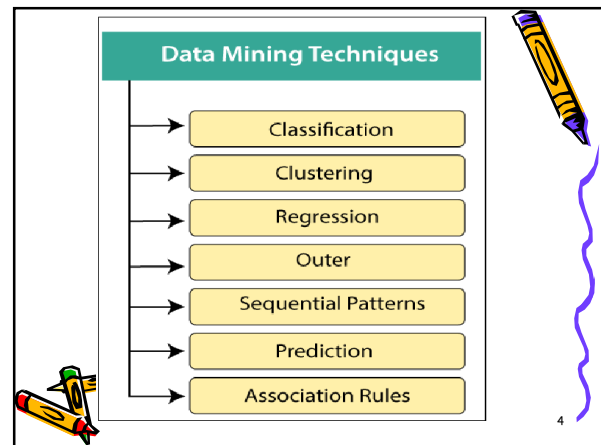# Data Mining Techniques

1

Introduction

- Data mining includes the utilization of refined data analysis tools to find previously unknown, valid patterns and relationships in huge data sets.
- These tools can incorporate statistical models, machine learning techniques, and mathematical algorithms, such as neural networks or decision trees.
- Thus, data mining incorporates analysis and prediction.

2

Introduction

- In recent data mining projects, various major data mining techniques have been developed and used, including association, classification, clustering, prediction, sequential patterns, and regression.

3

Data Mining Techniques

- Classification
- Clustering
- Regression
- Outer
- Sequential Patterns
- Prediction
- Association Rules

4

Classification:

- This technique is used to obtain important and relevant information about data and metadata.
- This data mining technique helps to classify data in different classes.
- Data mining techniques can be classified by different criteria, as follows:

5

Classification:

- **Classification of Data mining frameworks as per the type of data sources mined:**
  - This classification is as per the type of data handled.
  - For example, multimedia, spatial data, text data, time-series data, World Wide Web, and so on..

6

Classification:

- **Classification of data mining frameworks as per the database involved:**
  - This classification based on the data model involved.
  - For example. Object-oriented database, transactional database, relational database, and so on..

7

Classification:

- **Classification of data mining frameworks as per the kind of knowledge discovered:**
  - This classification depends on the types of knowledge discovered or data mining functionalities.
  - For example, discrimination, classification, clustering, characterization, etc.

8

Classification:

- **Classification of data mining frameworks according to data mining techniques used:**
  - This classification is as per the data analysis approach utilized, such as neural networks, machine learning, genetic algorithms, visualization, etc.
  - The classification can also take into account, the level of user interaction involved in the data mining procedure, such as query-driven systems, autonomous systems, or interactive exploratory systems.

9

Clustering

- Clustering is a division of information into groups of connected objects.
- Describing the data by a few clusters mainly loses certain confine details, but accomplishes improvement.
- Data modeling puts clustering from a historical point of view rooted in statistics, mathematics, and numerical analysis.

10

Clustering

- From a machine learning point of view, clusters relate to hidden patterns, the search for clusters is unsupervised learning, and the subsequent framework represents a data concept.
- From a practical point of view, clustering plays an extraordinary job in data mining applications.

11

Clustering

- Clustering analysis is a data mining technique to identify similar data.
- This technique helps to recognize the differences and similarities between the data.
- Clustering is very similar to the classification, but it involves grouping chunks of data together based on their similarities.

12

Regression

- Regression analysis is the data mining process used to identify and analyze the relationship between variables because of the presence of the other factor.
- It is used to define the probability of the specific variable.
- Regression, primarily is a form of planning and modeling.

13

Regression

- For example, we might use it to project certain costs, depending on other factors such as availability, consumer demand, and competition.
- Primarily it gives the exact relationship between two or more variables in the given data set.

14

Association Rules

- This data mining technique helps to discover a link between two or more items.
- It finds a hidden pattern in the data set.
- Association rules are if-then statements that support to show the probability of interactions between data items within large data sets in different types of databases.

15

Association Rules

- Association rule mining has several applications and is commonly used to help sales correlations in data or medical data sets.
- The way the algorithm works is that you have various data, For example, a list of grocery items that you have been buying for the last six months.
- It calculates a percentage of items being purchased together.

16

Association Rules

- There are three major measurements technique
  - **Lift:** This measurement technique measures the accuracy of the confidence over how often item B is purchased.

  (Confidence) / (item B)/ (Entire dataset)

17

- **Support:** This measurement technique measures how often multiple items are purchased and compared it to the overall dataset.

  (Item A + Item B) / (Entire dataset)

- **Confidence:** This measurement technique measures how often item B is purchased when item A is purchased as well.

  (Item A + Item B)/ (Item A)

18

Outer detection:

- This type of data mining technique relates to the observation of data items in the data set, which do not match an expected pattern or expected behavior.
- This technique may be used in various domains like intrusion, detection, fraud detection, etc. It is also known as Outlier Analysis or Outilier mining.

19

Outer detection

- The outlier is a data point that diverges too much from the rest of the dataset.
- The majority of the real-world datasets have an outlier.
- Outlier detection plays a significant role in the data mining field.
- Outlier detection is valuable in numerous fields like network interruption identification, credit or debit card fraud detection, detecting outlying in wireless sensor network data, etc.

20

Sequential Patterns

- The sequential pattern is a data mining technique specialized for **evaluating sequential data** to discover sequential patterns.
- It comprises of finding interesting subsequences in a set of sequences, where the stake of a sequence can be measured in terms of different criteria like length, occurrence frequency, etc.

21

Sequential Patterns

- In other words, this technique of data mining helps to discover or recognize similar patterns in transaction data over some time.

22

Prediction

- Prediction used a combination of other data mining techniques such as trends, clustering, classification, etc.
- It analyzes past events or instances in the right sequence to predict a future event.

23

Questions