Code No: R17A0524 **R17**

MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY

(Autonomous Institution – UGC, Govt. of India)

IV B.Tech I Semester Supplementary Examinations, Oct/Nov 2023

Data Warehousing and Data Mining

(CSE)

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.

***

SECTION-I

1	***A***	Give examples for defining star, snowflake and fact constellation schemas.	**[7M]**
	B	Discuss about a three-tier data warehouse architecture.	**[7M]**
		OR	
2	***A***	What are the various types of OLAP servers? Explain.	**[7M]**
	B	Explain the importance of Extraction-Transformation-Loading process.	**[7M]**
		SECTION-II	
3	***A***	Explain Data Mining Task Primitives with examples.	**[7M]**
	B	Discus the importance of dimensionality reduction.	**[7M]**
		OR	
4	***A***	Briefly compare and explain by taking an example of your point(s) Data cleaning, data transformation.	**[7M]**
	B	Explain about concept hierarchy generation for categorical data.	**[7M]**
		SECTION-III	
5	***A***	Find the frequent item sets **and generate** association rules on this. Assume that minimum support threshold (s = 33.33%) and minimum confident threshold (c = 60%)	**[10M]**

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs,Coke,Chips
T4	Chips,Coke
T5	Chips,ketchup
T6	Hot Dogs,Coke,Chips

	B	Explain in detail about Closed Frequent Itemset.	**[4M]**
		OR	
6	***A***	The given data is a hypothetical dataset of transactions with each letter representing an item. The minimum support given is 3.	**[10M]**

TID	Items Bought
100	*f, a, c, d, g, i, m, p*
200	*a, b, c, f, l, m, o*
300	*b, f, h, j, o*
400	*b, c, k, s, p*
500	*a, f, c, e, l, p, m, n*

[4M]

B Discuss the applications of Association rule mining

SECTION-IV

7 ***A*** Explain the Measures for Selecting the Best Split in Decision tree construction. **[7M]**

B We have data from the questionnaires survey to ask people opinion and objective testing with two attributes acid durability and strength to classify whether a special paper tissue is good or not. Here are four training samples. **[7M]**

X1= Acid durability(seconds)	X2=Strength (kg/square meter)	Y=Classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

OR

8 **[14M]**

Tid	Refund	Marital status	Taxable income	Evade
1	Yes	Single	125k	No
2	No	Married	100k	No
3	No	Single	70k	No
4	Yes	Married	120k	No
5	No	Divorced	95k	Yes
6	No	Married	60k	No
7	Yes	Divorced	220k	No
8	No	Single	85k	Yes
9	No	Married	75k	No
10	No	Single	90k	Yes

Find the class of this test record:

X= (Refund=yes , Marital status =Married, Income=85K)

SECTION-V

9 ***A*** Cluster the following eight points (with (x, y) representing locations) with K-means into into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9) **[10M]**

B Discuss the Key Issues strengths and weakness of Hierarchical Clustering **[4M]**

OR

10 ***A*** Hierarchical clustering Use single and complete link agglomerative clustering to group the data described by the following distance matrix. **[10M]**

	A	B	C	D
A	**0**	**1**	**4**	**5**
B		**0**	**2**	**6**
C			**0**	**3**
D				**0**

B Discuss the importance of Outlier Detection in clustering. **[4M]**

****