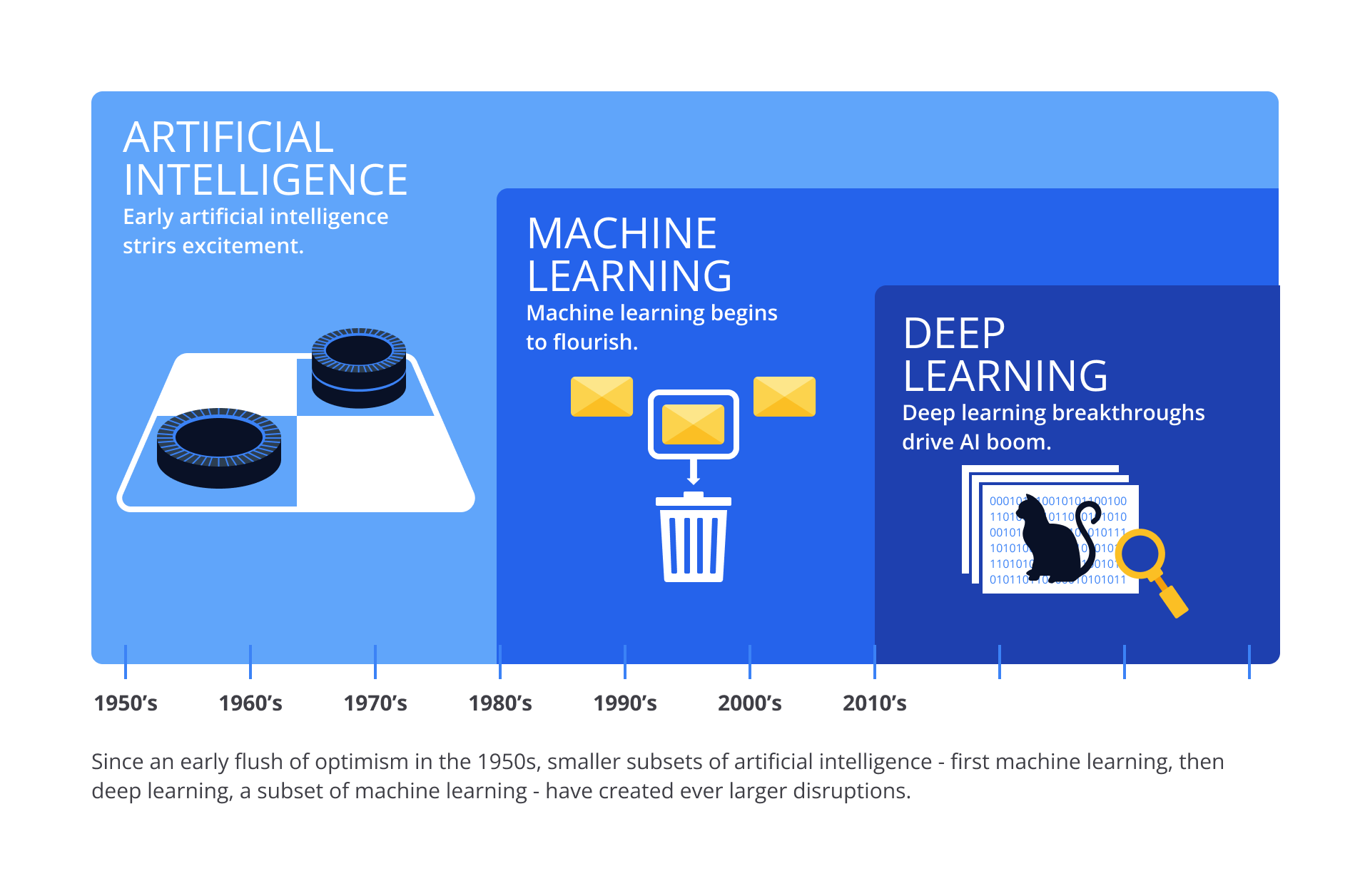
**Rangkuman Machine Learning untuk Data Science**

Menurut AWS (*Amazon Web Services*) bahwa machine learning adalah ilmu pengembangan algoritma dan model secara statistik yang digunakan sistem komputer untuk menjalankan tugas tanpa instruksi eksplisit dan mengandalkan pola serta inferensi sebagai gantinya.

Sederhananya, machine learning adalah ilmu pengembangan algoritma yang memanfaatkan konsep matematis dan statistik dalam menjalankan tugas tertentu tanpa instruksi eksplisit. Pada prosesnya, machine learning akan berusaha mengenal pola yang terdapat dalam sebuah data serta menggunakannya untuk menghasilkan prediksi.

**Scope Machine Learning**

[](https://www.dicoding.com/academies/615/tutorials/33213)

* ***Artificial Intelligence*(AI)**  
  Terlihat jelas pada gambar di atas bahwa eksistensi AI dimulai dari tahun 1950-an. AI menjadi langkah awal lahirnya machine learning dan deep learning di masa sekarang. Namun, sebenarnya apa sih istilah AI itu sendiri? AI adalah bidang ilmu komputer yang dikhususkan untuk memecahkan masalah kognitif yang umumnya terkait dengan kecerdasan manusia, seperti pembelajaran, pemecahan masalah, dan pengenalan pola.
* ***Machine Learning*(ML)**  
  Keberadaan machine learning sekitar tahun 1980an yang berarti terjadi gap selama 30 tahun setelah konsep AI dikenal. Lalu, apa yang dilakukan oleh ML? Pada dasarnya ML adalah subset AI yang membuat sistem agar mampu mengadaptasi kemampuan manusia untuk belajar.
* ***Deep Learning*(DL)**  
  Terhitung masih belia, ketika kelas ini ditulis, DL baru menginjak 13 tahun karena ia lahir sekitar tahun 2010-an. Namun, algoritma*Artificial Neural Network* yang merupakan bagian dari deep learning telah dikenal sejak lama, *lho.*Kira-kira apa definisi dari deep learning, ya?Deep learning adalah metode dalam AI yang mengajarkan komputer untuk memproses data dengan cara yang terinspirasi otak manusia. Model deep learning dapat mengenali pola kompleks dalam gambar, teks, suara, dan data lain untuk menghasilkan wawasan dan prediksi yang akurat.

**Faktor Popularitas Machine Learning**

1. *Mature Field*  
   Bidang machine learning telah banyak mengalami perubahan dan perkembangan selama beberapa dekade terakhir. Awalnya, machine learning tumbuh dari bidang artificial intelligence dan merupakan kumpulan metode yang dipelajari dari data atau pengalaman sebelumnya. Seiring berkembangnya zaman, machine learning berfokus pada metode statistik dan probabilitas yang dipadupadankan dengan data dan pengalaman yang telah dipelajari sebelumnya.
2. *Abundant Data*  
   Jumlah data yang tersimpan setiap harinya semakin bertambah. Keberadaan machine learning menjadi penyelamat untuk keberlimpahan data tersebut karena mampu membantu mengolah data yang besar dan kompleks menjadi sebuah informasi yang bermakna untuk mendorong pengambilan keputusan.
3. *Abundant Computation*  
   Machine learning populer karena sumberdaya komputasi yang berlimpah dan murah. Hal ini memungkinkan kita mengolah data yang berjumlah sangat besar dan melakukan eksperimen dengan algoritma machine learning yang lebih kompleks.

**Pentingnya Machine Learning**

Data adalah sumber kehidupan dari semua bidang terlebih lagi untuk sebuah bisnis. Keputusan berdasarkan data semakin membuat perbedaan antara mengikuti persaingan atau tertinggal lebih jauh. Keberadaan machine learning ini menjadi kunci untuk membuka nilai data perusahaan dan *customer*serta membuat keputusan yang membuat perusahaan tetap terdepan dalam persaingan. Terlebih lagi, machine learning dianggap begitu penting karena memberikan perusahaan sebuah pandangan tentang tren perilaku pelanggan (*Customer Behaviour*) dan pola operasional bisnis, serta mendukung pengembangan produk baru. Di bawah ini merupakan hasil dari penerapan machine learning di industri.

* Google Translate
* Gmail
* Google Photos

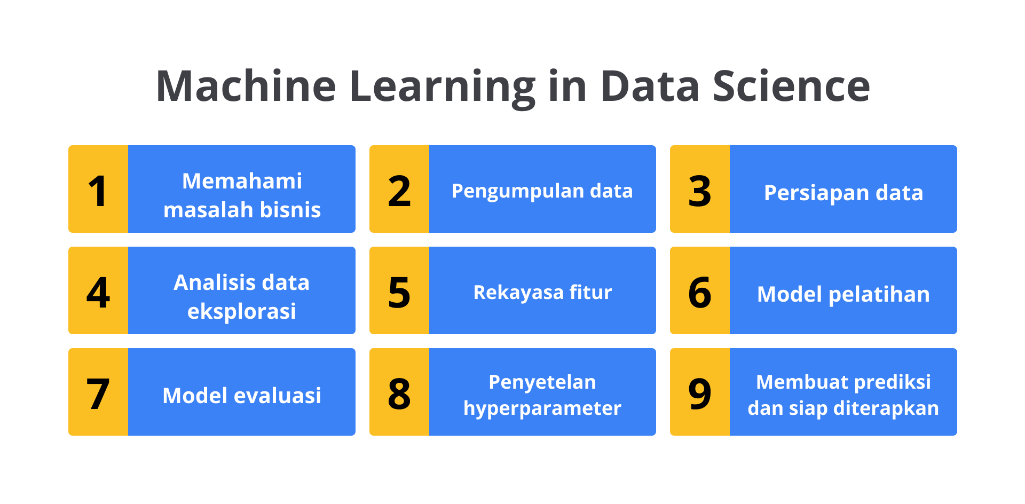
**Peran Machine Learning di Data Science**

Di bawah ini merupakan perbedaan antara machine learning dan data science yang perlu Anda ketahui.

| **Machine Learning** | **Data Science** |
| --- | --- |
| Machine learning adalah bidang artificial intelligence (AI) yang memungkinkan *software*belajar dari data untuk mengidentifikasi pola dan membuat prediksi secara otomatis dengan campur tangan manusia yang minimal. | Data Science adalah semua tentang proses dan sistem untuk mengekstrak insight dari data terstruktur dan semi terstruktur. |
| Machine learning memanfaatkan pengalaman masa lalu untuk mempelajari tentang data. | Data Science berurusan dengan data, baik masa lalu maupun *real-time.* |
| Machine learning terdiri dari tiga jenis: Supervised learning, Unsupervised learning, dan Reinforcement learning. | Data gathering, manipulasi, data cleaning, dll adalah operasi data science. |

**Alur Kerja Machine Learning Dalam Data Science**

Alur machine learning dalam data science umumnya terbagi menjadi 9 tahap seperti yang tertera pada gambar di bawah ini.

[](https://www.dicoding.com/academies/615/tutorials/33213)

1. **Memahami masalah bisnis**  
   Untuk membangun model bisnis yang sukses, sangat penting untuk memahami masalah bisnis yang dihadapi klien. Misalnya, destinasi wisata di Bali yang tidak merata, sehingga perlu peninjauan khusus terkait peningkatan kualitas destinasi wisata yang memiliki *rating* berada di bawah rata-rata.
2. **Pengumpulan data**  
   Setelah memahami pernyataan masalah, Anda harus mengumpulkan data yang relevan. Sesuai masalah bisnis, data dikategorikan berdasarkan data terstruktur, tidak terstruktur, dan semi-terstruktur dari basis data apa pun di seluruh sistem.
3. **Persiapan data**  
   Langkah pertama persiapan data adalah melakukan pembersihan data. Ini adalah langkah yang penting untuk dilakukan. Dalam persiapan data, Anda menghilangkan duplikat dan nilai null, tipe data yang tidak konsisten, entri yang tidak valid, data yang hilang, dan pemformatan yang tidak tepat.
4. **Exploratory Data Analysis (EDA)**  
   Exploratory Data Analysis memungkinkan Anda untuk mengungkap wawasan berharga yang akan berguna di fase berikutnya dari siklus hidup data science.
5. **Rekayasa fitur (Fitur Engineering)**  
   Rekayasa fitur adalah salah satu langkah penting dalam proyek data science. Ini membantu dalam membuat fitur baru, mengubah dan menskalakan fitur. Dalam domain ini, keahlian memainkan peran kunci (*key role*) dalam menghasilkan wawasan baru dari langkah eksplorasi data.
6. **Pelatihan Model**  
   Dalam pelatihan model, terdapat tahap menyesuaikan data pelatihan; di sinilah proses “belajar” dimulai. Pada tahap ini kita melatih model machine learning menggunakan data latih.
7. **Evaluasi Model**  
   Setelah pelatihan model selesai, saatnya untuk mengevaluasi kinerjanya. Jadi, mengevaluasi model pada kumpulan data baru akan memberi Anda gambaran tentang bagaimana kinerja model Anda di data mendatang.
8. **Penyetelan Hyperparameter**  
   Setelah model dilatih dan dievaluasi, kinerja model dapat ditingkatkan lagi dengan menyetel hyperparameternya. Penyesuaian hyperparameter model penting untuk meningkatkan kinerja model secara keseluruhan.
9. **Membuat prediksi dan siap diterapkan**  
   Ini adalah tahap akhir dari machine learning. Di sini, model digunakan untuk menjawab setiap pertanyaan Anda berdasarkan hasil pelatihannya. Setelah membuat prediksi yang akurat, model machine learning diterapkan ke dalam produksi.

**Tipe-tipe Machine Learning**

Secara umum, model machine learning dapat dibedakan berdasarkan penggunaannya, seperti *supervised, unsupervised,*dan *reinforcement.*Ketiganya merupakan istilah yang digunakan untuk memisahkan model dalam fungsi tertentu. Ingin tahu lebih lanjut? Yuk, kita simak materi berikutnya.

**Supervised Learning**

Machine learning supervised adalah pendekatan machine learning yang ditentukan oleh penggunaan kumpulan data yang berlabel.

Berikut merupakan fungsi dari supervised machine learning.

* Mengklasifikasikan berbagai jenis file, seperti gambar, dokumen, atau kata-kata tertulis.
* Memprediksi nilai dari data kontinu.
* Meramal tren dan hasil masa depan melalui pola pembelajaran dalam data penelitian.

**Unsupervised Learning**

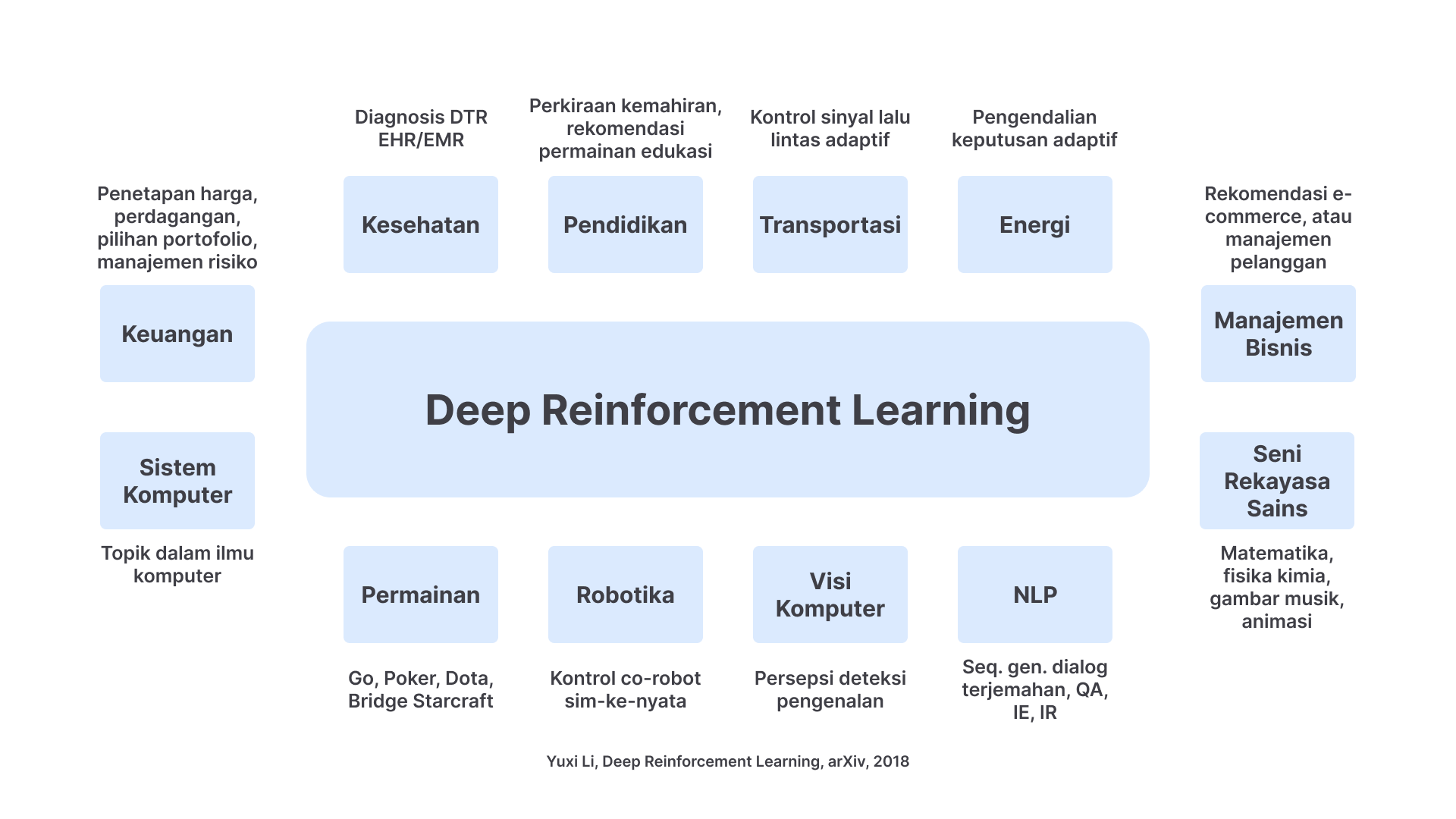
Bertolak belakang dengan supervised machine learning, tipe ini adalah pelatihan model pada data tidak berlabel.

Berikut merupakan fungsi dari unsupervised machine learning.

* Mengklasifikasikan dataset pada kesamaan antara fitur atau segmen pada data.
* Memahami hubungan antara berbagai titik data
* Melakukan analisis data awal.

**Reinforcement Learning**

Reinforcement machine learning adalah salah satu dari tiga paradigma machine learning disamping supervised dan unsupervised.

[](https://www.dicoding.com/academies/615/tutorials/33213)

**Algoritma Machine Learning Tradisional**

Di bawah ini merupakan bagian dari machine learning tradisional, yaitu *Linear Regression* dan *Logistic Regression.*

**Linear regression vs. Logistic regression**

| **Linear Regression** | **Logistic Regression** |
| --- | --- |
| Regresi linier digunakan untuk memprediksi variabel dependen kontinu menggunakan sekumpulan variabel independen tertentu. | Regresi logistik digunakan untuk memprediksi variabel dependen kategoris menggunakan seperangkat variabel independen yang diberikan. |
| Regresi linier digunakan untuk menyelesaikan masalah regresi. | Digunakan untuk memecahkan masalah klasifikasi. |
| Memprediksi nilai variabel kontinu. | Memprediksi nilai variabel kategoris. |
| Fit line (garis lurus). | Menggunakan s-curve. |
| Keluaran berupa nilai kontinu, seperti harga, umur, dll. | Keluaran berupa nilai kategoris. seperti 0 atau 1, Ya atau tidak, dll. |

**Algoritma Deep Learning**

Di bawah ini merupakan serba-serbi deep learning yang perlu Anda ketahui mulai dari definisi deep learning hingga berkenalan dengan neural network.

**Berkenalan dengan Deep Learning**

Deep learning adalah metode dalam artificial intelligence yang mengajarkan komputer untuk memproses data layaknya otak manusia ketika berpikir. Model *deep learning* dapat mengenali pola kompleks dalam gambar, teks, suara, dan data lain untuk menghasilkan wawasan serta prediksi akurat.

**Mengapa Perlu Deep Learning**

Dengan adanya deep learning ini sangat bermanfaat bagi para data scientist yang bertugas mengumpulkan, menganalisis, dan menafsirkan data dalam jumlah besar. Dan juga membuat proses apapun menjadi lebih cepat dan lebih mudah.

* Asisten digital (*chat bot*),
* Deteksi penipuan (*fraud detection*),
* Pengenalan wajah otomatis (*face recognition*), dst.

**Neural Network**

Neural network adalah metode dalam artificial intelligence yang mengajarkan komputer untuk memproses data yang terinspirasi dari cara kerja  otak manusia. Neural network adalah tipe proses machine learning yang termasuk dalam deep learning. Dalam prosesnya, ia menggunakan simpul atau neuron yang saling terhubung dalam struktur berlapis yang menyerupai otak manusia.

*Last but not least,*pada modul berikutnya Anda akan mempelajari cara untuk menembus lowongan pekerjaan mulai dari mencari peluang data scientist, membuat CV, dan sebagainya. *So,*tunggu apa lagi? Yuk! Kita langsung ke modul berikutnya.