

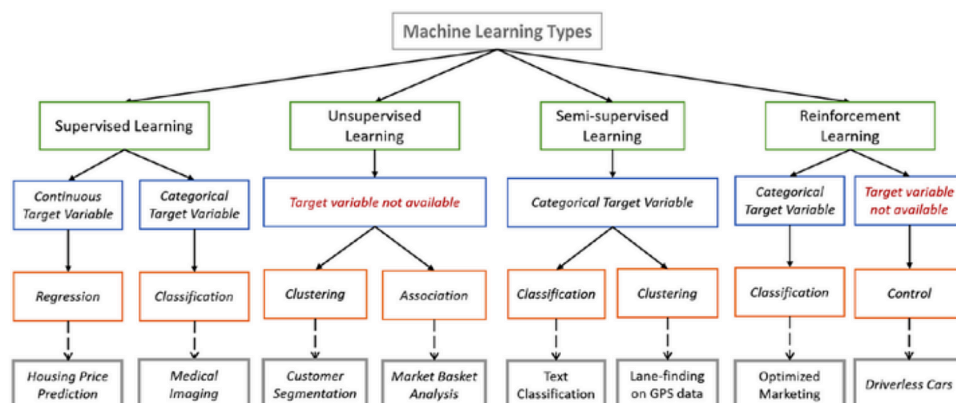
Machine Learning

Pengertian Machine Learning

Machine learning adalah cabang kecerdasan buatan (AI) yang memungkinkan sistem komputer belajar dan meningkatkan performa dari pengalaman, tanpa diprogram secara eksplisit. Sistem ini menggunakan algoritma untuk menganalisis data, mengidentifikasi pola, dan membuat keputusan.

Metode ini mirip dengan cara manusia belajar dari pengalaman. Jika seorang anak diajari membaca huruf berulang kali, maka ia akan mampu mengenali huruf tersebut. Dalam machine learning, mesin belajar dari data, bukan pengalaman, sehingga data menjadi komponen krusial.

Tipe Algoritma Machine Learning



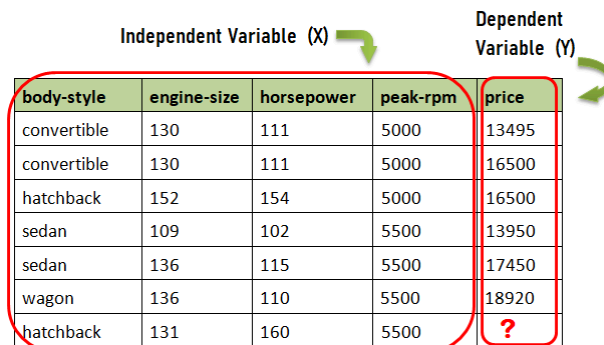
Source : <https://www.dataaksi.id/2023/09/regresi-berganda-dan-machine-learning.html>

Supervised Learning

Supervised Learning adalah model machine learning yang menggunakan data training yang berisi jawaban untuk masalah yang ingin diselesaikan. Mesin diharapkan meniru pola pada input data (prediktor) untuk menghasilkan output serupa.

REGRESI

Regresi : Metode yang digunakan untuk mempelajari hubungan antara satu atau lebih pengubah yaitu variable bebas (X) dengan variable tak bebas (Y)



body-style	engine-size	horsepower	peak-rpm	price
convertible	130	111	5000	13495
convertible	130	111	5000	16500
hatchback	152	154	5000	16500
sedan	109	102	5500	13950
sedan	136	115	5500	17450
wagon	136	110	5500	18920
hatchback	131	160	5500	?

Source : <https://ilmudatapy.com/apa-itu-regresi-klasifikasi-dan-clustering-klasterisasi/>

Jenis - Jenis Regresi

1. Regresi Linier

Regresi Linier merupakan metode statistik untuk memodelkan hubungan antara **variabel dependen (Y)** dengan satu atau lebih **variabel independen (X)**.

- 1 variabel X → Regresi Linier Sederhana
- Lebih dari 1 variabel X → Regresi Linier Berganda

$$Y = b_0 + b_1 \cdot X$$

Dimana :

- Y : Nilai prediksi dari variabel dependen (Y)
- X : Nilai variabel independen
- b_0 : Intersep (nilai Y saat $X = 0$)
- b_1 : Kemiringan garis (slope) → menunjukkan seberapa besar perubahan Y jika X naik 1 unit.

Rumus Umum

$$Y = b_0 + b_1 \cdot X$$

Dimana :

- Y : Nilai prediksi dari variabel dependen (Y)
- X : Nilai variabel independen
- b_0 : Intersep (nilai Y saat $X = 0$)
- b_1 : Kemiringan garis (slope) → menunjukkan seberapa besar perubahan Y jika X naik 1 unit.

Contoh implementasi : Prediksi harga rumah (Y) berdasarkan luas rumah (X)

$$Y = 50 + 3 \cdot X$$

Dimana :

- Intersep $b_0 = 50$: Rumah sekecil apapun (misal $X = 0$), harganya minimal 50 juta.
- Koefisien $b_1 = 3$: Setiap tambahan 1 m², harga naik 3 juta.

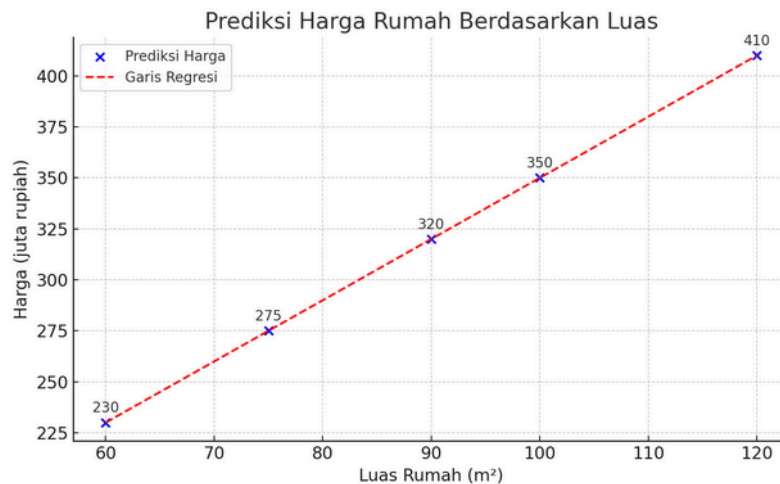
Perhitungan manual

Kalau rumah memiliki luas 100 m² :

$$Y = 50 + 3 \cdot 100 = 350$$

Sehingga, prediksi harga rumah adalah 350 juta.

Grafik hasil prediksi



Dimana :

- Titik biru: prediksi harga rumah untuk masing-masing luas.
- Garis merah putus-putus: garis regresi dari model $Y = 50 + 3X$.

2. Regresi Logistik

Regresi Logistik merupakan model statistik untuk memprediksi **variabel dependen kategorik** (biasanya biner: 0/1, ya/tidak, churn/tidak churn, dsb) berdasarkan satu atau lebih variabel independen.

Rumus Umum

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

Atau

$$\text{logit}(p) = \ln \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

Dimana :

- p : Probabilitas $Y = 1$
- β_0 : Intersep
- β_i : Koefisien fitur X_i

Contoh implementasi : Prediksi penderita diabetes

- Fitur
 - Umur (X_1)
 - BMI (X_2)
 - Glukosa (X_3)
- Model

$$\text{logit}(p) = -5 + 0.1 \cdot \text{umur} + 0.05 \cdot \text{BMI} + 0.08 \cdot \text{glukosa}$$

Semisal diberikan input :

- Umur : 50
- BMI : 30
- Glukosa : 120

Maka :

$$z = -5 + 0.1(50) + 0.05(30) + 0.08(120) = 6.5$$

$$p = \frac{1}{1 + e^{-6.5}} \approx 0.9985 \Rightarrow \text{Prediksi: Penderita Diabetes}$$

KLASIFIKASI

Klasifikasi: Teknik dalam machine learning untuk **memprediksi label kategori** dari suatu data berdasarkan fitur yang dimilikinya. Output dari klasifikasi berbentuk **kelas** atau **kategori**, bukan angka kontinu.

Width (cm)	Weight (g)	Color	Label
15	100	Red	Apple
7	150	Green	Apple
13	120	Yellow	Banana
10	155	Green-Yellowish	Banana
8	160	Red	Apple

Tipe Klasifikasi

1. Klasifikasi Biner (Binary Classification)
 - Hanya terdiri dari 2 kelas
 - Contoh : Sakit / Tidak, Lulus / Tidak
2. Klasifikasi Multikelas (Multiclass Classification)
 - Terdiri lebih dari 2 kelas
 - Contoh: Prediksi jenis bunga (Setosa, Versicolor, Virginica)
3. Klasifikasi Multilabel
 - Satu data bisa punya lebih dari satu label
 - Contoh: Film bisa diklasifikasi sebagai Action & Comedy sekaligus

Algoritma Klasifikasi Populer

Algoritma	Karakteristik
Logistic Regression	Sederhana, cepat, cocok untuk klasifikasi biner
K-Nearest Neighbors	Non-parametrik, prediksi berdasarkan tetangga terdekat
Decision Tree	Visual, mudah dipahami, rentan overfitting
Random Forest	Ensemble dari banyak decision tree, lebih stabil dan akurat
Naive Bayes	Berdasarkan probabilitas, cocok untuk data teks (seperti spam)
Support Vector Machine	Cari hyperplane terbaik, bagus untuk data high-dimensional
Neural Network	Kuat untuk data kompleks, butuh banyak data dan komputasi
