

END TERM EXAMINATION

SIXTH SEMESTER [BCA] MAY-JUNE-2016

Paper Code: BCA-302

Subject: Data Warehouse and Data Mining

Time : 3 Hours

Maximum Marks :75

Note: Attempt any six questions including Q.no.1 which is compulsory.

- Q1
- (a) Explain in brief how the evolution of database technology led to data mining?
 - (b) Write the name of steps involved in data mining when viewed as a process of knowledge discovery.
 - (c) Mention the name of the databases and information repositories on which data mining can be performed.
 - (d) How does classification work in data mining? How is (numeric) prediction different from classification?
 - (e) Mention the criteria for the comparison and evaluation of classification methods during data mining.
 - (f) Suppose that the data for analysis includes the attribute *age*. The *age* values for the data tuples are (in increasing order) 13, 21, 22, 25, 25, 30, 30, 33, 35, 35, 35, 36, 40, 52, 70. Estimate the *mean* of the data? Find the first quartile (Q1) and the third quartile (Q3) of the data? Give the *five-number summary* of the data.
 - (g) What is the use of summary tables in data warehouse?
 - (h) What can we do to secure the privacy of individuals while collecting and mining data?
 - (i) Suppose a group of 12 *sales price* records has been sorted as follows: 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215. Partition them into three bins by each of the following methods: (i) equal-frequency (equal-depth) partitioning OR Equal-width partitioning. (ii) Clustering
 - (j) Explain the use of meta data in data warehouse. **(2.5x10=25)**
- Q2
- Imagine that you need to analyze *All Electronics* sales and customer data (Data related to the sales of electronic itmes). You note that many tuples have no recorded value for several attributes, such as customer *income*. How can you go about filling in the missing values for this attribute? Explain some of the methods to handle the problem. **(10)**
- Q3
- Describe the major issues in data mining regarding mining methodology, user interaction, performance, and diverse data types in detail. **(10)**

P.T.O.

BCA-302
P.1/2

- Q4 Define Data warehouse. What are the features which distinguish data warehouses from other data repository systems, such as relational database systems, transaction processing systems, and file systems? **(10)**
- Q5 How do data warehousing and OLAP relate to data mining? Briefly compare between OLTP and OLAP systems from the following perspective: (i) Users and system orientation, (ii) Data contents, (iii) Database design. Draw a figure for Star schema and Snowflake schema of a data warehouse (Consider any data warehouse of your choice) for sales records. **(10)**
- Q6 Consider a database has five transactions. Let $min\ sup = 60\%$ and $min\ con\ f = 80\%$.

TID	Items bought
T100	{M,O,N,K,E,Y}
T200	{D,O,N,K,E,Y}
T300	{M,A,K,E}
T400	{M,U,C,K,Y}
T500	{C,O,O,K,I,E}

- Find all frequent itemsets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining processes. **(10)**
- Q7 Describes the major issues during preprocessing the data for classification and prediction. **(10)**
- Q8 Write the name of the types of data that often occur in cluster analysis and how to preprocess them for such an analysis? **(10)**
- Q9 Discuss the applications of data mining in following:- **(10)**
- Retail Industry
 - Telecommunication Industry
 - Biological Data Analysis

BCA-302
P2/2