I first created data maxtrixes as always by using csv file with parse data method.

Then trained the model as following:

Step 1 : take the whole training instances in the root.

Step 2 : if number of indices in the current node <= 15 -> make the current node a terminal node, calculate the mean value of the corresponding y values of indices in the current node and save it as node\_split value. (break the loop)

Step 3 : sort and find the X values which are unique.

Step 4 : find the center of each adjoining unique values.

A picture containing text

Description automatically generatedStep 5: for each center value calculate the score values with the formula below. (less is better).

Step 6 : find the center value which has minimum score and save the center value as node\_split value the node at that point.

Step 7 : create left and right nodes and repeat from (Step 2) for each.

(for terminal nodes node\_split is the mean value of y in this node, for non-terminal nodes node\_split is the center X value of the division)

After training the model, I created predict method such that, for each data to be trained.

Step 1: start from the root node.

Step 2 : if the current node is a terminal node -> return prediction as node split value in this node (break the loop)

Step3 : if the X data <= node\_split of the current node -> (go to Step 2 with left node)

Step4 : if the X data > node\_split of the current node -> (go to Step 2 with right node)

Then I plotted the data by creating an interval from 0 to 1601 and make a prediction for each points.

By using the formula for RMSE I calculated the RMSE value for minimum non-terminal element size 15.

Then by trying different minimum non-terminal element sizes (0, 5, 10, 15 … 50).I calculated RMSE for each and plotted them.