How to run the code

Libraries User:

Pandas, Numpy, Scikit Learn

Open the code D_tree.py file in PyCharm. Make sure that all the training, validation and test datasets are there in the same folder location.

The general command goes like this –

python **D_tree.py** -**algorithm_number** 1 -**train_data** train_c1000_d100.csv -**valid_data** valid_c1000_d100.csv -**test_data** test_c1000_d100.csv

D_tree.py is the name of the file where the code is written.

algorithm_number denotes which algorithm we want to run. Each number represents a different algorithm as shown below -

1 – **Naive Decision tree learner with Entropy as the impurity heuristic**

2 – **Naive Decision tree learner with Variance as the impurity heuristic.**

3 – **Decision tree learner with Entropy as the impurity heuristic and reduced error pruning**

4 – **Decision tree learner with Variance as the impurity heuristic and reduced error pruning**

5 – **Decision tree learner with Entropy as the impurity heuristic and depth-based pruning**

6 – **Decision tree learner with Variance as the impurity heuristic and depth-based pruning**

7 – **Random Forest**

train_data – Add the training dataset name to this parameter.

valid_data - Add the validation dataset name to this parameter.

test_data - Add the test dataset name to this parameter.

Sample command line input is shown below –

This is for the entire dataset involving 1000 clauses and 100 examples. We could use the same by modifying the files and run the same to generate outputs.

python D_tree.py -algorithm_number 1 -train_data train_c1000_d100.csv -valid_data valid_c1000_d100.csv -test_data test_c1000_d100.csv

python D_tree.py -algorithm_number 2 -train_data train_c1000_d100.csv -valid_data valid_c1000_d100.csv -test_data test_c1000_d100.csv

python D_tree.py -algorithm_number 3 -train_data train_c1000_d100.csv -valid_data valid_c1000_d100.csv -test_data test_c1000_d100.csv

python D_tree.py -algorithm_number 4 -train_data train_c1000_d100.csv -valid_data valid_c1000_d100.csv -test_data test_c1000_d100.csv

python D_tree.py -algorithm_number 5 -train_data train_c1000_d100.csv -valid_data valid_c1000_d100.csv -test_data test_c1000_d100.csv

python D_tree.py -algorithm_number 6 -train_data train_c1000_d100.csv -valid_data valid_c1000_d100.csv -test_data test_c1000_d100.csv

python D_tree.py -algorithm_number 7 -train_data train_c1000_d100.csv -valid_data valid_c1000_d100.csv -test_data test_c1000_d100.csv