Introduction to Machine Learning Program assignment #1

Problem: (70%)

For this assignment, you need to implement **ID3 algorithm** to construct a decision tree with C, C++, Java or python2/3, and use K-fold cross validation (K=5) to validate classification performance by outputting precision and recall for each class and total accuracy. **You CANNOT use packages to do the jobs for you.**

Note that, the instances of data must be randomly shuffled before constructing decision trees. The accuracy, precision and recall must be floating numbers within 0 and 1 and be arranged with the following format.

```
[Total accuracy]
[Precision of class 0] [Recall of class 0]
[Precision of class 1] [Recall of class 1]
[Precision of class 2] [Recall of class 2]
```

You should upload a single **[student-id].ZIP** file which contains a 'run.sh' shell script, source files, data and a report. The 'run.sh' should compile the source code (for C/C++ and java) and execute the program which output the results by a single **'./run.sh'** command.

Bonus: (20%)

Implement Random Forest algorithm and make a 'RF.sh' shell script to output the result with the same format.

Report: (10%)

The report should include the results, environment, using library and language, explain of your code and how to use it.

Accuracy, Precision and Recall: (20%)

We will test your source code and score base on your rank of following metrics:

$$1.5\times Accuracy + \sum_{i=0}^{2}(Precision_i + Recall_i)$$

Environment:

Your program will be executed on the following environment:

- Ubuntu 16.04.3 LTS
- gcc 5.4.0
- openjdk 1.8.0_131
- python 2.7.12
- python 3.5.2

Data:

https://archive.ics.uci.edu/ml/datasets/Iris

Including 150 number of instances with 4 attributes.

Attribute Information:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
    -- Iris Setosa
    -- Iris Versicolour
    -- Iris Virginica