PRIYADARSHINI COLLEGE OF ENGINEERING, NAGPUR

Department: Computer Technology Semester: VI Section: A and B

CAT-II (2022-23)

Subject:Data Warehousing and Mining SubjectCode:**BTCT602T**

Duration:1.5Hrs Max.Marks:35

Note:

1) All questions are compulsory. 2)All questions carry marks as indicated.

3)Due credits will be given on neatness. 4)Draw diagram wherever it is necessary.

	Questions	Marks	CO	BT Level
Q.1	A. Which one of the following correctly defines the term cluster? a) Group of similar objects that differ significantly from other objects b) Symbolic representation of facts or ideas from which information can potentially be extracted c). Operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm d). All of the above	1M	CO3	1
	B. The learning which is used for inferring a model from labeled training data is called? a) Unsupervised learning b) Reinforcement learning c)Supervised learning d) Missing data imputation	1M	CO3	1
	C. Consider the following dataset consisting of the scores of two variables on each of seven subjects. Design K Means clustering for the data set for two clusters.	5M	CO3	3

Subject	A	B
1	1	1
2	15	2
3	3	4
4	5	7
5	3.5	5
6	4.4	5
7	3.5	4.5

OR

Q.2	A. Identify the example of Nominal attributeone. a) Temperature b. Mass c. Salary d.Gender	1M	CO3	1
	B. Which is needed by K-means clustering? a) defined distance metric b) number of clusters c. initial guess as to cluster centroids d. all of these	1M	CO3	1
	C. What do you mean by hierarchical clustering approach? Explain agglomerative and divisive hierarchical clustering.	5M	CO3	2
Q.3	A. A collection of one or more items is called as _____recovery. a) Support b) Itemset c) Confidence d) Support Count	1M	CO4	1
	B. When do you consider an association rule interesting? a) If it only satisfies min_support b)If it only satisfies min_confidence c)If it satisfies both min_support and min_confidence	1M	CO4	1

	d)There are other measures to check so			
	C. What is Frequent pattern mining and Association Rules? What is the use of both? Explain.	5M	CO4	2
	D. A database has five transactions. Let min sup = 2. TID items bought T1 {A, B, C, D, E,} T2 {B, C, D} T3 {B, C, D, E} T4 {A, B, C, D, E} T5 {B, C, D, E} Find all frequent itemsets using FP-growth Algorithm.	7M	CO4	3
	OR			
Q.4	A. Apriori algorithm works on the principle? a) If a rule is infrequent, its specialized rules are also infrequent b) If a rule is infrequent, its generalized rules are also infrequent c) Both a and b d)None of these	1M	CO4	1
	B. What is association rule mining? a) Same as frequent itemset mining b) Finding of strong association rules using frequent itemsets c)Both a and b d)None of these	1M	CO4	1
	C. Explain Market Basket Analysis for mining frequent pattern set and association rules with suitable example.	5M	CO4	2
	D. Explain Apriori Algorithm with an example in detail.	7M	CO4	2
Q.5	A. Web content mining describes the discovery of useful information from the _____contents. a) Text b) Web c) Page d) Level	1M	CO5	1
	B. In web mining, _____ is used to know the order in which URLs tend to be accessed. a) Clustering b) Association c)Sequential Analysis d)Classification	1M	CO5	1
	C. Explain Web Mining in detail.	4M	CO5	2
	D. Discuss i) Web Content Mining ii) Web Usage Mining	8M	CO5	2
	OR			
Q.6	A. _____ mining is concerned with discovering the model underlying the link structures of the web. a)Data Structure b) Web Structure c) Text Structure d) Image Structure	1M	CO6	1
	B. A link is said to be _____ link if it is between pages with the same domain name a) intrinsic b) transverse c) direct d) contrast	1M	CO6	1
	C. Describe various Graph properties of Web.	4M	CO6	5
	D. Discuss i) Web Structure Mining ii) Visual Web Data Mining	8M	CO6	2

**********All The Best**********