

Chapter 1 : The Machine Learning Landscape

Bab ini memperkenalkan konsep dasar dan berbagai jenis machine learning, serta memberikan gambaran umum tentang alat dan aplikasi machine learning yang umum digunakan saat ini.

1. Tujuan Utama Bab Ini

- Memahami apa itu machine learning (ML)
- Mengetahui berbagai jenis ML
- Melihat contoh aplikasi ML di dunia nyata
- Mengetahui alur kerja umum dalam proyek ML
- Mengenal istilah-istilah penting seperti dataset, features, dan labels

2. Pengertian Machine Learning menurut pakar

- Definisi menurut Arthur Samuel: " Bidang studi yang memberi komputer kemampuan untuk belajar tanpa diprogram secara eksplisit."
- Definisi menurut Tom Mitchell: " Suatu program komputer dikatakan belajar dari pengalaman E berkenaan dengan beberapa tugas T dan beberapa ukuran kinerja P, jika kinerjanya pada T, sebagaimana diukur dengan P, meningkat seiring dengan pengalaman E.

Contoh:

Memprediksi harga rumah berdasarkan data historis penjualan rumah. Pengalaman = data, tugas = prediksi harga, kinerja = akurasi prediksi.

3. Jenis-jenis machine learning

- Supervised Learning

Input dan output sudah diketahui.

Contoh: regresi, klasifikasi

Algoritma umum: Linear Regression, Decision Trees, Random Forest, SVM, Neural Networks

- Unsupervised Learning

Tidak ada label/output.

Tujuan: menemukan struktur tersembunyi dalam data

Contoh: clustering, dimensionality reduction

Algoritma umum: K-Means, DBSCAN, PCA

- Semi-supervised Learning

Kombinasi data berlabel dan tak berlabel.

Contoh: Google Photos, YouTube Recommendations

- Reinforcement Learning

Agen belajar dengan mencoba-coba dan menerima reward/punishment.

Umum dipakai di game, robotik, sistem rekomendasi dinamis.

4. Contoh Aplikasi Machine Learning

- Deteksi spam
- Pengenalan wajah
- Prediksi penyakit
- Kendaraan otonom
- Rekomendasi produk

5. Alur Umum Proyek Machine Learning

- Tentukan masalah bisnis dan ubah menjadi masalah ML
- Dapatkan data dan pahami
- Bersihkan dan siapkan data
- Pilih model dan latih
- Evaluasi performa
- Tuning dan optimasi
- Deploy ke produksi

6. Terminologi Penting

- Feature: Input (misalnya luas rumah)
- Label: Output (misalnya harga rumah)
- Training set: Data untuk melatih model
- Test set: Data untuk menguji model
- Overfitting: Model terlalu cocok dengan data latih
- Underfitting: Model terlalu sederhana, tidak belajar cukup

7. Tools yang Dipakai di Buku Ini

- Scikit-Learn: untuk ML tradisional
- TensorFlow + Keras: untuk deep learning

8. Insight Kunci dari Bab Ini

- ML adalah pendekatan berbasis data, bukan pemrograman aturan eksplisit.
- Memahami jenis dan alur proyek ML sangat penting sebelum menyelami kode.
- Mengetahui batas dan tantangan: overfitting, kebutuhan data besar, interpretabilitas, dsb.