**1) Decision Tree 10 fold**

Folder : /org\_10\_fold

File: decision.ipynb

Tree Visualization Output folder

/org\_10\_fold/

In the first block of the notebook run

Comment or Uncomment lines to run respective dataset

# FOR DATASET 1

# String\_columns=[]

# filename="project3\_dataset1"

# FOR DATASET 2

# filename="project3\_dataset2"

# String\_columns=[4]

# FOR DEMO DATASET 4

filename="../project3\_dataset4"

String\_columns=[0,1,2,3]

k\_fold\_remove=1 #k\_fold=1 and k\_fold\_remove=1 to remove k fold it will run on training data as test data

k\_fold=1

**String\_columns** is the column number having categorical values

Change the value of **k\_fold** for the number of k folds

Pruning

There are two pruning parameters and a flag:

**Prune\_multiple** and **prune\_minimum**

And **prune\_multiple\_flag is** to enable or disable **prune\_multiple**

Results

Under the RESULTS Heading

We can see

Mean as well as for each fold iteration

accuracy, precision, recall, f\_measure

TREE Visualization

Proceed down in notebook

## Search for heading “TREE IMAGE CREATION” when it ends there will be pdf file of name demo.gv.pdf containing the visualization of the tree

**Note: To run the tree visualization first install graphviz:**

* **Run command “pip install graphviz”.**
* **Download graphviz.zip from-** [**http://www.graphviz.org/Download\_windows.php**](http://www.graphviz.org/Download_windows.php)
* **Unzip the file**
* **Give the path of unzipped folder to environment variables.**

**2) Decision Tree 10 fold with random forest**

Folder : /random\_10fold

File: decision.ipynb

Similarly Comment or Uncomment lines to run respective dataset

# FOR DATASET 1

# String\_columns=[]

# filename="project3\_dataset1"

# col\_random = 0

#FOR DATASET 2

String\_columns=[4]

filename="project3\_dataset2"

col\_random = 0

k\_fold=10

trees = 6

Trees: To control number of random forest

**3) Decision Tree 10 fold with Boosting**

Folder : /boost\_10\_fold\_new

File: decision.ipynb

Similarly Comment or Uncomment lines to run respective dataset

# FOR DATASET 1

# String\_columns=[]

# filename="project3\_dataset1"

# col\_random = 5

# FOR DATASET 2

String\_columns=[4]

filename="project3\_dataset2"

col\_random = 2

# filename="../project3\_dataset4"

# String\_columns=[0,1,2,3]

# col\_random = 2

k\_fold=10

trees = 3

Change the value of trees to control max week classifiers

**4) K nearest neighbor**

To run K nearest neighbor run the file called knn.py. It will give the result for dataset1.

To get the result for dataset2 comment the line:

filename="project3\_dataset1"

And uncomment the line:

#filename="project3\_dataset2"

To run the demo part change the value of “part” variable to be “demo”.

**5) Naive Bayes Classifier**

1. file to run- **naive1.ipynb**
2. Open the ipython notebook file using jupyter/canopy etc.
3. Run all cells.
4. Parameters such as filename, number of folds etc can be adjusted below the “SCRIPT STARTS HERE” heading in the python notebook.
5. Final result is found at the bottom under the heading “FINAL RESULT”

For demo:

1. run file **demo.ipynb**