Reg. No.															

Question Paper Code	13657

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2025

Seventh Semester

Information Technology

20ITPC702 – DATA SCIENCE WITH MACHINE LEARNING

Regulations - 2020

Duration: 3 Hours Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	State the primary purpose of Big Data in data science. (a) To store large amounts of unstructured data (b) To reduce data storage costs (c) To perform basic calculations (d) To create graphical user interfaces	1	K1	CO1
2.	In the context of data science, what does SQL stand for? (a) Structured Query Language (b) Sequential Query Language (c) Simple Query Language (d) Standard Query Language	1	K1	CO1
3.	In R, which function is used to combine elements into a vector? (a) combine( ) (b) c( ) (c) vector( ) (d) append( )	1	K1	CO2
4.	Which command is used to load a library/package in R? (a) use( ) (b) load( ) (c) library( ) (d) require( )	1	K1	CO2
5.	In supervised learning, the model learns from? (a) Unlabeled data (b) Rewards (c) Labeled data (d) Random guesses	1	K1	CO3
6.	In a univariate decision tree, each split is based on (a) A combination of multiple features (b) A single feature (c) Randomly chosen features (d) A clustering algorithm	1	K1	CO3
7.	In K-means clustering, the "K" refers to (a) The number of clusters (b) The number of data points (c) The number of iterations (d) The size of the dataset	1	K1	CO4
8.	Which method is commonly used to determine the optimal number of clusters in K-means clustering? (a) Silhouette method (b) Elbow method (c) Cross-validation (d) Random initialization	1	K1	CO4
9.	Which of the following matrices is used in PCA to compute principal components? (a) Correlation matrix (b) Covariance matrix (c) Distance matrix (d) Similarity matrix	1	K1	CO5
10.	Which of the following is a type of graphical model? (a) Bayesian Network (b) Linear Regression (c) Support Vector Machine (d) Decision Tree	1	K1	CO6

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

11.	List the key component technologies of Data Science.	2	K1	CO1
12.	State Business Intelligence (BI).	2	K1	CO1
13.	Differentiate between Scalars, vector, list, Matrix and Data frame.	2	K2	CO2
14.	List out some built-in functions.	2	K1	CO2
15.	Differentiate supervised and unsupervised Machine Learning algorithm.	2	K2	CO3
16.	To group an organization’s stakeholders, which type of clustering mechanism is used?	2	K2	CO3
17.	Differentiate Agglomerative clustering and Divisive Clustering.	2	K2	CO4
18.	Find the centroid of the given three points. A(4, 9), B(6, 15), and C(2, 6).	2	K2	CO4

19. State the scenario at which Bellman equation is used in Reinforcement learning. 2 K2 CO5

20. How does LDA differ from PCA in terms of its objective? 2 K2 CO5

21. State eigenvalues and eigenvectors. 2 K1 CO6

22. Differentiate between directed and undirected graphical models. 2 K2 CO6

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23. a) Brief about the life cycle of Data Science with a neat diagram. 11 K2 CO1

OR

b) Discuss in detail about hadoop framework. 11 K2 CO1

24. a) Discuss the key components of the R environment. How do you manage the workspace in R, including saving and loading the workspace, and removing objects? 11 K3 CO2

OR

b) Discuss conditional execution in R using if, else if, and else statements. Provide examples showing how these can be used to control the flow of a program. 11 K3 CO2

25. a) Classify the given sample using a decision tree algorithm. Apply the same to construct a decision tree for the data given below. 11 K3 CO3

Size	Color	Shape	Class
Small	Yellow	Round	A
Big	Yellow	Round	A
Big	Red	Round	A
Small	Red	Round	A
Small	Black	Round	B
Big	Black	Cube	B
Big	Yellow	Cube	B
Big	Black	Round	B
Small	Yellow	Cube	B

OR

b) Calculate the two regression equations of X on Y and Y on X from the data given below, taking deviations from a actual means of X and Y. Estimate the likely demand when the price is Rs.20. 11 K3 CO3

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

26. a) Discuss the various methods for choosing the number of clusters in clustering algorithms. 11 K2 CO4

OR

b) Discuss in detail about the various types of Clustering. 11 K2 CO4

27.	a)	State pruning in the context of decision trees? Discuss the different pruning methods (pre-pruning and post-pruning) and their effects on model performance and complexity. Provide examples of scenarios where each method might be preferable.	*11*	*K2*	*CO5*
		OR			
	b)	Discuss in detail about how PCA is used for dimensionality reduction with neat illustration.	*11*	*K2*	*CO5*
28.	a)	Discuss in detail about reinforcement learning with neat illustration.	*11*	*K2*	*CO6*
		OR			
	b)	Discuss with illustration about the graphical models in detail.	*11*	*K2*	*CO6*