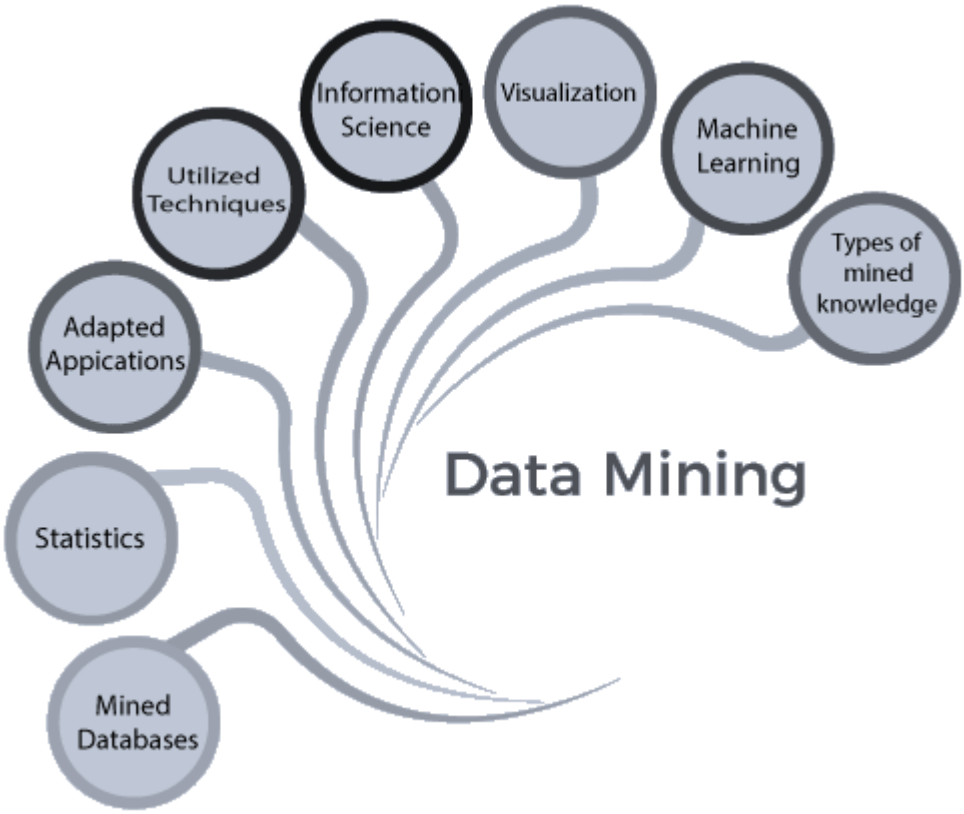
Classification of Data Mining Systems

Data Mining is considered as an interdisciplinary field. It includes a set of various disciplines such as statistics, database systems, machine learning, visualization and information sciences. Classification of the data mining system helps users to understand the system and match their requirements with such systems.

To understand the system and meet the desired requirements, data mining can be classified into the following systems:



1. **Classification based on the mined Databases**
2. **Classification based on the type of mined knowledge**
3. **Classification based on utilized techniques**
4. **Classification based on adapted applications**

Classification Based on the mined Databases

A data mining system can be classified based on the types of databases that have been mined. A database system can be further segmented based on distinct principles, such as data models, types of data, etc., which further assist in classifying a data mining system.

For example, if we want to classify a database based on the data model, we need to select either **relational, transactional, object-relational or data warehouse mining systems.**

Classification Based on the type of Knowledge Mined

A data mining system categorized based on the kind of knowledge mind may have the following functionalities:

- Characterization
- Discrimination
- Association and Correlation Analysis
- Classification
- Prediction
- Outlier Analysis

Classification Based on the Techniques Utilized

- A data mining system can also be classified based on the type of techniques that are being incorporated. These techniques can be assessed based on the involvement of user interaction involved or the methods of analysis employed.

- Statistics
- Machine Learning
- Neural Network
- Data Warehouse

Classification Based on the Applications Adapted

Data mining systems classified based on adapted applications adapted are as follows:

- Finance
- Telecommunications
- DNA
- Stock Markets
- E-mail

Data mining primitives.

A data mining query is defined in terms of the following primitives

1. Task-relevant data: This is the database portion to be investigated. For example, suppose that you are a manager of All Electronics in charge of sales in the India and China. In particular, you would like to study the buying trends of customers in China. Rather than mining on the entire database. These are referred to as relevant attributes

2. The kinds of knowledge to be mined: This specifies the data mining functions to be performed, such as characterization

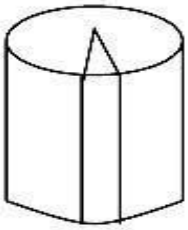
- Discrimination
- Association
- Classification
- Clustering
- Outlier analysis

For instance, if studying the buying habits of customers in India, you may choose to mine associations between customer profiles and the items that these customers like to buy

3. Background knowledge: Users can specify background knowledge, or knowledge about the domain to be mined. This knowledge is useful for guiding the knowledge discovery process, and for evaluating the patterns found. There are several kinds of background knowledge.

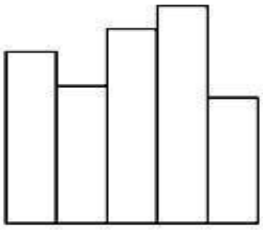
4. Interestingness measures: These functions are used to separate uninteresting patterns from knowledge. They may be used to guide the mining process, or after discovery, to evaluate the discovered patterns. Different kinds of knowledge may have different interestingness measures.

5. Presentation and visualization of discovered patterns: This refers to the form in which discovered patterns are to be displayed. Users can choose from different forms for knowledge presentation, such as rules, tables, charts, graphs, decision trees, and cubes.



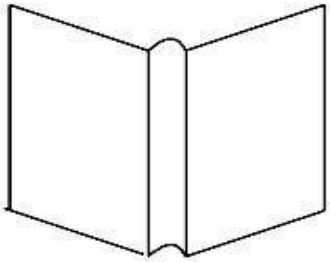
Task-relevant data

- database or data warehouse name
- database tables or data warehouse cubes
- conditions for data selection
- relevant attributes or dimensions
- data grouping criteria



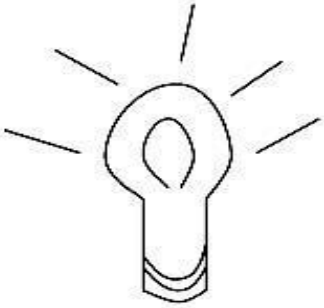
Knowledge type to be mined

- characterization
- discrimination
- association
- classification/prediction
- clustering



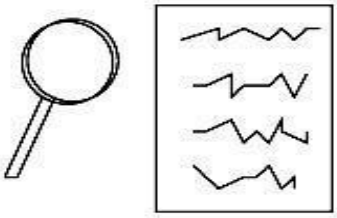
Background knowledge

- concept hierarchies
- user beliefs about relationships in the data



Pattern interestingness measurements

- simplicity
- certainty (e.g., confidence)
- utility (e.g., support)
- novelty



Visualization of discovered patterns

- rules, tables, reports, charts, graphs, decisison trees, and cubes
- drill-down and roll-up