Q. Code: 803951

Reg. No. | | | | | | | | | | | | | |

B.E. / B.TECH. DEGREE EXAMINATIONS, MAY 2024

Third Semester

CS22302 – DATAMINING AND DATA WAREHOUSING

(Common to CS & AD)

(Regulation 2022)

TIME: 3 HOURS **MAX. MARKS: 100**

COURSE OUTCOMES	STATEMENT	RBT LEVEL
CO 1	Understand data mining concepts and apply classification techniques.	**2**
CO 2	Do data analysis using frequent pattern and association rule mining techniques.	**3**
CO 3	Students will be able to apply various clustering techniques.	**3**
CO 4	Students will be able to understand data warehouse concepts, architecture and schema.	**2**
CO 5	Students will be able to understand various types of online analytical processing techniques.	**2**

PART- A (20 x 2 = 40 Marks)

(Answer all Questions)

		CO	RBT LEVEL
1.	List the need for Data Mining.	**1**	**2**
2.	What is a class label in a dataset?	**1**	**2**
3.	Discuss True positive and True Negative.	**1**	**2**
4.	Write about the various steps involved in Data Preprocessing.	**1**	**2**
5.	List the metrics used to evaluate the strength of association rules.	**2**	**2**
6.	Is itemset {A} closed frequent itemset?	**2**	**3**

Transaction ID Items

1	A B C D
2	A B D
3	A C D
4	B C D

7.	State the Apriori property of itemset.	**2**	**2**
8.	Define the use of Apriori Algorithm.	**2**	**2**

9.	Given two objects represented by the attribute values (1, 6, 2, 5, 3) and (3, 5, 2, 6, 6), compute the Euclidean distance and Manhattan distance between these two objects.	**3**	**3**
10.	Does clustering require labeled dataset? Why?	**3**	**2**
11.	Compare agglomerative algorithm with divisive algorithm.	**3**	**2**
12.	List any three weakness of DBSCAN algorithm	**3**	**2**
13.	Define Fact table.	**4**	**2**
14.	Mention the types of Schemas.	**4**	**2**
15.	Discuss about Dimension table with example.	**4**	**2**
16.	Describe the ETL process in data warehousing.	**4**	**2**
17.	Differentiate OLAP and OLTP.	**5**	**2**
18.	Mention the various OLAP Applications.	**5**	**2**
19.	List the various Challenges in Data Mining process.	**5**	**2**
20.	Outline how multidimensional data is represented.	**5**	**2**

PART- B (5 x 10 = 50 Marks)

		Marks	CO	RBT LEVEL
21. (a)	Explain in detail about various data mining techniques with suitable example.	**(10)**	**1**	**2**
	(OR)			
(b)	Explain decision tree algorithm with suitable example.	**(10)**	**1**	**2**
22. (a)	Explain the Apriori–TID algorithm with the given dataset for the minimum support of 50% and minimum confidence of 75%.	**(10)**	**2**	**3**

Transaction ID	Items bought

100	Bread, Cheese
200	Bread, Cheese, Juice
300	Bread, Milk
400	Cheese, Juice, Milk

(OR)

(b) Apply the FP growth algorithm to find the most frequent items pair(s). **(10) 2 3**

Transaction ID	Items
T1	{E, K, M, N, O, Y}
T2	{D, E, K, N, **O**, Y}
T3	{A, E, K, M}
T4	{C, K, M, U, Y}
T5	{C, E, I, K, O, O}

23. (a) Create four clusters for the given dataset using the K-means algorithm, indicate all the intermediate steps. **(10) 3 3**

Food Item	Protein content P	Fat content F
1	1.1	60
2	8.2	20
3	4.2	35
4	1.5	21
5	7.6	15
6	2	55
7	3.9	39

(OR)

(b) With suitable diagrams, explain the concepts used in the density-based algorithm to cluster the data. **(10) 3 3**

24. (a) With a neat sketch explain in detail about the Data Warehouse Architecture. **(10) 4 2**

(OR)

(b) Summarize the difference between OLTP, ODS and Data Warehouse. **(10) 4 2**

25. (a) Explain in detail about OLAP Servers with suitable example. **(10) 5 2**

(OR)

(b) Discuss in detail about ROLAP and MOLAP. **(10) 5 2**

PART- C (1 x 10 = 10 Marks)

(Q.No.26 is compulsory)

	Marks	CO	RBT LEVEL
26. A mobile theft dataset is given below.	**(10)**	1	**5**

ID	Color	Type	Origin	Stolen
1	Black	Android	Domestic	Yes
2	Black	Android	Domestic	No
3	Black	Android	Domestic	Yes
4	White	Android	Domestic	No
5	White	Android	Imported	Yes
6	White	iPhone	Imported	No
7	White	iPhone	Imported	Yes
8	White	iPhone	Domestic	No
9	Black	iPhone	Imported	No
10	Black	Android	Imported	Yes

Apply Naive Bayes Classification for a new test instance X = {Color =Black, Type= iPhone, and Origin = Domestic} and predict its label.

****