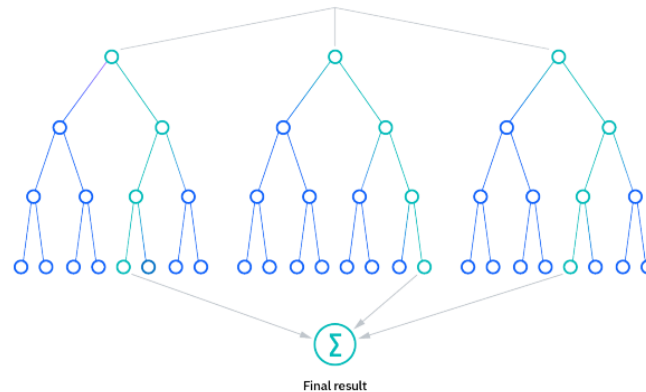


Mohamad Rayi Dwi Putra

1103213070

TK4504

Random Forest



Random Forest adalah metode pembelajaran mesin yang efektif, menggabungkan beberapa pohon keputusan untuk menghasilkan prediksi yang lebih kuat dan akurat. Untuk membangun Hutan Acak, beberapa langkah penting perlu dilakukan:

- Persiapan Data: Persiapkan data untuk tugas yang akan dilakukan, baik itu klasifikasi atau regresi.
- Penyetelan Hyperparameter: Atur parameter utama seperti:
 - Jumlah pohon ($n_estimators$): Nilai yang lebih tinggi umumnya meningkatkan kinerja tetapi membutuhkan waktu pelatihan yang lebih lama.
 - Kedalaman maksimum pohon (max_depth): Mengontrol kompleksitas pohon. Pohon yang lebih dalam dapat mengalami overfitting, sementara pohon yang dangkal mungkin mengalami underfitting.
 - Jumlah fitur yang dipertimbangkan pada setiap pemisahan ($max_features$): Memilih subset fitur secara acak pada setiap pemisahan dapat mengurangi overfitting.
 - Kriteria untuk pemisahan ($criterion$): Gini impurity (klasifikasi) atau mean squared error (regresi).
- Pelatihan Model: Algoritme membangun sejumlah pohon keputusan yang ditentukan, masing-masing dilatih pada subset data acak (bootstrapping) dan menggunakan subset fitur acak di setiap pemisahan.
- Prediksi: Untuk titik data baru, setiap pohon dalam forest membuat prediksi, dan prediksi akhir adalah voting mayoritas (klasifikasi) atau rata-rata (regresi) dari prediksi masing-masing pohon.

Menggunakan Random Forest:

- Klasifikasi: Memprediksi label kelas dari titik data baru berdasarkan voting mayoritas dari seluruh pohon.
- Regresi: Memprediksi nilai kontinu untuk data baru berdasarkan prediksi rata-rata dari seluruh pohon.
- Pentingnya Fitur: Menganalisis penurunan rata-rata ketidakmurnian di semua pohon untuk mengetahui fitur mana yang paling berkontribusi terhadap prediksi model.

Lalu Mengevaluasi Random Forest:

- Metrik Akurasi:
 - Klasifikasi: Akurasi, presisi, recall, F1-score, matriks kebingungan, ROC AUC.
 - Regresi: Mean squared error (MSE), R-squared.
- Validasi Silang (Cross-Validation): Memisahkan data menjadi set pelatihan dan validasi untuk mengestimasi kemampuan generalisasi model ke data yang belum pernah dilihat.
- Analisis Kesalahan: Menganalisis prediksi yang salah atau tidak akurat untuk mengidentifikasi potensi bias atau area untuk perbaikan.