

Random Forest

Teknik Ensemble untuk Klasifikasi
dan Regresi

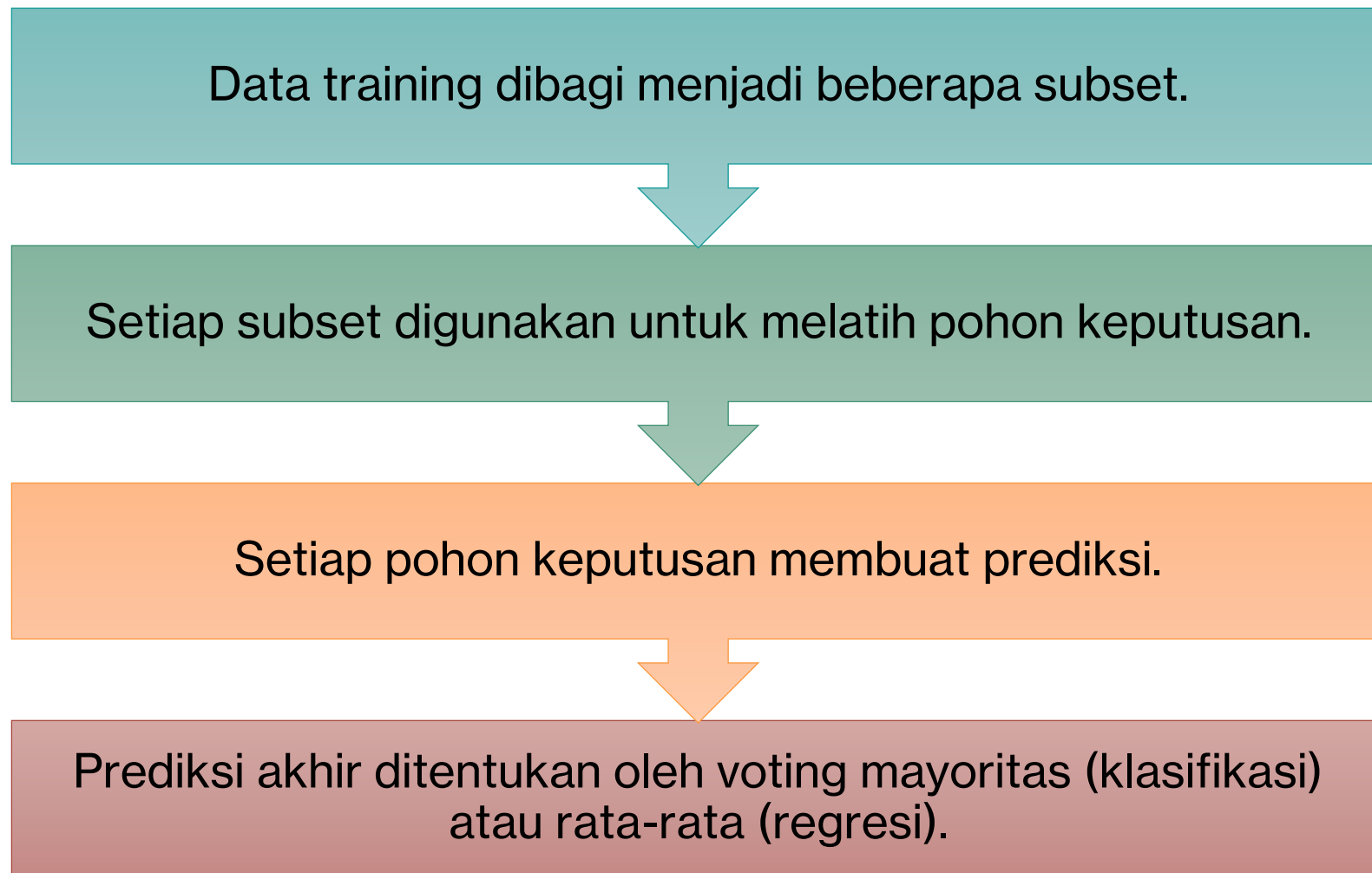
Ledy Elsera Astrianty, S.Kom., M.Kom

Random Forest

Teknik Ensemble, merupakan teknik dalam machine learning yang menggabungkan beberapa model untuk meningkatkan akurasi prediksi.

Random Forest, merupakan teknik ensemble yang menggunakan banyak pohon keputusan untuk meningkatkan akurasi prediksi.


Cara Kerja Random Forest



Kelebihan Random Forest

- 
- Mengurangi overfitting dengan menggunakan banyak pohon keputusan.

- 
- Dapat menangani data dengan dimensi tinggi.

- 
- Dapat menangani data dengan missing value.

- 
- Akurasi prediksi yang tinggi.

Kelebihan Random Forest



Menggabungkan prediksi dari beberapa model, sehingga dapat mengurangi efek noise.



Menggunakan median atau mean yang robust, sehingga dapat menangani outlier dengan baik.



Mampu memilih fitur yang relevan dan mengabaikan fitur yang tidak relevan, sehingga dapat mengurangi efek noise.

Kapan menggunakan Random Forest

Data dengan Dimensi Tinggi

- Random Forest sangat cocok untuk data dengan dimensi tinggi, karena dapat menangani fitur yang banyak dengan baik.

Data yang Mengandung Noise

- Random Forest dapat menangani data yang mengandung noise atau outlier dengan baik.

Data yang Hilang

- Random Forest dapat menangani data yang hilang dengan baik, bahkan jika beberapa fitur memiliki nilai yang hilang.

Evaluasi Model Random Forest

Akurasi (Accuracy)

- Mengukur proporsi sampel yang diklasifikasikan dengan benar.

Presisi (Precision)

- Mengukur proporsi sampel yang diklasifikasikan sebagai positif yang benar-benar positif.

Recall

- Mengukur proporsi sampel positif yang diklasifikasikan dengan benar.

F1-Score

- Mengukur keseimbangan antara presisi dan recall.

Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

- Mengukur kemampuan model untuk membedakan antara kelas positif dan negatif.

Evaluasi Model Random Forest

Mean Squared Error (MSE)

- Mengukur rata-rata kesalahan kuadrat antara nilai prediksi dan nilai sebenarnya.

Mean Absolute Error (MAE)

- Mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai sebenarnya.

Root Mean Squared Error (RMSE)

- Mengukur akar kuadrat dari rata-rata kesalahan kuadrat antara nilai prediksi dan nilai sebenarnya.

R-Squared (R^2)

- Mengukur proporsi variansi dalam data yang dijelaskan oleh model.



Thank You

Random Forest

Sains Data