

Praktikum Data Mining

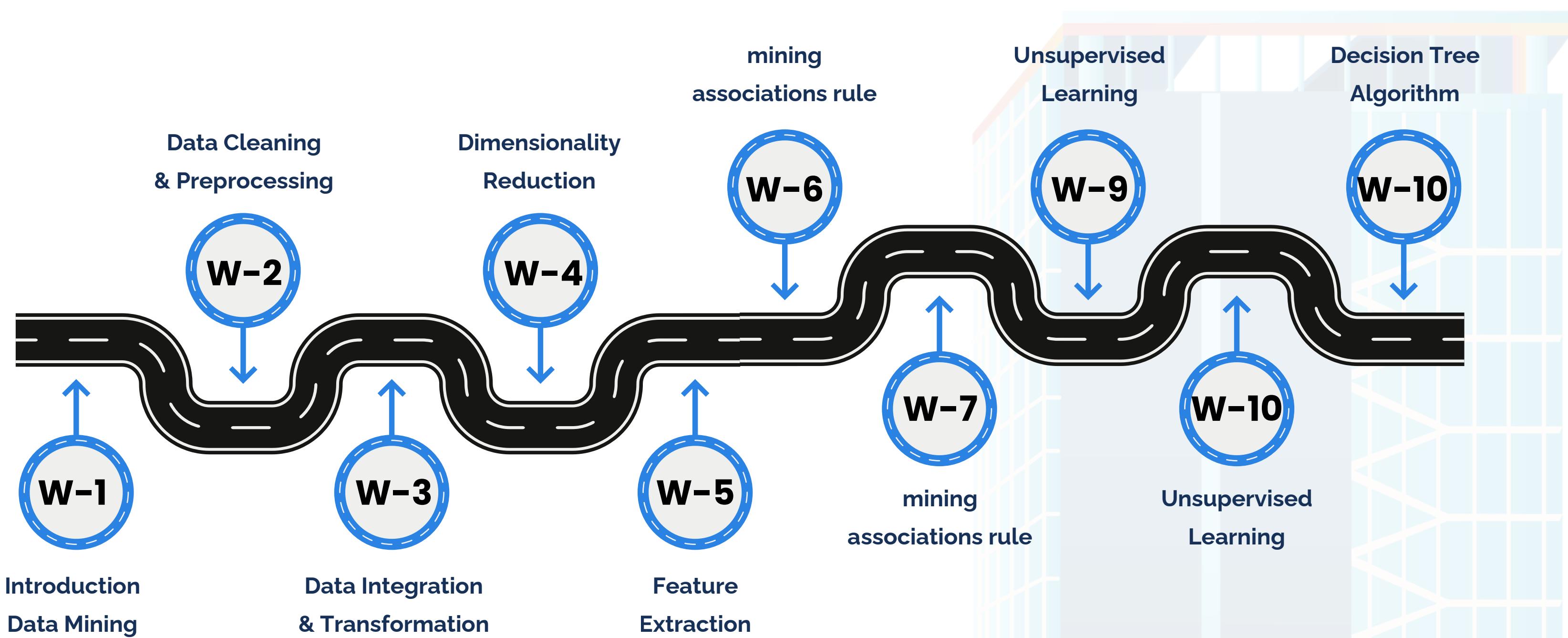
W1 - Introduction Data Mining



Danish Rafie Ekaputra
NIP

Yendra Wijayanto
5052231005

Roadmap Berdasarkan RPS

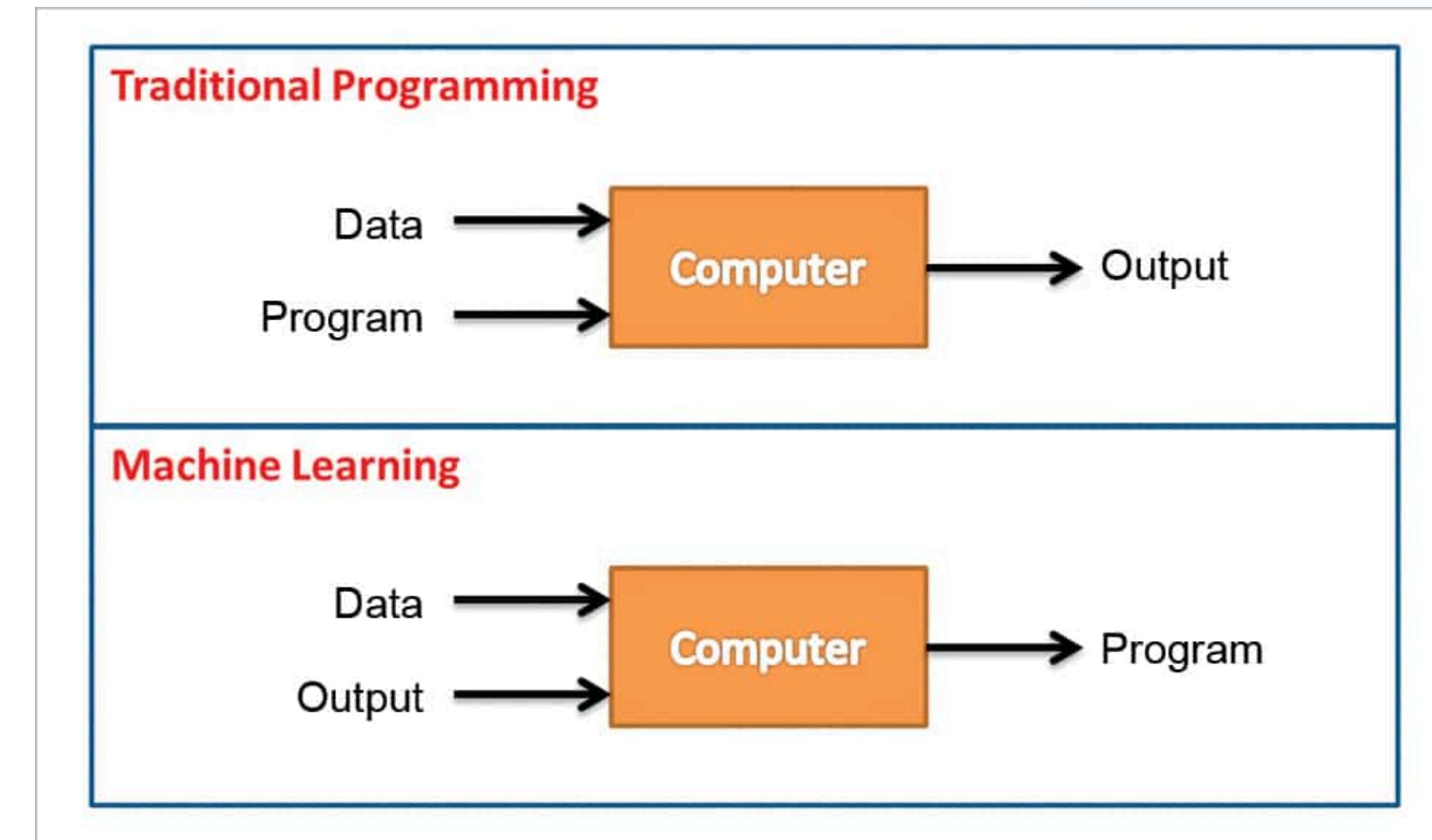


What is Data Mining ?

Data Mining adalah proses menemukan pola, model, atau pengetahuan yang menarik (interest patterns) dari set data yang sangat besar. Sering juga disebut sebagai **Knowledge Discovery from Data (KDD)**.

Tujuan utamanya adalah **untuk menemukan informasi tersembunyi (hidden patterns) yang dapat digunakan untuk pengambilan keputusan.**

Machine Learning



Machine learning adalah cabang dari AI, yaitu sebuah kemampuan ketika komputer dapat belajar dari data tanpa perlu diprogram secara eksplisit.

Step step on Data Mining

1. Data Preparation

- **Data Cleaning:** Membersihkan noise dan data yang tidak konsisten.
- **Data Integration:** Menggabungkan data dari berbagai sumber.
- **Data Transformation:** Mengubah data ke bentuk yang sesuai untuk proses mining.
- **Data Selection:** Mengambil data yang relevan dengan tugas analisis.

2. Data Mining

Langkah inti di mana metode cerdas diterapkan untuk mengekstrak pola.

3. Pattern/Model Evaluation

Mengidentifikasi pola yang benar-benar menarik

4. Knowledge Presentation

Menggunakan teknik visualisasi untuk menyajikan pengetahuan kepada pengguna.

***Sumber:** Han,J., Kamber, M.and J. Pei, Data Mining: Concepts and Techniques. Morgan Kaufmann, 4th. , 2023

Metode Dalam Data Mining

Classification

Membangun model untuk memprediksi label kelas kategorikal

Regression

Memprediksi nilai numerik kontinu, seperti prediksi harga rumah

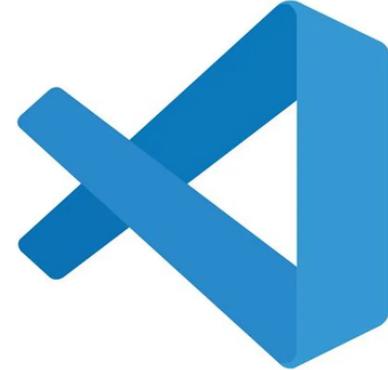
Clustering

Mengelompokkan data tanpa label kelas berdasarkan prinsip kemiripan antar objek (maximising intraclass similarity)

Association Rule Learning

Menemukan item yang sering muncul bersamaan, contohnya analisis keranjang belanja (market basket analysis)milarity

Data Mining Tools



Data Mining Library



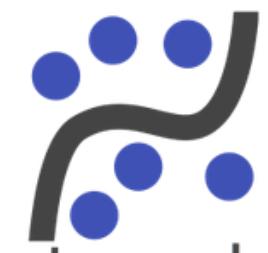
NumPy 

pandas 

matplotlib 

 scikit
learn

 TensorFlow


statsmodels

 Keras

 PyTorch

Virtual Environment



Virtual environment adalah lingkungan Python terisolasi yang digunakan untuk menjalankan suatu proyek dengan versi library dan dependensi yang spesifik, tanpa mempengaruhi Python atau proyek lain di sistem yang sama.

Virtual environment memungkinkan setiap proyek memiliki paket dan versinya sendiri.

Mengapa perlu virtual environment ?

- ketika kesalahan terjadi di environment terpisah, maka dengan mudah kita dapat menghapus environment tersebut dan tidak akan ada yang tersisa
- environment mudah dibagikan sehingga dapat membantu kerjasama untuk mengerjakan suatu projek dalam sebuah tim

Proyek	Library yang Dibutuhkan
Proyek A	numpy 1.20
Proyek B	numpy 1.26

Move To Notebook >>



Terima Kasih

Praktikum Data Mining Week 1

