

CS – 531: MACHINE LEARNING (Assignment# 1)

Date: 12th September, 2018
Due Date: 19th September, 2018

Max. Marks: 100

In this assignment you will train a decision tree to predict votes of US Congressmen based on their political party and on other votes they have made in the past. This is based on a [data set](#) from the UCI data repository. The data contains a total of 435 examples, one for each member of the US House of Representatives. Each example is described in terms of 17 attributes (all boolean) including "party" (democrat or republican) as a class attribute, and 16 other features, such as "education-spending" (with values "y", "n"). These are in comma-separated-file format, with the first line in the file listing the attribute names (separated by commas) and each remaining line in the file giving the values of these attributes for a single congress person.

You must train a decision tree to predict "party" (democrat or republican) based on the 16 votes of the congressperson.

Implement the ID3 decision tree learning algorithm described in class. To make life easy, you may use any programming language / libraries you like. You should also take care of missing attributes.

Your submission must include following things:

1. Exploratory data analysis on the dataset
2. Parameter tuning of your decision tree
3. Calculate Accuracy and F1-Score and mention which evaluation metric is best for the given problem and why
4. Mention size of your original tree and your pruned tree (in case you've done pruning)
5. Which of these trees (pruned or unpruned) would you recommend using to classify future data (justify your answer in terms of your actual observed accuracies)

For extra credits, you can try implementing well tuned Random Forest for better accuracy / f1-score. Note that, its only optional and not compulsory.

Submission:

- The file **AssignmentNo_YourName_YourRollNo.zip**. **Report and Code** should be submitted on Google Class Room.
- Submission other than Google Class Room should not be considered/accepted.

Note that if your implementation matches with your classmates or with any code available on Kaggle or Github, you will get a straight ZERO.

BEST OF LUCK !!