

B.Tech(CSE) – 6th Semester
Even Semester Examination 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 7 Marks.

Q.N.	Questions	Marks	CO	BL
1	Association rule mining often generates a large number of rules. Discuss effective methods that can be used to reduce the number of rules generated while still preserving most of the interesting rules.	5	CO1	L4
2	What do you mean by KDD process? Draw a clear flow diagram for it and explain it.	5	CO2	L2
3	Define Support factor and Confidence factor with reference to Association rule mining.	5	CO2	L2
4	Explain the use of data mining in retail industry citing suitable examples.	5	CO3	L2
5	Apply k-means clustering for the following data sets for 3 clusters: {1,2,6,7,8,10,15,17,20}	5	CO4	L3
6	Explain with reference to Data Warehouse: "Data inconsistencies are removed; data from diverse operational applications is integrated".	5	CO5	L2
7	Differentiate between Operational data and Decision Support data.	5	CO5	L2

SECTION B

(40 marks)

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

All Questions are of 10 Marks.

Q.N.	Questions	Marks	CO	BL																				
8	For the given Table 1. Transaction dataset, generate rules using Apriori Algorithm. Consider the values for Support = 22% and Confidence = 70%	10	CO4	L3																				
<table><tr><th>Transaction ID</th><th>Itemset</th></tr><tr><td>T1</td><td>I₁, I₂, I₅</td></tr><tr><td>T2</td><td>I₂, I₄</td></tr><tr><td>T3</td><td>I₂, I₃</td></tr><tr><td>T4</td><td>I₁, I₂, I₄</td></tr><tr><td>T5</td><td>I₁, I₃</td></tr><tr><td>T6</td><td>I₂, I₃</td></tr><tr><td>T7</td><td>I₁, I₃</td></tr><tr><td>T8</td><td>I₁, I₂, I₃, I₅</td></tr><tr><td>T9</td><td>I₁, I₂, I₃</td></tr></table>					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
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																							

9	(a) Find the weather prediction for the play(sunny, cool, high, true) considering the dataset given in table(Q.10) using Naïve Baye's classification. (b) Explain Linear Regression with example.					10	CO2	L6	
10	Construct a Decision Tree and find the classification rule for the weather data given in Table	Outlook	Temperature	Humidity	Windy	Play	10	CO6	L3
Sunny		Hot	High	False	No				
Sunny		Hot	High	True	No				
Overcast		Hot	High	False	Yes				
Rainy		Mild	High	False	Yes				
Rainy		Cool	Normal	False	Yes				
Rainy		Cool	Normal	True	No				
Overcast		Cool	Normal	True	Yes				
Sunny		Mild	High	False	No				
Sunny		Cool	Normal	False	Yes				
Rainy		Mild	Normal	False	Yes				
Sunny		Mild	Normal	True	Yes				
Overcast		Mild	High	True	Yes				
Overcast		Hot	Normal	False	Yes				
Rainy		Mild	High	True	No				
11	Predict the food type to which tomato(sweet = 6, crunch = 4) belongs to using KNN algorithm for below given dataset.					10	CO4	L3, L2	
		Item	Sweet	Crunch	Food type				
		Grape	8	5	Fruit				
		Beans	3	7	Vegetable				
		Nuts	3	6	Protein				
		Orange	7	3	Fruit				
12	Explain OLAP and OLTP. Differentiate between them and give the applications area where they are being used. Give the generalized architecture for Datawarehouse and explain its components.					10	CO2	L1, L4	
13	Write short notes on any TWO: a) Data Reduction b) SVM c) Genetic Algorithm c)) Data, Information & Knowledge					10	CO3	L2	