

LAPORAN ANALISIS DATA DAN IMPLEMENTASI MACHINE LEARNING

Topik: Klasifikasi Menggunakan Algoritma Random Forest

Dosen pengampu : Desi Anggreani, S.kom.,MT



Oleh :

Nama : PUTRI ARFINA

Nim : 105841116123

Kelas : 5 AI - A

**FAKULTAS TEKNIK
PROGRAM STUDI INFORMATIKA
UNIVERSITAS MUHAMMADIYAH MAKASSAR
2026**

1. Pendahuluan

Dalam era transformasi digital saat ini, pemanfaatan kecerdasan buatan menjadi sangat krusial, terutama dalam hal klasifikasi data yang kompleks. Proyek ini memfokuskan pada pengujian algoritma Random Forest, sebuah metode pembelajaran mesin yang bekerja dengan cara membangun sekumpulan pohon keputusan (*decision trees*) saat masa pelatihan. Alasan utama pemilihan algoritma ini adalah kemampuannya dalam menangani dataset yang memiliki banyak fitur serta keunggulannya dalam mencegah terjadinya overfitting dibandingkan dengan satu pohon keputusan tunggal. Melalui repository GitHub "*mathematic-for-AI*", proyek ini mendokumentasikan bagaimana prinsip matematika fundamental seperti probabilitas dan information gain diintegrasikan ke dalam baris kode pemrograman untuk menghasilkan model prediksi yang akurat.

2. Analisis data mentah

Proses pengolahan data dimulai dengan tahap eksplorasi terhadap file data mentah.xlsx yang menjadi basis utama penelitian ini. Dataset ini terdiri dari berbagai atribut input yang kemudian harus melalui tahap pra-pemrosesan (*preprocessing*) yang ketat untuk memastikan kualitas model yang dihasilkan. Hasil dari pembersihan data tersebut kemudian disimpan ke dalam file data_bersih_random_forest.xlsx, di mana data telah diubah menjadi format numerik yang siap diolah oleh algoritma. Tahapan ini sangat penting karena data mentah seringkali mengandung ketidakakonsistensi yang dapat menurunkan akurasi model, sehingga transformasi data menjadi langkah awal yang tidak boleh dilewatkan dalam siklus pengembangan AI.

3. Implementasi Program dan Algoritma

Model klasifikasi ini diimplementasikan menggunakan bahasa pemrograman Python melalui skrip random_forest.py yang memanfaatkan pustaka scikit-learn. Di dalam program tersebut, dataset yang telah bersih dibagi menjadi data latih dan data uji, di mana model dilatih untuk mengenali pola-pola matematis dari fitur-fitur yang tersedia. Program ini bekerja dengan membangun banyak pohon keputusan secara acak dan kemudian menggunakan mekanisme voting untuk menentukan hasil akhir klasifikasi. Dengan pendekatan ini, variansi model dapat dikurangi secara signifikan, sehingga menghasilkan prediksi yang lebih stabil dan handal ketika dihadapkan pada data baru yang belum pernah dilihat sebelumnya oleh sistem.

4. Analisis Perhitungan Manual dan Kesimpulan

Salah satu keunikan dari proyek ini adalah dilakukannya verifikasi matematis secara manual untuk memastikan bahwa logika yang dijalankan oleh komputer selaras dengan teori dasar. Proses ini dilakukan melalui lima iterasi perhitungan yang tercatat secara detail, mulai dari iterasi pertama hingga kelima, yang bertujuan untuk mensimulasikan pembentukan setiap pohon di dalam *forest*. Berdasarkan hasil akhir dari perhitungan manual tersebut, ditemukan bahwa mayoritas pohon memberikan hasil klasifikasi yang konsisten satu sama lain. Sebagai kesimpulan, proyek ini berhasil membuktikan bahwa algoritma Random Forest mampu memberikan hasil prediksi yang valid, di mana kesesuaian antara perhitungan manual dan output dari program Python menunjukkan bahwa model telah terkalibrasi dengan benar secara matematis dan teknis.