

END TERM EXAMINATION

SIXTH SEMESTER [BCA] JUNE 2024

Paper Code: BCA-302 Subject: Data Warehousing and Data Mining

(Batch-2021 Onwards)

Time: 3 Hours

Maximum Marks: 75

Note: Attempt five questions in all including Q. no.1 which is compulsory. Select one question from each unit.

TID	List of items
001	Pencil, sharpener, eraser, color papers
002	Color papers, charts, glue sticks
003	Pencil, glue stick, eraser, pen
004	Oil pastels, poster colours, correction tape
005	Whitener, pen, pencil, charts, glue stick
006	Colour pencils, crayons, eraser, pen

Find all frequent item sets using Apriori algorithm. List all the strong association rules.

UNIT-IV

- Q8 (a) How the data mining will be used in the telecommunication industry? (6.5)
 (b) How can one handle suspicious or missing data in a dataset while performing the analysis? (6)

OR

- Q9 (a) What are the techniques used in data reduction? Explain one of them in detail. (6.5)
 (b) Discuss the application of Data Mining in Retail Industry. (6)

UNIT-I

- Q1 Attempt any five of the following questions: (5x5=25)
 (a) Explain different data mining tasks.
 (b) Differentiate between designing a data warehouse and an OLAP cube.
 (c) Explain the Issues regarding Classification and Prediction.
 (d) Why data mining queries are more helpful? Explain.
 (e) Differentiate between Principal Component Analysis and Factor Analysis.
 (f) Explain Association Algorithm in Data Mining.
 (g) What are the differences between OLAP and OLTP?
 (h) Why Fuzzy logic is an important area for Data Mining?

UNIT-II

- Q2 (a) Explain three tier architecture of data ware housing. (6.5)
 (b) What are the most common transformations in ETL processes? (6)
- OR
- Q3 What are the various components of data warehouse? Explain their functionality in detail. (12.5)

UNIT-III

- Q4 What is the significance of OLAP in data warehouse? Describe OLAP operations. (12.5)

OR

- Q5 (a) Make comparisons between the MOLAP and HOLAP. (6.5)
 (b) Discuss the star and snowflake schema in detail with suitable example. (6)

UNIT-III

- Q6 Explain different data mining tasks for knowledge discovery. (12.5)
 OR

Q7 A database has six transactions. Let min-sup = 50% and min-conf = 75%.

BCA-302
P.1/2

BCA-302
P.1/2

END TERM EXAMINATION

SIXTH SEMESTER [BCA] JUNE 2024

Paper Code: BCA-302

Subject: Data Warehouse and Data Mining
(Upto-2020 Batch)

Time: 3 Hours

Maximum Marks: 75

Note: Attempt five questions in all including Q. no.1 which is compulsory. Select one question from each unit.

- Q1 Attempt **any Five** from the following (5x5=25)
- (a) Define the three-tier architecture of a data warehouse and explain the role of each tier in the overall system.
 - (b) Discuss the significance of metadata in a data warehouse environment and how it facilitates data management and retrieval.
 - (c) Explain the STAR schema and the Snowflake schema in dimensional modeling. Compare and contrast these two schemas, highlighting their advantages and disadvantages.
 - (d) What are the major features and functions of OLAP?
 - (e) Describe the process of data preprocessing in data mining?
 - (f) Explore the applications of data mining in retail and telecommunication industries?
 - (g) Discuss the role of open-source data mining tools such as Weka, RapidMiner, and IBM Watson in the analysis of datasets?

UNIT-I

- Q2 (a) How does the data warehouse architecture differ from that of a traditional database system? Discuss the specific components and their roles in a data warehouse environment. (6)
- (b) Explain the concept of dimensional analysis in defining business requirements for a data warehouse. What methods are commonly used for gathering requirements, and why is dimensional analysis important? (6.5)
- Q3 (a) Discuss the role of ETL tools in the data warehousing process. Provide examples of how these tools facilitate data integration, transformation, and loading tasks. (6)
- (b) What are the challenges associated with integrating a data mining system with a database or data warehouse system? How can these challenges be mitigated to ensure seamless integration? (6.5)

UNIT-II

- Q4 (a) Define dimensional modeling and its objectives in the context of designing a data warehouse schema. How does dimensional modeling help in translating business requirements into data design? (6)
- (b) Compare and contrast the STAR schema, Snowflake schema, and fact constellation schema. Provide examples of scenarios where each schema would be appropriate. (6.5)

P.T.O.

BCA-302
P_{1/2}

- Q5 (a) Explore the demand for OLAP in data warehousing. What are the limitations of other analysis methods, and how does OLAP address these limitations? **(6)**
- (b) Discuss the major OLAP operations such as drill-down, roll-up, slice-and-dice, and pivot or rotation. How do these operations support interactive data analysis in a data warehouse environment? **(6.5)**

UNIT-III

- Q6 (a) What are the key considerations in selecting the appropriate data mining technologies for a given application? Discuss the types of data that can be mined and the patterns that can be discovered. **(6)**
- (b) Provide an overview of data preprocessing techniques in data mining. How does data cleaning, integration, reduction, transformation, and discretization contribute to preparing data for analysis? **(6.5)**
- Q7 (a) Explain the significance of the Apriori algorithm for association rule mining. How does this algorithm identify frequent itemsets and generate association rules from transactional datasets? **(6)**
- (b) Discuss the major issues in data mining, including scalability, data quality, and privacy concerns. How can these issues be addressed to ensure successful data mining operations? **(6.5)**

UNIT-IV

- Q8 (a) Explore the various applications of data mining in industries such as retail and telecommunications. How does data mining contribute to improving business operations and decision-making in these domains? **(6)**
- (b) Introduce the concept of data mining tools, focusing on open-source options like Weka, RapidMiner, and IBM Watson. How do these tools facilitate classification and clustering algorithms, particularly when applied to datasets like IRIS? **(6.5)**
- Q9 (a) Discuss the role of recommender systems in data mining applications. How do recommender systems analyze user behavior and preferences to provide personalized recommendations? **(6)**
- (b) Investigate the social impact of data mining. What ethical considerations should be taken into account when collecting and analyzing data, and how can data mining be used responsibly to benefit society? **(6.5)**

BCA-302
P2/2