**UNIT**

SHORT QUESTIONS:

1. List and define the characteristics of data warehouse
2. Define metadata
3. What is Data Extraction?
4. Define data warehouse and data mart
5. Why to pre-process the data ? explain briefly
6. List and explain OLAP operations shortly
7. What are the characteristics of star schema
8. List any 4 differences between OLTP and OLAP
9. Compare OLTP and OLAP Systems.
10. What is data warehouse metadata?
11. Explain the differences between star and snowflake schema.
12. In the context of data warehousing what is data transformation?
13. Define Slice and Dice operation.
14. List the characteristics of a data ware house.
15. What are the various sources for data warehouse?
16. What is data warehouse?
17. Differentiate fact table and dimension table.
18. Briefly discuss the schemas for multidimensional databases.
19. How is a data warehouse different from a database? How are they similar?
20. What is dimensionality reduction?

LONG QUESTIONS:

1. . Draw and explain 3-tier architecture of a data warehouse
2. . List and explain OLAP operations with an example
3. a)explain different types of OLAP servers
4. b) draw and explain star, snowflake and fact constellation schema for Hospital management system
   1. Explain ETL process
   2. Explain fully additive ,semi additive and non additive measures with an example
5. List and explain the differences between OLTP and OLAP
6. Draw and explain star, snowflake and fact constellation schema for all electronics sales data
7. List the differences between OLAP and OLTP
8. Draw and explain the 3-tier architecture of data warehousing.
9. Write a short notes on following a) additive measures b) fact less facts c) Dimension table characteristics.
10. Explain star schema and snowflake schema with an example. List the advantages and disadvantages if any.
11. Design snowflake schema for hospital management system.
12. Explain OLAP Operations in detail.
13. What is Data warehouse? What are the characteristics of data warehouse?
14. Draw and explain types of the architecture
15. Briefly explain about ROLAP, MOLAP, HOLAP
16. What is Data Model? Types of the Data Model?
17. Give the difference between operational database and informational database
18. Write in detail about the architecture and implementation of the data warehouse.

# Diagrammatically illustrate and discuss the three tier data warehousing architecture.

(OR) Write a detailed diagram describe the general architecture of data warehouse. (OR) Describe the data warehouse architecture with a neat diagram.

# List and discuss the major features of a data warehouse.

1. Discuss the various types of warehouse schema with suitable example. (OR) What do you understand about database schemas? Explain.
2. Describe OLAP operations in multidimensional data model.

# Explain the types of OLAP server in detail.

1. Enumerate the building blocks of a data warehouse. Explain the importance of metadata in a data

# warehouse environment. What are the challenges in metadata management?

1. Compare and contrast the data warehouse and operational DB with various features.

# Explain in detail about the different kinds of data on which data mining can be applied.

UNIT 2

SHORT QUESTIONS:

* 1. List the methods for filling missing values
  2. Define KDD process
  3. Define concept Hierarchy
  4. Define the process of KDD
  5. Write and explain formulae to normalize the values using a)min-max normalization b)Z-score normalization
  6. Define transactional database with an example
  7. What are the different forms of data preprocessing
  8. Define bining methods
  9. What is descriptive and predictive data mining?
  10. List out the functions of OLAP servers in the data warehouse architecture.
  11. Differentiate data mining and data warehousing.
  12. What do you understand about knowledge discovery?
  13. What is the need for preprocessing the data?
  14. What is parallel mining of concept description?
  15. What is concept description?
  16. Mention the various tasks to be accomplished as part of data pre-processing.
  17. What is data cleaning?
  18. Define Data mining.

LONG QUESTIONS:

1. What is data mining? List and explain the motivating challenges of data mining
2. With the help of a neat diagram Explain data mining as a step process of knowledge discovery.
3. a)Given the following measurements for the variable age
4. 18,22,25,42,28,43,33,35,56,28
5. Standardize the variable by using Z-score
6. b) explain data reduction methods
7. Briefly discuss the data smoothing methods with an example
8. b)Explain how to handle missing values with an example
9. Explain the methods of attribute selection
10. Use the two methods below to normalize the following group of data: 200, 300, 400, 600, 1000
11. a)min-max normalization by setting min=0 and max=1

b)z-score normalization

1. Perform Decimal scaling normalization on sal attribute in below dataset



1. Explain challenges of DM
2. Explain Data Transformation methods\
3. Design and explain the process of KDD
4. List and explain the various ways of handling missing data
5. Perform binning on following set of data
6. 2,3,7,11,15,18,19,14,22,4,9,21,27.
7. Write a short notes on the following a)discretization and binarization b)PCA
8. Use the two methods below to normalize the following group of data a)min max normalization b) Z score normalization { 200,300,400,600,1000}
9. Explain data preprocessing techniques.

# Describe the architecture of typical data mining system with neat Sketch.

1. Explain the Steps of Knowledge Discovery in Databases with neat Sketch.

# Describe the data mining functionality and examine. What kinds of patterns can be mined.

1. Explain the classification of Data Mining Systems.

# Describe the various issues in datamining techniques.

1. Discuss the various data mining techniques.

# Explain the need and steps involved in data preoprocessing.

1. List out the primitives for specifying a data mining task.

# Describe how concepts hierarchies are usuful in data mining.

1. What are the various issued addressed during data integration?
2. Describe the various techniques for data preprocessing with examples.

# Explain the various primitives for specifying a data mining task.

1. Describe the various descriptive statistical measures for data mining.

# Briefly explain the data mining functionalities and examine,what kinds of patterns can be mined?

1. Elaborately explain the discretization and concepts hierarchy generation for numeric data and categorical data.

UNIT 3

SHORT QUESTIONS:

1. State apriori property
2. Define closed item set and maximal item set
3. Mention the importance of association rule mining
4. State the property of Apriori
5. Define support and confidence in Association rule mining with formulae
6. Mention the importance of association rule mining
7. Explain the technique join and prune in apriori algorithm
8. What is itemset, frequent itemset
9. List two interesting measures for association rules.
10. Frequent Item Set Generation
11. Give the difference between Boolean association rule and quantitative association rule.
12. List the techniques to improve the efficiency of Apriori algorithm.
13. Define support and confidence in Association rule mining.
14. What is FP growth?
15. How Meta rules are useful in constraint based association mining.
16. Mention few approaches to mining Multilevel Association Rules.
17. How rules do help in mining?
18. What is mean by the frequency item set property?

LONG QUESTIONS:

1. a)Explain the procedure to mining closed data item set

b)Explain fp-growth algorithm .find frequent itemsets using fp growth algorithm and construct fp tree for the below dataset



2. A) Explain Apriori algorithm for mining association rules

b) generate frequent itemset in below transactional dataset using apriori algorithm (min-sup=40% min-conf=70%)

|  |  |
| --- | --- |
| **Tid** | **Items brought** |
| T1 | Bread, butter, milk |
| T2 | Bread, butter |
| T3 | Biscuits, cookies,diapers |
| T4 | Bread, butter, milk, diapers |
| T5 | Biscuits,diapers |

1. Consider the following transactional data base , generate association rules and identify strong association rules using APRIORI algorithm with support=60% confidence=80%



1. Discuss compact representation of frequent itemset (closed and maximal) with min-sup=3

|  |  |
| --- | --- |
| **Tid** | **items** |
| T1 | {A,B,C,D} |
| T2 | {A,B,C,D} |
| T3 | {A,B,C} |
| T4 | {B,C,D} |
| T5 | {C,D} |

1. Consider the following transactional data base , generate association rules and identify strong association rules using APRIORI algorithm with support=50% confidence=70%

|  |  |
| --- | --- |
| **Tid** | **Items** |
| T1 | {1,3,4} |
| T2 | {2,3,5} |
| T3 | {1,2,3,5} |
| T4 | {2,5} |

1. Explain fp-growth algorithm .find frequent itemsets using fp growth algorithm and construct fp tree for the below dataset

|  |  |  |
| --- | --- | --- |
| **Tid** | | **Items brought** |
| T1 | F,A,C,D,G,I,M,P | |
| T2 | A,B,C,F,L,M,O | |
| T3 | B,F,H,J,Q,W | |
| T4 | B,C,K,S,P | |
| T5 | A,F,C,E,L,P,M,N | |

1. Consider the following transactional data base , generate association rules and identify strong association rules using APRIORI algorithm with support=2 confidence=60%

|  |  |
| --- | --- |
| **Tid** | **items** |
| T1 | Hotdogs, bun, ketchup |
| T2 | Hotdogs, bun |
| T3 | Hotdogs, coke, chips |
| T4 | coke, chips |
| T5 | Chips,ketchup |
| T6 | Hotdogs, coke, chips |

1. Write and explain FP-growth algorithm with an example
2. Write a short notes on a)super set b)frequent item set c) antimonotonocity d) closed frequent item set
3. A database has five transactions. Let min\_sup=60% and min\_confidence=80%
   1. TID LIST OF ITEMS
   2. T1 {M,O,N,K,E,Y}
   3. T2 {D,O,N,K,E,Y}
   4. T3 {M,A,K,E}
   5. T4 {M,U,C,K,Y}
   6. T5 {C,O,O,K,I,E}

a) generate frequent itemsets using apriori algorithm

b) list all the strong association rules

1. explain partitioning algorithm with an example.
2. Explain FP-Growth algorithm with an example.

UNIT 4

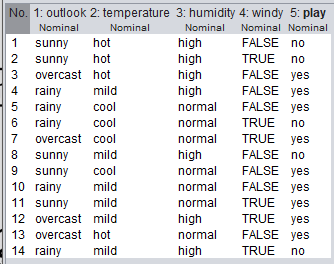
SHORT QUESTIONS:

1. What is rule classification
2. Define Decision tree
3. How prediction is different from classification
4. Write a short note on confusion matrix
5. Define two steps in classification
6. Define root node, internal node and leaf node
7. Write short note on entropy and information gain with formulae
8. List the measures for selecting best split in decision tree construction
9. What is tree pruning?
10. What is classification?
11. What is the objective function of the K-means algorithm?
12. The naïve Bayes classifier makes what assumption that motivates its name?
13. What is an outlier? (OR) Define outliers. List various outlier detection approaches.
14. What is Bayesian theorem?
15. What is Association based classification?
16. Compare the advantages of and disadvantages of eager classification (e.g., decision tree) versus lazy classification (k-nearest neighbor)
17. What is called Bayesian classification?

LONG QUESTIONS:

1. What is Bayesian belief network ? explain with an example
2. Write KNN classification algorithm with an example
3. Explain Naïve-Bayers classifiers and predict instance to the class (weather data set)

if outlook=rainy,temp=cool, humidity=high, windy=true, play =?



4. a)Explain how evaluation of classifiers will be done using confusion matrix

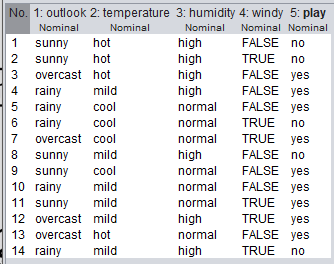
b)List the advantages and disadvantage of using decision trees

5. Describe the general approaches to solve a classification problem

6. Consider the following data, apply KNN algorithm to find class when acid durability=3 and strength=7

|  |  |  |  |
| --- | --- | --- | --- |
| name | acid durability | strength | class |
| Type1 | 7 | 7 | Bad |
| Type2 | 7 | 4 | Bad |
| Type3 | 3 | 4 | Good |
| Type4 | 1 | 4 | Good |

7. Construct a Decision tree using J48 for below data set



8. Explain Naïve-Bayers classifiers and predict instance to the class (weather data set)

if outlook=sunny play =?

1. Illustrate general approaches to solve classification.

10. Briefly outline naïve bayesian classifiers with an example

1. Write a short notes on decision tree induction
2. Explain the k-nearest neighbor classification and its characteristics.
3. Explain the methods for selecting best split.
4. Explain about the different measures used for selecting the splitting attribute.

# Write the algorithm to discover frequent itemsets without candidate generation and explain it with an exampled

# Discuss Apriori Algorithm with a suitable example and explain how its efficiency can be improved.

# Discuss mining of multi-level association rules from transactional databases.

1. What are classification rules? How is regression related to classification?

# Explain with example the various steps in Decision tree induction.

1. State Bayes theorem and discuss how Bayesian classifiers work.

# What back propagation? How does it work?

1. Describe the various techniques for improving classifiers accuracy.

UNIT 5

SHORT QUESTIONS:

1. Define clustering
2. What are the requirements for cluster analysis
3. What are outliers? What is the importance of outlier analysis?
4. Give a short note on K-Means algorithm
5. List the applications of clustering analysis
6. Differentiate clustering and classification
7. Define clustering
8. List agglomerative and hierarchical clustering techniques
9. Compare clustering and classification.
10. What is meant by hierarchical clustering?
11. Why tree pruning useful in decision tree induction?
12. What do you go for clustering analysis?
13. What are the requirements of cluster analysis?
14. What is mean by cluster analysis?
15. Mention the advantages of hierarchical clustering.

LONG QUESTIONS:

1. a) write about evaluation of clustering algorithms

b) write the key issues in hierarchical clustering algorithm

2. Explain agglomerative methods and construct a dendrogram for below distance matrix using single linkage

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| 1 | 0 |  |  |  |  |
| 2 | 9 | 0 |  |  |  |
| 3 | 3 | 7 | 0 |  |  |
| 4 | 6 | 5 | 9 | 0 |  |
| 5 | 11 | 10 | 2 | 8 | 0 |

3. a) Write about evaluation of clustering algorithm

b) what are the key issues in hierarchical clustering

4. Explain divisive method and construct a dendrogram for below distance matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| A | 0 |  |  |  |  |
| B | 1 | 0 |  |  |  |
| C | 2 | 2 | 0 |  |  |
| D | 2 | 4 | 1 | 0 |  |
| E | 3 | 3 | 5 | 3 | 0 |

5. Explain the outlier detection methods.

6. Discuss key issues involved in hierarchical and K-Means clustering

7. Explain agglomerative methods and construct a dendrogram for below distance matrix using complete linkage

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| 1 | 0 |  |  |  |  |
| 2 | 9 | 0 |  |  |  |
| 3 | 3 | 7 | 0 |  |  |
| 4 | 6 | 5 | 9 | 0 |  |
| 5 | 11 | 10 | 2 | 8 | 0 |

1. what are the key issues in hierarchical clustering
2. List the requirements of clustering in data mining.
3. Apply k-means algorithm on following values 12, 24, 20,42,15,14,11,9 where k=2.
4. Explain PAM algorithm
5. Write short notes on hierarchical clustering
6. Illustrate the key issues involved in hierarchical clustering
7. Briefly outline the outlier analysis.
8. Discuss the different types of clustering methods.

# Discuss the working of PAM algorithm.

1. Describe K-means clustering with an example.

# Explain hierarchical methods of clustering.

1. Explain the various methods for detecting outliers.

# Explain the mining of spatial databases.

1. Discuss the mining of text data mining.

# What are the salient features of times series data ming?

1. What is web mining? Discuss the various web mining techniques.

# Discuss in detail the application of Data mining for financial data analysis?

1. Discuss the application of data mining in business.

# Discuss in detail of applications of data mining for biomedical and DNA data analysis and telecommunication industry.

1. Discuss the social impacts of data mining sy**stems**