1. **Differentiate OLAP and OLTP:

- OLAP (Online Analytical Processing) is used for complex queries and data analysis, focusing on read-intensive operations.
- OLTP (Online Transaction Processing) handles day-to-day transactions, emphasizing high volume, speed, and accuracy of transactional data processing.

2. Differentiate ROLAP and MOLAP data mining servers:

- ROLAP (Relational OLAP) uses relational databases and SQL for dynamic data storage, suitable for large datasets.
- MOLAP (Multidimensional OLAP) uses multidimensional databases, pre-aggregating data for faster querying and analysis, ideal for smaller, well-defined datasets.

3. List the attribute types:

- Nominal
- Ordinal
- Interval
- Ratio
- Binary
- Categorical
- Continuous

4. List the Strengths of Association Rule Mining:

- Discovers hidden patterns
- Easy to understand results
- Applicable to various domains
- Identifies relationships between items
- Can handle large datasets
- Provides actionable insights

5. Define Confusion Matrix:

A confusion matrix is a table used to evaluate the performance of a classification model. It displays true positives, true negatives, false positives, and false negatives, helping to calculate accuracy, precision, recall, and other metrics.

6. List the Decision Tree Induction algorithms steps:

- Select the best attribute using a selection criterion
- Split the dataset into subsets based on the attribute
- Recursively apply the process to each subset
- Stop when all instances belong to the same class or no further attributes are available
- Prune the tree to improve generalization

Set-2

1. List different types of attributes. Give an example for each:

- Nominal: Categorical data without order (e.g., colors: red, blue, green).
- Ordinal: Categorical data with order (e.g., rankings: first, second, third).
- Interval: Numeric data without a true zero (e.g., temperature in Celsius).
- Ratio: Numeric data with a true zero (e.g., weight: 50 kg, 60 kg).
- Binary: Data with two categories (e.g., yes/no).
- Categorical: Data that can be divided into categories (e.g., types of fruit).
- Continuous: Data that can take any value within a range (e.g., height).

2. **Compute the distance between the two data objects given as X (22,1,24,10,46) and Y (12,2,24,23,46) using Manhattan and Euclidean distance:**

- **Manhattan Distance:**
  $|22-12|+|1-2|+|24-24|+|10-23|+|46-46| = 10+1+0+13+0 = 24$

- **Euclidean Distance:**
  $\sqrt{(22-12)^2+(1-2)^2+(24-24)^2+(10-23)^2+(46-46)^2}$
  $= \sqrt{100+1+0+169+0} = \sqrt{270} \approx 16.43$

4. Define ECLAT algorithm:

The ECLAT (Equivalence Class Clustering and bottom-up Lattice Traversal) algorithm is used in frequent itemset mining. It utilizes a depth-first search strategy to traverse itemsets, storing transaction IDs for each itemset to find frequent patterns efficiently.

5. Compare Predictive and Descriptive data mining techniques:

- Predictive: Focuses on predicting future outcomes based on historical data (e.g., regression, classification).
- Descriptive: Identifies patterns and relationships in data to describe its structure and contents (e.g., clustering, association rule mining).

6. What is item set and frequent item set:

- Item Set: A collection of one or more items (e.g., {milk, bread, butter}).
- Frequent Item Set: An item set that appears frequently in a dataset, satisfying a minimum support threshold (e.g., {milk, bread} appearing in 50% of transactions).