SRM Institute of Science and Technology
College of Engineering and Technology
School of Computing

Batch - 2

DEPARTMENT OF COMPUTING TECHNOLOGIES

SRM Nagar, Kattankulathur – 603203, Chengalpattu District, Tamilnadu

Academic Year: 2022-2023(ODD)

Test: CLAT-2 **Date:** 14.10.2022
Course Code & Title: 18CSE355T - Data Mining And Analytics **Duration:** 2 Periods
Year & Sem: III Year & 05th Semester **Max. Marks:** 50 Marks

Course Articulation Matrix: *(to be placed)*

S. No.	Course Outcome	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
1	CO2	3							3				
2	CO3		3						3				

Part – A
(10 x 1 = 10 Marks)
Answer all questions. The duration for answering the part A is 20 minutes (MCQ Answer sheet will be collected after 20 minutes)

Q. No	Question	Marks	BL	CO	PO	PI Code
1	Which of the following is not a frequent pattern mining algorithm? a) Apriori b) FP growth c) Decision trees d) Eclat	1	1	2	1	1.7.1
2	What does FP growth algorithm do? a) It mines all frequent patterns through pruning rules with lesser support b) It mines all frequent patterns through pruning rules with higher support c) It mines all frequent patterns by constructing a FP tree d) It mines all frequent patterns by constructing an item sets	1	1	2	1	1.7.1
3	You are a Data Scientist in an e-commerce company. You are analyzing all the transactions that happened over the past 1 week in your site. You observe that of the five hundred transactions that happened, two hundred of them had a mobile phone in them. What is the support for mobile phones in the last 1 week? a) 0.3	1	3	2	1	1.7.1

	b) 0.4 C) 0.5 d) 0.6					
4	How do you calculate Confidence (A -> B)? a) Support(A ∩ B) / Support (A) b) Support(A ∩ B) / Support (B) c) Support(A ∪ B) / Support (A) d) Support(A ∪ B) / Support (B)	1	2	2	1	1.7.1
5	What techniques can be used to improve the efficiency of apriori algorithm? a)Hash-based techniques b)Transaction Increases c)Sampling d)Cleaning	1	1	2	1	1.7.1
6	The problem of finding abstracted patterns in unlabeled dataset can be classified as _____ a) Supervised learning b) Unsupervised learning c) Hybrid learning d) Reinforcement learning	1	1	3	2	2.5.2
7	_____models continuous valued functions. a) Prediction b) Back Propagation c) Classification d) Data trends	1	1	3	2	2.5.2
8	_____is a statistical methodology that is most often used for numeric prediction a) Regression analysis b) Classification c) Class labels analysis d) decision tree classifiers	1	1	3	2	2.5.2
9	_____can be used to identify whether any two given attributes are statistically related. a) Relevance Analysis b) Regression Analysis c) Attribute subset selection d) Correlation analysis	1	1	3	2	2.5.2
10	Zero Probability value can be avoided using _____. a) Decision Trees b) If then Classification c) Laplacian smoothing d) Naïve Bayesian Classification	1	1	3	2	2.5.2

Part – B
(4 x 5 = 20 Marks)
Answer any 4 Questions

11	Consider the horizontal data format of the transaction database, D of a company. Show the transformed vertical data format. Mining can be performed on this data set by intersecting the TID sets of every pair of frequent single items. The minimum support count is 2. Because every single item is frequent in D. Table: I -Transactional Database 'D' for a company.	5	2	2	8	8.4.1

TID	LIST OF ITEM
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

12	What is Frequent Pattern Mining? Give example.	5	2	2	1	1.7.1
13	Bring out advantage of association rule mining in data mining.	5	2	2	1	1.7.1
14	Write algorithm for decision tree induction	5	2	3	2	2.6.4
15	Is clustering unsupervised or supervised classification? Give the reason for your answer.	5	2	3	2	2.6.4

Part – B
(2 x 10 = 20 Marks)

16	Compare FP growth and Apriori algorithm with suitable example?	10	3	2	1	1.7.1
	[OR]					
17	A database has five transactions. Let min support = 60% and min confidence = 80%. (a) Find all frequent item sets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining processes. (b) List all the strong association rules (with support s and confidence c) Matching the following metarule, where X is a variable representing customers, and item denotes variables representing items (e.g., "A," "B,"):	10	3	2	8	8.4.1

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}

	$\forall x \in$ transaction, buys(X, item1) $\wedge$ buys(X,item2) $\Rightarrow$ buys(X,item3) [s,c]					
18	Construct at least five decision tree from the dataset. Write 5 different rules derived from the constructed tree.	10	3	3	8	8.4.1

Outlook	Temperature	Humidity	Wind	Played football (Yes / No)
sunny	Hot	High	Weak	No
sunny	Hot	High	Strong	No
overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
overcast	Cool	Normal	Strong	Yes
sunny	Mild	High	Weak	No
sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
sunny	Mild	Normal	Strong	Yes
overcast	Mild	High	Strong	Yes
overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

	[OR]					
19	You are a data scientist which data mining task do you prefer under the following conditions. A) You are given with a dataset with 3 attributes. 1. Keyword, 2.Length of the document and 3. Spam or not. The attribute "keyword" has the values "accepted" and "Not accepted". Length of the document has the values "Less than 30" and "More than 30". B) A data table with 2 attributes Transaction Id and Items purchased. i) Justify the mining task chosen. ii) The algorithm you prefer to do the task. iii) The information which can be derived.	10	3	3	8	8.4.1

***Performance Indicators are available separately for Computer Science and Engineering in AICTE examination reforms policy.**

Course Outcome (CO) and Bloom's level (BL) Coverage in Questions