

B.Tech CSE 6th Semester (Supplementary), 2023

Paper Name : DATAMINING
Time : 3 Hours

Paper Code : BTCSE DED21
Maximum Marks : 75

SECTION A

(35 Marks)

All questions are compulsory in this Section. All Questions are of 5 Marks.

Q.N.	Questions	Marks	CO	BL
1	Explain the use of data mining in retail industry citing suitable examples.	7	CO1	L1
2	Define Support factor and Confidence factor with reference to Association rule mining.	7	CO1	L1
3	Explain the use of data mining in retail industry citing suitable examples.	7	CO2	L2
4	Discuss the demerits of Apriori technique and method used to improve it.	7	CO3	L3
5	Write the expression for Interestingness measure: lift. For the following data: (i) Compute lift(B,C) (ii) Compute lift(B, ^T C) (iii) find whether the values obtained in (i) & (ii) are positively or negatively correlated	7	CO3	L4

	Basketball	Not basketball	Sum (row)
Cereal	2000	1750	3750
Not cereal	1000	250	1250
Sum(col.)	3000	2000	5000

SECTION B

(40 marks)

Attempt **ANY FOUR** Questions from this Section.

All Questions are of 10 Marks.

Q.N.	Questions	Marks		
1	For the given Table Transaction dataset, generate rules using Apriori Algorithm. Consider the values for Support=22% and Confidence = 70%	10	CO5	L4

Transaction ID	Itemset
T1	I ₁ , I ₂ , I ₅
T2	I ₂ , I ₄
T3	I ₂ , I ₃
T4	I ₁ , I ₂ , I ₄
T5	I ₁ , I ₃
T6	I ₂ , I ₃
T7	I ₁ , I ₃
T8	I ₁ , I ₂ , I ₃ , I ₅
T9	I ₁ , I ₂ , I ₃

2	Construct a FP tree for the following dataset given in the table [minimum support=3]: <table><tr><td>Tid</td><td>Itemsets</td></tr><tr><td>T101</td><td>M, N, K, O, E, Y</td></tr><tr><td>T102</td><td>D, N, K, O, E, Y</td></tr><tr><td>T103</td><td>M, K, A, E</td></tr><tr><td>T104</td><td>M, C, K, U, Y</td></tr><tr><td>T105</td><td>C, O, I, K, E</td></tr></table>	Tid	Itemsets	T101	M, N, K, O, E, Y	T102	D, N, K, O, E, Y	T103	M, K, A, E	T104	M, C, K, U, Y	T105	C, O, I, K, E	10	CO2	L3								
Tid	Itemsets																							
T101	M, N, K, O, E, Y																							
T102	D, N, K, O, E, Y																							
T103	M, K, A, E																							
T104	M, C, K, U, Y																							
T105	C, O, I, K, E																							
3	(a) Differentiate between Classification and Clustering with examples. (b) Explain Linear Regression with example.	10	CO2	L4																				
4	(a) Differentiate between Supervised and Unsupervised learning. (b) Discuss Cluster analysis. Also write the steps involved in determining k-means algorithm.	10	CO3	L3																				
5	(a) Predict the food type to which tomato(sweet = 6, crunch = 4) belongs to using KNN algorithm for below given dataset. <table><tr><td>Item</td><td>Sweet</td><td>Crunch</td><td>Food type</td></tr><tr><td>Grape</td><td>8</td><td>5</td><td>Fruit</td></tr><tr><td>Beans</td><td>3</td><td>7</td><td>Vegetable</td></tr><tr><td>Nuts</td><td>3</td><td>6</td><td>Protein</td></tr><tr><td>Orange</td><td>7</td><td>3</td><td>Fruit</td></tr></table> (b) Apply k-means algorithm for dataset {2, 4, 10, 12, 3, 20, 30, 11, 25}. Given k=2.	Item	Sweet	Crunch	Food type	Grape	8	5	Fruit	Beans	3	7	Vegetable	Nuts	3	6	Protein	Orange	7	3	Fruit	10	CO2	L4
Item	Sweet	Crunch	Food type																					
Grape	8	5	Fruit																					
Beans	3	7	Vegetable																					
Nuts	3	6	Protein																					
Orange	7	3	Fruit																					
6	Write short notes on any TWO: (a) Artificial Neural Network (c) Data, Information & Knowledge (b) Genetic Algorithm (d) Lift and Correlation	10	CO1	L3																				