

Muhammad Harits

1103223159

Bab 1: Lanskap Machine Learning

Ringkasan Bab

Bab ini memberikan pengantar konsep Machine Learning (ML) tanpa melibatkan kode. Di dalamnya dijelaskan:

- Apa itu Machine Learning dan mengapa ML penting.
- Klasifikasi sistem ML berdasarkan:
 1. Pengawasan (supervision): Supervised, Unsupervised, Semi-supervised, dan Reinforcement Learning.
 2. Cara belajar: Batch (offline) learning dan Online learning.
 3. Cara generalisasi: Instance-based learning dan Model-based learning.
- Tantangan umum dalam ML seperti keterbatasan data, overfitting, dan underfitting.
- Konsep dasar evaluasi model: train/test split dan validation set.

Penjelasan Konsep Utama

1. Apa Itu Machine Learning?

Machine Learning (ML) adalah bidang ilmu yang fokus pada pembuatan sistem yang mampu belajar dari data—bukan hanya mengikuti instruksi yang diprogram secara eksplisit.

Tom Mitchell memberikan definisi populer:

“Sebuah program komputer dikatakan belajar dari pengalaman (E) untuk suatu tugas (T) berdasarkan ukuran kinerja (P), jika performanya pada tugas T—yang diukur dengan P—meningkat seiring bertambahnya pengalaman E.”

Mengapa Menggunakan ML?

Machine Learning sangat berguna untuk:

- Menyelesaikan masalah yang terlalu kompleks untuk didekati dengan algoritma tradisional.
- Menggantikan aturan-aturan manual yang panjang dan sulit dipelihara.
- Membuat sistem yang adaptif terhadap perubahan lingkungan.
- Menemukan pola dan wawasan dalam data yang besar dan kompleks (data mining).

2. Jenis-Jenis Sistem Machine Learning

A. Berdasarkan Pengawasan (Supervision)

1. Supervised Learning (Pembelajaran Terarah)
Data latih sudah memiliki label atau jawaban.
 - Klasifikasi → memprediksi label kategori (misalnya spam / bukan spam).
 - Regresi → memprediksi nilai kontinu (misalnya harga rumah).
Contoh algoritma: k-NN, Linear/Logistic Regression, SVM, Decision Trees, Random Forest, Neural Network.
2. Unsupervised Learning (Pembelajaran Tak Terarah)
Data tidak memiliki label; sistem mencari pola sendiri.
 - Clustering → mengelompokkan data yang mirip (misalnya segmentasi pelanggan).
 - Dimensionality Reduction & Visualization → menyederhanakan data tanpa kehilangan informasi penting (misalnya PCA, t-SNE).
 - Anomaly Detection → mendeteksi data yang menyimpang.
 - Association Rule Learning → menemukan hubungan antarfaktor dalam data (contoh: pola pembelian di supermarket).
3. Semi-supervised Learning (Pembelajaran Semi-terarah)
Mengombinasikan sedikit data berlabel dengan banyak data tidak berlabel.
Contoh: Google Photos yang mengelompokkan wajah-wajah yang sama, pengguna hanya memberi satu label nama.
4. Reinforcement Learning (Pembelajaran Penguatan)
Sistem (agent) belajar melalui percobaan langsung dengan lingkungan. Agent melakukan tindakan, lalu menerima reward atau penalty. Tujuannya adalah menemukan strategi terbaik untuk memaksimalkan reward jangka panjang.
Contoh: AlphaGo, robot berjalan.

B. Berdasarkan Cara Belajar

1. Batch Learning (Offline Learning)
Sistem dilatih sekali menggunakan seluruh data. Jika ada data baru, pelatihan harus diulang dari awal.
2. Online Learning (Incremental Learning)
Sistem belajar secara bertahap, menerima data sedikit demi sedikit (misalnya per mini-batch).
Berguna untuk data yang terus mengalir atau dataset besar yang tidak memuat seluruhnya di memori.

C. Berdasarkan Cara Model Menggeneralisasi

1. Instance-Based Learning
Sistem menyimpan contoh data latih, lalu membandingkan data baru berdasarkan tingkat kemiripan.
Contoh: k-Nearest Neighbors.

2. Model-Based Learning
Sistem membangun model (misalnya persamaan linear) dari data latih dan menggunakan model tersebut untuk prediksi.
Contoh: Linear Regression—mencari parameter terbaik untuk membuat garis prediksi.

3. Tantangan dalam Machine Learning

- Jumlah data tidak cukup → sebagian besar algoritma membutuhkan data besar.
- Data tidak representatif → mengakibatkan sampling bias.
- Data berkualitas buruk → berisi error, outlier, atau noise.
- Fitur tidak relevan → kualitas model bergantung pada kualitas fitur (feature engineering).

Overfitting

Model terlalu kompleks sehingga sangat bagus di data latih tapi buruk di data baru.
Solusi: sederhanakan model, tambah data, atau gunakan regularization.

Underfitting

Model terlalu sederhana sehingga gagal menangkap pola data.
Solusi: gunakan model lebih kompleks, tambahkan fitur, atau kurangi regularisasi.

4. Pengujian dan Validasi

- Training set & test set
Data dilatih pada training set, lalu dievaluasi pada test set untuk mengukur generalization error.
- Validation set
Digunakan untuk menyetel hyperparameter tanpa memengaruhi test set.
Prosesnya:
 1. Pisahkan sebagian data sebagai validation set.
 2. Latih beberapa model dan bandingkan performanya.
 3. Pilih model terbaik, latih ulang dengan seluruh training set, lalu evaluasi pada test set.