1. Unzip the folder and open the cmd in the folder level as main.py

2. The folder should contain main.py, tree folder (where the python code is ) and the data folder(datasets are inside this) at the same level.

3. Use python main.py 10 10 data/training\_set.csv data/validation\_set.csv data/test\_set.csv yes to display the stats of the tree before and after pruning for both the heuristics.

4. Yes value at the end of command prints the tree and no does not print the tree.

Accuracy of the decision tree using information gain before pruning: 0.7515

Accuracy of the decision tree using variance impurity before pruning: 0.6785

|  |  |  |  |
| --- | --- | --- | --- |
| L value | K value | Tree 1 pruned accuracy | Tree 2 pruned accuracy |
| 10 | 10 | 0.755 | 0.687 |
| 10 | 15 | 0.755 | 0.68 |
| 15 | 15 | 0.755 | 0.6845 |
| 25 | 25 | 0.7545 | 0.687 |
| 5 | 25 | 0.7545 | 0.6785 |
| 20 | 20 | 0.7545 | 0.6865 |
| 5 | 25 | 0.755 | 0.6835 |
| 12 | 22 | 0.755 | 0.6865 |
| 0 | 17 | 0.7545 | 0.6825 |
| 30 | 30 | 0.7548 | 0.6825 |

Output for 10 Possible values of L and K

1. L = 10 and K = 10

C:\Python27\mllib-master -changes\mllib-master>python main.py 10 10 data/training\_set.csv data/validation\_set.csv data/test\_set.csv no

Statistics of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1503

NEGATIVE CLASS: 497

OverAll Accuracy: 0.7515

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1510

NEGATIVE CLASS: 490

OverAll Accuracy: 0.755

---------------------------------------------------

Statistics of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1357

NEGATIVE CLASS: 643

OverAll Accuracy: 0.6785

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1374

NEGATIVE CLASS: 626

OverAll Accuracy: 0.687

---------------------------------------------------

2. L = 10 and K = 15

C:\Python27\mllib-master -changes\mllib-master>python main.py 10 10 data/training\_set.csv data/validation\_set.csv data/test\_set.csv no

Statistics of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1503

NEGATIVE CLASS: 497

OverAll Accuracy: 0.7515

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1510

NEGATIVE CLASS: 490

OverAll Accuracy: 0.755

---------------------------------------------------

Statistics of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1357

NEGATIVE CLASS: 643

OverAll Accuracy: 0.6785

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1374

NEGATIVE CLASS: 626

OverAll Accuracy: 0.687

---------------------------------------------------

3. L = 15 and K =15

C:\Python27\mllib-master -changes\mllib-master>python main.py 15 15 data/training\_set.csv data/validation\_set.csv data/test\_set.csv no

Statistics of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1503

NEGATIVE CLASS: 497

OverAll Accuracy: 0.7515

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1510

NEGATIVE CLASS: 490

OverAll Accuracy: 0.755

---------------------------------------------------

Statistics of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1357

NEGATIVE CLASS: 643

OverAll Accuracy: 0.6785

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1369

NEGATIVE CLASS: 631

OverAll Accuracy: 0.6845

---------------------------------------------------

4. L = 25 and K = 25

C:\Python27\mllib-master -changes\mllib-master>python main.py 25 25 data/training\_set.csv data/validation\_set.csv data/test\_set.csv no

Statistics of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1503

NEGATIVE CLASS: 497

OverAll Accuracy: 0.7515

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1509

NEGATIVE CLASS: 491

OverAll Accuracy: 0.7545

---------------------------------------------------

Statistics of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1357

NEGATIVE CLASS: 643

OverAll Accuracy: 0.6785

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1374

NEGATIVE CLASS: 626

OverAll Accuracy: 0.687

---------------------------------------------------

5. L =5 and K =25

C:\Python27\mllib-master -changes\mllib-master>python main.py 5 25 data/training\_set.csv data/validation\_set.csv data/test\_set.csv no

Statistics of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1503

NEGATIVE CLASS: 497

OverAll Accuracy: 0.7515

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1509

NEGATIVE CLASS: 491

OverAll Accuracy: 0.7545

---------------------------------------------------

Statistics of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1357

NEGATIVE CLASS: 643

OverAll Accuracy: 0.6785

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1374

NEGATIVE CLASS: 626

OverAll Accuracy: 0.687

---------------------------------------------------

6. L = 20 and K = 20

C:\Python27\mllib-master -changes\mllib-master>python main.py 20 20 data/training\_set.csv data/validation\_set.csv data/test\_set.csv no

Statistics of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1503

NEGATIVE CLASS: 497

OverAll Accuracy: 0.7515

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1509

NEGATIVE CLASS: 491

OverAll Accuracy: 0.7545

---------------------------------------------------

Statistics of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1357

NEGATIVE CLASS: 643

OverAll Accuracy: 0.6785

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1373

NEGATIVE CLASS: 627

OverAll Accuracy: 0.6865

---------------------------------------------------

7. L = 5 and K = 25

C:\Python27\mllib-master -changes\mllib-master>python main.py 5 25 data/training\_set.csv data/validation\_set.csv data/test\_set.csv no

Statistics of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1503

NEGATIVE CLASS: 497

OverAll Accuracy: 0.7515

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1510

NEGATIVE CLASS: 490

OverAll Accuracy: 0.755

---------------------------------------------------

Statistics of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1357

NEGATIVE CLASS: 643

OverAll Accuracy: 0.6785

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1367

NEGATIVE CLASS: 633

OverAll Accuracy: 0.6835

---------------------------------------------------

8. L = 12 and K = 22

C:\Python27\mllib-master -changes\mllib-master>python main.py 12 22 data/training\_set.csv data/validation\_set.csv data/test\_set.csv no

Statistics of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1503

NEGATIVE CLASS: 497

OverAll Accuracy: 0.7515

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1501

NEGATIVE CLASS: 499

OverAll Accuracy: 0.7505

---------------------------------------------------

Statistics of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1357

NEGATIVE CLASS: 643

OverAll Accuracy: 0.6785

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1374

NEGATIVE CLASS: 626

OverAll Accuracy: 0.687

---------------------------------------------------

9. L = 17 and K = 17

C:\Python27\mllib-master -changes\mllib-master>python main.py 17 17 data/training\_set.csv data/validation\_set.csv data/test\_set.csv no

Statistics of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1503

NEGATIVE CLASS: 497

OverAll Accuracy: 0.7515

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1509

NEGATIVE CLASS: 491

OverAll Accuracy: 0.7545

---------------------------------------------------

Statistics of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1357

NEGATIVE CLASS: 643

OverAll Accuracy: 0.6785

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1365

NEGATIVE CLASS: 635

OverAll Accuracy: 0.6825

---------------------------------------------------

10. L = 30 and K = 30

C:\Python27\mllib-master -changes\mllib-master>python main.py 17 17 data/training\_set.csv data/validation\_set.csv data/test\_set.csv no

Statistics of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1503

NEGATIVE CLASS: 497

OverAll Accuracy: 0.7515

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using information gain to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1509

NEGATIVE CLASS: 491

OverAll Accuracy: 0.7545

---------------------------------------------------

Statistics of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1357

NEGATIVE CLASS: 643

OverAll Accuracy: 0.6785

---------------------------------------------------

Statistics AFTER PRUNING of decision tree using variance impurity to calculate entropy

--------------------------------------------------

TESTING RESULTS on test dataset:data/test\_set.csv

POSITIVE CLASS : 1365

NEGATIVE CLASS: 635

OverAll Accuracy: 0.6825

---------------------------------------------------