Reg. No.															

Question Paper Code	13076

MBA - DEGREE EXAMINATIONS, NOV / DEC 2024

Third Semester

Master of Business Administration

20MBS304 - DATA MINING

Regulations - 2020

Duration: 3 Hours | Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	What is the relationship between metadata and data granularity?	2	K1	CO1
2.	How do enterprise conceptual entity models contribute to the design of robust data systems?	2	K1	CO1
3.	Outline the functions of clustering in data mining.	2	K2	CO2
4.	Illustrate the key elements to be considered while analyzing ethical issues in data mining.	2	K2	CO2
5.	List out the difference between regression and correlation in data analysis.	2	K3	CO3
6.	How linear regression helps in predicting continuous variables?	2	K3	CO3
7.	List the significance of historical data in the training of machine learning models.	2	K1	CO4
8.	List out the main differences between machine learning and artificial intelligence.	2	K1	CO4
9.	Organize how data mining enhances inventory management and product recommendations in retail.	2	K1	CO5
10.	Explain the ways in which data mining improves personalized healthcare and treatment plans.	2	K2	CO5

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

			Marks	K – Level	CO
11.	a)	Outline the role of an enterprise conceptual model in the development of a new information system for a healthcare organization. Explain how it helps in defining the scope and structure of the data.	13	K2	CO1
		OR			
	b)	Interpret how data granularity impacts decision-making in a financial institution, particularly in credit risk analysis and customer segmentation.	13	K2	CO1
12.	a)	Make use of the application of data mining in customer behaviour analysis. Explain the types of relationships identified and how they influence marketing strategies.	13	K3	CO2
		OR			

b) Identify the advantages of data mining in pharmaceutical research, particularly in drug discovery and patient care. Discuss the risks involved and how ethical guidelines can be followed. *13 K3 CO2*

13. a) Identify how regression analysis can be applied in predicting customer demand for a retail business, and compare it with correlation in identifying relationships between sales and marketing expenditure. *13 K3 CO3*

OR

b) Address how decision trees can be used for loan approval processes in banking, explaining the classification method and how pruning enhances the model's accuracy. *13 K3 CO3*

14. a) Examine how reinforcement learning can be used to train autonomous vehicles. Explain how the system learns optimal driving behaviours by interacting with its environment. *13 K4 CO4*

OR

b) Analyse how supervised learning can be applied in fraud detection systems for the banking industry. Explain how labeled data and classification algorithms are used to identify fraudulent transactions. *13 K4 CO4*

15. a) Explain the role of data mining in medical diagnostics and healthcare, focusing on predicting patient outcomes, disease progression, and optimizing treatment plans. *13 K5 CO5*

OR

b) Assess the application of data mining in production management for optimizing manufacturing processes, reducing downtime, and improving supply chain efficiency. *13 K5 CO5*

PART - C (1 × 15 = 15 Marks)
(Compulsory)

16. a) A large e-commerce company in India wants to optimize its product recommendation system. The goal is to recommend products that are not only similar to the user's previous purchases but also improve sales diversity across different product categories. The company has a vast amount of historical data related to user purchases, browsing patterns, and product categories. They decide to implement Ant Colony Optimization (ACO) for enhancing the recommendation algorithm, which would simulate the behavior of ants searching for optimal paths to balance customer preferences with the diversity of products offered. *15 K5 CO5*

Questions:

(i). Explain how the ACO algorithm will adapt to dynamic changes in customer preferences over time.

(ii). Propose how the algorithm might handle large-scale data while maintaining efficient processing.