017606

May, 2023
B.Tech. (EEIOT) - VI Semester
Data Mining (PEC-CS-DS-601)

Time: 3 Hours   **Max. Marks:75**

Instructions: 1. It is compulsory to answer all the questions (1.5 marks each) of Part -A in short.

2. Answer any four questions from Part -B in detail.

3. Different sub-parts of a question are to be attempted adjacent to each other.

PART -A

Q1 (a) Give any three applications of Datamining   (1.5)

(b) What are the advantages of DBSCAN over k-Means clustering algorithm?   (1.5)

(c) Suppose a data warehouse consists of three measures customer, account and   (1.5)
branch and two measures count (number of customers in the branch) and
balance. Draw the schema diagram using star schema.

(d) Give any two issues related to classification.   (1.5)

(e) "Frequent itemset mining in data streams is a challenging task",   (1.5)
Explain.

(f) Give reason for the following statement, "Data preprocessing is an important (1.5)
step of Datamining".

(g) What is a data cube?   (1.5)

(h) Define the term Frequent Itemset.   (1.5)

(i) Differentiate between Offspring and Crossover.   (1.5)

(j) What is spatial datamining.   (1.5)

PART -B

Q2 (a) Explain frequent subgraph mining using Apriori method.   (10)
(b) Explain the three tier architecture of Data warehouse.   (5)

Q3 (a) The following table shows the midterm and final exam grades obtained for students in a database course

X (Mid Term)	Y(Final Exam)
72	84
50	63
81	77
74	78
94	90
86	70
60	85
75	90
55	80
65	74

i) Use.the method of least squares to find an equation for the prediction of a student's final exam grade based on the student's midterm grade in the course. (10)

ii) Predict the final exam grade of a student who received 86 marks on the midterm exam with the above model. (5)

Q4 (a) Explain k-medoids partitioning method along with its advantages and disadvantages. (10)

(b) Explain major components for characterizing time-series data. (5)

Q5 (a) Given two objects represented by the tuples (26, 1, 45, 10) and (20, 0, 36, 4): (5)

(i) Compute the Euclidean distance between the two objects.
(ii) Compute the Manhattan distance between the two objects.

(b) Differentiate between (10)
- ROLAP, MOLAP and HOLAP

Q6 (a) Explain the characteristics of Social Networks. (5)

(b) The transaction details are given in the following table, what is the confidence and support of the association rule {Nuts} ⇒ {Egg, Milk}? (5)

TID	T1	T2	T3	T4	T5
Items	Beer, Nuts, Pencil	Nuts, Egg, Milk	Egg, Milk	Beer, Nuts, Pencil, Milk ,Egg	Nuts, Egg, Milk, Salt

(c) Using the above transaction table draw FP Growth tree. (5)

Q7 Write short notes on: (15)
- KDD
- Random Sampling
- Decision Tree Induction