**CATATAN WEEK 1**

**INTRODUCTION TO ML**

Dikaitkan dengan Arthur Samuel. Dia mendefinisikan **pembelajaran mesin** sebagai bidang studi yang memberi komputer kemampuan untuk belajar tanpa diprogram secara eksplisit.

**NOTE FROM QUICK QUIZ IN VIDEO**

Jika program permainan catur Arthur Samuel hanya diizinkan untuk memainkan 10 permainan (bukan puluhan ribu permainan) melawan dirinya sendiri, bagaimana hal ini akan mempengaruhi kinerjanya?

jawaban : Akan memperburuk keadaan

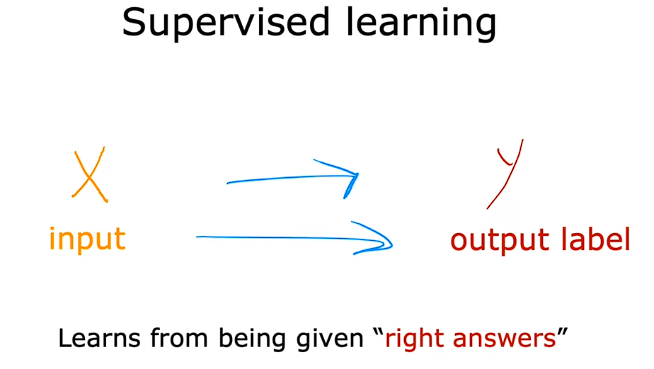
Alasan :

Ini karena jika semakin banyak peluang yang Anda berikan algoritma pembelajaran untuk belajar, semakin baik kinerjanya, karena di soal cuman 10 kali, ini terlalu sedikit sehingga memperburuk keadaan saja

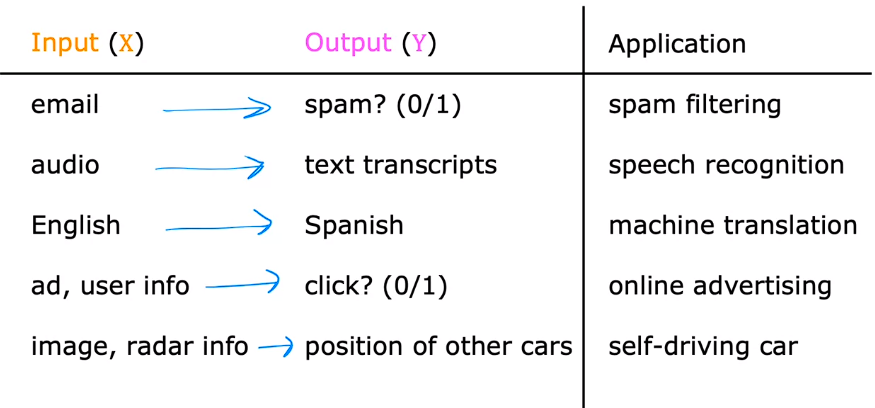
**Method Machine Learning :**

1. **SUPERVISED LEARNING**

**Supervised Learning (Pembelajaran yang diawasi),** mengacu pada algoritme yang belajar x ke y atau pemetaan input ke output.



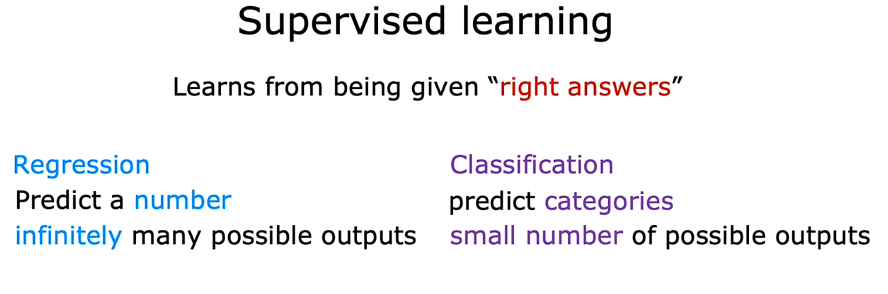
CONTOH :



Jenis utama pembelajaran supervised ada 2 yaitu **Regresi dan Klasifikasi**

Untuk mendefinisikan sedikit lebih banyak terminologi, prediksi harga perumahan adalah jenis pembelajaran terawasi tertentu yang disebut **regresi**. Dengan regresi, maksud saya, kami mencoba memprediksi angka dari banyak kemungkinan angka. Intinya regresi bertugas untuk memprediksi angka apapun yang tak terhingga banyaknya

**Klasifikasi**  merupakan membandingkan 2 output berdasarkan inputan, apakah suatu inputan tersebut merupakan output 1 atau 0 (bisa lebih dari 2 piilihan). Intinya **klasifiaksi** yakni memprediksi kategori



1. **UNSUPERVISED LEARNING**

Tugas kita adalah algoritma yang menemukan beberapa struktur atau pola atau hanya menemukan sesuatu yang menarik dalam data karena data hanya datang dengan input x tetapi bukan output y. Ini adalah **Unsupervised Learning**, disebut tanpa pengawasan karena tidak mencoba mengawasi algoritme. Untuk memberikan beberapa kutipan jawaban yang tepat untuk setiap masukan, sebagai gantinya, maka algoritma harus mencari tahu sendiri apa yang menarik. Atau pola atau struktur apa yang mungkin ada dalam data ini.

Jenis pembelajaran tanpa pengawasan (**Unsupervised Learning**) adalah **Clustering (Algoritma Pengelompokan), Anomaly Detection, dan Dimensionality Reduction**

**Contoh Clustering : Google news**



**Penjelasan**

judul artikel teratas, adalah panda raksasa melahirkan anak kembar di kebun binatang tertua di Jepang. Artikel ini sebenarnya menarik perhatian saya, karena putri saya suka panda dan jadi ada banyak hal mainan panda. Dan menonton video panda di rumah saya, dan melihat ini, Anda mungkin memperhatikan bahwa di bawah ini adalah artikel terkait lainnya. Mungkin dari berita utama saja, Anda bisa mulai menebak apa yang mungkin dilakukan pengelompokan. Perhatikan bahwa kata panda muncul di sini, di sini, di sini dan di sini dan perhatikan bahwa kata kembar juga muncul di kelima artikel. Dan kata Zoo juga muncul di semua artikel ini, jadi algoritma pengelompokan adalah menemukan artikel.

**algoritma Clustering (pengelompokan)** Yang merupakan jenis algoritma pembelajaran tanpa pengawasan, ini mengambil data tanpa label dan mencoba mengelompokkannya secara otomatis ke dalam kelompok.

**Annomaly Detection** digunakan untuk mendeteksi peristiwa yang tidak biasa, seperti deteksi penipuan dalam sistem keuangan, di mana peristiwa yang tidak biasa, transaksi yang tidak biasa bisa menjadi tanda penipuan dan untuk banyak aplikasi lainnya

**Dimensionality Reduction** Ini memungkinkan Anda mengambil kumpulan data besar dan hampir secara ajaib mengompresnya ke kumpulan data yang jauh lebih kecil sambil kehilangan informasi sesedikit mungkin.

Contoh yang menggunakan alogritma usupervised Learning :

1. Diberikan sekumpulan artikel berita yang ditemukan di web, kelompokkan artikel-artikel tersebut ke dalam kumpulan artikel tentang berita yang sama. (clustering)
2. Dengan database data pelanggan, secara otomatis menemukan segmen pasar dan mengelompokkan pelanggan ke dalam segmen pasar yang berbeda. (clustering)

**MODEL LINEAR REGRESSION**

Dataset yang digunakan untuk melatih model disebut **set pelatihan**

**Cost function** akan memberi tahu kami seberapa baik kinerja model sehingga kami dapat mencoba membuatnya menjadi lebih baik.

