R20

MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY
(Autonomous Institution – UGC, Govt. of India)
III B.Tech I Semester Supplementary Examinations, June/July 2024
Data Warehousing and Data Mining
(CSE-IOT & B.Tech-AIML)

Roll No										

Time: 3 hours   **Max. Marks: 70**

Note: This question paper Consists of 5 Sections. Answer **FIVE** Questions, Choosing ONE Question from each SECTION and each Question carries 14 marks.

***

SECTION-I

1 **A** Analyse the scalability and performance considerations between operational database systems and data warehouses. What factors contribute to their scalability, and how do they handle large volumes of data differently? **[7M]**

   B Compare and contrast different ETL (Extraction, Transformation, Loading) tools available in the market. Discuss their strengths, limitations, and considerations for choosing an appropriate tool for a specific data warehouse environment. **[7M]**

OR

2 **A** Discuss real-world use cases where the choice between a Star Schema and a Snowflake Schema significantly impacts the efficiency and effectiveness of data analysis and reporting. **[7M]**

   B Compare and contrast the architectural differences between ROLAP (Relational OLAP), MOLAP (Multidimensional OLAP), and HOLAP (Hybrid OLAP) servers. Highlight the strengths and limitations of each architecture. **[7M]**

SECTION-II

3 **A** Analyse the challenges associated with data quality in data mining. Explain how noise, missing values, and inconsistencies impact the accuracy and reliability of data mining results. **[7M]**

   B Define and discuss the fundamental Data Mining Task Primitives: Classification, Clustering, Association Rule Mining, and Sequential Pattern Mining. Provide examples illustrating each primitive's application. **[7M]**

OR

4 **A** Compare and contrast data integration and data transformation as essential steps in data preprocessing. How do these processes contribute to unifying disparate data sources and improving data quality? **[7M]**

   B Discuss the implications of scalability when preprocessing large datasets. How do preprocessing techniques adapt to handle big data efficiently? **[7M]**

SECTION-III

5 **A** Discuss the APRIORI principle in Association Rule Mining. Explain the significance of the downward closure property and how it contributes to efficient frequent item set generation. **[7M]**

   B Describe the APRIORI algorithm in detail. Discuss its steps and the strategy it employs to efficiently discover frequent item sets in large datasets. **[7M]**

OR

6	A	Discuss the advantages and limitations of the FP-Growth algorithm compared to the APRIORI algorithm. In what scenarios would one algorithm be preferred over the other?	**[7M]**
	B	Discuss the relationship between Closed Frequent Item Sets and Maximal Frequent Item Sets. Compare and contrast these concepts, highlighting their distinctive characteristics and applications.	**[7M]**

SECTION-IV

7	A	Explain the process of evaluating classifiers in the context of classification models. Discuss different evaluation metrics used to assess classifier performance and their significance in model selection.	**[7M]**
	B	Discuss the methods for expressing attribute test conditions in Decision Tree construction. How do different approaches, such as binary, multiway, or continuous attribute representation, impact tree structure and accuracy?	**[7M]**

OR

8	A	Define Bayesian Belief Networks (BBNs) and their application in probabilistic modelling. How do BBNs represent dependencies among variables using graphical structures?	**[7M]**
	B	Describe the K-Nearest Neighbour (K-NN) classification algorithm. How does it make predictions based on similarity measures and the majority voting principle?	**[7M]**

SECTION-V

9	A	Explain the differences between exclusive and overlapping clustering techniques. Provide examples illustrating scenarios where each type of clustering is advantageous.	**[7M]**
	B	Explain the Partitioning Around Medoids (PAM) algorithm. How does PAM differ from K-Means in its approach to finding representative points for clustering?	**[7M]**

OR

10	A	Discuss the strengths of Hierarchical Clustering compared to other clustering methods. Highlight its advantages in capturing cluster relationships and providing a visual hierarchy of clusters.	**[7M]**
	B	Describe the concept of outlier detection in the context of Hierarchical Clustering. How do outliers impact the cluster formation and dendrogram structures?	**[7M]**

***