

PROYEK DATA MINING
Klasifikasi Risiko Diabetes Menggunakan
Algoritma Decision Tree

Dosen :

Anna Baita, S.Kom., M.Kom.



Kelas : IF04

Nama : Kevin Rizki Irawan

Nim : 22.11.4870

PROGRAM STUDI SARJANA INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
2025

1. Cerita Singkat Proyek Ini

Saya ngerjain proyek ini supaya orang bisa cek sendiri risiko diabetes mereka secara mandiri. Caranya simpel, cukup masukan data seperti umur, BMI, dan kadar glukosa. Hasilnya langsung ditampilkan di aplikasi web yang saya buat pakai Streamlit.

Saya pakai algoritma Decision Tree karena menurut saya gampang dimengerti dan cukup cocok buat data jenis ini.

2. Dataset yang Digunakan

Dataset yang saya pakai adalah diabetes.csv, yang isinya data-data pasien dengan fitur seperti:

- Umur
- BMI (indeks massa tubuh)
- Glukosa

Dataset ini saya pisah jadi dua bagian:

- 80% untuk melatih model
 - 20% untuk ngetes hasilnya
Supaya bisa lebih objektif dan adil waktu evaluasi.
-

3. Algoritma dan Proses Pembentukan Model

Saya pilih Decision Tree Classifier karena lebih mudah dijelasin dan bisa menangani data numerik maupun kategorikal.

Langkah-langkah saya:

- Bersihin data yang kosong atau gak valid
- Lakuin scaling biar nilai tiap fitur seimbang
- Bagi data jadi training dan testing
- Latih modelnya pakai DecisionTreeClassifier dari Scikit-learn
- Cek hasilnya pakai confusion matrix, ROC AUC, dan classification report

Hasilnya lumayan bagus dan cukup layak buat prediksi awal.

4. Desain UI/UX & Wireframe

Saya bikin aplikasi ini tampilannya simpel dan gampang dipakai. Ada dua tampilan utama:

- Halaman Input: buat masukan data pengguna (umur, BMI, glukosa)
- Halaman Hasil: langsung nampilin risiko diabetes (Rendah / Sedang / Tinggi)

Wireframe-nya juga udah saya desain dan simpan dalam bentuk PDF.

')) [Link ke Wireframe PDF \(GitHub\)](#)

5. Deployment (Biar Bisa Diakses Online)

Aplikasinya saya deploy pakai Streamlit, jadi orang lain bisa langsung coba lewat browser. Tampilannya saya buat sederhana dan bersih supaya gak ribet dipakai.

³ [Link Aplikasi Streamlit](#)

⁴ [Link Repository GitHub](#)

6. Evaluasi Model

Model saya diuji pakai beberapa metrik, di antaranya:

- ROC AUC Score: ngukur seberapa baik model bisa bedain antara dua kelas
- Confusion Matrix: ngebandingin hasil prediksi sama kenyataan
- Classification Report: nyertain precision, recall, dan F1-score

Menurut saya, hasil model ini udah cukup oke buat jadi awal, meskipun kalau datanya lebih banyak dan bervariasi pasti bisa lebih bagus lagi.

7. Kesimpulan

Dari proyek ini, saya jadi lebih paham gimana proses kerja data mining dari awal banget sampai bisa ditampilin di web.

Saya juga belajar tentang pentingnya pemilihan algoritma yang tepat dan gimana cara nge-deploy aplikasi secara online.

Walaupun hasilnya belum sempurna, saya harap aplikasi ini bisa bantu ngasih gambaran awal ke orang-orang yang penasaran soal risiko diabetes mereka. Tapi tetap ya, untuk hasil yang pasti dan akurat, sebaiknya tetap periksa ke dokter.