**Introduction to Machine Learning**

**What is Artificial Intelligence?**

Seperti namanya, artificial intelligence yang biasa dikenal dengan AI adalah cara membuat komputer menjadi *intelligent*secara *artificially*. Dapatkah kalian memikirkan komputer atau mungkin robot yang dapat melakukan berbagai tugas yang mirip dengan manusia?

*According to the father of Artificial Intelligence, John McCarthy, it is "****The science and engineering of making intelligent machines, especially intelligent computer programs****".*

**Intuition & Motivation behind AI**

Otak manusia, intuisi & motivasi di balik AI berasal dari otak Manusia. Otak manusia adalah alat paling ampuh yang diberikan kepada umat manusia. Manusia adalah spesies paling cerdas di planet bumi. Dan rahasia di balik otak yang kuat ini adalah **Neocortex**.

Seperti yang kita semua ingat, ketika kita berusia 3-4 tahun, orang tua kita biasa membuat kita mempelajari 26 huruf alfabet, yang jika digabungkan menjadi sebuah kata, kata-kata jika digabungkan menjadi sebuah kalimat dan sebagainya. Seiring bertambahnya usia, kita mulai belajar dari lingkungan, seperti saat kita bermain. Jika kita jatuh, otak kita pasti ingat untuk tidak melakukan kesalahan yang sama lagi. Setelah kita mempelajari semua ini, kita kemudian akan dapat memprediksi hasilnya, katakanlah apa yang muncul sebelum Z tetapi setelah X adalah Y, atau 2 jika dikalikan dengan 12 sama dengan 24.

Jadi dengan intuisi yang sama, kita ingin komputer dapat belajar berdasarkan data yang kita masukkan ke dalamnya dan kemudian komputer harus cukup pintar untuk membuat keputusan berdasarkan apa yang telah dipelajari.

**Need for Intelligent Machines?**

Satu alasan utama di balik mengapa kita membutuhkan AI adalah untuk mengotomatiskan tugas-tugas yang menurut orang berlebihan. Ingat komputer jauh lebih cerdas dalam perhitungan ketika harus mengalikan dua atau lebih angka, komputer dapat memberikan hasil dalam hitungan detik, sementara manusia berjuang.  
  
Demikian pula, jika kita membuat komputer mempelajari beberapa data atau records, berdasarkan records, komputer dapat dengan mudah memprediksi hasil masa depan karena akan memahami beberapa pola & struktur dari data atau recordsmasa lalu.

**Machine Learning**

Machine learning adalah penerapan dari artificial intelligence (AI) yang memberikan kemampuan kepada komputer untuk belajar dan berkembang dari pengalaman tanpa harus diprogram secara explicit.

Machine learning dibutuhkan untuk menyelesaikan pekerjaan yang terlalu rumit bagi manusia untuk di-coding secara langsung. Jadi daripada kita membuat coding, kita memberikan data yang sangat banyak ke dalam sebuah machine learning algorithm dan membiarkan algoritma tersebut bekerja dengan mengeksplor data dan mencari sebuah model yang akan menyelesaikan masalah yang programmer punya.

*Traditionally, software engineering combined human created rules with data to create answers to a problem. Instead, machine learning uses data and answers to discover the rules behind a problem. (Chollet, 2017)*  
  
Ada berbagai algoritma Machine Learning seperti

* Decision trees,
* Naive Bayes,
* Random forest
* Support vector machine
* K-nearest neighbor,
* K-means clustering,
* Gaussian mixture model,
* Hidden Markov model etc.

Untuk saat ini, pahamilah bahwa dalam machine learning kita menggunakan salah satu algoritma seperti yang disebutkan di atas yang memberi komputer kemampuan untuk belajar dan memahami secara otomatis tanpa diprogram berkali-kali.

**Now the question is how will the computer learn automatically?**

Nah, jawabannya adalah **data**. Kita memasukkan data yang memiliki atribut atau fitur berbeda yang harus dipahami oleh algoritma dan memberi kita batasan keputusan berdasarkan data yang kita berikan. Setelah algoritma mempelajari dan menafsirkan data, yang berarti telah melatih dirinya sendiri, kita kemudian dapat menempatkan algoritma kita dalam fase pengujian dan tanpa memprogramnya secara eksplisit, masukkan titik data pengujian dan diharapkan memberi kita beberapa hasil.  
  
Contoh: Misalkan kita  harus memprediksi harga rumah, dengan kumpulan data yang terdiri dari biaya rumah dan jumlah kamar di rumah, serta 1000 rumah dengan atribut serupa. Baik harga dan jumlah kamar adalah fitur. Sekarang tujuan kita adalah memasukkan kedua fitur ini ke dalam algoritma. Input kita, dalam hal ini, adalah jumlah kamar, dan algoritma harus memprediksi harga rumah. Algoritma akan mencoba mempelajari hubungan antara jumlah kamar dan biaya rumah.

Sekarang, pada saat tes, kita akan memberikan algoritma katakanlah tiga (jumlah kamar) sebagai input, dan ML akan dapat memprediksi harga rumah secara akurat.

Terminology

* **Dataset**: A set of data examples, that contain features important to solving the problem.
* **Features**: Important pieces of data that help us understand a problem. These are fed in to a Machine Learning algorithm to help it learn.
* **Model**: The representation (internal model) of a phenomenon that a Machine Learning algorithm has learnt. It learns this from the data it is shown during training. The model is the output you get after training an algorithm. For example, a decision tree algorithm would be trained and produce a decision tree model.

Process

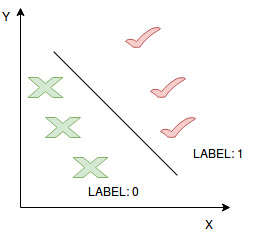
1. **Data Collection:** Collect the data that the algorithm will learn from.
2. **Data Preparation:** Format and engineer the data into the optimal format, extracting important features and performing dimensionality reduction.
3. **Training:** Also known as the fitting stage, this is where the Machine Learning algorithm actually learns by showing it the data that has been collected and prepared.
4. **Evaluation:** Test the model to see how well it performs.
5. **Tuning:** Fine tune the model to maximise it’s performance.

**Machine Learning Approaches**

Ada banyak pendekatan yang dapat diambil saat melakukan Machine Learning. Mereka biasanya dikelompokkan ke dalam area yang seperti di bawah ini. Supervised dan Unsupervised adalah pendekatan mapan dan paling umum digunakan.  Reinforcement Learning lebih baru dan lebih kompleks tetapi telah menunjukkan hasil yang mengesankan.  
  
**No Free Lunch theorem** terkenal dalam Machine Learning. **No Free Lunch theorem**menyatakan bahwa tidak ada algoritma tunggal yang akan bekerja dengan baik untuk semua tugas. Setiap tugas yang kita coba selesaikan memiliki keistimewaannya sendiri. Oleh karena itu, ada banyak algoritma dan pendekatan yang sesuai dengan setiap masalah individu.

**Supervised Learning**

Supervised sesuai namanya, merupakan suatu teknik pembelajaran yang didalam keseluruhan proses pembelajarannya, dibimbing atau di-supervisi. Tujuan utama learning algorithms ini adalah untuk memprediksi hasil yang diberikan serangkaian training samples bersama dengan training labels, yang juga dikenal sebagai classifying data point. Karena, kita memberi tahu algoritma pada saat pelatihan, apa yang harus diprediksi, karenanya hal itu disebut supervised learning. Mari kita pahami dengan sebuah contoh.  
  
Misalkan kita memiliki 10.000 gambar, 5.000 gambar kucing, dan 5.000 gambar anjing, masing-masing gambar kucing & anjing memiliki label nol dan satu. Nah, tujuan dari teknik ini adalah untuk menemukan pola dalam data yang diberi batasan sebagai label. Algoritma supervised learning akan mencoba menemukan batasan antara 10.000 gambar yang membaginya menjadi dua bagian. Sehingga pada saat test ketika gambar baru, misalkan gambar kucing datang sebagai input tanpa label apapun, algoritma memasukkannya ke dalam keranjang kucing yang berlabel 0, yang berarti algoritma tersebut mampu memprediksi / mengklasifikasikan gambar tersebut sebagai gambar kucing.



Mari ambil contoh lain di mana kalian adalah seorang real estate agent dan kalian meng-hire seorang trainee untuk membantu bisnis kalian berkembang. Untuk membantu trainee memprediksi harga rumah, kalian membuat data penjualan rumah untuk mereka. Dan yang paling penting, kalian menulis harga final rumahnya pada data yang ada.

| **Kamar** | **Luas (M2)** | **Tempat** | **Harga (USD)** |
| --- | --- | --- | --- |
| 3 | 200 | Jatinegara | 150.000 |
| 2 | 100 | Bintaro | 300.000 |
| 2 | 150 | Pondok Indah | 250.000 |
| 1 | 250 | Jatinegara | 50.000 |
| 4 | 100 | Bintaro | 450.000 |
| 2 | 100 | Pondok Indah | 250.000 |

Dengan data di atas kita bisa membuat sebuah program yang dapat memberikan estimasi harga.

| **Kamar** | **Luas (M2)** | **Tempat** | **Harga (USD)** |
| --- | --- | --- | --- |
| 2 | 100 | Bintaro | ??? |

Contoh permasalahan di atas merupakan kasus yang disebut dengan **supervised learning**. Permasalahan Supervised Learning dapat dikelompokkan menjadi **regression problem** dan **classification problems**.

Keluaran dari model supervised Machine Learning dapat berupa **category**dari finite set misalnya [rendah, sedang, tinggi] untuk jumlah pengunjung ke pantai:

Input [temperature=**20**] -> *Model* -> Output = [visitors=**high**]

Jika ini masalahnya, model akan memutuskan bagaimana **classify the input**, dan dikenal sebagai **classification**.  
  
Atau, keluarannya bisa berupa **real-world scalar** (keluaran berupa angka):

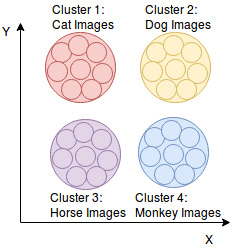
Input [temperature=**20**] -> *Model* -> Output = [visitors=**300**]

Jika demikian, ini dikenal sebagai **regression**.

**Unsupervised Learning**

Tidak seperti supervised learning, kita tidak memiliki label pelatihan untuk sampel pelatihan. Algoritma tersebut dirumuskan sedemikian rupa sehingga dapat menemukan struktur dan pola yang sesuai dalam data. Setelah pola yang konsisten ini menjadi jelas, titik data serupa dapat dikelompokkan bersama, dan titik data yang berbeda akan berada dalam kelompok yang berbeda. Ini sebagian besar digunakan untuk memproyeksikan data dimensi tinggi ke dalam dimensi rendah untuk tujuan visualisasi atau analisis.

Goal dari unsupervised learning adalah untuk memodelkan struktur dasar atau distribusi dalam data dengan tujuan untuk mempelajari data lebih jauh lagi, dengan kata lain, adalah menyimpulkan fungsi yang mendeskripsikan atau menjelaskan data. Berbeda dengan supervised learning, di sini tidak ada jawaban yang ‘dibenarkan secara terarah’ dan tidak ada ‘guru’.



Mari kita ambil contoh yang sama seperti kasus supervised learning, real estate agent. Bagaimana kalau kalian tidak tau harga jual tiap rumah?

| **Kamar** | **Luas (M2)** | **Tempat** |
| --- | --- | --- |
| 3 | 200 | Jatinegara |
| 2 | 100 | Bintaro |
| 2 | 150 | Pondok Indah |
| 1 | 250 | Jatinegara |
| 4 | 100 | Bintaro |
| 2 | 100 | Pondok Indah |

Apa yang bisa kita lakukan dengan data ini? Kita bisa membuat algoritma yang secara otomatis mengidentifikasi segmen pasar dari data tersebut. Mungkin nanti kalian bisa menemukan bahwa pembeli di sekitar kampus lebih suka rumah yang kecil dengan jumlah kamar yang banyak, tapi pembeli di kota lebih memilih rumah 3BR dengan tanah yang luas, dan lain sebagainya.

Contoh permasalahan di atas merupakan kasus yang disebut dengan **unsupervised learning**. Permasalahan unsupervised learning dapat dikelompokkan menjadi **clustering problems**, **association problems,** dan **dimensionality reduction**.

Contoh unsupervised learning dalam kehidupan nyata adalah menyortir koin warna yang berbeda menjadi tumpukan terpisah. Tidak ada yang mengajari kita cara memisahkannya, tetapi hanya dengan melihat fitur-fiturnya seperti warna, kita dapat melihat koin warna mana yang ter-**associated**dan mengelompokkannya/**cluster** ke dalam kelompok yang benar.

Untuk menemukan struktur yang menarik pada data yang tidak berlabel, kita menggunakan **density estimation**. Bentuk yang paling umum di antaranya adalah **clustering**. Diantaranya terdapat juga **dimensionality reduction**, **latent variable models** dan **anomaly detection**. Teknik unsupervised yang lebih kompleks melibatkan jaringan neural seperti Auto-encoder dan Deep Belief Networks, dll.

**Reinforcement Learning**

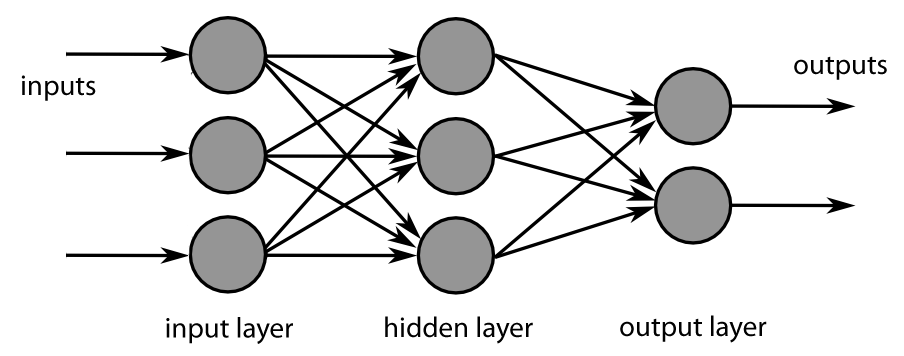
Reinforcement Learning adalah jenis machine learning yang memiliki agent (seperti robot) yang belajar bagaimana berperilaku di lingkungan dengan mengambil tindakan dan mengukur hasilnya. Jika agent memberikan tanggapan yang benar, ia mendapat reward point, yang meningkatkan kepercayaan diri agen untuk mengambil lebih banyak tindakan semacam itu.

Jika kalian terbiasa dengan psikologi, kalian pasti pernah mendengar tentang reinforcement learning. Jika belum, kita pasti sudah mengetahui konsepnya dari cara kita belajar dalam kehidupan sehari-hari. Dalam pendekatan ini, positive and negative feedback sesekali digunakan untuk **reinforce**perilaku. Anggap saja seperti melatih anjing, perilaku yang baik dihargai dengan treat dan menjadi lebih umum dilakukan. Perilaku buruk dihukum dan menjadi kurang umum dilakukan. Perilaku reward-motivated ini adalah kunci dalam reinforcement learning.

Beberapa aplikasi dari pembelajaran Reinforcement adalah game (Alpha Go, Chess, Mario), robotics, traffic light control system, dll.  
  
Itu semua tentang Machine Learning. Sekarang saatnya kita mengenal sedikit tentang **Deep Learning.**

**Deep Learning**

Deep Learning adalah subkategori machine learning. Seperti yang telah dibahas sebelumnya, ide AI terinspirasi dari otak manusia. Jadi, mari kita coba hubungkan titik-titik di sini, deep learning terinspirasi oleh artificial neural networks dan artificial neural networks biasa dikenal sebagai ANN terinspirasi oleh jaringan saraf biologis manusia.



Seperti yang kita lihat dari gambar di atas, ini adalah shallow neural network yang dapat disebut sebagai *Shallow Learning Network*. Neural network akan selalu memiliki:

* Input layer: It can be pixels of an image or a time series data
* Hidden layer: Commonly known as weights which are learned while the neural network is trained
* Output layer: Final layer mainly gives you a prediction of the input you fed into your network.

Jadi, neural network adalah approximation function di mana jaringan mencoba mempelajari parameter (bobot) di hidden layers yang bila dikalikan dengan input akan memberi kita predicted output mendekati outputyang diinginkan.

**Applications of Machine Learning and Deep Learning**

Machine learning dan deep learning banyak digunakan di banyak domain:

* Medical: For cancer cell detection, brain MRI image restoration, gene printing, etc.
* Document: Super-resolving historical document images, segmenting text in document images.
* Banks: Stock prediction, financial decisions.
* Natural Language Processing: recommendation systems: Netflix uses recommendation system to suggest movies to users based on their interest, sentiment analysis, photo tagging.
* Information Retrieval: search engines, both text search, and image search like the one used by Google, Amazon, Facebook, Linkedin, etc.