Sem-VI / Comp / 22-12-10 / Datawarehousing & mining.

ws Oct Sacn- 10 56

Con. 6135–10. **(REVISED COURSE)** **GT–7566**

(3 Hours) **[Total Marks : 100**

NOTE: (1) Question No.1 is **Compulsory**.
(2) Solve any **four** questions out of remaining six questions.
(3) All questions carry **equal** marks.
(4) Assume **suitable data** if **required**.
(5) **Figures** to the **right** indicate full marks.

Q.1) a) Explain ETL of data warehousing in detail. (10)

b) Explain data mining as a step in KDD. Give the architecture of typical DM system. (10)

Q.2) a) A dimension table is wide; the fact table is deep. Explain.
What is STAR schema and its advantages (10)

b) What is Clustering? Explain K-means clustering algorithm.
Suppose the data for clustering is {2,4,10,12,3,20,30,11,25} consider K=2,
cluster the given data using above algorithm. (10)

Q.3) a) Consider the transaction database given below. Use Apriori Algorithm with minimum support count 2, Generate the association rules along with its confidence : (10)

TID	List of items
T100	I1,I2,I5
T200	I2,I4
T300	I2,I3
T400	I1,I2,I4
T500	I1,I3
T600	I2,I3
T700	I1,I3
T800	I1,I2,I3,I5
T900	I1,I2,I3

b) Explain the characteristics of the data present in the data warehouse. (10)

[TURN OVER

Q.4) a) Explain HITS algorithm. (10)

b) Define Data Warehouse. Explain the architecture of data warehouse with suitable block diagram. (10)

Q.5) a) Distinguish between: (10)
(i) Top-Down and Bottom-Up Approach
(ii) OLAP and OLTP

b) Explain Partitioning Methods for Clustering. (10)

Q.6) a) Explain different OLAP operations. (10)

b) Given the training data for height classification, classify the tuple, t = < Rohit, M, 1.95m > using Bayesian Classification. (10)

Name	Gender	Height	Output
Kiran	F	1.6m	Short
Jatin	M	2m	Tall
Madhuri	F	109m	Medium
Manisha	F	1.88m	Medium
Shilpa	F	1.7m	Short
Bobby	M	1.85m	Medium
Kavita	F	1.6m	Short
Dinesh	M	1.7m	Short
Rahul	M	2.2m	Tall
Shree	M	2.1m	Tall
Divya	F	1.8m	Medium
Tushar	M	1.95m	Medium
Kim	F	1.9m	Medium
Aarti	F	1.8m	Medium
Rajashree	F	1.75m	Medium

Q.7) Write short notes on (Any **Four**): (20)

(a) Web Structure Mining
(b) Decision Tree based Classification approach
(c) Crawlers
(d) Metadata
(e) Web personalization