Reg. No.												

Question Paper Code	12354

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

Fifth Semester

Computer and Communication Engineering

20CSEL501 - 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.	What is a Data Ware House?	*2,K1,CO1*
2.	Define Multidimensional Database.	*2,K1,CO1*
3.	Define Data mining.	*2,K1,CO2*
4.	Name the steps in the process of knowledge discovery.	*2,K1,CO2*
5.	Define frequent patterns.	*2,K1,CO3*
6.	Define Support and Confidence.	*2,K1,CO3*
7.	State classification.	*2,K1,CO4*
8.	Define Clustering.	*2,K1,CO4*
9.	List the WEKA data formats.	*2,K1,CO6*
10.	Define Auto Imports in WEKA.	*2,K1,CO6*

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

11.	a)	Differentiate between OLAP and OLTP.	*13,K2,CO1*
		OR	
	b)	Write the types of OLAP operations in detail with examples.	*13,K2,CO1*
12.	a)	Explain the Various Data objects and Attributes in detail.	*13,K2,CO2*
		OR	
	b)	Briefly explain about Measuring Data Similarity and Dissimilarity methods.	*13,K3,CO2*
13.	a)	Explain the Apriori algorithm with suitable example.	*13,K3,CO3*
		OR	

b) Consider the Following Transaction. *13,K3,CO3*
T1 = { I1,I3,I6 } , T2 = { I2,I3,I5,I6 } , T3 = { I1,I2,I3,I5} ,
T4= { I2,I5 } , T5 = { I1,I3,I5 }
List all the Frequent Item set using FP-Growth Algorithm.

14. a) Discuss about basic decision tree induction algorithm. *13,K3,CO4*

OR

b) Explain briefly about the Naïve Bayesian Classification. *13,K3,CO4*

15. a) (i) Describe Iris Plant Dataset. *6,K2,CO6*
(ii) Describe Breast cancer Data set. *7,K2,CO6*

OR

b) How classification algorithms are modeled in WEKA tool? *13,K2,CO6*

PART - C (1 × 15 = 15 Marks)

16. a) Suppose that the data mining task is to cluster the following eight points (with (x,y) representing location) into three clusters. *15,K3,CO5*
A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8),B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9).
Suppose initially we assign A1, B1, and C1 as the center of each cluster respectively.
Use the K-Means algorithm to show only the three cluster centers after the first round of execution and The final three clusters.
(Consider The distance function as Manhattan distance)

OR

b) Explain hierarchical clustering in detail with an example. Draw the dendrogram using hierarchical clustering algorithm. *15,K3,CO5*