Sl. No.

B3.E4-R5 : DATA WAREHOUSING AND DATA MINING

NOTE :
1. **Answer question 1 and any FOUR questions from 2 to 7.**
2. **Parts of the same question should be answered together and in the same sequence.**

Total Time : 3 Hours   **Total Marks : 100**

---

1. (a) What is Data mining ? What are the major issues in Data Mining ?

    (b) What is Cuboid ? Explain any three OnLine Analytical Processing (OLAP) Operations on Data Cube with example.

    (c) Compare Extract, Transform, Load (ETL) and Extract, Load, Transform (ELT) processes.

    (d) What is Concept Hierarchy ? List and briefly explain types of Concept Hierarchy.

    (e) Explain KDD process with the help of Diagram.

    (f) Differentiate between OnLine Transactional Processing (OLTP) and OnLine Analytical Processing (OLAP).

    (g) Calculate covariance and relationship between x and y for the following data set :

    x : 2.1, 2.5, 3.6, 4.0

    y : 8, 10, 12, 14

    (i) What is the covariance for given data ?

    (ii) What is the correlation between x and y variables ?   **(7x4)**

2. (a) Differentiate Data mining and Data Warehousing.

    (b) Explain Star and Snowflake Schema for Multidimensional Databases.

    (c) Explain Extract, Transform, Load (ETL) process in Data Warehouse.   **(6+6+6)**

3. (a) Why naïve Bayesian classification is called "naïve"? Describe naïve Bayesian classification with an example.

    (b) Explain Linear regression with an example.   **(10+8)**

4. (a) What is Data pre-processing ? Explain procedure of Data Reduction technique of pre-processing with an example.

    (b) Draw and Explain Data warehouse architecture.

    (c) What is classification and prediction ? List out issues regarding classification and prediction.   **(6+6+6)**

---

5. (a) What do you mean by multilevel association rules ? Discuss different variations to this approach.

(b) Find all frequent itemsets using the Apriori Algorithm for the following set of transactions. Consider support threshold s = 33.34% and confidence threshold c = 60%.

Transaction ID	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

(10+8)

6. (a) Explain Hierarchical Clustering with the help of an example.

(b) Define Text mining. Explain any 2 application of Text mining. What are the issues in text mining ? **(10+8)**

7. (a) What is Web Content Mining ? List and explain the web content mining techniques using agents and databases. List out the application of web content mining.

(b) What is outlier analysis ? Why outlier mining is important ? Briefly describe the different approaches: statistical-based outlier detection, distance-based outlier detection and deviation-based outlier detection.

(10+8)

- o O o -