

# Tugas 1: Algoritma Machine Learning

**Raffa Yuda Pratama - 0110224081 <sup>1\*</sup>, Aan Adriyana - 0110224014 <sup>1</sup> and  
Muhammad Varrel Mahardika - 0110224123 <sup>1</sup>**

<sup>1</sup> Teknik Informatika, STT Terpadu Nurul Fikri, Depok

\*E-mail: [0110224081@student.nurulfikri.ac.id](mailto:0110224081@student.nurulfikri.ac.id)

**Abstract.** Laporan ini membahas konsep dasar dalam pembelajaran mesin, termasuk supervised learning, unsupervised learning, dan reinforcement learning. Selain itu, disajikan 18 studi kasus dunia nyata yang mencakup algoritma regresi, klasifikasi, klustering, dan asosiasi. Setiap kasus dilengkapi dengan penjelasan solusi dan sitasi dari jurnal atau penelitian terkait. Tujuan laporan ini adalah untuk memberikan pemahaman yang komprehensif mengenai penerapan algoritma machine learning dalam berbagai bidang.

## 1. Konsep Dasar Machine Learning

Machine learning (pembelajaran mesin) merupakan sub-bidang dari kecerdasan buatan (AI) yang berfokus pada pengembangan algoritma dan model yang memungkinkan komputer untuk "belajar" dari data tanpa diprogram secara eksplisit. Konsep intinya adalah membangun sistem yang dapat mengidentifikasi pola, membuat keputusan, dan meningkatkan kinerjanya secara otomatis melalui pengalaman (data). Berbeda dengan pemrograman tradisional di mana manusia harus menuliskan semua aturan, dalam machine learning, komputer belajar aturan-aturan tersebut sendiri dari contoh-contoh data yang diberikan (Shaveta, 2023).

Perkembangan machine learning didorong oleh ketersediaan data dalam jumlah besar (big data), peningkatan daya komputasi, dan kemajuan dalam algoritma. Teknologi ini sekarang menjadi tulang punggung dalam berbagai aplikasi modern seperti kendaraan otonom, asisten virtual, sistem rekomendasi, dan diagnosis medis. Terdapat tiga pendekatan utama: supervised learning, unsupervised learning, dan reinforcement learning.

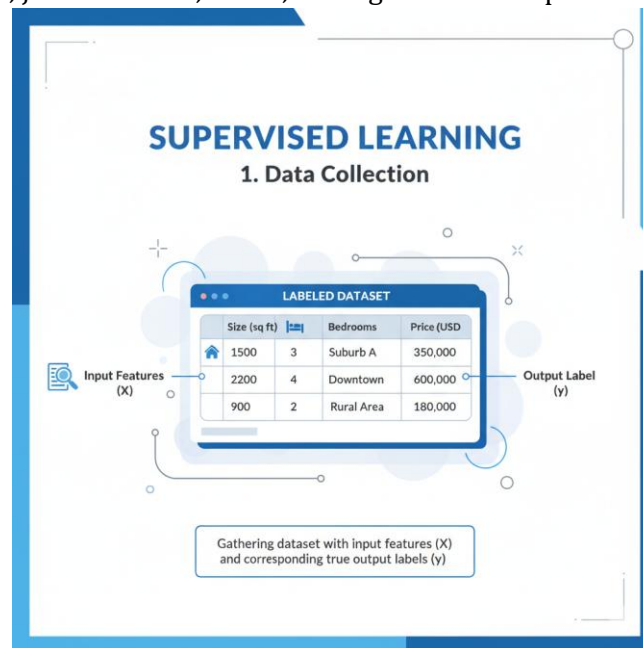
### *1.1 Supervised Learning*

Supervised Learning adalah jenis pembelajaran mesin di mana model dilatih menggunakan data yang telah diberi label (berpasangan antara input dan output yang benar). Ibaratnya, Anda memiliki seorang guru yang menunjukkan contoh soal beserta jawabannya. Model belajar dari contoh-contoh ini untuk membuat prediksi pada data baru yang belum pernah dilihat (Santoso, Abijono, & Anggreini, 2021).

## Tahap-tahap Supervised Learning:

### 1. Pengumpulan Data Berlabel

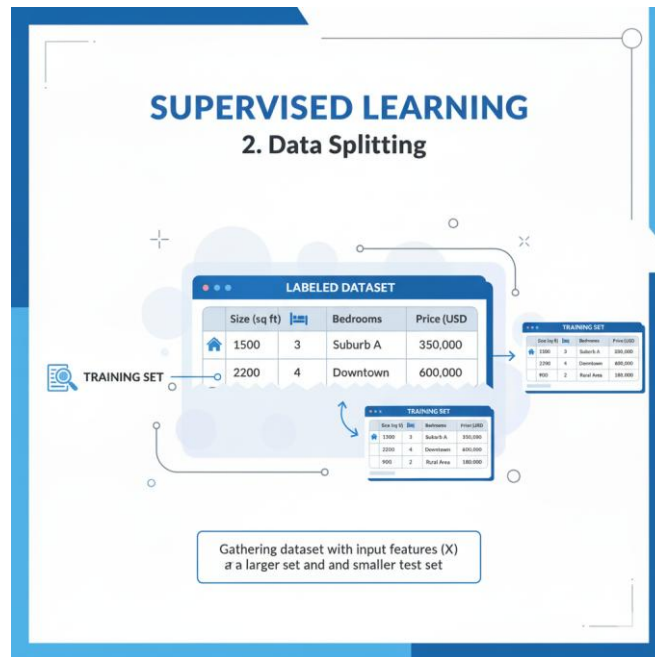
Tahap awal adalah mengumpulkan dataset yang berisi fitur input dan label output yang sesuai. Misalnya, jika Anda ingin memprediksi harga rumah, fitur input bisa berupa ukuran rumah, jumlah kamar, lokasi, sedangkan label output adalah harga sebenarnya.



Gambar 1 Supervised Learning Data Collection

### 2. Pembagian Data

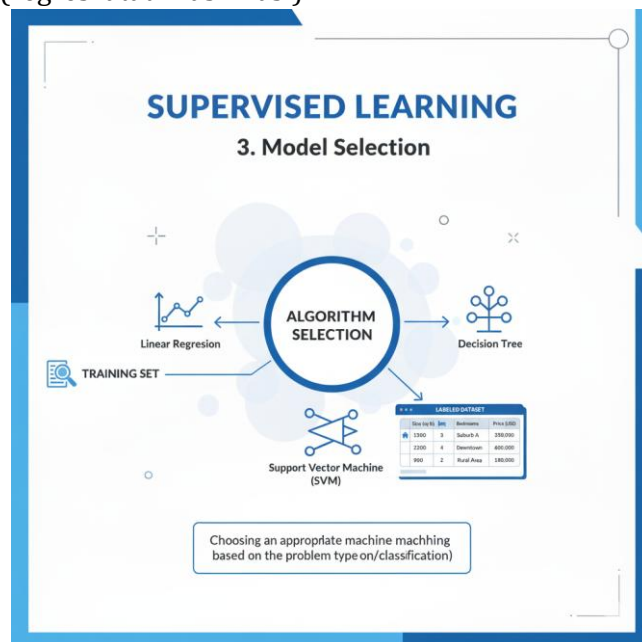
- **Training Set:** Digunakan untuk melatih model.
- **Test Set:** Digunakan untuk mengevaluasi kinerja model pada data yang belum pernah dilihat sebelumnya.



Gambar 2 Supervised Learning Data splitting

### 3. Pemilihan Model

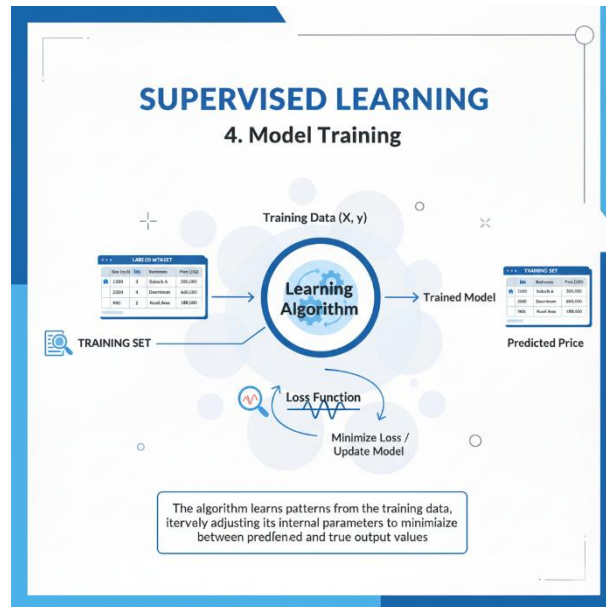
Memilih algoritma pembelajaran mesin yang sesuai (misalnya, Regresi Linear, Support Vector Machine, Pohon Keputusan, Neural Network) tergantung pada jenis masalah yang ingin dipecahkan (regresi atau klasifikasi).



Gambar 3 Supervised Learning Model Selection

#### 4. Pelatihan Model

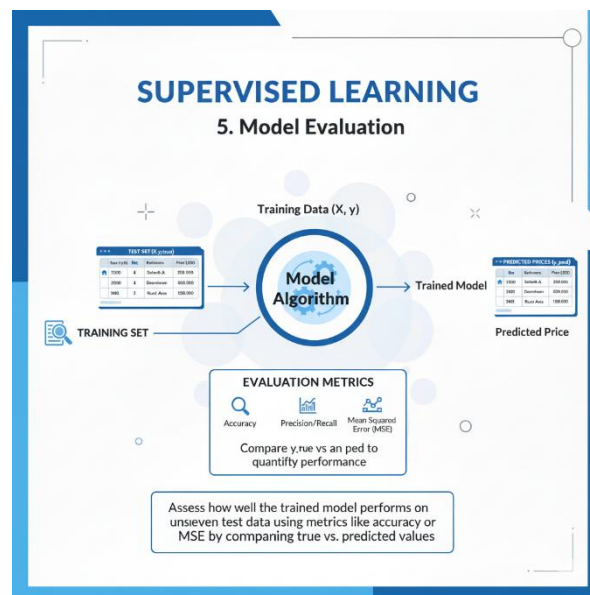
Model dilatih menggunakan training set. Selama pelatihan, model belajar pola dan hubungan antara input dan output dengan meminimalkan fungsi kerugian (loss function) yang mengukur seberapa jauh prediksi model dari label sebenarnya.



Gambar 4 Supervised Learning Model Training

#### 5. Evaluasi Model

Setelah dilatih, model dievaluasi menggunakan test set untuk mengukur seberapa baik kinerjanya pada data yang tidak terlihat. Metrik evaluasi yang umum meliputi akurasi, presisi, recall, F1-score untuk klasifikasi, dan Mean Squared Error (MSE) atau R-squared untuk regresi.



Gambar 5 Supervised Learning Model Evaluation

## 1.2 Unsupervised Learning

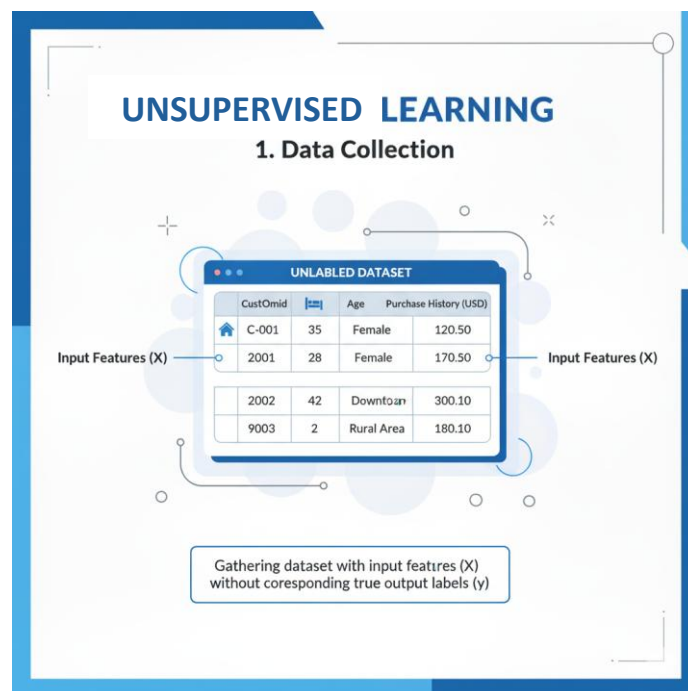
Unsupervised learning atau pembelajaran tidak terawasi merupakan pendekatan dalam machine learning yang digunakan ketika data yang tersedia tidak memiliki label atau target output. Berbeda dengan supervised learning yang memiliki "guru" berupa label data, unsupervised learning harus menemukan pola dan struktur tersembunyi secara mandiri dari data yang diberikan. Pendekatan ini menyerupai cara manusia belajar secara alamiah melalui observasi dan pengelompokan pola tanpa bimbingan eksplisit (Santoso, Abijono, & Anggreini, 2021).

Tujuan utama dari unsupervised learning adalah untuk mengeksplorasi struktur dasar data, mengidentifikasi pola yang tidak terlihat, serta mengelompokkan atau mereduksi kompleksitas data tanpa panduan sebelumnya. Pendekatan ini sangat berharga ketika kita tidak mengetahui dengan pasti apa yang harus dicari dalam data, atau ketika kita ingin melakukan eksplorasi data secara menyeluruh.

### Tahapan Utama dalam Unsupervised Learning:

#### 1. Pengumpulan Data Tanpa Label

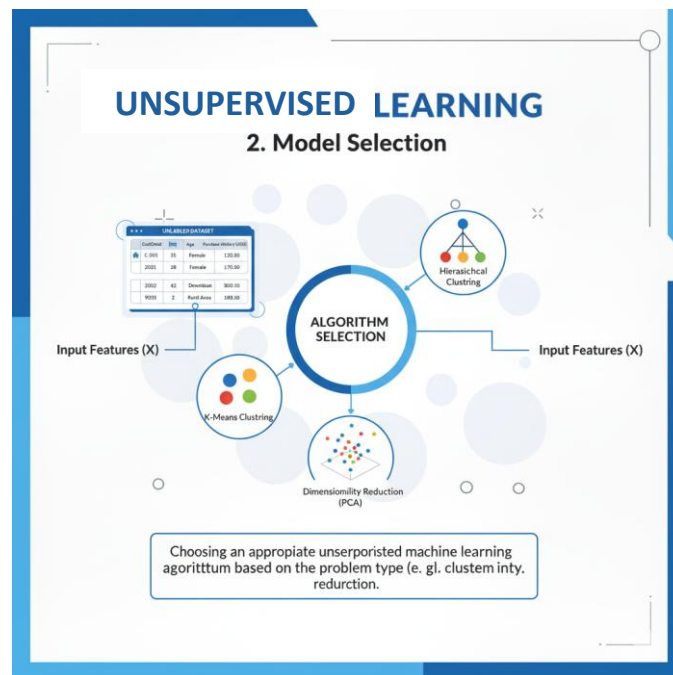
Tahap awal adalah mengumpulkan dataset yang hanya berisi fitur input tanpa label output yang sesuai. Misalnya, Anda memiliki data pelanggan dan ingin mengelompokkan mereka tanpa tahu sebelumnya kelompok apa yang ada.



Gambar 6 Unsupervised Learning Data Collection

#### 2. Pemilihan Model

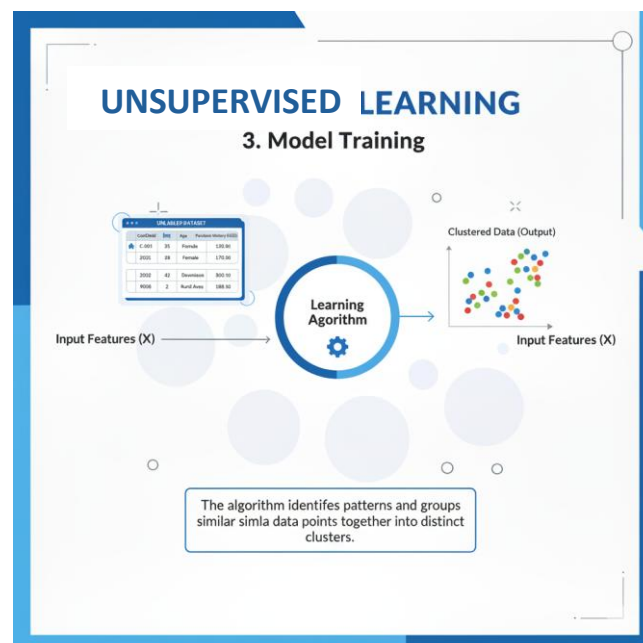
Memilih algoritma pembelajaran tanpa terawasi yang sesuai. Algoritma umum meliputi K-Means, Hierarchical Clustering untuk clustering, dan Principal Component Analysis (PCA) untuk pengurangan dimensi.



Gambar 7 Unsupervised Learning Model Selection

### 3. Pelatihan Model

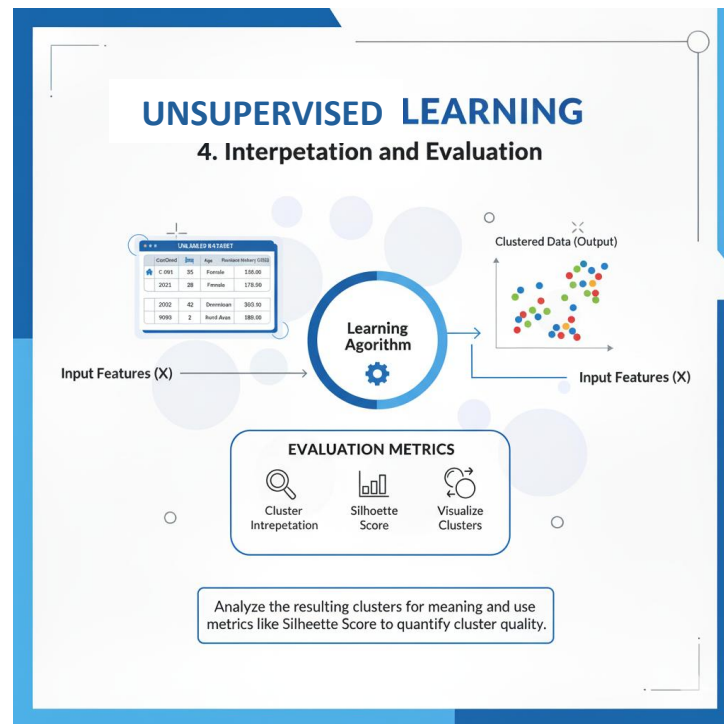
Model dilatih menggunakan data tanpa label. Tujuannya adalah untuk menemukan pola atau struktur intrinsik dalam data. Misalnya, dalam clustering, algoritma akan mengelompokkan titik data yang serupa bersama-sama.



Gambar 8 Unsupervised Learning Model Training

#### 4. Interpretasi dan Evaluasi

Karena tidak ada label yang benar, evaluasi dalam unsupervised learning seringkali lebih subjektif dan melibatkan interpretasi hasil. Misalnya, dalam clustering, Anda perlu menganalisis karakteristik setiap kluster yang ditemukan untuk memahami apa artinya. Metrik seperti Silhouette Score dapat digunakan untuk mengukur kualitas kluster.



Gambar 9 Unsupervised Learning Interpretation and Evaluation

### 1.2 Reinforcement Learning

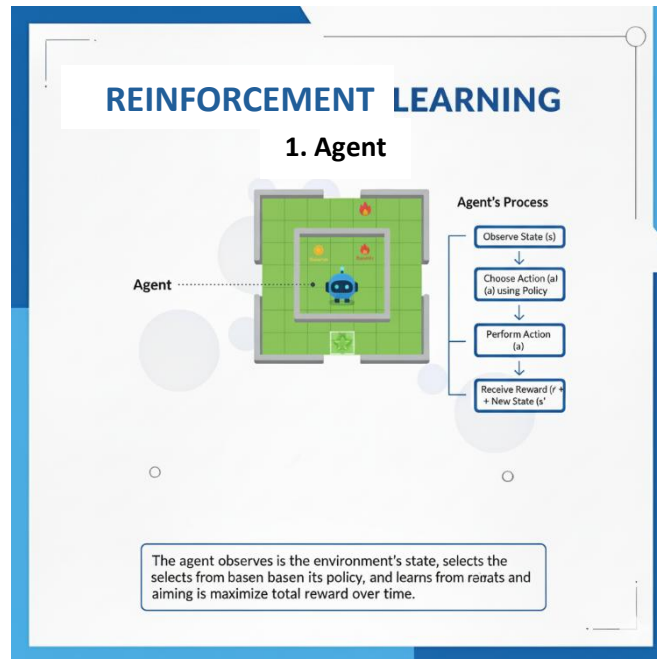
*Reinforcement Learning* (RL) adalah cabang dari *machine learning* di mana sebuah **agen (agent)** belajar untuk mengambil **tindakan (action)** di dalam suatu **lingkungan (environment)** untuk memaksimalkan **imbalan (reward)** yang diterimanya. Agen belajar dari coba-coba (*trial and error*), mirip seperti cara manusia atau hewan belajar dari pengalaman.

Tujuan utamanya adalah menemukan kebijakan (*policy*) terbaik, yaitu strategi yang memberi tahu agen tindakan apa yang harus diambil pada setiap keadaan (*state*) untuk mendapatkan imbalan kumulatif tertinggi (Vaish, U.D., Tewari, & S.M., 2021).

## Tahapan Utama dalam Reinforcement Learning:

### 1. Agen

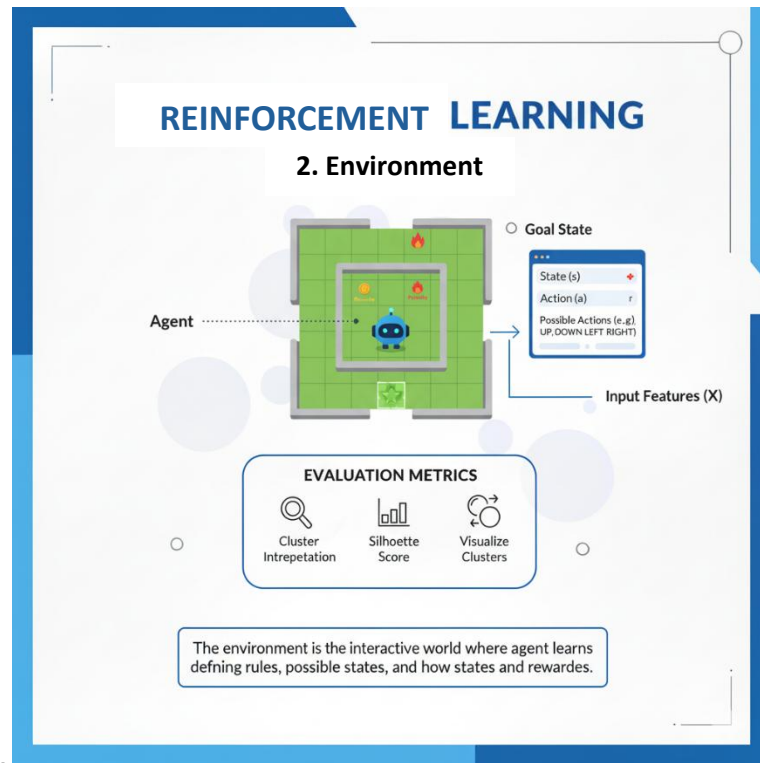
Agen adalah entitas yang belajar atau pembuat keputusan. Ia mengamati lingkungan dan memilih tindakan berdasarkan pengamatannya untuk mencapai tujuan tertentu.



### 2. Lingkungan (Environment)

Lingkungan adalah dunia atau konteks di mana agen beroperasi. Agen berinteraksi dengan lingkungan, dan lingkungan merespons tindakan agen dengan memberikan imbalan dan mengubah keadaannya. Dalam permainan catur, lingkungan adalah papan catur dan lawannya

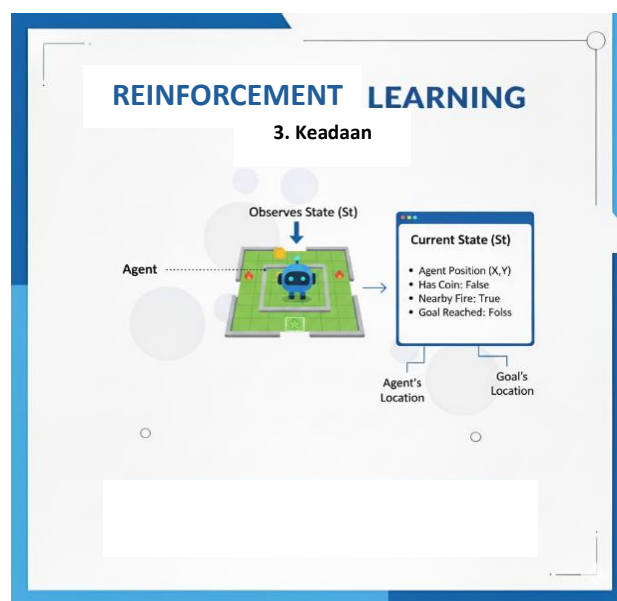




Gambar 10 Reinforcement Learning Environment

### 3. Keadaan (State)

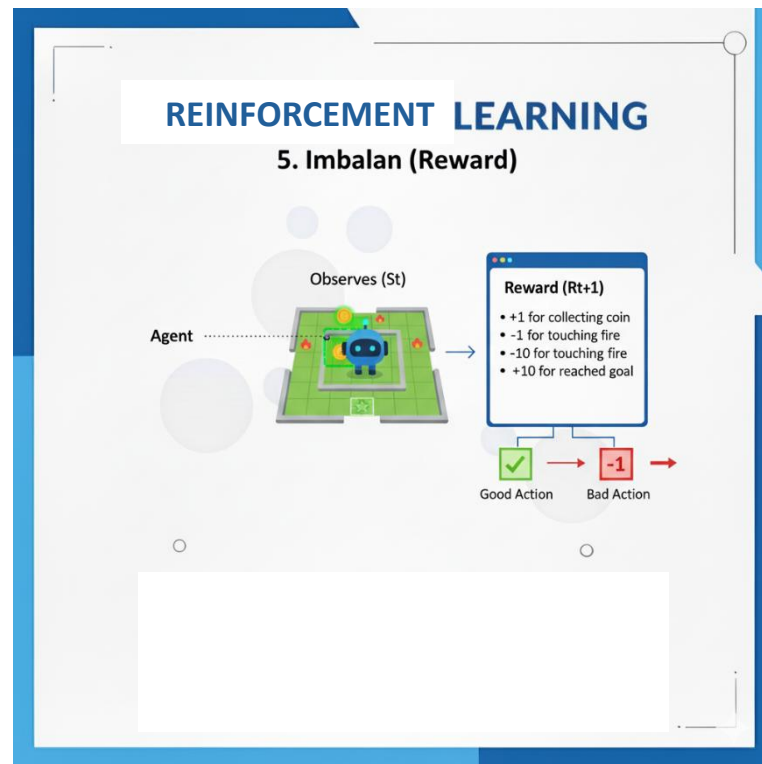
Keadaan ( $S_t$ ) adalah representasi dari lingkungan pada satu waktu tertentu yang diamati oleh agen. Misalnya, dalam permainan *video game*, keadaan bisa berupa posisi karakter, musuh, dan item di layar. Agen menggunakan keadaan ini untuk memutuskan tindakan selanjutnya.



Gambar 11 Reinforcement Learning State

#### 4. Imbalan (Reward)

Imbalan ( $R_{t+1}$ ) adalah sinyal umpan balik yang dikirim oleh lingkungan kepada agen setelah agen melakukan suatu tindakan. Imbalan bisa positif (penghargaan) atau negatif (hukuman). Sinyal inilah yang memandu agen untuk belajar; tujuannya adalah memaksimalkan total imbalan dari waktu ke waktu.



Gambar 12 Reinforcement Learning Reward

## 2. Perbandingan Pendekatan Utama dalam Machine Learning

Dalam *machine learning*, pendekatan yang digunakan sangat bergantung pada tujuan analisis data. Tabel berikut menyajikan perbedaan mendasar antara empat pendekatan utama: **Regression, Classification, Clustering, dan Association**, beserta aplikasi khas untuk masing-masing.

Karakteristik	Regression	Classification	Clustering	Association
<b>Tujuan Utama</b>	Memprediksi nilai yang berkelanjutan (kontinu).	Memberi label atau mengkategorikan data ke dalam kelas tertentu.	Mengelompokkan data yang serupa ke dalam grup.	Menemukan hubungan atau pola antar item dalam data.
<b>Output</b>	Angka/Nilai numerik	Kategori/Kelas	Grup/Cluster	Aturan (misal: Jika A, maka B)

Contoh Aplikasi	Prediksi harga properti.	Klasifikasi teks dan gambar.	Segmentasi pelanggan.	Analisis keranjang belanja ( <i>Market basket analysis</i> ).
Contoh Algoritma	Random Forest, XGBoost	SVM ( <i>Support Vector Machine</i> )	K-Means	Apriori

Table 1 Perbandingan Pendekatan Utama dalam Machine Learning

### 3. Studi Kasus dan Solusi Algoritma Machine Learning

Machine learning telah terbukti efektif dalam menyelesaikan berbagai masalah kompleks di dunia nyata. Studi kasus berikut menunjukkan bagaimana algoritma yang tepat, ketika diterapkan dengan baik, dapat memberikan solusi inovatif across berbagai industri.

Pemilihan algoritma dalam setiap kasus didasarkan pada karakteristik data, jenis masalah, dan tujuan yang ingin dicapai. Keberagaman studi kasus ini mencakup empat pendekatan utama: regresi untuk prediksi nilai kontinu, klasifikasi untuk kategorisasi data, klastering untuk pengelompokan tanpa label, dan asosiasi untuk menemukan pola hubungan. Berikut adalah 18 studi kasus nyata yang diselesaikan dengan algoritma machine learning.

#### 3.1 Studi Kasus Regression

1. **Prediksi Harga Murah**

Algoritma yang digunakan **Regresi Linier** untuk memprediksi harga rumah berdasarkan fitur seperti luas, lokasi, dan jumlah kamar. Model ini mampu memetakan hubungan linear antara karakteristik properti dengan harga pasar secara efektif. Implementasinya membantu dalam estimasi harga properti yang akurat dan kompetitif (Uyanik & Guler, 2013).

2. **Peramalan Cuaca**

Algoritma yang digunakan **ARIMA** untuk memprediksi parameter cuaca seperti suhu dan curah hujan berdasarkan data time series historis. Model ini efektif dalam menangkap pola musiman dan tren dalam data iklim (Misshuari, Kurniyaningrum, & Saily, 2023).

3. **Estimasi Produksi Panen**

Algoritma yang digunakan **Random Forest Regression** untuk memperkirakan hasil panen berdasarkan data iklim, tanah, dan karakteristik tanaman. Algoritma ini mampu menangani hubungan non-linear dengan robust (Salami, Babamuratov, Khudayberganov, & Singh, 2025).

4. **Prediksi Kebutuhan Energi**

Algoritma yang digunakan **Support Vector Regression (SVR)** berhasil memprediksi konsumsi energi listrik harian berdasarkan pola penggunaan dan faktor eksternal. Model ini efektif dalam ruang fitur berdimensi tinggi (Athaya, Raharjo, & Rizal, 2023).

5. **Peramalan Penjualan**

Algoritma yang digunakan **XGBoost Regression** untuk memperkirakan volume penjualan berdasarkan data historis dan variabel pemasaran. Algoritma ini menunjukkan akurasi tinggi dalam forecasting time series (Gustiyandi, Ardea, & Sulastri, 2025).

### ***3.2 Studi Kasus Classification***

#### **1. Deteksi Spam Email**

Algoritma yang digunakan **Naïve Bayes** untuk mengklasifikasikan email sebagai spam atau non-spam berdasarkan konten teks dan metadata. Algoritma ini efisien untuk pemrosesan data teks dalam volume besar (Ghani & Sulaiman, 2023).

#### **2. Diagnosis Penyakit Jantung**

Algoritma yang digunakan **Decision Tree** untuk mengidentifikasi pasien berisiko penyakit jantung berdasarkan rekam medis dan faktor klinis. Model ini memberikan interpretasi yang transparan untuk diagnosis (Yosrita & Cahyaningtyas, 2021).

#### **3. Klasifikasi Sentimen**

Algoritma yang digunakan **Logistic Regression** berhasil mengklasifikasikan ulasan produk sebagai positif atau negatif berdasarkan analisis teks. Algoritma ini efektif untuk masalah klasifikasi biner pada data teks (Burrhanuddin & Rozi, 2025).

#### **4. Pengenalan Tulisan Tangan**

Algoritma yang digunakan **Convolutional Neural Network (CNN)** untuk mengenali digit tulisan tangan pada dataset MNIST. Arsitektur deep learning ini optimal untuk ekstraksi fitur visual (Ameng, Adytia, & Khair, 2024).

#### **5. Identifikasi Jenis Tanaman**

Algoritma yang digunakan **Random Forest** berhasil mengklasifikasikan spesies tanaman berdasarkan citra daun dan karakteristik morfologi. Algoritma ini robust terhadap variasi dalam data citra (Wibowo, 2021).

### ***3.3 Studi Kasus Clustering***

#### **1. Segmentasi Pelanggan**

Algoritma yang digunakan **K-Means clustering** untuk mengelompokkan pelanggan berdasarkan perilaku pembelian dan demografi. Algoritma ini efektif untuk segmentasi pasar yang targeted (Fadhillah, Suyoso, & Puspitasari, 2025).

#### **2. Pengelompokan Berita**

Algoritma yang digunakan **Hierarchical Clustering** untuk mengelompokkan artikel berita berdasarkan kesamaan topik dan konten. Pendekatan ini memungkinkan analisis hierarkis dokumen teks (Arnelawati, Rini, & Ermatita, 2025).

#### **3. Analisis Genetik**

Algoritma yang digunakan **DBSCAN** untuk mengidentifikasi kelompok gen dengan ekspresi serupa berdasarkan data microarray. Algoritma ini mampu menangani noise dalam data biologis (Sari & Primajaya, 2019).

#### **4. Segmentasi Pasar Real Estate**

Algoritma yang digunakan **Gaussian Mixture Model (GMM)** berhasil membagi wilayah berdasarkan harga properti dan fasilitas. Pendekatan probabilistic ini fleksibel untuk berbagai bentuk kluster (Raditya, Indwiarti, & Rohmawati, 2022).

## 5. Pengelompokan Gangguan Jaringan

Algoritma yang digunakan **Mean Shift clustering** untuk mendeteksi anomali dalam jaringan komputer berdasarkan pola traffic. Algoritma ini tidak memerlukan jumlah kluster predefined (Utari, Fitri, Agustian, & Afriyanti, 2024).

### 3.4 Studi Kasus Association

#### 1. Analisis Market Basket

Algoritma yang digunakan **Algoritma Apriori** untuk menemukan produk yang sering dibeli bersamaan dalam transaksi retail. Teknik ini optimal untuk mining association rules (Brighton & Hariyanto, 2024).

#### 2. Rekomendasi Film

Algoritma yang digunakan **FP-Growth** untuk merekomendasikan film berdasarkan riwayat penonton dan preferensi. Algoritma ini efisien untuk frequent pattern mining tanpa candidate generation (Firdaus & Rahmatulloh, 2025).

#### 3. Asosiasi Gejala Penyakit

Algoritma yang digunakan **Algoritma Eclat** berhasil menemukan hubungan antara gejala medis berdasarkan data rekam medis. Pendekatan ini cepat untuk vertical data format (Mardatillah, Nazir, Fikry, Haerani, & Syafria, 2020).

## Referensi

- Ameng, A. I., Adytia, P., & Khair, A. A. (2024). Sistem Pengenalan Tulisan Tangan (Handwriting Recognition) menggunakan. *Sebatik*, 240-280. <https://doi.org/10.46984/sebatik.v28i2.0000>
- Arnelawati, Rini, D. P., & Ermatita. (2025). Penerapan Hierarchical Agglomerative Clustering Untuk Penentuan Faktor Penyebab Ketidaktuntasan Belajar Matematika. *JURNAL SISTEM INFORMASI*, 1-14. <https://doi.org/10.18495/jsi.v17i1.200>
- Athaya, K. S., Raharjo, J., & Rizal, S. (2023). Support Vector Regression untuk Prediksi Beban Listrik Jangka. *Journal of Multidisciplinary Research and Development*, 247-253. <https://doi.org/10.38035/rrj.v5i4>
- Brighton, K., & Hariyanto, S. (2024). Penerapan Metode Market Basket Analisis Dengan Algoritma Apriori Pada Toko Ritel Elektronik. *JOURNAL Bit-Tech*, 38-46. <https://doi.org/10.32877/bt.v7i1.1417>

- Burhanuddin, N. I., & Rozi, A. F. (2025). SENTIMENT ANALYSIS AND CLASSIFICATION OF USER. *JURNAL INOVTEK POLBENG - SERI INFORMATIKA*, 877-887.
- Fadhillah, M. F., Suyoso, A. L., & Puspitasari, I. (2025). Segmentasi Pelanggan dengan Algoritma Clustering Berdasarkan Atribut Recency, Frequency dan Monetary (RFM). *Indonesian Journal of Machine Learning and Computer Science*, 48-56.  
<https://doi.org/10.57152/malcom.v5i1.1491>
- Firdaus, M. A., & Rahmatulloh, A. (2025). IMPLEMENTASI STEGANOGRAFI CITRA DIGITAL LSB MENGGUNAKAN ENKRIPSI AES-256 DAN EMBEDDINGPSEUDORANDOM. *Jurnal Informatika dan Teknik Elektro Terapan*, 411-418. <https://doi.org/10.23960/jitet.v13i1.5621>
- Ghani, M. A., & Sulaiman, H. (2023). Deteksi Spam Email dengan Metode Naive Bayes dan Particle Swarm Optimization (PSO). *INFOTEK*, 11-20. <https://doi.org/10.29408/jit.v6i1.7049>
- Gustiandi, Z., Ardea, & Sulastri, H. (2025). Prediksi Penjualan Produk Menggunakan Algoritma Xtreme. *JURTI*, 20-29. <http://dx.doi.org/10.30872/jurti.v9i1>
- Mardatillah, P., Nazir, A., Fikry, M., Haerani, E., & Syafria, F. (2020). Penerapan Algoritma Equivalence Class Transformation (Eclat) Dalam Pencarian Adverse Event Obat Diphenhydramine. *Jurnal RESTIKOM : Riset Teknik Informatika dan Komputer*, 143-155.  
<https://doi.org/10.52005/RESTIKOM.V2I3.74>
- Misshuari, I. W., Kurniyaningrum, E., & Saily, R. (2023). STUDI PERAMALAN CURAH HUJAN MENGGUNAKAN. *Construction Engineering and Sustainable Development*, 22-28.  
<https://doi.org/10.25105/cesd.v6i2.18815>
- Raditya, M. H., Indwiarti, & Rohmawati, A. A. (2022). House Prices Segmentation Using Gaussian Mixture Model-Based Clustering. *Jurnal Resti Rekayasa Sistem dan Teknologi Informasi*, 866-871. <https://doi.org/10.29207/resti.v6i5.4459>
- Salami, Z. A., Babamuratov, B., Khudayberganov, K., & Singh, C. (2025). Modelling Crop Yield Prediction with Random Forest and Remote Sensing Data. *Natural and Engineering Sciences*, 67-78. <https://doi.org/10.28978/nesciences.1763843>
- Sari, B. N., & Primajaya, A. (2019). PENERAPAN CLUSTERING DBSCAN UNTUK PERTANIAN PADI DI KABUPATEN KARAWANG. *JIKO : Jurnal Informatika dan Komputer*, 28-34.  
<http://dx.doi.org/10.26798/jiko.v4i1.178>
- Utari, R. F., Fitri, I., Agustian, S., & Afriyanti, L. (2024). Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mean Shift. *Indonesian Journal of Machine Learning and Computer Science*, 1015-1023. <https://doi.org/10.57152/malcom.v4i3.1428>
- Uyanik, G. K., & Guler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, 234-240.
- Wibowo, A. P. (2021). Komparasi Algoritma Klasifikasi untuk Penentuan Jenis Spesies Tanaman Hutan. *JURNAL CAKRAWALA INFORMASI*, 12-19. <https://doi.org/10.54066/jci.v1i1.209>

Yosrita, E., & Cahyaningtyas, R. (2021). Sistem Diagnosis Penyakit Jantung Koroner Dengan Menggunakan. *Jurnal Pengkajian dan Penerapan Teknik Informatika*, 214-222.  
<https://doi.org/10.33322/petir.v14i2.1388>

Santoso, P., Abijono, H., & Anggreini, N. L. (2021). Algoritma Supervised Learning dan Unsupervised Learning dalam Pengolahan Data. *G-Tech*, 315-318.

Shaveta. (2023). A review on machine learning . *International Journal of Science and Research Archive*, 281-285.

Vaish, R., U.D., D., Tewari, S., & S.M., T. (2021). Machine learning applications in power system fault diagnosis: Research advancements and perspectives. *Engineering Applications of Artificial Intelligence*, 1-20.

Vaish, R., U.D., D., Tewari, S., & S.M., T. (2021). Machine learning applications in power system fault diagnosis: Research advancements and perspectives. *Engineering Applications of Artificial Intelligence*, 1-20. <https://doi.org/10.1016/j.engappai.2021.104504>