❑ Decision Tree

A decision tree is a tree where each node represents a feature(attribute), each link(branch) represents a decision(rule) and each leaf represents an outcome.

Part-1

- Decision tree with K-fold cross validation using Entropy and Information gain

$$S = -\sum_{i=1}^{N} p_i \log_2 p_i,$$

This is the formulae used to find the entropy

- Information Gain is the difference of parent entropy and average entropy of childs
- Decision tree i formed is bivariate and split the data by finding the max value in each column, and then divide that by 10, numbers greater than that value will go in right subtree and numbers which are lesser than that will go in left.
- Accuracy i got using this method- 78.22

Part-2

- Decision tree with K-fold cross validation and using gini_index
- The formulae for calculating gini_index is given below

$$= 1 - \sum_{t=0}^{t=1} P_t^2$$

- The gini gain is the difference of parent gini_index and the average gini index of childs
- Decision tree i formed is bivariate and split the data by finding the max value in each column, and then divide that by 10, numbers greater than that value will go in right subtree and numbers which are lesser than that will go in left
- I tried for trivariate also, but the accuracy is decreasing using that method
- Accuracy i got using this method- 78.39