

MCA-501(N)

M. C. A. (Fifth Semester)
EXAMINATION, Nov.-Dec., 2007

(New Course)

DATA MINING AND WAREHOUSING

[MCA-501(N)]

Time : Three Hours

Maximum Marks : 100

Minimum Pass Marks : 40

Note : Attempt all questions. Attempt any two parts in each question.

1. (a) What do you understand by Data mining and knowledge discovery ? Explain how the evolution of database technology led to data mining. 10
 - (b) Define the following data mining functionalities: characterization, discrimination, association and correlation analysis, classification, prediction and clustering. Give example of each using a real life database. 10
 - (c) Discuss major issues in data mining regarding mining methodology, user interaction, performance and diverse data types. 10
2. (a) What is a data warehouse ? How it is different from an operational database ? Explain data marts. 10

- (b) Discuss 3-tier data warehousing architecture and explain ROLAP, MOLAP and HOLAP servers. 10
 - (c) Compare the following : 5 each
 - (i) Star schema and Snow flake schema.
 - (ii) Data cleaning and Data transformation.
3. (a) Describe various methods for : 5 each
 - (i) Data preprocessing
 - (ii) Explain how concept hierarchies are useful in data mining.
 - (b) Why analytical characterization is needed and how it can be performed ? Compare the result of two induction methods with relevance analysis and without relevance analysis. 10
 - (c) Describe various data mining primitives for specifying a data mining task. 10
4. (a) Explain a priori algorithm for association rule mining taking suitable example. 10
 - (b) A database has four transactions. Let min. support = 60% and min. conf = 80% :

T ID	Date	Item Bought
T 100	15/10/07	{K, A, D, B}
T 200	15/10/07	{D, A, C, E, B}
T 300	19/10/07	{C, A, B, E}
T 400	20/10/07	{B, A, D}

Find all frequent item sets using A priori and FP-growth respectively. Compare the efficiency of the two methods. 10

- (c) Discuss mining of multilevel association rules and explain how to check redundant multilevel association rules. 10
- 5. (a) What is Classification ? Discuss any *one* method of Decision tree induction. 10
- (b) Write an algorithm for kNN classification. 10
- (c) Describe the partitioning and density based methods of clustering. Write applications of clustering. 10