Code: 20DS501PC | SET-I | HT NO: 7R

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

B. Tech. V Semester Supply End Examinations, June-2023

Data Mining

Common to CSM & CSD

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

Note

i. This Question paper contains Part- A and Part- B.
ii. All the Questions in Part A are to be answered compulsorily.
iii. All Questions from Part B are to be answered with internal choice among them.

****

PART-A

10 X 02 = 20 Marks

			Marks	CO	BL
1.	a	What is a Data Warehouse?	2	CO1	L1
	b	List Data Mining Task Primitives.	2	CO1	L1
	c	How association rules mined from large databases?	2	CO2	L1
	d	Explain Correlation Analysis.	2	CO2	L2
	e	Compare Classification and Prediction	2	CO3	L2
	f	How do you evaluate the accuracy of a classifier?	2	CO3	L2
	g	Explain types of data used in Clustering?	2	CO4	L2
	h	Define dendrogram	2	CO4	L1
	i	What are the techniques used to mine streaming data?	2	CO5	L1
	j	Explain Spatial Data Mining.	2	CO5	L2

PART- B

5 X 10 = 50 Marks

			Marks	CO	BL
2.	a	Discuss in detail about the Data Mining Functionalities	5	CO1	L2
	b	Explain various data pre-processing techniques. How data reduction helps in data pre-processing	5	CO1	L2
		OR			
3	a	Explain the major issues in data mining.	5	CO1	L2
	b	Describe the problem of data quality with examples. Explain the usage of feature subset selection in data pre-processing	5	CO1	L2
4	a	State Apriori principle. Write apriori algorithm for frequent itemsets. Explain with an example	4	CO2	L2

	b	Apply the following transaction data set that shows few transactions and list of items using FP Growth Approach to find frequent itemset with min-support =3	6	CO2	

TID	Items
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}

OR

5	a	A database has the following five transactions. Let min_support = 80% and min_confidence = 20%. Find the Frequent item sets using Apriori Algorithm	6	CO2	L

TID	items_bought
T100	{K, A, D, B}
T200	{D, A, C, E, B}
T300	{C, A, B, E}
T400	{B, A, D }

	b	Describe in detail about Constraint based Association mining.	4	CO2	L2
6	a	How classification is defined as a two-step process. Explain in detail	4	CO3	L2
	b	Define Bayes Theorem? With an example, predict a class label using Naïve Bayesian Classification.	6	CO3	L2

OR

7	a	Write K-Nearest Neighbour classification algorithm and explain its characteristics	6	CO3	L2
	b	What are the advantages and disadvantages of decision trees over other classification methods?	4	CO3	L2
8	a	Explain K-Means algorithm for following problem instance. A1(2,10), A2(2,5), A3(8,4), A4(5,8), A5(7,5), A6(6,4), A7(1,2), A8(4,9).	6	CO4	L2
	b	What is Outlier? Explain Outlier detection techniques.	4	CO4	L2

OR

9	a	Describe how categorization of major clustering methods is being done.	4	CO4	L2
	b	Discuss DBSCAN algorithm used for clustering	6	CO4	L2
10	a	Demonstrate the application of data mining on Time-series data	5	CO5	L2

ect Code: 20CS501PC | SET-I | HT NO: 7 R

	b	Describe the functionalities of DB Miner-case and how it can be applied to text database	5	CO5	L2
		OR			
11	a	Discuss in detail about web mining. Explain the algorithms used in mining the structure and content of the web with suitable applications	5	CO5	L2
	b	Explain the procedure for Mining sequence patterns in Transactional databases	5	CO5	L2

CO : Course Outcomes

BL : Bloom's Taxonomy Levels L 1 : Remembering L 2 : Understanding L 3 : Applying L 4 : Analysing L 5 : Evaluating L 6 : Creating

*****

Code: 20CS501PC | SET-II | HT NO: | | | 7 | R | | | | | |

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

B. Tech. V Semester Regular End Examinations, Dec-2022

Data Mining

Common to CSM& CSD

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

Note

i. This Question paper contains Part- A and Part- B.
ii. All the Questions in Part A are to be answered compulsorily.
iii. All Questions from Part B are to be answered with internal choice among them.

****

PART-A

10 X 02 = 20 Marks

			Marks	CO	BL
1.	a	How data mining is differed from data warehousing?	2	CO1	L2
	b	What do you mean by Characterization in data mining functionalities?	2	CO1	L1
	c	What is graph pattern mining and write its types.	2	CO2	L1
	d	Define association rules.	2	CO2	L1
	e	How is decision tree used for classification?	2	CO3	L2
	f	What is pre-pruning and post-pruning?	2	CO3	L1
	g	Write the grid based methods of clustering	2	CO4	L2
	h	Write a short notes on agglomerative hierarchical clustering method	2	CO4	L2
	i	What do you mean by spatial data mining?	2	CO5	L1
	j	What kinds of associations can be mined in multimedia data?	2	CO5	L1

PART- B

5 X 10 = 50 Marks

			Marks	CO	BL
2.	a	How data mining is classified? Explain the classification.	5	CO1	L2
	b	Explain the data mining task primitives	5	CO1	L2
		OR			
3	a	Describe how integration of Data mining system with Data warehouse is achieved.	5	CO1	L4
	b	Discuss about Data reduction process in Data mining	5	CO1	L6
4	a	How will you mining the frequent itemsets without candidate generation?	4	CO2	L2
	b	Write the algorithm for FP Growth in pattern mining the frequent itemsets	6	CO2	L1
		OR			
5	a	Find the frequent itemset using Apriori algorithm for the following dataset with a minimum support 2.	5	CO2	L6

Trans ID	Items
10	A,C,D
20	B,C,E
30	A,B,C,E
40	B,E

	b	Write the methods of constraint based association mining and explain it.	5	CO2	L2
6	a	Write the Sequential covering algorithm and explain it.	6	CO3	L2
	b	Define the following terms. i. True Positive Rate ii. True Negative Rate iii. False Positive Rate iv. False Negative Rate	4	CO3	L1
		OR			
7	a	Write the Adaboost algorithm and identify its significance in Classification.	7	CO3	L3
	b	Write a short note on Random forest ensemble method.	3	CO3	L2
8	a	Write a detailed notes on Density based outlier detection	5	CO4	L2
	b	Explain Deviation based outlier detection	5	CO4	L2
		OR			
9	a	Write the DBSCAN algorithm and explain how it works in clustering.	6	CO4	L3
	b	Write the probabilistic hierarchical clustering algorithm.	4	CO4	L2
10	a	Explain the scalable methods for Mining Sequence Patterns in Transactional Databases.	6	CO5	L2
	b	How will you measure Precision & Recall for text retrieval?	4	CO5	L3
		OR			
11	a	Explain the approaches of similarity based retrieval in image databases, based on image signature.	5	CO5	L2
	b	Explain the text mining approaches based on the kind of data taken as input.	5	CO5	L2

CO : Course Outcomes

BL : Bloom's Taxonomy Levels **L 1: Remembering** **L 2: Understanding** **L 3: Applying** **L 4: Analysing** **L 5: Evaluating** **L 6: Creating**

*****

Subject Code: 20DS501PC	SET-II	HT NO:			7	R					

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

B. Tech. V Semester Regular &Supply End Examinations, January-2024

Data Mining

Common to CSM, CSD, AIML&CSG

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

Note

i. This Question paper contains Part- A and Part- B.
ii. All the Questions in Part A are to be answered compulsorily.
iii. All Questions from Part B are to be answered with internal choice among them.

****

PART-A

10 X 02 = 20 Marks

			Marks	CO	BL
1.	a	How data mining systems are classified.	2	CO1	L1
	b	Define Data Pre-Processing	2	CO1	L1
	c	Define Association Rule Measures	2	CO2	L1
	d	Explain Graph Pattern Mining	2	CO2	L2
	e	What is a Decision Tree?	2	CO3	L1
	f	Explain Lazy Learner Classification	2	CO3	L2
	g	Explain types of data used in Clustering?	2	CO4	L2
	h	List major Clustering Methods	2	CO4	L1
	i	What is Multimedia Data Mining?	2	CO5	L1
	j	Explain Precision and Recall measures.	2	CO5	L2

PART- B

5 X 10 = 50 Marks

			Marks	CO	BL
2.	a	What is Data Mining? Explain Data Mining Task Primitives	5	CO1	L2
	b	Explain the major issues in Data Mining.	5	CO1	L2
		OR			
3	a	Discuss in detail about the Data Mining Functionalities	5	CO1	L2
	b	What is Data Pre-Processing? Explain the various approaches used for Data Cleaning	5	CO1	L2
4	a	Define the terms frequent item sets, closed item sets and association rules.	4	CO2	L2

	b	Apply the following transaction data set that shows 6 transactions and list of items using Apriori Algorithm to find frequent itemset with min-support =2.	6	CO2	L3

T Id	List of items
001	I1,I3,I5,I7
002	I1,I5,I6,I7
003	I6,I7
004	I2,I3,I6,I7
005	I8,I1,I6
006	I1,I5,I8

OR

5	a	A database has the following five transactions. Let min_support = 60% and min_confidence = 80%.	6	CO2	L3

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 item sets, using FP-Growth.

	b	How can we mine multilevel Association rules efficiently using concept hierarchies? Explain in detail.	4	CO2	L2
6	a	What is classification? Explain Bayesian classification with suitable example	4	CO3	L2
	b	What is Decision tree? With an example, briefly describe the algorithm for generating decision tree.	6	CO3	L2

OR

7	a	What are Bayesian classifiers? With an example, describe how to predict a class label using naïve Bayesian classification.	6	CO3	L2
	b	Explain about classifier accuracy? Explain the process of measuring the accuracy of a classifier?	4	CO3	L2
8	a	Describe how categorization of major clustering methods is being done.	6	CO4	L2
	b	Explain K-Means algorithm for following problem instance. A1(2,10), A2(2,5), A3(8,4), A4(5,8), A5(7,5), A6(6,4), A7(1,2), A8(4,9).	4	CO4	L3

OR

9	a	Given two objects represented by the tuples (22,1,42,10) and (20,0,36,8): (a) Compute the Euclidean Distance between the two objects. (b) Compute the Manhattan distance between the two objects.	6	CO4	L3

		(c) Compute the Minkowski distance between the two objects, using q = 3.			
	b	What is Outlier? Explain about the Statistical-based outlier detection?	4	CO4	L2
10	a	Define Information retrieval. What are basic measures for text retrieval?	5	CO5	L2
	b	Discuss about mining time-series and sequence data.	5	CO5	L2
		OR			
11	a	Explain in detail about Spatial Data Mining.	5	CO5	L2
	b	Briefly discuss about mining the World Wide Web.	5	CO5	L2

CO : Course Outcomes

BL : Bloom's Taxonomy Levels **L 1 : Remembering** **L 2 : Understanding**

L 3 : Applying **L 4 : Analysing**

L 5 : Evaluating **L 6 : Creating**

*****