BE/sem VIII/ comp/ CBGS / DWM

Datawarehousing & Mining / 02·12·16

Q.P. Code : 733601

(3 Hours) **[Total Marks : 80**

Note: 1. Question No.1 is compulsory
2. Attempt any **Three** questions out of remaining questions
3. Assume suitable data wherever necessary and state them clearly

Q1 a) Consider following dimensions for a Hypermarket chain: Product, Store, Time and Promotion. With respect to this business scenario, answer the following questions. Clearly state any reasonable assumptions you make. Design a star schema. Whether the star schema can be converted to snowflake schema? Justify your answer and draw snowflake schema for the data warehouse (clearly mention the Fact table(s), Dimension table(s), their attributes and measures). [10]

b) Define linear, non-linear and multiple regressions. Plan a regression model for Disease development with respect to change in weather parameters. [10]

Q2 a) What is meant by metadata in the context of a Data warehouse? Explain the different types of meta data stored in a data warehouse. Illustrate with a suitable example. [10]

b) Describe the various functionalities of Data mining as a step in the process of knowledge Discovery. [10]

Q3 a) In what way ETL cycle can be used in typical data ware house, explain with suitable instance. [10]

b) What is Clustering Technique? Discuss the Agglomerative algorithm with the following data and plot a Dendrogram using single link approach. The table below comprises sample data items indicting the distance between the elements. [10]

Item	E	A	C	B	D
E	0	1	2	2	3
A	1	0	2	5	3
C	2	2	0	1	6
B	2	5	1	0	3
D	3	3	6	3	0

TURN OVER

Q4 a) Discuss how computations can be performed efficiently on data cubes. [10]

b) A database has five transactions. Let min-support=60% and min-confidence = 80%. [10]
Find all Find frequent item sets by using Apriori Algorithm. T_ID is the transaction ID.

T_ID	Items bought
T-1000	M, O, N , K, E, Y
T-1001	D, O, N, K, E, Y
T-1002	M, A, K, E
T-1003	M, U, C, K, Y
T-1004	C, O, O, K, E

Q5 a) Differentiate [10]

i. OLTP Vs. OLAP
ii. Data Warehouse Vs. Data Mart

b) Why naive Bayesian classification is called "naive"? Briefly outline the major ideas of naive Bayesian classification. [10]

Q6 Write short notes on any **four** of the following: [20]

i. Application of Data Mining to Financial Analysis
ii. Fact less Fact Table
iii. Indexing OLAP data
iv. Data Quality
v. Decision Tree based Classification Approach
