



Naive Bayes

Siska Atmawan Oktavia, S.T., M.Cs.





Abstract

- ❑ Konsep Dasar Naïve Bayes
- ❑ Naïve Bayes Untuk Data Kategorial
- ❑ Naïve Bayes Untuk Data Kontinue



Introduction

Thomas Bayes ialah ilmuwan berketurunan Inggris dan sukses menciptakan Algoritma Naïve Bayes. Naïve Bayes yang meramal suatu peluang di masa depan berdasarkan data di masa lebih dahulu.



Pengertian Naïve Bayes

Algoritma *Naïve Bayes* (NB) merupakan penerapan dari *Bayes Theorem*. Ide dasar dari Bayes Theorem adalah bahwa hasil dari hipotesis (H) dapat diperkirakan berdasarkan pada beberapa evidence (E) yang diamati. Secara umum didefinisikan sebagai berikut:

$$P(H|E) = \frac{P(E|H)*P(H)}{P(E)}$$

P= Probabilitas

H= Hipotesis

E= Evidence



Naïve Baiyes merupakan salah satu algoritma lazy learning (melakukan learning/ pelatihan saat ada data yang akan diprediksi) yang dapat digunakan pada masalah klasifikasi, penyaringan pesan *spam*, analisis sentimen, hingga sistem rekomendasi. Algoritma Naïve Bayes membutuhkan fitur atau prediktor independen.

Contoh Data Kontinue

Apabila $C1 = 300$, $C2 = 17$, $C3 = \text{Tidak}$, maka tentukan apakah lokasi akan di bangun perumahan?

Aturan ke-	Harga tanah (C1)	Jarak dari pusat kota (C2)	Ada angkutan umum (C3)	Dipilih untuk Perumahan (C4)
1	100	2	Tidak	Iya
2	200	1	Tidak	Iya
3	500	3	Tidak	Iya
4	600	20	Tidak	Tidak
5	550	8	Tidak	Tidak
6	250	25	Ada	Tidak
7	75	15	Ada	Tidak
8	80	10	Tidak	Iya
9	700	18	Ada	Tidak
10	180	8	Ada	Iya

Hal yang pertama kamu lakukan adalah mencari probabilitas kemunculan setiap nilai untuk atribut (*class*).

- Probabilitas kemunculan setiap nilai untuk atribut Harga tanah (C1).
- Probabilitas kemunculan setiap nilai untuk atribut Jarak dari pusat kota (C2).

Rumus perhitungan yang digunakan

$$f(C1 = 300 | \text{ya}) = \frac{1}{\sqrt{2\pi}(168,8787)} e^{\frac{-(300-212)^2}{2(168,8787)^2}} = 0,0021.$$

$$f(C1 = 300 | \text{tidak}) = \frac{1}{\sqrt{2\pi}(261.9637)} e^{\frac{-(300-435)^2}{2(261.9637)^2}} = 0,0013.$$

$$f(C2 = 17 | \text{ya}) = \frac{1}{\sqrt{2\pi}(3.9623)} e^{\frac{-(17-4,8)^2}{2(3.9623)^2}} = 0,0009.$$

$$f(C2 = 17 | \text{tidak}) = \frac{1}{\sqrt{2\pi}(6,3008)} e^{\frac{-(17-17,2)^2}{2(6,3008)^2}} = 0,0633.$$

Ketika nilai C1 dan C2 sudah dapat dengan rumus Gaussian, jadi kita lanjut untuk mencari nilai *likelihood*.

Maka dari hasil akhir kita dapat menyimpulkan bahwa pada lokasi tersebut tidak dibangun perumahan.

Tipe-tipe Naive Bayes Classifier

1. Multinomial Naive Bayes

Digunakan untuk menyelesaikan masalah klasifikasi dokumen. Algoritma ini akan membantu Anda mengetahui kategori suatu dokumen, apakah termasuk dokumen penting, dokumen yang dapat dialihkan, atau justru dokumen sampah yang berbahaya.

2. Bernoulli Naive Bayes

Mimiliki kemiripan dengan Multinomial Naive Bayes. Perbedaannya terletak pada fitur atau prediktornya. Alih-alih menggunakan frekuensi kata, algoritma ini menggunakan variabel boolean. Parameter yang kita gunakan untuk memprediksi variabel kelas hanya mengambil nilai ya atau tidak.

3. Gaussian Naive Bayes

Fitur atau prediktor mengambil nilai yang kontinu (tidak diskrit). Saat data diplot, maka akan menampilkan sebuah kurva dengan bentuk lonceng yang simetris. Kurva tersebut menunjukkan rata-rata nilai fitur.

Penerapan Algoritma Naive Bayes Classifier

1. Naive Bayes Classifier banyak digunakan pada tugas klasifikasi seperti:

1. Pengenalan Wajah

Mengidentifikasi wajah atau fitur lainnya, seperti hidung, mulut, mata.

2. Prediksi cuaca

Memprediksi apakah cuaca akan baik atau buruk.

3. Diagnosa medis

Dokter dan profesional kesehatan dapat menggunakan Naive Bayes untuk mendiagnosis apakah pasien berisiko tinggi untuk penyakit dan kondisi tertentu, seperti penyakit jantung, kanker, dan penyakit lainnya.

4. Klasifikasi berita

Google News dapat mengenali apakah sebuah berita bersifat politik, berita dunia, dan sebagainya.

Kelebihan Naïve Bayes

Algoritma ini cukup memerlukan sejumlah data pelatihan (Training Data) dalam menaksir

1. perkiraan skala yang dibutuhkan dalam proses pengelompokan. Naïve Bayes kerap beroperasi jauh lebih baik dalam mayoritas suasana dunia nyata yang kompleks dari pada yang diperlukan

Cepat dalam perhitungan, algoritma simpel dan presisi. Naïve Bayes lebih akurat

2. diaplikasikan dalam data yang besar dan dapat mengerjakan data yang tidak utuh (missing value) serta kuat atas atribut yang tidak bermakna dan noise pada data

3. Mudah diimplementasikan dan dalam banyak kasus memberikan hasil yang baik



Kekurangan Naïve Bayes

1. Probabilitasnya tidak dapat menilai seberapa presisi suatu pengelompokkan. Algoritma klasifikasi Naïve Bayes mempunyai kelemahan pada penetapan atribut yang terdapat di dalam data sehingga dapat mempengaruhi produk akhir berupa tingkat presisi.
2. Amat rentan pada fitur yang berlebih, sehingga membuat tingkat presisi pengelompokkan menjadi rendah.
3. Tidak valid jika probabilitas semmentaranya adalah kosong, jika kosong maka probabilitas prakiraan akan bernilai kosong juga. Serta memperhitungkan variabel independent.





Terima kasih

