



# **CIS 635 - Knowledge Discovery & Data Mining**

Introduction to Data Mining

# What is Data Mining?

- **Process of extracting and discovering patterns in large datasets**
- Involves methods: ML, Statistics, DBMS
- Interdisciplinary field: CS, Statistics
- Overall goal:
  - Extracting information from dataset
  - Transform into a comprehensive structure for further use
- Data mining is the analysis step of
  - The KDD
  - Aside from raw analysis, it also involves
    - Database and data management aspects
    - Data preprocessing
    - Modeling and inference considerations
    - Evaluation and metrics
    - Post processing of discovered structures and visualizations

[Wikipedia](#)



# What is Data Mining?

- Process of extracting and discovering patterns in large datasets
- **Involves methods: ML, Statistics, DBMS**
- Interdisciplinary field: CS, Statistics
- Overall goal:
  - Extracting information from dataset
  - Transform into a comprehensive structure for further use
- Data mining is the analysis step of
  - The KDD
  - Aside from raw analysis, it also involves
    - Database and data management aspects
    - Data preprocessing
    - Modeling and inference considerations
    - Evaluation and metrics
    - Post processing of discovered structures and visualizations

[Wikipedia](#)



# What is Data Mining?

- Process of extracting and discovering patterns in large datasets
- Involves methods: ML, Statistics, DBMS
- **Interdisciplinary field: CS, Statistics**
- Overall goal:
  - Extracting information from dataset
  - Transform into a comprehensive structure for further use
- Data mining is the analysis step of
  - The KDD
  - Aside from raw analysis, it also involves
    - Database and data management aspects
    - Data preprocessing
    - Modeling and inference considerations
    - Evaluation and metrics
    - Post processing of discovered structures and visualizations

[Wikipedia](#)



# What is Data Mining?

- Process of extracting and discovering patterns in large datasets
- Involves methods: ML, Statistics, DBMS
- Interdisciplinary field: CS, Statistics
- **Overall goal:**
  - **Extracting information from dataset**
  - **Transform into a comprehensive structure for further use**
- Data mining is the analysis step of
  - The KDD
  - Aside from raw analysis, it also involves
    - Database and data management aspects
    - Data preprocessing
    - Modeling and inference considerations
    - Evaluation and metrics
    - Post processing of discovered structures and visualizations

[Wikipedia](#)



# What is Data Mining?

- Process of extracting and discovering patterns in large datasets
- Involves methods: ML, Statistics, DBMS
- Interdisciplinary field: CS, Statistics
- Overall goal:
  - Extracting information from dataset
  - Transform into a comprehensive structure for further use
- **Data mining is the analysis step of**
  - The KDD
  - Aside from raw analysis, it also involves
    - Database and data management aspects
    - Data preprocessing
    - Modeling and inference considerations
    - Evaluation and metrics
    - Post processing of discovered structures and visualizations

[Wikipedia](https://en.wikipedia.org/wiki/Data_mining)





# Knowledge Discovery Process (KDP) Models

## Academic Research Models

- Introduced in mid 1990s
- Several models available
- Suggested steps are similar

Data Mining - A Knowledge Discovery Approach by Cis Pedrycz,  
and Swiniarski

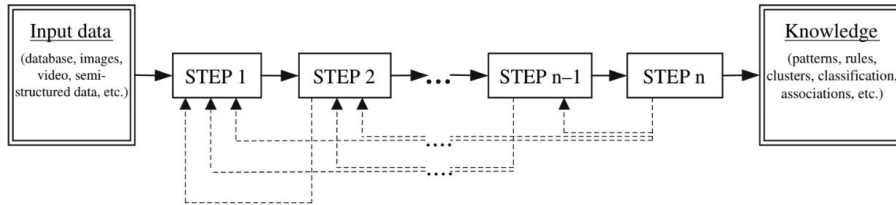
## 9 steps: Fayyad et al KDP model:

- 1) Understanding the application domain
- 2) Creating a target dataset
- 3) Data cleaning and preprocessing
- 4) Data reduction and projection
- 5) Choosing the data mining task
- 6) Choosing the algorithm
- 7) **Data mining**
- 8) Interpreting mined patterns
- 9) Consolidating discovered patterns

# Knowledge Discovery Process (KDP) Models

## Academic Research Models

- Introduced in mid 1990s
- Several models available
- Suggested steps are similar



## 9 steps: Fayyad et al KDP model:

- 1) Understanding the application domain
- 2) Creating a target dataset
- 3) Data cleaning and preprocessing
- 4) Data reduction and projection
- 5) Choosing the data mining task
- 6) Choosing the algorithm
- 7) **Data mining**
- 8) Interpreting mined patterns
- 9) Consolidating discovered patterns

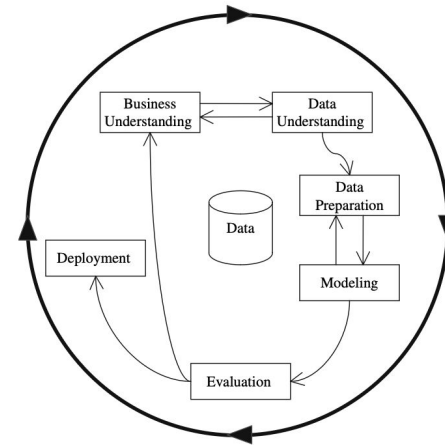


# Knowledge Discovery Process (KDP) Models

## Industrial Models

- Business understanding
- Data Understanding
- Data preparation
- Modeling
- Evaluation
- Deployment

Data Mining - A Knowledge Discovery Approach by Cis Pedrycz, and Swiniarski



The CRISP-DM KD process model (source: <http://www.crisp-dm.org/>).



**QA**