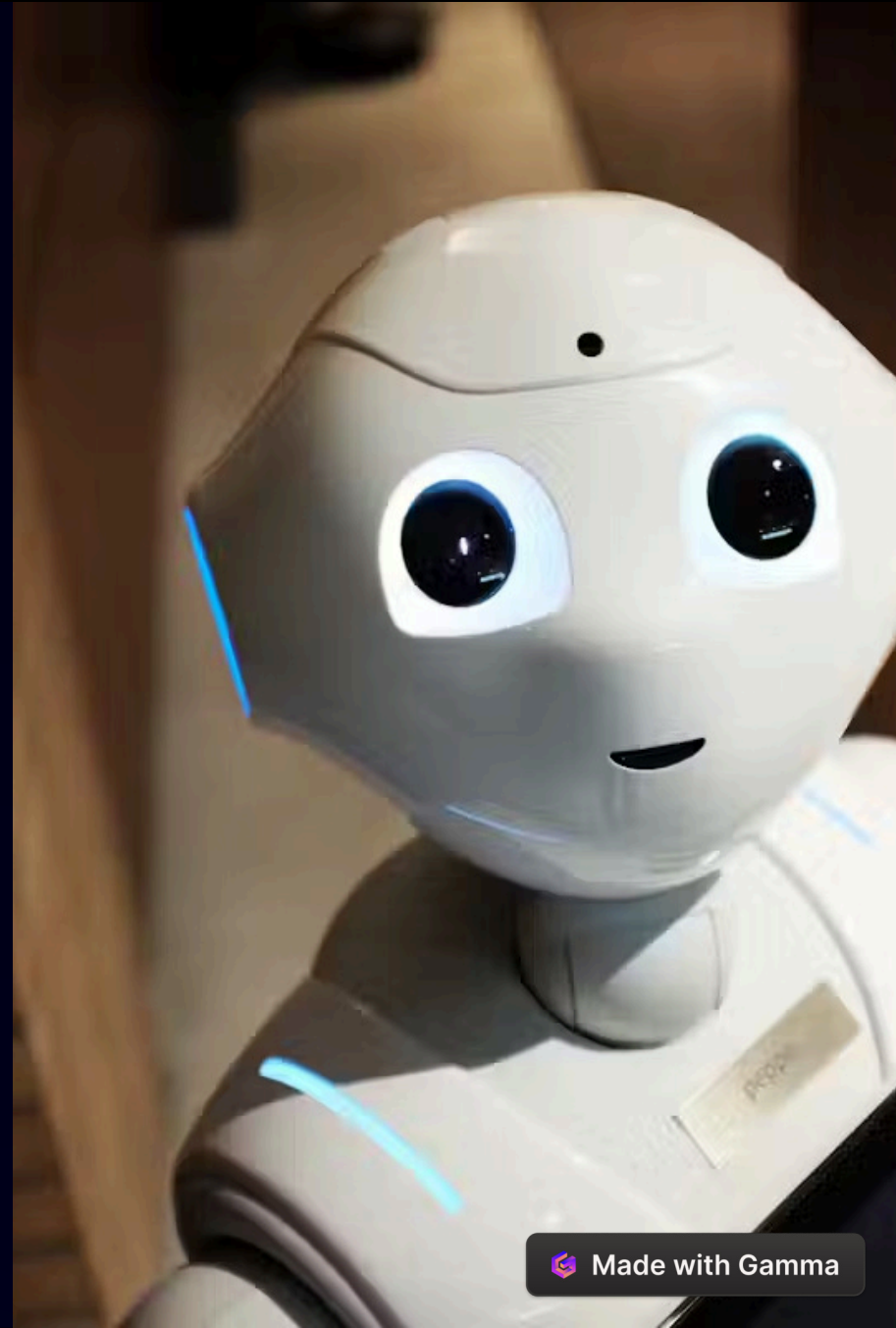


Pengenalan Machine Learning

Data Mining IPB
2025



Apa itu Machine Learning?

Definisi

Machine Learning (ML) adalah subbidang Artificial Intelligence (AI) yang memungkinkan komputer untuk "belajar" dari data tanpa diprogram secara eksplisit.

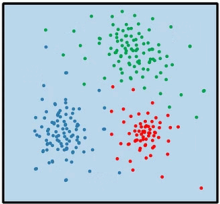
Konsep Dasar

ML menggunakan algoritma untuk mengidentifikasi pola dan membuat prediksi berdasarkan data historis. Semakin banyak data, semakin baik prediksi.

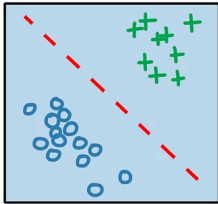
Jenis-jenis Machine Learning

machine learning

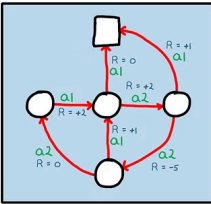
unsupervised learning



supervised learning



reinforcement learning



1

Supervised Learning

Model dilatih dengan data berlabel (input dan output). Contoh: Klasifikasi gambar, prediksi harga.

2

Unsupervised Learning

Model mencari pola dan struktur dalam data tanpa label. Contoh: Clustering, pengurangan dimensi.

3

Reinforcement Learning

Model belajar melalui trial and error, mendapat reward untuk tindakan yang benar. Contoh: Permainan, robot yang belajar berjalan.

Pipeline Machine Learning

Steps in a full machine learning project



Problem Definition

What business problem are we trying to solve? How can it be phrased as a machine learning problem?

Tentukan masalah bisnis yang hendak dipecahkan dan ubah menjadi masalah machine learning yang terdefinisi dengan jelas, termasuk identifikasi tujuan dan pertanyaan utama.

Contoh

Sebuah perusahaan properti ingin memprediksi harga rumah.
Pertanyaan utama adalah: **"Bagaimana cara memprediksi harga rumah berdasarkan fitur seperti ukuran, lokasi, dan jumlah kamar?"**



Data: The New Gold

If machine learning is getting insights out of data, what data we have? How does it match the problem definition? Is our data structured or unstructured? Static or streaming?



Pentingnya Data

Data adalah bahan utama untuk Machine Learning. Kualitas data menentukan hasil model.



Kualitas Data

Pastikan data bersih, akurat, dan relevan dengan masalah yang ingin dipecahkan.



Analisis Data

Pahami karakteristik data, termasuk distribusi, outliers, dan hubungan antar variabel.

Evaluation: What defines success? Is a 95% accurate machine learning model good enough?

Berikut adalah beberapa hal lain yang perlu dipertimbangkan untuk masalah klasifikasi:

- **False negatives** – Model memprediksi negatif, padahal sebenarnya positif. Dalam beberapa kasus, seperti prediksi spam email, false negatives tidak terlalu mengkhawatirkan. Namun, jika sistem computer vision pada mobil self-driving memprediksi tidak ada pejalan kaki padahal sebenarnya ada, hal ini sangat tidak baik.
- **False positives** – Model memprediksi positif, padahal sebenarnya negatif. Memprediksi seseorang menderita penyakit jantung padahal sebenarnya tidak, mungkin terdengar bisa diterima. Namun, lebih baik aman, bukan? Terutama jika prediksi tersebut berdampak negatif pada gaya hidup orang tersebut atau menjerumuskan mereka ke dalam rencana perawatan yang tidak perlu.
- **True negatives** – Model memprediksi negatif, dan memang negatif. Ini merupakan hasil yang baik.
- **True positives** – Model memprediksi positif, dan memang positif. Ini merupakan hasil yang baik.
- **Precision** – Proporsi dari prediksi positif yang benar-benar tepat. Model yang tidak menghasilkan false positives memiliki precision sebesar 1.0
- **Recall** – Proporsi dari kasus positif aktual yang diprediksi dengan benar. Model yang tidak menghasilkan false negatives memiliki recall sebesar 1.0.

Features: Kenali Ciri-ciri Data Anda

1

Identifikasi Features

Tentukan ciri-ciri atau variabel yang relevan dalam data Anda.

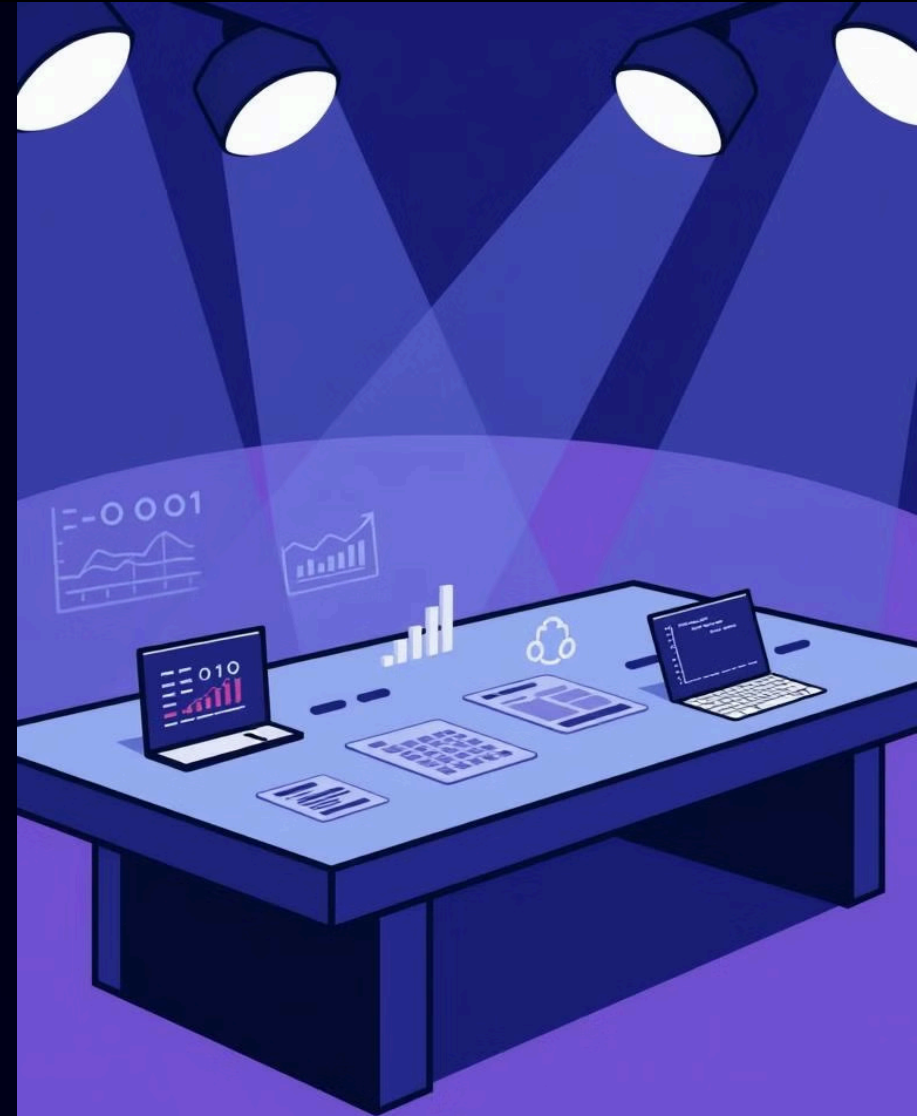
2

Feature Engineering

Transformasikan features untuk meningkatkan performa model. Contoh: Standarisasi, encoding.

Contoh

Dari dataset, memilih fitur-fitur penting seperti luas rumah, jumlah kamar, dan indeks keamanan lingkungan, sambil mengabaikan fitur yang kurang relevan seperti warna cat.



Modelling: Which model should you choose?

1

Linear Regression

Untuk memprediksi nilai kontinu.

2

Logistic Regression

Untuk klasifikasi biner.

3

Decision Tree

Untuk klasifikasi dan regresi.

4

Support Vector Machine

Untuk klasifikasi dan regresi.

5

Neural Network

Untuk masalah kompleks dengan banyak data.

6

1000+ Model Lainnya!

Tools you can use

Data Science

