# IEMS5723 Assignment 2

## The Problem

This assignment requires the use of Natural Language Tool Kit (nltk). By training a classifier (Naïve Bayes Classifier) with business.txt and health.txt, classify the news category (business or health) of words in test.txt. The feature extraction function should be designed by considering multiple features (e.g. Part-of-Speech).

## The Input

The input would be three text files business.txt, health.txt and test.txt.

## The Output

Identify the category of each word in test files and output the words and respective category in dictionary order to file “output.txt”.

## Submission

A runnable Python program file with the use of nltk should be submitted. Additionally, discussion on rationale in defining the features is required.

## Sample Input

See attached.

## Sample Output

bank: business

biology: health

brain: health

child: health

deliver: business

food: health

profit: business

treatment: health

virus: health

warn: business