

Laporan Jobsheet 12

Teknologi Data (Naïve Bayes Classifier)



Disusun oleh :

Nama :Dhuta Pamungkas Ibnusiqin

Nim : 1941723014

Class:TI-3D

Jurusan Teknologi Informasi
Politeknik Negeri Malang
2021

Naïve Bayesian Classifier

Naive Bayes adalah keluarga algoritme probabilistik yang memanfaatkan teori probabilitas dan Teorema Bayes untuk memprediksi tag teks (seperti berita atau ulasan pelanggan). Mereka bersifat probabilistik, yang berarti bahwa mereka menghitung probabilitas setiap tag untuk teks tertentu, dan kemudian mengeluarkan tag dengan yang tertinggi. Cara mereka mendapatkan probabilitas ini adalah dengan menggunakan Teorema Bayes, yang mendeskripsikan probabilitas fitur, berdasarkan pengetahuan sebelumnya tentang kondisi yang mungkin terkait dengan fitur tersebut.

Dengan kata lain, **Naïve Bayesian Classifier** mengasumsikan bahwa keberadaan sebuah atribut (variabel) tidak ada kaitannya dengan keberadaan atribut (variabel) yang lain.

- Karena asumsi atribut tidak saling terkait (conditionally independent), maka:

$$P(X | C_i) = \prod_{k=1}^n P(x_k | C_i)$$

- Bila $P(X|C_i)$ dapat diketahui melalui perhitungan di atas, maka klas (label) dari data sampel X adalah klas (label) yang memiliki $P(X|C_i) \cdot P(C_i)$ maksimum

Dalam setiap metode pastinya memiliki kelebihan dan kekurangan dalam setiap prosesnya. Hal ini juga terdapat pada Naïve Bayesian Classifier yaitu :

- Kelebihan :
 - Mudah diimplementasi
 - Memberikan hasil yang baik untuk banyak kasus
 - Hanya memerlukan jumlah data sedikit yang dibutuhkan untuk klasifikasi
 - Cepat, efisien, dan mudah untuk dibuat
- Kekurangan :
 - Harus mengasumsi bahwa antar fitur tidak terkait (independent) Dalam realita, keterkaitan itu ada
 - Independence antar atribut membuat akurasi menjadi berkurang
 - Tidak berlaku jika nilai probabilitasnya adalah nol (0)
 - Keterkaitan tersebut tidak dapat dimodelkan oleh Naïve Bayesian Classifier

Contoh Studi Kasus :

- Data Mobil Pribadi

Warna a1	Tipe a2	Asal a4	Tercuri vj
merah	Sport	Domestik	ya
merah	Sport	Domestik	tidak
merah	Sport	Domestik	ya
kuning	Sport	Domestik	tidak
kuning	Sport	Import	ya
kuning	SUV	Import	tidak
kuning	SUV	Import	ya
kuning	SUV	Domestik	tidak
merah	SUV	Import	tidak
merah	Sport	Import	ya

- Cara perhitungan probabilitas

Menghitung Probabilitas Mobil tercuri YA & TIDAK (P(vj))			
$P(v_j) = \frac{N}{\text{jumlah}}$			
P(YA) = 5/10			
P(TIDAK) = 5/10			

- Perhitungan probabilitas yang tercuri

Menghitung Probabilitas yang tercuri pada mobil warna merah, tipe SUV, dan asal domestik			
P(Merah YA) = 3/5 = 0.6			
P(SUV YA) = 1/5 = 0.2			
P(Domestik YA) = 2/5 = 0.4			

- Perhitungan probabilitas tidak dicuri

Menghitung Probabilitas tidak tercuri pada mobil warna merah, tipe SUV, dan asal domestik			
P(Merah TIDAK) = 2/5 = 0.4			
P(SUV TIDAK) = 3/5 = 0.6			
P(Domestik TIDAK) = 3/5 = 0.6			

- Hasil

Menentukan berapa persen mobil tercuri dan tidak tercuri

$$\begin{aligned} \text{a. tercuri (ya)} &= P(\text{ya}) * P(\text{merah}|\text{ya}) * P(\text{SUV}|\text{ya}) \\ &\quad * P(\text{domestik}|\text{ya}) \\ &= 0.5 * 0.6 * 0.2 * 0.4 \\ &= 0.024 \text{ atau } 2,4\% \end{aligned}$$

$$\begin{aligned} \text{b. tercuri (tidak)} &= P(\text{tidak}) * P(\text{merah}|\text{tidak}) \\ &\quad * P(\text{SUV}|\text{tidak}) * P(\text{domestik}|\text{tidak}) \\ &= 0.5 * 0.4 * 0.6 * 0.6 \\ &= 0.072 \text{ atau } 7,2\% \end{aligned}$$

- Kesimpulan

Jadi, berdasarkan hasil perhitungan tercuri di atas dengan hasil tercuri (tidak) > tercuri (ya) yaitu 7.2% > 2.4% maka dapat disimpulkan mobil dengan warna merah, tipe SUV, dan asal domestik TIDAK TERCURI.