Assignment 4: Decision Tree

Steps:

(i) Load the required libraries

(ii) Import the dataset

(iii) Separate dependent and Independent variables

(iv) Label Encoding of target values

(v) Split the dataset into testing and training dataset

(vi) Define maximum depth.

(vii) Initialize the tree with Original training set as root node.

(viii) for each iteration (until max depth is not achieved).

(a) Iterate through very unused attribute of the dataset and calculate entropy and Information Gain of the attribute.

(b) Select the attribute which has the lowest Entropy or highest Information Gain.

(c) Split the tree by selected attribute to produce Child Node (Subset of Data).

(d) Continue to recur on each subset until max-depth is not achieved.

Entropy $H_S = \sum -p_i \log_2 p_i$ where p_i is the probability of features of state S.

Information Gain = Entropy(T) - Entropy(T, X)

Difference blw Entropy of parent Node from Sum of Entropies of Child Nodes.

(ix) Predict the value of y for X-test by using the decision tree obtained from model training.

(x) Compute the Confusion Matrix, Classification Report and Accuracy.