ReadMe
Information Retrieval
Assignment 3

Submitted By:

Himanshu Aggarwal

MT17015

Repository of this Assignment 2 contains two python code file namely, **generator.py** and **testing.py**, and few JSON files, containing processed data for the code.

generator.py:

This code generates tokens from the given dataset, preprocess it, compute tf-idf values for each term and document, construct an inverted index using default dictionary. Trains the naive bayes model on the training data.

testing.py

This is the main file for running and evaluating this assignment. It tests the model on various splits of data and computes accuracy and confusion matrix.