

Gaussian Naive Bayes

Cara Kerja:

1. Hitung rata-rata dan variansi pada setiap fitur untuk setiap kelas.
2. Hitung probabilitas posterior menggunakan Teorema Bayes.
3. Klasifikasikan data baru dengan nilai probabilitas posterior tertinggi.

Perbandingan:

- *Scratch*

```
[31]: from gnb import GaussianNaiveBayesScratch

gnb_scratch = GaussianNaiveBayesScratch()
gnb_scratch.fit(X_train, y_train)
y_pred_gnb_scratch = gnb_scratch.predict(X_test)

validate_model(gnb_scratch, method_name="Gaussian Naive Bayes from Scratch")
```

Hold-Out Validation (Gaussian Naive Bayes from Scratch):
F1 Score: 0.8888888888888888

	precision	recall	f1-score	support
0	0.92	0.96	0.94	72
1	0.92	0.86	0.89	42
accuracy			0.92	114
macro avg	0.92	0.91	0.91	114
weighted avg	0.92	0.92	0.92	114

K-Fold Cross-Validation (Gaussian Naive Bayes from Scratch):
F1 Scores for each fold: [0.8955223880597015, 0.9846153846153847, 0.9014084507042254, 0.8888888888888888, 0.9]
Mean F1 Score: 0.9140870224536402
Standard Deviation of F1 Score: 0.03553317277632346

- *Library*

```
[32]: from sklearn.naive_bayes import GaussianNB

gnb = GaussianNB()
gnb.fit(X_train, y_train)
y_pred_gnb = gnb.predict(X_test)

validate_model(gnb, method_name="Gaussian Naive Bayes from Library")
```

Hold-Out Validation (Gaussian Naive Bayes from Library):
F1 Score: 0.9090909090909091

	precision	recall	f1-score	support
0	0.91	1.00	0.95	72
1	1.00	0.83	0.91	42
accuracy			0.94	114
macro avg	0.96	0.92	0.93	114
weighted avg	0.94	0.94	0.94	114

K-Fold Cross-Validation (Gaussian Naive Bayes from Library):
F1 Scores for each fold: [0.9230769230769231, 0.9846153846153847, 0.8985507246376812, 0.9014084507042254, 0.9]
Mean F1 Score: 0.9215302966068428
Standard Deviation of F1 Score: 0.03279824222386503

Dari implementasi secara *scratch* dan *library*, terlihat bahwa penggunaan *library* memiliki nilai F1 score yang lebih tinggi. Ini juga didukung dengan presisi memprediksi diagnosis kanker(1) dengan skor 1.

Improvement:

Improvement yang dapat dilakukan pada algoritma *Gaussian Naive Bayes* secara *scratch* dapat dilakukan dengan menggunakan vektorisasi saat melakukan perhitungan matematis, terutama untuk data yang memiliki variansi sangat kecil.