

IMPLEMENTASI MODEL KLASIFIKASI UNTUK DETEKSI RESIKO DIABETES



Disusun Oleh :

SIANE SANTOSO

A11.2022.13956

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI TEKNIK INFORMATIKA

UNIVERSITAS DIAN NUSWANTORO SEMARANG

2024

A. Deskripsi Singkat

Diabetes adalah kondisi kesehatan kronis yang mempengaruhi jutaan orang di seluruh dunia. Penyakit ini terjadi ketika kadar glukosa darah seseorang meningkat di luar batas normal, menyebabkan berbagai masalah kesehatan serius, termasuk gangguan jantung, kerusakan saraf, gangguan penglihatan, hingga komplikasi yang mengancam nyawa jika tidak ditangani dengan baik. Peningkatan angka kejadian diabetes juga membawa dampak ekonomi yang signifikan karena perawatan jangka panjang dan pengelolaan penyakit ini memerlukan sumber daya kesehatan yang besar.

Salah satu tantangan utama dalam mengatasi penyebaran diabetes adalah deteksi dini. Seringkali, diabetes berkembang secara perlahan dan tidak menunjukkan gejala yang jelas hingga mencapai tahap yang lebih serius. Akibatnya, banyak orang yang tidak menyadari risiko mereka, dan hal ini menyebabkan kurangnya tindakan pencegahan yang dapat dilakukan sejak dini. Pengembangan sistem prediktif yang dapat mendeteksi risiko diabetes berdasarkan data klinis umum menjadi solusi untuk mendukung deteksi dini dan pencegahan penyakit ini.

Dalam penelitian ini, saya berusaha membangun model prediksi yang mampu mendeteksi risiko diabetes dengan cepat dan akurat. Model ini dikembangkan dengan memanfaatkan dataset yang memuat informasi yang relevan, seperti usia, jenis kelamin, riwayat hipertensi, riwayat penyakit jantung, dan beberapa parameter fisik lainnya yang berhubungan dengan kondisi diabetes. Dengan model ini, diharapkan dapat diperoleh wawasan yang bermanfaat untuk memahami serta mengidentifikasi risiko diabetes secara lebih dini dan mudah.

B. Masalah dan Tujuan yang Ingin Dicapai

1. Masalah

Diabetes adalah salah satu penyebab utama kematian yang berdampak serius pada kualitas hidup. Banyak orang tidak menyadari risiko diabetes yang mereka hadapi hingga penyakit

ini mencapai tahap lanjut, di mana komplikasi serius mungkin sudah terjadi.

2. Tujuan

Tujuan dari penelitian ini adalah untuk mengembangkan model prediksi yang dapat membantu mengidentifikasi faktor-faktor klinis yang berkaitan dengan risiko diabetes. Dengan model ini, diharapkan dapat diperoleh pemahaman lebih mendalam mengenai pengaruh faktor-faktor seperti usia, hipertensi, riwayat penyakit jantung, dan indeks massa tubuh (BMI) terhadap risiko diabetes. Model ini juga diharapkan memberikan wawasan yang bermanfaat bagi deteksi dini risiko diabetes.

C. Kerangka Eksperimen

1. Persiapan Dataset

a. Pengunduhan dan Eksplorasi Dataset

Dataset yang digunakan akan diambil dari Kaggle. Dataset ini berisi berbagai data klinis yang relevan dalam mendeteksi risiko diabetes. Pada tahap ini, dilakukan eksplorasi awal untuk memahami struktur dataset, melihat rentang nilai dari setiap fitur, serta mengidentifikasi pola-pola tertentu yang mungkin muncul pada individu dengan risiko tinggi atau rendah terhadap diabetes.

b. Analisis Awal Data

Langkah ini mencakup pemeriksaan distribusi data untuk setiap fitur, melihat korelasi antar fitur, dan mengidentifikasi pola tertentu yang bisa muncul pada data individu yang memiliki dan tidak memiliki diabetes. Hal ini bertujuan untuk memperoleh wawasan awal mengenai hubungan antara fitur-fitur dalam dataset dan label diabetes, yang dapat menjadi dasar dalam membangun model prediksi yang lebih akurat.

c. Pembersihan Data

Dalam proses ini, nilai yang hilang atau ekstrem (outlier) akan ditangani dengan metode seperti imputasi nilai median atau

pembuangan data yang sangat ekstrim, agar kualitas data lebih baik dan performa model dapat ditingkatkan.

2. Preprocessing Data

a. Normalisasi Data

Sebagian besar algoritma klasifikasi sensitif terhadap skala fitur. Oleh karena itu, semua fitur numerik dalam dataset dinormalisasi ke rentang yang sama. Normalisasi ini bertujuan agar proses analisis dan klasifikasi tidak didominasi oleh satu fitur tertentu yang memiliki rentang nilai lebih besar

b. Pembagian Data

Dataset kemudian dibagi menjadi data latih dan data uji untuk melatih model dan mengukur performanya pada data baru yang tidak terlihat selama pelatihan. Rasio pembagian ini akan ditentukan untuk memastikan bahwa model tidak mengalami overfitting atau underfitting.

3. Pemodelan dan Evaluasi

a. Pelatihan Model

Pada tahap ini, model dilatih dengan data latih. Model yang dibangun bertujuan untuk memprediksi risiko diabetes berdasarkan fitur-fitur yang telah dianalisis. Model ini akan melakukan klasifikasi apakah individu memiliki risiko diabetes atau tidak berdasarkan data yang dimasukkan.

b. Evaluasi Model

Setelah pelatihan selesai, model diuji pada data uji untuk mengevaluasi akurasi prediksi. Menggunakan metrik seperti akurasi untuk mengukur performa model, sehingga dapat dipastikan bahwa model mampu memberikan prediksi yang handal.

D. Penjelasan Dataset

1. Sumber Data

Dataset yang digunakan diambil dari platform Kaggle, yang menyediakan data terkait diabetes dari berbagai sumber. Dataset ini memuat data klinis serta informasi yang relevan untuk penelitian dan pengembangan model prediksi risiko diabetes.

2. Penjelasan Atribut dalam Dataset

a. Gender

Jenis kelamin individu, dikodekan sebagai 0 untuk pria dan 1 untuk wanita.

b. Age

Usia individu dalam tahun, yang merupakan salah satu faktor risiko penting untuk diabetes.

c. Hypertension

Indikator hipertensi, dengan 0 menunjukkan tidak ada riwayat hipertensi dan 1 menunjukkan ada riwayat hipertensi. Hipertensi berkaitan erat dengan peningkatan risiko diabetes.

d. heart_disease

Indikator riwayat penyakit jantung, dengan 0 untuk tidak ada riwayat dan 1 untuk ada riwayat. Penyakit jantung sering kali berhubungan dengan kondisi diabetes yang kurang terkontrol.

e. smoking_history

Riwayat merokok individu, dikodekan dalam beberapa kategori seperti tidak pernah merokok, mantan perokok, atau perokok aktif. Riwayat merokok seringkali dianggap sebagai faktor risiko tambahan untuk diabetes.

f. Bmi

Indeks massa tubuh, yang merupakan ukuran hubungan antara berat dan tinggi badan. BMI tinggi sering dikaitkan dengan risiko diabetes yang lebih tinggi.

g. HbA1c_level

Tingkat Hemoglobin A1c, yang menunjukkan kadar gula darah rata-rata dalam tiga bulan terakhir. HbA1c yang tinggi merupakan salah satu indikasi kuat dari diabetes.

h. `blood_glucose_level`

Tingkat glukosa darah saat ini, yang bisa menjadi indikasi risiko diabetes apabila kadarnya terlalu tinggi atau rendah.

i. Diabetes

Label target, dengan 0 untuk individu tanpa diabetes dan 1 untuk individu dengan diabetes. Label ini yang akan menjadi target prediksi dalam model.

E. TimeLine Pengerjaan :

1. Minggu 1

Menyiapkan dataset dengan mengunduh data dari Kaggle yang berisi informasi klinis terkait risiko diabetes. Melakukan eksplorasi awal untuk memahami struktur dataset, melihat rentang nilai dari setiap fitur, dan mencari pola-pola awal pada individu dengan dan tanpa risiko diabetes.

2. Minggu 2

Melakukan analisis awal data dengan memeriksa distribusi data untuk setiap fitur dan melihat korelasi antar fitur untuk menemukan hubungan yang mungkin berpengaruh pada risiko diabetes.

3. Minggu 3

Membersihkan data. Nilai yang hilang dan nilai ekstrem (outlier) ditangani, misalnya dengan cara mengisi nilai median atau menghapus data yang sangat ekstrem.

4. Minggu 4

Memproses data lebih lanjut. Semua fitur numerik dinormalisasi supaya berada dalam skala yang sama, sehingga analisis tidak didominasi oleh satu fitur tertentu. Setelah itu, dataset dibagi menjadi data latih dan data uji agar model bisa diuji pada data baru yang belum pernah dilihat selama pelatihan.

5. Minggu 5

Melatih model menggunakan data latih. Data latih digunakan untuk melatih model prediksi risiko diabetes berdasarkan fitur-fitur klinis.

6. Minggu 6

Menguji performa model yang sudah dilatih sebelumnya. Data uji digunakan untuk melihat akurasi prediksi model, dan performanya diukur menggunakan metrik seperti akurasi. Jika hasilnya belum memuaskan, model akan disesuaikan lagi agar hasilnya maksimal.

7. Minggu 7

Menyimpulkan hasil penelitian. Model yang sudah diuji dievaluasi lebih mendalam untuk melihat seberapa efektif model dalam mendeteksi risiko diabetes.