

Assignment #2 Decision Trees

Problem# 1(5 pts): Consider the following dataset, where we want to predict if a student will get an A in the course. Given the five attributes on the left, we want to predict if the student got an A in the course.

Early registration	Finished homework II	Senior	Likes Coffee	Liked The Last homework	A
1	1	0	0	1	1
1	1	1	0	1	1
0	0	1	0	0	0
0	1	1	0	1	0
0	1	1	0	0	1
0	0	1	1	1	1
1	0	0	0	1	0
0	1	0	1	1	1
0	0	1	0	1	1
1	0	0	0	0	0
1	1	1	0	0	1
0	1	1	1	1	0
0	0	0	0	1	0
1	0	0	1	0	1

[1] Create 2 decision trees for this dataset. For the first, only go to depth 1. For the second go to depth 2. For all trees, use the ID3 entropy algorithm from class. For each node of the tree, show the decision, the number of positive and negative examples and show the entropy at that node.

[2] Recommend another type of trees than ID3 that would build a less deep tree than ID3, Assume that you are building a complete ID3 tree. Justify your choice.

**Problem# 2(5 pts+ 2 Bonus pts for second part):**

[1] Implement the ID3 decision tree learning algorithm described in class (without post-pruning). You may use any programming language you like. Your program can assume that all features/ attributes take only discrete values (no real-valued attributes), that the data contains no missing attributes, and that legal values for each attribute are known in advance. Use Kaggle Cardiovascular disease Dataset, you can download the dataset through this [link](#), please use the first 90% and 10% as training and testing dataset respectively. Use the accuracy as the performance parameter to test the performance of the learnt tree.

[2](Bonus) Implement bagging Ensemble Learning. Use TrainingData.txt for building the tree, while TestData.txt to test the performance of the learnt tree.

**General Instructions**

- 1- This is a individual based assignment.
- 2- The source code as well as the report describing your functions and output should be submitted through Google Classroom.
- 3- The due date for the submission of this phase is Tuesday, April, 09, 2020 at 12:00 am.
- 4- Please Review the definition of cheating in the first presentation.

Best Regards,

*Inas A. Yassine*