

DATA MINING

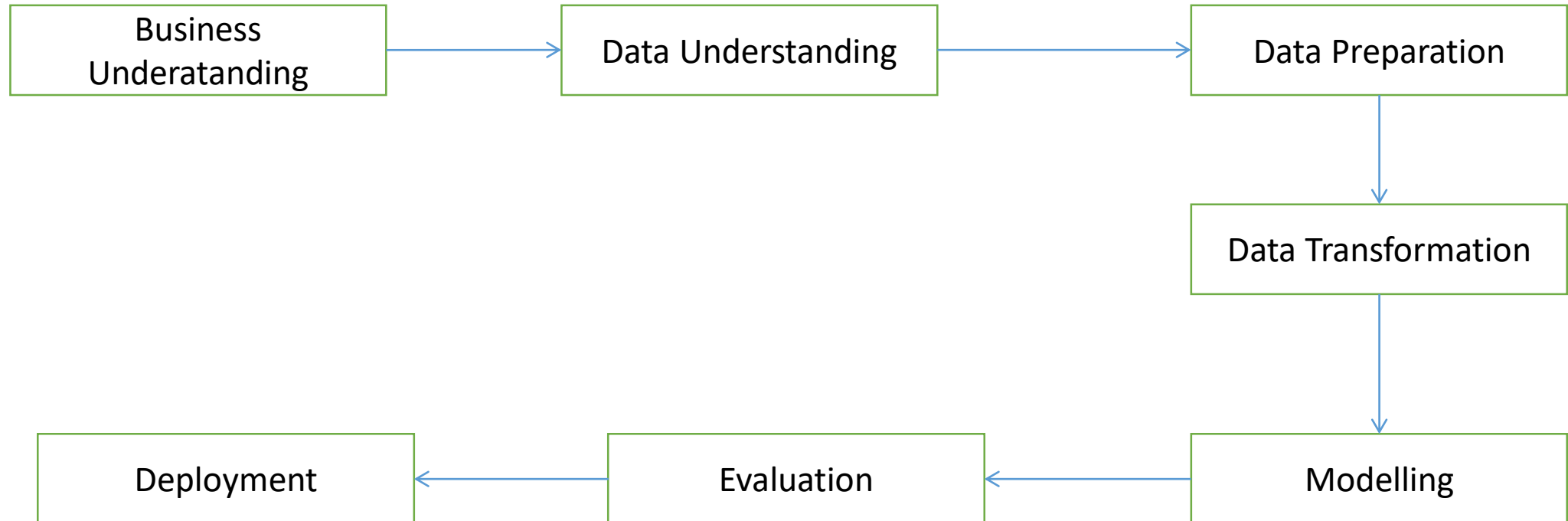
Introduction

- Data mining is the technology of discovering structures and patterns in large data sets.
- It can be distinguish from other analytic disciplines by the size of data sets.
- Data mining is looking for hidden, valid, and potentially useful patterns in huge data sets.

Importance

- Understand what is relevant and then make good use of that information to assess likely outcomes.
- Accelerate the pace of making informed decisions.

Implementation Process



Business Understanding

- Understand business and client objectives
- Take stock of the current data mining scenario
- Define data mining goal

Data Understanding

- Data is collected from multiple data sources available in the organization.
- These data sources may include multiple databases. Issues like object matching and schema integration can arise during Data Integration process.
- Metadata should be used to reduce errors in the data integration process.
- Search for properties of acquired data.
- Based on the results of query, the data quality should be ascertained

Data Preparation

- The data from different sources should be selected, cleaned, transformed, formatted, anonymized, and constructed.
- Data cleaning is a process to clean the data by smoothing noisy data and filling in missing values.

Data Transformation

- Smoothing
- Aggregation
- Generalization
- Normalization
- Attribute Construction

Modelling

- Suitable modeling techniques should be selected for the prepared dataset.
- Create a scenario to test check the quality and validity of the model.
- Run the model on the prepared dataset.

Evaluation

- Results generated by the data mining model should be evaluated.

Deployment

- The knowledge or information discovered during data mining process should be made easy to understand for non-technical stakeholders.

Exercises and Tools

- Model Building
- Pattern Detection

Model Building

- It summarizes the data in a convenient form.
- Choice between models is typically based on the relative size of data.
- Practical data mining often has a preference for combining the results of different models.

Pattern Detection

- It seeks regularities in data set.
- It is interested only in particular small localities of data.
- Pattern matching
- Supervised pattern detection
- pattern discovery

Data Mining Techniques

- Classification
- Clustering
- Regression
- Associative Rules
- Outlier Detection
- Sequential Patterns
- Prediction

Advantages

- Data mining technique helps companies to get knowledge-based information.
- It helps organizations to make the profitable adjustments in operation and production.
- It helps with the decision-making process.
- Facilitates automated prediction of trends and behaviors as well as automated discovery of hidden patterns.
- It can be implemented in new systems as well as existing platforms
- It is the speedy process which makes it easy for the users to analyze huge amount of data in less time.

Disadvantages

- Many data mining analytics software is difficult to operate and requires advance training to work on.
- Different data mining tools work in different manners due to different algorithms employed in their design. Therefore, the selection of correct data mining tool is a very difficult task.
- The data mining techniques are not accurate, and so it can cause serious consequences in certain conditions

Applications

- Communication
- Insurance
- Education
- Manufacturing
- Banking
- Supermarket
- Others

Thank you!