

Apa itu supervised learning?

Supervised learning adalah pembelajaran sebuah model akan sebuah dataset yang sudah dilabeli dengan benar. Dalam kata lain, sebuah model tersebut belajar dengan sebuah dataset yang sudah memiliki “kunci jawabannya”. Dengan begitu, model dapat memprediksi input-input yang belum pernah dilihatnya berdasarkan data-data lain yang sudah dipelajari.

Jelaskan bagaimana cara kerja dari algoritma yang anda implementasikan!

Algoritma KNN

Algoritma KNN memiliki sebuah nilai K, yaitu jumlah “tetangga” terdekat yang jaraknya dihitung secara euclidean. Dari K tetangga tersebut, akan dilihat perbandingan label tetangganya, dari hasil perbandingannya, label akan ditentukan dari label tetangga terbanyak.

Logistic Regression

Algoritma Logistic Regression yang diimplementasikan pada awalnya akan menskala seluruh data menjadi range 0-1 agar bisa diproses menggunakan library math tanpa Overflow error. Setelah diskala, dibuat sebuah array berisi koefisien (diinisialisasi dengan 0) yang berupa nilai pengali data untuk dimasukkan ke dalam fungsi sigmoid. Koefisien-koefisien tersebut diulang ulang penghitungannya sebanyak epoch yang ditentukan, dengan setiap kali iterasi digunakan perhitungan dengan faktor learning rate yang juga sudah ditentukan. Pada setiap iterasi juga dilakukan penghitungan loss dengan metode stochastic gradient descent. Koefisien pada akhir epoch merepresentasikan bentuk akhir model tersebut. Digunakan threshold ≥ 0.5 untuk diklasifikasikan menjadi 1 dan sebaliknya untuk diklasifikasikan menjadi 0. Fungsi sigmoid yang digunakan adalah sebagai berikut.

$$S(x) = \frac{1}{1 + e^{-x}}.$$

ID3

Algoritma ID3 menghitung information gain dari setiap atribut non target dengan memperhitungkan entropinya dan kemudian menjadikan atribut dengan information gain tertinggi sebagai basis pembagi kategori berdasarkan kelas-kelas yang dimilikinya. Jika setiap anggota dari sebuah kelas dalam atribut terpetakan ke dalam kategori yang sama dalam target, kelas tersebut akan menjadi daun dari decision tree. Penentuan decision tree kemudian akan lanjut secara rekursif sampai ditemukan simpul daun.

Bandingkan ketiga algoritma tersebut, kemudian tuliskan kelebihan dan kekurahannya!

Algoritma	Kelebihan	Kekurangan
KNN	<ul style="list-style-type: none"> • Tidak ada asumsi tentang data • Sederhana dan mudah dipahami • Bisa dipakai untuk klasifikasi dan regresi 	<ul style="list-style-type: none"> • Memakan banyak memori • Sensitif terhadap feature yang tidak penting

Logistic Regression	<ul style="list-style-type: none">• Mudah untuk diimplementasi• Akurasi baik untuk dataset yang simpel• Dengan mudah dijadikan klasifikasi untuk beberapa kelas	<ul style="list-style-type: none">• Hanya bisa digunakan untuk memprediksi fungsi diskrit• Dapat overfit jika jumlah observasi kurang dari jumlah fitur• Memiliki batasan linear
ID3	<ul style="list-style-type: none">• Membangun tree yang paling cepat dan relatif pendek.• Search space yang komplit• Instance yang memiliki atribut missing value bisa dipakai untuk training.• Kurang terpengaruh dengan noise	<ul style="list-style-type: none">• Tidak bisa menggunakan atribut yang bertipe kontinu.• Tidak mempedulikan value-value yang mungkin hilang.• Tidak baik untuk dataset berukuran kecil.• Masalah overfit kurang diperdulikan.

Jelaskan penerapan dari algoritma supervised! (misal di bidang kesehatan atau industri)

Algoritma supervised bisa diaplikasikan di berbagai bidang. Contohnya dalam bidang kesehatan, diagnosis penyakit tertentu bisa saja diprediksi dengan melihat ukuran-ukuran kondisi tubuh, misalkan tekanan darah, kadar gula, dan sebagainya. Dari ukuran-ukuran tubuh tersebut dapat diklasifikasikan mana orang yang beresiko terkena suatu penyakit dan mana yang tidak dalam sebuah kategori biner.