

Resume 04

Machine Learning

Nama : Andi Nurul Qalbi

NIM : H071191014

Prodi : Sistem Informasi

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

Naïve bayes,

Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema Bayes . Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes , yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes . Ciri utama dr Naïve Bayes Classifier ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian.

Naive Bayes Classifier bekerja sangat baik dibanding dengan model classifier lainnya. Hal ini dibuktikan oleh Xhemali , Hinde Stone dalam jurnalnya “Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages” mengatakan bahwa “Naïve Bayes Classifier memiliki tingkat akurasi yg lebih baik dibanding model classifier lainnya”.

Keuntungan penggunaan adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan (training data) yg kecil unt menentukan estimasi parameter yg diperlukan dalam proses pengklasifikasian. Karena yg diasumsikan sebagai variable independent, maka hanya varians dr suatu variable dalam sebuah kelas yg dibutuhkan unt menentukan klasifikasi, bukan keseluruhan dr matriks kovarians.

Kelebihan Naïve Bayes :

- Bisa dipakai untuk data kuantitatif maupun kualitatif
- Tidak memerlukan jumlah data yang banyak
- Tidak perlu melakukan data training yang banyak
- Jika ada nilai yang hilang, maka bisa diabaikan dalam perhitungan.
- Perhitungannya cepat dan efisien
- Mudah dipahami
- Mudah dibuat

- Pengklasifikasian dokumen bisa dipersonalisasi, disesuaikan dengan kebutuhan setiap orang
- Jika digunakan dalam bahasa pemrograman, code-nya sederhana
- Bisa digunakan untuk klasifikasi masalah biner ataupun multiclass

Kekurangan Naïve Bayes :

- Apabila probabilitas kondisionalnya bernilai nol, maka probabilitas prediksi juga akan bernilai nol
- Asumsi bahwa masing-masing variabel independen membuat berkurangnya akurasi, karena biasanya ada korelasi antara variabel yang satu dengan variabel yang lain
- Keakuratannya tidak bisa diukur menggunakan satu probabilitas saja. Butuh bukti-bukti lain untuk membuktikannya.
- Untuk membuat keputusan, diperlukan pengetahuan awal atau pengetahuan mengenai masa sebelumnya. Keberhasilannya sangat bergantung pada pengetahuan awal tersebut Banyak celah yang bisa mengurangi efektivitasnya
- Dirancang untuk mendeteksi kata-kata saja, tidak bisa berupa gambar