

a. While deciding threshold, if higher threshold is showing same weighted gini index, higher threshold was chosen always as it gave simpler decision tree for this data. While choosing the attribute if weighted gini was same for two different attributes which did not occur for provided training data, trainer program chooses attribute which was seen first as next node.

b. Gini index of subset of data is used as a stopping criteria. If gini index is found to be 0, it means all records in subset belong to one output class hence decision is made. The node is marked as a leaf node and corresponding output class is stored.

c. Following is the structure of final decision tree:

```
if record['Attr2'][i] <= 7.9:
    if record['Attr1'][i] <= 5.01:
        return 1
    else:
        if record['Attr2'][i] <= 4.85:
            return 0
        else:
            return 1
else:
    return 0
```

d. Training data and testing data give 100% accuracy.

e. Trainer program generates classifier which can be used to test any same type of data in csv.

f. I learnt that, decision tree can be efficiently used for predicting classes of continuous data as well. I handled data in data-frames. Splitting, sorting it was done using pandas module. I learnt later that while splitting data-frame based on values, I could have used filters which is easier.