

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY

M. E. - SEMESTER – II • EXAMINATION – WINTER • 2014

Subject code: 1722304

Date: 04-12-2014

Subject Name: Data Warehouse and Data Mining

Time: 02:30 pm - 05:00 pm

Total Marks: 70

Instructions:

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

- Q.1 (a) Discuss the major issues in Data mining. 07
(b) Define Data Mart, Meta Data. Discuss the performance issues in Data mart. 07
Q.2 (a) Give architecture of data warehouse. Compare OLTP and OLAP systems based on various features. 07
(b) Discuss Arbor Essbase Web, Microstrategy DSS Web in detail 07

OR

- Q.2 (b) Explain types of OLAP servers. Explain various data transformation techniques in brief. 07
Q.3 (a) How might the efficiency of Apriori be improved? Discuss with example. 07
(b) A data warehouse is required to build for a telecommunication company to model customer billing data. Each month, the operational billing system generates a bill for each phone number also known as service line. Each service line is associated with a single customer. Each service line has its own set of billing metrics, such as the number of calls, number of minutes used and monthly service charge, Roaming charge, long-distance charge. There is a single rate plan associated with each service line on a given bill. Design multidimensional data warehouse schema to model customer billing data. 07

OR

- Q.3 (a) Define Data warehouse and compare the following concepts: 07
1. Enterprise data warehouse, virtual data warehouse
2. DBMS, KDD and data mining
(b) Discuss the method that mines the complete set of frequent itemsets without candidate generation. 07
Q.4 (a) A database has four transactions. Let min_sup=50% and min_conf=80% 07

TID	DATE	ITEM_BOUGHT
T100	10/15/99	{K,A,D,C,B}
T200	10/15/99	{D,A,C,B}
T300	10/19/99	{C,A,B,E}
T400	10/22/99	{B,A,D,E,K}

1. Find all frequent itemsets using Apriori and FP-growth, respectively
 2. List all the strong association rules
- (b) Discuss Web Usage Mining. 07

OR

- Q.4 (a) Discuss the centroid base and object-based partitioning technique 07
(b) Discuss Web Content mining. Define Ontology. 07
Q.5 (a) How does the BIRCH algorithm work? How effective is BIRCH? 07
(b) Discuss the GSP algorithm. Discuss basic concepts of temporal data mining 07

OR

- Q.5 (a) Compare prepruning vs. postpruning with decision tree induction. 07
(b) Discuss any two approaches for density based clustering method. 07
