JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATIONS-2022

M.Tech-I Semester (CS)

COURSE CODE (CREDITS): 22M11CI112 (3) MAX. MARKS: 15

COURSE NAME: Introduction to Data Science

COURSE INSTRUCTORS: Ms. Simran Setia MAX. TIME: 1 Hour

Note: *All questions are compulsory. Marks are indicated against each question in square brackets.*

Q1. Suppose the following data sample is given to you regarding some students enrolled in a MOOC named "Python Programming". The description of the columns is as follows:

Name of the student: It represents the anonymous name of the students.
Marks scored by the student in Exam: It describes the marks obtained by the students in the particular MOOC exam. The data in this column follows normal distribution with mean as 45 and standard deviation as 2.
Internal marks scored by the student: It represents the internal marks scored by the student. The data in this column follows normal distribution with mean 12 and standard deviation 3.
Number of other MOOCs that the student was enrolled in: It represents the number of other MOOCs (apart from the one named "Python programming") that the particular student was enrolled in.
Number of placements secured by the student: It describes the number of placements secured by the student post the MOOC "Python programming".

Find and remove the outliers (if any) from the given data. **[10]**

Name of the Student	Marks scored by the student in Exam	Internal marks scored by the student	Number of other MOOCs that the student was enrolled in	Number of placements secured by the student
A	42.50	12.5	2	5
B	47.0	14.5	2	6
C	98.25	12	2	4
D	42.50	11.5	3	5
E	10.50	7	1	6
F	44.50	14	2	5
G	46.50	11	2	5

H	48.50	12.5	1	3
I	47.25	12	2	8
J	43.0	13.25	3	5

Q2. Suppose you have been the following greyscale image with pixel values as given in the figure.

255	0	0	0	255
255	0	0	0	255
255	0	0	0	255
255	0	0	0	255
255	255	255	255	255

Now, if a mask (with black color) is applied on the whole image. What will be the pixel values of the resulting image? Explain. **[3]**

Q3. What are Stemming and Lemmatization in text preprocessing? Explain the difference between these two with appropriate examples. **[2]**