

# Unsupervised Learning

Oleh : Dodo Zaenal Abidin

*Unsupervised Learning* adalah salah satu jenis pembelajaran dalam *machine learning* di mana **algoritma** mencoba **menemukan pola** atau struktur dalam data tanpa menggunakan **label** atau jawaban yang sudah ada. Dalam pembelajaran ini, algoritma **tidak** diberikan **output** yang benar untuk dilatih, melainkan hanya data mentah yang tidak berlabel. Tujuannya adalah untuk menemukan hubungan atau struktur yang tersembunyi dalam data tersebut.

Beberapa metode dan teknik yang umum digunakan dalam Unsupervised Learning meliputi:

## 1. Clustering

**Clustering** adalah teknik untuk mengelompokkan data ke dalam beberapa kelompok (cluster) sehingga data dalam satu kelompok lebih mirip satu sama lain daripada data dalam kelompok lain. Algoritma clustering yang populer termasuk:

- **K-Means Clustering**: Mengelompokkan data menjadi KKK cluster berdasarkan jarak rata-rata dari titik tengah cluster.
- **Hierarchical Clustering**: Mengelompokkan data dalam hierarki bertingkat dari cluster.
- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**: Mengelompokkan data berdasarkan kepadatan titik.

## 2. Dimensionality Reduction

Teknik ini digunakan untuk **mengurangi jumlah fitur dalam dataset sambil mempertahankan informasi yang paling penting**. Tujuannya adalah untuk **menyederhanakan data dan mempermudah visualisasi**. Metode yang umum termasuk:

- **Principal Component Analysis (PCA)**: Mengurangi dimensi data dengan mengubah variabel asli menjadi variabel baru yang tidak berkorelasi satu sama lain (komponen utama).
- **t-Distributed Stochastic Neighbor Embedding (t-SNE)**: Mengurangi dimensi data sambil menjaga hubungan jarak non-linear antara data.

## 3. Association Rules

**Association Rules** digunakan untuk **menemukan hubungan atau pola yang sering muncul dalam dataset**. Metode ini sering digunakan dalam analisis keranjang belanja untuk menemukan produk yang sering dibeli bersama. Algoritma yang umum digunakan adalah:

- **Apriori Algorithm**: Mengidentifikasi item-item yang sering muncul bersama dalam dataset.

## 4. Anomaly Detection

**Anomaly Detection** adalah teknik untuk mendeteksi data yang tidak biasa atau berbeda secara signifikan dari mayoritas data. Ini berguna dalam mendeteksi penipuan, kesalahan dalam sistem, atau kejadian tidak biasa lainnya.

### Contoh Penggunaan Unsupervised Learning:

- **Segmentasi Pelanggan:** Mengelompokkan pelanggan berdasarkan perilaku pembelian mereka untuk tujuan pemasaran.
- **Deteksi Penipuan:** Mengidentifikasi transaksi yang mencurigakan berdasarkan pola transaksi normal.
- **Rekomendasi Produk:** Menggunakan pola perilaku pengguna untuk merekomendasikan produk yang mungkin menarik bagi mereka.
- **Visualisasi Data:** Mengurangi dimensi data untuk mempermudah visualisasi dan eksplorasi.

*Unsupervised Learning* sangat berguna dalam situasi di mana kita tidak memiliki label data yang tersedia atau kita ingin menemukan pola tersembunyi dalam data yang kompleks.

Berikut adalah tabel yang mencantumkan beberapa algoritma *Unsupervised Learning* berdasarkan kelompoknya:

Kelompok	Algoritma
Clustering	✓ K-Means Clustering
	✓ Hierarchical Clustering
	✓ DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
	✓ Mean Shift
	✓ Gaussian Mixture Models (GMM)
	✓ Agglomerative Clustering
	✓ OPTICS (Ordering Points to Identify the Clustering Structure)
	✓ BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)
Dimensionality Reduction	• Principal Component Analysis (PCA)
	• t-Distributed Stochastic Neighbor Embedding (t-SNE)
	• Independent Component Analysis (ICA)
	• Singular Value Decomposition (SVD)
	• Linear Discriminant Analysis (LDA)
	• Non-negative Matrix Factorization (NMF)
	• UMAP (Uniform Manifold Approximation and Projection)
Association Rules	○ Apriori Algorithm
	○ Eclat Algorithm
	○ FP-Growth (Frequent Pattern Growth)
Anomaly Detection	✓ Isolation Forest
	✓ One-Class SVM
	✓ Local Outlier Factor (LOF)
	✓ Robust Covariance Estimation
	✓ Ensemble Methods (e.g., Combination of Isolation Forest and LOF)
Other Methods	• Autoencoders
	• Self-Organizing Maps (SOM)
	• Restricted Boltzmann Machines (RBM)
	• Generative Adversarial Networks (GANs)
	• K-Nearest Neighbors (K-NN) for Density Estimation

Tabel ini mencakup berbagai algoritma *Unsupervised Learning* yang digunakan untuk clustering, reduksi dimensi, aturan asosiasi, deteksi anomali, dan metode lainnya.

Berikut adalah tabel yang mencantumkan beberapa contoh dataset yang sering digunakan untuk pekerjaan *Unsupervised Learning*, beserta [link](#) web untuk mengaksesnya:

Kategori	Dataset	Deskripsi	Link
Clustering	Iris Dataset	Data tentang panjang dan lebar sepal serta petal dari tiga spesies bunga Iris.	<a href="#">Iris Dataset</a>
	Mall Customers Dataset	Data tentang pengeluaran tahunan pelanggan dan skor spending mereka untuk segmentasi pelanggan.	Mall Customers Dataset
	Wine Dataset	Informasi kimia dari berbagai jenis anggur dari tiga wilayah yang berbeda.	<a href="#">Wine Dataset</a>
Dimensionality Reduction	MNIST Dataset	Gambar digit tangan (0-9) yang dapat digunakan untuk reduksi dimensi dan visualisasi.	MNIST Dataset
	Fashion MNIST	Gambar item pakaian, berguna untuk tugas reduksi dimensi dan visualisasi.	Fashion MNIST
	Digits Dataset	Gambar digit dari 0-9, sering digunakan dalam PCA dan t-SNE untuk visualisasi.	<a href="#">Digits Dataset</a>
Association Rules	Groceries Dataset	Data daftar belanja dari berbagai pelanggan, digunakan untuk menemukan aturan asosiasi dan pola pembelian.	Groceries Dataset
	Retail Market Basket Dataset	Data transaksi ritel untuk analisis aturan asosiasi dan penambangan pola frekuen.	Retail Market Basket Dataset
Anomaly Detection	KDD Cup 1999 Dataset	Data serangan jaringan dan data normal, digunakan untuk deteksi anomali dalam jaringan.	<a href="#">KDD Cup 1999 Dataset</a>
	Credit Card Fraud Detection Dataset	Data transaksi kartu kredit yang termasuk transaksi normal dan transaksi penipuan.	Credit Card Fraud Detection Dataset
	NYC Taxi Dataset	Data perjalanan taksi di New York City, digunakan untuk mendeteksi perjalanan taksi yang tidak biasa.	NYC Taxi Dataset
Other Methods	CIFAR-10 Dataset	Gambar kecil dari 10 kelas berbeda, digunakan dalam autoencoders untuk pembelajaran representasi.	CIFAR-10 Dataset

	20 Newsgroups Dataset	Kumpulan dokumen dari 20 kelompok berita berbeda, cocok untuk clustering dan reduksi dimensi dalam teks.	20 Newsgroups Dataset
	Amazon Reviews Dataset	Ulasan pelanggan tentang produk Amazon, digunakan untuk clustering ulasan atau deteksi anomali dalam teks.	Amazon Reviews Dataset

Tabel ini mencakup berbagai dataset yang digunakan untuk clustering, reduksi dimensi, aturan asosiasi, deteksi anomali, dan metode lainnya dalam Unsupervised Learning, beserta link untuk mengaksesnya.