

Nama : Syauqi Atrobi

NIM : 221011400029

Kelas : 05TPLP001

1. Alur Umum Penggunaan Random Forest

Algoritma Random Forest umumnya digunakan dalam masalah klasifikasi dan regresi. Berikut adalah alur umum penggunaannya:

1. Pengumpulan Data:

- Kumpulkan data yang relevan dengan masalah yang ingin diselesaikan.
- Pastikan data memiliki fitur-fitur yang dapat digunakan untuk membuat model prediksi.

2. Preprocessing Data:

- **Pembersihan Data:** Hapus data yang duplikat, hilang, atau tidak konsisten.
- **Transformasi Data:** Ubah data ke dalam format yang dapat diproses oleh algoritma, misalnya mengubah data kategorikal menjadi numerik.
- **Normalisasi Data:** Skala data agar memiliki rentang yang sama, sehingga fitur-fitur tidak didominasi oleh fitur dengan nilai yang sangat besar.
- **Pemilihan Fitur:** Pilih fitur-fitur yang paling relevan dengan target variabel.

3. Pembuatan Model:

- **Pembagian Data:** Bagi data menjadi data latih dan data uji.
- **Pembuatan Pohon Keputusan:** Random Forest membangun banyak pohon keputusan. Setiap pohon dibuat dengan sampel data yang berbeda secara acak dan fitur-fitur yang dipilih secara acak.
- **Penggabungan Prediksi:** Prediksi dari setiap pohon digabungkan melalui voting (untuk klasifikasi) atau rata-rata (untuk regresi) untuk mendapatkan prediksi akhir.

4. Evaluasi Model:

- **Metrik Evaluasi:** Gunakan metrik seperti akurasi, presisi, recall, F1-score (untuk klasifikasi), atau Mean Squared Error (MSE) (untuk regresi) untuk mengukur kinerja model pada data uji.
- **Visualisasi:** Gunakan visualisasi seperti matriks konfusi atau kurva ROC untuk memahami kinerja model secara lebih baik.

5. Penyetelan Hyperparameter:

- **Tuning Hyperparameter:** Sesuaikan parameter-parameter model seperti jumlah pohon, kedalaman pohon, dan jumlah fitur yang dipilih secara acak untuk meningkatkan kinerja model.

- **Cross-Validation:** Gunakan teknik cross-validation untuk mengevaluasi kinerja model secara lebih objektif.

Hasil dan Evaluasi

- **Hasil:** Hasil dari Random Forest adalah model prediksi yang dapat digunakan untuk memprediksi nilai baru.
- **Evaluasi:** Evaluasi bertujuan untuk mengukur seberapa baik model tersebut dalam melakukan prediksi. Metrik evaluasi yang digunakan akan bergantung pada jenis masalah (klasifikasi atau regresi) dan tujuan dari model.

Contoh Kasus dan Referensi

- **Analisis Sentimen:** Dalam penelitian "Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based dan Random Forest" (Universitas Putera Batam), Random Forest digunakan untuk mengklasifikasi sentimen teks menjadi positif, negatif, atau netral. Preprocessing yang dilakukan meliputi normalisasi, stemming, dan penggunaan stopwords. Hasilnya menunjukkan akurasi model mencapai 87,23%.
- **Deteksi Sarkasme:** Penelitian lain menggunakan Random Forest untuk mendeteksi sarkasme pada teks. Preprocessing meliputi case folding, tokenisasi, dan penghapusan stopwords. Hasilnya menunjukkan bahwa Random Forest dapat mendeteksi sarkasme dengan akurasi yang cukup baik.

Referensi:

- **Jurnal Universitas Putera Batam:**
<https://ejournal.upbatam.ac.id/index.php/jif/article/download/7987/3319>
- **Jurnal Djtechno:**
<https://jurnal.dharmawangsa.ac.id/index.php/djtechno/article/view/3361/0>
- **Jurnal Komputasi:**
<https://komputasi.fmipa.unila.ac.id/index.php/komputasi/article/download/166/166/191>

2. Berikut 3 jurnal internasional yang bersumber dari ieeexplore.ieee.org

1. The Effective Breast Cancer Classification with the Random Forest Algorithm
LINK : <https://ieeexplore.ieee.org/document/10364529>
2. Chronic Kidney Disease for Collaborative Healthcare Data Analytics using Random Forest Classification Algorithms
LINK : <https://ieeexplore.ieee.org/document/9402574>
3. Optimizing Lung Cancer Classification Prediction based on Random Forest and SMOTE Techniques
LINK : <https://ieeexplore.ieee.org/document/10718616>