DATA MINING ASSIGNMENT 3

Name : Pavani Rangineni

CWID: A20516359

b) What is the default class label associated with each vertex?

Level 0, Vertex 1: Default class label is setosa

Level 1, Vertex 1: Default class label is setosa

Level 1, Vertex 2: Default class label is versicolor

Level 2, Vertex 1: Default class label is versicolor

Level 2, Vertex 2: Default class label is virginica

(c) Starting from the root note, what is the name of the first attribute used for a decision, and what are the split points?

Level 0, split on attribute: Petal.Length

Split points: <2.5 left subtree, >= 2.5 right subtree

Level 1, split on attribute: Petal.Width

Split points: < 1.8 left subtree, >= 1.8 right subtree

(d) Each vertex has three lines

(i) At each vertex, what do the three numbers in the middle line signify?

The middle line shows the dataset's setosa, versicolor, and virginica observations. It shows number of times Observations are connected per each class in that specific node. The distribution of our three classes Setosa, Versicolor, and Virginica are as follows respectively:.00 91.09

The above line suggests that 91% of the observations connected with that node are the leftover and versicolor which are represented by the line.09 have the label virginica.

(ii) At each vertex, what does the last line signify?

The final line shows how many observations are stored at each node, or how many observations are present across the entire dataset. Percentage is used to indicate it. As an illustration, if the value is 67%, then 67% of the dataset's observations are divided up into that node.