

KDD AND CRISP



Methodologies for Data Analytics



- ❖ Methodologies for planning and developing projects in data analytics.
 1. KDD - Knowledge Discovery in Databases - a methodology from Academia
 2. CRISP-DM - CRoss-Industry Standard Process for Data Mining - a methodology from industry



KDD Methodology

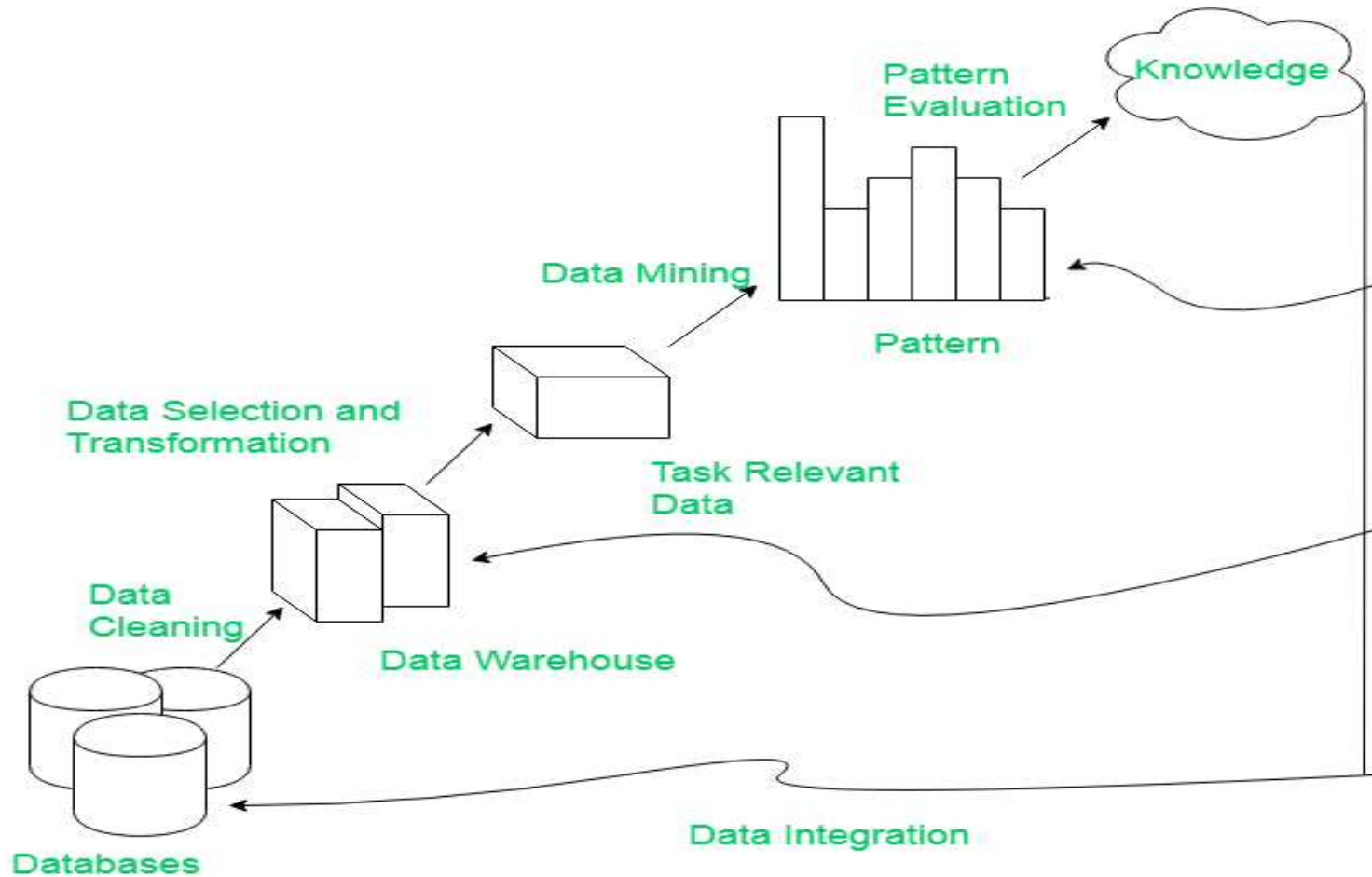


KDD – Knowledge Discovery in Databases :

- 1. Learning the application domain
- 2. Creating a target dataset
- 3. Data cleaning and pre-processing
- 4. Data reduction and projection
- 5. Choosing the data mining function - summarization, clustering, classification and regression
- 6. Choosing the data mining algorithm(s)
- 7. Data mining
- 8. Interpretation
- 9. Using discovered knowledge



KDD PROCESS





CRISP

The CRISP-DM Methodology :

➤ Cross-Industry Standard Process for Data Mining (CRISP-DM) is a six-step method:

1. Business understanding
2. Data understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Deployment



The CRISP-DM Methodology

