Data Mining Lab Assignment-08

CSE 6th semester

(Language/Platform: Python)

Objective 1:
To design a tree-based machine learning algorithm and preprocess input dataset into a compatible format for the algorithm.

Outcome:
Students will able to learn various preprocessing techniques, design fit functions to train the model   on   the   input   dataset,   and   then   apply   the   trained   model   for   predicting unseen\unknown data into appropriate class label

Load vehicle0.dat file in Data Frame using pandas.read_csv ("vehicle.csv") and prepare a user defined decision tree module. Make suitable preprocessing in the data dataset if required.

1. Compute the split point for each attribute in the dataset using the following
   strategies: a. Information Gain
      b. Gini Indexs
      c. Gain Ratio
2. Design module for creating the decision tree and its representation in graphical
   format for the following cases:
   a. Binary Tree (each node split into exactly two branches).
   b. General Tree (each node may split into more than two branches depending on
      count nominal labels corresponding attributes).

3. Design module which predicts the class label of unknown and unseen data using
   tree traversal or any other techniques.

(General structure of standard machine learning-based model)