

## BAB III

### ANALISA SISTEM BERJALAN

#### 1.1. Penjelasan Dataset

Sumber data sekunder studi ini adalah "Diabetes Connection Care: Diabetes Health Analysis". Sebanyak 1.879 riwayat medis lengkap pasien, gaya hidup, dan karakteristik klinis tercakup dalam set data ini. Kondisi kesehatan setiap pasien diwakili oleh salah satu dari 45 kolom atau atribut dalam format CSV terstruktur. Analisis mendalam terhadap karakteristik dan variabel yang memengaruhi diabetes dimungkinkan karena setiap baris dalam set data ini mewakili satu entri pasien. Atribut set data ini mencakup variabel-variabel signifikan yang dapat menjelaskan hubungan antara hasil kesehatan, kondisi medis, dan gaya hidup pasien. Oleh karena itu, set data ini merupakan alat yang sangat baik untuk penelitian kesehatan di masa mendatang, terutama di bidang manajemen dan pemahaman diabetes.

Setiap dataset merepresentasikan satu entri pasien. Atribut-atribut dalam dataset mencakup:

**Tabel III.1**  
**Atribut Dataset Tahun 2025**

No	Nama Atribut	Tipe Data	Rentang Nilai	Keterangan
1	Pregnancies	Integer	0 – 17	Jumlah kehamilan yang pernah dialami pasien
2	Glucose	Integer	0 – 199	Kadar glukosa darah (mg/dL)
3	BloodPressure	Integer	0 – 122	Tekanan darah diastolik pasien (mmHg)
4	SkinThickness	Integer	0 – 99	Ketebalan lipatan kulit triceps (mm)
5	Insulin	Integer	0 – 846	Kadar insulin dalam darah (mu U/ml)
6	BMI	Float	0.0 – 67.1	Indeks Massa Tubuh (kg/m <sup>2</sup> )
7	DiabetesPedigreeFunction	Float	0.078 – 2.42	Skor riwayat keluarga/genetik terhadap diabetes
8	Age	Integer	21 – 81	Usia pasien (tahun)
9	Outcome	Integer	0 / 1	Target label: 0 = tidak diabetes, 1 = diabetes

Sumber: Dataset Prima Indians

Tabel III.1 menunjukkan Sembilan karakteristik penting yang secara akurat merepresentasikan kesehatan pasien dan digunakan sebagai input untuk proses klasifikasi risiko diabetes membentuk dataset yang digunakan dalam penelitian ini. Dataset Diabetes Pima

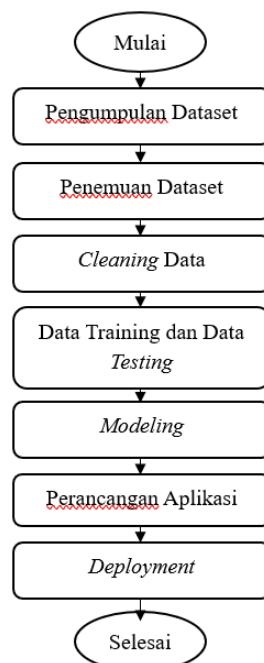
Indians, yang tersedia untuk umum dan terkenal karena penggunaannya dalam pembuatan sistem prediksi penyakit diabetes, dimodifikasi dalam dataset ini. Dataset ini distandarisasi secara ketat sebelum digunakan dalam analisis untuk menjamin kesesuaian dan kompatibilitasnya dengan *Support Vector Machine* (SVM), sebuah algoritma klasifikasi yang terkenal karena kemampuannya dalam menangani masalah klasifikasi non-linier.

Karakteristik yang dianalisis meliputi: jumlah kelahiran prenatal (kehamilan), tekanan darah (*blood pressure*), ketebalan kulit (*skin thickness*), insulin (*insulin*), indeks massa tubuh (IMT), skor riwayat keluarga/genetik terhadap diabetes (*DiabetesPedigreeFunction*), usia pasien (*age*), dan luaran sebagai prediksi target label. Setiap atribut masukan berupa variabel numerik yang dapat berupa bilangan bulat (*float*) atau bilangan bulat (*integer*), sedangkan keluaran berupa bilangan bulat dengan dua klasifikasi: 0 untuk pasien yang belum terdiagnosis diabetes dan 1 untuk pasien yang telah terdiagnosis diabetes.

Sebanyak 268 pasien dengan diagnosis diabetes (positif/label 1) dan 501 pasien tanpa diagnosis diabetes (negatif/label 0) membentuk total 769 data pasien dalam set data ini. Meskipun terdapat ketidakseimbangan yang signifikan dalam distribusi kelas ini, distribusi ini masih cocok untuk melatih model klasifikasi. Pemilihan atribut-atribut ini memberikan kontribusi substansial terhadap upaya masyarakat dalam mencegah dan mengelola diabetes, yang didasarkan pada relevansi klinis yang tinggi dan telah terbukti mendukung tingkat akurasi yang tinggi dalam deteksi dini risiko diabetes.

## 1.2. Alur Penelitian

Penelitian ini menggunakan metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang terdiri dari:



### 1.3. Use Case Diagram

Interaksi pengguna dengan sistem deteksi diabetes digambarkan dalam Use Case Diagram. Aktor utama adalah sistem dan pengguna, yang melakukan tugas utama seperti memasukkan data, memproses prediksi, dan menampilkan hasil risiko diabetes.

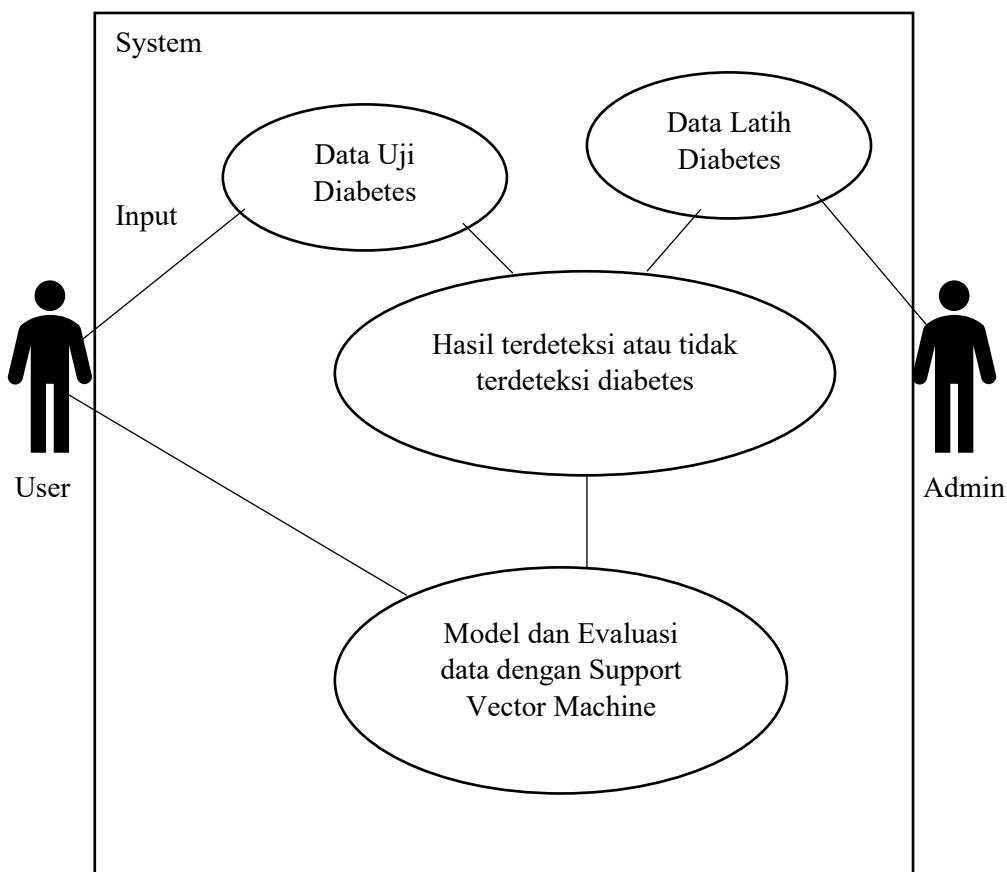


Diagram Kasus pada Gambar III.2 mengilustrasikan interaksi antara sistem, pengguna, dan administrator dalam konteks pengujian dan pelatihan data diabetes. Sistem menerima masukan berupa data uji dan pelatihan diabetes, kemudian menghasilkan keluaran yang menunjukkan apakah diabetes terdeteksi atau tidak. Dalam proses ini, data dianalisis dan dievaluasi oleh Support Vector Machines (SVM). Diagram ini menunjukkan alur kerja sistem dan hubungan antar komponen penting dalam proses deteksi diabetes. Pengguna berinteraksi dengan sistem untuk mendapatkan hasil deteksi diabetes.

### 1.4. Activity Diagram

Pada tahap ini, menampilkan Activity Diagram atau skema alur kegiatan yang menggambarkan proses awal berjalannya aplikasi atau sistem hingga selesai.

**Tabel III.2**  
**Activity Diagram**

No	Aktor	Aktivitas
1	User	Memulai pengujian dengan memasukkan hasil tes kesehatan terkait diabetes
2	User	Menekan tombol "Deteksi Diabetes" pada tampilan aplikasi
3	System	Membuka tampilan aplikasi di peramban web
4	System	Memulai proses deteksi berdasarkan input data pengguna
5	Model Machine Learning	Membaca dan memproses data uji untuk memprediksi risiko diabetes
6	System	Menampilkan hasil deteksi data kepada pengguna

Sumber: Pengumpulan Data Penelitian

Tabel IV.2 menggambarkan diagram aktivitas yang menjelaskan alur interaksi antara pengguna dan sistem dalam proses deteksi diabetes. Proses ini dimulai ketika pengguna memasukkan hasil tes kesehatan terkait diabetes ke dalam aplikasi dan menekan tombol "Deteksi Diabetes" pada antarmuka. Kemudian, sistem merespons dengan membuka tampilan aplikasi di peramban *web* dan memulai proses deteksi berdasarkan data yang dimasukkan pengguna. Selanjutnya, model pembelajaran mesin (seperti *Support Vector Machine* yang telah disebutkan sebelumnya) membaca dan memproses data tes untuk memprediksi risiko diabetes, dan hasil prediksi ini kemudian ditampilkan kembali kepada pengguna melalui antarmuka aplikasi. Diagram ini menjelaskan tahapan pengumpulan, transmisi, dan distribusi data.

## 1.5. Alat dan Bahan Penelitian

Komponen perangkat keras serta perangkat lunak yang dipakai untuk penelitian ini ditunjukan khusus untuk perancangan aplikasi.

**Tabel III.3**  
**Perangkat Keras untuk Perancangan Aplikasi**

Perangkat Keras
Processor: AMD A9-9425 RADEON R5 3.10GHz
RAM: 4 GB
HDD: 932 GB
GPU: AMD Radeon R5

Sumber: Pengumpulan Data Penelitian

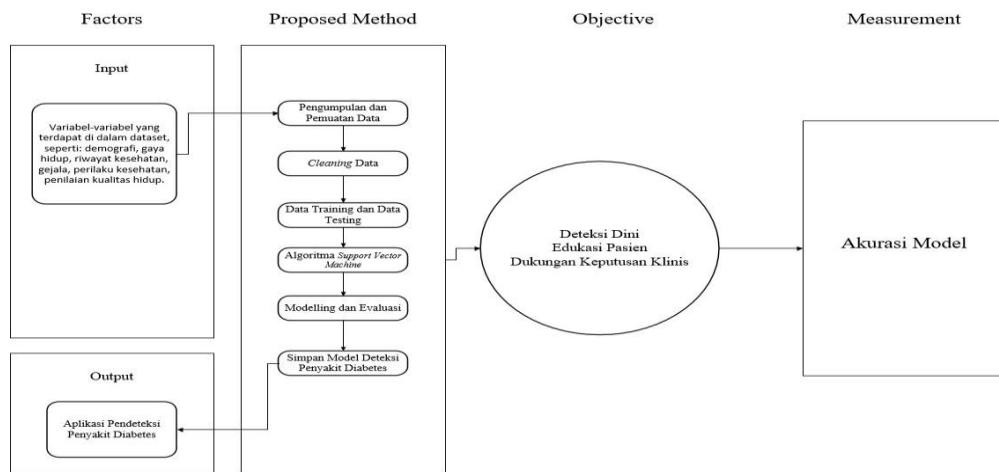
**Tabel III.4**  
**Perangkat Lunak untuk Perancangan Aplikasi**

Perangkat Lunak
Windows: 10 Pro 64 bit
Python 3.11.9
Visual Studio Code
Jupyter Notebook
Steamlit
Library: Pandas, NumPy, Scikit-Learn, Seaborn, dan Matplotlib

Sumber: Pengumpulan Data Penelitian

## 1.6. Kerangka Pemikiran

Rasional di balik penelitian ini adalah bahwa *algoritma Support Vector Machine* (SVM) memiliki kemampuan untuk menciptakan model klasifikasi yang memaksimalkan jarak antara dua kelas diagnostik. Diharapkan sistem yang dikembangkan dapat memberikan prediksi dini terkait risiko diabetes menggunakan dataset dengan kelengkapan terjamin dan menerapkan prosedur praproses yang ideal. Sistem ini juga dapat membantu pengambilan keputusan manajemen kesehatan yang lebih baik dan mengedukasi pengguna.



**Gambar III.5 Kerangka Pemikiran Penelitian**

Tujuan penelitian ini adalah untuk membuat sistem yang dapat menggunakan analisis prediktif Support Vector Machine (SVM) untuk mendeteksi diabetes dini. Dalam dataset yang digunakan, terdapat 45 variabel input, meliputi faktor demografi (usia, jenis kelamin), klinis (tekanan darah, HbA1c), gaya hidup (aktivitas fisik, pola makan), dan gejala diabetes spesifik (poliuria, polidipsia). Total pasien yang digunakan adalah 1.127 negatif dan 752 positif. Berdasarkan data prediktif, sistem ini dirancang untuk meningkatkan manajemen kesehatan dengan mendeteksi penyakit dini dan membantu pengambilan keputusan klinis. Menggunakan

Python (Scikit-Learn, Streamlit), aplikasi ini diuji berdasarkan akurasi model, sehingga hasilnya dapat diandalkan untuk skrining risiko diabetes yang efektif.