REPORT

PART-01

- Filtering the loaded MNIST dataset from the provided link based on labels belonging to classes 0 and 1. And relabeling them as -1 and 1. The final dimensions of Train and Test data are:

```
Dimensions of Test and Train filtered data:
(12665, 784)
(2115, 784)
```

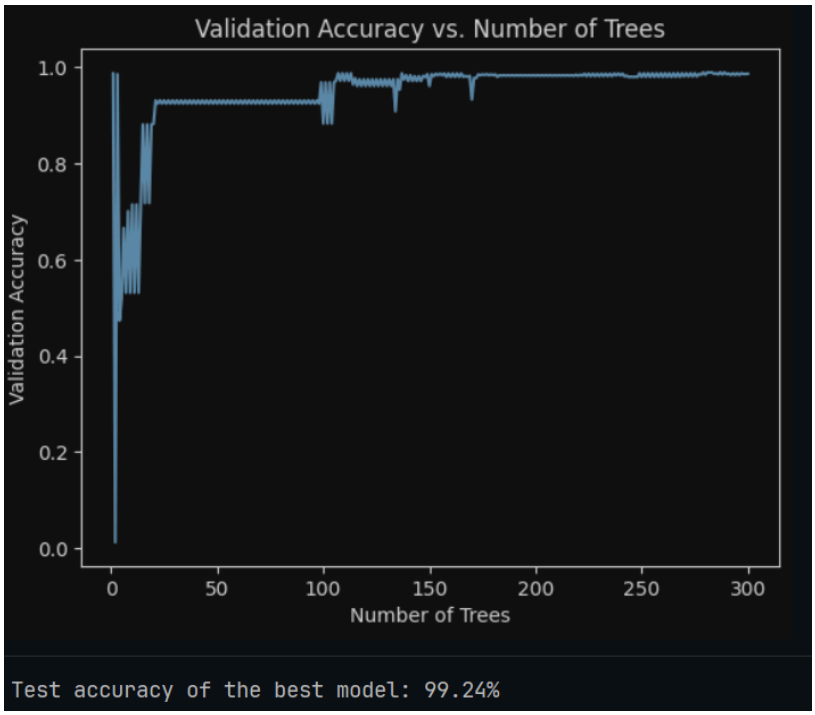
- Partitioning the Train Dataset into train and validation datasets. Selecting 1000 data points for each class -1 and 1 for validation and the rest for the training dataset. The dimensions of training and validation are:

```
Dimensions of Train set: (10665, 784)
Dimensions of Validation set: (2000, 784)
```

- Applying PCA to reduce the features from 784 to 5 of training, validation, and testing. The training dataset should only be used in pca to reduce dimensions.

```
Dimension of Up
(784, 5)
New Dimension of Train Dataset
(10665, 5)
New Dimension of Validation Dataset
(2000, 5)
New Dimensions of Test Dataset
(2115, 5)
```

- Now growing 300 decision stumps. For each stump for each dimension, the best split was evaluated randomly, choosing 1000 midpoints to make computations faster. Alpha is determined after finding the best split that minimizes the weighted misclassification error, and the previous weights are updated with new weights. Again, continuing the process till 300 decision stumps grow. Then choose the best decision stump having the Maximum validation dataset accuracy finally evaluate the Test Dataset's Accuracy. Which came out to be 99.5% (after running for several times). The validation accuracy vs iteration graph and Testing accuracy is next page.



The max validation accuracy noted for this run was 99.00% and for test was 99.24%.

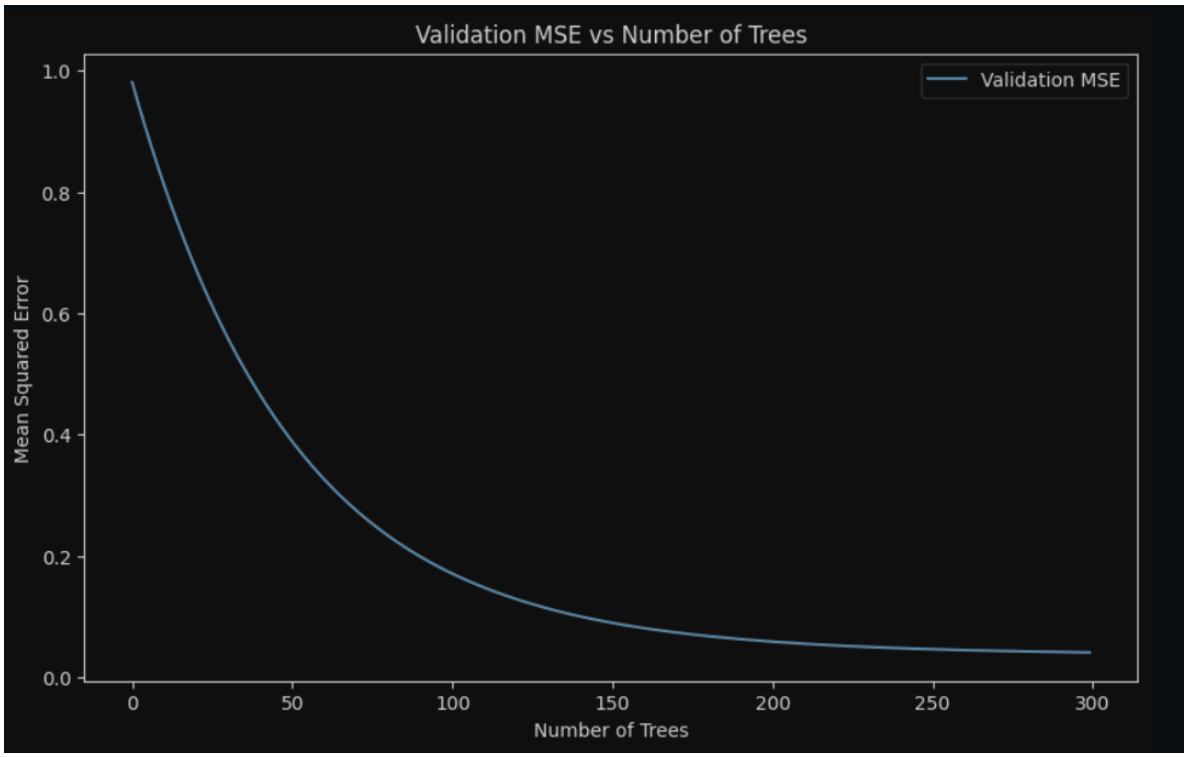
PART-02

- The steps till applying PCA from part01 are the same.

```
Dimension of Training dataset:
(10665, 5)
Dimension of Validation dataset:
(2000, 5)
Dimension of Testing dataset:
(2115, 5)
```

- Similarly, growing 300 decision stumps but this time splitting on the basis of minimum SSR and determining the residue using y-0.01*h1(x). Then, update the labels after every iteration. At each iteration determine the MSE on the validation dataset. Then select the decision stump returning the minimum MSE to compute the MSE for test dataset.

```
Best model test MSE: 0.019771437749838 18
```


Validation MSE vs Number of Trees
Validation MSE
Mean Squared Error
Number of Trees
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300