GANDHINAGAR INSTITUTE OF TECHNOLOGY
INFORMATION TECHNOLOGY & COMPUTER ENGINEERING DEPARTMENT

QUESTION BANK
SEMESTER: 7th
Data Mining & Business Intelligence (2170715)

Chapter 1: Introduction to Data Warehousing

1. What is Cuboid? Explain various OLAP Operations on Data Cube with example.
2. Explain following Terms:
   - Concept Hierarchy.
   - Histogram.
3. Compare OLTP & OLAP systems.
4. Explain Star, Snowflake, and Fact Constellation Schema for Multidimensional Database.
5. Explain Meta data repository.
6. Explain three-tier data warehouse architecture.

   OR

   With the help of a neat diagram explain the 3-tier architecture of a data warehouse.
7. Clearly state the differences between "Data Warehouses" and "Operational Database Systems".

Chapter 2: Introduction to Data Mining

8. What is Data Mining? Explain Data mining as one step of Knowledge Discovery Process.

   OR

   Define the term "Data Mining". With the help of a suitable diagram explain the process of knowledge discovery from databases.
9. List and describe major issues in data mining.
10. Explain the KDD process in details.
11. Explain Architecture of a typical data mining system.
12. What kind of data mined in Data Mining?
13. Explain data mining Functionalities.
14. Classification of data mining system.
15. Explain Data mining task primitives.

Chapter 3: Data Pre-processing

16. Explain Mean, Median, Mode, and Variance & Standard Deviation in brief.
17. Why pre-process the data?
18. Explain Data Cleaning process for missing values & Noisy data.
19. What is noise? Describe the possible reasons for noisy data. Explain the different techniques to remove the noise from data.

    OR

    List and describe the methods for handling the missing values in data cleaning.
20. Short Note: Distributive and Holistic measures.
21. Explain data transformation in data mining.
22. Explain with example how continuous numerical data values can be discretized.

23. Explain methods for data normalization.
24. Explain Entropy-Based Discretization.
25. Explain the 3-4-5 rule.

OR

Explain Discretization by Intuitive Partitioning.

26. Use two methods below to normalize following group of data :
    200, 300, 400, 600, 1000
    - Min-max normalization by setting min=0 & max=1.
    - z-score normalization
27. Suppose a group of 12 sales price records has been sorted as follows:
    5,10,11,13,15,35,50,55,72,92,204,215
    Partition them into three bins by each of following methods:
    - equal frequency partitioning
    - equal width partitioning

Chapter 4: Concept Description and Association Rule Mining

28. Explain BUC algorithm for the Computation of Sparse or Iceberg Cube.
29. Write & Explain Apriori algorithm for discovering frequent itemsets for mining Boolean Association Rules.
30. What is Market Basket Analysis? Explain Association Rules with Confidence & Support.
31. Write an algorithm for finding frequent item-sets using candidate generation.
32. Describe the list of techniques for improving the efficiency of Apriori-based mining.
33. A database has five transactions. Let Min_sup=60%.

TID	items_bought
T100	{M,N,O,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}

    Find all frequent itemsets using Apriori algorithm.

Chapter 5: Classification and Prediction

34. Explain Information Gain, Gain Ratio & Gini Index.
35. What is "Information Gain"? Explain the steps required to generate a Decision Tree from a training data set.
36. Explain the Classification by Decision Tree Induction Algorithm.
37. Explain Linear & Non-Linear Regression methods of Predictions.
38. Explain k-means and k-medoids algorithm of clustering.
39. Explain Rule-based Classification in brief.
40. Write the typical requirements of clustering in data mining.
41. Explain k-means and k-medoids algorithms of clustering.
42. Explain rule based classification and case based reasoning in details.
43. Write the steps of the k-means clustering algorithm. Also state its limitations.
44. Explain how the accuracy of a classifier can be measured.
45. Explain how the accuracy of a predictor can be measured.
46. Write a short note on hierarchical clustering.
47. Explain "Linear Regression" using suitable example.
48. Explain how the topology of a neural network is designed.

49. Discuss applications of “Fuzzy Logic”.
50. Explain Naïve Bayesian Classification algorithm.
51. Explain Rough Set Approach.
52. What is cluster analysis? Explain types of data in cluster analysis.

Chapter 6: Data Mining for Business Intelligence Applications

53. Explain about below data mining Applications:
    - Data Mining for Financial Data Analysis.
    - Data Mining for the Retail Industry.
    - Data Mining for the Telecommunication Industry.
    - Data Mining for Biological Data Analysis.
54. Explain different types of Web Mining with example.
55. Explain different types of data on which mining can be performed.
56. Explain the methodologies for stream data processing and stream data Systems.
57. What are the challenges for effective resource and knowledge discovery in mining the World Wide Web?
58. Explain the information retrieval methods used in text mining.
59. Explain how a search engine automatically identifies authoritative web pages on a user’s search topic.
60. Explain trend analysis in mining time-series data.