[ECS-707]
M.Tech. Degree Examination
Computer Science & Technology
I SEMESTER
DATA WAREHOUSING AND DATA MINING
(Effective from the admitted batch 2015-16)

Time: 3 Hours Max.Marks: 60

Instructions: Each Module carries 12 marks.
Answer all modules choosing one question from each module.
All parts of the module must be answered in one place only.
Figures in the right hand margin indicate marks allotted.

MODULE-I

1. a) Explain the process and steps of knowledge discovery of data 7
   b) What is object relational database and give with an example? 5

   OR

2. a) What is data Reduction? Explain the attribute subset selection strategy for data reduction 7
   b) In real world data types with missing values for some attributes are a common occurrence. Describe various methods for handling this problem 5

MODULE-II

3. a) Explain the process of OLAP Queries 6
   b) How the process of data from tables, spreadsheets to data cubes? 6

   OR

4. Briefly compare for the following:
   a) Snow flake schema and fact constellation
   b) Enterprise data ware house and data mart 12

MODULE-III

5. a) Write a basic Algorithm for inducing a decision tree from training tuples 4
   b) Write an Algorithm for K-nearest-neighbour classification given ‘K’ and ‘n’, no. of attributes describing each tuple 4
   c) Explain briefly about support-vector machines 4
6. a) Explain in detail about the Bayesian classifiers
   b) What is a decision tree? Explain with an example

    MODULE-IV

7. a) Briefly explain about the various objective measures of interestingness for association patterns
    b) Briefly explain about the finding frequent item sets using candidate generation

    OR

8. A database has 5 transactions. Let minimum-support=60% and minimum-confidence=80% for the following table

    TABLES

    | TID | Items-bought |
    |-----|--------------|
    | T100 | {C,O,O,K,I,E} |
    | T200 | {C,O,M,P,U,T,E,R} |
    | T300 | {S,O,F,T,W,A,R,E} |
    | T400 | {W,O,O,D} |
    | T500 | {G,A,D,G,E,T} |

    Table: 8.1

    Find all frequent item sets using Apriori and FP-Growth, respectively, compare the efficiency of the two meaning process.

    MODULE-V

9. a) Give two objects represented by the tuples (22,1,42,10) and (20,0,36,8):
    i) Compute the Euclidean distance between the two objects
    ii) Compute the Manhattan distance between the two objects
    b) Briefly describe the following approaches to clustering:
       K – means, DBSCAN

    OR

10. a) What is cluster analysis? Explain its types
     b) Briefly explain about the Agglomerative Hierarchical Clustering