Reg. No.															

Question Paper Code	13515

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

Sixth Semester

Information Technology

20CSEL501 - DATA WAREHOUSING AND DATA MINING

Regulations - 2020

Duration: 3 Hours | Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	Which of the following is not a basic data mining task? (a) Classification (b) Prediction (c) Spooling (d) Clustering	1	K1	CO1
2.	What is the primary purpose of OLTP systems? (a) Data analysis (b) Business reporting (c) Day-to-day transaction processing (d) Complex queries and aggregations	1	K1	CO1
3.	What is the main goal of the Knowledge Discovery Process? (a) Storing data in databases (b) Extracting useful patterns from large datasets (c) Encrypting sensitive information (d) Designing relational schemas	1	K1	CO2
4.	What is the role of data visualization in DBMS? (a) Indexing database records (b) Securing stored procedures (c) Representing query results graphically (d) Managing database schemas	1	K1	CO2
5.	Which of the following tools is widely used for data mining in databases? (a) Microsoft Access (b) Weka (c) SQL Server Management Studio (d) phpMyAdmin	1	K1	CO3
6.	Which of the following methods is commonly used for classification using frequent patterns? (a) Apriori (b) FP-Growth (c) Naive Bayes (d) Both A and B	1	K1	CO3
7.	In decision tree induction, the root node is selected based on: (a) Alphabetical order (b) Random choice (c) Highest information gain (d) Lowest support count	1	K1	CO4
8.	Which of the following techniques is commonly used to improve classification accuracy in a machine learning model? (a) Feature selection (b) Normalization (c) Data augmentation (d) All of the above	1	K1	CO4
9.	Cluster analysis is primarily used to: (a) Classify data into pre-defined categories (b) Group similar data points together (c) Predict future outcomes based on past data (d) Identify anomalies or outliers in the data	1	K1	CO5
10.	WEKA stands for: (a) Western Electronic Knowledge Analysis (b) Waikato Environment for Knowledge Analysis (c) Wide Expert Knowledge for Analysis (d) Weka Environmental Knowledge Architecture	1	K1	CO6

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

11.	List the characteristic of Data warehousing.	2	K1	CO1
12.	Compare OLTP and OLAP system.	2	K2	CO1
13.	What is Data Objects?	2	K1	CO2
14.	Distinguish between data cleaning and noisy data.	2	K2	CO2
15.	What is the difference between association and correlation in data analysis?	2	K1	CO3

16.	Show the Multi Dimensional Space.	*2*	*K2*	*CO3*
17.	Give the features of Decision tree induction.	*2*	*K2*	*CO4*
18.	Define Lazy learners with an example.	*2*	*K1*	*CO4*
19.	List the major clustering methods.	*2*	*K1*	*CO5*
20.	Examine the challenges of outlier detection	*2*	*K2*	*CO5*
21.	Mention the Iris plants database.	*2*	*K1*	*CO6*
22.	Write concept about Rule Learners.	*2*	*K2*	*CO6*

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23.	a) i)	Draw the data warehouse architecture and explain its components.	*6*	*K2*	*CO1*
	ii)	Explain the steps in building a data warehouse.	*5*	*K2*	*CO1*
		OR			
	b) i)	Explain the Multidimensional Data Model.	*6*	*K2*	*CO1*
	ii)	Explain the different types of OLAP tools.	*5*	*K2*	*CO1*
24.	a)	Demonstrate in detail about data mining steps in the process of knowledge discovery.	*11*	*K3*	*CO2*
		OR			
	b) i)	Explain data visualization. How it helps data warehousing.	*6*	*K3*	*CO2*
	ii)	Summarize the Data Similarity and Dissimilarity Measures.	*5*	*K3*	*CO2*
25.	a)	Analyze the effectiveness of Pattern Evaluation Method and Pattern Mining Multilevel.	*11*	*K3*	*CO3*
		OR			
	b)	Analyze the advantages of using frequent pattern mining for classification compared to traditional classification techniques.	*11*	*K3*	*CO3*
26.	a)	Generalize the Bayes theorem of posterior probability and explain the working of a Bayesian classifier with an example.	*11*	*K3*	*CO4*
		OR			
	b)	Design a neural network model for binary classification using back propagation and explain its architecture.	*11*	*K3*	*CO4*
27.	a)	Explain in detail about density based methods and Grid Based Methods.	*11*	*K2*	*CO5*
		OR			
	b)	Demonstrate on clustering high dimensional data.	*11*	*K2*	*CO5*
28.	a)	Compare the Breast Cancer dataset with other medical datasets in terms of completeness and usability.	*11*	*K3*	*CO6*
		OR			
	b)	Evaluate and compare the Learning algorithms and Clustering algorithms.	*11*	*K3*	*CO6*