<u>CS561 - ARTIFICIAL INTELLIGENCE LAB</u>

Assignment-3: Naive Bayes Classifier

(Read all the instructions carefully & adhere to them.)

Date: 08-09-2021   **Deadline: 14-09-2021**

Total Credit: 30 (Implementation: 20; Documentation & Explanation: 10)

Problem: Designing a spam filtering based on Naive Bayes classifier. You have to implement for both multinomial and multivariate Naive Bayes classifier versions. To avoid zero counts, make sure you also implement the add-one smoothing

Reading the dataset

Download and unpack **this zip file (smsspamcollection.zip)**. The SMS Spam Collection v.1 is a set of SMS tagged messages that have been collected for SMS Spam research. It contains one set of SMS messages in English of 5,574 messages, tagged according to being ham (legitimate) or spam. It has a total of 4,827 SMS legitimate messages (86.6%) and a total of 747 (13.4%) spam messages. The files contain one message per line. Each line is composed of two columns: one with a label (ham or spam) and the other with the raw text. Here are some examples:

- *ham   What you doing?how are you?*
- *spam   FreeMsg: Txt: CALL to No: 86888 & claim your reward of 3 hours talk time to use from your phone now! ubscribe6GBP/ month inc 3hrs 16 stop?txtStop*

Evaluating the classifier:

Report the 5-fold cross-validation results in terms of accuracy.

Instructions:

1. You can't use any existing implementation. You have to code it by yourself.
2. Compute the accuracy and make your observations to compare between multinomial and multivariate models.
3. The assignment should be completed and uploaded by **14th Oct 2021, 11:59**

PM IST.

4. Markings will be based on the correctness and soundness of the outputs. Marks will be deducted in case of plagiarism.

5. Proper indentation and appropriate comments are mandatory.

6. You should zip all the required files and name the zip file as: **roll_no_of_all_group_members.zip,eg. 2101cs11_2101cs03_2021cs05.zip.**

7. Upload your assignment (**the zip file**) in the following link: