

# Introduction to ML, Fall 2022

## Term Project#1 Classification Tree

Due Date: 11:50PM Jan 9, 2023

### Objective of Project:

Implement a classification tree learning algorithm to make predictions. By “implement” it means you *write your own code from scratch*. You are not allowed to include anyone else’s code (partial or complete) in your project. Any form of plagiarism will incur severe penalty. As a CS major (or a non-CS major but taking this CS course), you are expected to build your own version of Classification Tree right from the beginning rather than simply know how to run classification tree in Scikit Learn or WEKA because I sincerely believe you are much better than that.

### Programming Language:

You can *only* use one of the programming languages, C/C++, Java or Python.

### Evaluation:

Your classification tree will be tested on several real-world data sets with mixed types of features (categorical and numerical) and missing values. To deal with missing values, you can simply preprocess the data by filling in means or modes accordingly as explained in class, or for *extra credits*, you can implement more advanced missing data handlers if you wish.

TAs will post some sample data sets on e3 later for you to play with, and you can evaluate your code.

For the real evaluation, TAs will post the final training and testing data sets on e3 from which you build your classification trees, and test them on the testing data.

### What to turn in:

- (a) A copy of your source code (be correct, i.e. it can be compiled to run),
- (b) A document of your code (be clear, e.g. how to deal with missing data and missed types of features, pruning, etc.)
- (c) A snapshot of your input and output (i.e. user interface).
- (d) A list of your predictions of the testing data.

### Grading policy:

- (a) You will get reasonable points for your project if you turn in what you are supposed to even though your AI program is not among the top-ranking programs in class.
- (b) The higher ranking, the more points, of course.
- (c) And yes, to be fair, an incomplete project will receive a low grade.