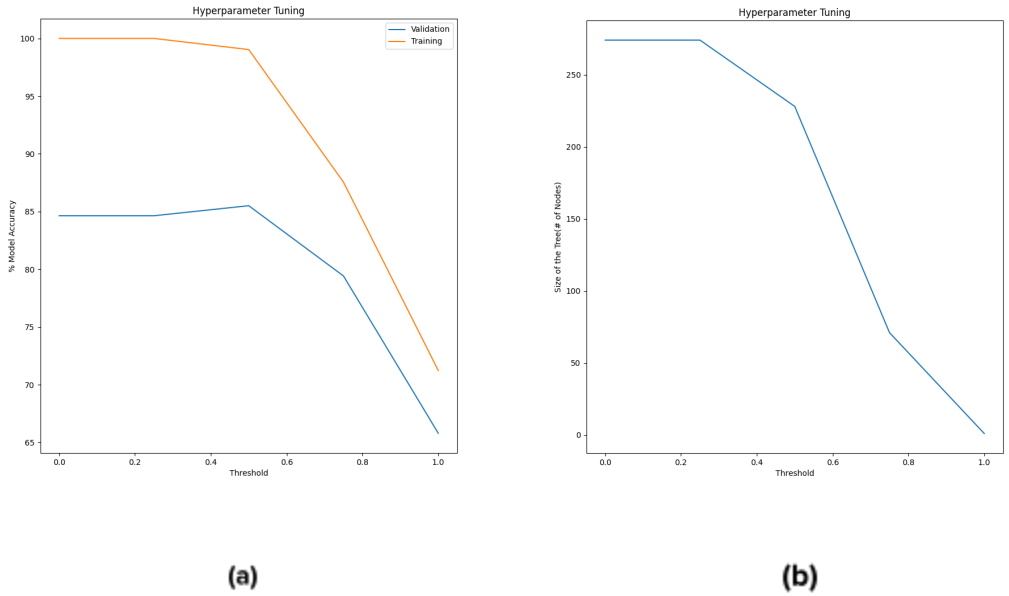
MLFA LAB
Assignment - 4

Name: Aryan Satpathy

Roll: 20EC10014

Experiment 1

Effect of threshold on Accuracy and Size of Decision Tree:



(a) (b)

Figure 1: Effect of threshold

Remarks:

1. At low values of threshold, the tree gets very **high accuracy on the training set** and **low accuracy on.the validation set**. This is a sign of **Overfitting**.
2. At high values of threshold, **accuracy suffers** on **both training and validation data**. This is a sign of **Underfitting**.
3. Optimal value of threshold is **0.5** (**looking at accuracy alone**).
4. **Number of nodes reduces with increase in threshold**. This is because the tree becomes **more relaxed**.

Experiment 2

Relationship between **model accuracy and branching:**

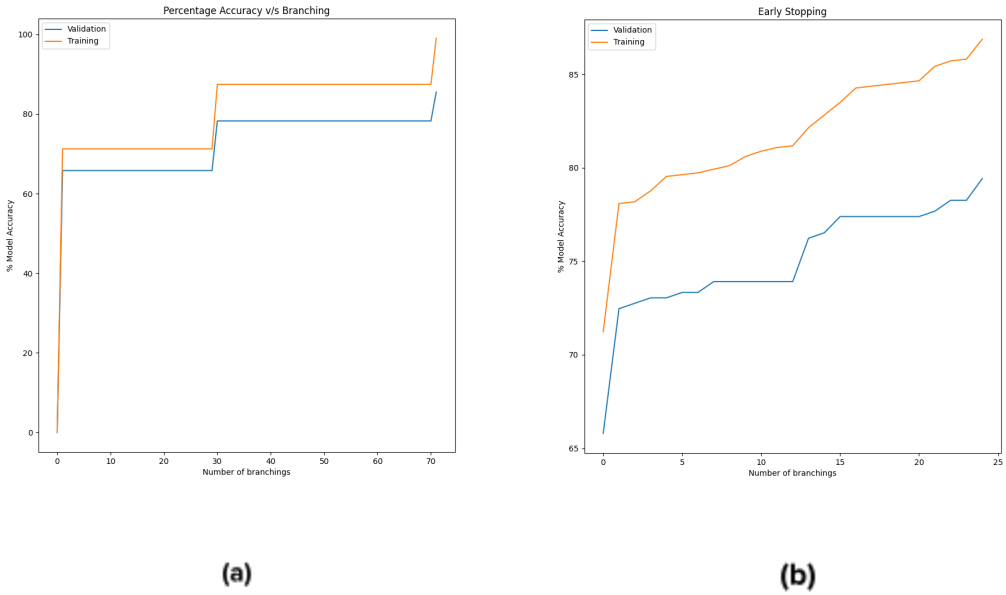


Figure 2: Percentage Accuracy v/s Branching

Percentage Accuracy for the two approaches:

Set	Early Stopping	Without Early Stopping
Train Set	86.87	*99.03*
Test Set	84.64	*89.56*

Table 1: Percent Accuracy

Remarks:

1. Early Stopping significantly reduces the number of nodes (**228 → 86**).
2. Both the accuracies of the model take a hit. This means that the **data requires a complex model** and **Early Stopping** might work **better** with **smaller threshold values where** we could notice **Overfitting**. Here, Early Stopping simply causes the model to underfit.

Experiment 3

Output for this experiment is very long. I have skipped it in the report but it is printed in a very readable format in the output of the code. I will just put a sample explaining how to read the output.

```
discriminant feature:  Safety
   |-- if value: high => discriminant feature:  Persons
      |-- if value: 2 => class: unacc
      |-- end
      |-- if value: 4 => discriminant feature: Price Buying
         |-- if value: high => class: acc
         |-- end
         |-- if value: low => discriminant feature:  Price Maintenance
            |-- if value: high => class: acc
            |-- end
            |-- if value: low => class: vgood
            |-- end
            |-- if value: med => class: good
            |-- end
            |-- if value: vhigh => class: acc
            |-- end
         |-- end
         |-- if value: med => class: good
         |-- end
         |-- if value: vhigh => class: acc
         |-- end
      |-- end
      |-- if value: more => class: acc
      |-- end
   |-- end
   |-- if value: low => class: unacc
   |-- end
   |-- if value: med => class: unacc
   |-- end
end
```

Figure 3: Sample Output

This basically translates to the following tree:

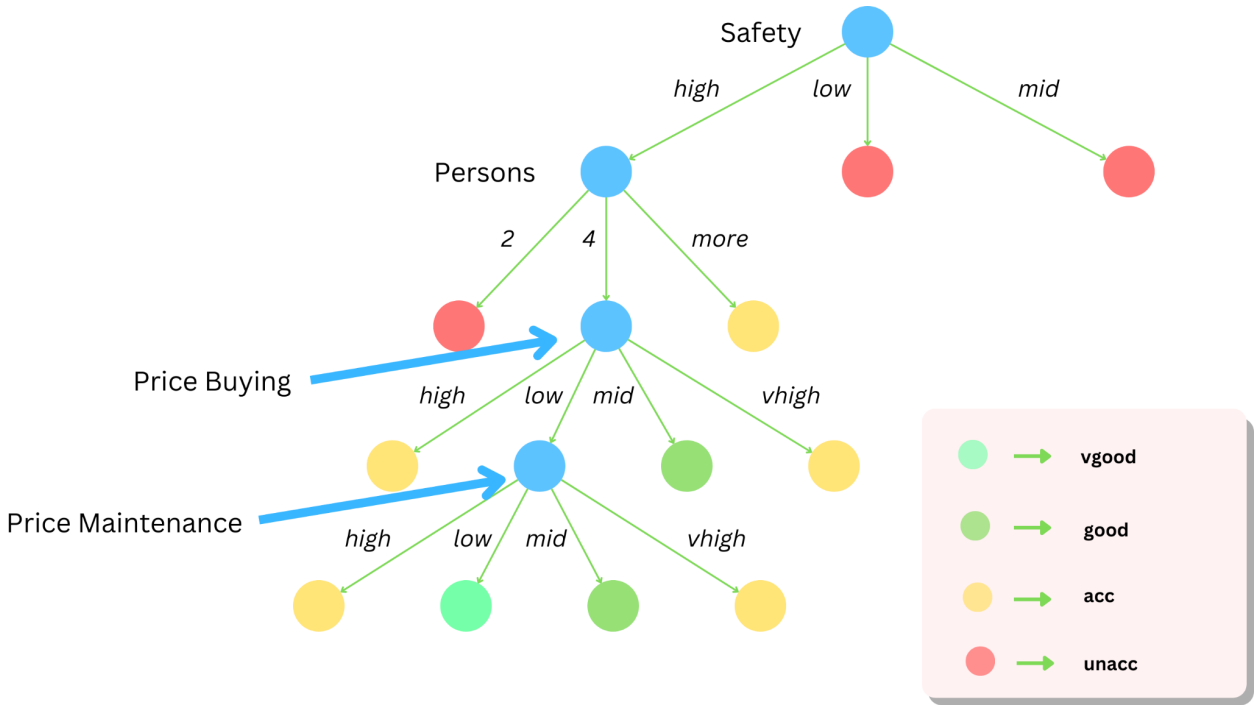


Figure 4: Decision Tree