

## Algoritma Machine Learning

Nama : I Putu Ananta Wijaya

Program : Introduction to Python for Data Science

Instruktur : Raka Ardhi

### 1) Pengenalan Algoritma Machine Learning (ML)

- Kemampuan kita untuk belajar dari pengalaman untuk menjadi lebih baik dari sebelumnya. Ketika kita lahir, kita tidak mengetahui apa-apa dan tidak bisa melakukan apa-apa sendiri. Tapi seiring berjalannya waktu kita belajar dan menjadi lebih baik dari sebelumnya. Komputer bisa meniru ini, machine learning memanfaatkan statistika dan CS untuk mempelajari pekerjaan yang diinginkan tanpa perlu diprogram secara statis. Komputer membedakan gambar kucing dan anjing. Bagi kita mudah, tapi untuk komputer tidak demikian. Seperti anak kecil, komputer harus belajar dari pengalaman. Kita bisa memberinya data dan komputer akan mempelajari statistical pattern dalam data untuk memungkinkan komputer membedakan kucing dengan anjing. Komputer dapat belajar dengan sendirinya mungkin informasi seperti hidung lebih pendek dan info yang lain. Yang paling penting bukan manusia yang membuat pattern itu tapi komputer berdasarkan data.
- Dari waktu ke waktu makin banyak algoritma dan datanya, makin presisi algoritmanya makin presisi hasilnya.
- Diterapkan dalam Facial Recognition, Recommendation Product, dll.
- Machine learning akan mengambil alih kita? Tidak sekarang jamannya manusia dan mesin melawan permasalahan yang dihadapi, dengan lebih cepat. Task repetitive misalnya dengan manusia sebagai pembuat strateginya.

### 2) Prediksi : Linear Regression (LR)

Berapa banyak film yang kalian tonton bulan ini? Berapa banyak makanan yang dimakan bulan ini atau pendapatan kalian tahun depan? Jika ingin menjawab pertanyaan tersebut maka menggunakan metode populer di DS, yaitu estimasi. Teknik dasarnya adalah Linear Regression (LR)

Contoh Kasus

- Bayangkan kalian pergi ke suatu taman bermain, duduk dan mengamati ada banyak ibu-ibu yang datang dengan anaknya.
- Tantangan: bisa memprediksi umur ibu yang akan datang selanjutnya dengan umur anaknya. Bisa mengamati ibu dan anak yang ada di taman
- Saat ada anak yang terlihat umurnya 3 tahun, ibunya terlihat 30 tahun.
- Saat ada anak yang terlihat umurnya 10 tahun, ibunya terlihat 40 tahun.
- Dengan kalkulasi matematika yang simpel, anak yang baru lahir umur ibunya adalah 28. Jika ada tambahan anak 1 tahun, ibunya bertambah 1,5 tahun lebih tua. Jika ada anak yang umurnya 5 tahun, maka ibunya  $28 + (1,5 \times 5)$  tahun. Disini ada dua variabel, variabel independen adalah umur anaknya dan umur ibu adalah dependen variabel.

Variabel

- Variabel independen (bebas) adalah variabel yang mempengaruhi atau sebab perubahan timbulnya variabel terikat (dependen).
- Variabel dependen (terikat) adalah variabel yang dipengaruhi, akibat dari adanya variabel bebas.

Cara kerja LR

- Planning 1 trip ke kota lain (semarang) – 7 jam perjalanan (442 km). Kalau pulang pergi 14 jam (884 km).
- Berapa jumlah uang yang harus dialokasikan untuk beli bensin ?
- Data dari tahun lalu untuk total uang yang dikeluarkan dengan total jaraknya. Kalau datanya cuma dibaca aja akan rancu. Tapi kalau di plot akan bisa lebih mudah dibaca dengan adanya diagram, garis, dan sumbu. LR yaitu garis yang nantinya akan memprediksi jumlah uang yang dikeluarkan.
- Cari sumbu uang dikeluarkan, ambil mean total berkendara dan mean total dibayar dari data yang ada pada titik itu (titik 442 km). Hasilnya 430 ribu.

### 3) Klustersasi: K-Means Clustering

Algoritma clustering yang paling simpel dan banyak dipakai. Berusaha mengelompokkan data yang ada dalam beberapa kelompok. Dalam satu kelompok mempunyai karakteristik yang sama, sebaliknya jika antar kelompok.

- Analoginya, buka warteg, ingin menentukan membuka warung dimana. Terlihat ada kelompok masyarakat yang tinggal di sebelah kanan, kiri, bawah. Jika kita buat bintang diantara kelompok tersebut, terlihat masyarakat akan mendekat ke bintang terdekat. Pindahkan bintang, maka akan berubah juga lokasi masyarakatnya (beradaptasi lagi). Setelah semua tidak ada berubah maka sudah didapatkan kelompok masyarakat yang efisien.
- Menentukan jumlah kelompok, tidak bisa dilakukan oleh komputer, maka digunakan Elbow Method untuk menentukan pembagian yang paling bagus untuk membagi kelompok tersebut. Setelah dibuatkan diagram oleh elbow, maka ambil jarak yang paling jauh dari setiap titik masyarakat, tampilkan garis tersebut dan buat penghubung di atasnya, garis yang bengkok akan menjadi kategori yang diambil sebagai jumlah kelompok.

### 4) Klasifikasi: K Nearest Neighbours (KNN)

Algoritma supervised learning dimana hasil dari data yang baru diklasifikasikan berdasarkan mayoritas dari kategori k-tetangga terdekat. Tujuan untuk klasifikasi objek baru berdasarkan atributnya. Misalnya sebuah rumah yang terletak di sebuah perbatasan antara kota semarang dan kabupaten semarang. Kesulitan rumah tersebut masuk ke wilayah mana. Dapat ditentukan dengan KNN. Dengan menentukan tetangga rumah. Analoginya adalah

- Hitung jarak setiap tetangga. Urutkan
- Ambil 3 tetangga yang paling dekat. Lihat masing-masing, apakah termasuk wilayah kota atau kabupaten.
- Jika ada 2 rumah yang masuk ke kota semarang, maka rumah ini masuk ke kota semarang. Dan sebaliknya.

Ada tiga dots (anjing (merah), kucing (biru), jerapah (kuning)). Lihat grouping berdasarkan tinggi badan dan warna. Ada group merah di sebelah kiri, diikuti group biru di tengah, group kuning di kanan. Jika ada titik ungu nyempil, lihat tetangganya, mana yang lebih banyak. Kalau data biru lebih banyak, ya masuk ke kelompok biru. Tergantung jumlah berapa tetangga yang kita ambil.