

GUJARAT TECHNOLOGICAL UNIVERSITY

BE-6 SEMESTER – OLD PAPER – S22 TO W25 – QUESTION BANK

Subject Name & Code: **DATA MINING (3160714)**

Unit 1: Introduction to Data Mining

Repeated Questions:

1. Explain KDD process. (W24 - Q1c, 07 marks) (S23 - Q1b, 07 marks) (W23 - Q1c, 07 marks)
2. What is data mining? List/explain data mining functionalities. (W25 - Q1a, 03 marks) (W24 - Q1a, 03 marks) (S22 - Q1c, 07 marks) (S23 - Q1a, 03 marks)
3. Draw and explain the data mining architecture. (S25 - Q1c, 07 marks) (S22 - Q1c, 07 marks)
4. Briefly explain the issues in data mining. (W25 - Q1b, 04 marks) (S23 - Q1c, 07 marks) (W23 - Q5c, 07 marks)

Other Important Questions:

1. What is the difference between discrimination and classification? (S25 - Q1a, 03 marks)
 2. Describe the steps involved in data mining when viewed as a process of knowledge discovery. (S25 - Q1b, 04 marks)
 3. How is a data warehouse different from database? (W24 - Q1b, 04 marks)
 4. Compare descriptive and predictive data mining. (S22 - Q1b, 04 marks) (W22 - Q1a, 03 marks)
 5. Define Data Warehouse. State its features. (W23 - Q1a, 03 marks)
 6. Differentiate between OLAP and OLTP. (W23 - Q1b, 04 marks)
 7. What are the types of data? (S22 - Q1a, 03 marks)
 8. What is market basket analysis? (W24 - Q3a, 03 marks) (S23 - Q1a, 03 marks)
 9. Define each of the following data mining functionalities: characterization, discrimination, regression. (W24 - Q1a, 03 marks)
 10. What is data mining? Describe three challenges to data mining regarding data mining methodology and user interaction issues. (S24 - Q1a, 03 marks)
 11. Explain the steps in knowledge discovery. (S24 - Q1b, 07 marks)
 12. Explain major requirements and challenges in data mining. (W22 - Q1c, 07 marks)
-

Unit 2: Data Pre-processing

Repeated Questions:

1. How to handle missing values? Explain various methods. (W25 - Q4a, 03 marks) (W24 - Q2a, 03 marks) (S23 - Q1b, 04 marks) (S24 - Q2c, 07 marks) (W22 - Q2c, 07 marks)
2. Explain data pre-processing functionalities / data cleaning methods. (W25 - Q2 OR c, 07 marks) (S22 - Q2c, 07 marks) (W22 - Q2c, 07 marks)
3. Discuss dimensionality reduction techniques. (W25 - Q2b, 04 marks) (S22 - Q2a, 03 marks)
4. Explain normalization methods with example/data. (S25 - Q2c, 07 marks) (S23 - Q3c, 07 marks) (S24 - Q2b, 07 marks)

Other Important Questions:

1. Explain k-fold cross validation. (W25 - Q2a, 03 marks)
 2. Discuss issues to consider during data integration. (S25 - Q2a, 03 marks) (S22 - Q2 OR c, 07 marks)
 3. Use smoothing by bin means to smooth these data, using a bin depth of 3... [Figure or Table given - refer Question paper of S25]. (S25 - Q2 OR c, 07 marks)
 4. How to handle noisy data? (W24 - Q2b, 04 marks)
 5. What is missing values and noise? How to handle them? (W25 - Q4a, 03 marks)
 6. What are the smoothing techniques available to remove noise? (S24 - Q2a, 03 marks)
 7. What is Data Cleaning? Describe various methods of Data Cleaning. (W22 - Q2c, 07 marks)
 8. Why to preprocess the data in data Mining? (W23 - Q2a, 03 marks)
 9. Explain Binning method with the help of example. (W23 - Q2b, 04 marks)
 10. What is dimensionality reduction? (S22 - Q2a, 03 marks)
 11. What are the types of concept hierarchies? (S22 - Q2b, 04 marks)
 12. What do you mean by concept hierarchy? (W22 - Q2a, 03 marks)
 13. Explain the smoothing techniques. (W22 - Q2b, 04 marks)
 14. Discuss data discretization and concept hierarchy generation. (S24 - Q2 OR c, 07 marks)
 15. Explain about the different Data Reduction techniques. (W22 - Q2 OR c, 07 marks)
-

Unit 3: Concept Description, Mining Frequent Patterns, Associations and Correlations

Repeated Questions:

1. Explain Apriori algorithm for finding frequent itemsets with an example. (S25 - Q3c, 07 marks) (W24 - Q2c, 07 marks) (W23 - Q2c, 07 marks) (W22 - Q3c, 07 marks) (S24 - Q3c, 07 marks)
2. Explain frequent itemset mining / association rule mining with example. (W25 - Q1c, 07 marks) (S23 - Q2c, 07 marks) (S22 - Q3b, 04 marks)
3. What is meant by association rule? (S25 - Q3a, 03 marks) (S22 - Q3a, 03 marks)
4. Discuss support, confidence, frequent itemset and strong association rules. (W25 - Q3b, 04 marks) (W23 - Q2c, 07 marks)

Other Important Questions:

1. What is meant by Maximal Frequent Item Set? (S25 - Q3 OR a, 03 marks)
2. Write a short note on: Web content mining. (S25 - Q3b, 04 marks)
3. Write a short note on: Web usage mining. (S25 - Q3 OR b, 04 marks)
4. Which patterns are interesting? Explain based on Pattern Evaluation Methods. (S25 - Q3 OR c, 07 marks)
5. Give a short example to show that items in a strong association rule are not always interesting. (S23 - Q2a, 03 marks)
6. Briefly describe how partitioning technique may improve the efficiency of Apriori algorithm. (S23 - Q2b, 04 marks)
7. Discuss how frequent itemsets can be generated using FP-Growth algorithm... [Figure or Table given - refer Question paper of S23]. (S23 - Q2c, 07 marks)
8. Find all frequent itemsets and also generate the strong association rules using Apriori algorithm... [Figure or Table given - refer Question paper of S23]. (S23 - Q2 OR c, 07 marks)
9. Describe any three primitives for specifying a data mining task. (S23 - Q3a, 03 marks)
10. What are the techniques to improve the efficiency of Apriori algorithm? (W22 - Q3a, 03 marks)
11. What is an Itemset? What is a Frequent Itemset? (W22 - Q3b, 04 marks)
12. For the given data find the frequent itemsets and generate association rules... [Figure or Table given - refer Question paper of W22]. (W22 - Q3c, 07 marks)
13. Describe the different classifications of Association rule mining. (W22 - Q3 OR a, 03 marks)
14. What is meant by Reduced Minimum Support? (W22 - Q3 OR b, 04 marks)
15. How are association rules mined from large databases? (S24 - Q3a, 03 marks)
16. Give the difference between Boolean association rule and quantitative association rule. (S24 - Q3b, 04 marks)
17. What are the limitations of the apriori approach for mining? Briefly describe the techniques to improve the efficiency of apriori algorithm. (S24 - Q3c, 07 marks)
18. Describe two interesting measures for association rules. (S24 - Q3 OR a, 03 marks)
19. How Meta rules are useful in constraint based association mining. (S24 - Q3 OR b, 04 marks)
20. Write an algorithm for finding frequent item-sets using candidate generation. (S24 - Q3 OR c, 07 marks)
21. Explain the various criteria for the classification of frequent pattern mining. (S22 - Q3c, 07 marks)
22. List two interesting measures for association rules. (S22 - Q3 OR a, 03 marks)
23. What is meant by multidimensional association rules? (S22 - Q3 OR b, 04 marks)
24. Write short notes on Maximal Frequent Item Set & Closed Frequent Item Set. (S22 - Q3 OR c, 07 marks)
25. What is closed item-set and maximal frequent item-set? (W25 - Q4 OR a, 03 marks)

Unit 4: Classification and Prediction

Repeated Questions:

1. Explain decision tree algorithm with suitable example. (W25 - Q3c, 07 marks) (W24 - Q3c, 07 marks) (S23 - Q4c, 07 marks) (S24 - Q4c, 07 marks)
2. Differentiate classification and clustering / prediction. (W25 - Q5b, 04 marks) (W24 - Q4a, 03 marks) (S23 - Q4a, 03 marks) (S22 - Q4a, 03 marks)
3. Explain classification vs. prediction. (W25 - Q3 OR a, 03 marks) (S22 - Q4 OR a, 03 marks)
4. Discuss issues regarding classification and prediction. (W25 - Q4 OR c, 07 marks) (S25 - Q4 OR b, 04 marks) (S22 - Q4 OR c, 07 marks) (W22 - Q4 OR c, 07 marks)
5. Explain neural network-based algorithms / ANN / Backpropagation. (W25 - Q3 OR c, 07 marks) (W24 - Q4 OR c, 07 marks) (S25 - Q4 OR c, 07 marks)

Other Important Questions:

1. What is the difference between "supervised" and "unsupervised" learning? (S25 - Q4a, 03 marks) (S22 - Q4 OR b, 04 marks) (S24 - Q4a, 03 marks) (W23 - Q3a, 03 marks)
2. What are the attribute selection measures? Explain any one in detail. (S25 - Q4b, 04 marks)
3. Discuss about Decision tree induction algorithm with an example. (S25 - Q4c, 07 marks)
4. What do you mean by Bayesian Classification? (S25 - Q4 OR a, 03 marks)
5. Explain proximity measures used in distance based classifiers. (W25 - Q3 OR b, 04 marks)
6. Differentiate linear and logistic regression. (W25 - Q4b, 04 marks)
7. Define accuracy, precision, recall and G-mean. (W25 - Q4 OR b, 04 marks)
8. Explain Linear regression. (W24 - Q3b, 04 marks)
9. Explain logistic regression. (W24 - Q3 OR b, 04 marks)
10. Which metrics used for evaluating classifier performance? (W24 - Q4b, 04 marks)
11. Explain Principal Component Analysis. (W24 - Q4c, 07 marks)
12. Explain outlier detection. (W24 - Q4 OR b, 04 marks)
13. Use the method of least squares to find an equation for the prediction... [Figure or Table given - refer Question paper of S23]. (S23 - Q3b, 07 marks)
14. What is noise? Describe the possible reasons for noisy data. Explain the different techniques to remove the noise from data. (S23 - Q3c, 07 marks)
15. Discuss outlier analysis as a data mining functionality with the help of an example. (S23 - Q3 OR a, 03 marks)
16. Explain how classification rules are extracted from a decision tree with the help of an example. (S23 - Q3 OR b, 04 marks)
17. Apply ID3 classification algorithm on the following data and construct a decision tree... [Figure or Table given - refer Question paper of S23]. (S23 - Q4c, 07 marks)
18. Discuss cross-validation method for evaluating the accuracy of a classifier. (S23 - Q4 OR a, 03 marks)
19. Predict class label of the tuple X with the help of Naive Bayesian classification method... [Figure or Table given - refer Question paper of S23]. (S23 - Q4 OR c, 07 marks)
20. What are Bayesian Classifiers? (W22 - Q4a, 03 marks)
21. What are the hierarchical methods used in classification? (W22 - Q4b, 04 marks)
22. Describe in detail about Rule based Classification. (W22 - Q4c, 07 marks)
23. What is attribute selection measure? (W22 - Q4 OR a, 03 marks)
24. Write down short note on backpropagation. (S24 - Q4b, 04 marks)
25. What is information gain? Explain the steps required to generate a decision tree from a training data set. (S24 - Q4c, 07 marks)
26. Differentiate between linear regression and nonlinear regression. (S24 - Q4 OR a, 03 marks)
27. Explain various methods of evaluating accuracy of classifier. (S24 - Q4 OR b, 04 marks)
28. Write a short on: web content mining. (S24 - Q4 OR c, 07 marks)
29. What is an outlier? (S22 - Q4a, 03 marks)
30. What is Bayesian theorem? (S22 - Q4b, 04 marks)
31. Demonstrate how Bayesian classification helps in predicting class membership probabilities. (S22 - Q4c, 07 marks)

32. Explain CART Classification Method. (W24 - Q3 OR c, 07 marks)
 33. What is Regression? Explain Linear Regression in short. (W23 - Q3b, 04 marks)
 34. Explain Naive Bayes Classifier with example. (W23 - Q3c, 07 marks)
 35. What do you mean by Tree Pruning? Explain with example. (W23 - Q4a, 03 marks)
 36. Explain the following as attribute selection measure: (i) Information Gain (ii) Gain Ratio. (W23 - Q4b, 04 marks)
 37. What do you mean by learning-by-observation? Explain k-Means clustering algorithm in detail. (W23 - Q4c, 07 marks)
 38. Define Data Cube. Explain any two operations on it. (W23 - Q4 OR a, 03 marks)
-

Unit 5: Cluster Analysis

Repeated Questions:

1. Explain K-means algorithm with example. (W25 - Q5c, 07 marks) (S22 - Q5c, 07 marks) (W23 - Q4c, 07 marks)
2. Explain hierarchical and agglomerative method of clustering. (W25 - Q4c, 07 marks) (S23 - Q5 OR a, 03 marks)
3. How to evaluate clustering algorithm? (W25 - Q5 OR a, 03 marks) (S25 - Q5a, 03 marks)
4. Differentiate partitioning and hierarchical methods for clustering. (S24 - Q5b, 04 marks) (W23 - Q4 OR b, 04 marks)
5. Explain PAM algorithm. (W25 - Q5 OR c, 07 marks) (W24 - Q5c, 07 marks)

Other Important Questions:

1. What are the requirements of clustering? (S25 - Q5a, 03 marks) (W22 - Q5a, 03 marks) (W23 - Q5 OR c, 07 marks)
 2. Discuss about Outlier Detection. (S25 - Q5b, 04 marks)
 3. Discuss about k-nearest neighbor classification algorithm with an example? (S25 - Q5c, 07 marks)
 4. State the categories of clustering methods? (S25 - Q5 OR a, 03 marks)
 5. Explain the concept hierarchy. (S25 - Q5 OR b, 04 marks)
 6. Explain in detail about partitioned Clustering method. (S25 - Q5 OR c, 07 marks)
 7. Briefly explain outlier analysis. (W25 - Q5a, 03 marks)
 8. Discuss data matrix and dissimilarity matrix with respect to clustering. (S23 - Q4b, 04 marks)
 9. How k-means clustering method differs from k-medoids clustering method? Discuss major drawbacks of k-means clustering method. (S23 - Q4 OR b, 04 marks)
 10. Suppose that the data mining task is to cluster the following eight points into three clusters... [Figure or Table given - refer Question paper of S23]. (S23 - Q5c, 07 marks)
 11. Explain the any four typical requirements of clustering in data mining. (S23 - Q5 OR b, 04 marks)
 12. Differentiate Agglomerative and Divisive Hierarchical Clustering? (W22 - Q5b, 04 marks)
 13. Write a short note: Web content mining. (W22 - Q5c, 07 marks)
 14. What is meant by hierarchical clustering? (W22 - Q5 OR a, 03 marks)
 15. Illustrate strength and weakness of k-mean in comparison with k-medoid algorithm. (W22 - Q5 OR b, 04 marks) (S24 - Q5 OR b, 04 marks)
 16. Write a short note: Web usage mining. (W22 - Q5 OR c, 07 marks)
 17. Explain following clustering algorithm in details: 1) CLARA 2) BIRCH. (S24 - Q5c, 07 marks)
 18. List out the applications of distributed and parallel data mining. (S24 - Q5 OR a, 03 marks)
 19. Explain the typical requirements of clustering in data mining. (S24 - Q5 OR c, 07 marks)
 20. What do you meant by Clustering? Explain the requirements used in Clustering? (S22 - Q5 OR c, 07 marks)
 21. How K-Mean clustering method differs from K-Medoid clustering method? (W23 - Q5a, 04 marks)
 22. Draw and explain the topology of a multilayer, feed-forward Neural Network. (W23 - Q5b, 07 marks)
 23. Give difference between text mining and web mining. (W23 - Q5 OR a, 03 marks)
 24. Why Hadoop is important? (W23 - Q5 OR b, 07 marks)
-

Unit 6: Web Mining and Other Data Mining

Repeated Questions:

1. Explain web usage mining. (S22 - Q5b, 04 marks) (S23 - Q5 OR c, 07 marks) (W24 - Q5b, 04 marks)
2. Explain web content mining. (S22 - Q5 OR b, 04 marks) (W22 - Q5c, 07 marks) (S24 - Q4 OR c, 07 marks)
3. Discuss application of web mining. (W25 - Q5 OR b, 04 marks) (W24 - Q5 OR c, 07 marks)
4. Discuss spatial Data Mining, Temporal Mining, and Multimedia Mining with their application. (W25 - Q4 OR c, 07 marks) (S22 - Q5a, 03 marks) (S23 - Q5b, 04 marks)

Other Important Questions:

1. What is spatial data mining? Explain with its application. (W25 - Q3a, 03 marks)
2. What is web mining? Explain web usage mining in detail. (S23 - Q5 OR c, 07 marks)
3. Discuss web structure mining. (S23 - Q5a, 03 marks) (W24 - Q5 OR b, 04 marks)
4. Discuss multimedia mining. (S23 - Q5b, 04 marks)
5. What is temporal mining? (S22 - Q5a, 03 marks) (S24 - Q5a, 03 marks)
6. Explain temporal mining. (S24 - Q5a, 03 marks)
7. What is multimedia mining? (S22 - Q5 OR a, 03 marks)
8. What is Web log structure? And discuss issues regarding web logs. (W24 - Q5b, 04 marks)
9. Write Applications of Distributed and parallel Data Mining. (W24 - Q5 OR c, 07 marks)
10. Write similarity measures for clustering graph and network data. (W24 - Q5 OR a, 03 marks)
11. Write applications of clustering graph and network data. (W24 - Q5a, 03 marks)
12. What is web log? Explain web structure mining and web usage mining in detail. (W23 - Q5 OR c, 07 marks)
