**INTRODUCTION TO MACHINE LEARNING PROJECT**

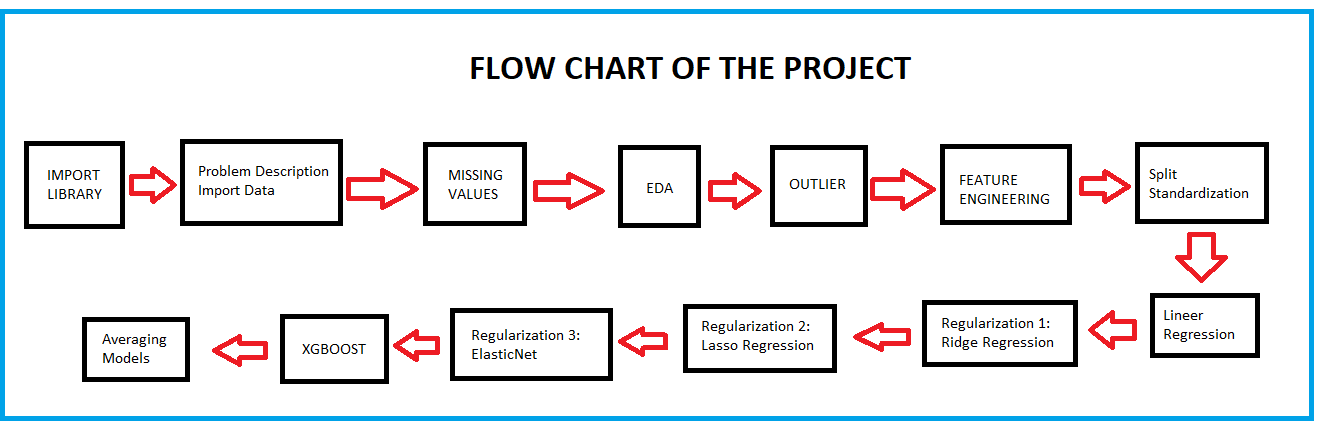
**LECTURER:** DR. MUHAMMED DAVUD

**STUDENT:** MEHMET ŞENER – 030717050

**DEPARTMENT:** SOFTWARE ENGINEERING

**EXPLANATION OF THE PROJECT**

I did a project in introduction to machine learning course , which is Fuel consumption prediction of the vehicles. In my Project, I wanted to use Python programming language because of fact that it gives us many advantages.

I did a lot of operation for my project as a respectively ;

**DATASET INFORMATION**

This dataset is a slightly modified version of the dataset provided in the StatLib library. In line with the use by Ross Quinlan (1993) in predicting the attribute "mpg", 8 of the original instances were removed because they had unknown values for the "mpg" attribute. The original dataset is available in the file "auto-mpg.data-original".  
  
"The data concerns city-cycle fuel consumption in miles per gallon, to be predicted in terms of 3 multivalued discrete and 5 continuous attributes." (Quinlan, 1993)

**Attribute Information of the Dataset**

1 - mpg : continuous (MPG , which means mile per galon)

2 - cylinders : multi-valued discrete

3 - displacement : continuous

4 - horsepower : continuous

5 -weight : continuous

6 - acceleration : continuous

7 - model-year : multi-valued discrete

8 - origin : multi-valued discrete

9 - car name : string (unique for each instance)

**DATASET LINK**

<https://archive.ics.uci.edu/ml/datasets/Auto+MPG>

**Libraries I Use in this Project;**

* Numpy
* Seaborn
* Pandas
* Matplotlib
* Scipy
* Scipy.stats
* Sklearn
* Xgboost

**STEPS OF THE PROJECTS**

**1 – Imputing Missing Value :** In the dataset , Horsepower column has 6 Nan . I have been changed the missing values with mean.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

**2 – EDA :** Expolaratory Data Analysize

**3- Outlier detection and removel.**

**4 – Feature Engineering :** Skewness

**5 – Feature Engineering** **:** One hot encoding

**6 – Prepprocess :** Training / Testing Separation and Standardization

**7 – Lineer Regression**

**8 – Regularization 1 :** Ridge Regression

**9 – Regularization 2 :** Lasso Regression

**10 – Regularization 3:**

**ElasticNet**

Elastic net linear regression uses the penalties from both the lasso and ridge techniques to regularize regression models. The technique combines both the lasso and ridge regression methods by learning from their shortcomings to improve on the regularization of statistical models.

**11 – XGBOOST Algorithm :**

XGBoost is an algorithm that has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

**12 – Averaging Models**

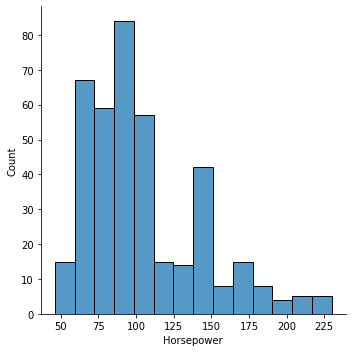
**GRAHPS OF THE OPERATIONS IN THE PROJECT**

FIGURE 1

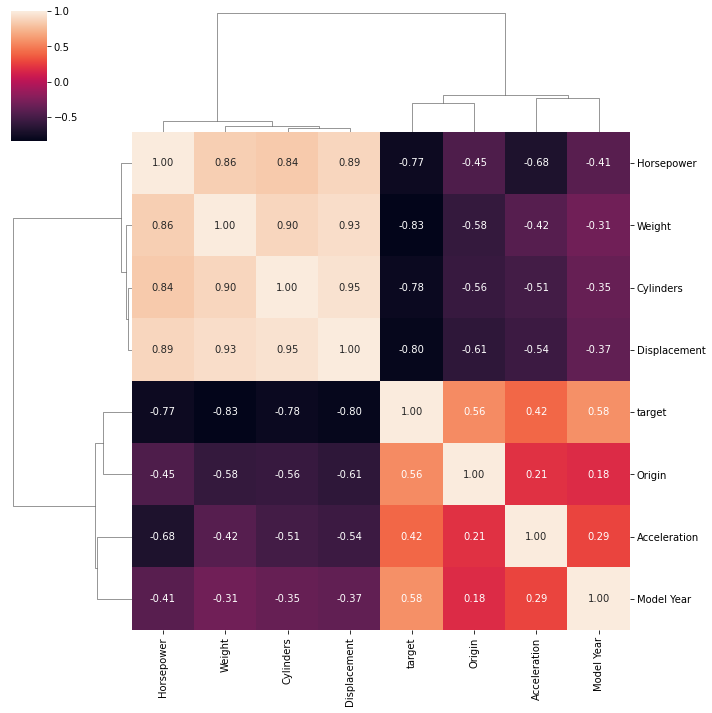
****

FIGURE 2

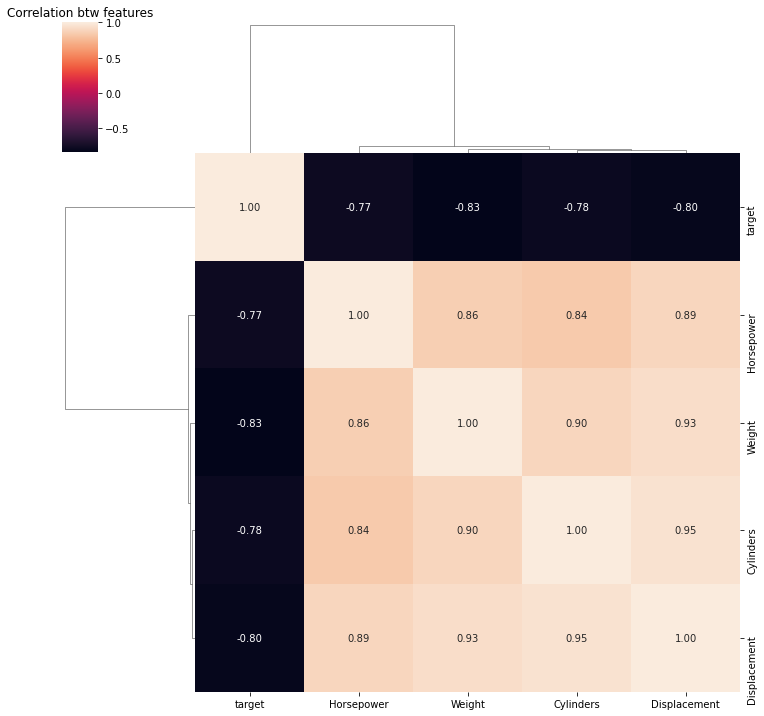
****

FIGURE 3

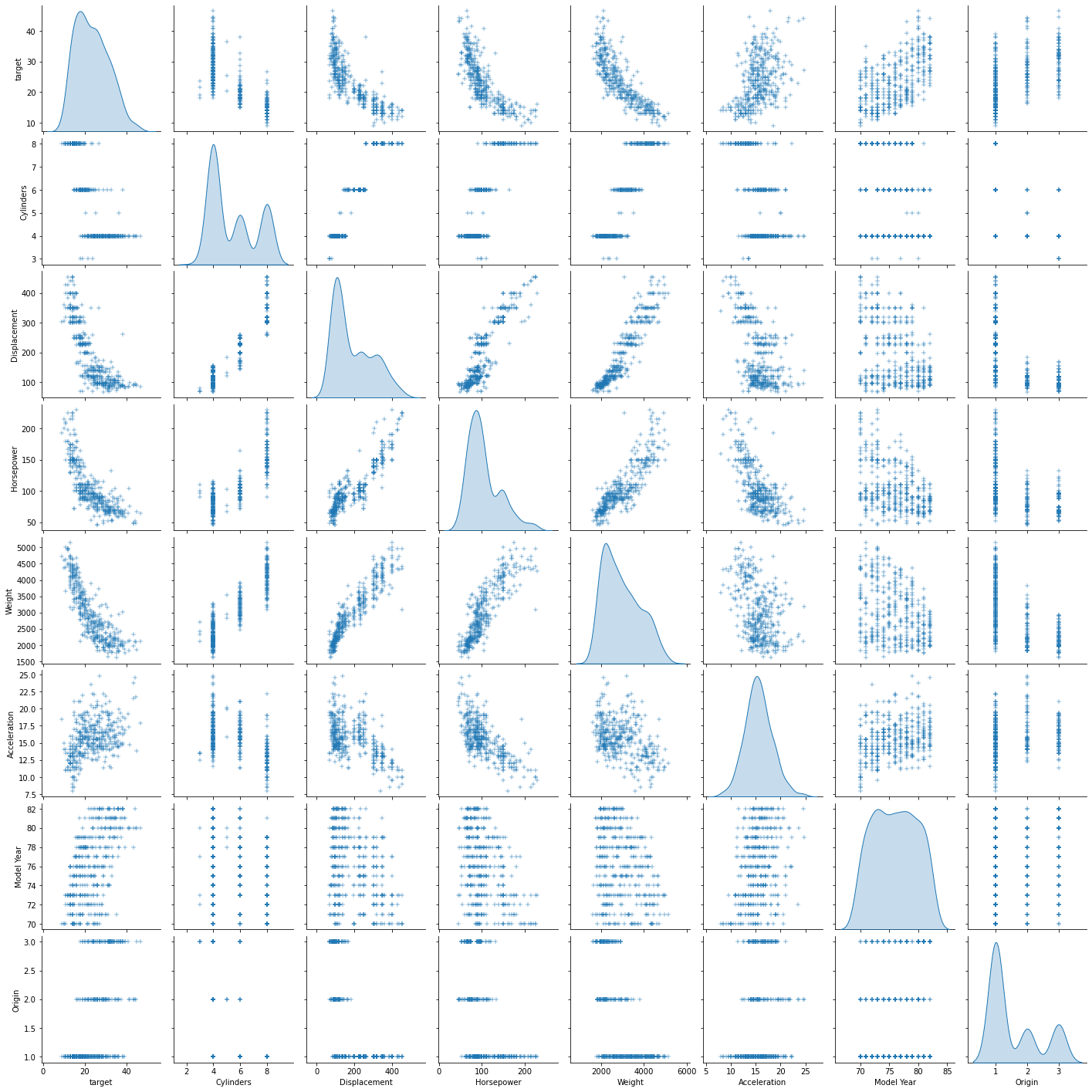


FIGURE 4

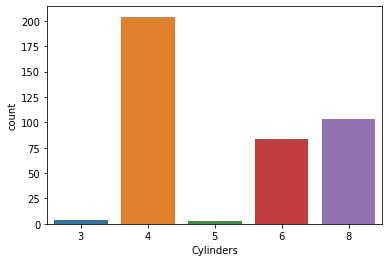


FIGURE 5

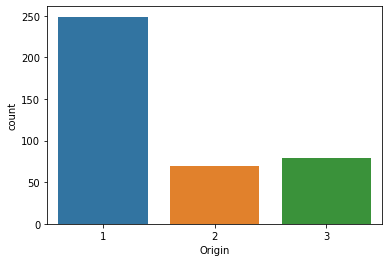


FIGURE 6

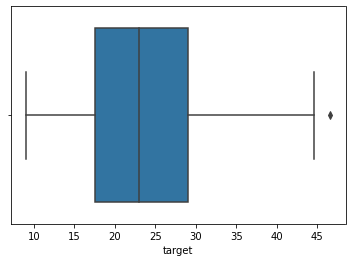


FIGURE 7

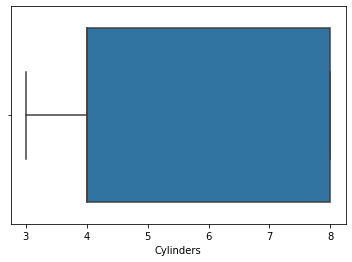


FIGURE 8

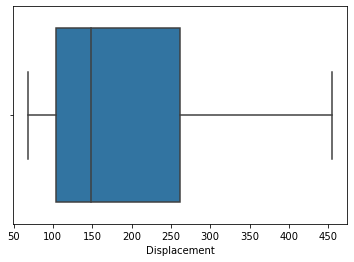


FIGURE 9

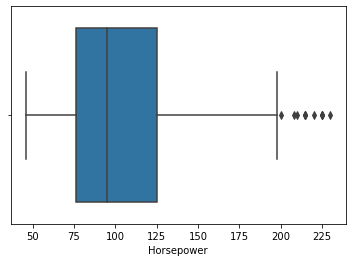


FIGURE 10

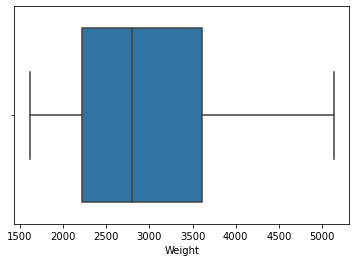


FIGURE 11

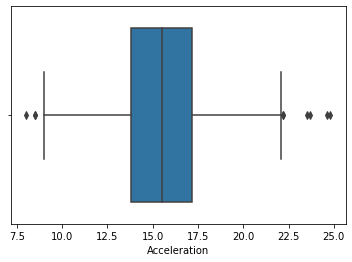


FIGURE 12

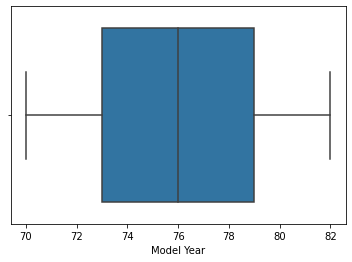


FIGURE 13

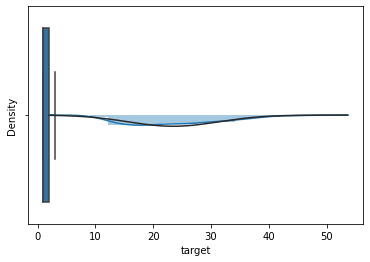


FIGURE 14

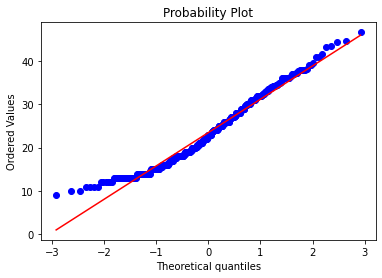


FIGURE 15

metin, siyah, levha içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

****

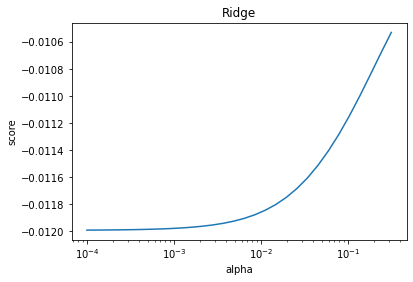
****

FIGURE 16

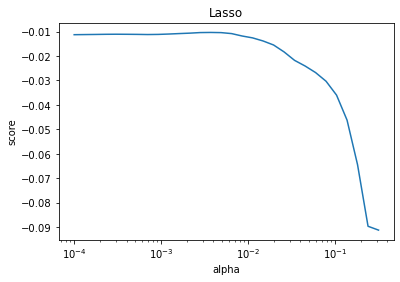


FIGURE 17

**FINAL**

StandardScaler:

Linear Regression MSE: 0.020632204780133015

Ridge MSE: 0.019725338010801216

Lasso MSE: 0.017521594770822522

ElasticNet MSE: 0.01749609249317252

XGBRegressor MSE: 0.017167257713690008

Averaged Models MSE: 0.016034769734972223

RobustScaler:

Linear Regression MSE: 0.020984711065869643

Ridge MSE: 0.018839299330570554

Lasso MSE: 0.016597127172690837

ElasticNet MSE: 0.017234676963922273

XGBRegressor MSE: 0.01753270469361755

Averaged Models MSE: 0.0156928574668921