

Gaussian Naive Bayes

1. Cara kerja algoritma

Gaussian naive bayes adalah metode klasifikasi berbasis probabilitas yang menggunakan teorema bayes dengan asumsi independensi antar fitur. Untuk binary classification, model ini memprediksi kelas dari data baru berdasarkan probabilitas fitur-fitur yang ada.

1. Menghitung probabilitas prior

Probabilitas prior adalah probabilitas awal dari setiap kelas sebelum melihat data fitur. Dalam binary classification, ada dua kelas C_0 dan C_1 . Probabilitas prior dihitung sebagai rasio jumlah data dari setiap kelas terhadap total jumlah data.

2. Menghitung probabilitas likelihood.

Untuk setiap fitur, model ini mengasumsikan bahwa fitur-fitur tersebut terdistribusi secara Gaussian. Oleh karena itu, kita menghitung probabilitas likelihood fitur dengan menggunakan distribusi normal.

3. Memprediksi kelas

Untuk databaru, kita menghitung probabilitas posterior untuk setiap kelas menggunakan teorema Bayes. Kemudian, kelas dengan probabilitas posterior tertinggi dipilih sebagai prediksi

4. Evaluasi model

Model GNB dapat dievaluasi dengan menggunakan metrik seperti accuracy, precision, recall, dan F1-score. Evaluasi ini dilakukan dengan menggunakan data uji yang belum pernah dilihat oleh model selama pelatihan.

2. Perbandingan model buatan sendiri dan library

Berdasarkan hasil evaluasi, kedua model memiliki nilai Accuracy yang sama, yaitu 0.86. Ini menunjukkan bahwa kedua model memiliki performa yang setara dalam metrik yang diukur.

3. Improvement yang bisa dilakukan

- Handling zero variance

Tambahkan smoothing (misalnya Laplace smoothing) untuk menangani kasus di mana varians dari suatu fitur adalah 0, yang dapat menyebabkan pembagian dengan nol.

- Numerical stability

Saat menghitung log-likelihood, tambahkan epsilon kecil ke varians untuk mencegah pembagian dengan nol dan meningkatkan stabilitas numerik.

- Kalibrasi

Rizqika Mulia Pratama
13522126

Kalibrasi probabilitas prediksi menggunakan metode seperti Platt scaling atau isotonic regression untuk mendapatkan probabilitas yang lebih akurat.