

**LAPORAN UJIAN AKHIR SEMESTER**

**PROYEK DATA MINING**

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



**Disusun Oleh:**

Nama : Rizky Nanda Anggia

Kelas : 22S1IF-ProyekD4(ST167)

NIM : 22.11.4825

Nomor Kursi : 137

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**2025**

# Analisis Performa Model dan Deployment

## 1. Analisis Performa Model

Model yang saya kembangkan adalah model klasifikasi diabetes tipe 2 menggunakan algoritma Random Forest yang dioptimalkan dengan teknik SMOTE.

- **Penanganan Data Tidak Seimbang:** Saya berhasil mengatasi masalah ketidakseimbangan data pada dataset diabetes\_data.csv. Awalnya, data menunjukkan ketidakseimbangan yang signifikan antara 1127 sampel non-diabetes (kelas 0) dan 752 sampel diabetes (kelas 1). Setelah menerapkan SMOTE, jumlah sampel pada kedua kelas berhasil diseimbangkan menjadi masing-masing 1127, yang sangat penting untuk mencegah model menjadi bias terhadap kelas mayoritas.
- **Pemilihan Fitur:** Model saya menggunakan 10 fitur terpenting yang diidentifikasi dari dataset, seperti FastingBloodSugar, HbA1c, SleepQuality, CholesterolHDL, FatigueLevels, CholesterolLDL, MedicationAdherence, QualityOfLifeScore, DiastolicBP, dan Age. Analisis korelasi menunjukkan bahwa FastingBloodSugar memiliki korelasi positif terkuat dengan diagnosis, yang konsisten dengan pengetahuan medis.
- **Optimalisasi dan Hasil:** Model Random Forest dioptimalkan menggunakan GridSearchCV. Model yang telah di-*tuning* menunjukkan performa yang sangat baik pada data uji dengan akurasi 89,80%. Evaluasi lebih lanjut dari *classification report* menunjukkan:
  - **Presisi (Kelas 1 - Diabetes):** 0.94. Ini berarti 94% dari semua pasien yang diprediksi diabetes oleh model saya memang benar-benar diabetes.
  - **Recall (Kelas 1 - Diabetes):** 0.85. Ini menunjukkan model saya berhasil mengidentifikasi 85% dari semua kasus diabetes yang ada di data uji, yang merupakan metrik penting untuk diagnosis dini.

## 2. Analisis Deployment

Model yang telah dilatih dan dioptimalkan berhasil saya *deploy* sebagai aplikasi web yang dapat diakses publik.

- **Komponen Aplikasi:**

- **Backend:** Saya menggunakan framework Flask dan Python untuk logika pemrosesan data dan prediksi. Model dan objek scaler disimpan dalam format .pkl untuk penggunaan kembali.
- **Frontend:** Saya menggunakan Tailwind CSS untuk tampilan yang intuitif dan responsif. Antarmuka ini memiliki formulir input untuk 10 fitur dan menampilkan hasil prediksi di area terpisah.

- **Tampilan Aplikasi:**

**Klasifikasi Diabetes Tipe 2**  
Gunakan AI untuk memprediksi kemungkinan diabetes berdasarkan data klinis dan gaya hidup

**Prediktor Risiko Diabetes**  
Masukkan data klinis dan gaya hidup untuk mendapatkan prediksi.

Gula Darah Puasa (mg/dL) Contoh: 90.0	Kadar HbA1c (%) Contoh: 5.7
Kualitas Tidur (1-10) Contoh: 8.0	Kolesterol HDL (mg/dL) Contoh: 40.0
Tingkat Kelelahan (1-10) Contoh: 3.0	Kolesterol LDL (mg/dL) Contoh: 100.0
Kepatuhan Minum Obat (1-10) Contoh: 9.0	Skor Kualitas Hidup (0-100) Contoh: 75.0
Tekanan Darah Diastolik (mmHg) Contoh: 80.0	Usia (Tahun) Contoh: 45

**Dapatkan Prediksi**

Output Hasil Prediksi Akan Muncul di Sini

© 2025 - Dibuat untuk Proyek Data Mining oleh Rizky Nanda Anggia (22.11.4825)  
Universitas AMIKOM Yogyakarta

**Gambar 1:** Menunjukkan tampilan awal aplikasi web "Prediktor Risiko Diabetes" dengan 10 kolom input, sebuah tombol "Dapatkan Prediksi", dan area output yang masih kosong.

**Klasifikasi Diabetes Tipe 2**  
Gunakan AI untuk memprediksi kemungkinan diabetes berdasarkan data klinis dan gaya hidup

### Prediktor Risiko Diabetes

Masukkan data klinis dan gaya hidup untuk mendapatkan prediksi.

Gula Darah Puasa (mg/dL) Contoh: 90.0	Kadar HbA1c (%) Contoh: 5.7
Kualitas Tidur (1-10) Contoh: 8.0	Kolesterol HDL (mg/dL) Contoh: 40.0
Tingkat Kelelahan (1-10) Contoh: 3.0	Kolesterol LDL (mg/dL) Contoh: 100.0
Kepatuhan Minum Obat (1-10) Contoh: 9.0	Skor Kualitas Hidup (0-100) Contoh: 75.0
Tekanan Darah Diastolik (mmHg) Contoh: 80.0	Usia (Tahun) Contoh: 45

**Dapatkan Prediksi**

Hasil: Risiko Rendah Terkena Diabetes (Keyakinan: 87.00%)

© 2025 - Dibuat untuk Proyek Data Mining oleh Rizky Nanda Anggia (22.11.4823)  
Universitas AMIKOM Yogyakarta

**Gambar 2:** Menunjukkan contoh hasil prediksi untuk kasus "Risiko Rendah Terkena Diabetes". Hasil ditampilkan dengan teks berwarna hijau, disertai dengan persentase keyakinan model sebesar 87.00%.

**Klasifikasi Diabetes Tipe 2**  
Gunakan AI untuk memprediksi kemungkinan diabetes berdasarkan data klinis dan gaya hidup

### Prediktor Risiko Diabetes

Masukkan data klinis dan gaya hidup untuk mendapatkan prediksi.

Gula Darah Puasa (mg/dL) Contoh: 90.0	Kadar HbA1c (%) Contoh: 5.7
Kualitas Tidur (1-10) Contoh: 8.0	Kolesterol HDL (mg/dL) Contoh: 40.0
Tingkat Kelelahan (1-10) Contoh: 3.0	Kolesterol LDL (mg/dL) Contoh: 100.0
Kepatuhan Minum Obat (1-10) Contoh: 9.0	Skor Kualitas Hidup (0-100) Contoh: 75.0
Tekanan Darah Diastolik (mmHg) Contoh: 80.0	Usia (Tahun) Contoh: 45

**Dapatkan Prediksi**

Hasil: Berisiko Tinggi Terkena Diabetes (Keyakinan: 94.50%)

© 2025 - Dibuat untuk Proyek Data Mining oleh Rizky Nanda Anggia (22.11.4823)  
Universitas AMIKOM Yogyakarta

**Gambar 3:** Menunjukkan contoh hasil prediksi untuk kasus "Risiko Tinggi Terkena Diabetes" dengan teks berwarna merah, disertai dengan persentase keyakinan model sebesar 94.50%.

- **Validasi *Deployment*:** Model yang saya muat ulang ke dalam aplikasi web telah divalidasi dan akurasi mencapai 90%. Hal ini menegaskan bahwa proses *deployment* tidak mengurangi kinerja model.

## Kesimpulan

Secara keseluruhan, proyek ini berhasil menciptakan sebuah model klasifikasi yang efektif dan terintegrasi dalam sebuah aplikasi web yang fungsional. Peningkatan akurasi model melalui penggunaan SMOTE dan GridSearchCV, serta keberhasilan *deployment* ke PythonAnywhere menunjukkan pemahaman yang mendalam tentang siklus proyek *data mining*.

## Saran dan Pengembangan Lanjutan:

- **Dataset:** Saya akan mempertimbangkan penggunaan dataset yang lebih dinamis dan bervariasi untuk menguji ketahanan model.
- **Algoritma:** Eksplorasi perbandingan dengan algoritma lain seperti XGBoost atau SVM dapat dilakukan untuk melihat potensi peningkatan performa.
- **Interpretasi Model:** Saya akan melakukan analisis interpretasi model yang lebih mendalam menggunakan teknik seperti SHAP atau LIME untuk memahami lebih lanjut pengaruh setiap fitur.