

Gaussian Naïve Bayes:

1. Pertama, model menghitung beberapa statistic penting untuk tiap kelas, seperti rata-rata, variansi, dan probabilitas sebuah kelas muncul. Kemudian, dari data statistic sebelumnya, akan dibuat distribusi gaussian/ distribusi normalnya untuk setiap kelas. Setelah itu, akan dihitung Kemungkinan suatu sampel berada dalam suatu kelas. Kelas dengan probabilitas tertinggi akan dipilih sebagai output dari hasil prediksi. Proses ini dilakukan untuk semua data atau sampel lain yang perlu diprediksi.
2. Model Gaussian Naïve Bayes dari scikit-learn memiliki akurasi yang lebih tinggi. Hal ini karena Gaussian NB scikit learn penanganan *edge-case* yang jauh lebih baik, seperti penanganan pembagian dengan nol dengan menggunakan epsilon serta penanganan underflow Ketika probabilitas yang dihasilkan sangat kecil.
3. Hal yang dapat ditingkatkan dari model saat ini adalah dengan mengoptimasi algoritmanya sehingga waktu eksekusi yang dibutuhkan bisa diminimalisir. Selain itu, penanganan *edge-case*, seperti penggunaan nilai epsilon yang tepat untuk mencegah pembagian dengan nol, juga diperlukan sehingga akurasi prediksi yang dihasilkan bisa mendekati model bawaan scikit-learn. Model ini adalah model yang menghasilkan tingkat akurasi terendah pada dataset ini, hal ini disebabkan oleh penggunaan one-hot encoding yang menyebabkan *curse of dimensionality* sehingga model ini menjadi susah untuk dilatih. Solusinya

adalah dengan mengganti one-hot encoder dengan label encoding. Namun, hal ini tidak dilakukan untuk dataset ini karena label encoding digunakan jika ada hubungan hierarkikal pada suatu fitur di suatu dataset.