MAGIC Gamma Telescope

Dataset: MAGIC Gamma Telescope Dataset
Source: https://archive.ics.uci.edu/ml/datasets/MAGIC+Gamma+Telescope

Dataset information:
The data are Monte-Carlo generated to simulate registration of high energy gamma particles in a ground-based atmospheric Cherenkov gamma telescope using the imaging technique. The dataset contains 19,020 records. The records have 10 feature attributes and one class attribute.

Objective:
Train decision trees based on the MAGIC Gamma Telescope Dataset to predict whether the collected patterns are caused by the primary gamma (signal) or hadron (background), and compared the performance (e.g., accuracy) of the classifiers.

Task requirements:

1. (1) Implement four (4) decision tree (DT) classifiers MIG, MGR, MVA and MGI that use the information gain, gain ratio, variance and Gini index as split criteria, respectively.
   Note that you can use either binary-split or multiple-split; you can also use any reasonable pre- processing method and any reasonable early stopping criteria (e.g. pre-pruned parameters) but need to explain your reasons.
   It is recommended that each classifier includes a tree induction function, a classification function and any other auxiliary functions.

2. (2) Implement an ensembled DT classifier M^* whose predictions are based on a voting function between the three classifiers MIG, MGR and MVA.

3. (3) Implement a 10-fold cross-validation to evaluate M^* and MGI, and implement Student's t-test to determine the statistical significance of the error rate difference between the two classifiers.
   Note that you should set a suitable significance level.

4. (4) Implement a post-pruning method for MGI by using (approximately) 1/3 dataset for training, 1/3 for pruning and 1/3 for evaluation. Report the TP rates and FP rates for both the unpruned MGI and the pruned MGI.
   Note that you should generate the three subsets randomly.