ML Mini-Project

# Car Insurance Approval using Decision Tree-based Learning Model

**Group 9**

**19CS10031 | 19CS3001 | 18IE10037**

**Objective: To develop a decision tree-based model for a Car Insurance Company, responsible for making insurance approval and rejection decision making both for TD3 and Gini impurity algorithm and reduced overfitting using Reduce Error pruning**

## Node:

Its a class used by decision tree building model as a basic unit using it as a tree node consisting of one input node and many output nodes

## Decision Tree:

Contains various functions for calculating accuracy, prediction, Entropy, Gini Impurity.

print_DT
Prints decision Tree in a recursive way like expand one node until it ends and then move on to the next node

Fit_Tree
Recursively calculate entropy or Gini impurity depending on the type of Decision Tree and creates a node of the best parameter until it fits the training data completely by recursively created a Node for every Attribute Split.

# Reduced error pruning

Developed a Reduced error pruning function where training data was partitioned into grow and validation set. Then built a complete tree for the grow data. For each non-leaf node in the tree Temporarily prune the tree below and replace it by majority vote. Test the accuracy of the hypothesis on the validation set, Permanently prune the node with the greatest increase in accuracy on the validation test. This process continues till accuracy on validation set increases.

# K-fold stratified cross-validation

Splitting the train set into 90 per cent and 10 per cent 10 times and then training on 90 per cent of training data and testing on the validation set(10 per cent of the training set). Taking the average of this accuracy.

Accuracy is increasing from 92.82 to 93.52 per cent after pruning the decision tree in both cases,i.e, information gain and Gini.

# Accuracy plots of the 2 models

Training the decision tree with and without pruning on 10 per cent to 100 per cent of the training set and plotting the accuracy of the decision tree by testing it on test data.

## 1.) ID3
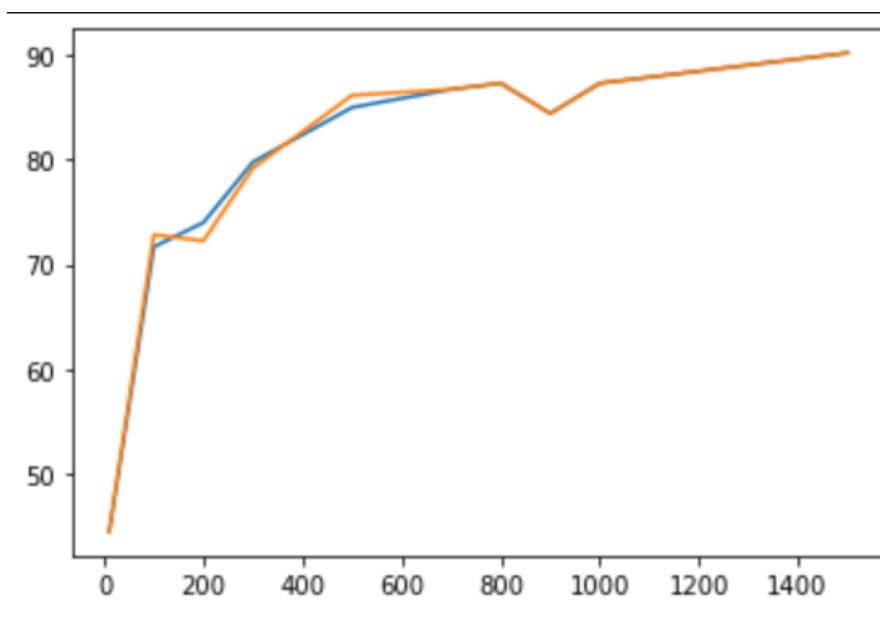
Red denotes plot of decision tree accuracy with pruning(Accuracy is calculated based on Validation set).

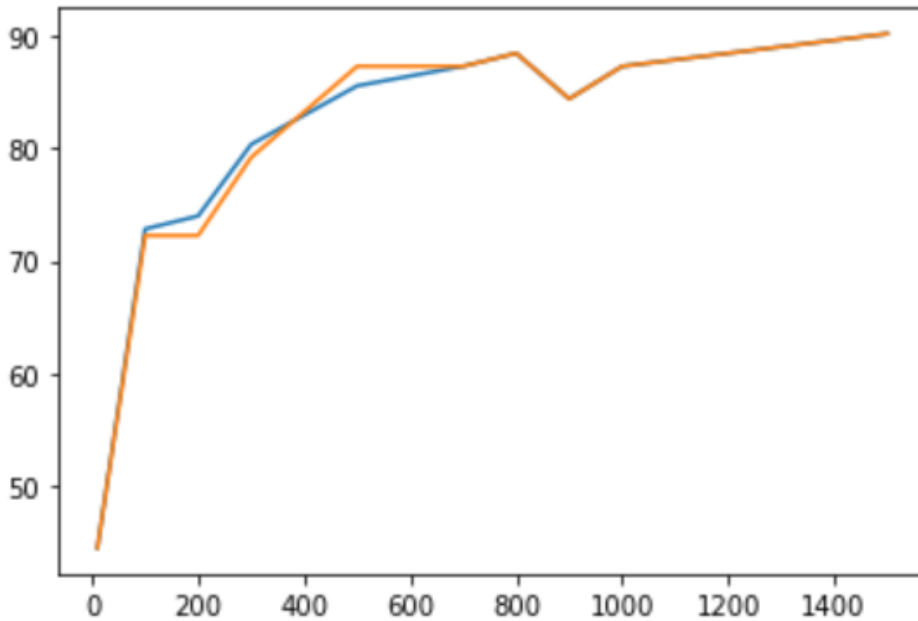Blue denotes plot of decision tree accuracy without pruning(Accuracy is calculated based on Validation set).

# 2.) Gini



Red denotes plot of decision tree accuracy with pruning(Accuracy is calculated based on Validation set).

Blue denotes plot of decision tree accuracy without pruning(Accuracy is calculated based on Validation set).

# Classification Report

```
Unpruned DT Classification Report for gini

              precision    recall  f1-score   support

         bad       0.96      0.97      0.97       112
        good       0.36      0.57      0.44         7
          ok       0.90      0.82      0.86        44
       vgood       0.89      0.80      0.84        10

    accuracy                           0.91       173
   macro avg       0.78      0.79      0.78       173
weighted avg       0.92      0.91      0.91       173
```

```
Unpruned DT classification Report for ID3

              precision    recall  f1-score   support

         bad       0.96      0.97      0.97       112
        good       0.36      0.57      0.44         7
          ok       0.90      0.82      0.86        44
       vgood       0.89      0.80      0.84        10

    accuracy                           0.91       173
   macro avg       0.78      0.79      0.78       173
weighted avg       0.92      0.91      0.91       173
```

```
pruned DT classification Report for gini

              precision    recall  f1-score   support

         bad       0.97      0.96      0.97       112
        good       0.36      0.57      0.44         7
          ok       0.89      0.75      0.81        44
       vgood       0.64      0.90      0.75        10

    accuracy                           0.89       173
   macro avg       0.72      0.80      0.74       173
weighted avg       0.91      0.89      0.90       173
```

```
pruned DT classification Report for ID3

              precision    recall  f1-score   support

         bad       0.97      0.96      0.97       112
        good       0.36      0.57      0.44         7
          ok       0.89      0.75      0.81        44
       vgood       0.64      0.90      0.75        10

    accuracy                           0.89       173
   macro avg       0.72      0.80      0.74       173
weighted avg       0.91      0.89      0.90       173
```