Algoritma Naïve bayes adalah salah satu algoritma klasifikasi berdasarkan teorema Bayesian pada statistika (Suntoro, Wahyu, & Indriawati, 2018). Algoritma Naïve Bayes dapat digunakan untuk meprediksi probabilitas keanggotaan suatu kelas (Han and Kamber, 2012).

Teorema Bayesian menghitung nilai posterior probability P(H|X) menggunakan probabilitas P(H), P(X), dan P(X|H) (Kantardzic, 2011), di mana nilai X adalah data testing yang kelasnya belum diketahui. Nilai H adalah hipotesis data X yang merupakan suatu kelas yang lebih spesifik. Nilai P(X|H) atau disebut juga dengan likelihood, adalah probabilitas hipotesis X berdasarkan kondisi H. Nilai P(H) atau disebut juga dengan prior probability adalah probabilitas hipotesis H. sedangkan nila P(X) yang disebut juga dengan predictor prior probability, adalah probabilitas X.

Algoritma Naïve Bayes sangat cocok untuk melakukan klasifikasi pada dataset bertipe nominal.

Langkah-langkah algoritma naïve bayes :

1. Siapkan dataset.
2. Hitung jumlah kelas pada data training.
3. Hitung jumlah kasus yang sama denan kelas yang sama.
4. Kalikan semua hasil sesuat dengan data testing yang akan dicari kelasnya.
5. Bandingkan hasil per kelas, nilai tertinggi ditetapkan sebagai kelas baru.

Perhitungan manual Tipe data numeric dan nominal Algoritma Naïve Bayes

1. Siapkan dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cuaca | Temperature | Kelembaban | Angin | Play |
| Cerah | 85 | 85 | Tidak | Tidak |
| Cerah | 80 | 90 | Ada | Tidak |
| Mendung | 83 | 78 | Tidak | Ya |
| Hujan | 70 | 96 | Tidak | Ya |
| Hujan | 68 | 80 | Tidak | Ya |
| Hujan | 65 | 70 | Ada | Tidak |
| Mendung | 64 | 65 | Ada | Ya |
| Cerah | 72 | 95 | Tidak | Tidak |
| Cerah | 69 | 70 | Tidak | Ya |
| Hujan | 75 | 80 | Tidak | Ya |
| Cerah | 75 | 70 | Ada | Ya |
| Mendung | 72 | 90 | Ada | Ya |
| Mendung | 81 | 75 | Tidak | Ya |
| Hujan | 71 | 80 | Ada | Tidak |

Data testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cuaca | Temperature | Kelembaban | Angina | Play |
| Cerah | 73 | 80 | Tidak | ? |

1. Hitung jumlah kelas pada data training

Jumlah kelas bermain golf = 9

Jumlah kelas tidak bermain golf = 5

Maka:

1. Hitung jumlah kasus yang sama dengan kelas yang sama
2. Atribut cuaca
3. Atribut Temperatur
4. Atribut kelembaban
5. Atribut angin
6. Kalikan semua hasil sesuai dengan data testing yang akan dicari kelasnya
7. Bandingkan hasil per kelas

Dapat disimpulkan bahwa data cuaca = cerah, temperaur = 73, kelembaban = 80, dan angin = tidak, masuk ke dalam kelas bermain golf karena probabilitas bermain golf (0,00023418) lebih tinggi disbandingkan probabilitas tidak bermain golf (0,00017850).