

ALCCS – NEW SCHEME

Time: 3 Hours

FEBRUARY 2012

Max. Marks: 100

PLEASE WRITE YOUR ROLL NO. AT THE SPACE PROVIDED ON EACH PAGE IMMEDIATELY AFTER RECEIVING THE QUESTION PAPER.

NOTE:

- Question 1 is compulsory and carries 28 marks. Answer any FOUR questions from the rest. Marks are indicated against each question.
- Parts of a question should be answered at the same place.

-
- Q.1** a. Briefly explain “data granularity” with the help of example.
b. What is ETL (Extraction /Transformation /Loading) process? Discuss in brief.
c. List the major benefits of Data mining.
d. “Every data in Data warehouse is time stamped.” Discuss.
e. How is “Data mining” different from the “OLAP”?
f. Briefly outline the major steps in decision tree classification.
g. Differentiate between ‘Operational’ and ‘Decision Support’ systems. (7×4)
- Q.2** a. How are “data warehouse” different from a “database”? How are they similar? (9)
b. What do you mean by data reduction? What are the strategies of the data reduction? (9)
- Q.3** a. What do you mean by association rule mining? Give an example of market basket analysis from the real world. (9)
b. Differentiate between:
(i) OLAP and OLTP
(ii) Data warehouse and Data Mart. (9)
- Q.4** a. What do you mean by cluster analysis? What is the typical requirement of clustering in data mining? (9)
b. Since many data mining products are available in the market, mention any five multiple dimensional features which shall be kept in mind before choosing such a system. (9)

Code: CT75

Subject: DATA WAREHOUSING AND DATA MINING

Q.5 a. Why is naïve Bayesian Classification called “naïve”? Briefly outline the major ideas of naïve Bayesian classification. (9)

b. What are the different types of OLAP servers? Discuss essential differences between them. (9)

Q.6 a. Explain the features of Neural Network based classification. (4)

b. Discuss about the concept of data warehousing and the web. (4)

c. Six observation on two variables are available as shown in the following table:

Obs.	X1	X2
a	3	2
b	4	1
c	2	5
d	5	2
e	1	6
f	4	2

(i) Plot the observations in a scatter diagram. How many groups would you say there are and what are their members?

(ii) Apply the nearest neighbour method and the squared Euclidean distance as a measure of dissimilarity. Use a dendrogram to arrive at the number of groups and their membership. (5×2)

Q.7 Write short note on any **THREE**

(i) Data mining Applications

(ii) Drill Down analysis

(iii) Hyper cubes

(iv) Metadata

(3×6)