QP CODE: D2BCS2402	(Pages: 2)	Reg. No : .....
		Name : .....

SECOND SEMESTER FYUGP EXAMINATION, APRIL 2025

MINOR COURSE

CSC2MN102 : Introduction to Data Science

(Credits: 4)

Time: 2 Hours **Maximum Marks: 70**

Section A

Answer the following questions. Each carries 3 marks (Ceiling: 24 marks)

1.	Describe the role of data preprocessing in data science.	BL1	CO2, CO6
2.	Describe the characteristics of a uniform distribution.	BL1	CO1
3.	What is the role of diagnostic data analysis in data science?	BL1	CO1
4.	Explain how k-means assigns data points to clusters.	BL2	CO4
5.	What is the difference between data standardization and normalization?	BL2	CO1, CO2, CO6
6.	Explain Principal Component Analysis (PCA). What is its main purpose?	BL2	CO1, CO2, CO6
7.	Explain how Naive Bayes is used for spam filtering.	BL2	CO4
8.	What is a bar chart?	BL1	CO3
9.	Explain the k-Nearest Neighbors (k-NN) algorithm.	BL1	CO4
10.	Describe the concept of overfitting in machine learning. How can it be prevented during model training?	BL2	CO1, CO2, CO6

Section B

Answer the following questions. Each carries 6 marks (Ceiling: 36 Marks)

11.	Define structured, semi-structured, and unstructured data. Provide examples for each.	BL1	CO1
12.	What are some ethical considerations in data science, particularly regarding data privacy and fairness?	BL2	CO2

(PTO)

13.	What is dispersion, and why is it an important concept in data analysis? Discuss the different measures of dispersion.	BL2	CO2
14.	What are outliers? Explain how they can affect data analysis. Provide two methods to address them.	BL2	CO2
15.	How can you do spam filtration using the Naive Bayes algorithm? Could you explain the steps involved with a suitable example?	BL2	CO1, CO4
16.	Compare and Contrast Supervised, Unsupervised and Reinforcement Learning with suitable examples.	BL2	CO1
17.	Discuss the importance of evaluating machine learning models.	BL2	CO4
18.	What is the k-means clustering algorithm? How does it work, and in which scenarios can it be applied?	BL2	CO5
	Section C		
	Answer any one question. Each carries 10 marks (1 x 10 = 10 Marks)		
19.	Explain the different types of Exploratory Data Analysis (EDA) and their applications in data analysis. Provide examples of each type.	BL2	CO1, CO2
20.	Explain the different types of data and their roles in data science applications, using examples to illustrate your points.	BL2	CO1
	CO : Course Outcome		
	BL : Bloom's Taxonomy Levels (1 – Remember, 2 – Understand, 3 – Apply, 4 – Analyse, 5 – Evaluate, 6 – Create)		