

Capstone proposal: Seid Ahmed

Problem

The first capstone project will focus on Auto Mpg for different cars of origin.

Client

The client could be any one who cares about car efficiency including the Federal government, car dealers who want to sell efficient cars, and others.

Dataset

I will get the data from the UCI website. I use the Auto-Mpg data set

Outlines

Import the required libraries

Download the data from the website

Analyze the dataset and check if there are missing values

Check the Data types if there is a non-numerical value.

Check the skewness of the data distribution and scale the data if its skewed.

Explore the target with different features to determine which feature is more important and decisive.

The scatter plots and heat map could give as a good hint how the features are related.

We evaluate the dataset with some baseline algorithms with the non-scaled data.

After checking the scores of the algorithms, we will proceed to evaluate the same algorithms with a scaled data.

We tune the algorithms we chose based on the scores we get.

We prepare a pipeline to show the reproducibility of the above process.

Present the model.

Deliverables

A code that do all the above