



**KOÇ  
UNIVERSITY**



**DASC521: Introduction to  
Machine Learning**

Fall 2021

**Instructor:** Mehmet Gönen

**Homework 05-** Decision Tree Regression

Gamze Keçibaş- 60211

06.12.2021

There are three main steps in the homework as importing and preparing datasets, writing three functions to make predictions and calculating errors. Firstly, provided dataset that is “*The Old Faithful geyser in Yellowstone National Park, Wyoming, USA*” is imported to script as pandas dataframe. After that, the set is separated two parts as training and testing set using suggested size as seen below:

```
Size of given dataset: (272, 2)
First five rows:
[[ 3.6  79. ]
 [ 1.8  54. ]
 [ 3.333 74. ]
 [ 2.283 62. ]
 [ 4.533 85. ]]
Size of training data: (150, 1)
First five rows:
[[3.6 ]
 [1.8 ]
 [3.333]
 [2.283]
 [4.533]]
Size of training labels: (150, 1)
First five rows:
[[79.]
 [54.]
 [74.]
 [62.]
 [85.]]
Size of test data: (122, 1)
Size of test label: (122, 1)
```

Figure 1: Size of datasets and first five rows of first three sets

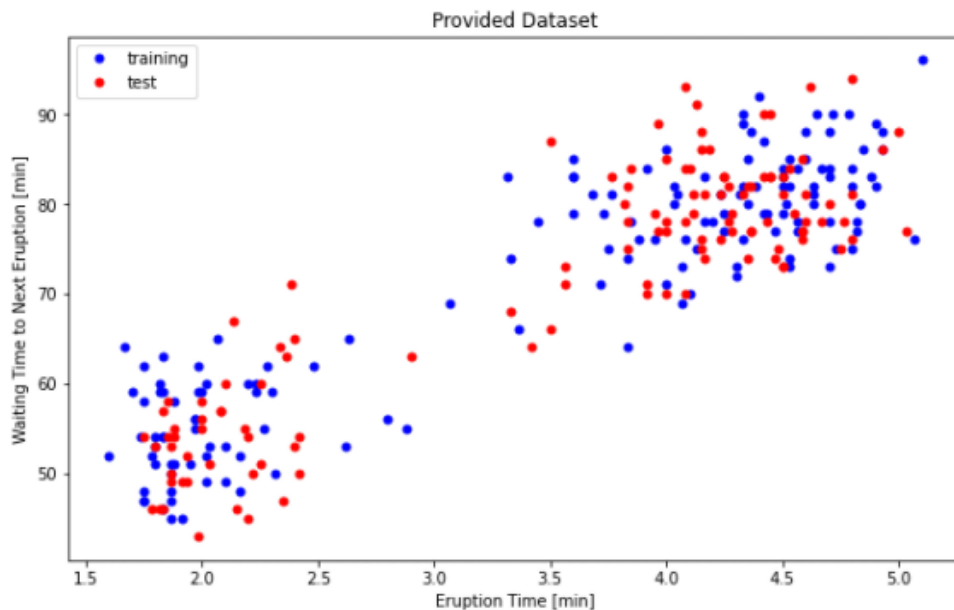


Figure 2: Using dataset

After that, appropriate learning algorithm is written using reference PS code and making prediction is presented in Figure 3 that is black line where pre-pruning parameter is 25:

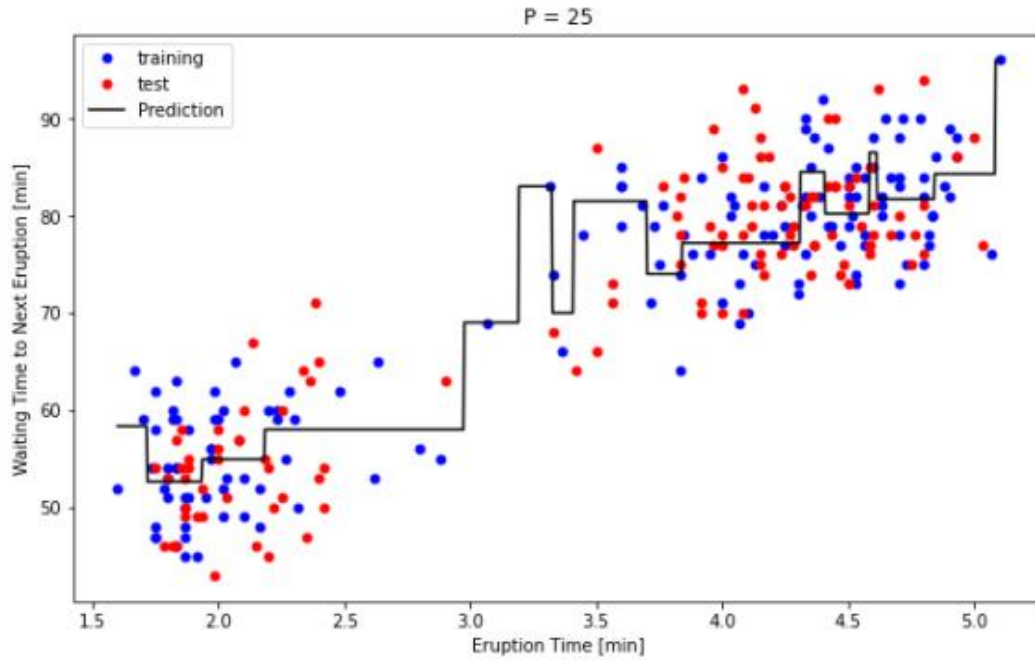


Figure 3: Making prediction by decision tree regression algorithm

Error rate of the algorithm is checked by root mean squared error (RMSE) equation. Its results are equal to 4.57 for training set and approximately 6.50 for test set.

RMSE for training set= 4.573257748671775  
RMSE for test set= 6.4991477594218345

Figure 4: RMSE values

When the pre-pruning parameter is changed between five and fifty five with five increments, RMSE value is changed as seen below:

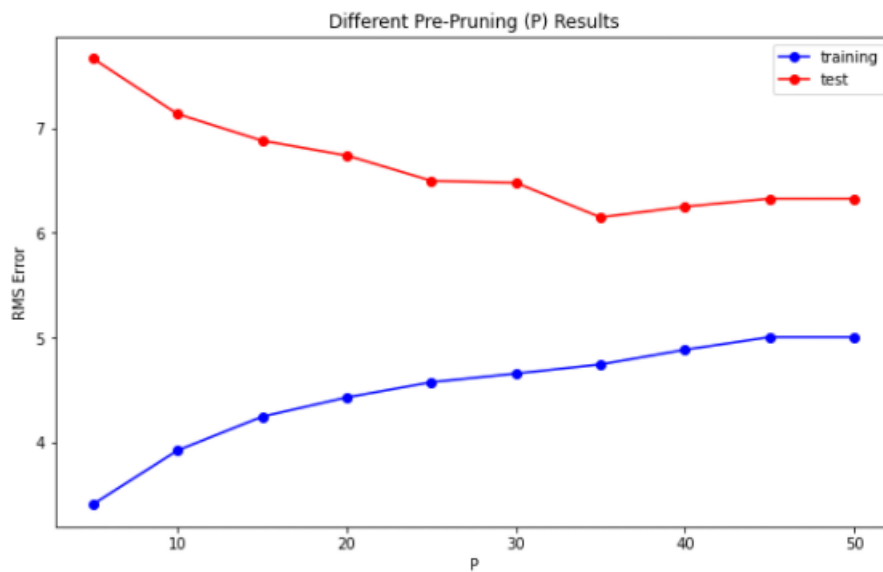


Figure 5: RMSE vs pre-pruning parameter