

**CS571: Artificial Intelligence**  
**Assignment 3: Naive Bayes**  
**Due: Friday, March 11, 11:59pm**

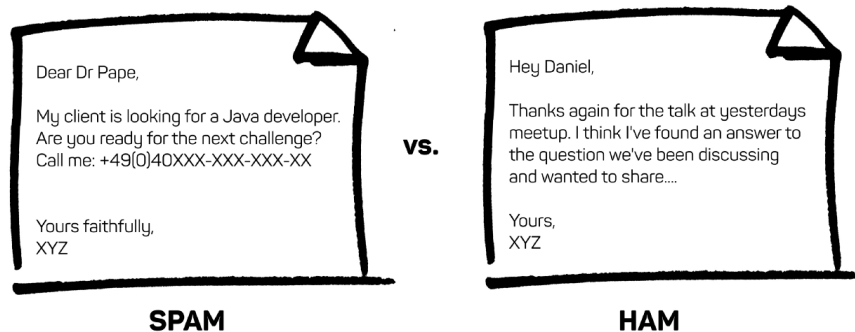
In this assignment, you will implement the Naive Bayes classification method using Python.

You can either do this project individually or in a group of two. Use Google Colab for your code, plots, and comments. Copy the link of the Colab notebook and paste it in the same notebook and also in the comments section of CC. When you finish editing, re-run all the cells to make sure they work and then convert your notebook into a pdf (using print function). You can upload the pdf file on Cougar Courses.

The Naive Bayes code must be your own. You are not allowed to use any machine learning toolkits such as scikit-learn.

---

**Goal is to classify a given message either a “spam” or a “ham”, using a Naive Bayes classifier.**



For this assignment, you will be working with a Spam Collection dataset, consisting of text messages that have been collected for Spam research.

Dataset can be downloaded from here:

<https://drive.google.com/file/d/1mCfliFZt1tbvpEMDEABWRSZpnZCiF1Zh/view?usp=sharing>

The csv file contains one message per line with a total of 30 messages tagged either being ham (legitimate) or spam. Each line is composed of two columns: column 1 contains the label (ham or spam) and column 2 contains raw text.

**Consider the first 20 samples as your training set and the rest 10 samples for your testing.**

Your program should

- Predict the class (spam or ham) for each sample in the test set
- Report the test accuracy