**Email Spam Filter**

In this project, we are implementing Naïve Bayes’ classifier to detect an email as a spam or a ham i.e., non-spam. For this, we teach our filter hat a spam email looks like and then evaluate our model using factors like accuracy and precision.

**Github Repo**

If you have access to ccs github, click the below link to access the project:

https://github.ccs.neu.edu/shahnisarg/CS5100

**Getting Started**

Download the zipped file. Unzip the contents in a particular location. Inside the CS5100 folder, there is a mail-spam-filter folder that consists of our data set and source code.

Resources folder consists of our training and test data set. bareTraining consists of training set and bareTesting consists of test set.

Source code is contained within 3 python files: main.py, test.py and train.py.

Running main.py will automatically train the data and test it.

**Prerequisites**

Need to have any python running IDE. PyCharm or Sublime or any IDE capable of running python code.

**Version**

We have used Python version 2.7 for our project.

**Running**

*Through IDE*

Since there are two models implemented, bag-of-words and tf-idf, while running the main file, click on “Edit Configurations”. Under the configuration tabs:

* To run the bag-of-words model, type in parameter as “bow”.
* To run tf-idf model, type in parameters as “tfidf”.

After typing in parameter, run the main file.

*Through Command-Line*

To run the project through command line. Go to the project directory through terminal.

Make sure you are in mail-spam-filter directory while running the code.

* To run the bag of model, run:
  + python main.py bow
* To run tf-idf model, run:
  + python main.py tfidf

The results will be displayed according to the model chosen.

**Authors**

The project is designed by Nisarg Hareshbhai Shah and Sugandha Kher.