SML ASSIGNMENT-03 REPORT(2022156)

Using MNIST dataset for Training and Testing of my Decision Tree Model. The Number of features are 784 for one datapoint.

After filtering out my training dataset and testing dataset for only classes 0, 1 and 2 the Shape of my training and testing data is:

```
Dimensions of Test and Train filtered data:
(18623, 784)
(3147, 784)
```

Now applying PCA on my datasets to reduce features from 784 to 10. The shape i got is;

```
Dimension of Up
(784, 10)
New Dimension of Train Dataset
(18623, 10)
New Dimension of Test Dataset
(3147, 10)
```

Decision Tree Model:

For determining the best-split cut using the Gini Index, the point giving minimum Gini index was selected as the best split for that dimension. Then, after checking for each dimension, the best split was taken.

Along a particular dimension, select 20 random points between minimum and maximum values, and the point that results in the minimum Gini index is chosen as the best split and that value as the best Threshold value.

The formula for the Gini Index used 1-(Sum of Squares of all classes)

The classwise and Overall Accuracy is:

```
prediction Vector [2 1 0 ... 0 1 2]
Class 0 Accuracy: 95.00 %
Class 1 Accuracy: 80.00 %
Class 2 Accuracy: 89.15 %
Total Accuracy: 87.67 %
```

For Bagging creating 5 Different datasets with replacement and training 5 different trees then predicting on those different trees and majority results taken in considerations to make final prediction vector, The final accuracies are: