ReadMe file: Data Mining Assignment 2

Naïve Bayes and Decision Tree

Author: Harsh Gupta (gupta.749)

Date: 10/09/2016

* The code is compatible with Python 2.7
* You can run the code using the command
* **./python –W ignore NB\_DT.py** followed by a parameter depending on the number of samples you want to run the code on. There are 8 samples of 2720 documents each.
* **‘/home/8/athmakuri.1/athmakuri.1\_DM\_Lab\_2/NB\_DT’**

All the relevant output files are in ‘.csv’ and ‘.txt’ format are in the folder above. They can be viewed using the ‘cat’ command or be opened through any other application.

* The parameter number will decide the number of samples you want to check the code. Running the code for more than 1 sample will take considerable amount of time but can be done if required.
* We use –W ignore in order to ignore any warning which will arise due to assigning precision, recall values to 0 when there is no sample.
* The parameters is only for the number of samples you want to run the code on
* The parameter can be given from 1 to 8.
* Example:
* **Python –W ignore ./NB\_DT.py 1**

The above command will execute only on 1 sample out of 8 samples in sparse bag of words file and TF-IDF matrix.

* **python –W ignore ./NB\_DT.py 2**

The above command will execute only on 2 samples out of 8 samples in sparse bag of words file and TF-IDF matrix.

**Disclaimer:**

* If you don’t give any parameter or give more than 1 parameter, it won’t run anything and give an error
* When the code is running for the first time it might generate an alarm of setting up precision, recall, f score and rank to be ill defined. It is essentially because when we don’t have any input test data, those metrics are assigned to 0.0 which is technically correct but the program raises a warning. Hence the run command is having –W ignore , to ignore the warning.