

Nama : Devi Intan Nurisma Putri

NIM : 123220184

Makul : Bioinformatika

Kelas : IF-A

1. Supervised learning menggunakan data berlabel untuk melatih model dalam melakukan Prediksi. Sedangkan unsupervised learning memproses data tanpa label untuk Menemukan pola atau struktur tersembunyi.

2. Confusion matrix

	Predicted Positif	Predicted Negatif
Actual Positif	4 (TP)	1 (FN)
Actual Negatif	2 (FP)	3 (TN)

$$\text{Accuracy} = (TP + TN) / \text{Total} = 7/10 = 0.7 (70\%)$$

$$\text{Precision} = TP / (TP + FP) = 4/4+2 = 0.6667 (66.67\%)$$

$$\text{Recall} = TP / (TP + FN) = 4/(4+1) = 0.8 (80\%)$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) = 2 \times (0.6667 \times 0.8) / (0.6667 + 0.8) \\ = \approx 0.727$$

Insight :

- Model memiliki akurasi sedang (70%)
- Recall tinggi menunjukkan model cukup baik dalam mengenali kasus positif (covid-19)
- Precision sedang mengindikasikan beberapa negatif diprediksi salah sebagai positif (False positive)
- Cocok untuk kasus dimana deteksi awal lebih penting daripada kesalahan positif (seperti covid-19)

3. Problem : Mencari senyawa herbal Indonesia potensial sebagai terapi pendukung covid-19 menggunakan pendekatan machine learning dan farmakofor

Data : 223 senyawa dari database herbal Indonesia.

Label : aktivitas antivirus (aktif atau tidak).

Fitur : molekuler (deskriptor kimia)

Metode : Random forest : Hyperparameter : jumlah pohon (n-estimators), kedalaman maksimum.

- Support Vector Machine (SVM) : Hyperparameter : Kernel (RBF), nilai C, gamma
- Naive Bayes
- K-Nearest Neighbors (K-NN) : Hyperparameter : Jumlah tetangga (K).

No. _____

Date _____

Fold



Ukuran Performa model :

- Accuracy : Ketepatan keseluruhan prediksi.
- Precision : Proporsi prediksi positif yang benar
- Recall : Kemampuan mendeteksi data positif
- F1 Score : harmonisasi antara precision dan recall.
- AUC-ROC : Kemampuan memisahkan kelas.

Model Terbaik : Random Forest memberikan performa terbaik, karena memiliki keseimbangan tinggi antara precision dan recall, serta nilai AUC yang unggul. Sehingga cocok untuk data dengan dimensi tinggi dan kompleksitas fitur kimia.