R20

Code No: R20A1206

MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY

(Autonomous Institution – UGC, Govt. of India)

III B.Tech I Semester Supplementary Examinations, June 2025

Data Warehousing and Data Mining

(CSE-IOT & B.Tech-AIML)

Roll No										

Time: 3 hours **Max. Marks: 70**

Note: This question paper Consists of 5 Sections. Answer **FIVE** Questions, Choosing ONE Question from each SECTION and each Question carries 14 marks.

***

			BCLL	CO(s)	Marks
		SECTION-I			
1	***A***	Define Data warehouse. List the characteristics of data warehouse.	**L1**	**CO-I**	**[4M]**
	B	Explain about Data ware house architecture and its components.	**L2**	**CO-I**	**[10M]**
		OR			
2	***A***	Explain various OLAP operations used in Data warehousing.	**L2**	**CO-I**	**[10M]**
	B	List characteristics of Dimension table.	**L1**	**CO-I**	**[4M]**
		SECTION-II			
3	***A***	Summarize data mining functionalities.	**L2**	**CO-II**	**[7M]**
	B	Explain data mining task primitives.	**L4**	**CO-II**	**[7M]**
		OR			
4	***A***	What is data cleaning? Outline the advantages of Data cleaning.	**L2**	**CO-II**	**[7M]**
	B	(i)Categorize the major approaches for data integration. (ii)What are the issues to consider during data integration.	**L4** **L2**	**CO-II**	**[7M]**
		SECTION-III			
5	***A***	Consider the following dataset to analyse and find frequent item sets.	**L5**	**CO-III**	**[10M]**

TID	items
T1	I1, I2 , I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

Explain the step-by-step process to find frequent itemssets and generate association rules using the Apriori algorithm:
Minimum support=2, Minimum Confidence=60%

	B	Define Support and confidence.	**L1**	**CO-III**	**[4M]**
		OR			
6	***A***	Illustrate FP Growth algorithm with an example.	**L2**	**CO-III**	**[10M]**
	B	Outline the disadvantages of Apriori algorithm.	**L2**	**CO-III**	**[4M]**
		SECTION-IV			
7	***A***	Explain decision tree with suitable example.	**L2**	**CO-IV**	**[7M]**
	B	Explain Bayesian Belief networks algorithm with an example.	**L4**	**CO-IV**	**[7M]**
		OR			
8	***A***	Explain working of KNN algorithm in detail.	**L4**	**CO-IV**	**[10M]**
	B	List the disadvantages of KNN algorithm.	**L1**	**CO-IV**	**[4M]**
		SECTION-V			
9	***A***	Apply K(=2)-Means algorithm over the data (185, 72), (170, 56), (168, 60), (179,68), (182,72), (188,77) up to two iterations and show the clusters. Initially choose first two objects as initial centroids.	**L3**	**CO-V**	**[10M]**
	B	Outline the applications of cluster analysis.	**L2**	**CO-V**	**[4M]**
		OR			
10	***A***	List out the key Issues in Hierarchical Clustering.	**L1**	**CO-V**	**[4M]**
	B	Explain the process of Agglomerative clustering in detail.	**L4**	**CO-V**	**[10M]**

***