COMP338: Artificial Intelligence
Fall 2024/2025
Project Machine Learning: Decision Tree Learning
Deadline: 10 January 2025 (before midnight)

You will build a predictive model based on Decision Tree algorithm. This is a team-work of groups (**each group of two students**). You should submit the code of the work as well a report containing a proper cover page and the details of your work.

The dataset will be used to construct the decision tree is provided under this link: https://archive.ics.uci.edu/dataset/19/car+evaluation . Full description of the dataset can be found in that link.

In this work you are required to build decision tree classifiers using Weka library. You can use the weka tool using its interface or you can integrate it within your java code as explained in this link: https://waikato.github.io/weka-wiki/use_weka_in_your_java_code/.

You should provide a report of no more than 5 pages containing the following:

1. Print out the distribution of the target class (i.e., the percentage of each of the outcomes of the target class).
2. Split the dataset into 70% training and 30% test data.
3. Use the Decision Tree algorithm (i.e., C4.5 or C5.0 or their alternative open source implementation) to train a model called M1.
4. Use the 30% of the test data to test the accuracy of the generated model
5. Print the accuracy and the F1-Score of the test set.
6. Create another model (called M2) on a new data split of 50% training and 50% test and show the accuracy and F1-Score of this new model and compare it with the accuracy of model M1. Explain any differences in the results between the two models.
7. Generate (plot) the generated decision tree of model M1.
8. Generate (plot) the generated decision tree of model M2.

Make sure to shuffle the data before you split into training/test sets.

As for points (1, 4, 5, 6, 7, 8), have them written properly and explained in the report. The report should contain your explanation of the work, details of your implementation and the results.

You have to turn in a Word document containing the information required to as specified above. Please response to the thread where this assignment was sent with your solution. This is a group project and you should specify the file name as St1Name_St2Name_AIPML_YourSecNo.zip