

## Implementation:

1). `run_decision_tree()`: It is a global function definition, called upon in line number 192 inside main clause

This function contains the following action set:

- Loading the wine dataset (Line 136-139)
- Creating cross validation sets based on value of 'K' (Line 149-160)
- Building the decision tree for each K training set (line 163-173)
- Classifying based on fitted tree and printing the accuracy of each K-validation set (Line 177-182)
- Printing the average accuracy of the K-sets of validations (line number 183-188)

2). Node Class: This class contains function called `__init__()` and assigns class variables through constructor call

3). DecisionTree: This class encapsulates all the functions to fit a tree and classify validation sets

3). `Fit()`: This functions runs first in order to call `learn()` function and saves instance of the root node in variable 'tree'

4). `Learn()`: Following are its tasks

- Initialize the predicted class, total cases and cases per class, and current node impurity (using entropy function)
- Run recursive calls to itself to initialize left and right child tree until it reaches maximum depth. Values of the child nodes are determined by calling `calc_impurity` function

5) `Calc_impurity()`: Calculate the current impurity based on entropy/gini, calculate the impurity of child, compare the two return the lowest impurity values and the corresponding split value and index where the lowest impurity was measured

6). `Classify()`: Traverse through the created nodes and return the predicted class of the last node with no childs

**Code\_name:** `Srijit_Sen_cs21mds14009_entropy_final`

**Average Accuracy attained:** 77.98%

## Improvements:

**1). Introducing Pruning:** (Instead of running the recursion till no child nodes are left, use `max_depth` hyper parameter to determine the max depth of the tree. This will pre-prune the tree and stop before reaching end node (Line 106)

**Code\_name:** `Srijit_Sen_cs21mds14009_entropy_pruning`

**Accuracy attained:** 78.39% (depth=50, K=10)

**Reason:** Pruning reduces overfitting and reduces generalization error, which improves more when cross validated is implemented

**2). Replacing Entropy with Gini Index:** Replace formula for entropy with Gini in function (Calc\_impurity)

**Code\_name:** Srijit\_Sen\_cs21mds14009\_gini\_final

**Accuracy attained:** 84.27% (depth=50, K=10)

**Reason:** Possible reason of improvement could be that some entropy values might be impossible to calculate if total class cases in a split is 0 since log of 0 is not defined. There is no such issue with gini index calculation as it is exponential of probability