National University of Computer and Emerging Sciences 0141

Lahore Campus

Data Science (CS4048)

Sessional-II Exam

Date: Nov 5th 2024

Course Instructor(s)

Ms. Eesha tur Razia Babar

Ms. Maimoona Akram

Mr. Muhammad Saif ul Islam

Total Time (Hrs): 1

Total Marks: 36

Total Questions: 6

Roll No _____ Section _____ Student Signature _____

Do not write below this line

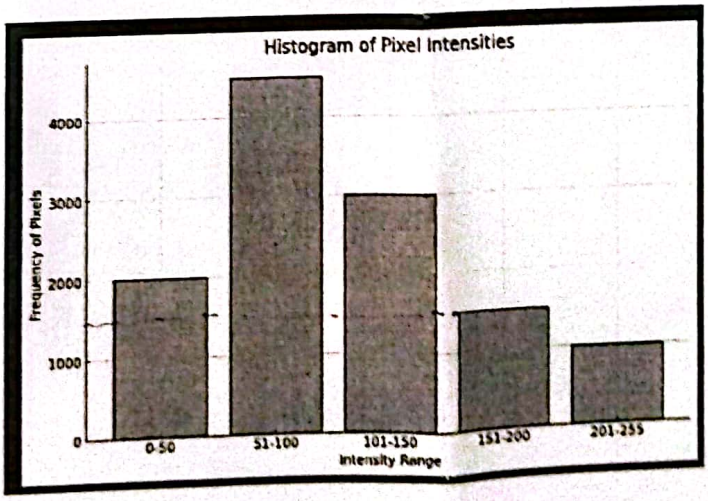
Attempt all the questions on answer sheet. Summarize your descriptive answers into 2-3 sentences. Please show all calculations. Make sure your hand writing is readable.

CLO #1: Extract, clean, and transform data for analysis

Q1. Convert the following sentences into a Bag-of-Words vector representation, where each vector not only stores the occurrence but also the frequency count of each unique token. Remove the stop words before proceeding. [Stop words: is, a, the, of, can, be, or from] **Marks = 6**

a. data Science is a multidisciplinary field
b. data Science is a process of extracting hidden insights from the data
c. data can be structured or unstructured

Q2. You are given a grayscale image with pixel intensities ranging from 0 to 255. Answer the following questions related to it: **Marks = 3+3**



1. Calculate the total number of pixels in the image.

2. Suppose we replicate this image thrice to create an RGB image. What would be the resultant image? Gray scale, Colored or Binary?

Q3: Discretize the **"Age"** attribute using **equi-frequency** binning. Create three bins with labels "teenager", "middle-aged", and "old". Show complete working. **Marks = 6**

ID	1	2	3	4	5	6	7	8	9	10
Age	22	34	15	67	54	75	29	45	63	18

***CLO #2: Apply* tools for performing exploratory data analysis and visualization.**

Q4: Using the following data set and critical value of 5.991, determine if there is any relationship between gender and preference for pets.

Marks = 6

	Cat	Dog	Total
Male	250	50	300
Female	200	1000	1200
Total	450	1050	1500

Q5: Feature Selection **Marks: 3+3**

1. On what basis the wrapper-based feature selection methods select features?
2. Given a set of features, how would you determine if there is multicollinearity, and what impact might this have on your model?

CLO #3: Build and evaluate machine learning models.

Sr.	X1	X2	Y
1	4	2	40
2	6	3	45
3	8	5	55
4	9	7	65
5	10	7	75
6	15	10	85

Q6. Answer the following questions related to optimization algorithms. **Marks = 6**

1. Given a dataset D(X1, X2) and corresponding target values (Y) implement a linear regression model using batch gradient descent. Perform ONE epoch of training with a learning rate of 0.1. Update the weights and bias and display the updated values. Assume initial weights=0, and bias=1.