Roll No. _____ [Total No. of Pages : 2

5E1821

5E1821

B.Tech. V Sem. (Main & Back) Examination, January/February - 2024

Artificial Intelligence and Data Science

5AID3-01 Data Mining -Concepts and Techniques

AID, CAI, CDS

Time : 3 Hours **Maximum Marks : 70**

Instructions to Candidates:

Attempt all Ten questions from Part A, five question out of Seven from Part B and Three questions out of five questions from Part C.

Schematic diagrams must be shown wherever necessary. Any data you feel missing suitably be assumed and stated clearly. Units of quantities used/ calculated must be stated clearly.

Use of following supporting material is permitted during examination. (Mentioned in form No.205)

PART - A

(Answer should be given up to 25 words only)

ALL questions are Compulsory. **(10×2=20)**

1. What is concept description?
2. Define Data characterization.
3. What is the need for preprocessing the data?
4. List the important issues that have to be addressed during data integration.
5. Define Data mining as KDD (knowledge Discovery in Databases).
6. Write the strategies for data reduction.
7. List some variants of Apriori Algorithm.
8. What is Clustering?
9. State two clustering methods that are used in "Grid and density based" method.
10. List some applications of Data Mining.

PART - B

(Analytical/Problem Solving questions)

Attempt any FIVE questions. **(5× 4 = 20)**

1. Explain the differences between "explorative data mining" and "Predictive data mining" and give one example of each.
2. How concept Hierarchies are useful in data mining?
3. What do you mean by market basket analysis and how it can help in a supermarket?
4. Give brief description of the following:
   i) Binning
   ii) Regression.
5. What is the drawback of k-means algorithm? How can we modify the algorithm to diminish that problem?
6. Explain how data mining helps in detection of money Laundering and other financial crimes?
7. Compare SCAN with DBSCAN. What are their similarities and differences?

PART - C

(Descriptive/Analytical/Problem Solving/Design question)

Attempt any THREE questions. **(3× 10 = 30)**

1. Explain data mining as kDD process. Describe various functionalities of data mining.
2. Write and explain the algorithm for mining frequent item sets without candidate generation. Give relevant example.
3. What is Bayesian belief network? Explain in detail.
4. a) Classify various clustering methods.
   b) Explain any one partitioning based clustering methods.
5. What are various applications of data mining in Science and engineering? Describe various challenges of emerging scientific applications of data mining.
