

DATA MINING & WAREHOUSING QUESTION BANK

Semester :6th

- Q.1 Explain different data mining tasks.
- Q.2 What is the relation between data warehousing and data mining?
- Q.3 Explain the differences between “Explorative Data Mining” and “Predictive Data Mining” and give one example of each.
- Q.4 What are the application areas of data Mining?
- Q.5 Explain the differences between Knowledge discovery and data mining.
- Q.6 How is data warehouse different from a database? How are they similar?
- Q.7 What type of benefit you might hope to get from data mining?
- Q.8 What are the key issues in data Mining?
- Q.9 How can Data Mining help business analyst?
- Q.10 What are the limitations of data Mining?
- Q.11 Discuss the need of human intervention in data mining process.
- Q.12 As a bank manager, how would you decide whether to give loan to an applicant or not?
- Q.13 What steps you would follow to identify a fraud for a credit card company.
- Q.14 What is data Mining?
- Q.15 State three different application for which data mining techniques seem appropriate. Informally explain each application.
- Q.16 Explain briefly the differences between “classification” and “clustering” and give an informal example of an application that would benefit from each techniques.
- Q.17 What do you mean by Data Processing?
- Q.18 Explain data cleaning.
- Q.19 Describe different data cleaning approaches.
- Q.20 How can we handle missing values?
- Q.21 Explain Noisy Data.
- Q.22 Give Brief description of following:
- (a) Binning
 - (b) regression
 - (c) Clustering
 - (d) Smoothing

(e) Generalization

(f) Aggregation

Q.24 Can you briefly describe the four stages of knowledge discovery(KDD)? Can you describe the multi-tiered data warehouse architecture?

Q.25 A data set for analysis includes only one attribute X:

$X = \{7, 12, 5, 8, 5, 9, 13, 12, 19, 7, 12, 12, 13, 3, 4, 5, 13, 8, 7, 6\}$

- (a) What is the mean of the data set X?
- (b) What is the median?
- (c) Find the standard deviation for X.

Q.26 Define Frequent sets, confidence, support and association rule.

Q.27 What do you mean by Market Basket analysis and how it can help in a supermarket?

Q.28 Explain whether association rule mining is supervised or unsupervised type of learning.

Q.29 Name some variants of Apriori Algorithm.

Q.30 Discuss the importance of Association Rule Mining.

Q.31 Consider the Data set D. Given the minimum support₂, apply apriori algorithm on this dataset.

Transaction ID	Items
100	A,C,D
200	B,C,E
300	A,B,C,E
400	B,E

Q.32 Describe example of data set for which apriori check would actually increase the cost? By describe I mean either show an instance of the data set or describe how would it look like.

Q.33 Same question for MaxMiner. When does MaxMiner perform worse than apriori. How does MaxMiner generate the frequency counts for every itemset which meets support constraints?

Q.34 With a neat sketch explain the architecture of a data warehouse

Q.35 Discuss the typical OLAP operations with an example

Q.36 (i) Discuss how computations can be performed efficiently on data cubes.

(ii) Write short notes on data warehouse meta data.

Q.37 (i) Explain various methods of data cleaning in detail.

(ii) Give an account on data mining Query language.

Q.38 How is Attribute-Oriented Induction implemented? Explain in detail.

Q.39 (a) Write and explain the algorithm for mining frequent item sets without candidate generation. Give relevant example.

Q.40 Discuss the approaches for mining multi level association rules from the transactional databases. Give relevant example.

Q.41 (i) Explain the algorithm for constructing a decision tree from training samples.

(ii) Explain Bayes theorem.

Q.42 Explain the following clustering methods in detail:

- (i) BIRCH
- (ii) CURE

Q.43 Classification is supervised learning. Justify.

Q.44 Explain different classification Techniques.

Q.45 Entropy is an important concept in information theory. Explain its significance in mining context.

Q.46 What are over fitted models? Explain their effects on performance.

Q.47 Explain Naive Baye's Classification.

Q.48 Describe the essential features of decision trees in context of classification.

Q.49 What are the advantages and disadvantages of decision trees over other classification methods?

Q.50 Explain ID3 Algorithm.

Q.51 Explain the methods for computing best split.

Q.52 What is Clustering? What are different types of clustering?

Q.53 Explain different data types used in clustering.

Q.54 Define Association Rule Mining

Q.55 When can we say the association rules are interesting?

Q.56 Explain Association rule in mathematical notations.

Q.57 Define support and confidence in Association rule mining.

Q.58 How are association rules mined from large databases?

Q.59 Describe the different classifications of Association rule mining.

Q.60 What is the purpose of Apriori Algorithm?

Q.61 Define anti-monotone property.

Q.62 How to generate association rules from frequent item sets?

Q.63 Give few techniques to improve the efficiency of Apriori algorithm.

Q.64 Mention few approaches to mining Multilevel Association Rules

Q.65 What are multidimensional association rules

Q.66. List out the differences between OLTP and OLAP.

Q.67. Explain mining Multi-dimensional Boolean association rules from transaction

Q.68. Explain constraint-based association mining?

Q.69 Specify the 5 criteria for the evaluation of classification & prediction?

Q.70 State two clustering methods that are used in "grid and density based method?"

Q.71 Why every data structure in the data warehouse contains the time element.

Q.72 How does a snowflake schema differ from a star schema? Name two advantages and two disadvantages of the snowflake schema.

Q.73 What are the essential differences between the MOLAP and ROLAP models? Also list a few similarities.

Q.74 Why is the entity-relationship modelling technique not suitable for the data warehouse.

Q.75 How is Data Mining different from OLAP? Explain Briefly.

