

Nama : Ferza Reyaldi
NIM : 09021281924060
Mata Kuliah : Data Mining

Tugas Pertemuan 1 (Pertemuan 11)

Pelajari & tuliskan penjelasan tentang metode atau algoritma Naive Bayes!

Jawab:

Algoritma Naive Bayes didasarkan pada teorema Bayes dengan asumsi independensi di antara prediktor. Secara sederhana, pengklasifikasi Naive Bayes mengasumsikan bahwa kehadiran fitur tertentu di kelas tidak terkait dengan kehadiran fitur lainnya.

Teorema Bayes menyediakan cara menghitung probabilitas posterior $P(c|x)$ dari $P(c)$, $P(x)$ dan $P(x|c)$. Perhatikan persamaan di bawah ini:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

The diagram shows the formula with arrows pointing to each term: $P(c|x)$ is labeled 'Posterior Probability', $P(x|c)$ is labeled 'Likelihood', $P(c)$ is labeled 'Class Prior Probability', and $P(x)$ is labeled 'Predictor Prior Probability'.

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Keterangan:

- $P(c|x)$ adalah probabilitas posterior kelas (c , target) yang diberikan prediktor (x , atribut).
- $P(c)$ adalah probabilitas prior kelas.
- $P(x|c)$ adalah probabilitas yang merupakan probabilitas dari kelas yang diberikan oleh prediktor.
- $P(x)$ adalah probabilitas prior dari prediktor.

Cara Kerja Algoritma Naive Bayes:

- 1) Ubah kumpulan data menjadi tabel frekuensi
- 2) Buat tabel Likelihood dengan mencari probabilitas fiturnya.
- 3) Gunakan persamaan Naive Bayesian untuk menghitung probabilitas posterior untuk setiap kelas. Kelas dengan probabilitas posterior tertinggi adalah hasil prediksi.

Pros:

- Mudah dan cepat untuk memprediksi kelas kumpulan data uji. Ini juga berkinerja baik dalam prediksi multi-kelas
- Ketika asumsi independensi berlaku, pengklasifikasi Naive Bayes berkinerja lebih baik dibandingkan dengan model lain seperti regresi logistik dan Anda memerlukan lebih sedikit data pelatihan.

- Ini berkinerja baik dalam hal variabel input kategoris dibandingkan dengan variabel numerik. Untuk variabel numerik, diasumsikan distribusi normal (kurva lonceng, yang merupakan asumsi kuat).

Cons:

- Jika variabel kategori memiliki kategori (dalam kumpulan data uji), yang tidak diamati dalam kumpulan data pelatihan, maka model akan menetapkan probabilitas 0 (nol) dan tidak akan dapat membuat prediksi. Ini sering dikenal sebagai "Frekuensi Nol". Untuk mengatasi hal ini, kita dapat menggunakan teknik smoothing. Salah satu teknik pemulusan yang paling sederhana disebut estimasi Laplace.
- Di sisi lain naive bayes juga dikenal sebagai estimator yang buruk, sehingga keluaran probabilitas tidak perlu dianggap terlalu serius.
- Keterbatasan lain dari Naive Bayes adalah asumsi prediktor independen. Dalam kehidupan nyata, hampir tidak mungkin kita mendapatkan seperangkat prediktor yang benar-benar independen.