Conclusion

Name: Harnish Savadia

Assignment work:

- Quantized Data as required of Training and Validation data.
- Generated Test Suites as initial practice
- Created Trainer/Mentor Program Manually for each node.
- Created Recursive Trainer/Mentor Program
- Created Classifier Using Output of Trainer/Mentor Program
- Generated Classifications on the Validation set.
- Generated Confusion Matrix
- Computed Accuracy on Validation set.

Maximum Call Depth: Maximum Call Depth of the Final Classifier reaching leaf nodes is 3. And I did not use maximum number levels.

Decision of Final Classifier:

Threshold Computed:

- BangLn: 6.0
- HairLn: 11.0
- TailLn: 9.0
- EarLobes: 1
- Ht: 180
- Reach: 127.0
- Age: 46.0

Generated Confusion Matrix and Accuracy:

```
Creating Confusion Matrix of the Labelled Training Data

[[ 957  243]
 [ 127 1073]]

Accuracy achieved:  84.58333333333333
```

Assam classified as Assam: 957
Assam classified as Bhuttan: 243
Bhuttan classified as Assam: 127
Bhuttan classified as Bhuttan: 1073

Conclusion: There was a lot of back and forth in understanding how the actual threshold will be computed while splitting data on the decision tree. There were a lot of doubts and enjoyed the research behind it. Finally, the entire data was not classified correctly, nevertheless I received an accuracy of about 86%. I revised the quantization of data, parsing the data and computing Gini indices. The new stuff that I learnt was building a decision tree from scratch, splitting data into data frames, building a recursive decision function, as well as generating a confusion matrix.