DATA MINING - TERMINOLOGIES

Data Mining

Data mining is defined as extracting the information from a huge set of data. In other words we can say that data mining is mining the knowledge from data. This information can be used for any of the following applications −

- Market Analysis
- Fraud Detection
- Customer Retention
- Production Control
- Science Exploration

Data Mining Engine

Data mining engine is very essential to the data mining system. It consists of a set of functional modules that perform the following functions −

- Characterization
- Association and Correlation Analysis
- Classification
- Prediction
- Cluster analysis
- Outlier analysis
- Evolution analysis

Knowledge Base

This is the domain knowledge. This knowledge is used to guide the search or evaluate the interestingness of the resulting patterns.

Knowledge Discovery

Some people treat data mining same as knowledge discovery, while others view data mining as an essential step in the process of knowledge discovery. Here is the list of steps involved in the knowledge discovery process −

- Data Cleaning
- Data Integration
- Data Selection
- Data Transformation
- Data Mining
- Pattern Evaluation
- Knowledge Presentation

User interface

User interface is the module of data mining system that helps the communication between users and the data mining system. User Interface allows the following functionalities −

- Interact with the system by specifying a data mining query task.

- Providing information to help focus the search.
- Mining based on the intermediate data mining results.
- Browse database and data warehouse schemas or data structures.
- Evaluate mined patterns.
- Visualize the patterns in different forms.

Data Integration

Data Integration is a data preprocessing technique that merges the data from multiple heterogeneous data sources into a coherent data store. Data integration may involve inconsistent data and therefore needs data cleaning.

Data Cleaning

Data cleaning is a technique that is applied to remove the noisy data and correct the inconsistencies in data. Data cleaning involves transformations to correct the wrong data. Data cleaning is performed as a data preprocessing step while preparing the data for a data warehouse.

Data Selection

Data Selection is the process where data relevant to the analysis task are retrieved from the database. Sometimes data transformation and consolidation are performed before the data selection process.

Clusters

Cluster refers to a group of similar kind of objects. Cluster analysis refers to forming group of objects that are very similar to each other but are highly different from the objects in other clusters.

Data Transformation

In this step, data is transformed or consolidated into forms appropriate for mining, by performing summary or aggregation operations.