

USN

--	--	--	--	--	--	--	--	--	--

13MCA442

**Fourth Semester MCA Degree Examination, June/July 2017**  
**Data Warehousing and Data Mining**

Time: 3 hrs.

Max. Marks:100

**Note: Answer any FIVE full questions.**

- 1
  - a. What is a Data Warehouse? List down the differences between operational database systems and data warehouses. (06 Marks)
  - b. With a neat diagram, explain in detail about Three – tier data warehousing architecture. (06 Marks)
  - c. Discuss in detail about Star, Snowflake and Fact constellation schemas in detail. (08 Marks)
- 2
  - a. Define Data Mining. Explain the challenges that motivated the development of Data Mining. (06 Marks)
  - b. With a neat diagram, explain the Process of knowledge Discovery in Databases. (04 Marks)
  - c. Discuss Data Mining tasks in detail. (10 Marks)
- 3
  - a. Describe in detail about various types of data sets. (10 Marks)
  - b. What is Data Preprocessing? Explain the following techniques in detail : (10 Marks)
    - i) Sampling
    - ii) Dimensionality Reduction
    - iii) Discretization and Binarization.
- 4
  - a. Write down the Apriori principle and explain the Pseudo code for the frequent itemset generation part of the Apriori algorithm. (07 Marks)
  - b. Write a detailed note on Maximal frequent itemsets and Closed frequent itemsets. (05 Marks)
  - c. Discuss FP Growth algorithm in detail. (08 Marks)
- 5
  - a. What is a Decision Tree? Write an algorithm for Decision Tree induction. (07 Marks)
  - b. Explain Sequential Covering Algorithm in detail. (06 Marks)
  - c. Discuss K – nearest neighbour classification algorithm with characteristics of Nearest neighbour classifiers. (07 Marks)
- 6
  - a. Discuss in detail about various techniques for improving the accuracy of classification methods. (07 Marks)
  - b. Explain in detail about various evaluation criteria for classification methods. (06 Marks)
  - c. Describe Multiclass Problem in detail. (07 Marks)
- 7
  - a. What is Cluster Analysis? Explain Agglomerative clustering method in detail. (06 Marks)
  - b. Discuss K – means method in detail, with an example. (08 Marks)
  - c. Describe DBSCAN method in detail. (06 Marks)
- 8
  - a. What are Outliers? Explain statistical approaches in detail. (10 Marks)
  - b. Discuss Clustering – based approaches in detail. (10 Marks)

\*\*\*\*\*

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.  
 2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.