Total No. of Questions: 22

Enrollment No : _____

MEDICAPS UNIVERSITY

Faculty of Engineering

End Semester Examination May 2025

IT3ED02 Data Mining & Warehousing

Programme	:	B.Tech.	**Branch/Specialisation**	:	IT
Duration	:	3 hours	**Maximum Marks**	:	60

Note: *All questions are compulsory. Internal choices, if any, are indicated. Assume suitable data if necessary. Notations and symbols have their usual meaning.*

Section 1 (Answer all question(s))

		Marks	CO	BL
Q1.	Operational database is based on- ○ Star schema ◉ ER-model ○ Snowflake schema ○ Fact constellation schema	1	1	1
Q2.	A star schema has what type of relationship between a dimension and fact table? ○ Many-to-many ○ One-to-one ◉ One-to-many ○ All of the above	1	1	1
Q3.	In _____, the value of an attribute is examined as it varies over time. ○ Prediction ○ Regression ◉ Time series analysis ○ Sequence discovery	1	2	2
Q4.	Incorrect or invalid data is known as- ○ Rich data ○ Outliers ◉ Noisy data ○ Missing data	1	2	1
Q5.	When a sub-node divides into other sub-nodes, it is referred to as a- ○ Pruning ◉ Decision node ○ Splitting ○ Terminal node	1	3	1
Q6.	A set of together is called _____. ○ Frequent Itemset ◉ Itemset ○ K-itemset ○ None of these	1	3	2
Q7.	Given two items (4, 6) that are represented as tuples (2, 0). Between two objects, what is the Manhattan distance? ○ 2 ○ 4 ○ 6 ◉ 8	1	4	2
Q8.	Given two items (3, 5) that is represented as tuples (2, 0). How far apart are two things in Euclidean distance? ○ 5 ◉ 5.1 ○ 6 ○ None of the above	1	4	1
Q9.	Data that can be modelled as dimension attributes and measure attributes are called _____. ○ Mono-dimensional data ◉ Multi-dimensional data ○ Measurable data ○ Efficient data	1	5	2

Q10. The process of viewing the cross-tab (Single dimensional) with a fixed value of one attribute is- **1 1 1**

- (●) Slicing
- ( ) Dicing
- ( ) Pivoting
- ( ) Both (A) and (B)

Section 2 (Answer all question(s))

Marks CO BL

Q11. Explain data warehousing components. **2 1 1**

Rubric	Marks
data warehousing components	2

Q12. Define data mart and its types. **3 1 2**

Rubric	Marks
Define data mart 1 mark, its types 2 marks.	3

Q13. (a) Discuss star, snowflake, and galaxy schema for multidimensional databases. **5 1 2**

Rubric	Marks
Star, Snowflake and Galaxy schema 5 marks	5

(OR)

(b) Explain ETL process in data warehouse.

Rubric	Marks
ETL Process describe 5 marks	5

Section 3 (Answer all question(s))

Marks CO BL

Q14. Explain application and challenges of data mining. **4 2 2**

Rubric	Marks
Application 2 marks, Challenges 2 marks.	4

Q15. (a) Explain data mining architecture. Write some applications of data mining. **6 2 2**

Rubric	Marks
Architecture 3 marks, Application 3 marks	6

(OR)

(b) Explain the knowledge discovery process.

Rubric	Marks
Knowledge Discovery Process with diagram 6 marks	6

Section 4 (Answer all question(s))

Marks CO BL

Q16. What are an itemset and a frequent itemset? **4 3 2**

Rubric	Marks
Itemset 2 marks, Frequent Itemset 2 marks.	4

Q17. (a) A database has five transactions. Let minimum support=60% and minimum confidence=80%. **6 3 2**

TID ITEMS_BOUGHT
T100 {M, O, N, K, E, Y}
T200 {D, O, N, K, E, Y }
T300 {M, A, K, E}
T400 {M, U, C, K, Y}
T500 {C, O, O, K, I, E}
Find all frequent itemsets using the Apriori algorithm.

Rubric	Marks
Find all frequent itemsets using the Apriori algorithm.	6

(OR)

(b) Explain pattern evaluation methods.

Rubric	Marks
Pattern evaluation methods minimum 3 methods	6

Section 5 (Answer all question(s))

Marks CO BL

Q18. Explain DBSCAN with application. **4 4 2**

Rubric	Marks
DBSCAN 2 marks, applications 2 marks	4

Q19. (a) Difference between clustering and classification. Briefly describe the hierarchical clustering method. **6 4 2**

Rubric	Marks
Difference between clustering and classification. 4 marks, Briefly describe the hierarchical clustering method. 2 marks	6

(OR)

(b) Suppose we have the following points: (1,1), (2,4), (3,4), (5,8), (6,2), (7,8). Use k - means algorithm (k = 2) to find two cluster. The distance function is Euclidean distance.

Rubric	Marks
Solution to find two clusters	6

Section 6 (Answer any 2 question(s))

Marks CO BL

Q20. What do you mean by reporting and query tools? **5 5 1**

Rubric	Marks
Reporting and Query Tools	5

Q21. What do you KNIME, ORANGE and ETL? Explain use in business analysis. **5 5 1**

Rubric	Marks
KNIME, ORANGE and ETL 2 marks, use in business analysis 3 marks.	5

Q22. Explain OLAP operations in brief. **5 5 2**

Rubric	Marks
Explain OLAP operations in brief.	5

******