

Roll No.

Total Pages : 05

003603

May 2025

B. Tech. (Sixth Semester)

Data Mining (PEC-CS-D-601)

Time : 3 Hours

[Maximum Marks : 75

Note : It is compulsory to answer all the questions (1.5 marks each) of Part A in short. Answer any *four* questions from Part B in detail. Different sub-parts of a question are to be attempted adjacent to each other. Use of Scientific Calculator is allowed.

Part A

1. (a) List some features of Data Warehousing which make it different from Information Processing ? 1.5
- (b) What is the difference between supervised and unsupervised learning in data mining ? 1.5
- (c) Discuss bitmap indexing in data warehouse. 1.5

- (d) What is the difference between Manhattan Distance and Euclidean Distance ? 1.5
- (e) Discuss the importance of support and confidence. 1.5
- (f) What is meant by concept hierarchy ? Explain its need. 1.5
- (g) Define Prior Probability, Conditional Probability and Posterior Probability in context of Bayes theorem. 1.5
- (h) You are given a cluster containing the following 3 data points in 2D space : 1.5
 $P1 = (2, 3)$, $P2 = (4, 5)$, $P3 = (6, 7)$
 Calculate the Clustering Feature (CF) for this cluster.
- (i) Explain the process of Crossover and Mutation in Genetic algorithms. 1.5
- (j) List some challenges in mining data streams. 1.5

Part B

2. (a) Explain in detail the three-tier architecture of Data warehouse. 7

(b) What are different schemas supported by a data warehouse ? Explain in detail with the help on an example scenario. 8

3. (a) Find the frequent itemsets and generate association rules on the following transactional database. Assume that minimum support count is 2 and minimum confidence is 60%. Use A-Priori or FP-Growth method : 10

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	Hot Dogs, Coke, Chips

(b) Differentiate between ROLAP, MOLAP and HOLAP servers. 5

4. (a) What are Decision trees ? How they assist in classifying data ? Calculate Information Gain for attributes Weather and Temperature using

following database with class label attribute

Play :

10

Weather	Temperature	Play
Sunny	Hot	No
Sunny	Hot	No
Overcast	Hot	Yes
Rainy	Mild	Yes
Rainy	Cool	Yes
Rainy	Cool	No
Overcast	Cool	Yes
Sunny	Mild	No

- (b) What is sequential pattern mining ? Explain in brief. 5
5. (a) What is backpropagation in neural networks ? Explain classification by backpropagation by giving mathematical formulation of forward propagation, loss calculations and backward propagations. 7
- (b) For the given data of 7 points in 2D space with coordinates (A,B), compute 2 clusters using K-means algorithm for clustering where

initial cluster centers are (1.0, 1.0) and (5.0, 7.0). Execute for two iterations and find the next centroids. 8

Points	A	B
R1	1.0	1.0
R2	1.5	2.0
R3	3.0	4.0
R4	5.0	7.0
R5	3.5	5.0
R6	4.5	5.0
R7	3.5	4.5

6. (a) Explain Trend Analysis in time-series data with all its components. 7
- (b) What is Hierarchical Clustering and its types? Describe single linkage, complete linkage and average linkage methods used in hierarchical clustering. 8
7. (a) Describe class imbalance problem and its solutions. 5
- (b) What is web mining? What are its categories? 5
- (c) Write a short note on Social Network Analysis. 5

