No. of Printed Pages : **2**

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) Why does every structure in the data warehouse contain the time element ?

(b) Selection of an appropriate ETL Tools is an important decision. Discuss why ?

(c) What is graph mining ? Briefly discuss.

(d) Relational OLAP (ROLAP) is one of the OLAP servers. Briefly discuss it.

(e) How to select the value of K in the K-Nearest Neighbor Algorithm ?

(f) Discuss the limitations of Apriori algorithm.

(g) Discuss the major advantages of star schemas in a decision-support environment. **(7×4)**

2. (a) Briefly discuss and differentiate spatial data mining and temporal data mining.

(b) What is multi-dimensional data model ? Through an example, show how the data in 2D table is represented using 3D data cube.

(c) Suppose that a data warehouse consists of the four dimensions, date, spectator, location, and game, and the two measures, count and charge, where charge is the fare that a spectator pays when watching a game on a given date. Spectators may be students, adults, or seniors, with each category having its own charge rate. Draw a star schema diagram for the data warehouse. **(8+5+5)**

3. (a) Highlight the advantages and disadvantages of data warehousing and data mining. Also, on different basis, give the comparison between data warehousing and data mining.

(b) Briefly discuss each step of the ETL process. Also highlight the advantages and disadvantages of the ETL process in data warehousing. **(9+9)**

4. (a) Roll-up, drill-down, slice, dice, and pivot are five basic analytical operations that can be performed on an OLAP cube. Illustrate the working of these operations through examples.

(b) List and describe the five primitives for specifying a data mining task. **(10+8)**

5. (a) What is density-based clustering ? Briefly discuss the working of density-based clustering and highlight its major features.

(b) Find Frequent, Closed Frequent and Maximal Frequent Itemsets using Apriori algorithm for the following database with 5 transactions. Minimum support is 2.

TID	Items
1	ABC
2	ABCD
3	BCE
4	ACDE
5	DE

(9+9)

6. (a) How does the Decision Tree algorithm Work ? Briefly discuss. Also highlight the strengths and weaknesses of the Decision Tree approach.

(b) Discuss the advantages and disadvantages of Web usage Mining.

(c) Are data mining and data warehousing related ? Discuss. **(8+5+5)**

7. (a) What is Snowflake schema ? Diagrammatically present a snowflake schema with two dimensions, each having three levels. Also discuss the advantages and disadvantages of snowflake schema.

(b) How ELT is different than ETL ? Discuss.

(c) Highlighting the advantages and disadvantages of KDD, differentiate between KDD and Data Mining over various parameters. **(6+4+8)**

- o O o -