Question Paper Code	12584

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2024
Seventh Semester
Information Technology
20ITPC702 - DATA SCIENCE WITH MACHINE LEARNING
Regulations - 2020

Duration: 3 Hours   Max. Marks: 100

PART - A (10 × 2 = 20 Marks)
Answer ALL Questions

		Marks	K-Level	CO
1.	Define Data Science.	2	K1	CO1
2.	What is Bigdata? List the four types of data.	2	K1	CO1
3.	Write a program to add two matrices.	2	K2	CO2
4.	How R can be used for predictive analysis?	2	K2	CO2
5.	Differentiate supervised and unsupervised Machine Learning.	2	K2	CO3
6.	Consider the following statements: Statement 1: Decision Trees are linear non-parametric models. Statement 2: A decision tree may be used to explain the complex function learned by a neural network. which statements are true ? Justify your answer.	2	K2	CO3
7.	Which type of clustering mechanism is used to group an organization's stake holders?	2	K2	CO4
8.	Define Centroid.	2	K1	CO4
9.	Justify your answer for why pruning is necessary?	2	K2	CO5
10.	Differentiate Univariate trees and Multivariate trees.	2	K2	CO5

PART - B (5 × 13 = 65 Marks)
Answer ALL Questions

11.	a)	Describe the life cycle of Data Science with a neat diagram.	13	K2	CO1
		OR			
	b)	Explain the role of Business Intelligence in data science.	13	K2	CO1
12.	a) i)	Write a R program to find the maximum and the minimum value of a given vector.	7	K3	CO2
	ii)	Check whether the value of the element of a given vector greater than 10 or not. Return TRUE or FALSE.	6	K3	CO2
		OR			
	b) i)	Write the R code for the following a. Calling a function with default arguments.	7	K2	CO2

K1 – Remember; K2 – Understand; K3 – Apply; K4 – Analyze; K5 – Evaluate; K6 – Create   ***12584***

b. Calling a function with arguments.

c. Calling a function without arguments.

ii) Illustrate a R code using the following functions: 6   K2   CO2
seq() , paste(), print(), format(), mode(), order().

13. a) Discuss in detail about supervised machine learning algorithm with   13   K2   CO3
neat sketch.

OR

b) Calculate the two regression equations of X on Y and Y on X from   13   K3   CO3
the data given below, taking deviations from a actual means of X and
Y.

Prize	10	12	13	12	16	15
Amount Demanded	40	38	43	45	37	43

Estimate the likely demand when the price is Rs.20

14. a) List the various types of Clustering. Explain any two types of   13   K2   CO4
Clustering with real time examples.

OR

b) Cluster the following fifteen points (with (x, y) representing locations)   13   K3   CO4
into three clusters. A1(2,10), A2(2,6), A3(11,11), A4(6,9), A5(6,4),
A6(1,2), A7(5,10), A8(4,9), A9(10,12), A10(7,5), A11(9,11),
A12(4,6), A13(3,10), A14(3,8), A15(6,11).

15. a) Elaborate in detail about Reinforcement Learning with an example.   13   K2   CO5

OR

b) Explain in detail about Multivariate trees and its applications.   13   K2   CO5

PART - C (1 × 15 = 15 Marks)

16. a) Given data = {12,13,14,15,16;1,5,3,6,7,8}.   15   K3   CO6
Compute the principal component using PCA Algorithm.

OR

b) Describe how dimensionality reduction methods can be applied as a   15   K3   CO6
preprocessing step before constructing graphical models. Discuss how
this approach can help mitigate the curse of dimensionality and
improve the performance of the graphical model. Provide a step-by-
step explanation of the overall process, including the order of
operations.