

Introduction to Data Mining Project

Data Mining:
Data Mining Project
with Dr. Qin Lv



Master of Science in Data Science
UNIVERSITY OF COLORADO BOULDER

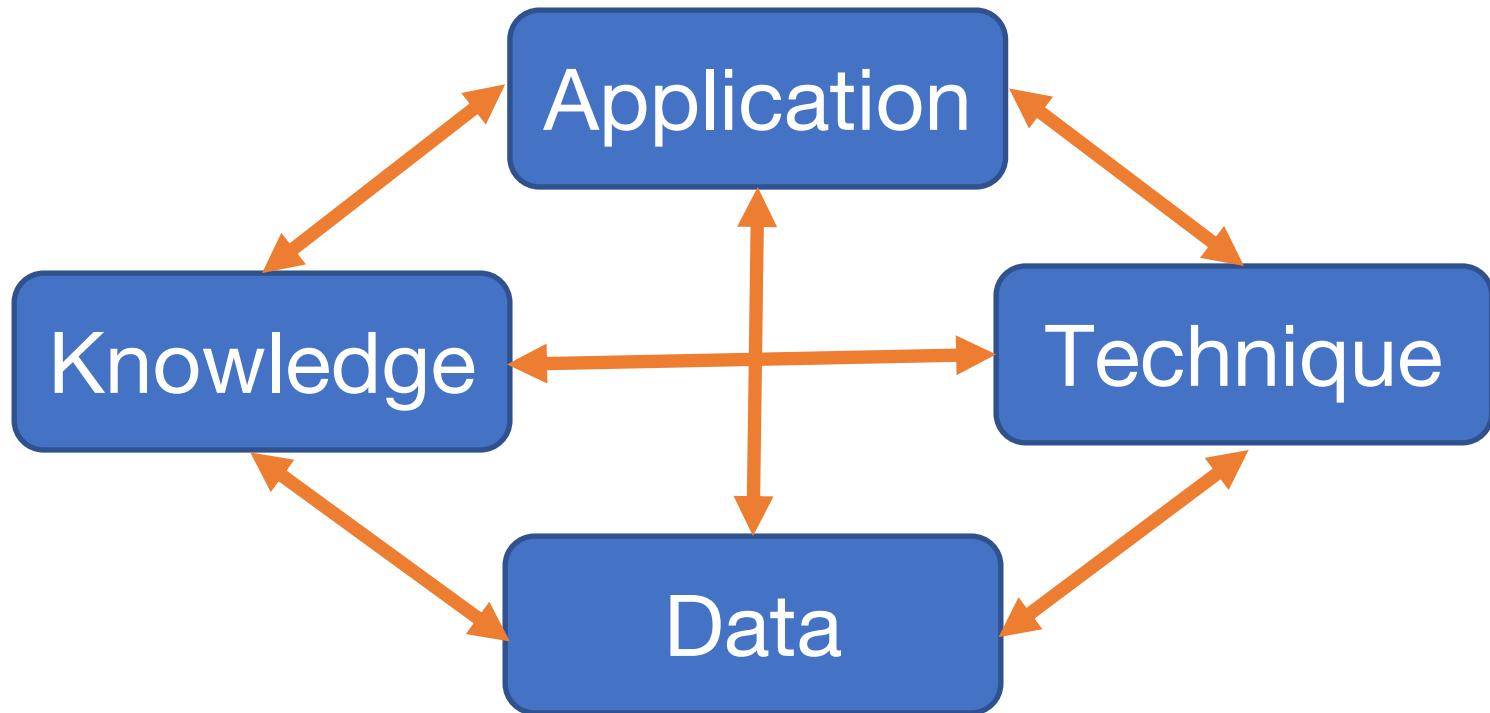


Learning objective: Identify the key components of real-world data mining projects and propose your own project.

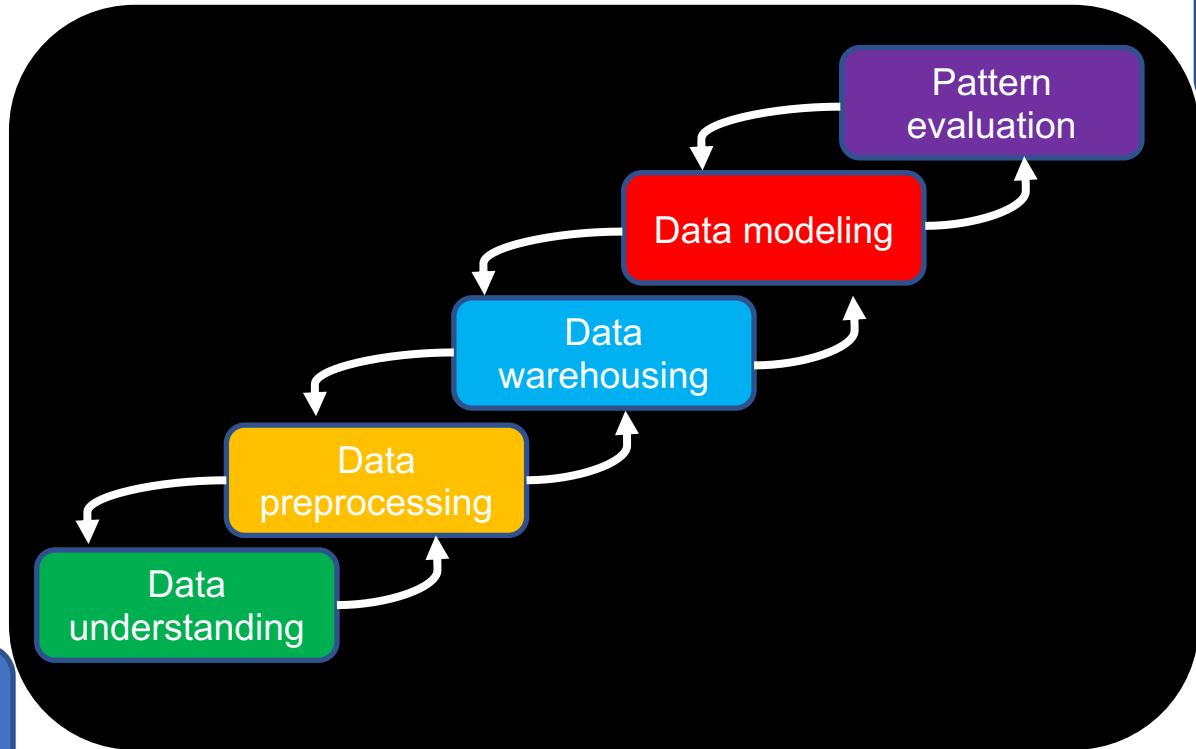
Data Mining Specialization

- Data mining pipeline
- Data mining methods
- Data mining project

Data Mining: Four Views



Data Mining Pipeline



Application

Knowledge

Technique

Data

Technique View

- Frequent pattern analysis
- Classification, prediction
- Clustering
- Anomaly detection
- Trend and evolution analysis

Data Mining Project

➤ Task

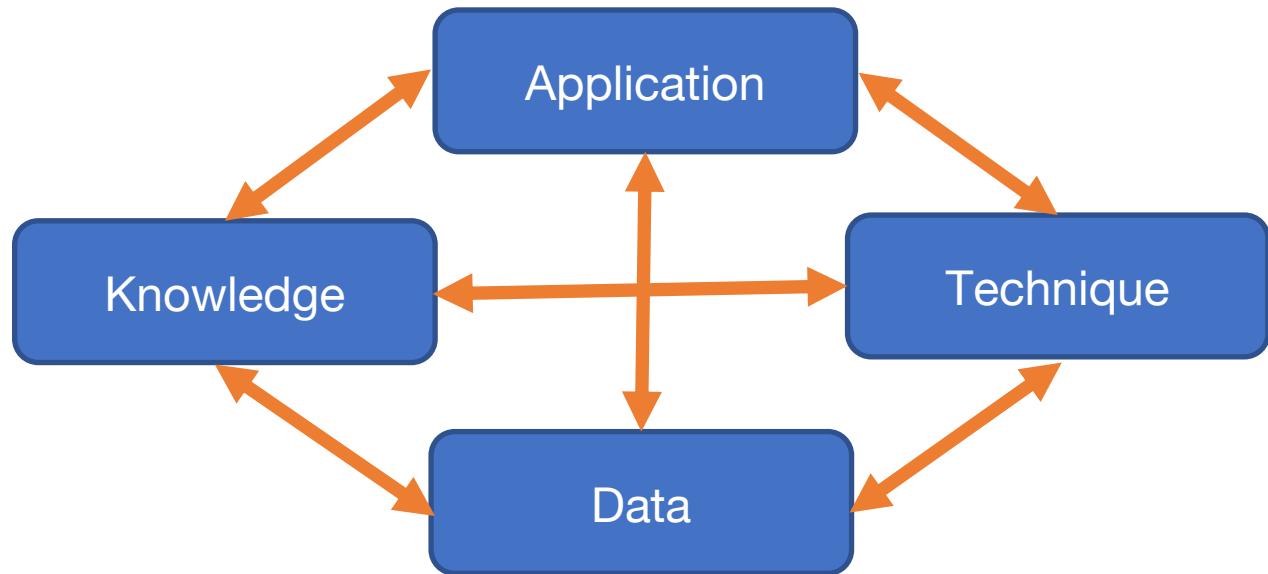
- Specific input => expected output
- E.g., scatter plot, correlation analysis, image classifier

➤ Architect

- big-picture view, analytical thinking
- Identify and propose a data mining project

Where Do We Start?

- What interests you?
- Application
- Knowledge
- Data
- Technique



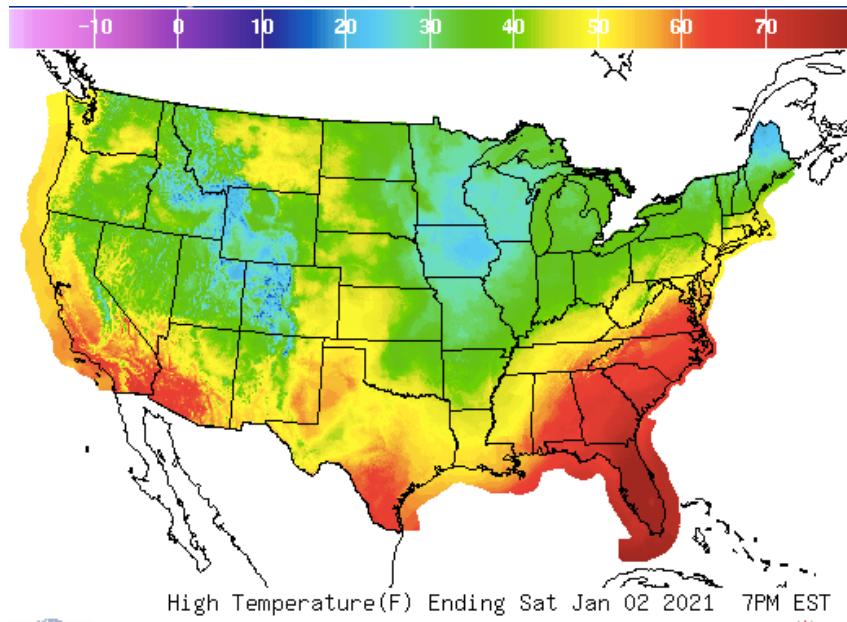
Application

- Wide range of application domains
- Domain knowledge
- What is the challenge?
- What is the value of solving it?



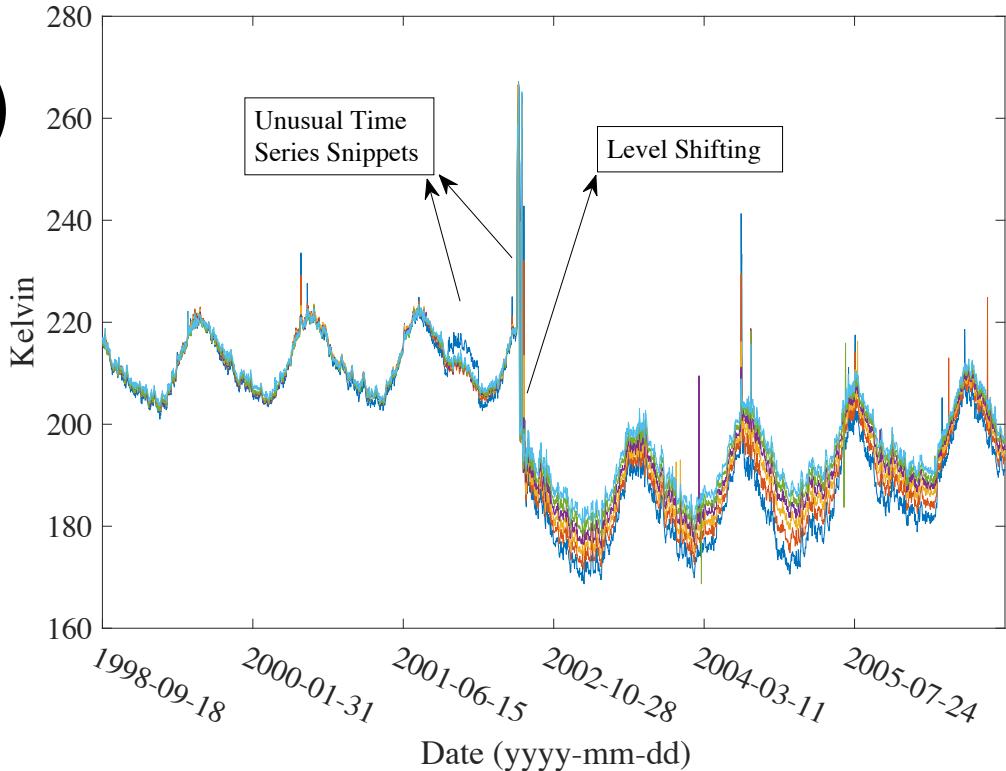
Knowledge

- What are we trying to learn?
- General patterns
- Anomalies
- Specific pattern or anything of interest



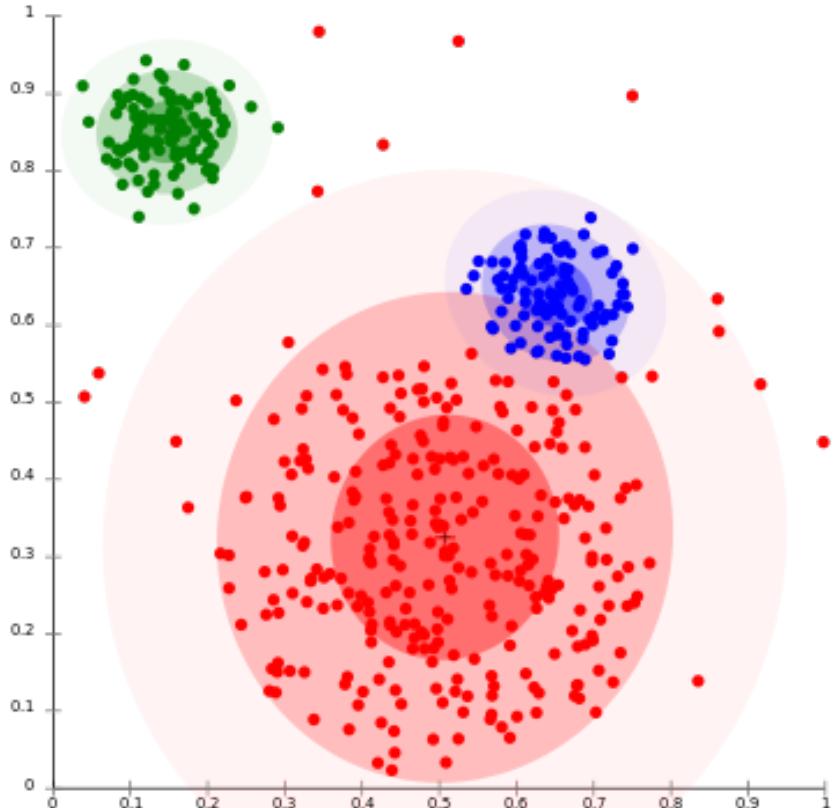
Data

- What data? (4Vs)
 - Volume
 - Variety
 - Velocity
 - Veracity
- Data availability

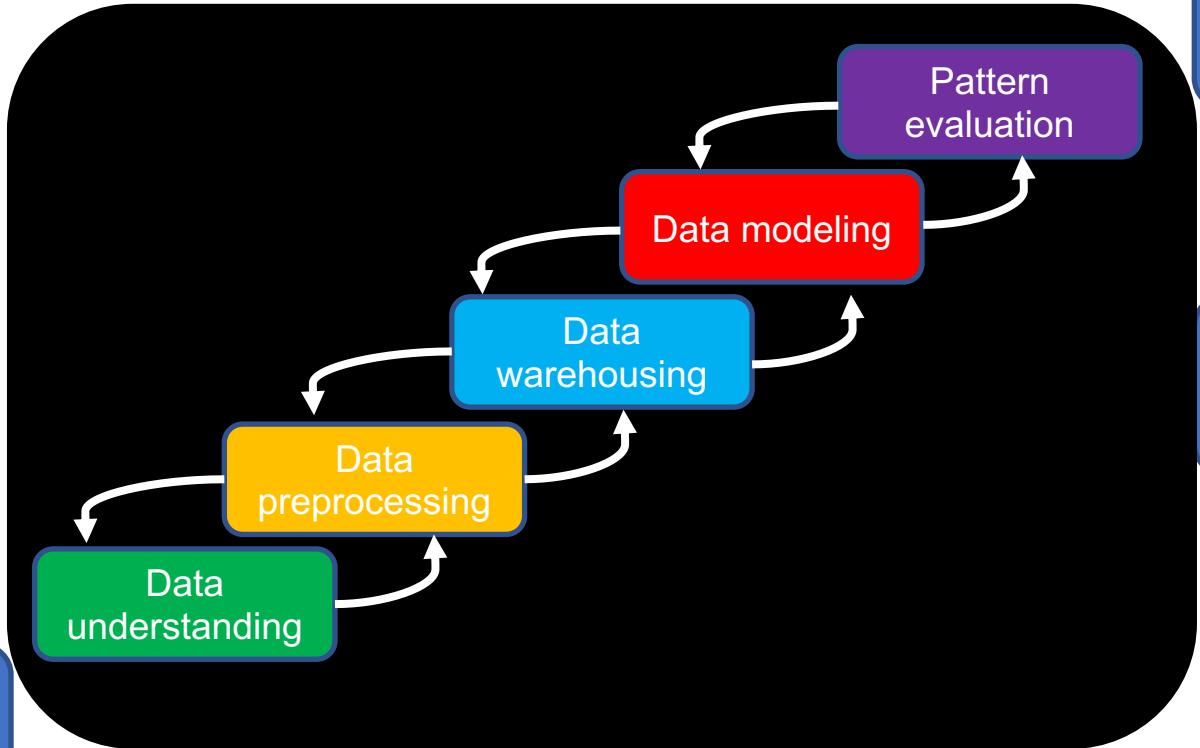


Technique

- What techniques to use?
- Evaluation metrics
- Comparison
- Innovation



Data Mining Pipeline



Application

Knowledge

Technique

Data

Project Scoping

- Timeline
- Prioritized list of tasks
- Expected outcome (evaluation)
- Analytical thinking

Project Brainstorming

- Pick some **initial ideas**
- Check **feasibility**: data, evaluation, ...
- **Discuss** with family, friends, colleagues, ...
- **Iterate**: add/modify/remove ideas