

Resume 03

Machine Learning

Nama : Andi Nurul Qalbi

NIM : H071191014

Prodi : Sistem Informasi

Github : <https://github.com/Andiqal/Machine-Learning.git>

Decision tree,

Decision tree membangun model klasifikasi dan regresi dalam bentuk struktur pohon. Algoritma ini menguraikan kumpulan data menjadi himpunan bagian yang lebih kecil dan menghubungkannya menjadi pohon keputusan yang terkait. Tujuan utama dari algoritma decision tree adalah untuk membangun model pelatihan yang digunakan untuk memprediksi nilai variabel target dengan mempelajari aturan keputusan. Aturan ini disimpulkan dari data training yang sebelumnya telah diinput. Keuntungan algoritma ini adalah mudah dimengerti, mudah menghasilkan aturan, tidak mengandung hiper-parameter, dan model decision tree yang kompleks dapat disederhanakan secara signifikan dengan visualisasinya.

Decision Tree adalah salah satu metode klasifikasi yang paling populer, karena mudah untuk diinterpretasi oleh manusia. Decision Tree adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Algoritma Machine Learning jenis ini melakukan tugasnya dengan menggunakan konsep struktur flowchart bercabang menggunakan decision rules atau aturan-aturan keputusan yang dibuat oleh desainernya.

Pada dasarnya, Decision Tree dimulai dengan satu node atau simpul. Kemudian, node tersebut bercabang untuk menyatakan pilihan-pilihan yang ada. Selanjutnya, setiap cabang tersebut akan memiliki cabang-cabang baru. Mengutip dari Venngage, Decision Tree memiliki tiga elemen di dalamnya, yaitu:

- Root node (akar), Tujuan akhir atau keputusan besar yang ingin diambil.
- Branches (ranting), Berbagai pilihan tindakan.
- Leaf node (daun), Kemungkinan hasil atas setiap tindakan.

Hierarchical Clustering,

Hierarchical Clustering adalah teknik clustering dengan algoritma Machine Learning yang membentuk hirarki atau berdasarkan tingkatan tertentu sehingga

menyerupai struktur pohon. Dengan demikian proses pengelompokannya dilakukan secara bertingkat atau bertahap. Biasanya, metode ini digunakan pada data yang jumlahnya tidak terlalu banyak dan jumlah cluster yang akan dibentuk belum diketahui.

Secara prinsip, Hierarchical Clustering ini akan melakukan clustering secara berjenjang berdasarkan kemiripan tiap data. Sehingga pada akhirnya, pada ujung hierarki akan terbentuk cluster-cluster yang karakteristiknya berbeda satu sama lain, dan objek di satu cluster yang sama memiliki kemiripan satu sama lain.

Di dalam metode hirarki, terdapat dua jenis strategi pengelompokan yaitu Agglomerative dan Divisive. Agglomerative Clustering (metode penggabungan) adalah strategi pengelompokan hirarki yang dimulai dengan setiap objek dalam satu cluster yang terpisah kemudian membentuk cluster yang semakin membesar. Jadi, banyaknya cluster awal adalah sama dengan banyaknya objek. Sedangkan Divisive Clustering (metode pembagian) adalah strategi pengelompokan hirarki yang dimulai dari semua objek dikelompokkan menjadi cluster tunggal kemudian dipisah sampai setiap objek berada dalam cluster yang terpisah.

K-Means Clustering,

Seperti namanya, algoritma ini biasa digunakan untuk kasus clustering. K-means Clustering adalah salah satu contoh algoritma Unsupervised Machine Learning yang paling sederhana dan populer. Metode K-Means Clustering berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain.

Cara kerja algoritma ini mula-mula adalah dengan membentuk sejumlah k titik, yang disebut dengan centroid (dimana nilai k merepresentasikan jumlah cluster). Kemudian titik-titik data (data points) yang ada akan membentuk cluster dengan centroid terdekat darinya. Otomatis, titik pusat (centroid) akan berubah seiring dengan penambahan anggota tiap cluster-nya (yang mana adalah data points tadi). Oleh karena itu, tiap-tiap cluster yang telah terbentuk akan mencari titik centroid barunya. Proses ini terus menerus dilakukan hingga diperoleh kondisi konvergensi, contohnya jika posisi centroid sudah tidak berubah.

Terdapat dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu Hierarchical dan Non-Hierarchical, dan K-Means merupakan salah satu metode data clustering non-hierarchical atau Partitional Clustering.

Naive Bayes Classifier,

Naive bayes classifier merupakan algoritma klasifikasi yang sangat sederhana berdasarkan apa yang disebut pada teorema bayesian. Algoritma ini memiliki satu sifat umum, yaitu setiap data diklasifikasikan tidak bergantung pada fitur lain yang terikat pada kelas atau biasa disebut dengan independen. Artinya, satu data tidak berdampak pada data yang lain. Meskipun algoritma ini merupakan algoritma yang tergolong sederhana, namun naive bayes dapat mengalahkan beberapa metode klasifikasi yang lebih canggih. Algoritma ini biasa digunakan untuk deteksi spam dan klasifikasi dokumen teks. Kelebihan algoritma ini adalah sederhana dan mudah diterapkan, tidak sensitif terhadap fitur yang tidak relevan, cepat, hanya membutuhkan sedikit data training, dan dapat digunakan untuk masalah klasifikasi multi-class dan biner.

Random Forest,

Algoritma Random Forest Classifier merupakan salah satu algoritma klasifikasi machine learning yang paling populer. Seperti namanya, algoritma ini bekerja dengan cara membuat hutan pohon secara acak. Semakin banyak pohon yang dibuat, maka hasilnya akan semakin akurat. Dasar dari algoritma random forest adalah algoritma decision tree. Keuntungan dari algoritma ini adalah dapat digunakan untuk rekayasa fitur seperti mengidentifikasi fitur yang paling penting diantara semua fitur yang tersedia dalam dataset training, bekerja sangat baik pada database berukuran besar, sangat fleksibel, dan memiliki akurasi yang tinggi.