Auto-Mpg

Gibril

Introduction

The Auto-MPG dataset contains information about nine (9) attributes of cars. It can be used for testing different supervised learning algorithms mainly for regression.

The aim of this project is to predict the fuel consumption in miles per gallon using regression techniques.

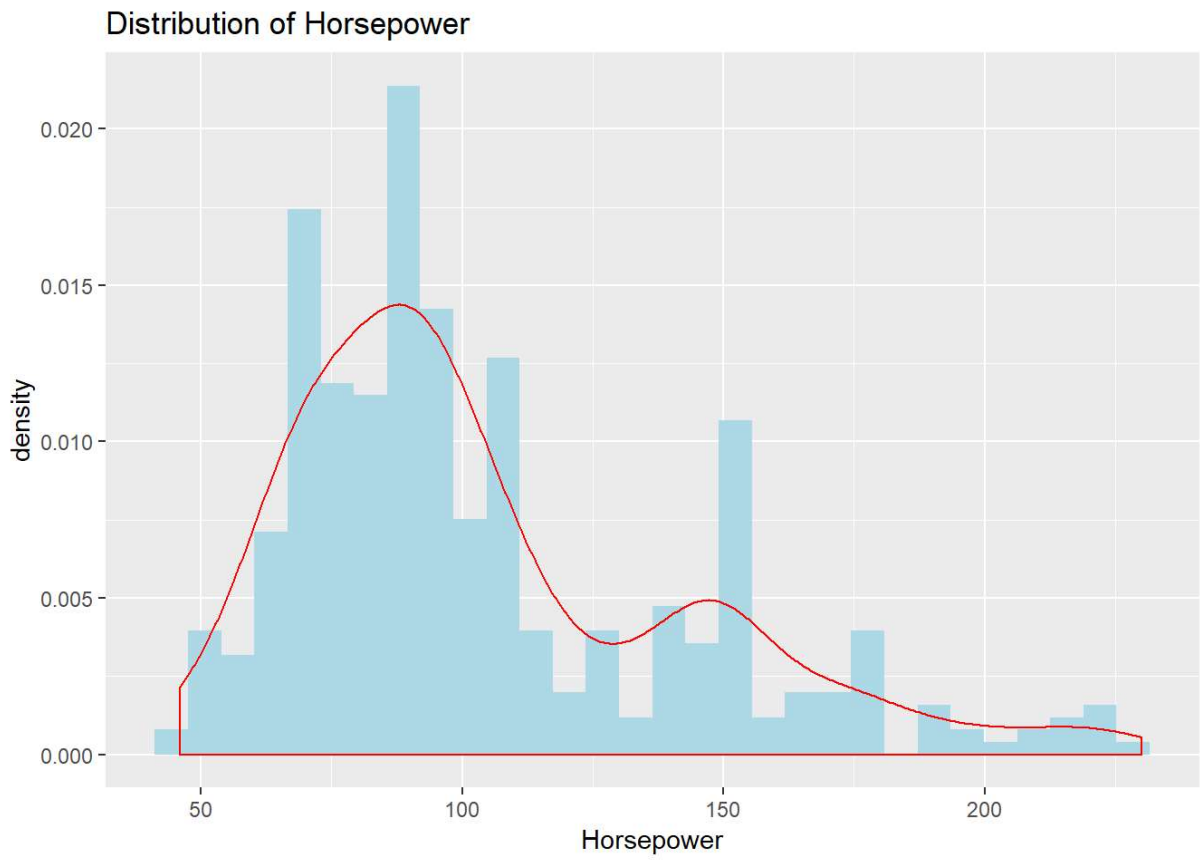
Dataset

This dataset was obtained from UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/auto+mpg). It has 398 observations of 9 variables of cars. Of these, 3 are multivalued discrete and 6 are continuous.

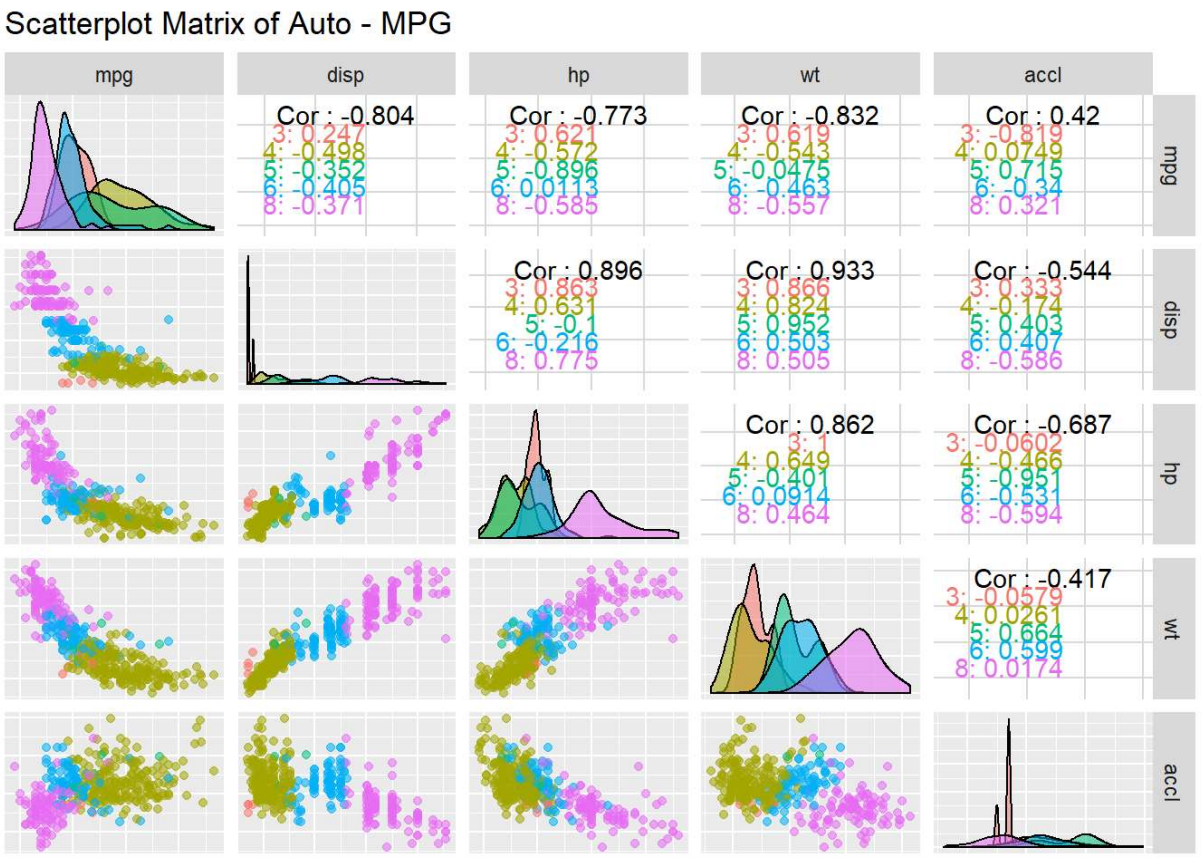
The variables are:

- mpg : city-cyle fuel consumption in miles per gallon
- cylinders : number of cylinders in a car (categorical)
    - 3 cylinders
    - 4 cylinders
    - 5 cylinders
    - 6 cylinders
    - 8 cylinders
- disp : displacement
- hp : engine horsepower
- wt : weight (lbs)
- accl : acceleration
- model_yr : model year
- origin
- car name : name of each car

The data was cleaned by converting the variables: `cylinders`, `model year` and `origin` into factors and the variable `car name` was used to extract manufacturer names of the cars. A new variable `manuc` stored manufacturer names of the cars and then `car names` was dropped. Futhermore, `horsepower` has 6 missing values which were imputed using the median values.



The processed data was stored in a CSV file for analysis. The following graph summarizes the pairwise relationship between the continuous variables in the dataset.



Training Methods

Linear Model

Results

Summary