## Kecerdasan Buatan (Artificial Intelligence / AI)

* **AI** adalah ilmu yang mengajarkan mesin untuk belajar, bertindak, dan berpikir seperti manusia dalam menyelesaikan tugas kehidupan nyata.
* **AI** menerapkan peniruan perilaku manusia pada komputer agar dapat mempelajari dan melakukan tugas tanpa bantuan eksplisit mengenai output yang diharapkan.
* **Machine Learning (ML)** merupakan cabang AI yang menggunakan metode statistika untuk membuat komputer mampu mengenali pola pada data tanpa diprogram secara langsung.

## Tahapan Pengembangan AI

1. **Digitalise & Collect**  
   Mengumpulkan serta menyimpan data yang akan digunakan dalam pembangunan AI.
2. **Transform**
   * Menyiapkan dan mengubah data ke dalam format yang sesuai.
   * Melakukan evaluasi untuk mengidentifikasi data yang tidak relevan.
3. **Train**  
   Melatih komputer dengan data agar mampu belajar dan melakukan tugas berdasarkan pola yang dipelajari.
4. **Execute**  
   Menggunakan model AI yang sudah dilatih untuk menyelesaikan tugas.  
   Proses eksekusi kemudian dievaluasi kembali untuk memastikan hasil sesuai harapan.
5. **Provide Insights to Make Decisions**  
   Mengekstraksi pengetahuan dari model untuk membantu pengambilan keputusan dan meningkatkan pemahaman dalam pengembangan AI berikutnya.

## Data

* **Data**: Fakta atau informasi yang dapat berupa teks, angka, gambar, suara, dan bentuk lainnya.
* **Dataset**: Kumpulan data yang disusun terstruktur, biasanya berbentuk tabel (baris & kolom) dan disimpan dalam format seperti CSV, Excel, JSON, dll.
* **Basis Data (Database)**: Kumpulan data yang diatur secara terorganisir agar mudah diakses, disimpan, dan dikelola.

## Jenis Data

1. **Data Terstruktur**  
   Data yang memiliki format tetap dan mudah diproses.
   * **Data Kuantitatif**
     + **Data Kontinu**: Nilai numerik yang dapat berupa desimal atau bilangan bulat. Contoh: tinggi, berat, suhu, waktu, usia.
     + **Data Diskrit**: Hanya berupa bilangan bulat, tidak bisa dipecah lebih kecil. Contoh: jumlah anak, jumlah mobil.
2. **Data Kategorikal**
   * **Ordinal**: Data dengan urutan atau peringkat tertentu. Contoh: tingkat pendidikan, ranking lomba.
   * **Nominal**: Data kategori tanpa urutan. Contoh: warna, jenis kelamin.
3. **Data Tidak Terstruktur**  
   Data yang tidak memiliki format jelas, sulit diatur dalam tabel. Contoh: teks bebas, gambar, video, audio.

## Pengenalan Machine Learning

* **Machine Learning** adalah cabang AI yang memberi komputer kemampuan untuk belajar secara otomatis dari data tanpa perlu diprogram secara eksplisit.

## Tipe-Tipe Machine Learning

### 1. Supervised Learning

* **Definisi**: Proses pembelajaran menggunakan data yang memiliki **label** atau jawaban.
* **Cara Kerja**:
  + Data berlabel dimasukkan ke model.
  + Model melakukan perhitungan berulang untuk mengenali pola input.
  + Hasilnya berupa prediksi label yang sesuai dengan data aktual.
* **Metode utama**:
  + **Classification** → mengelompokkan data (contoh: klasifikasi email spam/tidak spam).
  + **Regression** → memprediksi nilai numerik (contoh: memprediksi harga rumah).
* **Contoh penerapan**:
  + Klasifikasi jenis buah.
  + Deteksi spam pada pesan.
  + Prediksi harga rumah.

### 2. Unsupervised Learning

* **Definisi**: Proses pembelajaran **tanpa label data**, model mencoba menemukan pola atau struktur secara mandiri.
* **Cara Kerja**:
  + Algoritma menginterpretasi data untuk mencari kesamaan atau perbedaan.
  + Tidak ada jawaban yang disediakan manusia.
* **Contoh penerapan**:
  + Clustering (mengelompokkan pelanggan berdasarkan kebiasaan belanja).
  + Market basket analysis (mencari pola produk yang sering dibeli bersamaan).
  + Segmentasi gambar.

### 3. Reinforcement Learning

* **Definisi**: Teknik pembelajaran berdasarkan proses **trial and error**, di mana agent (mesin) belajar mengambil keputusan untuk memaksimalkan reward.
* **Cara Kerja**:
  + Agent melakukan suatu tindakan.
  + Jika benar → mendapat **reward (Rt+1)**.
  + Jika salah → mendapat **hukuman/sanksi (St+1)**.
  + Tujuannya adalah memilih tindakan terbaik secara berurutan untuk hasil optimal.
* **Contoh penerapan**:
  + Robot navigasi (belajar bergerak menghindari rintangan).
  + Game AI (belajar strategi untuk menang).
  + Sistem rekomendasi yang menyesuaikan sesuai feedback pengguna.

## Machine Learning Workflow

### 1. Pengumpulan Data

* Dataset bisa diperoleh dari sumber publik seperti **Kaggle, Zenodo, UCI**, atau dari sumber internal perusahaan (database, file, sensor, dll).
* Data yang dikumpulkan bisa berupa **structured data** (Excel, SQL database) maupun **unstructured data** (teks, email, video, audio, gambar, data sensor).
* Faktanya, lebih dari **80% data** yang digunakan saat ini adalah unstructured data.
* Dataset mentah perlu diproses lebih lanjut sebelum bisa digunakan untuk membangun model ML → inilah fungsi **data preprocessing**.

### 2. Data Preprocessing

Tahap ini bertujuan mengubah data mentah menjadi lebih siap dipakai oleh algoritma ML.  
Terdiri dari beberapa bagian:

**a. Data Cleaning**

* Menangani missing value.
* Menghapus data duplikat.
* Memperbaiki data tidak konsisten.
* Mengatasi imbalance data.

**b. Data Transformation**

* **Scaling**
  + Standardisasi: mean = 0, std = 1 → cocok untuk data dengan distribusi normal.
  + Normalisasi: rentang [0,1] atau [-1,1] → cocok untuk data dengan skala berbeda.
* Mengonversi format data:
  + Kolom tanggal → diubah jadi format datetime.
  + Data kategorikal → diubah menjadi kategori (bukan numeric palsu).

**c. Data Integration**

* Menggabungkan beberapa dataset.
* Menghapus fitur duplikat.
* Menyamakan format data.

### 3. Model Development

* Tahap ini mencakup **pemilihan algoritma ML** serta **tuning hyperparameter**.
* Tidak ada satu model yang cocok untuk semua data (No Free Lunch Theorem).
* Contoh kasus: klasifikasi biner → bisa dipilih logistic regression atau SVM classifier, tergantung data.
* **Hyperparameter** yang diatur misalnya: epochs, learning rate, optimizer, jumlah K pada KNN, dll.
* Proses tuning hyperparameter bertujuan menemukan kombinasi terbaik untuk performa model → disebut juga **model development**.

### 4. Model Evaluation

* Dilakukan menggunakan **data uji** (test set) yang belum pernah dipakai model sebelumnya.
* Langkah evaluasi:
  1. Model memprediksi label data uji.
  2. Bandingkan hasil prediksi dengan label aktual.
  3. Hitung metrik evaluasi seperti akurasi, precision, recall, F1-score, atau error rate.
* Tujuan: mengukur kemampuan generalisasi model terhadap data baru.

### 5. Model Deployment

* Setelah model memiliki performa yang baik, langkah berikutnya adalah **deployment**, yaitu mengintegrasikan model ke dalam sistem nyata agar bisa digunakan oleh pengguna atau aplikasi.
* **Tujuan deployment**: memastikan model tidak hanya bekerja di lingkungan eksperimen, tetapi juga dapat memberi nilai nyata di dunia produksi.
* **Langkah umum deployment**:
  1. **Mempersiapkan environment** → memilih platform deployment (server, cloud, edge device).
  2. **Membungkus model** dalam bentuk API, service, atau aplikasi agar bisa diakses oleh sistem lain.
  3. **Monitoring & maintenance** → memantau performa model di dunia nyata, mendeteksi data drift atau penurunan akurasi.
  4. **Feedback loop** → data baru dari pengguna/lingkungan dikumpulkan untuk melatih ulang (retraining) model agar tetap relevan.
* **Contoh penerapan**:
  1. Model prediksi harga rumah di-deploy ke aplikasi web.
  2. Model rekomendasi produk diintegrasikan dengan e-commerce.
  3. Model deteksi fraud dipasang di sistem transaksi bank secara real-time.