Reg. No.												

Question Paper Code	12324

B.E. / B.Tech. / M.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2023

Fifth Semester

Computer Science and Engineering (5 Years Integrated)

20CJPC501 - DATA WAREHOUSING AND DATA MINING

(Regulations 2020)

Duration: 3 Hours Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

		Marks, K-Level, CO
1.	Define Metadata in Data warehousing.	*2,K1,CO1*
2.	What are the two approaches to build the data warehouse?	*2,K1,CO1*
3.	Name the some of the data mining techniques.	*2,K1,CO2*
4.	Define association and correlations.	*2,K1,CO2*
5.	Define Lift and write the formula.	*2,K1,CO3*
6.	What is strong association Rule explains with an example?	*2,K1,CO3*
7.	Define precision and recall.	*2,K1,CO4*
8.	What is Lazy learners algorithm?	*2,K1,CO4*
9.	How do you explain the similarity in clustering?	*2,K2,CO6*
10.	Illustrate some Application of outlier analysis.	*2,K2,CO6*

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

11.	a)	Explain the Data warehouse architecture with neat diagram.	*13,K2,CO1*
		OR	
	b)	Explain mapping data warehouse with multiprocessor architecture with the concept of parallelism and data partitioning.	*13,K2,CO1*
12.	a)	Explain the major data preprocessing techniques and explain in detail with examples.	*13,K2,CO2*
		OR	
	b)	(i) Demonstrate in detail about data mining steps in the process of knowledge discovery. (ii) List the application area of data mining.	*7,K2,CO2* *6,K2,CO2*
13.	a)	Apply FP Growth algorithm to find the Frequent item set for the following Transactional database with minimum support count value=2.	*13,K3,CO3*

Transaction ID	Items
T100	I1,I2,I5
T200	I2,I4
T300	I2,I3
T400	I1,I2,I4
T500	I1,I3
T600	I2,I3
T700	I1,I3
T800	I1,I2,I3,I5
T900	I1,I2,I3

OR

b) Apply Apriori Algorithm and find the Frequent item set for the following Transactional database using Apriori Algorithm with the minimum support count value=2. *13,K3,CO3*

Transaction ID	Items
T100	I1,I3,I4
T200	I2,I3,I5
T300	I1,I2,I3,I5
T400	I2,I5

14. a) Explain the classification of Decision tree induction. *13,K2,CO4*

OR

b) Define classification. With an example explain how support vector machines can be used for classification. *13,K2,CO4*

15. a) Discuss the following clustering algorithm with an example.
(i) K.means. *7,K2,CO6*
(ii) K-medoid. *6,K2,CO6*

OR

b) Explain briefly about hierarchical clustering. *13,K2,CO6*

PART - C (1 × 15 = 15 Marks)

16. a) Apply R program to perform the Decision tree classification and explain with an example. *15,K3,CO5*

OR

b) Apply R program to perform the K means algorithm and explain with an example. *15,K3,CO5*