Name: Naveenraj Palanisamy

NetId: NXP154130

Machine Learning Assignment -1 (Decision Tree)

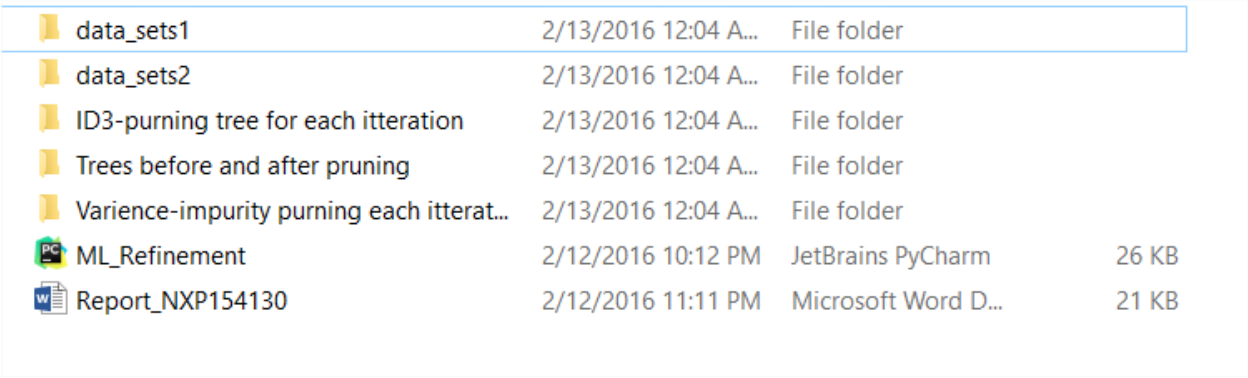
(Note: Runs in python 3.5)

(Note: File Path should be full path)

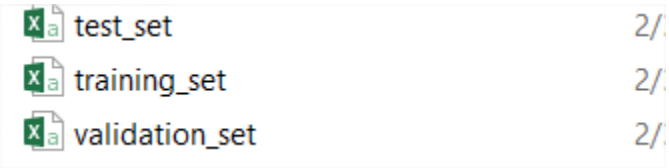
Zip file name: Naveenraj_NXP154130_ML_Assignment_1.zip

Folder Name: Naveenraj_NXP154130_ML_Assignment_1

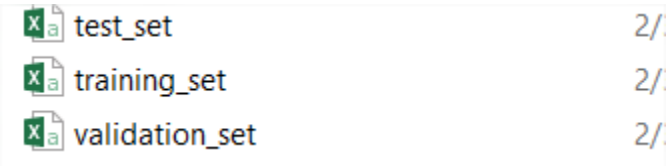
Folder Structure:



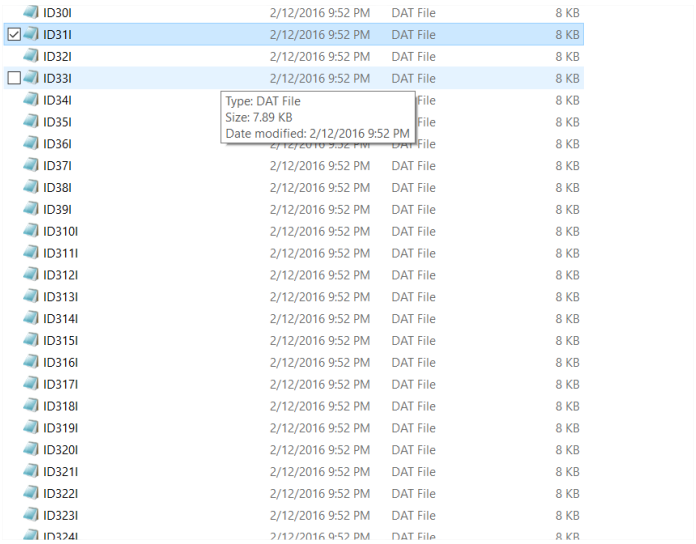
Data_set1-> Contains first set of given data



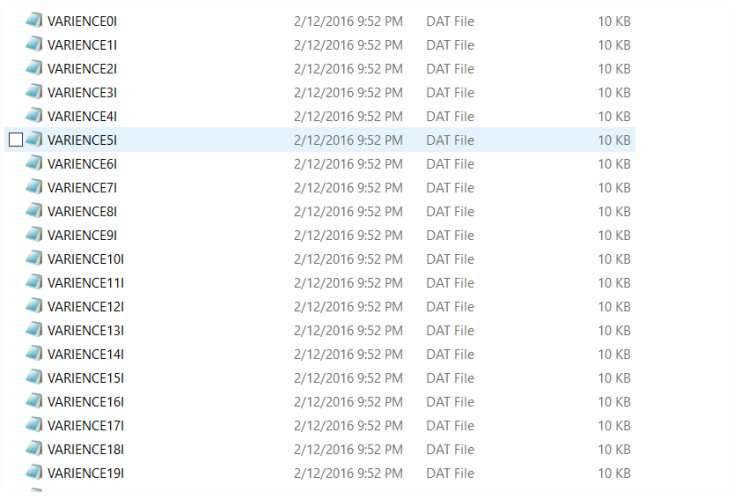
Data_set2->Contains second set of given data



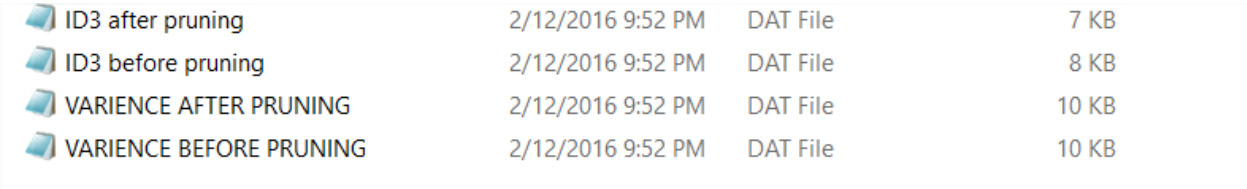
ID3-purning tree for each iteration-> Contains all tree got during pruning for ID3 for each L (L is given as 50)



Varience-impurity purning each iteration -> Contains all tree got during pruning for Impurity variance for each L (L is given as 50)



Trees before and after pruning-> Contains tree for ID3 and Impurity Variance (Both before and after pruning)



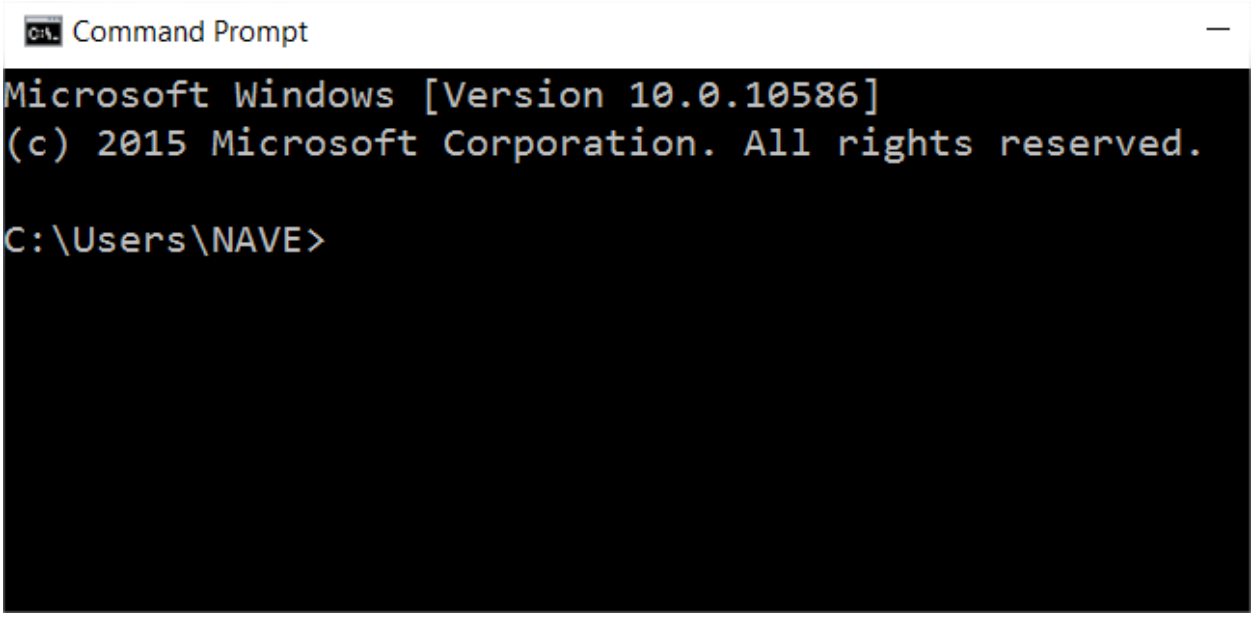
ML_Refinement.py -> python program that need to run.

Running the program steps: ->

1) **Go to ‘command prompt’**
2) **Go to folder where python is installed.**
3) **Given command as (python.exe ‘full path to-> ML_Refinement.py’ filelocatio1 filelocation2 filelocation3 K L YES/NO)**
4) **Giving YES to last command line argument will display the tree in command prompt and also prints Trees before and after in the current path.**
5) **Giving NO will only prints Trees before and after in the current path.**

Sample Run:

1) **Start -> cmd**

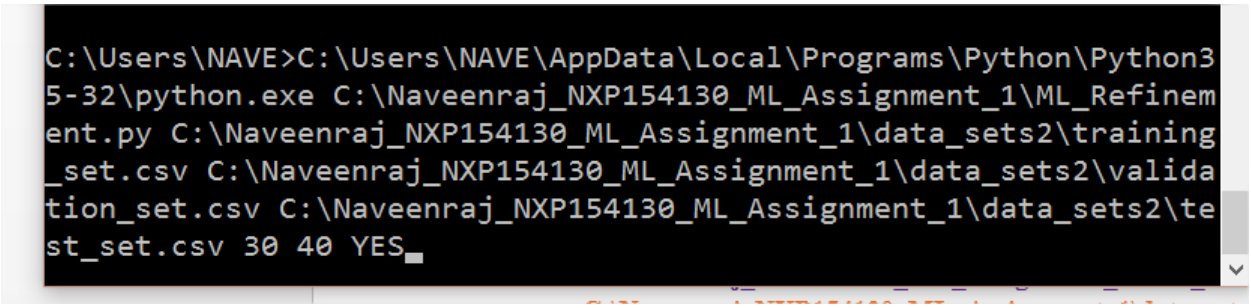


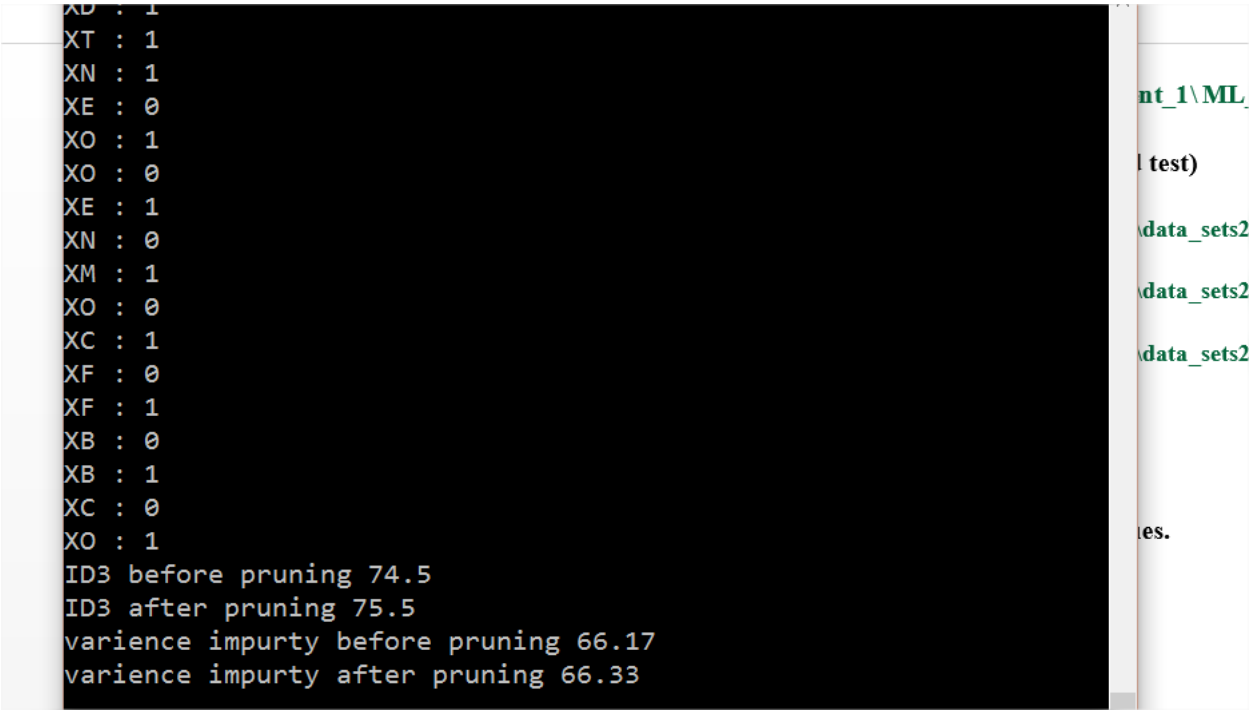
2) **Copy the path where python 3.5 is installed.**
   For me path is:
   C:\Users\NAVE\AppData\Local\Programs\Python\Python35-32\python.exe
3) **Copy the path where folder is downloaded.**
   For me it is:
   C:\Naveenraj_NXP154130_ML_Assignment_1\ ML_Refinement.py
4) **Copy the file path for all the files in the data set**
   For me it is: (in the order of training, validation and test)
   First ->
   C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\ training_set.csv
   Second->
   C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\ validation_set.csv
   Third ->
   C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\ test_set.csv
5) **Decide values for L and K (I gave as 20 4)**

6) **Decide need to print tree or not (I gave YES)**

Now run the program with all the above mentioned values.

C:\Users\NAVE\AppData\Local\Programs\Python\Python35-32\python.exe
C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinement.py
C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training_set.csv
C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\validation_set.csv
C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\test_set.csv 20 4 YES





Different try and results are given below:

```
C:\Users\NAVE>C:\Users\NAVE\AppData\Local\Programs\Python\Python3
5-32\python.exe C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinem
ent.py C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training
_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\valida
tion_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\te
st_set.csv 20 4 NO
ID3 before pruning 74.5
ID3 after pruning 74.83
varience impurty before pruning 66.17
varience impurty after pruning 66.33
```

```
C:\Users\NAVE>C:\Users\NAVE\AppData\Local\Programs\Python\Python3
5-32\python.exe C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinem
ent.py C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training
_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\valida
tion_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\te
st_set.csv 10 14 NO
ID3 before pruning 74.5
ID3 after pruning 74.5
varience impurty before pruning 66.17
varience impurty after pruning 66.17
```

```
C:\Users\NAVE>C:\Users\NAVE\AppData\Local\Programs\Python\Python3
5-32\python.exe C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinem
ent.py C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training
_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\valida
tion_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\te
st_set.csv 19 9 NO
ID3 before pruning 74.5
ID3 after pruning 75.33
varience impurty before pruning 66.17
varience impurty after pruning 66.5
```

```
C:\Users\NAVE>C:\Users\NAVE\AppData\Local\Programs\Python\Python3
5-32\python.exe C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinem
ent.py C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training
_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\valida
tion_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\te
st_set.csv 190 9 NO
ID3 before pruning 74.5
ID3 after pruning 75.83
varience impurty before pruning 66.17
varience impurty after pruning 67.33
```

```
C:\Users\NAVE>C:\Users\NAVE\AppData\Local\Programs\Python\Python3
5-32\python.exe C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinem
ent.py C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training
_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\valida
tion_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\te
st_set.csv 190 190 NO
ID3 before pruning 74.5
ID3 after pruning 74.83
varience impurty before pruning 66.17
varience impurty after pruning 68.83
```

ID3 tree after pruning will look like (for one random condition):

XI: 1

| XK: 0

| | XC: 0

| | | XS: 0

| | | | XG: 1

| | | | | XO: 1

| | | | | | XL: 0

| | | | | | | XH: 0 1 (goes on)(…… to file the full file see in the trees before and after folder)