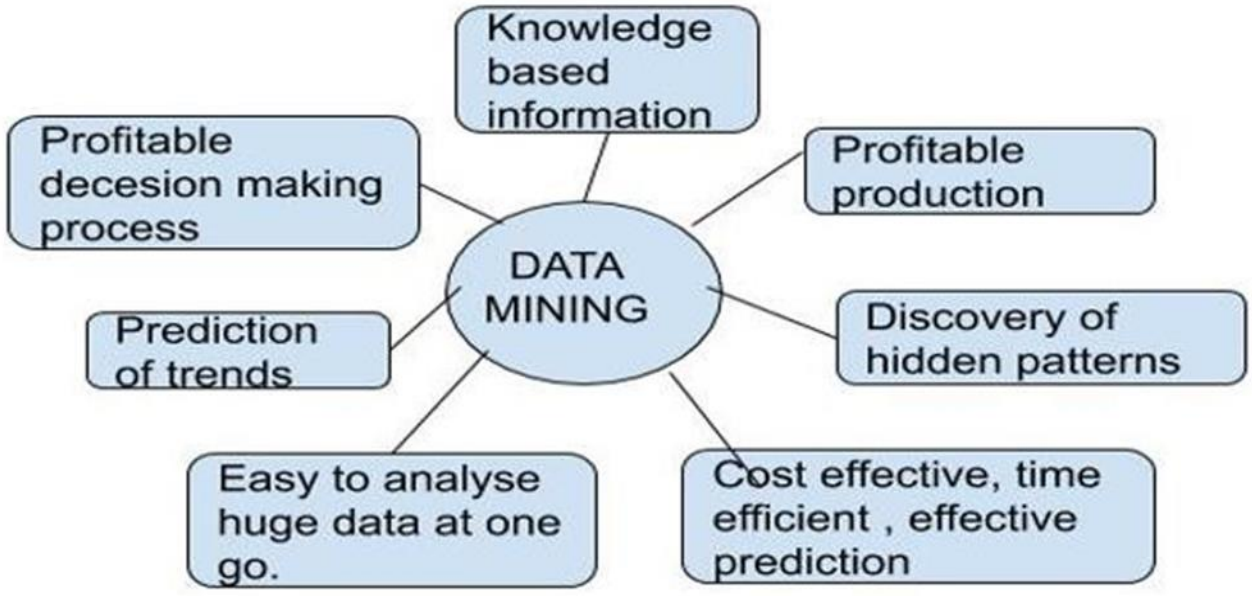
Data Mining Concepts

What Is Data Mining?

Data mining refers to extracting or mining knowledge from large amounts of data. The term is actually a misnomer. Thus, data mining should have been more appropriately named as knowledge mining which emphasis on mining from large amounts of data. There are various important parameters in Data Mining, such as association rules, classification, clustering, and forecasting. Some of the key features of Data Mining are

- Prediction of Patterns based on trends in the data.
- Calculating the predictions for the outcomes.
- Creating information in response to the analysis
- Focusing on greater databases.
- Clustering the visual data



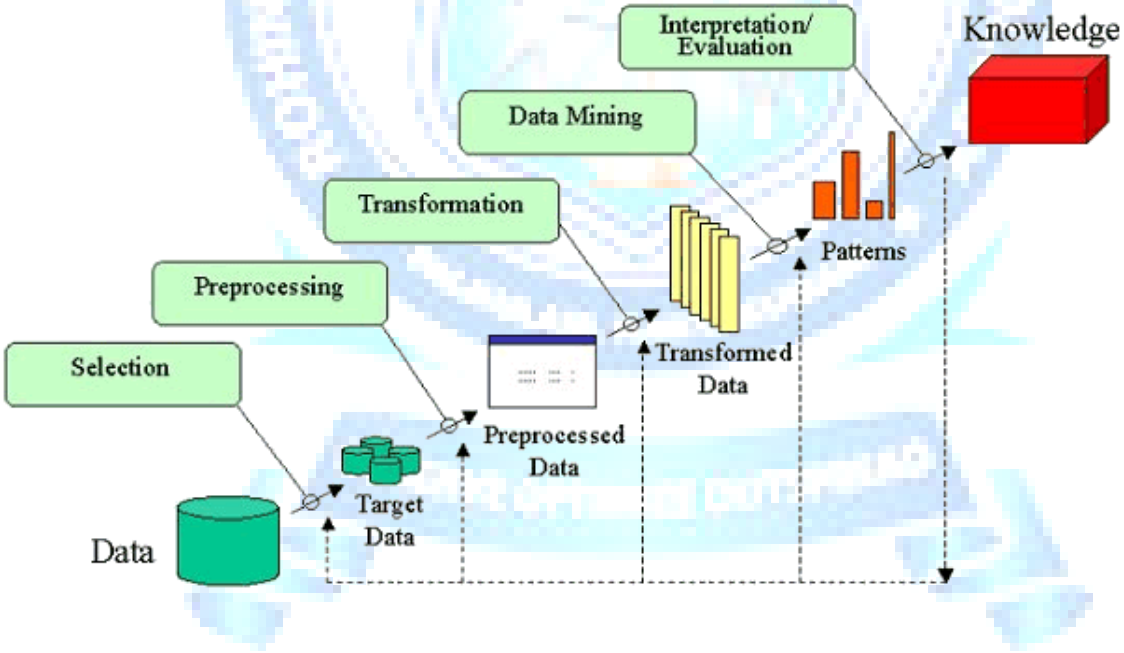
Data Mining:

Extracting and 'Mining' knowledge from large amounts of data. "Gold Mining from rock or sand" is same as "Knowledge mining from data" Other terms for Data Mining:

- Knowledge Mining
- Knowledge Extraction
- Pattern Analysis
- Data Archeology
- Data Dredging

Data Mining is not same as KDD (Knowledge Discovery from Data)

Data Mining is a step in KDD



- Data Cleaning – In Data Cleaning the noise and inconsistent data is removed.
- Data Integration − multiple data sources are combined.
- Data Selection − only the relevant data is selected from the database.
- Data Transformation − data is consolidated into appropriate forms for mining by performing summary or aggregation operations.

- Data Mining − this is an intelligent step in which various methods are applied to extract data patterns.
- Pattern Evaluation − data patterns, which can be in different forms like trees, associations, clusters, etc. are evaluated.
- Knowledge Presentation − this step finally provides knowledge

ADVANTAGES OF DATA MINING

- It is helpful to predict future trends
- It signifies customer habits
- Helps in decision making
- Increase company revenue
- It depends upon market-based analysis
- Quick fraud detection

DISADVANTAGES

- It violates user privacy
- Additional irrelevant information
- Misuse of information
- Accuracy of data

DBMS VERSUS DATA MINING

DBMS and Data Mining

	DBMS	Data Mining
Task	Extraction of detailed and summary data	Knowledge discovery of hidden patterns and insights
Type of result	Information	Insight and Prediction
Method	Deduction (Ask the question, verify with data)	Induction (Build the model, apply it to new data, get the result)
Example question	Who purchased mutual funds in the last 3 years?	Who will buy a mutual fund in the next 6 months and why?

❖ **DBMS**

- DBMS, sometimes just called a database manager, is a collection of computer programs that is dedicated for the management (i.e. organization, storage and retrieval) of all databases that are installed in a system (i.e. hard drive or network).
- There are different types of Database Management Systems existing in the world, and some of them are designed for the proper management of databases configured for specific purposes.
- Most popular commercial Database Management Systems are Oracle, DB2 and Microsoft Access.
- All these products provide means of allocation of different levels of privileges for different users, making it possible for a DBMS to be controlled centrally by a single

administrator or to be allocated to several different people. There are four important elements in any Database Management System.

- They are the modeling language, data structures, query language and mechanism for transactions. The modeling language defines the language of each database hosted in the DBMS.
- Currently several popular approaches like hierarchal, network, relational and object are in practice.
- Data structures help organize the data such as individual records, files, fields and their definitions and objects such as visual media

❖ **DATA MINING:**

- Data mining is also known as Knowledge Discovery in Data (KDD). As mentioned above, it is a felid of computer science, which deals with the extraction of previously unknown and interesting information from raw data.
- Due to the exponential growth of data, especially in areas such as business, data mining has become very important tool to convert this large wealth of data in to business intelligence, as manual extraction of patterns has become seemingly impossible in the past few decades.
- For example, it is currently been used for various applications such as social network analysis, fraud detection and marketing. Data mining usually deals with following four tasks: clustering, classification, regression, and association.
- Clustering is identifying similar groups from unstructured data.
- Classification is learning rules that can be applied to new data and will typically include following steps:
- Preprocessing of data, designing modeling, learning/feature selection and Evaluation/validation.
- Regression is finding functions with minimal error to model data. And association is looking for relationships between variables