2/21

# 0105N115

Candidate's Seat No :_____

**MSc Sem.-2 Examination**

**407**

**CS-DW & DM**

**May-2025**

**Time : 2-30 Hours]** **[Max. Marks : 70**

Q:1 (A) Explain the need for reporting and analyzing data. How does raw data become valuable information? Describe the lifecycle of data in the context of data warehousing. (7)

Q:1 (B) Differentiate between a data warehouse and a data mart with examples. What is metadata in a data warehouse? Explain its types and roles. (7)

OR

Q:1 (A) Compare OLTP and OLAP systems in terms of architecture, performance, and application. (7)

Q:1 (B) Describe the star schema with a neat diagram. How is it used for multidimensional analysis? (7)

Q:2 (A) Describe the types of data suitable for data mining with relevant examples. Explain the various functionalities of data mining and the patterns it can discover. (7)

Q:2 (B) What is the concept description? How is data summarization useful in knowledge discovery? Explain the role of data generalization in concept description. (7)

OR

Q:2 (A) Describe association rule mining. What is market basket analysis? And Explain the Apriori algorithm for frequent itemset generation with an example. (7)

Q:2 (B) Discuss the various applications of classification in data mining. And Explain the process of preparing data for classification and prediction. (7)

Q:3 (A) Compare Information Gain, GINI Index, Entropy, and Classification Error as attribute selection measures. (7)

Q:3 (B) Explain direct and indirect methods for building classification rules.also Compare common classification algorithms. (7)

OR

(P.T.O)

MIS-2

| | | |
|---|---|---|
| Q:3 (A) | Define clustering and explain its real-world applications. and Explain different types of data variables used in clustering. | (7) |
| Q:3 (B) | Describe the k-means algorithm and compare it with the k-medoids method. | (7) |
| Q:4 (A) | Explain linear regression and how it is used for data prediction. Also Describe different techniques used for outlier analysis. | (7) |
| Q:4 (B) | What is deviation-based outlier detection? How is it different from statistical-based Methods? Explain data characterization and generalization in conceptual data mining. | (7) |

OR

| | | |
|---|---|---|
| Q:4 (A) | What is Big Data? Explain the Four Vs and the need for Big Data analytics. Also Describe the Hadoop architecture and its core components. | (7) |
| Q:4 (B) | Explain the function of HDFS and its key components: NameNode, DataNode, and Secondary NameNode. | (7) |

Q:5 True/False Attempt any seven out of Twelve (Each carries 2 Marks) (14)

1 Data marts are smaller, more focused versions of data warehouses.

2 Raw data requires processing and transformation to become valuable information in data warehousing.

3 Metadata in a data warehouse refers to the data that describes other data within the system.

4 OLAP systems are primarily designed for analytical querying and decision support, not transaction processing.

5 A snowflake schema in OLAP reduces redundancy by normalizing data into multiple related tables.

6 Data mining can be applied to both structured and unstructured data, not just structured data.

7 The Apriori algorithm is used for association rule mining and identifies frequent item sets.

8 Classification is a supervised learning technique used for predicting categorical outcomes.

9 Clustering is a type of unsupervised learning used to group similar data points.

10 Pruning in decision trees is used to reduce overfitting and improve generalization.

11 The k-means algorithm assigns data points to the nearest cluster centroid.

12 Clustering can be applied to both numerical and categorical data types, though techniques differ.