Government Polytechnic Nagpur

DEPARTMENT OF INFORMATION TECHNOLOGY

QUESTION BANK

SUBJECT   : IT504E-DATAWAREHOUSINGAND MINING

Q.NO	QUESTIONS	COMPETENCE
	1.Introduction to data warehouse and Data mining	
1.	How is data ware house different from a database? **Identify** the similarity.	Remember
2.	**Differentiate** metadata and data mart.	Understand
3.	**Analyze** why one of the biggest challenges when designing a data ware house is the selection and cleaning.	Analyze
4.	How would you **evaluate** the goals of data mining?	Evaluate
5.	**List** different data warehouse tool.	Remember
6.	What elements would you use to **relate** the design of data warehouse?	Apply
7.	**Define** Data mart	Remember
8.	**Define** star schema	Remember
9.	What is Data warehousing? **Explain** the benefits of Data warehousing.	Evaluate
10.	Why data transformation is essential in the process of Knowledge discovery? **Describe** it.	Remember
11.	**Describe** Knowledge discovery in data.(KDD Steps)	Understand
12.	**Distinguish** datawarehouse and database	Understand
13.	How would you **show** your understanding in Multidimensional data model?	Apply
14.	**Formulate** what is data discretization.	Create
15	**List** Application of Data Warehouse and Data Mining.	Remember
16	**Describe** Issues in Data Warehouse and Data Mining	Understand
	2.Datawarehouse and OLAP Technology	
17	**List** features of Metadata repository in data warehousing	Remember
18	**Define** Metadata repository	Remember
19	**Illustrate** the benefits of metadata repository.	Apply
20	**Design** the data warehouse architecture.	Create

21	What is data warehouse? **Give** the Steps for design and construction of Data Warehouses and explain with three tier architecture diagram.	Understand
22	Diagrammatically **illustrate** and discuss the following preprocessing techniques: (i) Data cleaning (ii) Data Integration (iii) Data transformation (iv) Data reduction	Apply
23	Draw the data warehouse architecture and **explain** its components.	Analyze
24	**Analyze** the information needed to support DBMS schemas for Decision support.	Analyze
25	**Describe** the overall architecture of data warehouse?	understand
26	Suppose that a data warehouse consists of four dimensions customer, product, salesperson and sales time, and the three measure sales Amt(in rupees), VAT(in rupees) and payment_type(in rupees). **Draw** star schema that is popularly used for modeling data warehouses and **explain** it.	Evaluate
27	**Define** data Preprocessing**? Describe** the various data pre processing techniques.	Understand
28	**Describe** the steps involved in Knowledge discovery in databases(KDD).	Understand
29	**Describe** in detail about Data marts.	Understand
30	**Compare** the similarities and differences between the database and datawarehous.	Evaluate
31	**Compare** OLTP and OLAP.	Analyze
32	**Describe** Different OLAP Operation with example.	Understand
33	**Summarize** the distinct features of OLTP with OLAP.	Understand
34	**What** is multidimensional data model?. Give example.	Remember
35	**List** Different OLAP Server.	Remember
36	**Describe** MOLAP and ROLAP.	Understand
37	.**Summarize** the various OLAP operations in the Multidimensional Data Model.	Evaluate

	UNIT 3- DATA MINING	
1	**Define** Data mining. **List** out the steps in data mining?	Remember
2	**List** the steps involved in the process of KDD. How it is different from data mining?	Remember
3	**List** application of data mining	Remember
4	**Compare** drill down with roll up approach.	Analyze
5	**Describe** what the other kinds of data are.	Understand
6	**Define** association and correlations.	Remember
7	**Evaluate** the major tasks of data preprocessing.	Evaluate
8	Consider the following set of data X = {15,27,62,35,39,50,44,44,22,98} Do preprocessing using smoothing by bin means and bin boundary to smooth the data, using a bin of depth 3.**Evaluate** it.	Evaluate
9	**Formulate** why do we need data transformation. Mention the ways by which data Can be transformed.	Create
10	**Discuss** whether or not each of the following activities is a data mining task. 1. Credit card fraud detection using transaction records. 2. Dividing the customers of a company according to their gender. 3. Computing the total sales of a company 4. Predicting the future stock price of a company using historical records. 5. Monitoring seismic waves for earthquake activities.	Understand

	UNIT-4- Mining Association Rules in Large Databases	
1	**Define** correlation and market basket analysis.	Remember
2	**Formulate** the principle frequent itemset .	Create
3	How would you **explain** the principle of Apriori algorithm? How can the efficiency of an Apriori algorithm be improved?	Evaluate
4	**Define** Data pruning. State the need for pruning phase in decision tree construction.	Remember
5	**List** advantages of apriori algorithm.	Remember
6	**How** will you generate association rules from frequent itemsets	Analyze
7	**Discuss** association rule mining .list the two interesting measures of an association rule.	Understand
8	**List** Pros and Cons of Apriori Algorithm	Understand
9	**Describe** why Association Mining is necessary	Understand
10	**List** the two interesting measures of an association rule. **1.Support 2.Confidence**	Remember
11	**Giving** concrete example , explain a method that performs frequent item set mining by using the prior knowledge of frequent item set properties.	Understand

12	**Describe** in detail about frequent pattern classification.	Understand
13	Describe Apriori Algorithm with example.	Understand
14	A**pply** the Apriori algorithm for discovering frequent item sets of the table. Use 0.3 for the minimum support value. Illustrate each step of the Apriori Algorithm. Trans ID — Items Purchased 101 — Milk,bread,eggs 102 — Milk,juice 103 — Juice,butter 104 — Milk,bread,eggs 105 — Coffee,eggs 106 — Coffee 107 — Coffee,Juice 108 — Milk,bread,cookies,eggs 109 — Cookies,butter 110 — Milk,bread	Analyze
15	Find all frequent item sets for the given training set using Apriori. Use 0.2 for the minimum support value. Illustrate each step of the Apriori Algorithm TID ITEMS BROUGHT T100 {M , O, N , K , E , Y } T200 {D , O, N, K, E, Y } T300 {M , A K, E } T400 {M ,U , C , K ,Y } T500 {C , O , O ,K , I , E }	Analyze
16	Find all frequent item sets for the given training set using Apriori. 	Transaction ID	Items Bought	 	2000	A,B,C	 	1000	A,C	 	4000	A,D	 	5000	B,E,F	 Let minimum support 50% and minimum confidence 50%.	Analyze

UNIT-5- Classification and Prediction

1	**Compare** Classification and Prediction.	Understand
2	**Describe** the issues regarding classification and prediction.	Understand
3	**Illustrate** an algorithm for classification using decision trees.	Apply
4	What approach would you use to **apply** decision tree induction?	Apply
5	**State** the types of Regression	Remember
6	**Illustrate** Issues regarding Classification & Prediction	Apply
7	**Define** Regression	Remember

8	A Simple example from the stock market involving only discrete range has profit as categorically attribute with values {UP,DOWN} and training data is				Apply

Age	Competition	Type	Profit
old	Yes	Software	Down
old	No	Software	Down
old	No	Hardware	Down
mid	Yes	Software	Down
mid	Yes	Hardware	Down
mid	No	Hardware	Up
mid	No	Software	Up
new	Yes	Software	Up
new	No	Hardware	Up
new	No	Software	Up

1. Apply Decision Tree Algorithm –ID3 & Construct Decision Tree

2. Enlist rules from Decision Tree.

9	Apply Decision Tree Algorithm –ID3 & Construct Decision Tree for below given data.					Apply

Outlook	Temp	Humidity	Windy	PlayGolf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

10	Apply Decision Tree Algorithm –ID3 & Construct Decision Tree for below given data.					Apply

Id	Age	Income	Student	Credit Ratings	Buys Computer
1	<=30	High	No	Fair	No
2	<=30	high	No	Excellent	No
3	31-40	high	No	Fair	Yes
4	>40	Medium	No	Fair	Yes
5	>40	Low	Yes	Fair	Yes
6	>40	low	Yes	Excellent	No
7	31-40	Low	Yes	Excellent	Yes

8	<=30	Medium	No	Fair	No
9	<=30	low	Yes	Fair	Yes
10	31-40	Medium	Yes	Fair	Yes
11	<=30	Medium	Yes	Excellent	Yes
12	31-40	Medium	No	Excellent	Yes
13	31-40	High	Yes	Fair	Yes
14	>40	Medium	No	Excellent	No

UNIT 6- CLUSTERING AND TRENDS IN DATA MINING		
1	**Define** Clustering.	Remember
2	**Define** K-means partitioning.	Remember
3	**Explain** why a cluster has to be evaluated.	Analyze
4	**Define** what is meant by K nearest neighbor algorithm.	Remember
5	**Illustrate** some applications of data mining.	Apply
6	**Formulate** the role of application and challenges in clustering.	Create
7	**Classify** the hierarchical clustering methods.	Analyze
8	**Define** outlier. Examine the outlier analysis.	Remember
9	**Discuss** the challenges of outlier detection.	Understand
10	**Distinguish** between Classification and clustering.	Understand
11	**Evaluate** what information is used by outlier detection method.	Evaluate
12	**Give** the categorization of major clustering methods.	Understand
13	**What is clustering? Describe** in detail about the features of K- means partitioning method.	Remember
14	**Explain** in detail about hierarchical based method.	Analyze
15	Consider five points { X_1, X_2, X_3, X_4, X_5} with the following coordinates as a two dimensional sample for clustering: X_1 = (0,2.5); X_2 = (0,0); X_3= (1.5,0); X_4 = (5,0); X_5 = (5,2) **Illustrate** the K-means partitioning algorithm using the above data set.	Apply

16	Discuss the steps in K-means algorithm and **evaluate** the following table | Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5 | using K- means.	Evaluate
17	**Analyze** and elaborate the current trends in data mining in any **three** fields . 1. Financial data analysis 2. Biological data analysis 3. Telecommunication industry 4. Intrusion detection 5. Retail industry	Create