

Nama : Ghina Khoerunnisa

Program : Data Scientist Intern

Algoritma Machine Learning

1. Pengenalan Machine Learning

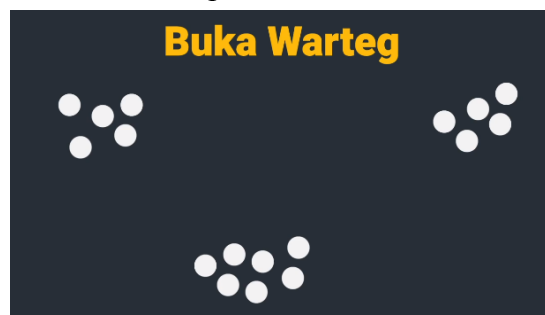
Apa itu machine learning? Machine learning memanfaatkan statistika dan computer science untuk mempelajari bagaimana cara untuk menyelesaikan pekerjaan yang diberikan tanpa harus diprogram. ML belajar dari data hingga akhirnya menemukan statistical pattern yang sesuai. Penerapan ML ini ada di belakang face recognition, text to speech recognition, spam detection, recommender system, dll. AI & ML digunakan bukan untuk memusnahkan manusia tetapi ML dan Manusia saling bekerja sama untuk memecahkan suatu masalah dengan waktu yang lebih singkat dan menyelesaikan repetitive task.

2. Linear Regression

Linear Regression digunakan untuk melakukan regresi, estimasi. Misal di taman ada banyak ibu-ibu yang membawa anaknya. Nah ada Anak A berumur 3 thn dan Ibu A berumur 30 thn, kemudian ada Anak B berumur 10 thn dan Ibunya berumur 40 tahun. Dari sini kita dapat menghitung umur Ibu C jika anaknya berumur 1 tahun dengan memberikan bobot. Nah bobot inilah yang dicari dengan mesin. Sebelumnya terdapat variabel independen dan dependen, independen berarti dapat berdiri sendiri atau tidak bergantung dengan variabel lainnya misalnya dalam contoh di atas adalah umur anak, sedangkan variabel dependen merupakan variabel yang bergantung pada variabel lain contohnya umur ibu yang bergantung pada umur anaknya pada kasus ini. Linear regression bekerja dengan mencari bobot sedemikian hingga menghasilkan cost function seminimal mungkin.

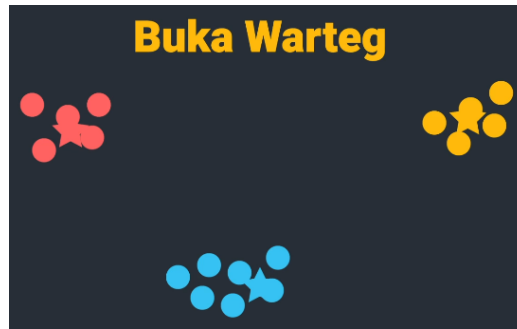
3. K-Means

K-Means digunakan untuk clustering. Clustering bekerja untuk mengumpulkan data menjadi beberapa kelompok/cluster yang mana data yang berada di satu kelompok memiliki karakteristik yang sama satu sama lainnya dan punya karakteristik yang berbeda dengan kelompok lainnya. Contoh Studi kasusnya adalah pembukaan warteg pada suatu tempat/wilayah. Misalnya terdapat 3 kelompok penduduk di suatu wilayah. Nah kita ingin menentukan mau membuka warteg dimana tempatnya. Sehingga warteg tersebut bisa lebih optimal.



Mengenerate 3 initial centroid (representasi cluster) secara random. Kemudian menghitung jarak semua titik ke masing-masing centroid. Lalu assign titik-titik tersebut ke dalam centroid terdekat. Update nilai centroid dengan menghitung

rata-rata tiap cluster. Lakukan perhitungan kembali hingga tidak ada perubahan yang terjadi di tiap centroid.



Jika datanya terlihat seperti di atas kita dapat dengan mudah mengetahui bahwa ada 3 cluster. Tetapi di dunia nyata tidak semudah itu oleh karena itu kita perlu mengetahui jumlah K yang dibutuhkan dengan menggunakan Elbow Method. Elbow method ini kita menghitung jarak antar titik terjauh di cluster kemudian menampilkan grafiknya dan garis yang bengkok atau elbownya merupakan K yang optimum.

4. K-Nearest Neighbors

KNN merupakan algoritma supervised learning dimana hasil dari data yang baru diklasifikasikan berdasarkan mayoritas dari kategori k-tetangga terdekatnya. Misal kita ada sebuah rumah yang berada di tengah antara Kota A dan Kota B sehingga sulit untuk mengetahui rumah tersebut masuk ke kota mana. Disini kita bisa menggunakan KNN dengan cara menghitung semua jarak semua rumah ke rumah tersebut kemudian mengambil observasi sebanyak K terdekat (ganjil), misalnya 3 tetangga terdekat. Dari 3 tetangga tersebut mayoritasnya apakah ada di Kota A atau Kota B. Jika mayoritasnya dari Kota A maka rumah tersebut masuk ke Kota A dan sebaliknya.