Engr 421 – HW #4 Report

At first, I divided the data set to into train and test by allocating first 150 of them to train and remaining 122 to test. N_test denotes for the number of test data.

I set the P value to 25 as mentioned in the pdf. I determined the best splits for each node. Inside the loop, I implemented the decision tree regression algorithm with pre-pruning rule which is stop splitting if a node has P or fewer data points. Also, we need to convert that node to a terminal node.

I defined 2 functions which are for maintaining the decission tree regression algorithm and returning the results from this algorithm. I draw the plot with the P value using these functions.

I calculate the root mean squared errors by defining a function which uses the formula given in the pdf for P values from 5 to 50 where step size is 5.

Lastly, I draw the plot of the graph RMSE versus P values.