

Roll No

IT-840**B.E. VIII Semester**

Examination, June 2016

Data Mining and Warehousing

(Elective-IV)

Time : Three Hours

Maximum Marks : 70

- Note: i) Answer five questions. In each question part A, B, C is compulsory and D part has internal choice.
 ii) All parts of each questions are to be attempted at one place.
 iii) All questions carry equal marks, out of which part A and B (Max.50 words) carry 2 marks, part C (Max.100 words) carry 3 marks, part D (Max.400 words) carry 7 marks.
 iv) Except numericals, Derivation, Design and Drawing etc.

Unit - I

1. a) What is Data mart? What are the types of data mart?
- b) What are steps involved in clean and transformation of data?
- c) List the contexts of dimension table.
- d) Draw the data warehouse architecture and explain its components.

OR

Give reason, why it is necessary to separate data warehouse from operational database.

Unit - II

2. a) What is the difference between OLTP and data warehouse?
- b) List the difference between OLAP and OLTP.
- c) How can the data warehouse data be accessed efficiently?
- d) Discuss the methods for efficient computation of data cubes.

OR

Write short notes on

- i) ROLAP
- ii) MOLAP

IT-840

PTO

Unit - III

3. a) List any four data mining application.
- b) Write the difference between data base and knowledge base.
- c) What are goals of web usage mining?
- d) Explain in detail about text mining applications.

OR

What is data mining functionality? Explain different types of data mining functionality with examples.

Unit - IV

4. a) Define support and confidence in association rule mining.
- b) What are the latest trends in association rules mining.
- c) Define FP-growth algorithm.
- d) Explain the algorithm for mining frequent item sets without candidate generation for the given dataset minimum support value is 2.

| TID | Items bought |
|-----|--------------------------|
| 100 | (a, c, d, f, g, i, m, p) |
| 200 | (a, b, c, f, l, m, o) |
| 300 | (b, f, n, j, o, w) |
| 400 | (b, c, k, s, p) |
| 500 | (a, f, c, e, l, p, m, n) |

OR

Describe the algorithm for time series mining association rules.

Unit - V

5. a) What is meant by outlier?
- b) How is the zero frequency problem handled in naive bayes classifier?
- c) List out difference between clustering and classification.
- d) What is Clustering? Briefly describe the partitioning and hierarchical clustering methods. Give examples in each case.

OR

Discuss in detail about the Bayesian and decision tree classifier.

IT-840