E-Governance

Er. Prahlad Panthi

Data Warehousing and Data Mining in Government

Data Warehousing

A DBMS stores up to date data in the form of tables, uses ER model and the goal is ACID (Atomicity, Consistency, Isolation and Durability) properties. A data warehouse stores huge amount of data which is typically collected from multiple heterogeneous sources like files, databases, etc. It is mainly used for producing statistical results that may help in decision making.
In other words, a data warehouse can be explained as a subject-oriented, integrated, time-variant, non-volatile collection of data, cutting across the enterprise.

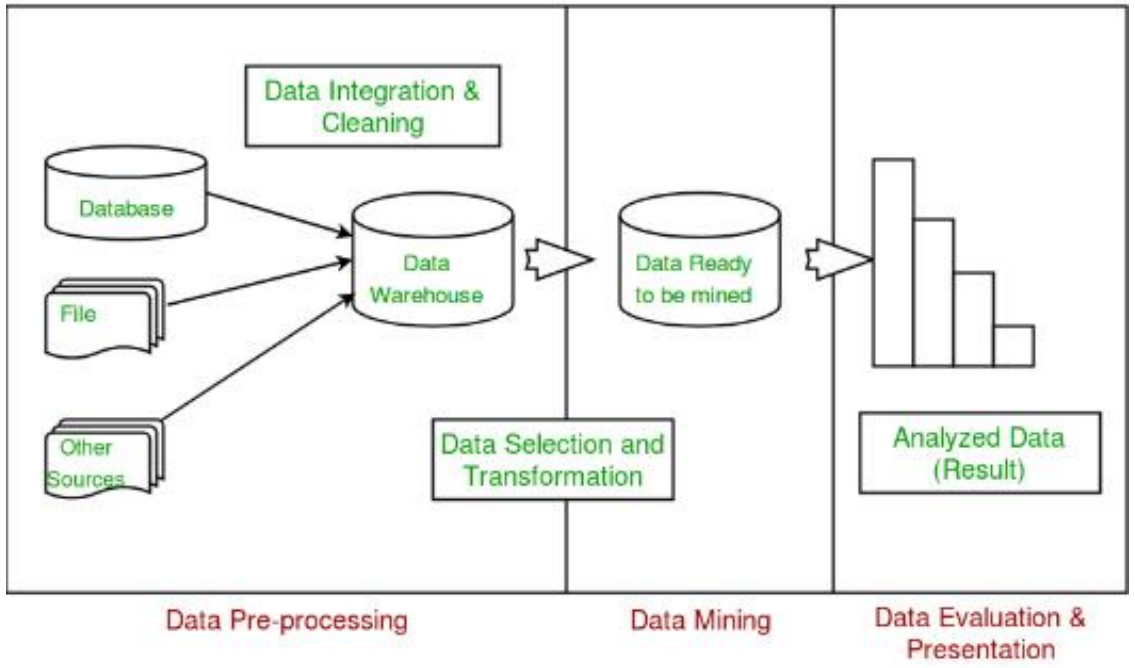
Advantages

- more cost effective decision making
- better enterprise intelligence
- enhanced customer service
- business reengineering

Data Mining

It is the extraction of useful information from data warehouse, databases or the internet, also known as **Knowledge Discovery in Databases (KDD)**. Data mining tools predict future trends and behaviours, allowing businesses to make proactive, knowledge driven decisions. Its applications are:

- financial analysis
- biological analysis
- scientific analysis
- fraud detection
- research analysis, etc.



Advantages

- automated prediction of trends and behaviours
- automated discovery of previously unknown patterns

Steps involved in Knowledge Discovery Process (KDD Process)

1. **Data Cleaning**: Removal of noisy and irrelevant data from collection.
   - Cleaning in case of Missing values.
   - Cleaning noisy data, where noise is a random or variance error.
   - Cleaning with Data discrepancy detection and Data transformation tools.

2. **Data Integration**: Combining heterogeneous data from multiple sources into a single source (Data Warehouse).
   Various tools are used for data integration.

3. **Data Selection**: The process where data relevant to the analysis is decided and retrieved from the data collection.
   It can be done using various techniques.

4. **Data Transformation**: The process of transforming data into appropriate form required by mining procedure.

5. **Data Mining**: Clever techniques that are applied to extract patterns potentially useful.
   - Transforms task relevant data into patterns.
   - Decides purpose of model using classification or characterization.

6. **Pattern Evaluation**: Identifying strictly increasing patterns representing knowledge based on given measures.
   - Find interestingness score of each pattern.
   - Uses summarization and Visualization to make data understandable by user.

7. Knowledge Representation: Technique which utilizes visualization tools to represent data mining results.

- Generate reports.
- Generate tables.
- Generate discriminant rules, classification rules, characterization rules, etc.

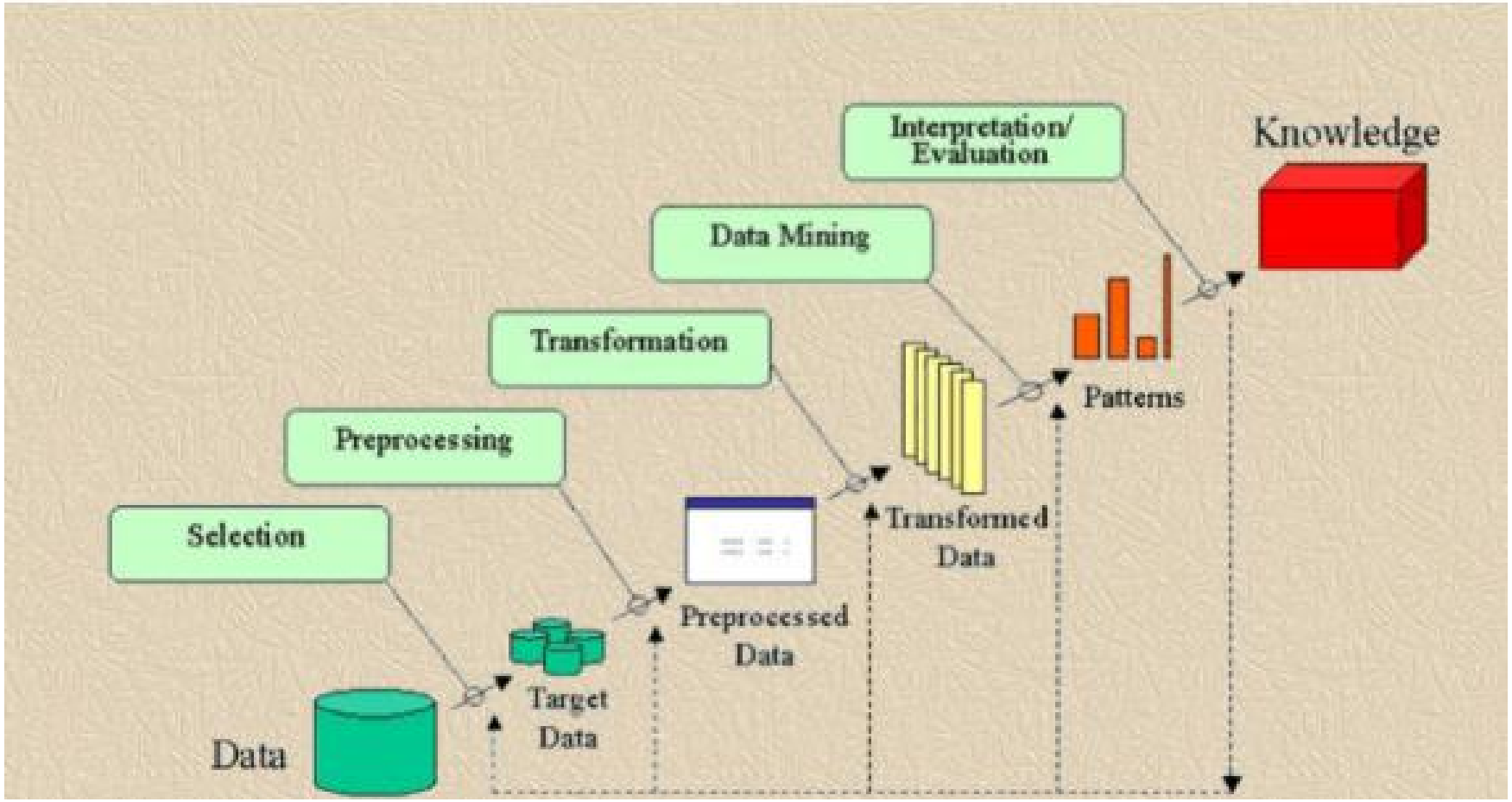


fig: Knowledge Discovery from Data (KDD) Process

National Data Warehouses

A data warehouse can be explained as a subject-oriented, integrated, time-variant, non-volatile collection of data, cutting across the enterprise.

Census Data:
The Central Bureau of Statistics under National Planning Commission under the Government of Nepal keeps track of all the census data of Nepal. In India, the Registrar General and Census Commissioner of India compiles information of all individuals, villages, population groups, etc. Similarly, in USA, United States Census Bureau serves as the leading provider of quality data about its people and economy.

Prices of Essential Commodities:
Monitored by the Department of Food Technology and Quality Control of Ministry of Agriculture and Livestock Department under the Government of Nepal.

Other areas of Data Warehousing and Data Mining

- Agriculture
- Rural Development
- Health
- Planning
- Education
- Commerce and Trade
- Other sectors
    - Tourism
    - Revenue
    - Economic affairs: budget and expenditure data, annual economic survey
    - Audit and Accounts

To Be Continued...