

Data mining and Warehousing

CSC5901



UNIVERSITY
of **KIGALI**

Unit1:

Introduction on

Data mining and Warehousing

Introduction on Data mining

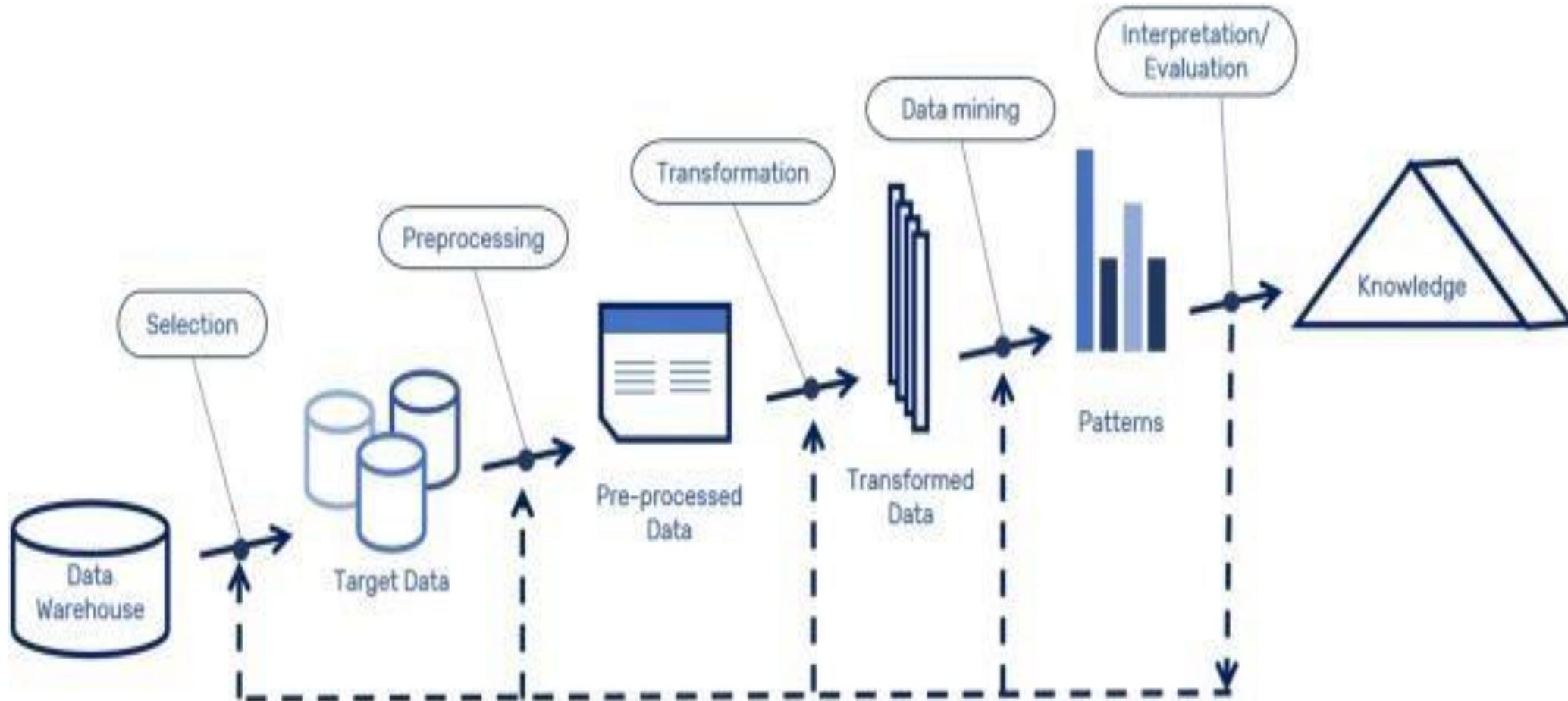
What is Data Mining?

Data mining is techniques that help entrepreneurs, researchers, and government and individuals to extract valuable information from huge sets of data explaining the past and predicting the future via Data analysis.

What is Data Mining?

Data mining is also called *Knowledge Discovery in Database* (**KDD**). The knowledge discovery process includes Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern evaluation, and Knowledge presentation.

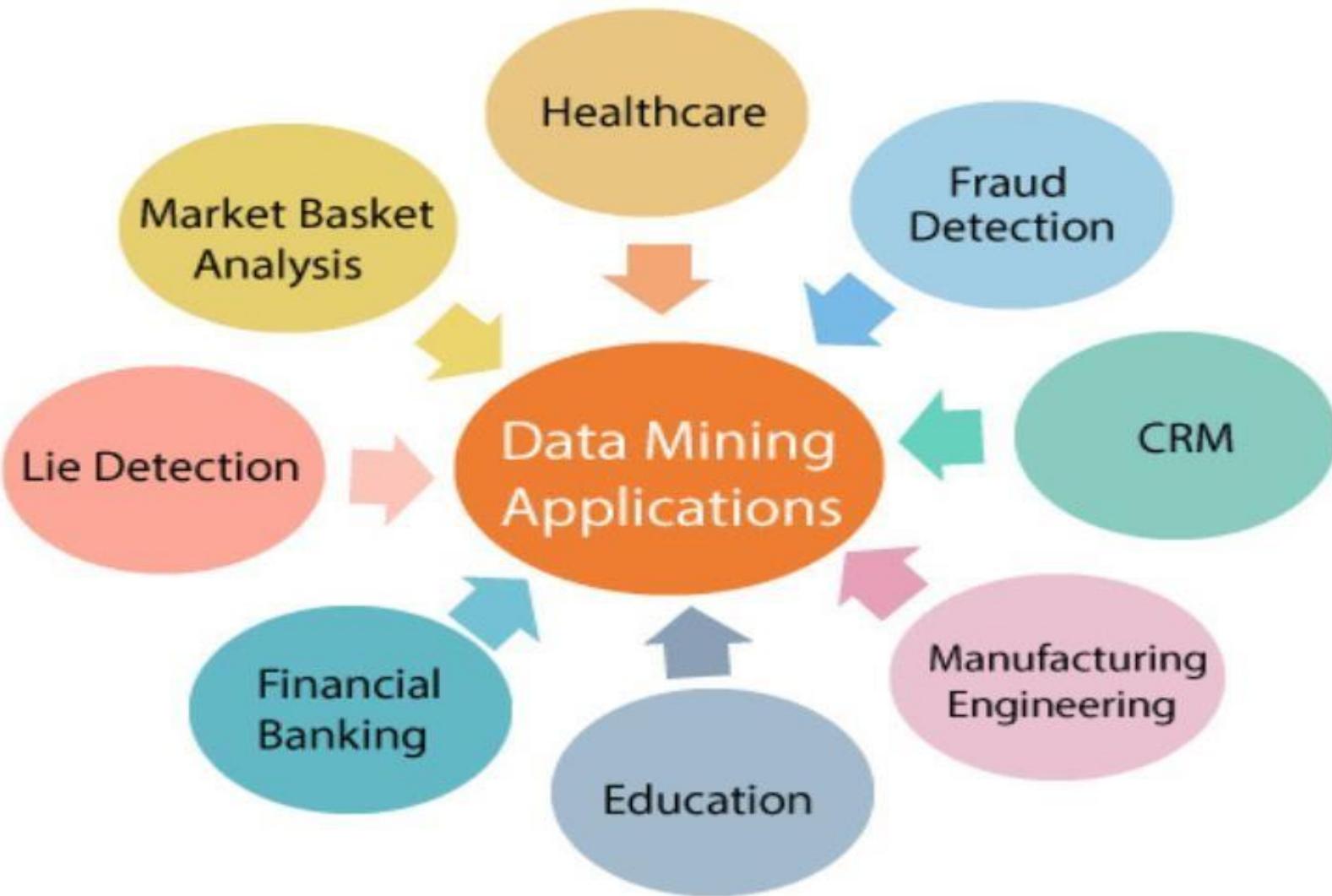
Steps In The Data Mining Process



How this technique work?

uses machine learning, statistics, and AI to extract information to evaluate future events probability

Data Mining Applications



Data mining is highly useful in:

Domains:

- Market Analysis and Management
- Corporate Analysis & Risk Management
- Fraud Detection

Data mining is highly useful in:(cont..)

Market Analysis and Management (Domain)

- **Customer Profiling** – helps determine what kind of people buy what kind of products.
- **Identifying Customer Requirements** – helps in identifying the best products for different customers. It uses prediction to find the factors that may attract new customers.
- **Cross Market Analysis** – Data mining performs Association/correlations between product sales.

Data mining is highly useful in:(cont..)

Market Analysis and Management(Cont..)

- **Target Marketing** – helps to find clusters of model customers who share the same characteristics such as interests, spending habits, income, etc.
- **Determining Customer purchasing pattern** – helps in determining customer purchasing pattern.
- **Providing Summary Information** – Data mining provides us various multidimensional summary reports.

Data mining is highly useful in:(cont..)

Corporate Analysis and Risk Management

- **Finance Planning and Asset Evaluation** – It involves cash flow analysis and prediction, contingent claim analysis to evaluate assets.
- **Resource Planning** – It involves summarizing and comparing the resources and spending.
- **Competition** – It involves monitoring competitors and market directions.

Challenges of Implementation in Data mining



Challenges of Implementation in Data mining

- **Incomplete and noisy data:** The data in the real-world is heterogeneous, incomplete, and noisy. Data in huge quantities will usually be inaccurate or unreliable.
- **Data Distribution:** Real-worlds data is usually stored on various platforms eg email, individual systems, It is not easy to arrange and store as useful information
- **Complex Data:** Real-world data is heterogeneous(different format), and it is not easy to extract useful information.

Challenges of Implementation in Data mining

- **Performance:** The data mining system's performance relies on the efficiency of algorithms and techniques used. If the designed algorithm and techniques used present weakness it will affect performance .
- **Data Privacy and Security:** Data mining usually leads to serious issues in terms of data security, governance, and privacy.
- **Data Visualization:** efficient, and successful data visualization processes is very dificult

Introduction on Warehousing

What is Data Warehousing?

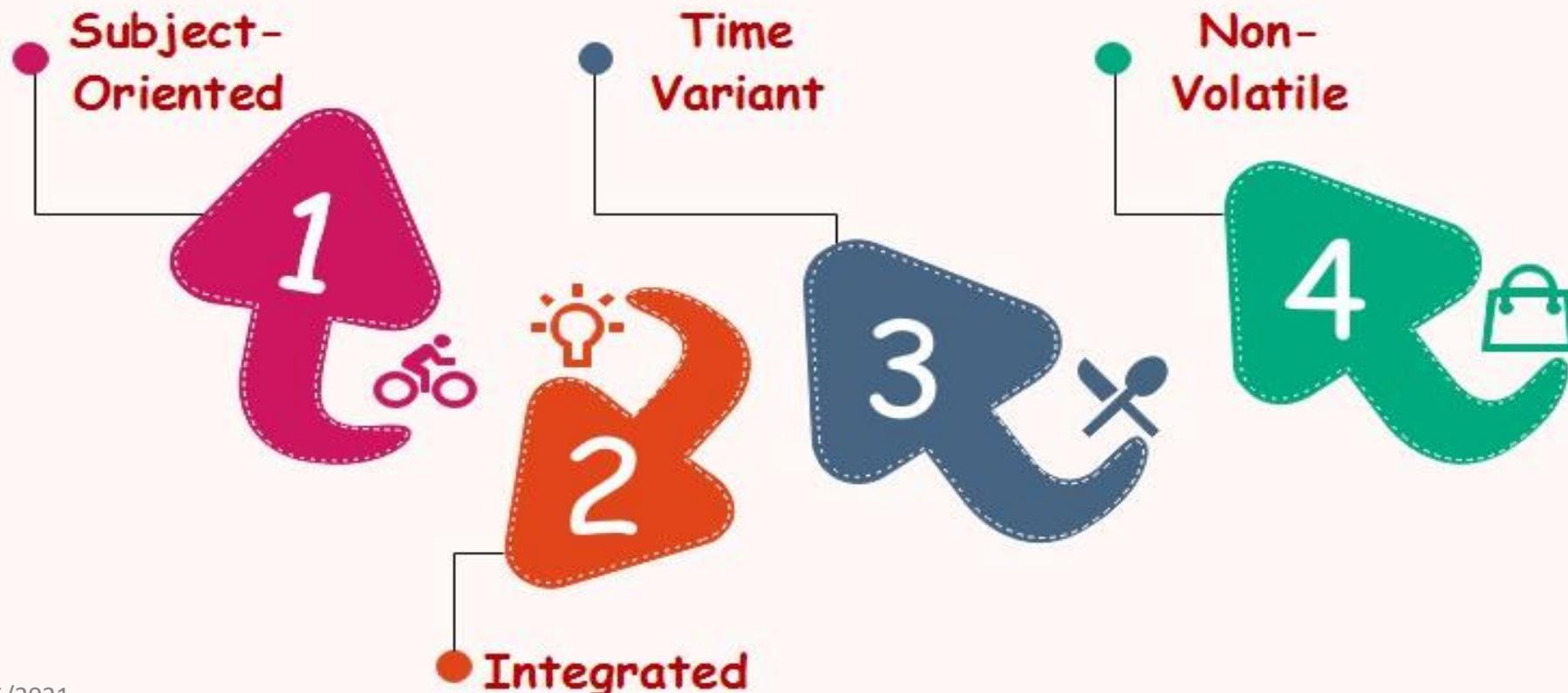
- It is a technique for collecting and managing data from varied sources to provide meaningful business insights to make a difference.
- It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. in an organization.

What is Data Warehousing?

- A Data Warehouse provides integrated, enterprise-wide, historical data and focuses on providing support for decision makers for data modeling and analysis.
- A Data Warehouse is a group of data specific to the entire organization, not only to a particular group of users.
- It is not used for daily operations and transaction processing but used for making decisions.

Characteristics/ Features of Data Warehouse

The key features of Data Warehouse are:



Characteristics/ Features of Data Warehouse

- **Subject Oriented** – it provides information around a subject rather than the organization's ongoing operations. These subjects can be product, customers, suppliers, sales, revenue, etc.
- **Integrated** – A data warehouse is constructed by integrating data from heterogeneous sources such as relational databases, flat files, etc.

Characteristics/ Features of Data Warehouse

- **Time Variant** – The data collected in a data warehouse is identified with a particular time period. The data in a data warehouse provides information from the historical point of view.
- **Non-volatile** – Non-volatile means the previous data is not erased when new data is added to it.

Major functions of a data warehouse are:

- Maintaining past and present records for future decision and prediction.
- Helping organizations to take effective business decisions with precise data analysis.

Why data warehouse?

- Since a data warehouse can gather information quickly and efficiently, it can **enhance business productivity**.
- A data warehouse provides overview on subject , hence, it **helps us in subject management**.—
- **A Data Warehouse Saves Time:** Since business users can quickly access critical data from a number of sources—all in one place

Why data warehouse?

- **A Data Warehouse Enhances Data Quality and Consistency:** A data warehouse implementation includes the conversion of data from numerous source systems into a common format.
- **A Data Warehouse Provides Historical Intelligence:** A data warehouse stores large amounts of historical data of different time periods, to be analyzed and make future predictions.

Need for Data Warehouse



Need for Data Warehouse

- **Business User:** need to view summarized data from the past.
- **Store historical data:** is required to store the time variable data from the past.
- **Make strategic decisions:** Some strategies depending upon the data in the data warehouse. So, data warehouse contributes to making strategic decisions.

Need for Data Warehouse

- **For data consistency and quality:** Bringing the data from different sources at a commonplace, sound as good solution.
- **High response time:** demands of a significant degree of flexibility and quick response time is solution for future production and strategic decisions .

Data Warehouse Applications

- Banking services
- Consumer goods
- Financial Services
- Retail sectors
- Controlled manufacturing

Types of data warehouse

Data mart

Operation data store

Enterprise data warehouse

1. Enterprise Data Warehouse (EDW):

Enterprise Data Warehouse (EDW) is a centralized warehouse. It provides decision support service across the enterprise. It offers a unified approach for organizing and representing data. It also provide the ability to classify data according to the subject and give access according to those divisions.

2. Operational Data Store:

Operational Data Store, which is also called ODS, are nothing but data store required when neither Data warehouse nor OLTP systems support organizations reporting needs. In ODS, Data warehouse is refreshed in real time. Hence, it is widely preferred for routine activities like storing records of the Employees.

3. Data Mart:

A data mart is a subset of the data warehouse. It specially designed for a particular line of business, such as sales, finance, sales or finance. In an independent data mart, data can collect directly from sources.

Thank you



UNIVERSITY
of **KIGALI**