(Following Paper ID and Roll No. to be filled in your Answer Books)		
PAPER ID :	**Roll No.**	

M.C.A.(DD)

Theory Examination (Semester-VI) 2015-16

DATA WAREHOUSING & MINING

Time : 3 Hours   *Max. Marks : 100*

Note: Attempt questions from all Sections as per directions.

SECTION-A

1. Attempt *all parts* of this section. Answer in brief. [2X10=20]

a) Define KDD. Describe the phases in KDD process.

b) Write methods for handling noisy data.

c) Differentiate between descriptive data mining and predictive data mining.

d) How is prediction different from classification?

e) Explain data mining interface in brief.

f) Why mean is sensitive to extreme (outlier) values. How can we deal with these outliers?

g) Discuss various types of web mining techniques.

h) Explain the client/server computing model.

i) Explain the slice and dice operations of OLAP.

j) What is Data Warehousing? Also explain the Data Warehousing components.

SECTION-B

2. Attempt *any five questions* from this section. [10X5=50]

a) Write the algorithm of decision tree induction. What are the methods that can be used for selecting the splitting criteria?

b) Association rule mining often generates a large number of rules. Discuss effective methods that can be used to reduce the number of rules generated while still preserving most of the interesting rules

c) Give the difference between OLAP and OLTP. Describe various OLAP operations.

d) Briefly outline the major steps of decision tree classification. Why is tree pruning useful in decision tree induction? What is drawback of using a separate set of tuples to evaluate pruning?

e) Explain K-means clustering algorithm.

f) A database has 9 transactions. Let min support=2/9 and min confidence =70%:

TID	Items Bought
T100	1,2,5
T200	2,4
T300	2,3
T400	1,2,4
T500	1,3
T600	2,3
T700	1,3
T800	1,2,3,5
T900	1,2,3

Find all frequent item sets using Apriori Algorithm.

g) Implement Apriori algorithm and discover association rules in transactional databases.

h) Describe star and snowflake schema design methods.Suppose that a data warehouse for Big University consists of the following four dimensions: student, course, semester, and instructor, and two measures count

and avg grade. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the avg grade measure stores the actual course grade of the student. At higher conceptual levels, avg grade stores the average grade for the given combination. Draw a star and snowflake schema diagram for the data warehouse.
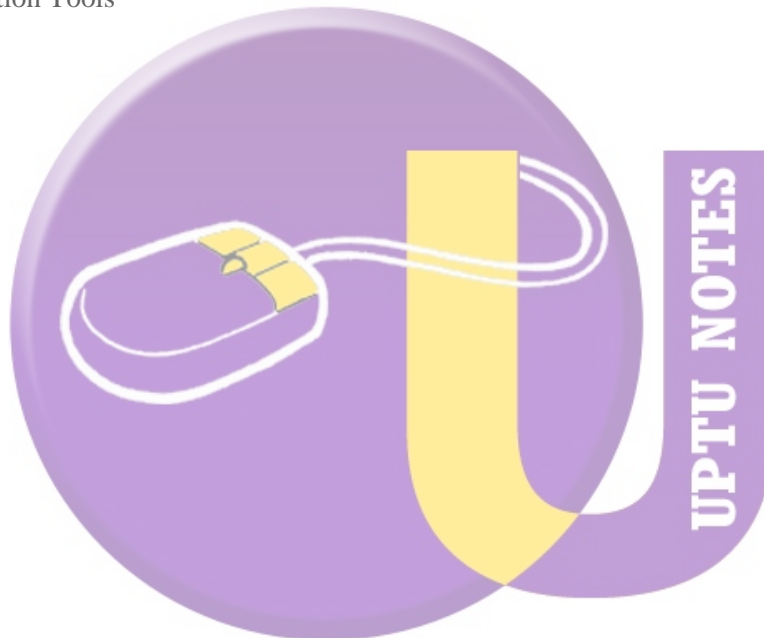
SECTION-C

Attempt *any two questions* **from this section. [15X2=30]**

3. Give a decision tree; you have the option of (a) converting the decision tree to rules and then pruning the resulting rules, or (b) Pruning the decision tree and then converting the pruned tree to rules. What advantages does (a) have over (b)? Explain with a suitable example.

4. (i) Explain the various types of OLAP servers. What are the steps for efficient processing of OLAP queries?
   (ii) How nearest neighbours prediction algorithm used in Data mining?

5. Write short notes on any three of the following:
a) Parallel Processors and Cluster systems
b) Data Transformation Tools
c) Meta-Data
d) Data Extraction.