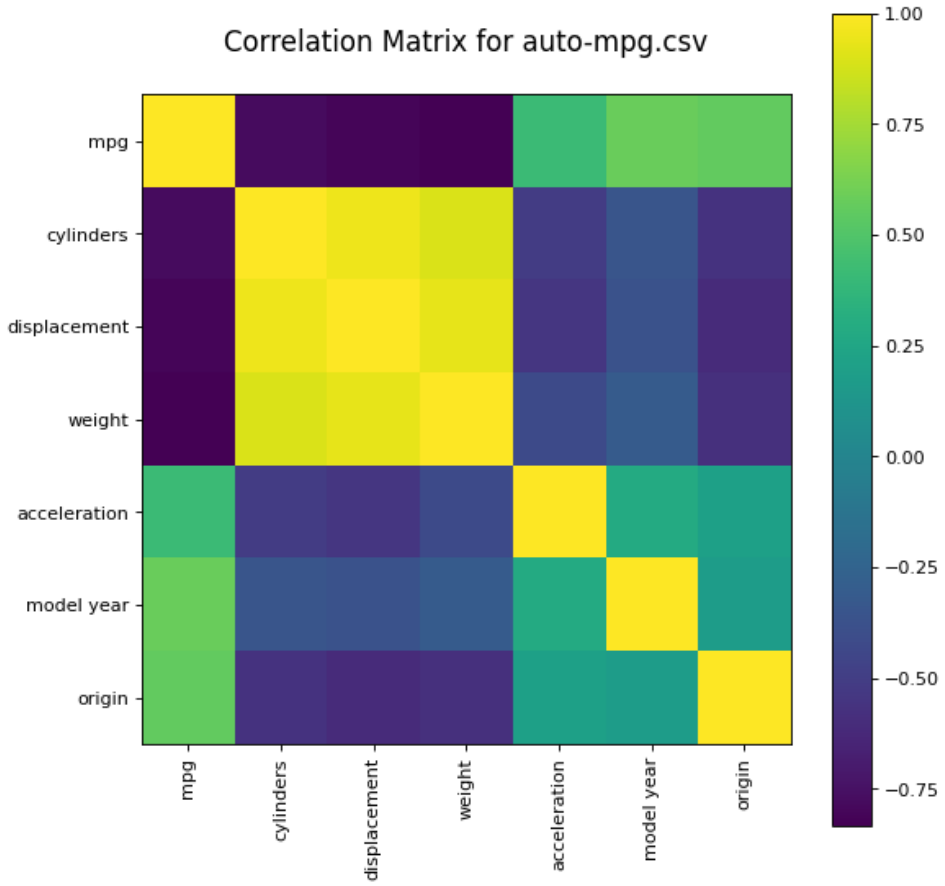
While we cannot perfectly clean the dataset, we can at least follow some basics steps to ensure that our dataset has the best possible chance of generating a good model.

4. EDA

The purpose of EDA is to enhance our understanding of trends in the dataset without involving complicated machine learning models. Oftentimes, we can see obvious traits using graphs and charts just from plotting columns of the dataset against each other.

We've completed the necessary preprocessing steps, so let's create a correlation map to see the relations between different features.

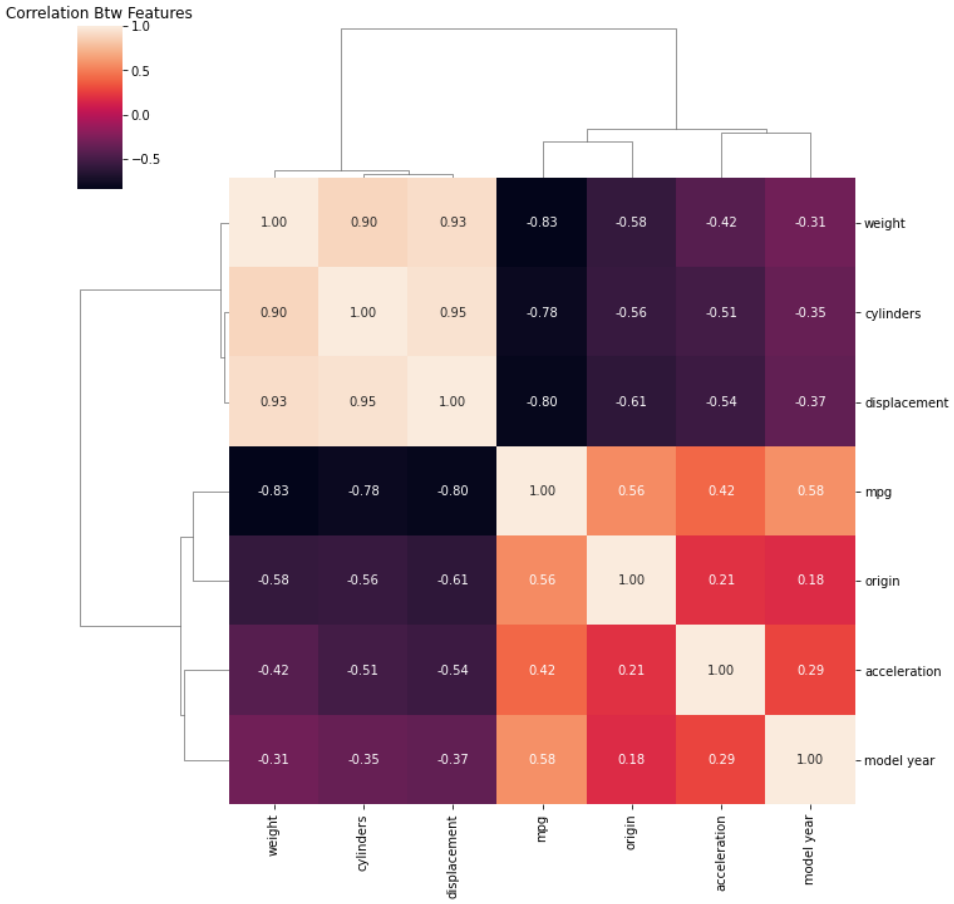
A correlation map (or correlation matrix) is a visual tool that illustrates the relationship between different columns of the dataset.



5. Model Training:

We will be creating and training our model for predicting what a car's mpg will be. Since there are multiple algorithms we can use to build our model, we will compare the accuracy scores after testing and pick the most accurate algorithm.

- At last compare all the types of regressions with there mean square error.



This Image shows the corrections between the two features.

Models used for the prediction:

- Linear Regression Algorithm
- Neural network regression
- Polynomial Regression
- Kernel Ridge Regression
- Bayesian Ridge Regression
- Decision Tree Algorithm
- Lasso Regression
- Ridge Regression
- SVM Algorithm
- Random forest Regressor Algorithm
- Gaussian Regression
- K-Nearest Neighbour (KNN) Algorithm

By training these features for the above mentioned regressors we will find the MPG of the particular car .

- **Linear Regression is implemented by deriving the prediction for each column respectively**
- **Polynomial Regression is implemented by different degree values.**
- **Remaining Regressions implemented by using their features**

Result:

- Linear Regression mean square error= 4.289901161854227e+21
- Lasso mean square root error= 7.437308400407741
- Polynomial mean square error= 4.289901161854227e+21
- Kernel Ridge mean square error= 3.6691420983815473
- Bayesian Ridge mean square error= 3.6691420983815473
- Decision Tree mean square error= 12.822000000000003

Conclusion:

To predict the fuel consumption of a passenger car powered by internal combustion engine, already at the stage of its design on the basis of the technical parameters specified in the assumptions for the design is confirmed in this publication. It is a multi-faceted issue, manifested by the tendency of the automotive market to minimise fuel consumption in relation to the reduction of emissions of harmful chemical compounds into the environment.

Code:

Colab link:

https://colab.research.google.com/drive/1pv1MKkMjQCdaFiQjMVvz92v8QXgRkJc-?usp=sharing