National University
of Computer & Emerging Sciences-Islamabad
Chiniot-Faisalabad Campus

Dr. Hashim Yasin
Department of Computer Science.

CS4104 – Applied Machine Learning
Assignment No. 3

Instructions:

1. Submit your assignment in **soft** as well as in **hard form (Report)** within the due date and time. Soft form does **not** mean the photos of the hard copy. Late submission will result in a deduction of marks.
2. Report should necessarily contain the discussion, comments, and conclusion about the solution. Without a report, you will not get full marks.
3. Mention your names and roll numbers clearly on your document.
4. Name your zip or other folder/file that you want to submit according to the following format: **AML_A3_RollNo_FirstName**
5. Try to solve each task of the assignment on your own.
6. No excuse or resubmission is permissible.
7. Do your assignment in a group of a maximum of two members.
8. There is no restriction of the language for the programming tasks.
9. **In programming tasks, you are NOT allowed to use any built-in function of any library for specific tasks.**

Question No. 1: Bayes Theorem

(a) A medical test for a disease D has outcomes + and −. The probabilities are:

	D	$\neg D$
+	0.009	0.099
−	0.001	0.891

What is the probability that the test will return a positive result for a sick person? If you go for a test and get a positive, what is the probability that you have the disease D?

(b) Suppose you divide your email into three categories: A1 = spam, A2 =low priority, A3 =high priority. From previous experience you know that P(A1) = 0.7, P(A2) = 0.2, P(A3) = 0.1. Let B be the event that the email contains the word “free”. What is the probability that it is spam?

Question No. 2: Naïve Bayes Classifier

Implement the naïve bayes classifier on the following dataset. Your code should be flexible enough that may handle such types of data with different number of examples or the variable with different values.

	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY GOLF
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

Question No. 3: K-mean and K-mediod Clustering

Load the dataset available in the folder with the name “dataset” where there are four features A1, A2, A3, and A4. Your task is to implement the following,

(a) **K-mean clustering** algorithm on the given dataset when the value for **K** is user-defined, that may be 2, 3, or 4, etc.

(b) **K-mediod clustering** algorithm on the given dataset when the value for **K** is user-defined, that may be 2, 3, or 4, etc.

(c) For visualization of the cluster, draw a scatter plot of the dataset and assign colors based on clusters computed through **K-means** and **K-mediod** methods.