Data Mining and Analytics

Question Bank (Unit-1 and 2)

1. Explain the concept of data mining and its importance in modern organizations.
2. Describe the Knowledge Discovery in Databases (KDD) process in detail.
3. Explain different types of data mining tasks with suitable examples.
4. Discuss various data mining functionalities.
5. Explain classification of data mining systems.
6. Describe major issues in data mining.
7. Discuss applications of data mining in various domains.
8. Apply the KDD process steps to a real-world business problem.
9. Apply data mining functionalities to solve a retail business problem.
10. Describe the data warehouse process in detail.
11. Describe OLAP operations with multidimensional data cube using suitable example.
12. Draw steps of data mining process of knowledge discovery.
13. Draw and discuss multi-tiered architecture of data warehouse.
14. Explain different types of predictive data mining.
15. Explain different types of Descriptive data mining.
16. Discuss various Data Mining Task Primitives.
17. Explain the concept of association rule mining with a suitable example.
18. Describe support, confidence, and lift measures used in association rules.
19. Explain the Apriori principle and its importance in frequent itemset mining.
20. Describe the steps involved in the Apriori algorithm.
21. Explain the FP-Growth algorithm and its working mechanism.
22. Compare Apriori and FP-Growth algorithms.
23. Explain market basket analysis and its applications.

24. Describe different types of association rules such as multilevel and multidimensional rules.

25. Apply the Apriori algorithm to generate frequent itemsets from a given transaction database.

26. Illustrate how to construct an FP-Tree using a sample dataset.

27. Apply confidence and support calculations to determine strong association rules.

28. Demonstrate rule generation from frequent itemsets with a suitable example.

29. Apply association rule mining techniques in real-world business scenarios.

30. Given the transaction dataset: T1={Milk, Bread}, T2={Milk, Diaper, Beer, Eggs}, T3={Milk, Diaper, Beer, Coke}, T4={Bread, Milk, Diaper, Beer}, T5={Bread, Milk, Diaper, Coke}. Calculate support for {Milk, Diaper}.

31. Using the dataset T1={Milk, Bread}, T2={Milk, Diaper, Beer, Eggs}, T3={Milk, Diaper, Beer, Coke}, T4={Bread, Milk, Diaper, Beer}, T5={Bread, Milk, Diaper, Coke}. Calculate support for {Milk, Diaper}, calculate confidence for the rule {Milk} → {Diaper}.

32. If Given support(Milk)=30%, support(Bread)=60%, and support(Milk, Bread)=40%, find confidence for rule Bread → Milk.

33. For a dataset of 100 transactions, item A appears in 40 transactions and item B appears in 30 transactions, while both appear together in 20 transactions. Calculate support, confidence, and lift for rule A → B.

34. Apply Apriori algorithm with minimum support = 20% on the following transactions: T1={A,B,C}, T2={A,C}, T3={A,B}, T4={B,C}, T5={A,B,C}.

35. Explain the concept of frequent pattern mining.

36. Describe the FP-Growth algorithm with a neat diagram.

37. What is an FP-Tree? Explain its structure and components.

38. Describe the steps involved in constructing an FP-Tree.

39. Discuss advantages of FP-Growth over Apriori.

40. Demonstrate how frequent patterns are mined from an FP-Tree.

41. Apply FP-Growth algorithm to find frequent itemsets from a sample database.

42. Illustrate conditional FP-Tree construction using an example.

43. Apply pattern growth technique to generate frequent itemsets.

44. Given transactions: T1={A,B}, T2={B,C,D}, T3={A,C,D,E}, T4={A,D,E}, T5={A,B,C}. Construct FP-Tree and find frequent itemsets with minimum support = 2.

45. Given dataset: T1={Milk,Bread}, T2={Bread,Diaper,Beer,Eggs}, T3={Milk,Diaper,Beer,Coke}, T4={Bread,Milk,Diaper,Beer}, T5={Bread,Milk,Diaper,Coke}. Construct FP-Tree with minimum support count = 30%.

46. Find all frequent patterns using FP-Growth method for the dataset: T1={A,B,C}, T2={A,C}, T3={B,C}, T4={A,B,C}, T5={A,B} with minimum support = 2.

48. Explain Class/Concept Descriptions with Data Characterization and Data Discrimination.