

3019113

Roll No.

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

B.E / B.Tech (Full Time) DEGREE END SEMESTER EXAMINATIONS APRIL / MAY 2013

COMPUTER SCIENCE AND ENGINEERING

SIXTH Semester

19

CS9027 – DATA WAREHOUSING AND DATA MINING

(Regulation 2009)

Time: 3 Hours

Answer ALL Questions

Max. Marks 100

PART-A (10 x 2 = 20 Marks)

1. What are the reasons for implementing a separate ODS and a separate data warehouse in an enterprise?
2. Give two reasons for dirty data being extracted from source systems.
3. What are the likely benefits of building an enterprise data warehouse?
4. Discuss some of the reasons for the growth in enterprise data.
5. Give an example of use of data mining in astronomy.
6. What is a candidate itemset and the frequent itemset?
7. List three most important evaluation criteria for classification methods.
8. Can a decision tree be built from numerical data? Give an example.
9. What is the difference between Single-link and Complete link?
10. List any three most widely used Data Mining Softwares.

Part – B (5 x 16 = 80 marks)

11. (i) List some of the major differences between OLTP System and Data Warehouse system. (8)
- (ii) Explain the different schema available in the Data Warehouse. (8)
12. a) Discuss in detail on any one of the Reporting and Query Tools of a business analysis system. Further mention its applications.

OR

- b) Give a detailed note on the Multi-dimensional data model. Explain the different operations that could be performed on a Multi-dimensional model.

13. a) Describe the steps that are required in a typical data mining process and explain the importance of each step.

OR

- b) Explore one area of data mining and show how the developments in these areas would not have been possible without a rapid growth in electronic data.

Roll No.

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

14. a) Explain the naïve algorithm for finding the association rules.

OR

b) What kind of data is the decision tree method most suitable for? Briefly outline the major steps of the algorithm to construct a decision tree. Explain each step.

15. a) List four desirable features of the cluster analysis method. Which of them are important for large datasets? Discuss.

OR

b) Suppose that the data mining task is to cluster the following eight points into three clusters:

A1(4,6), A2(2,5), A3(9,3), A4(6,9), A5(7,5), A6(5,7), A7(2,2), A8(6,6)

Suppose initially we assign A1, A2, A3 as the seeds of three clusters that we wish to find. Use the K-means method to show:

- (i) The three cluster centroids after the first iteration using Manhattan distance.
- (ii) The final three clusters.
- (iii) Find the same for the Euclidean distance.