

Gaussian Naïve Bayes

Procedure 方法

Gaussian Naïve Bayes adalah model machine learning yang menggunakan bayes theorem dan gaussian distribution untuk melakukan prediksi. Naïve disini adalah karena model mengasumsi semua fitur independen satu sama lain, dan mengikuti distribusi gaussian.

1. **Inisialisasi Prior**

Model menghitung probabilitas prior, yaitu probabilitas munculnya kelas target.

2. **Hitung probabilitas fitur berdasarkan kelas**

Setiap fitur dihitung probabilitasnya yang sebelumnya sudah dipisah berdasarkan kelas.

3. **Prediksi y**

Prediksi dilakukan dengan mengalikan semua nilai probabilitasnya berdasarkan nilai fitur yang dimilikinya. Hasil kelas yang diprediksi adalah kelas yang memiliki nilai probabilitas terbesar setelah perkalian tersebut.

VS Sklearn

Hasil di notebook DoE menghasilkan nilai F1 yang jauh lebih rendah dibandingkan sklearn, walaupun testing terpisah lainnya menghasilkan nilai yang lebih mirip. Hal ini mungkin dikarenakan implementasi sklearn dapat handle fitur dengan distribusi yang tidak gaussian seperti one-hot features, atau adanya kesalahan dalam perhitungan probabilitas pada implementasi from scratch.

Potential Improvements

- Perbaikan dari perhitungan probabilitas