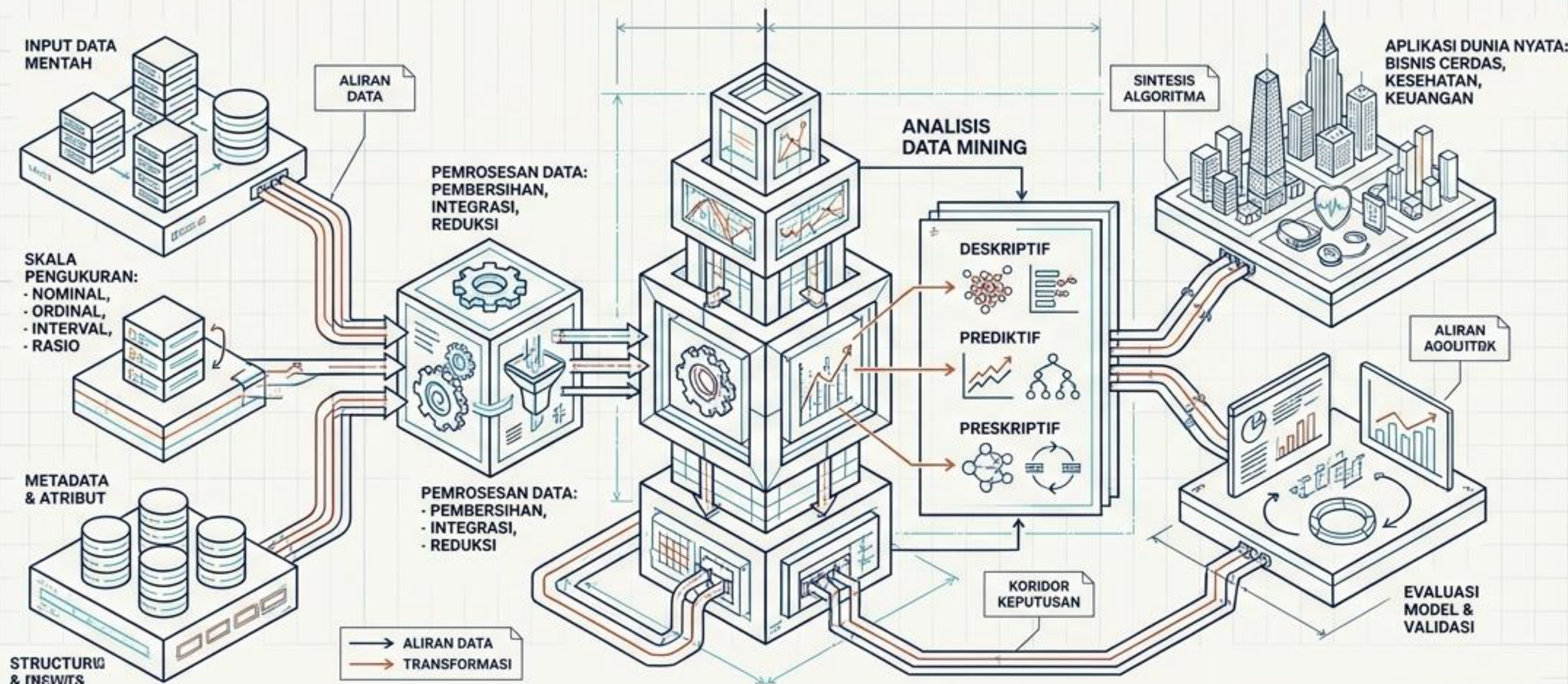


# Arsitektur Pembelajaran Mesin Dimulai dari Pemahaman Fondasi Data

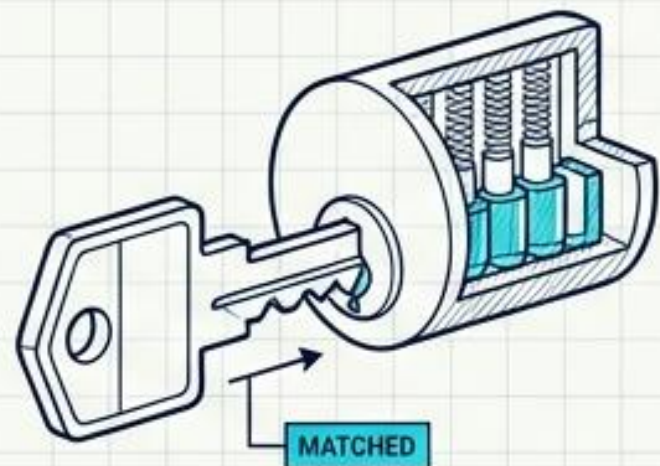
Sebuah cetak biru komprehensif yang membedah anatomi Data Mining: dari struktur skala pengukuran hingga aplikasi algoritma di dunia nyata.



# Empat Kuadran Utama Membentuk Lanskap Pembelajaran Mesin

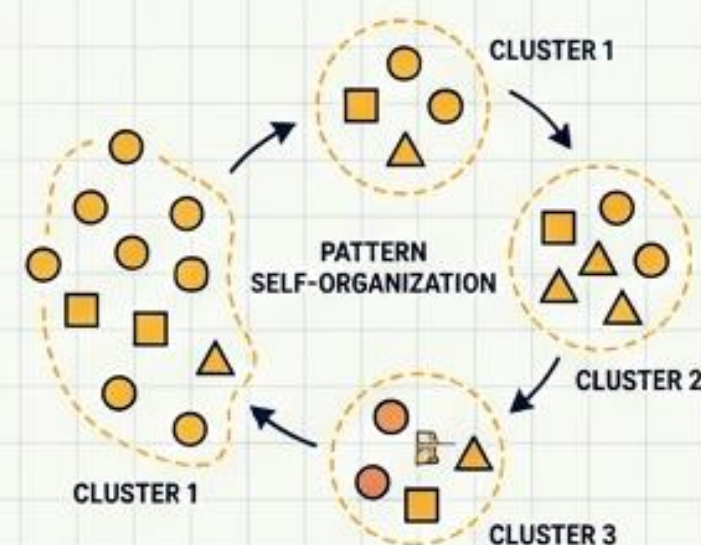
## Supervised Learning

Belajar menggunakan dataset yang memiliki target, label, atau kelas yang jelas.



## Unsupervised Learning

Mencari pola tersembunyi dari seluruh variabel tanpa adanya label target.



## Semi-Supervised Learning

Bekerja dengan skala kecil data berlabel untuk memandu pemrosesan data berskala besar yang tidak berlabel.

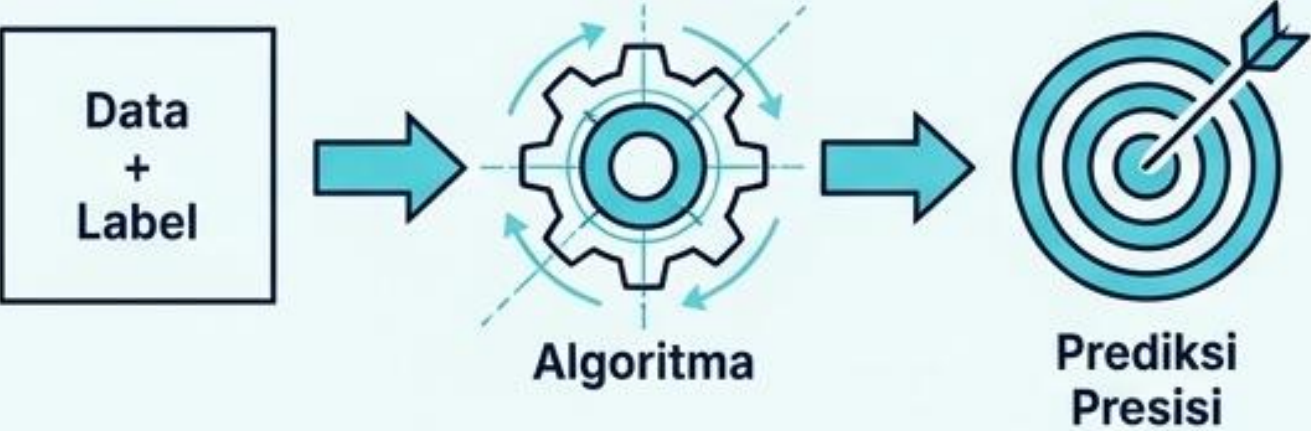
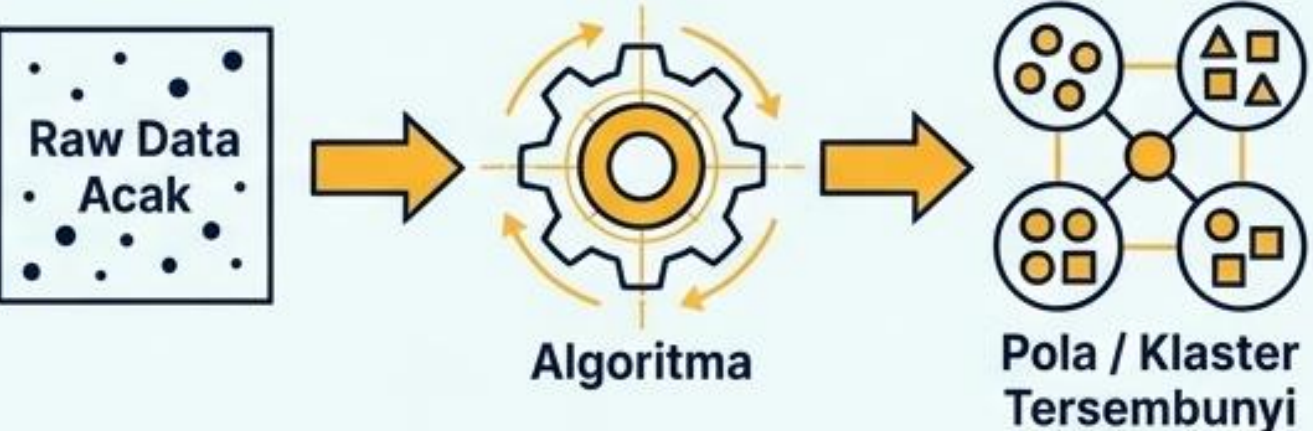


## Reinforcement Learning

Menemukan aksi optimal melalui uji coba berulang (*trial and error*) berdasarkan respons lingkungan.

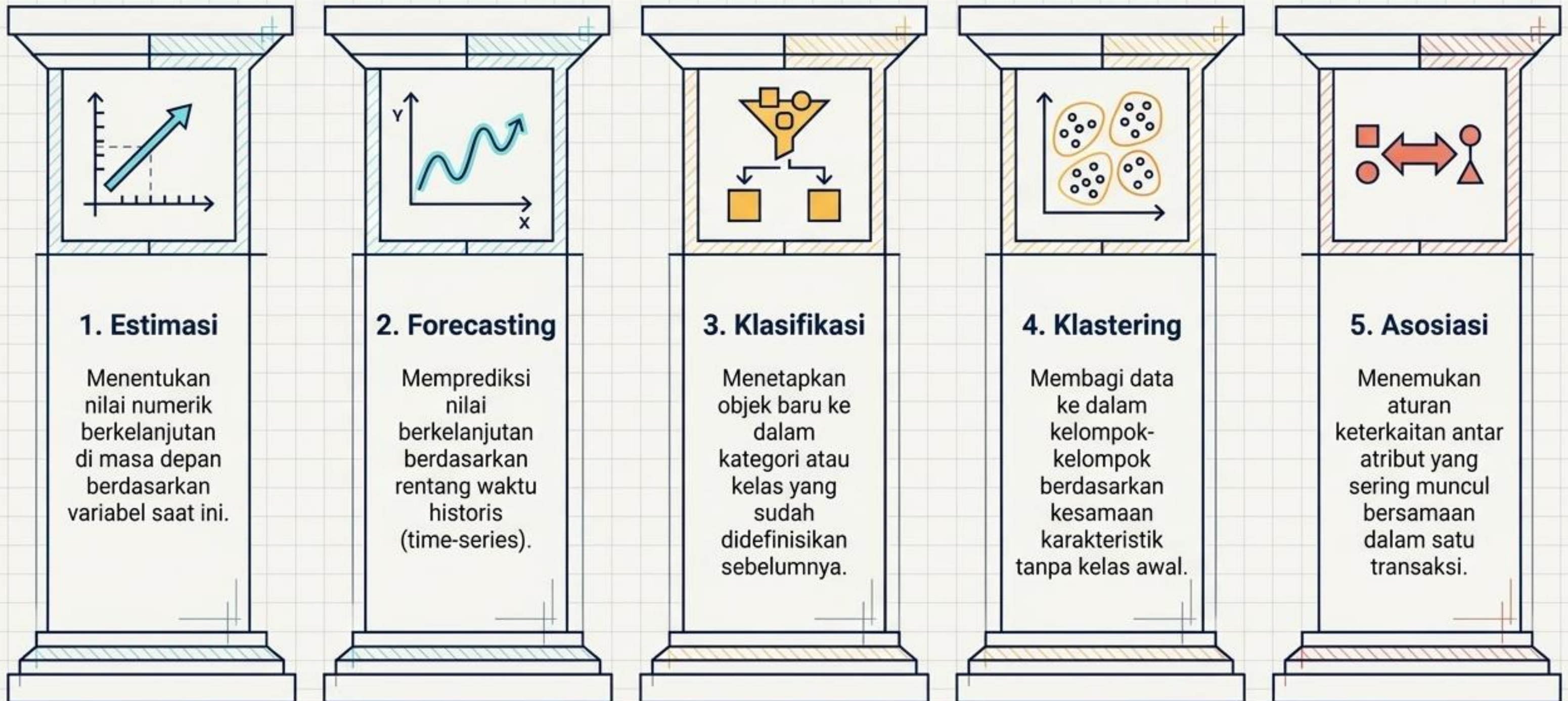


# Supervised dan Unsupervised Learning Memiliki Pendekatan yang Bertolak Belakang

Supervised Learning	Unsupervised Learning
<p><b>1</b> Karakteristik Utama</p> <p>Menggunakan data historis berlabel untuk melatih algoritma.</p>	<p><b>1</b> Karakteristik Utama</p> <p>Mengeksplorasi data tanpa label untuk menemukan struktur intrinsik.</p>
<p><b>2</b> Tujuan Akhir</p> <p><b>Memprediksi output (klasifikasi/estimasi)</b> untuk data baru di masa depan.</p>	<p><b>2</b> Tujuan Akhir</p> <p>Mengidentifikasi pengelompokan (<i>clustering</i>) atau aturan asosiasi.</p>
<p><b>3</b> Ilustrasi Konseptual</p> 	<p><b>3</b> Ilustrasi Konseptual</p> 

# Lima Pilar Fungsional Menjadi Mesin Penggerak Data Mining

Setiap algoritma di dalam ekosistem Data Mining dieksekusi untuk memenuhi satu dari lima peran utama berikut:



# Fungsi Prediktif Menerjemahkan Variabel Menjadi Proyeksi Angka

## Estimasi (Regresi Linier)

Menghitung output pasti dari variabel prediktor yang ada.

**Studi Kasus:** Memprediksi Waktu Tempuh Pengiriman Pizza (T).

Pesanan (P)

Lampu Lalu Lintas (TL)

Jarak (J)

Waktu Tempuh (T)

$$T = 0.48(P) + 0.23(TL) + 0.5(J)$$

## Forecasting (Neural Network)

Memproyeksikan data rentet waktu (time-series).

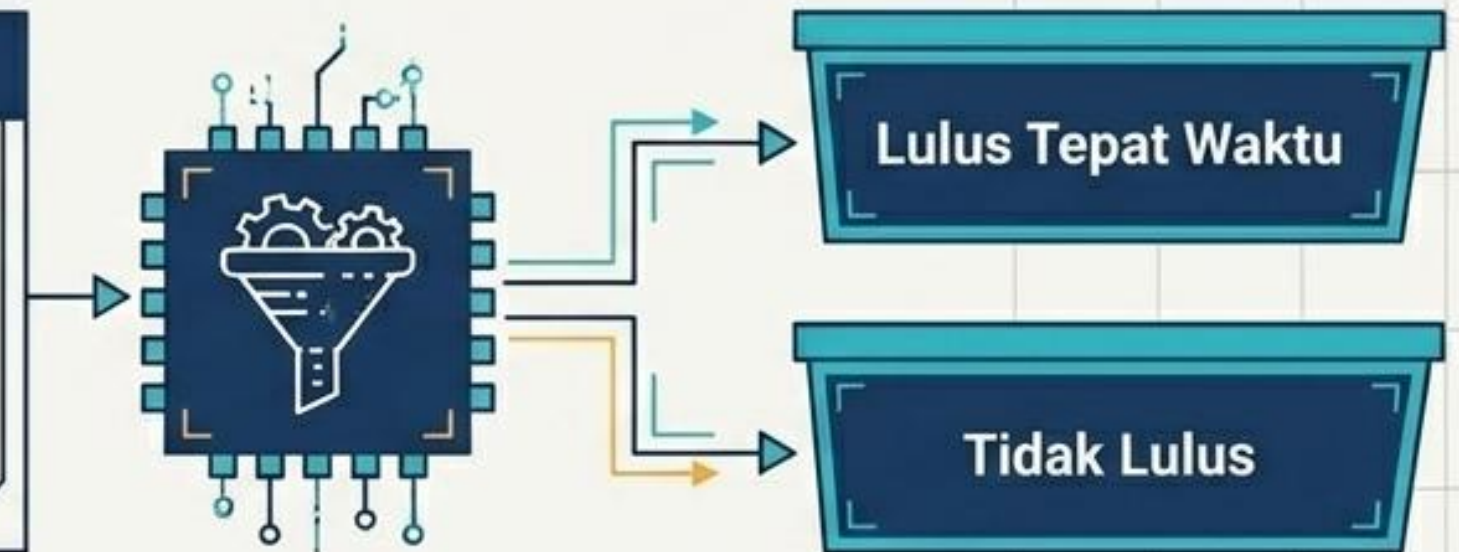
**Studi Kasus:** Memprediksi pergerakan dan fluktuasi harga saham di bursa berdasarkan tren historis.



# Pengelompokan Data Membutuhkan Pendekatan Berbasis Label atau Pola Ekstraktif

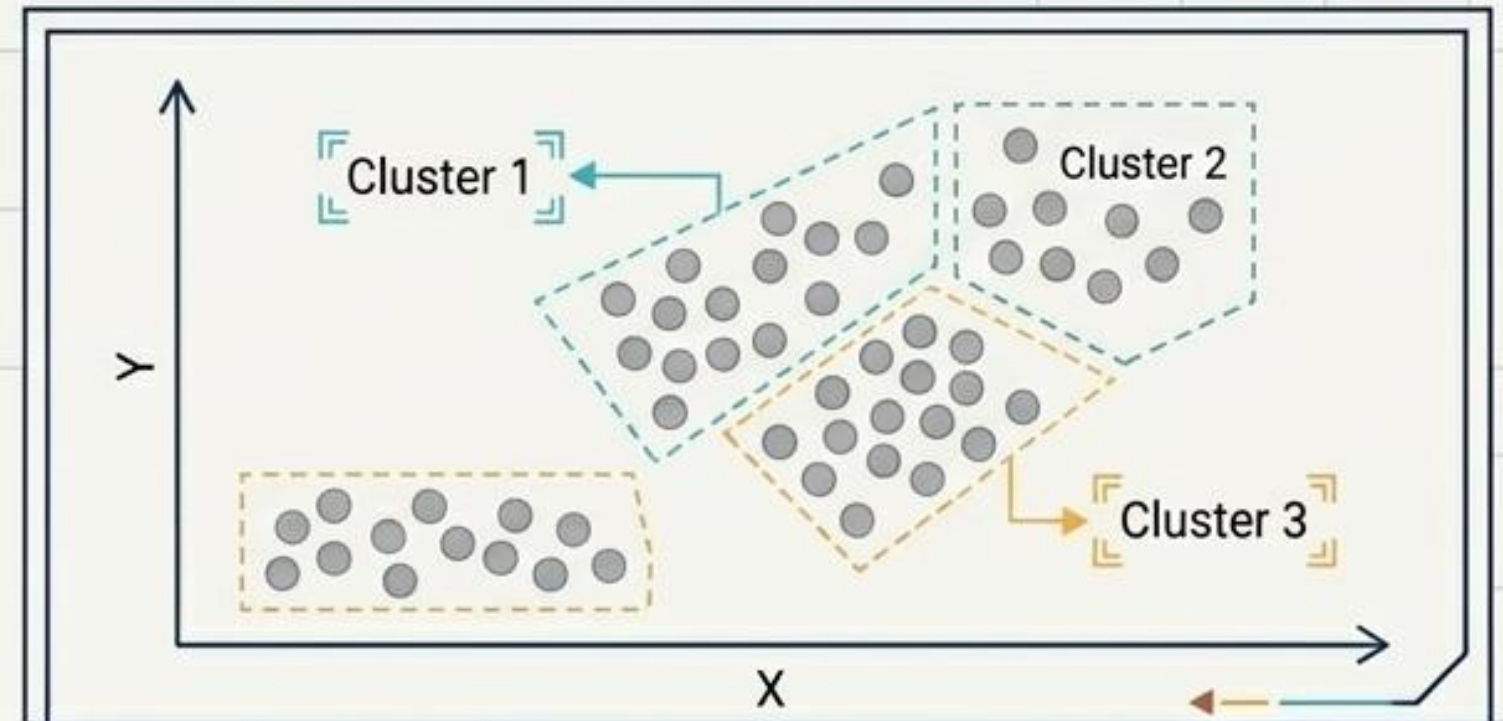
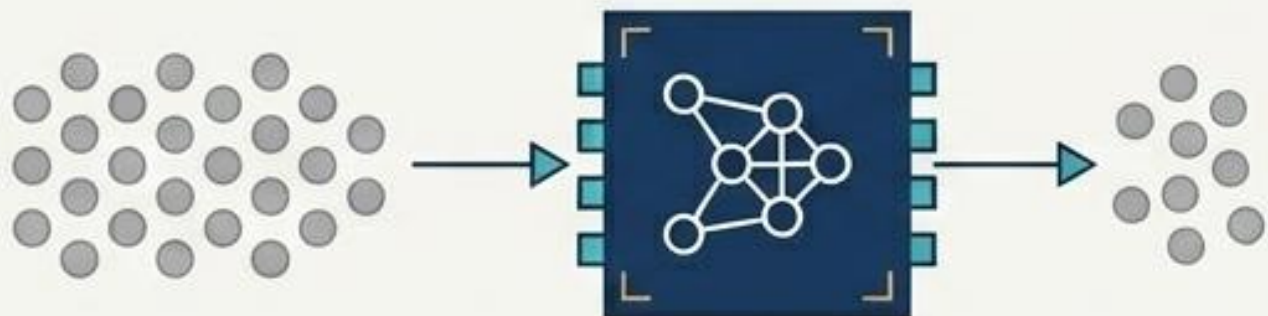
## Klasifikasi (Supervised - C4.5)

Atribut kelas sudah diketahui. Mesin mempelajari profil untuk memilah probabilitas kelulusan mahasiswa berdasarkan rekam jejak akademik.



## Klustering (Unsupervised - K-Means)

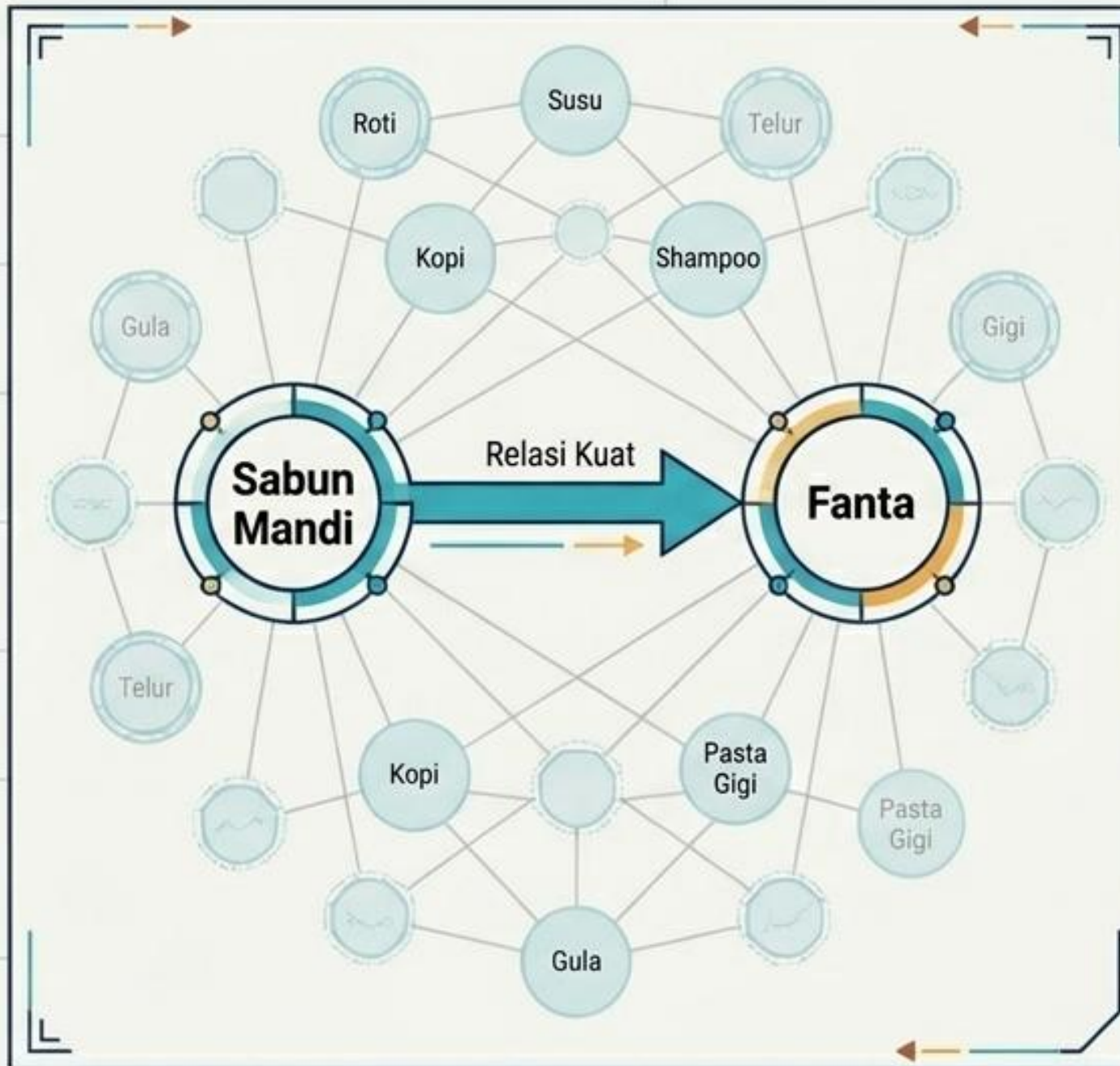
Kelas tidak diketahui. Mesin mendeteksi kedekatan atribut secara mandiri untuk memisahkan spesies bunga Iris berdasarkan struktur fisiknya.



# Aturan Asosiasi Menemukan Relasi Tersembunyi di Balik Transaksi Acak

## Mekanisme FP-Growth / A Priori

Menemukan **atribut** yang **muncul bersamaan** dalam kumpulan data bervolume tinggi.



## Dekonstruksi Kasus Supermarket ABC

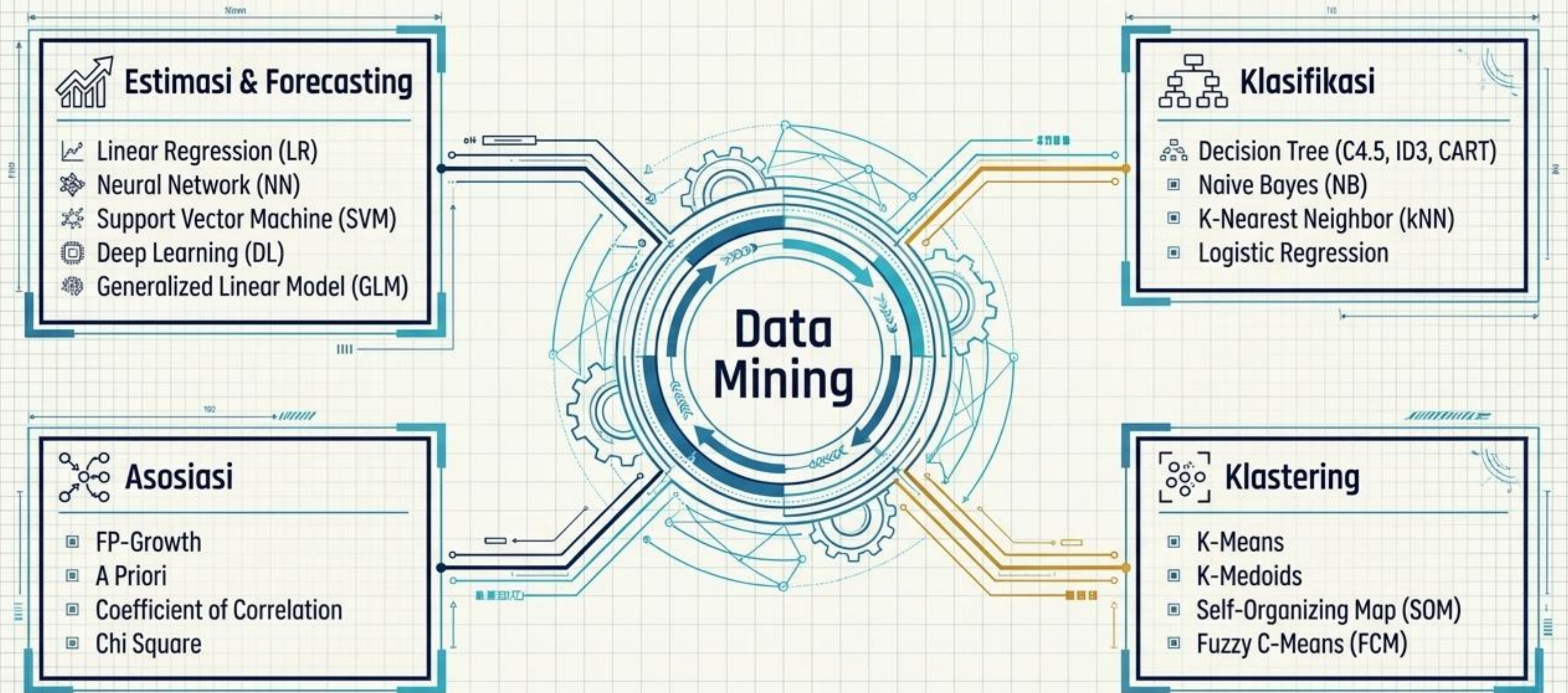
**Data Input:** Dari 1000 pelanggan Kamis malam, 200 membeli Sabun Mandi. Dari 200 orang tersebut, 50 juga membeli Fanta.

**Ekstraksi Aturan:** Jika membeli Sabun Mandi, maka probabilitas membeli Fanta meningkat.

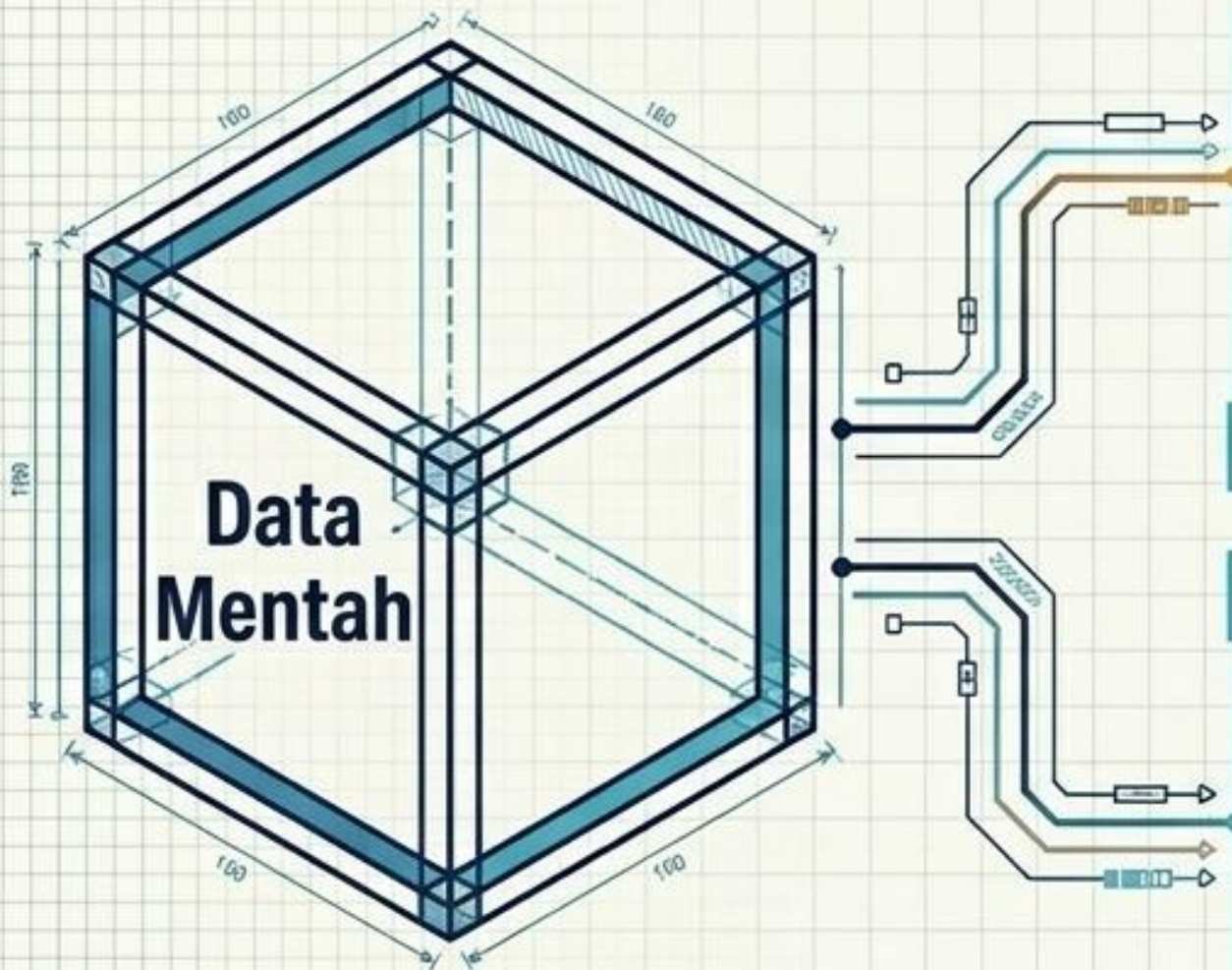
Nilai Support: **20%** (200/1000)  
Nilai Confidence: **25%** (50/200)

# Setiap Peran Fungsional Membutuhkan Instrumen Algoritma yang Spesifik

Pemilihan algoritma sepenuhnya bergantung pada kualitas dan jenis data yang diumpankan.



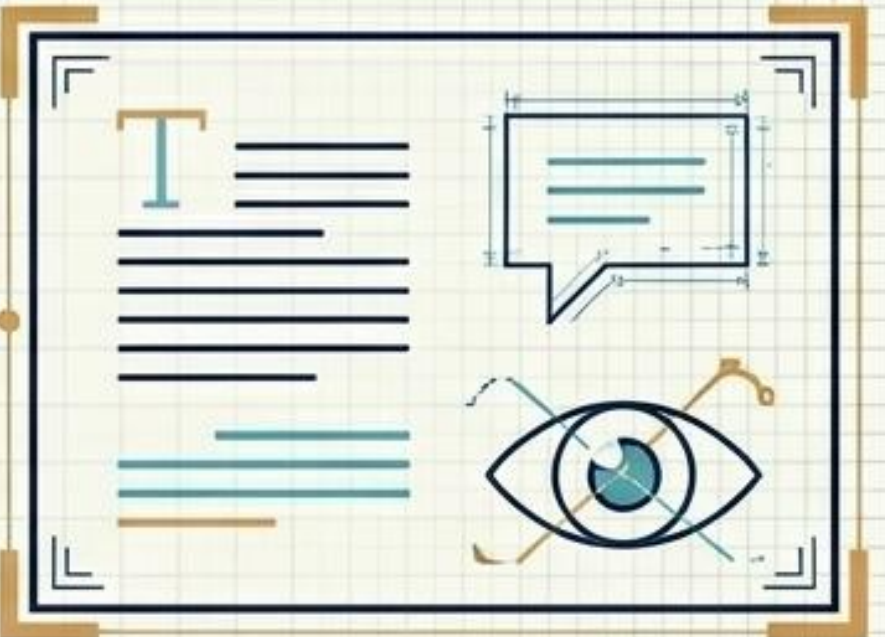
# Kualitas Output Pembelajaran Mesin Bergantung pada Anatomi Jenis Data



## Data Kualitatif

Disajikan dalam bentuk kata, kalimat, gambar, atau makna.

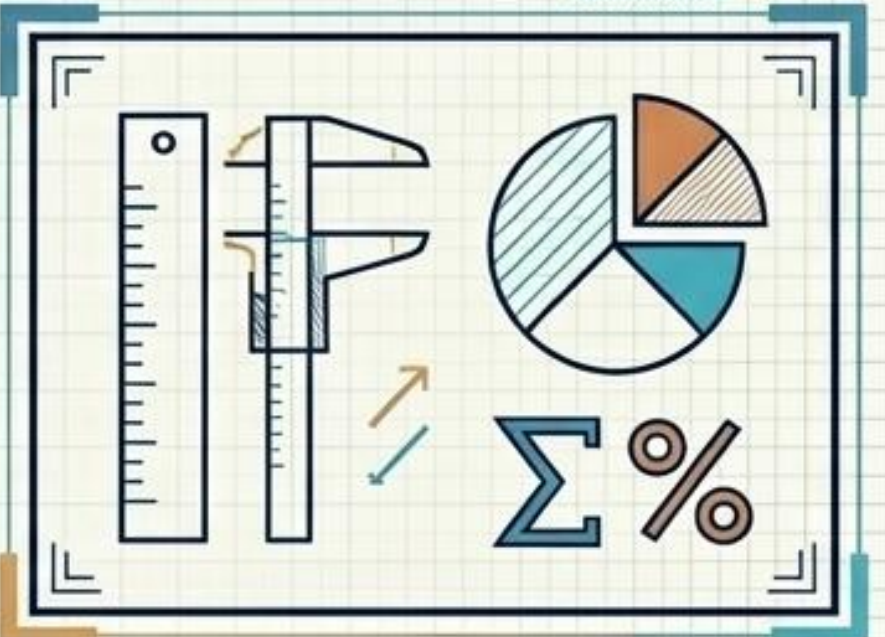
- ▣ **Fungsi:** Menangkap persepsi, anggapan, atau kategorisasi.
- ▣ **Contoh:** Persepsi konsumen terhadap kemasan produk; klasifikasi sentimen pasar.



## Data Kuantitatif

Dipaparkan mutlak dalam bentuk angka-angka terukur.

- ▣ **Fungsi:** Memungkinkan kalkulasi algoritma yang presisi dan pemodelan matematis.
- ▣ **Contoh:** Jumlah pembeli harian; fluktuasi nilai tukar Rupiah; dimensi fisik produk.



# Eskalasi Skala Pengukuran Menentukan Batas Operasi Matematis

## 1. Nominal (Label)

Hanya menggolongkan kategori tanpa tingkatan. (Contoh: Jenis kelamin, Kode Pos).

## 2. Ordinal (Peringkat)

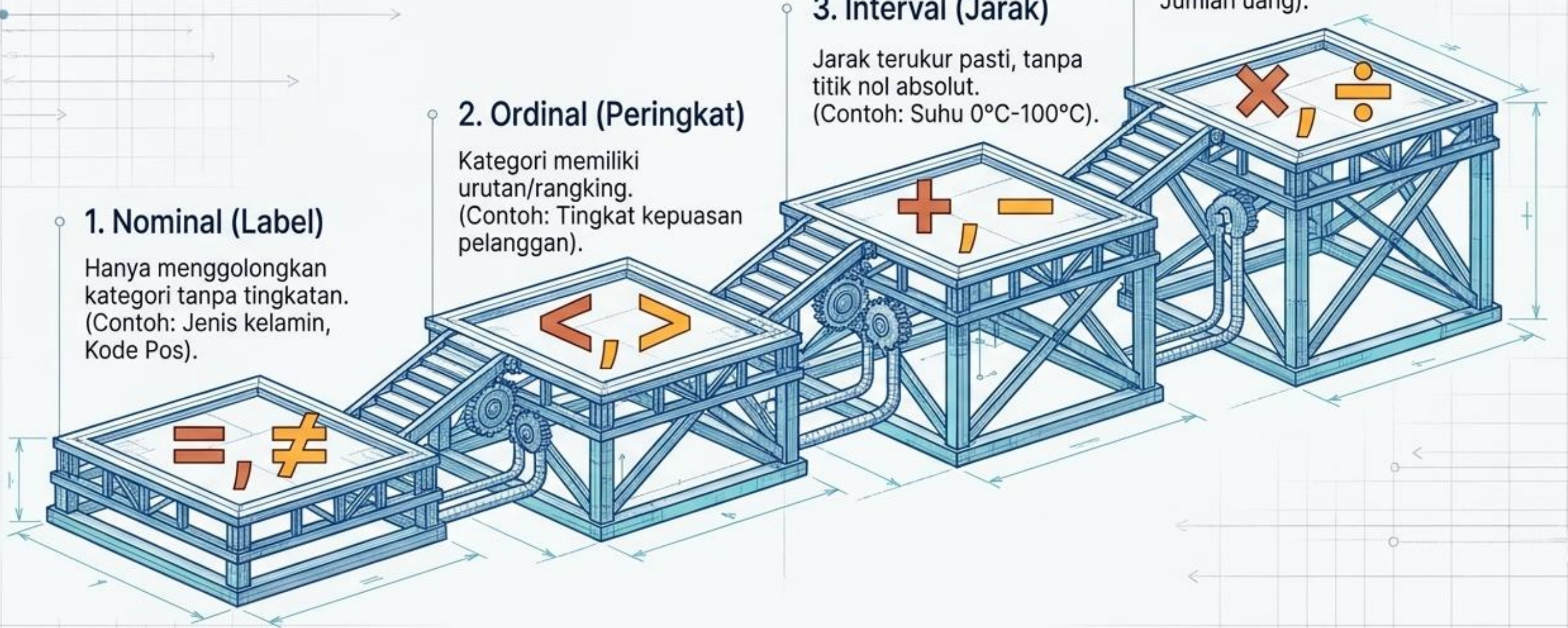
Kategori memiliki urutan/rangking. (Contoh: Tingkat kepuasan pelanggan).

## 3. Interval (Jarak)

Jarak terukur pasti, tanpa titik nol absolut. (Contoh: Suhu  $0^{\circ}\text{C}$ - $100^{\circ}\text{C}$ ).

## 4. Rasio (Mutlak)

Pengukuran absolut dengan titik nol mutlak. (Contoh: Berat badan, Jumlah uang).



# Dimensi Ekstraksi Menentukan Konteks dan Validitas Analisis

## Berdasarkan Cara Memperoleh

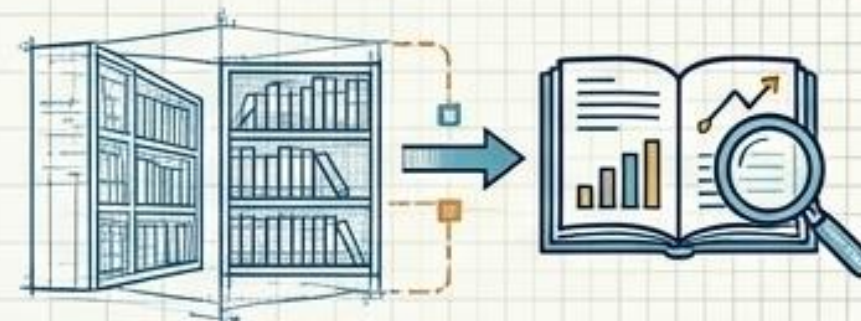
### Primer

Diambil langsung dari objek (misal: Wawancara penonton bioskop).



### Sekunder

Data yang sudah dikumpulkan pihak lain (misal: Statistik riset majalah).



## Berdasarkan Sumber Lokasi

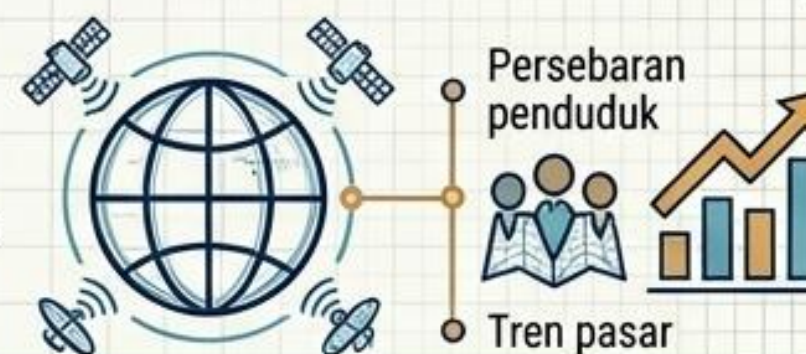
### Internal

Menggambarkan kondisi dalam organisasi (Keuangan, Pegawai).



### Eksternal

Menggambarkan kondisi luar (Persebaran penduduk, Tren pasar).



## Berdasarkan Waktu & Sifat

### Cross Section vs Time Series

Titik waktu spesifik (Laporan Q4)

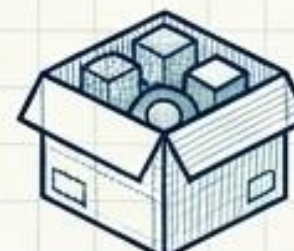
Berkelanjutan historis (Nilai tukar Dollar-Euro 2004-2006).



### Diskrit vs Kontinyu

Bilangan bulat mutlak

Nilai pada interval tertentu (kira-kira 850 ton).



# Supervised Learning Mengotomatisasi Keputusan di Sektor Transportasi dan Medis

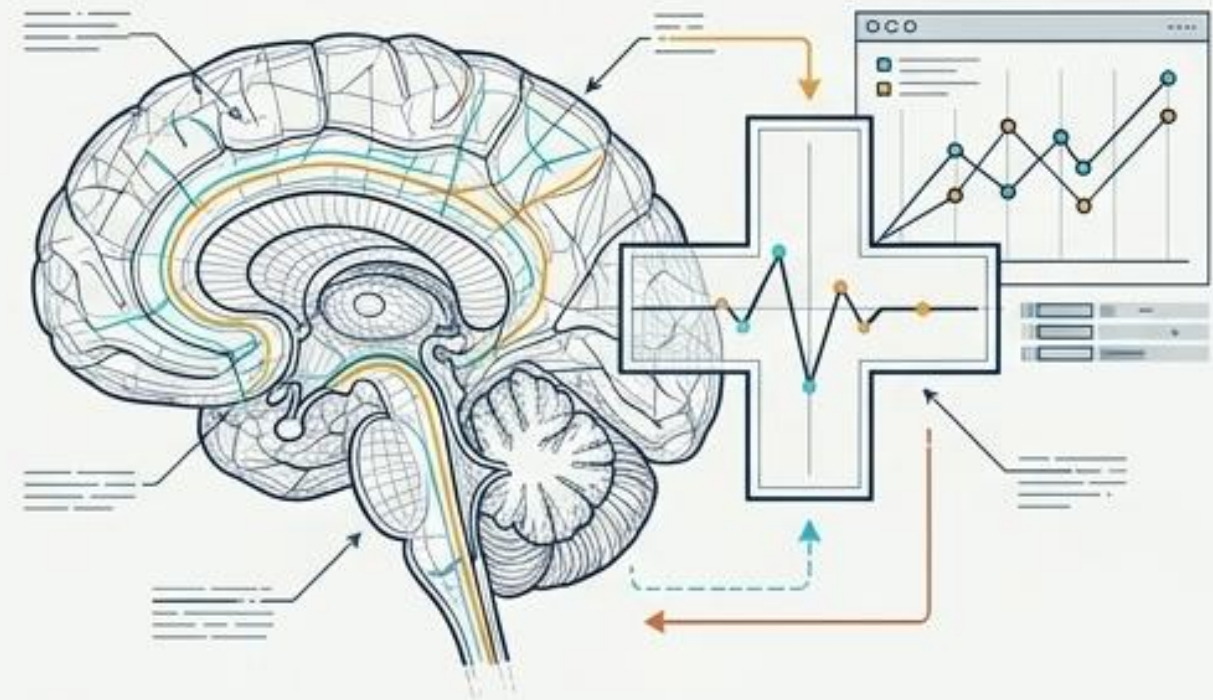
## Optimalisasi Infrastruktur Transportasi



**Algoritma:** Convolutional Neural Network (CNN) dipadukan dengan Support Vector Machine (SVM).

**Aplikasi:** CNN mengekstraksi fitur visual, SVM mengklasifikasi plat nomor kendaraan. Solusi ini diterapkan untuk manajemen parkir pintar dan sistem pembayaran tol otomatis.

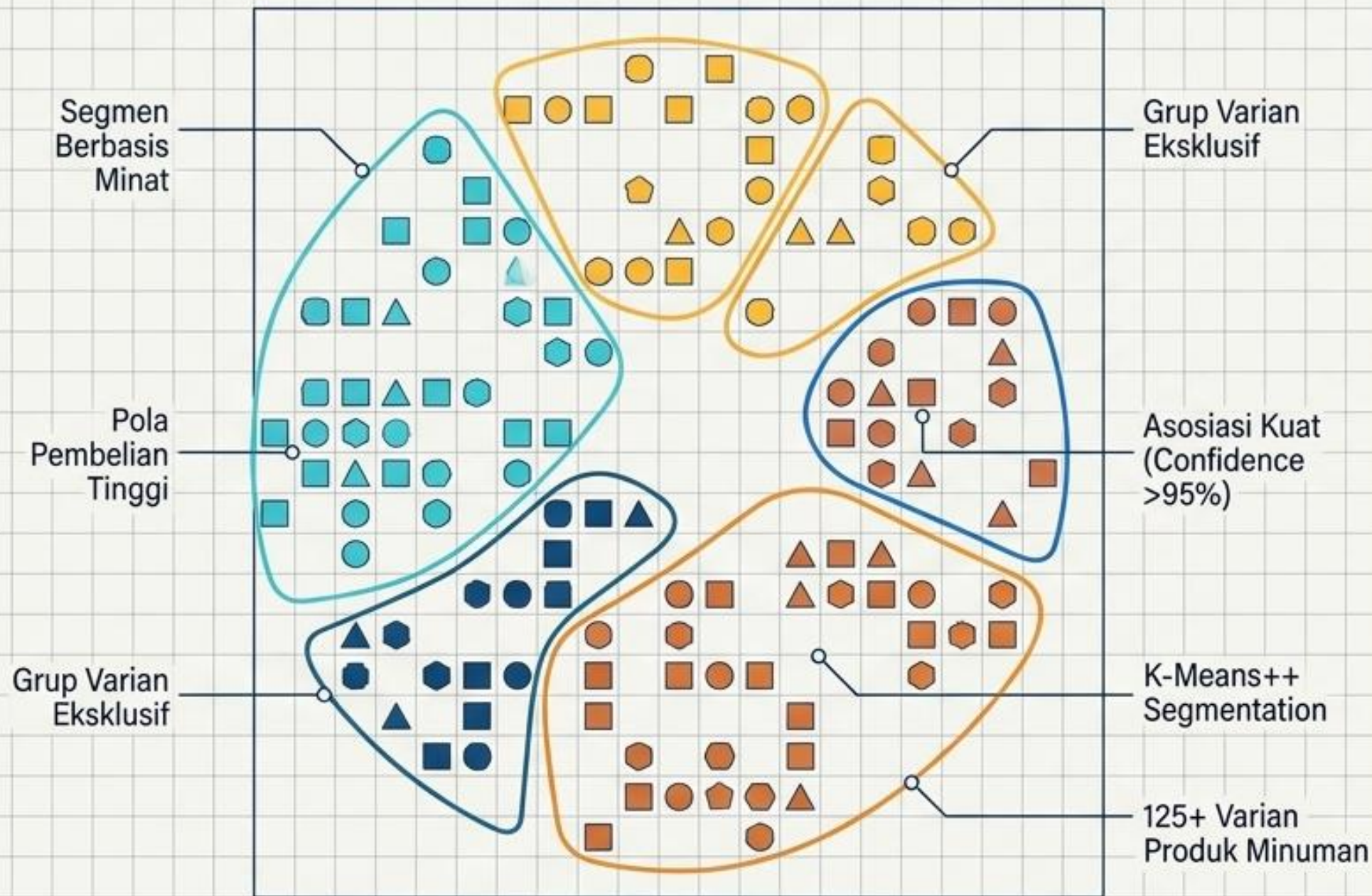
## Presisi Diagnostik Kedokteran



**Algoritma:** Regresi Logistik dan Decision Stump (C4.5).

**Aplikasi:** Mengkategorikan pola penyakit dari citra medis dan mencapai klasifikasi akurasi 100% pada evaluasi gaya berjalan pasien hemiplegia untuk menentukan dosis terapi lanjutan.

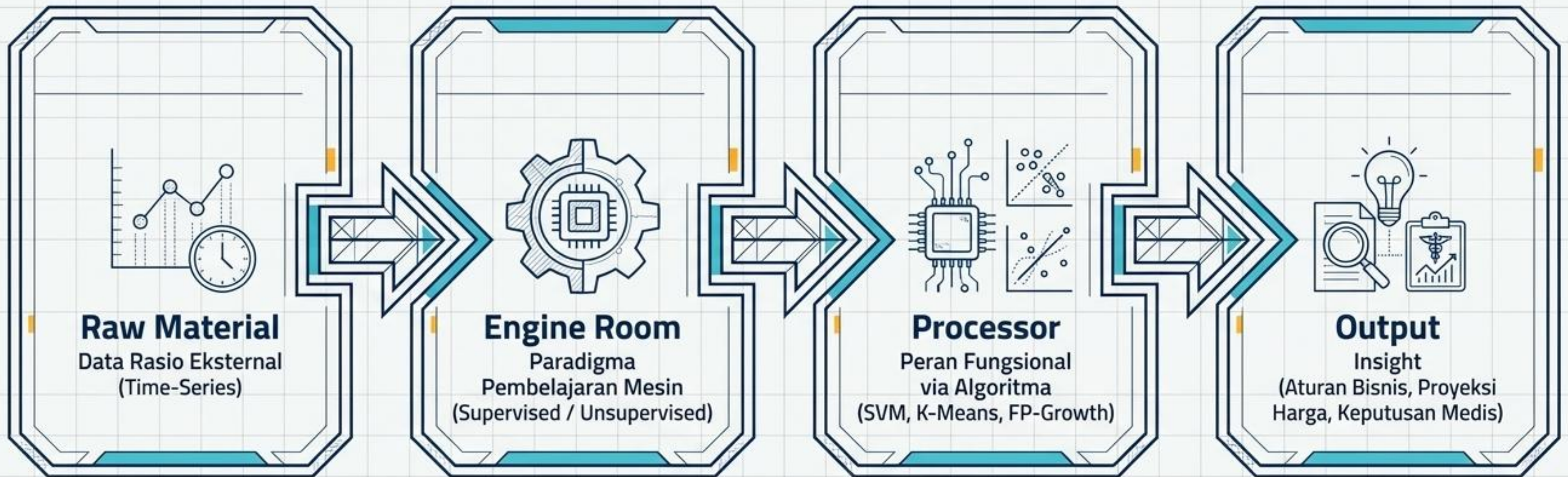
# Unsupervised Learning Mengungkap Segmen Pasar yang Belum Terpetakan



## Analisis Keranjang Konsumen (Industri FMCG)

- **Tantangan:** Merancang strategi promosi dari ratusan varian produk tanpa adanya label klasifikasi awal.
- **Eksekusi K-Means++:** Secara mandiri mengelompokkan 125 jenis produk minuman berdasarkan minat konsumen tanpa umpan balik eksternal.
- **Integrasi Aturan Asosiasi:** Memetakan kombinasi penjualan dengan nilai confidence tertinggi (hingga 100%).
- **Hasil:** Menghasilkan rekomendasi produk cerdas untuk optimalisasi stok dan promosi.

# Siklus Arsitektur Menyatukan Data Mentah Menjadi Kecerdasan Bisnis



Ekosistem tidak bekerja secara terisolasi. **Skala pengukuran data** mendikte **batasan matematika**. Batasan tersebut memfilter **algoritma**. Algoritma tersebut mendefinisikan **peran mining**, yang pada akhirnya menentukan **arsitektur** solusi bisnis yang dihasilkan.

# Tiga Prinsip Utama Implementasi Data Mining yang Komprehensif

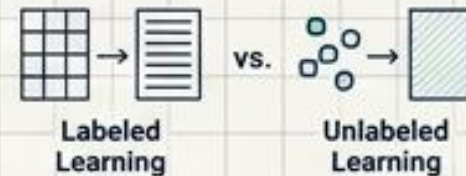
## 1. Hierarki Data Menentukan Kapasitas Analitik

Jangan memaksakan perhitungan kompleks pada data berlabel (Nominal). Validasi skala pengukuran adalah langkah pertama sebelum memilih model.



## 2. Paradigma Bergantung pada Ketersediaan Label

Supervised Learning adalah pilihan absolut untuk prediksi presisi, namun Unsupervised Learning adalah instrumen terbaik untuk eksplorasi area buta bisnis.



## 3. Algoritma Adalah Alat, Bukan Tujuan

Penggunaan CNN-SVM atau K-Means++ harus selalu dikaitkan dengan pilar fungsional dasar demi menghasilkan *insight* yang dapat dieksekusi.

