

**PERANCANGAN SISTEM DIAGNOSA PENYAKIT  
DIABETES MENGGUNAKAN APLIKASI DIATECT  
BERBASIS WEB**

**SKRIPSI**

Diajukan Untuk Memenuhi Syarat Mendapatkan Gelar Sarjana Strata Satu Pada  
Program Studi Teknik Elektro Fakultas Teknik Universitas Islam Nusantara



Oleh :

**ANISA FEBRIANTI**

**41037002211015**

**PROGRAM STUDI TEKNIK ELEKTRO  
FAKULTAS TEKNIK  
UNIVERSITAS ISLAM NUSANTARA  
2025**

## LEMBAR KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini:

Nama : Anisa Febrianti  
NIM : 41037002211015  
Program Studi : Teknik Elektro

Menyatakan bahwa Skripsi yang berjudul:

**PERANCANGAN SISTEM DIAGNOSA PENYAKIT DIABETES MENGGUNAKAN APLIKASI DIATECT BERBASIS WEB** dengan demikian, saya menyatakan bahwa karya ini disusun secara mandiri berdasarkan hasil penelitian, pemikiran, dan pemaparan saya sendiri Karya ini diperlukan untuk mendapatkan gelar Sarjana (S1) di Program Studi Teknik Elektro di Fakultas Teknik Universitas Islam Nusantara Bandung. Sejauh yang saya ketahui, karya ini bukanlah replika atau duplikasi dari skripsi yang telah diterbitkan atau digunakan untuk mendapatkan gelar Sarjana (S1) di Fakultas Teknik Universitas Islam Nusantara Bandung atau di institusi lain, kecuali kecuali bagian yang sumber informasi dicantumkan sebagaimana mestinya.

Bandung, 23 April 2025

Yang membuat pernyataan,

**ANISA FEBRIANTI**

**NIM 41037002211015**

**LEMBAR PENGESAHAN**  
**PERANCANGAN SISTEM DIAGNOSA PENYAKIT DIABETES**  
**MENGGUNAKAN APLIKASI DIATECT BERBASIS WEB.**

Disusun dan diajukan oleh :

**ANISAFEBRIANTI**  
**41037002211015**

Disetujui dan disahkan pada sidang  
skripsi Pada tanggal :

**Bandung, 23 April 2025**

Pembimbing 1

Pembimbing 2

**Dr. Iksal Rachman, M.T.**

**Agung Muhamad Toha, S.T., M.T.**

Mengetahui :

Dekan Fakultas Teknik

Kepala Program Studi

**Dr. Ricky Yoseptry, S.T., M.M.Pd.**

**Muhammad Zimamul Adli, S.Si., M.Si.**

**LEMBAR PENGESAHAN**

**REVISIAN SKRIPSI**

**PERANCANGAN SISTEM DIAGNOSA PENYAKIT DIABETES  
MENGUNAKAN APLIKASI DIATECT BERBASIS WEB.**

**Telah Direvisi**

**Oleh :**

**ANISAFEBRIANTI**

**41037002211004**

**Bandung, 17 Juli 2025**

Mengesahkan,

Penguji 1

Penguji 2

**Ganis Sanhaji S.Si., M.Sc.**

**Muhammad Zimamul Adli, S.Si., M.Si.**

Ketua Sidang

**Dr. Ricky Yoseptry, S.T., M.M.Pd.**

## BIODATA PENULIS



Nama : Anisa Febrianti  
Tempat, Tanggal Lahir : Bekasi, 02 Februari 2003  
Telepon : 0856-9316-3795  
Email : nisafeb020203@gmail.com  
Riwayat Pendidikan : SDN Karedok  
SMPN 1 Jatigede  
SMKN 1 Sumedang

## KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT, Tuhan Yang maha esa, atas limpahan rahmat, kasih sayang, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Perancangan Sistem Diagnosa Penyakit Diabetes Menggunakan Aplikasi DIATECT Berbasis Web”. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan program pendidikan Strata Satu (S1) di Universitas Islam Nusantara, Fakultas Teknik, Program Studi Teknik Elektro.

Penyusunan skripsi ini merupakan perjalanan yang panjang dan penuh tantangan. Penulis menyadari bahwa pencapaian ini tidak akan terwujud tanpa doa, dukungan, serta bantuan dari berbagai pihak. Dengan penuh rasa syukur dan hormat, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, yang senantiasa melimpahkan kekuatan, kesehatan, dan petunjuk di setiap langkah perjalanan penulis. Segala sesuatu terjadi atas kehendak-Nya, dan penulis berdoa semoga karya ini dapat memberikan manfaat
2. Ibu Ririn dan Almarhum Ayah saya Bapak Warya selaku orang tua tercinta yang sangat saya cinta dan banggakan, terima kasih atas dukungan serta doa selama penulis menyelesaikan skripsi dan menyelesaikan masa perkuliahan selama empat tahun. Serta kakak tercinta Mba Yuli dan Mba Yani serta keluarga besar penulis yang telah mendukung dengan memberika pendampingan moral dan semangat.
3. Bapak Dr. Ricky Yoseptry, S.T., M.M.Pd. Selaku Dekan Fakultas Teknik, yang telah memberikan dukungan dan motivasi kepada penulis selama masa perkuliahan.
4. Bapak Muhammad Zimamul Adli, S.Si., M.Si. Selaku Ketua Program Studi Teknik Elektro sekaligus dosen pembimbing selama masa perkuliahan, atas arahan, bimbingan dan kesabaran dalam membantu penulis menyelesaikan perkuliahan.
5. Bapak Dr. Iksal Rachman, M.T. Selaku dosen pembimbing 1 skripsi yang telah memberi arahan selama peneliti.

6. Bapak Agung Muhamad Toha, S.T., M.T. Selaku dosen pembimbing 2 skripsi yang telah memberi arahan selama peneliti.

Bapak Ganis Sanhaji, S.Si., M.Sc. Selaku Dosen Teknik Elektro, yang telah memberikan dukungan dan motivasi kepada penulis selama masa perkuliahan.

7. Seluruh dosen Fakultas Teknik, yang dengan tulus telah membagikan ilmu dan pengalaman berharga sebagai bekal yang sangat berarti bagi penulis.

8. Seluruh staf Tata Usaha Fakultas Teknik, yang dengan penuh kesabaran telah membantu penulis dalam menyelesaikan berbagai keperluan administrasi kemahasiswaan.

9. Vito Dwi Nur Hidayat, Selaku Pasangan dan Partner Saya atas dukungan dan bantuan selama penulis melakukan penelitian.

10. Rekan seperjuangan Herlan Syah, Julian, Decky dan Hidayat, yang telah menjadi teman diskusi, penyemangat, dan partner perjuangan selama menyelesaikan skripsi ini.

11. Seluruh teman-teman Teknik Elektro angkatan 2021, yang telah memberikan, dukungan, dan kebersamaan selama masa perkuliahan.

12. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah membantu, mendukung, dan mendoakan penulis selama ini.

Penulis menyadari sepenuhnya bahwa skripsi ini masih memiliki kekurangan dan belum mencapai kesempurnaan. Oleh karena itu, penulis sangat mengharapkan saran dan kritik yang konstruktif untuk perbaikan pada masa mendatang. Semoga karya ini dapat memberikan manfaat bagi semua pihak yang membacanya.

Bandung, 23 April 2025

**Anisa Febrianti**

**Nim 41037002211015**

## ABSTRAK

Nama : Anisa Febrianti Program

Studi : Teknik Elektro

Judul : PERANCANGAN SISTEM DIAGNOSA PENYAKIT

DIABETES MENGGUNAKAN APLIKASI DIATECT BERBASIS WEB

Penelitian ini menunjukkan efek yang signifikan dari diabetes. Tujuan aplikasi ini untuk menemukan gejala diabetes. Fokus dari penelitian ini adalah peningkatan jumlah orang yang menderita diabetes. Organisasi Kesehatan Dunia ( WHO ) telah melaporkan bahwa dalam beberapa dekade terakhir, jumlah orang yang menderita diabetes telah mengalami peningkatan yang signifikan, menjadikannya masalah besar bagi kesehatan dunia dengan tingkat kejadian yang terus meningkat. Jumlah orang dewasa penderita diabetes di dunia akan meningkat dari 135 juta pada tahun 1995 menjadi 300 juta pada tahun 2025. Indonesia merupakan negara dengan penderita penyakit diabetes mellitus cukup tinggi. Dengan Prevalensi 8,6 % dari total penduduk, terdapat 4,5 juta pengidap diabetes dan pada tahun 2025 diperkirakan meningkat menjadi 12,4 juta penderita. Diabetes menjadi penyakit ke-3 faktor penyebab kematian di Indonesia.

Penelitian ini memperkenalkan aplikasi berbasis web yang disebut DIATECT (Perancangan Sistem Diagnosa Penyakit Diabetes Menggunakan Aplikasi DIATECT Berbasis Web) yang dirancang untuk memprediksi risiko penyakit diabetes. Aplikasi ini memanfaatkan algoritme Support Vector Machine (SVM) untuk mengolah data kesehatan yang luas dan beragam. DIATECT dapat memprediksi kemungkinan bahwa seseorang memiliki diabetes. Berdasarkan informasi yang diberikan. Penelitian ini menggunakan metode kualitatif, dan 768 data sekunder terdiri dari individu yang menderita diabetes dan tidak menderita diabetes. Hasil penelitian menunjukkan bahwa aplikasi ini memiliki performa yang kuat, dengan presisi mencapai 77%, *recall* mencapai 71%, dan F1-skor mencapai 77%. Aplikasi ini memiliki potensi besar untuk mendukung masyarakat dalam mendeteksi dan mengidentifikasi penyakit melalui platform web yang mudah diakses.

Kata Kunci : SVM, Diabetes, Deteksi, Aplikasi, DIATECT

## ***ABSTRACT***

Name : Anisa Febrianti  
Departement : Electrical Engineering  
Title : DESIGN OF A DIABETES DIAGNOSIS SYSTEM USING THE  
WEB-BASED DIATECT APPLICATION.

This study shows the significant effects of diabetes. The purpose of this application is to find the symptoms of diabetes. The focus of this study is the increasing number of people suffering from diabetes. The World Health Organization (WHO) has reported that in recent decades, the number of people suffering from diabetes has increased significantly, making it a major problem for world health with an increasing incidence rate. The number of adults with diabetes in the world will increase from 135 million in 1995 to 300 million in 2025. Indonesia is a country with quite high diabetes mellitus sufferers. With a prevalence of 8.6% of the total population, there are 4.5 million people with diabetes and in 2025 it is estimated to increase to 12.4 million sufferers. They are susceptible to other diseases such as heart attacks, strokes, blindness, kidney failure, paralysis, and even death due to this condition. Diabetes is the 3rd leading cause of death in Indonesia.

This study introduces a web-based application called DIATECT (Design of a Diabetes Disease Diagnosis System Using the Web-Based DIATECT Application) which is designed to predict the risk of diabetes. This application utilizes the Support Vector Machine (SVM) algorithm to process extensive and diverse health data. DIATECT can predict the likelihood that someone has diabetes. Based on the information provided. This study used a qualitative method, and 768 secondary data consisting of individuals with and without diabetes. The results showed that this application has strong performance, with a precision of 77%, recall of 71%, and F1-skor of 77%. This application has great potential to support the community in detecting and identifying diseases through an easily accessible web platform.

*Keywords:* SVM, Diabetes, *Detection, Application, DIATECT*

## DAFTAR ISI

<b>LEMBAR KEASLIAN SKRIPSI.....</b>	<b>i</b>
<b>LEMBAR PENGESAHAN.....</b>	<b>ii</b>
<b>LEMBAR PENGESAHAN REVISI SKRIPSI.....</b>	<b>iii</b>
<b>BIODATA PENULIS .....</b>	<b>iv</b>
<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vii</b>
<b>DAFTAR ISI.....</b>	<b>ix</b>
<b>DAFTAR GAMBAR .....</b>	<b>xi</b>
<b>DAFTAR TABEL.....</b>	<b>xii</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	4
1.3. Tujuan Penelitian .....	4
1.4. Manfaat Penelitian .....	5
1.5. Metode Penelitian .....	6
1.6. Metode Studi Pusaka .....	7
1.7. Sistematika Penulisan .....	7
<b>BAB 2 LANDASAN TEORI.....</b>	<b>8</b>
2.1. Penelitian Terdahulu.....	8
2.2. Definisi Penyakit Diabetes.....	18
2.3. Klasifikasi Penyakit Diabetes .....	20
2.4. Faktor Risiko Penyebab Penyakit Diabetes .....	22
2.5. Pencegahan dan Deteksi Dini .....	22
2.6. Teknologi dalam Pencegahan dan Pengelolaan Diabetes.....	23
2.7. Algoritma Support Vector Machine ( SVM ) .....	24
2.7.1. Definisi Support Vector Machine ( SVM ).....	24
2.7.2. Komponen Utama Pemrosesan Algoritma SVM .....	25
2.7.3. Prinsip Kerja Support Vector Machine ( SVM ).....	26
2.7.4. Kelebihan dan Kelemahan Support Vector Machine ( SVM ).....	32
2.7.5. Kerangka Berpikir .....	33
<b>BAB 3 PERANCANGAN SISTEM DAN IMPLEMENTASI SISTEM.....</b>	<b>34</b>
3.1. Metode Penelitian .....	34

3.1.1. Pendekatan Penelitian.....	35
3.1.2. Jenis Penelitian .....	36
3.1.3 Populasi dan Sampel Penelitian.....	37
3.1.4. Sumber Data.....	38
3.1.5. Pengumpulan Data.....	39
3.2. Perencanaan Blok Diagram .....	39
<b>BAB 4 PEMBAHASAN DAN ANALISIS.....</b>	<b>41</b>
4.1. Analisis Data .....	41
4.1.1 Pengumpulan Data.....	41
4.1.2. Pelabelan Atribut Data.....	42
4.1.3. <i>Preprocessing</i> Data .....	43
4.1.4. Klasifikasi Data Training & Testing .....	44
4.2. <i>Load Data Set</i> .....	46
4.3. Standardisasi Data.....	48
4.4. Pembagian Data.....	49
4.5. <i>Modelling</i> .....	50
4.6. Akurasi .....	52
4.7. Performa Aplikasi .....	53
4.8. Tampilan Aplikasi.....	56
<b>BAB 5 KESIMPULAN DAN SARAN .....</b>	<b>60</b>
5.1. Kesimpulan.....	60
5.2. Saran.....	62
DAFTAR PUSAKA .....	63
LAMPIRAN.....	65

## DAFTAR GAMBAR

Gambar 1.1 Data Faktor Penyebab Kematian Tertinggi Di Indonesia.....	1
Gambar 3.2. <i>Flowchart</i> Sistem.....	40
Gambar 4.1.4. Proses Pemisahan Data Training dan Testing.....	45
Gambar 4.2. Proses <i>Load Data Set</i> .....	47
Gambar 4.2. Proses Baris Data Set.....	47
Gambar 4.3. Proses Standarisai Data.....	48
Gambar 4.4. Proses Pembagian Data.....	49
Gambar 4.5. <i>Modelling</i> .....	51
Gambar 4.6. Akurasi Data.....	52
Gambar 4.7. Performa Aplikasi.....	56
Gambar 4.8. Tampilan Aplikasi.....	56

## DAFTAR TABEL

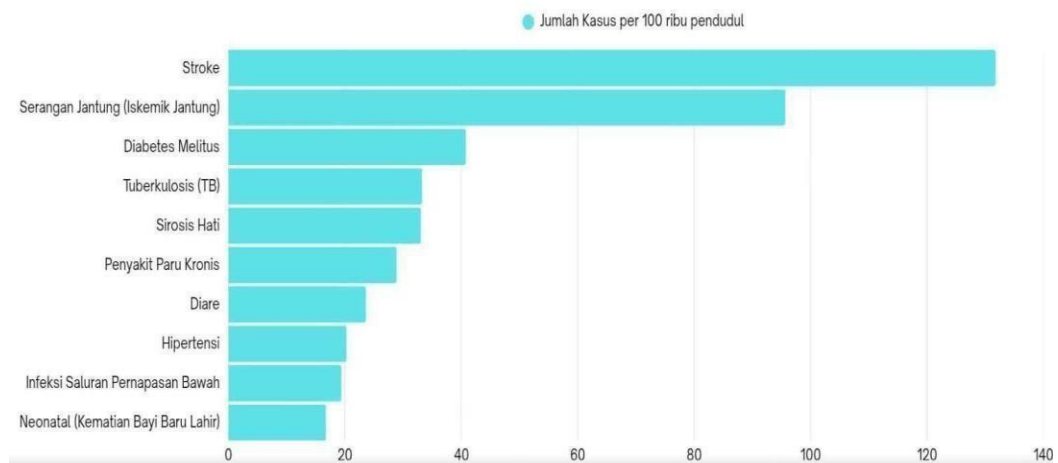
Tabel 2.1. Penelitian Terdahulu.....	8
Tabel 4.1.1. Jenis Keterangan Yang Digunakan.....	41
Tabel 4.1.2. Jenis prediksi beserta Labelnya .....	42
Tabel 4.1.4. Data <i>Training</i> dan Data <i>Testing</i> .....	45

# BAB 1

## PENDAHULUAN

### 1.1. Latar Belakang

Organisasi Kesehatan Dunia (WHO) mengatakan bahwa jumlah orang yang menderita diabetes telah meningkat secara signifikan dalam beberapa dekade terakhir. Hal ini menjadikannya masalah besar bagi kesehatan global dengan tingkat kejadian yang terus meningkat. Pada tahun 2025, jumlah orang dewasa yang menderita diabetes di seluruh dunia akan meningkat dari 135 juta pada tahun 1995 menjadi 300 juta, Sebagian besar peningkatan ini akan terjadi di negara berkembang ( King H, Aubert RE, Herman WH, 1998 ). Negara maju akan mengalami peningkatan sebesar 42% dari 51 menjadi 72 juta, dan negara berkembang akan mengalami peningkatan sebesar 170% dari 84 menjadi 228 juta. Indonesia memiliki tingkat prevalensi diabetes mellitus yang tinggi. Dengan 4,5 juta orang yang menderita diabetes, yang merupakan 8,6 persen dari total penduduk, jumlah ini diperkirakan akan meningkat menjadi 12,4 juta pada tahun 2025.



**Gambar 1.1.** Data Faktor Penyebab Kematian Tertinggi Di Indonesia Tahun 2024

( Sumber : Data WHO 2019 / Inilah.com )

Diabetes melitus merupakan salah satu penyakit kronis yang ditandai dengan meningkatnya kadar glukosa dalam darah yang melebihi batas normal. Kondisi ini dapat disebabkan oleh resistensi sel tubuh terhadap insulin atau gangguan pada produksi hormon insulin pankreas. Hormon utama yang bertanggung jawab untuk mengatur metabolisme glukosa agar sel-sel tubuh dapat menggunakannya sebagai sumber energi adalah insulin. Jika fungsi insulin terganggu, sel tidak dapat menyerap glukosa dengan baik, yang menyebabkan hiperglikemia, yang dapat menyebabkan banyak masalah serius, Hiperglikemia jangka panjang dapat menyebabkan berbagai komplikasi serius, baik akut maupun kronis, yang memengaruhi sistem organ tubuh secara keseluruhan ( F. Amira Mumtaz, 2020 )

Pola makan yang tidak sehat, yang mencakup makanan yang tinggi gula dan lemak serta rendah serat, serta kurangnya aktivitas fisik, stres berlebihan, dan gaya hidup sedentari, adalah salah satu penyebab utama peningkatan prevalensi diabetes. Namun, diabetes termasuk dalam kategori penyakit tidak menular yang dapat dicegah hingga 80% dengan menerapkan pola hidup sehat secara teratur. Perawatan pencegahan yang efektif, seperti pendidikan kesehatan, pola makan yang sehat, aktivitas fisik yang teratur, dan pemeriksaan kesehatan rutin, tidak hanya dapat mencegah peningkatan jumlah penderita diabetes setiap tahunnya, tetapi juga meningkatkan kualitas hidup Masyarakat.

Diabetes melitus adalah salah satu penyakit tidak menular (PTM) yang memiliki pengaruh besar terhadap kesehatan masyarakat di seluruh dunia karena dapat menyebabkan berbagai komplikasi serius yang dapat mengganggu kualitas hidup pasien

Pembelajaran mesin dan kemajuan teknologi informasi saat ini sangat membantu mencegah dan mengelola diabetes melitus. Teknik ini tidak hanya memungkinkan pengolahan data dalam jumlah besar secara efisien, tetapi juga memungkinkan pengembangan sistem cerdas yang dapat membantu dalam diagnosis dan deteksi dini faktor-faktor yang berpotensi menyebabkan diabetes. Pengembangan sistem berbasis web yang disebut DIATECT adalah salah satu implementasi teknologi ini. Aplikasi ini dimaksudkan untuk memprediksi risiko diabetes pada individu. Aplikasi ini menggunakan algoritma Support Vector Machine (SVM), sebuah metode pembelajaran mesin yang telah terbukti sangat akurat untuk menganalisis dan mengklasifikasikan data kesehatan ( M. D. Purbolaksono, 2021 ).

Algoritma SVM bekerja dengan cara menentukan *hyperplane*, optimal yang memisahkan kelas data sejauh mungkin. Ini disebut margin maksimum, dan jarak antar kelas harus sejauh mungkin. SVM dapat digunakan untuk mentransformasikan data ke dalam ruang dimensi yang lebih besar dengan menggunakan pendekatan trik kernel. Ini memungkinkan pemisahan kelas bahkan pada data yang tidak dapat dipisahkan secara linear di ruang awal. Oleh karena itu, SVM dapat digunakan untuk membuat model prediksi yang efektif baik dalam kasus linear maupun non-linear. Selain itu, karena SVM hanya bergantung pada kumpulan data pelatihan yang kecil, vector pendukung meningkatkan efisiensi komputasi model dan memberikan generalisasi yang baik untuk data baru ( A. Abimayu, 2023 ).

Sistem prediksi risiko diabetes DIATECT menggunakan data kesehatan dan metode pembelajaran mesin yang canggih. Selain itu, DIATECT dirancang untuk berfungsi sebagai sumber informasi interaktif dan panduan bagi mereka yang telah didiagnosa diabetes. Melalui antarmuka berbasis web yang mudah digunakan dan dapat diakses dari berbagai perangkat, DIATECT menawarkan saran tentang pengelolaan gaya hidup, membantu orang mengakses informasi kesehatan, dan memberi tahu mereka tentang komplikasi yang mungkin terjadi jika pengobatan diabetes tidak dilakukan dengan benar.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang yang telah dijelaskan, terdapat beberapa permasalahan utama yang menjadi fokus penelitian ini, yaitu:

1. Bagaimana aplikasi berbasis web seperti DIATECT dapat membantu dalam menurunkan angka diabetes melalui pencegahan dan deteksi dini?
2. Bagaimana integrasi teknologi informasi dan pembelajaran mesin, seperti SVM, dapat meningkatkan efektivitas deteksi dini dan manajemen penyakit diabetes?
3. Bagaimana memanfaatkan algoritma Support Vector Machin (SVM) untuk memprediksi risiko diabetes secara akurat dalam sebuah aplikasi berbasis web?

## **1.3. Tujuan Penelitian**

Penelitian ini memiliki beberapa tujuan utama, yaitu:

1. Support Vector Machine (SVM) untuk memprediksi kemungkinan timbulnya diabetes.
2. Menganalisis efektivitas algoritma SVM dalam mengolah data kesehatan untuk menghasilkan prediksi yang andal terkait risiko diabetes.
3. Mendukung upaya pencegahan dan pengelolaan diabetes melalui inovasi teknologi informasi, khususnya dalam mendeteksi dini risiko diabetes.

## **1.4. Manfaat Penelitian**

### **Manfaat Teoretis:**

1. Pengembangan Ilmu Pengetahuan : Penelitian ini akan memberikan kontribusi bagi perkembangan ilmu pengetahuan, terutama di bidang teknologi kesehatan, dengan fokus pada penggunaan algoritma Support Vector Machine (SVM) untuk mendiagnosis penyakit. Hasil penelitian ini diharapkan dapat menjadi landasan untuk penelitian lebih lanjut mengenai penggunaan teknologi canggih dalam bidang kesehatan.
2. Penerapan Teknologi dalam Kesehatan : Penelitian ini memberikan gambaran yang lebih jelas tentang bagaimana teknologi informasi dan pembelajaran mesin dapat diterapkan untuk mencegah dan mengelola penyakit diabetes, sehingga membuka peluang bagi pengembangan model prediksi berbasis web pada masa depan.
3. Pemahaman Tentang Diabetes : Penelitian ini juga akan memperdalam pemahaman kita mengenai penyakit diabetes, termasuk bagaimana teknologi dapat membantu dalam deteksi dini dan pengelolaan yang lebih baik.

### **Manfaat Praktis:**

1. Alat Prediksi Risiko Diabetes : Aplikasi DIATECT yang dikembangkan akan memberikan kemudahan bagi individu untuk memeriksa risiko terkena diabetes. Dengan adanya aplikasi berbasis web ini, orang-orang dapat dengan cepat mengetahui kondisi kesehatan mereka dan mengambil langkah pencegahan jika diperlukan.
2. Penyuluhan dan Edukasi Kesehatan : Aplikasi ini juga berfungsi sebagai sarana edukasi yang bermanfaat, memberikan informasi mengenai cara mencegah dan mengelola diabetes dengan lebih baik, serta membantu orang-orang mengubah gaya hidup mereka ke arah yang lebih sehat.

3. Meningkatkan Pengelolaan Penyakit : Bagi mereka yang sudah menderita diabetes, aplikasi ini bisa menjadi panduan yang sangat membantu dalam mengelola penyakit mereka dengan lebih efektif, memberikan rekomendasi berdasarkan data kesehatan pribadi.

**Manfaat Sosial:**

1. Pencegahan Penyakit : Dengan meningkatkan kesadaran masyarakat tentang pentingnya deteksi dini dan pencegahan diabetes, penelitian ini dapat membantu menurunkan angka kasus diabetes di masyarakat, yang pada akhirnya akan mengurangi beban kesehatan global.
2. Meningkatkan Kualitas Hidup : Melalui informasi yang mudah diakses tentang pencegahan dan pengelolaan diabetes, aplikasi ini akan membantu orang-orang menjalani gaya hidup yang lebih sehat, sehingga kualitas hidup mereka dapat meningkat.
3. Mengurangi Beban Kesehatan Masyarakat : Dengan mencegah dan mendeteksi diabetes lebih dini, aplikasi ini dapat mengurangi risiko komplikasi serius, seperti penyakit jantung, stroke, dan gagal ginjal. Ini pada gilirannya dapat mengurangi biaya pengobatan serta beban perawatan medis yang harus ditanggung oleh masyarakat dan sistem kesehatan.

**1.5. Metode Penelitian**

**Pendekatan kualitatif**

Metode penelitian yang digunakan dalam penelitian ini adalah pendekatan kualitatif dengan metode eksperimen. Penelitian ini bertujuan untuk menilai efektivitas algoritma Support Vector Machine (SVM) dalam memprediksi risiko diabetes menggunakan data numerik yang relevan. Penelitian ini menggunakan data sekunder dari platform Kaggle dengan jumlah sampel 768 data, terdiri dari 500 data non-diabetes dan 268 data diabetes.

### **1.5. Metode Studi Pustaka**

Untuk mendukung metode eksperimen, penulis menerapkan metode studi pustaka dengan mencari referensi yang relevan dengan judul penelitian.

### **1.6. Sistematika Penulisan**

Berikut sistematika penulisan yang digunakan pada penelitian ini :

#### **BAB I: PENDAHULUAN**

Bab ini membahas latar belakang penelitian, tujuan dan manfaat penelitian, ruang lingkup kajian, serta sistematika penulisan laporan ini.

#### **BAB II: LANDASAN TEORI**

Bab ini menjelaskan dasar-dasar teori yang mendukung penelitian, termasuk konsep dan teknologi yang digunakan dalam sistem yang akan dibahas.

#### **BAB III: PERANCANGAN SISTEM DAN IMPLEMENTASI**

Bab ini menjelaskan perancangan sistem yang mencakup teknologi yang digunakan, metode implementasi, serta pengembangan perangkat lunak berbasis algoritma svm.

#### **BAB IV: PEMBAHASAN DAN ANALISIS**

Bab ini berisi hasil pengujian sistem, analisis data yang diperoleh, serta evaluasi terhadap kinerja sistem yang telah dibuat.

#### **BAB V: KESIMPULAN DAN SARAN**

Bab terakhir menyajikan kesimpulan dari penelitian yang dilakukan serta saran untuk pengembangan lebih lanjut berdasarkan hasil yang diperoleh.

## BAB 2

### LANDASAN TEORI

#### 2.1. Penelitian Terdahulu

Metode deteksi dini diabetes dan penyakit kronis lainnya telah berubah seiring perkembangan jaman. Dengan kemajuan dalam teknologi kesehatan, metode komputasi cerdas dapat membantu proses diagnosis menjadi lebih akurat dan efisien. Penggunaan algoritma pembelajaran mesin, khususnya Support Vector Machine (SVM), adalah salah satu pendekatan yang mendapat perhatian ilmiah. Penelitian telah menunjukkan bahwa SVM adalah salah satu metode yang paling mungkin untuk membangun sistem pendukung keputusan di bidang kesehatan. Hasil penelitian telah menunjukkan bahwa SVM efektif dalam menemukan pola medis yang terkait dengan risiko diabetes.

**Tabel 2.1.** Penelitian Terdahulu

No	Nama Peneliti	Judul Penelitian	Metode Penelitian	Isi Penelitian
1.	( Nazin Ahmed , Rayhan Ahammed , Md. Manowarul Islam , Md. Ashraf Uddin , Arnisha Akhter , Md. Alamin Talukder, Bikash Kumar	Machine learning based diabetes prediction and development of smart web application.	Penelitian terapan dengan pendekatan eksperimen komputasional	Aplikasi prediksi diabetes berbasis web menggunakan algoritma SVM, Decision Tree, dan Logistic Regression. Tujuan aplikasi ini adalah untuk memungkinkan pengguna memasukkan data

	Paul, 2021 )			secara mandiri, lalu sistem memprediksi apakah mereka berisiko terkena diabetes. Model evaluasi menunjukkan bahwa algoritma SVM paling akurat dibandingkan dengan algoritma lainnya.
2.	( Abdullellah Alotaibi, Ali Al-Ganmi, Leila Golizadeh, Lin Perry, 2016)	Diabetes Knowledge of Nurses in Different Countries.	Systematic review dari 25 studi dimasukkan dalam tinjauan dan disintesis berdasarkan karakteristik studi, desain, dan temuan	Studi ini menekankan bahwa literasi digital yang rendah dan pelatihan tenaga kesehatan yang kurang berbasis teknologi adalah penghalang utama dalam penanganan diabetes. Oleh karena itu, artikel ini menekankan pentingnya

				<p>mendukung pelayanan kesehatan preventif dengan sistem edukatif berbasis teknologi, yang mencakup platform dan aplikasi digital.</p>
3.	<p>( Ichwanul Muslim Karo Karo, Hendriyana, 2022 )</p>	<p>Klasifikasi Penderita Diabetes Menggunakan Algoritma Machine Learning dan Z-Score.</p>	<p>Eksperimental kuantitatif, dengan membangun dan menguji model klasifikasi diabetes dengan kombinasi preprocessing Z-Score dan algoritma machine learning.</p>	<p>Dalam Penelitian ini, Support Vector Machine , Decision Tree, dan Naive Bayes dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1- skor. Hasil menunjukkan bahwa SVM memberikan hasil terbaik, dengan akurasi 80,73% dan F1-skor 76,6%. Penulis juga menekankan bahwa penerapan normalisasi Z-</p>

				Score secara signifikan dapat meningkatkan kinerja klasifikasi model dibandingkan dengan metode normalisasi lain,
4.	( Favorisen R. Lumbanraja, Fanni Lufiana, Yunda Heningtyas & Kurnia Muludi, 2022 )	Implementasi Support Vector Machine (SVM) untuk Klasifikasi Penderita Diabetes Mellitus.	Metode kuantitatif dengan pendekatan machine learning.	Hasil ini menunjukkan bahwa SVM dengan kernel RBF dapat menangani data skala besar tentang klasifikasi diabetes. Mereka juga menunjukkan bahwa teknik validasi silang sangat penting untuk memastikan keakuratan model prediksi medis.

5.	( Praveen Talari, Mana Saleh Al Reshan , Bharathiraja , Adel Sulaiman , Gaganpreet Kaur , Hani Alshahrani , Asadullah Shaikh, 2024 )	Hybrid feature selection and classification technique for early prediction and severity of diabetes type 2.	Pendekatan eksperimental berbasis machine learning hibrid.	Bertujuan meningkatkan akurasi prediksi diabetes tipe 2 melalui pendekatan hybrid machine learning berbasis model fusion. Menggunakan Pima Indian Diabetes Dataset (PIDD) dari UCI Machine Learning Repository yang berisi 768 data pasien dengan 8 fitur medis, penelitian ini melakukan pra-pemrosesan data (penanganan missing values, normalisasi, dan pembagian data train-test) sebelum membangun empat model
----	--	---	--	---

				<p>pembelajaran mesin individu, yaitu Logistic Regression (LR), Random Forest (RF), XGBoost (XGB), dan Support Vector Machine (SVM). Model fusion yang dihasilkan mencapai akurasi 88,7% dan ROC-AUC 0,91, melampaui performa model tunggal (LR: 76,3%; RF: 82,3%; XGB: 83,2%; SVM: 80,1%), sehingga membuktikan bahwa kombinasi model mampu memberikan hasil prediksi lebih akurat dan berpotensi diimplementasikan</p>
--	--	--	--	--

				n dalam sistem pendukung keputusan di bidang kesehatan.
6.	( Sonya Dian Wahyuni, R. Hadapiningradja Kusumodestoni, 2024 )	Optimalisasi Algoritma Support Vector Machine	Pendekatan eksperimental dalam ranah machine learning.	Bertujuan mengatasi permasalahan stunting di Indonesia melalui penerapan algoritma SVM dengan empat jenis kernel, yaitu Linear, Polynomial, RBF, dan Sigmoid. Menggunakan dataset
7.	( Mahade Hasan, Farhana Yasmin, 2025)	Predicting Diabetes Using Machine Learning: A Comparative Study of Classifiers.	Metode eksperimental kuantitatif berbasis machine learning, dengan pendekatan komparatif terhadap berbagai algoritma	Menggabungkan algoritma konvensional (Logistic Regression, SVM, Naïve Bayes, dan Random Forest) dengan model ensambeldan

			klasifikasi.	arsitektur hybrid CNN-LSTM yang canggih bernama DNet. Hasil evaluasi dataset Kaggle menunjukkan bahwa DNet memiliki akurasi deteksi diabetes dini.
8.	( Nurlaelatul Maulidah, Riki Supriyadi, Dwi Yuni Utami, Fuad Nur Hasan, Ahmad Fauzi, Ade Christian,2021 )	Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector.	Pendekatan eksperimental berbasis machine learning.	Dua algoritma pembelajaran mesin, Support Vector Machine (SVM) dan Naive Bayes, digunakan untuk memproses data dalam penelitian ini. Hasilnya menunjukkan bahwa model SVM memiliki akurasi 78,04%, sedangkan Naive Bayes memiliki

				76,98%, dan perbedaan performa sekitar.
9.	( Ilham Arif Kuncoro Adi & Wahyu Aji Eko Prabowo, 2023)	Klasifikasi Penyakit Diabetes Menggunakan Pendekatan Pembelajaran Mesin dengan Model Non-Linear	Eksperimen komputasional dalam domain machine learning Support Vector Machine (SVM), Random Forest (RF), dan K-Nearest Neighbors (K-NN).	Evaluasi dari tiga algoritma pembelajaran mesin: Support Vector Machine (SVM), Random Forest (RF), dan K-Nearest Neighbors (K-NN). Studi ini bertujuan untuk mengidentifikasi algoritma yang paling efisien untuk klasifikasi risiko diabetes berdasarkan glukosa darah, tekanan darah, BMI, dan usia, menggunakan dataset Kaggle NIH yang berisi 768 entri data pasien.

10.	( Trisya Septiana, Mona Arif Muda, Deny Budiyanto, dan Mahendra Pratama, 2024)	Analisis Penggunaan Support Vector Machine pada Deteksi Dini Penyakit Diabetes Melitus	Eksperimen berbasis machine learning Support Vector Machine (SVM)	Support Vector relevan sebagai metode untuk mengklasifikasikan penyakit diabetes, terutama jika digunakan dengan tahapan preprocessing yang ketat dan validasi model yang cukup matang. Ini juga menunjukkan kemungkinan integrasi SVM ke dalam sistem diagnosis berbasis aplikasi web yang umumnya dapat diakses.
-----	--	--	---	--

## 2.2. Definisi Penyakit Diabetes

Kemampuan tubuh untuk mengatur gula darah yang tidak normal adalah penyebab penyakit jangka panjang yang dikenal sebagai diabetes. Penyakit ini disebabkan oleh tubuh menggunakan insulin yang tidak tepat atau pankreas tidak dapat membuat jumlah insulin yang cukup. Insulin adalah hormon penting yang membantu sel-sel tubuh menyerap glukosa dari makanan dan mengubahnya menjadi energi. Jika insulin yang tersedia tidak mencukupi atau tubuh tidak dapat bereaksi dengan tepat terhadap insulin, glukosa akan terakumulasi dalam darah, menyebabkan peningkatan kadar gula darah (King H, Aubert RE, Herman WH, 1998).

Salah satu akibat serius dari penyakit diabetes melitus yang tidak terkontrol adalah kerusakan pembuluh darah akibat kadar glukosa darah yang tinggi secara terus-menerus. Kondisi ini dapat menyebabkan hipertensi atau tekanan darah tinggi, penyakit jantung koroner, serangan jantung (infark miokard), stroke iskemik maupun hemoragik, dan komplikasi kardiovaskular lainnya. Hipoglikemia yang berlangsung lama memperburuk disfungsi endotel, peradangan sistemik, dan peningkatan stres oksidatif, yang semuanya mempercepat proses aterosklerosis, yaitu penumpukan plak di dinding arteri yang dapat menyumbat aliran darah ke organ penting.

Pada organ ginjal, diabetes melitus yang tidak terkontrol dapat menyebabkan komplikasi serius berupa nefropati diabetik, suatu kondisi yang menyebabkan kerusakan progresif pada glomerulus, unit penyaring ginjal, akibat kadar glukosa darah yang tinggi dalam jangka waktu lama. Nefropati diabetik adalah komplikasi serius diabetes melitus yang tidak terkontrol, yang dapat menyebabkan gagal ginjal kronis jika tidak segera dikenali dan ditangani dengan benar. Untuk mempertahankan fungsi vital tubuh, penderita mungkin memerlukan terapi pengganti ginjal, seperti hemodialisis atau bahkan transplantasi ginjal, di kemudian hari.

Pada kesehatan mata melalui komplikasi retinopati diabetik. Mikroangiopati, kerusakan pada lapisan tipis jaringan di bagian belakang mata yang menangkap cahaya dan mengirimkan sinyal visual ke otak, menyebabkan kondisi ini. Pembuluh darah retina dapat melemah, bocor, atau tumbuh secara abnormal ketika kadar gula darah tubuh tidak terkendali. Retinopati diabetik dapat berkembang menjadi edema makula diabetik atau retinopati proliferatif, keduanya sangat berbahaya dan dapat menyebabkan kebutaan permanen ( K. Khoirudin Mohammad Burhan Hanif, 2020 ).

Individu yang menderita diabetes melitus seringkali menunjukkan luka pada kulit yang lebih lambat dibandingkan dengan orang yang sehat. Hal ini disebabkan oleh kerusakan pembuluh darah dan masalah sistem kekebalan yang sering terjadi bersamaan dengan diabetes kronis. Luka yang tidak diobati dapat berkembang menjadi infeksi yang serius, yang dalam beberapa kasus dapat menyebabkan amputasi. Situasi ini menjadikan diabetes bukan sekadar masalah kesehatan individu tetapi juga masalah besar bagi sistem kesehatan masyarakat secara global karena biaya pengobatan yang tinggi dan efek jangka panjangnya terhadap produktivitas penderita.

### **2.3. Klasifikasi Penyakit Diabetes**

Diabetes diklasifikasikan menjadi tiga jenis utama :

1. Diabetes tipe 1 adalah penyakit autoimun di mana sistem kekebalan tubuh menghasilkan sel beta pankreas, atau sel yang bertanggung jawab untuk memproduksi insulin. Kondisi ini mengakibatkan pankreas tidak dapat memproduksi cukup insulin, atau bahkan kadang-kadang tidak dapat memproduksi insulin secara konsisten. Insulin adalah hormon yang sangat penting karena membantu sel-sel tubuh menyerap glukosa dari makanan. Diabetes tipe 1 lebih umum ditemukan pada anak- anak dan remaja, meskipun juga dapat memengaruhi orang dewasa. Suplementasi insulin harus diberikan kepada individu dengan diabetes tipe 1 untuk menjaga kadar gula darah tetap normal dan mencegah komplikasi. Ini menunjukkan bahwa pengobatan diabetes tipe 1 memerlukan pendekatan yang terfokus dan penuh perhatian.
2. Diabetes tipe 2 adalah jenis penyakit yang paling umum dan berkembang secara bertahap dalam jangka waktu tertentu. Dalam kondisi ini, tubuh mulai mengembangkan resistensi terhadap insulin, yang berarti bahwa tubuh tidak merespons insulin dengan efektif. Selain itu, pankreas juga menghambat kemampuan tubuh untuk memproduksi insulin dalam jumlah yang sesuai untuk memenuhi kebutuhan tubuh. Beberapa faktor utama yang berkontribusi pada diabetes tipe 2 adalah obesitas, kebiasaan makan yang tidak sehat, kurangnya aktivitas fisik, dan faktor genetik. Berlawanan dengan diabetes tipe 1, yang lebih umum terjadi pada orang muda, diabetes tipe 2 biasanya ditemukan pada orang dewasa.

3. Diabetes gestasional adalah jenis diabetes yang hanya muncul selama masa kanak-kanak. Kondisi ini disebabkan oleh perubahan hormon yang mengganggu kemampuan tubuh untuk mengatur kadar gula darah, sehingga kadar gula dalam darah menjadi lebih tinggi dari normal. Ibu hamil yang menderita diabetes gestasional membutuhkan perhatian khusus karena kondisi ini memengaruhi tidak hanya kesehatan ibu tetapi juga janin yang masih dalam kandungan. Meskipun diabetes gestasional biasanya hilang selama proses penyembuhan, wanita yang sebelumnya mengalaminya memiliki risiko lebih tinggi untuk mengembangkan diabetes tipe 2 sepanjang siklus menstruasi. Dampak diabetes gestasional terhadap ibu dan bayi dapat dikurangi dengan perawatan dan pengobatan yang tepat.

## **2.4. Faktor Risiko Penyebab Penyakit Diabetes**

Obesitas adalah salah satu faktor risiko terbesar, terutama untuk diabetes tipe 2. Salah satu penyebab utama diabetes adalah resistensi insulin, yang dapat disebabkan oleh lemak dalam tubuh, terutama di area perut. Selain itu, riwayat keluarga dengan diabetes juga memainkan peran penting. Jika seseorang memiliki anggota keluarga, seperti orang tua atau saudara kandung, yang menderita diabetes, kemungkinan kondisi ini makin memburuk. Faktor usia juga tidak bisa dihilangkan, karena diabetes meningkatkan risiko kejang sebagai respons terhadap perubahan perubahan fungsi tubuh dan aktivitas fisik ( K. Nugrogo, 2023 ).

## **2.5. Pencegahan dan Deteksi Dini**

Dengan mendeteksi diabetes lebih awal, pencegahan dapat mencegah berbagai komplikasi yang mungkin terjadi dan mengurangi frekuensi penyakit tersebut. Mendidik tentang pentingnya menjaga gaya hidup sehat adalah komponen penting dalam pengobatan diabetes. Ini meningkatkan kesadaran tentang pentingnya mengadopsi kebiasaan makan sehat, seperti mengurangi konsumsi daging dan ikan serta meningkatkan konsumsi makanan tanpa daging. Manfaat berolahraga dalam menjaga kadar gula darah yang stabil juga penting untuk diketahui oleh semua orang. Deteksi dini melalui pemeriksaan kadar glukosa darah berkala juga berfungsi sebagai alat strategis dalam perjuangan melawan diabetes.

Penelitian ini dapat membantu mengidentifikasi mereka yang memiliki risiko tinggi mengembangkan diabetes atau yang sudah berada dalam keadaan pra-diabetes. Diabetes tipe 2 dapat dicegah atau dibalikkan melalui intervensi diet, seperti perubahan gaya hidup atau pengobatan. Penerapan perubahan gaya hidup sehat seperti pola makan seimbang dan aktivitas fisik teratur, dan peningkatan akses ke layanan kesehatan berkualitas tinggi, seperti deteksi dini diabetes, pengobatan yang tepat, dan pemantauan kondisi yang berkelanjutan, adalah cara yang lebih efektif untuk melawan dan mengelola diabetes.

## 2.6. Teknologi dalam Pencegahan dan Pengelolaan Diabetes

Terobosan teknologi, terutama dalam pendidikan pembelajaran mesin, telah membuka peluang besar untuk pengobatan diabetes. Salah satu inovasi teknologi terbaru yang digunakan untuk memprediksi kemungkinan terkena diabetes adalah Algoritma Vector Support (SVM). Algoritma ini menganalisis data kesehatan dan membaginya ke dalam kategori yang sesuai dengan menggunakan *hyperplane* dengan margin maksimum. Ini memungkinkan analisis data yang lebih akurat dan efisien ( H. Ichwanul Muslim Karo Karo, 2022 ).

Program ini menggunakan teknologi support vector machines (SVM) untuk menilai berbagai data kesehatan pengguna, seperti kadar gula darah, asupan makanan, tingkat aktivitas fisik, dan faktor risiko lainnya. DIATECT dapat menghasilkan hasil prediksi yang akurat dengan menggabungkan kumpulan data yang besar dan beragam. DIATECT tidak hanya berfungsi sebagai alat prediksi, tetapi juga memberikan informasi tentang prediksi dini diabetes. Program ini menurunkan faktor risiko diabetes, meningkatkan manajemen masalah kesehatan masyarakat umum, dan menyebarkan informasi tentang risiko penyakit melalui platform berbasis web.

Inovasi seperti ini menunjukkan bagaimana teknologi dapat menjadi alat yang sangat berguna dalam mengatasi masalah kesehatan global seperti diabetes dan mendorong orang untuk mengendalikan diri mereka sendiri. Sebuah inovasi teknologi yang disebut DIATECT (Perancangan Sistem Diagnosa Penyakit Diabetes. Menggunakan Aplikasi DIATECT Berbasis Web) telah dikembangkan untuk membantu orang memahami risiko terkena diabetes.

## 2.7. Algoritma Support Vector Machine (SVM)

### 2.7.1. Definisi Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu algoritma pembelajaran mesin yang termasuk dalam metode *supervised learning* dan telah menjadi pilihan utama dalam berbagai aplikasi kecerdasan buatan. Algoritma ini dirancang khusus untuk menangani tugas-tugas seperti klasifikasi dan regresi dengan tingkat akurasi yang tinggi. Dalam penggunaannya, SVM bekerja dengan mencari sebuah *hyperplane* atau bidang pemisah yang paling optimal. *Hyperplane* ini bertujuan untuk memisahkan data dari berbagai kelas dalam ruang fitur sehingga model dapat mengenali pola atau hubungan dalam data tersebut ( F. R. Lumbanraja, 2023 ).

Proses pencarian *hyperplane* terbaik dalam algoritma Support Vector Machine (SVM) mencakup tahap optimasi yang bertujuan untuk memaksimalkan jarak atau margin antara titik- titik data dari masing-masing kelas terhadap *hyperplane* yang memisahkan mereka. Jarak minimum antara titik-titik data dari masing-masing kelas ke *hyperplane* ini secara teoretis berfungsi sebagai batas pemisah optimal antara dua kelompok data yang berbeda. SVM dapat memaksimalkan margin ini untuk membuat batas klasifikasi yang akurat untuk data pelatihan dan generalisatif untuk data baru yang belum pernah dilihat oleh model sebelumnya ( M. I Gunawan, 2020 ).

Salah satu kekuatan utama SVM adalah kemampuannya untuk bekerja tidak hanya dengan data yang dapat dipisahkan secara linear, tetapi juga dengan data yang memiliki pola non- linear. Untuk mengatasi keterbatasan pemisahan linear, SVM menggunakan pendekatan yang dikenal sebagai *kernel trick*. Melalui teknik ini, SVM memproyeksikan data dari ruang asli ke ruang dimensi yang lebih tinggi, di mana data tersebut dapat dipisahkan secara linear. Fungsi kernel yang umum digunakan meliputi kernel linear, kernel polinomial, dan kernel Radial Basis Function (RBF), masing-masing dipilih sesuai dengan karakteristik dataset yang sedang dianalisis.

SVM telah banyak digunakan dalam berbagai bidang. Termasuk klasifikasi gambar, pengenalan wajah, analisis teks, dan bahkan di bidang medis untuk analisis genomik atau prediksi penyakit. Meskipun algoritma ini memiliki banyak keunggulan, penggunaannya juga memiliki tantangan. Beberapa tantangan termasuk kebutuhan komputasi yang lebih besar saat bekerja dengan dataset besar dan kebutuhan untuk memilih parameter kernel yang tepat untuk mendapatkan hasil terbaik.

### 2.7.2. Komponen Utama Pemrosesan Algoritma SVM

Dalam mencapai tujuan sebuah pemrosesan, ada beberapa komponen utama yang digunakan dalam bahasa pemrograman system SVM.

#### a. *Hyperplane*

*Hyperplane*, yang secara umum didefinisikan sebagai sebuah batas atau garis pemisah dalam ruang fitur, berfungsi untuk memisahkan data ke dalam kelas-kelas yang berbeda berdasarkan karakteristik atau atribut yang dimiliki data tersebut. Dalam penerapannya, pemilihan *hyperplane* dilakukan dengan sangat cermat untuk memastikan bahwa jarak antara *hyperplane* dan titik data terdekat dari setiap kelas, yang dikenal sebagai margin, dapat dimaksimalkan. Margin yang lebih besar menunjukkan bahwa model memiliki kemampuan yang lebih baik untuk membedakan antara kelas-kelas data secara jelas, yang pada gilirannya meningkatkan generalisasi model.

Oleh karena itu, keberadaan dan posisi *hyperplane* yang ideal dalam algoritma Support Vector Machine (SVM) memastikan bahwa model bekerja dengan baik terhadap data pelatihan dan memiliki kemampuan generalisasi yang tinggi terhadap data baru atau belum pernah dilihat sebelumnya. Akibatnya, model yang dihasilkan dapat membuat prediksi yang lebih akurat dan andal dalam berbagai situasi nyata. *Hyperplane* berfungsi sebagai batas yang efektif antara kelas-kelas data, dan ketika dipilih dengan benar, memungkinkan SVM untuk mengenali dan menyesuaikan diri terhadap pola penting yang tersembunyi dalam data.

*Hyperplane* ini didefinisikan oleh persamaan linear:

$$w \cdot x + b = 0$$

di mana :

$w$  adalah vektor bobot (vektor normal) yang menentukan orientasi dari

*hyperplane*,  $b$  adalah bias yang menggeser posisi *hyperplane*,

$x$  adalah vektor data input.

b. Margin

Dalam konteks Support Vector Machine (SVM), margin didefinisikan sebagai jarak antara *hyperplane* dan titik data terdekat dari masing-masing kelas, merupakan komponen krusial dalam proses pemisahan data. Margin ini tidak hanya berfungsi sebagai pengukur seberapa jauh *hyperplane* dapat memisahkan kelas-kelas data, tetapi juga menentukan seberapa baik model mampu menangkap pola-pola yang ada dalam data tanpa terlalu terfokus pada data pelatihan tertentu, margin yang lebih besar menunjukkan bahwa model memiliki ruang toleransi yang lebih besar untuk menangani variasi data yang lebih besar ( H. Amir, 2024 ).

Model Support Vector Machine (SVM) mencapai struktur pemisahan yang ideal dengan memaksimalkan jarak antara *hyperplane* dan titik data terdekat dari masing-masing kelas. Margin yang optimal ini meningkatkan ketahanan model terhadap variasi data dan membantunya memprediksi data yang belum pernah dilihat sebelumnya dengan akurat. Selain itu, optimalisasi margin adalah bagian penting dari algoritma SVM karena berperan penting dalam menghasilkan model pembelajaran mesin yang tidak hanya akurat, tetapi juga andal dan dapat digeneralisasikan di berbagai kondisi. *Overfitting* kondisi di mana model terlalu menyesuaikan diri terhadap data pelatihan sehingga tidak dapat melakukan generalisasi yang efektif pada data baru ( Muttaqin, 2023).

c. Support vectors

Support vectors, yang didefinisikan sebagai titik-titik data yang berada paling dekat dengan *hyperplane*. Karena titik-titik data ini secara langsung memengaruhi posisi dan orientasi *hyperplane*, support vectors sangat penting dalam algoritma Support Vector Machine (SVM). Dalam hal ini, support vectors bukan hanya titik data biasa mereka adalah elemen penting yang menentukan kemampuan *hyperplane* untuk memisahkan kelas data. Karena hanya titik-titik data ini yang berkontribusi langsung pada pembentukan model, sementara titik-titik data lainnya tidak memiliki pengaruh yang signifikan terhadap posisi *hyperplane*, pemilihan vector pendukung sangat penting dalam proses pembelajaran SVM.

d. Kernel

Kernel, yang merupakan sebuah fungsi matematika yang digunakan untuk memproyeksikan data ke dalam ruang dimensi yang lebih tinggi, memainkan peran yang sangat krusial dalam memperluas kemampuan Support Vector Machine (SVM) untuk menangani data yang memiliki pola non-linear. Pada dasarnya, kernel memungkinkan SVM untuk melakukan transformasi terhadap data asli ke dimensi yang lebih tinggi, di mana data yang sebelumnya tidak dapat dipisahkan secara linear, menjadi terpisah dengan jelas dan dapat dibedakan satu sama lain ( N. Pratiwi dan Y. Setyawan, 2021 ).

Teknik kernel memungkinkan Support Vector Machine (SVM) untuk mengubah representasi data ke dalam ruang lebih besar di mana pemisahan linear dapat dilakukan. Ini terjadi meskipun data asli tidak dapat dipisahkan secara linear dalam ruang fitur aslinya. Sebagai hasilnya, transformasi ini tetap efisien secara komputasi karena dilakukan secara implisit tanpa perlu memetakan data secara manual. Oleh karena itu, fungsi kernel sangat penting untuk keberhasilan SVM dalam menangani masalah klasifikasi yang kompleks. Ini terutama berlaku ketika pola hubungan antar fitur dalam data tidak linear dan sulit dipisahkan menggunakan pendekatan tradisional.

e. Parameter C

Parameter C, yang memiliki peran krusial dalam algoritma Support Vector Machine (SVM), berfungsi untuk mengontrol keseimbangan antara dua faktor penting, yaitu ukuran margin yang besar dan tingkat kesalahan klasifikasi pada data pelatihan. Dengan nilai C yang lebih besar, model akan lebih fokus pada upaya untuk meminimalkan kesalahan klasifikasi, sehingga margin yang dihasilkan cenderung lebih kecil dan lebih ketat dalam memisahkan data. Sebaliknya, dengan nilai C yang lebih kecil, model akan memberikan toleransi lebih terhadap kesalahan klasifikasi, memungkinkan margin yang lebih besar dan lebih fleksibel ( F. Rosyking Lumbanraja, 2023 ).

Pengaturan nilai C yang tepat sangat penting karena dapat memengaruhi kemampuan model dalam menggeneralisasi data baru, dengan nilai C yang terlalu besar berisiko menyebabkan *overfitting*, sementara nilai C yang terlalu kecil bisa menyebabkan *underfitting*, di mana model tidak cukup belajar dari data pelatihan. Dengan demikian, parameter C memainkan peran sentral dalam menentukan kualitas dan akurasi model SVM dalam menghadapi berbagai jenis data dan permasalahan klasifikasi.

f. Gamma

Gamma, yang merupakan salah satu parameter penting dalam kernel Radial Basis Function (RBF) dan Polynomial, memiliki peran yang sangat menentukan dalam mengontrol sejauh mana pengaruh sebuah titik data terhadap titik data lainnya dalam ruang fitur. Pemilihan nilai gamma yang tepat sangat penting dalam menyeimbangkan antara kemampuan model untuk menangkap pola-pola yang sangat spesifik di data serta kemampuan untuk melakukan generalisasi terhadap data yang lebih luas, sehingga memengaruhi akurasi dan efektivitas model dalam berbagai konteks aplikasi pembelajaran mesin ( H. Ichwanul Muslim Karo Karo, 2022 ).

Secara khusus, gamma menentukan seberapa besar kontribusi titik data tertentu terhadap pembentukan model, dengan nilai gamma yang lebih tinggi membuat pengaruh titik data lebih terkonsentrasi pada titik-titik yang lebih

dekat, sehingga model akan lebih sensitif terhadap perubahan lokal dan lebih mudah mengidentifikasi pola atau hubungan yang ada pada data yang sangat spesifik. Sebaliknya, nilai gamma yang lebih rendah memperluas pengaruhnya, sehingga titik data yang lebih jauh juga ikut memengaruhi keputusan model.

g. Loss Function

Loss function adalah sebuah fungsi matematika yang berperan penting dalam proses pelatihan model pembelajaran mesin, khususnya dalam mengukur sejauh mana kesalahan terjadi ketika model mencoba memisahkan kelas-kelas data pada data pelatihan. Fungsi ini memberikan penilaian terhadap seberapa baik atau buruk model dalam mengklasifikasikan data, dengan menghitung selisih antara hasil prediksi model dan nilai yang sebenarnya. Dalam konteks SVM, loss function digunakan untuk menghitung kesalahan yang terjadi ketika titik data tidak terpisah dengan sempurna oleh *hyperplane*, dan tujuan utama dari pelatihan adalah untuk meminimalkan nilai fungsi loss ini ( J. Wira, 2020 ).

Dengan cara ini, model dapat diperbaiki secara bertahap untuk mengurangi kesalahan klasifikasi, sehingga meningkatkan kemampuan model dalam melakukan generalisasi pada data yang tidak terlihat sebelumnya, dan menghasilkan prediksi yang lebih akurat dan andal. Fungsi loss yang tepat sangat berpengaruh terhadap kualitas model, karena memandu proses optimasi agar model dapat menemukan solusi terbaik untuk masalah klasifikasi yang dihadapi.

h. Optimization

Optimization algorithm adalah sebuah metode matematika yang sangat penting dalam proses pelatihan Support Vector Machine (SVM), yang bertujuan untuk menemukan *hyperplane* terbaik yang dapat memisahkan kelas-kelas data dengan cara yang paling optimal. Proses ini dilakukan dengan memaksimalkan margin antara kelas-kelas data, yang berarti memperbesar jarak antara *hyperplane* dan titik data terdekat dari setiap kelas. ( N. Pratiwi dan Y. Setyawan, 2021 ).

Algoritma optimasi sangat penting untuk menentukan efektivitas model dalam membedakan dua kelas data berbeda dalam Support Vector Machine (SVM). Dalam proses ini, tujuan adalah untuk mencapai keseimbangan antara dua tujuan utama yang saling berhubungan tetapi seringkali bertentangan. Tujuan pertama adalah memaksimalkan margin, yaitu jarak antara *hyperplane* ( garis atau bidang pemisah ) dengan titik data kelas terdekat.

### 2.7.3. Prinsip Kerja Support Vector Machine ( SVM )

Prinsip kerja algoritma ini melibatkan beberapa langkah utama:

- a. Pencarian *Hyperplane* : Proses identifikasi *hyperplane* yang optimal, yang mampu memisahkan dua kelas data dengan margin yang paling lebar.
- b. Pemaksimalan Margin : Fokus pada upaya untuk memaksimalkan jarak antara *hyperplane* dan titik data terdekat dari setiap kelas, yang dikenal sebagai support vectors.
- c. Klasifikasi : Penentuan kelas suatu data dilakukan dengan mengidentifikasi sisi *hyperplane* tempat data tersebut berada, yang mengarah pada keputusan klasifikasi.
- d. Margin : Implementasi variabel slack untuk menangani kasus di mana data tidak dapat dipisahkan secara linear, dengan tujuan untuk mencapai keseimbangan antara margin dan toleransi terhadap kesalahan klasifikasi.
- e. Kernel : Penggunaan fungsi kernel untuk memetakan data ke ruang dimensi yang lebih tinggi, memungkinkan pemisahan linear meskipun data dalam ruang asli tidak dapat dipisahkan dengan garis lurus.
- f. Dual Formulasi : Penerapan teknik Lagrange multipliers dalam formulasi dual untuk menyelesaikan masalah optimasi dengan lebih efisien, memungkinkan penanganan masalah yang lebih kompleks dengan lebih efektif.

#### 2.7.4. Kelebihan dan Kelemahan Support Vector Machine ( SVM )

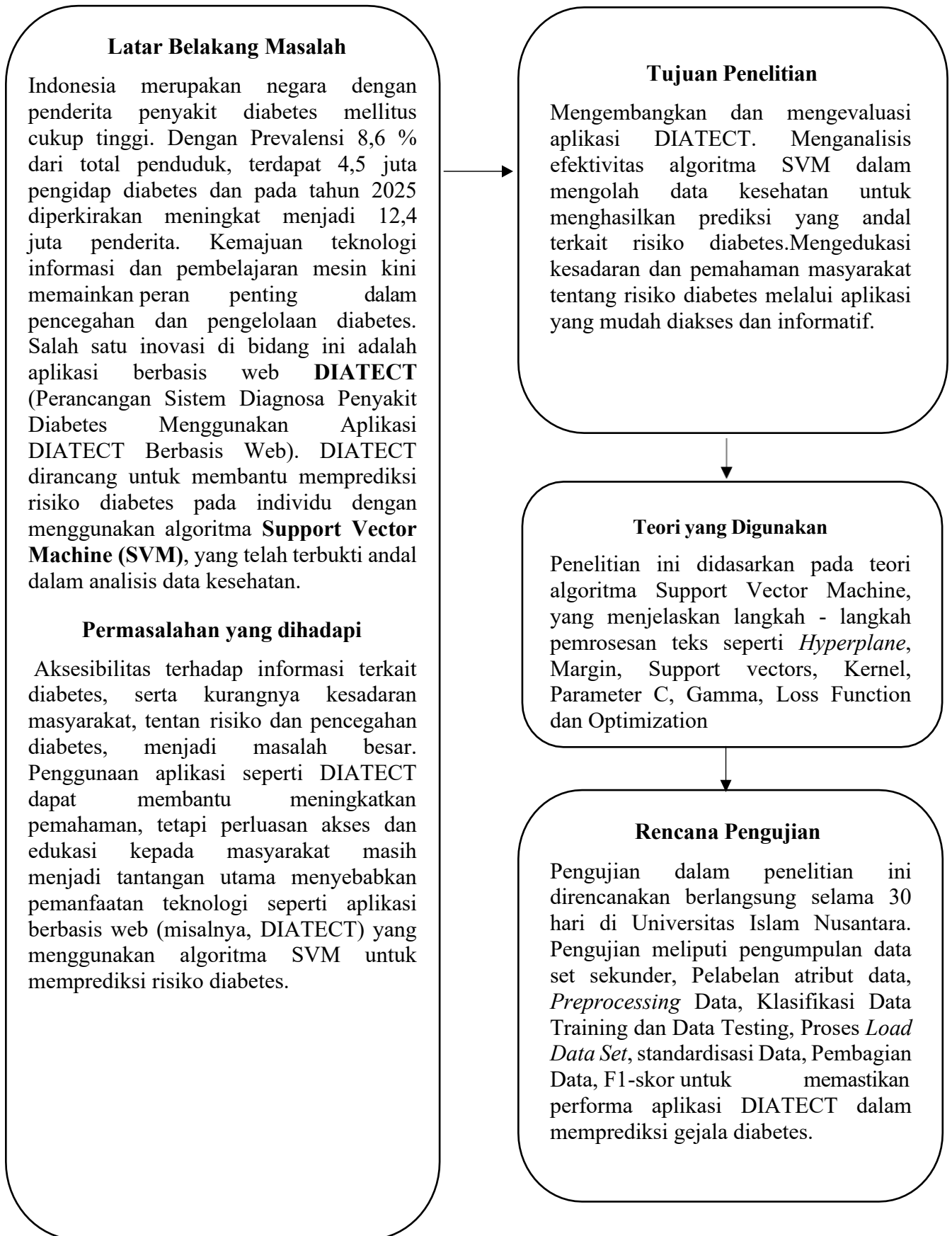
##### a. Kelebihan

1. Komputasi yang mahal untuk dataset besar.
2. Memerlukan tuning parameter yang cermat untuk menghindari *overfitting* atau *underfitting*.
3. Tidak efisien untuk data yang sangat besar atau noisy.
4. Kesulitan dalam penerapan multiklasifikasi tanpa modifikasi.
5. Pemilihan kernel yang tepat bisa menjadi tantangan.

##### b. Kelemahan

1. Kompleksitas komputasi yang tinggi pada dataset besar.
2. Sensitivitas terhadap pemilihan parameter yang tepat (C dan gamma).
3. Kinerja buruk pada data yang sangat besar.
4. Kesulitan menangani data yang sangat noisy atau tumpang tindih.
5. Kesulitan dalam multiklasifikasi tanpa modifikasi.
6. Pemilihan kernel yang tepat bisa menjadi tantangan.
7. Pemrosesan data yang tidak terpisahkan secara linear.

### 2.7.5. Kerangka Berpikir



## BAB 3

### PERANCANGAN SISTEM DAN IMPLEMENTASI SISTEM

#### 3.1. Metode Penelitian

##### 3.1.1. Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kualitatif yang dirancang untuk mengembangkan dan menguji efektivitas aplikasi berbasis web DIATECT dalam memprediksi risiko diabetes. Pendekatan kualitatif dipilih karena kemampuannya untuk memberikan analisis yang sistematis, terukur, dan berbasis data numerik, sehingga mampu menghasilkan informasi yang akurat dan dapat diandalkan dalam mengevaluasi kinerja algoritma Support Vector Machine (SVM). Penelitian ini bertujuan untuk menganalisis hubungan antara pola hidup sehat dan tidak sehat terhadap risiko diabetes, dengan memanfaatkan data numerik yang diolah secara komprehensif melalui algoritma SVM, yang telah terbukti andal dalam klasifikasi dan prediksi data kesehatan yang kompleks.

Dalam penelitian ini, algoritma SVM digunakan untuk memproses data kesehatan dengan pendekatan yang sangat terstruktur, memungkinkan identifikasi *Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age* dan *Outcome*. Melalui penggunaan trik kernel, algoritma SVM mampu memproyeksikan data ke dalam ruang fitur yang lebih tinggi, sehingga dapat membangun model prediksi yang lebih fleksibel dan akurat, bahkan ketika data yang tersedia memiliki distribusi yang kompleks. Proses ini tidak hanya menghasilkan prediksi yang lebih presisi, tetapi juga memungkinkan evaluasi menyeluruh terhadap efektivitas algoritma dalam mendukung pencegahan dan manajemen diabetes secara lebih luas.

### 3.1.2. Jenis Penelitian

Penelitian ini mengadopsi pendekatan kualitatif dengan metode eksperimen untuk menilai efektivitas algoritma Support Vector Machine (SVM) dalam memprediksi risiko diabetes menggunakan data numerik yang relevan. Proses penelitian ini dilakukan melalui serangkaian tahapan yang sistematis, dimulai dengan pengumpulan dataset yang berisi berbagai atribut yang berperan dalam menentukan risiko seseorang terkena diabetes. Atribut-atribut tersebut mencakup *Pregnancies* (jumlah kehamilan), *Glucose* (kadar glukosa dalam darah), *Blood Pressure* (tekanan darah), *Skin Thickness* (ketebalan lipatan kulit), *Insulin* (tingkat insulin dalam tubuh), *Body Mass Index* (Indeks Massa Tubuh), *Diabetes Pedigree Function* (faktor keturunan diabetes), *Age* (usia), serta *Outcome*, yang menunjukkan apakah individu tersebut didiagnosis menderita diabetes atau tidak.

Algoritma SVM diterapkan untuk menganalisis data tersebut, yang bertujuan untuk menghasilkan model prediksi yang dapat mengidentifikasi individu yang berisiko tinggi terkena diabetes. Akurasi dari algoritma SVM diukur menggunakan berbagai metrik evaluasi, seperti *precision*, *recall*, dan F1-skor, yang memberikan gambaran seberapa baik algoritma ini dalam melakukan prediksi yang tepat. Selain itu, penelitian ini juga mencakup perbandingan kinerja algoritma SVM dengan metode prediksi lainnya, untuk menilai keunggulan dan keandalan SVM dalam konteks prediksi risiko diabetes.

Penelitian ini juga mengungkap komponen evaluatif, yang berfokus pada penilaian efektivitas aplikasi berbasis web, DIATECT, dalam mendukung pencegahan dan manajemen diabetes. Penelitian ini akan mengevaluasi beberapa aspek dari aplikasi DIATECT, dimulai dengan penilaian dari pengguna mengenai antarmuka aplikasi dan kemudahan penggunaannya.

Hal ini bertujuan untuk mengetahui sejauh mana aplikasi ini dapat diakses dan digunakan secara efektif oleh berbagai kalangan. Selain itu, penelitian ini juga akan menganalisis dampak penggunaan aplikasi terhadap peningkatan kesadaran pengguna mengenai risiko diabetes, serta seberapa besar aplikasi ini mampu memberikan prediksi yang akurat dan relevan. Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam memahami dan meningkatkan efektivitas aplikasi DIATECT dalam mengurangi memprediksi dini penyakit diabetes.

### **3.1.3 Populasi dan Sampel Penelitian**

Populasi yang digunakan pada penelitian ini terdiri dari 2 kategori yaitu terkena diabetes dan tidak terkena diabetes, yang teridentifikasi dari 9 atribut yaitu *Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Body Mass Index, Diabetes Pedigree Function, Age* dan *Outcome*. Proses pemilihan dan pelabelan data dilakukan secara manual untuk memastikan akurasi dan konsistensi yang tinggi.

Penelitian ini menggunakan stratified sampling, peneliti dapat membagi populasi menjadi beberapa strata ( sub kelompok ) yang relevan berdasarkan usia, dan status kesehatan, kemudian memilih sampel secara acak dari setiap strata untuk memastikan representatif yang seimbang. Jumlah sampel yang digunakan dalam penelitian ini adalah 768 data sekunder yang terdiri dari:

1. 500 data teridentifikasi tidak terkena diabetes
2. 268 data teridentifikasi terkena diabetes

### 3.1.4. Sumber Data

Penelitian ini menggunakan data sekunder yang diambil dari platform Kaggle, dengan judul *Diabetes Data Set*. Dataset ini dipilih karena relevansi dan kelengkapannya untuk mendukung pengembangan model prediksi risiko diabetes yang akurat. Dataset tersebut berisi 768 entri data, yang masing-masing memiliki 9 atribut utama, yaitu *Pregnancies* (jumlah kehamilan), *Glucose* (kadar glukosa), *Blood Pressure* (tekanan darah), *Skin Thickness* (ketebalan kulit), *Insulin* (kadar insulin), *Body Mass Index* (indeks massa tubuh), *Diabetes Pedigree Function* (faktor genetik diabetes), *Age* (usia), dan *Outcome* (hasil diagnosis). Atribut-atribut ini dipilih karena memiliki keterkaitan signifikan dengan risiko seseorang menderita diabetes.

Data ini selanjutnya dikonversi ke dalam format CSV agar dapat diolah dan dianalisis menggunakan pustaka pemrograman Python, seperti Pandas untuk manipulasi data, NumPy untuk perhitungan numerik, dan pustaka pembelajaran mesin untuk pengembangan model. Proses pelabelan pada dataset dilakukan secara manual dengan menggunakan sistem klasifikasi biner. Label 0 digunakan untuk mengindikasikan bahwa pasien tidak menderita diabetes, sedangkan label 1 menunjukkan bahwa pasien terdiagnosis diabetes. Proses pelabelan ini dilakukan dengan hati-hati untuk memastikan akurasi dalam pengelompokan data, yang sangat penting dalam membangun model prediksi yang andal.

Pemilihan dataset ini dilakukan dengan pertimbangan bahwa data bersifat lengkap dan mencakup atribut-atribut penting yang relevan dengan tujuan penelitian, yaitu memprediksi risiko diabetes secara akurat. Dataset ini memberikan fleksibilitas kepada peneliti untuk menerapkan berbagai teknik analisis, seperti eksplorasi data, visualisasi, dan pembentukan model pembelajaran mesin. Selain itu, ukuran dataset yang terdiri dari 768 entri dianggap cukup representatif untuk melatih dan menguji model, tanpa mengorbankan efisiensi komputasi.

### 3.1.5. Pengumpulan Data

Penelitian ini memanfaatkan data sekunder yang diambil dari platform Kaggle, yang berjudul Diabetes Data Set. Pemilihan dataset ini didasarkan pada relevansi dan kelengkapannya dalam mendukung pengembangan model prediksi risiko diabetes yang lebih akurat. Dataset tersebut terdiri dari 768 entri data, masing-masing berisi 9 atribut penting yang meliputi :

- a. Pregnancies (jumlah kehamilan)
- b. Glucose (kadar glukosa darah)
- c. Blood Pressure (tekanan darah)
- d. Skin Thickness (ketebalan kulit)
- e. BMI (indeks massa tubuh)
- f. Insulin (kadar insulin dalam tubuh)
- g. Diabetes Pedigree Function (faktor genetik diabete)
- h. Age (usia)
- i. Outcome (hasil diagnosis yang menunjukkan apakah individu terdiagnosis diabetes atau tidak).

Dataset yang diperoleh kemudian dikonversi menjadi format CSV agar lebih mudah untuk diproses dan dianalisis menggunakan alat pemrograman Python, seperti pustaka pandas untuk manipulasi data dan NumPy untuk perhitungan numerik. Selain itu, pustaka- pustaka lain yang berkaitan dengan pembelajaran mesin juga digunakan untuk membangun model prediksi. Proses pelabelan dataset dilakukan secara manual dengan menggunakan sistem klasifikasi biner, di mana :

1. label 0 menunjukkan bahwa pasien tidak menderita diabetes.
2. label 1 menunjukkan bahwa pasien terdiagnosis menderita diabetes.

Pelabelan ini dilakukan dengan sangat teliti untuk memastikan keakuratan data, yang penting dalam menghasilkan model yang tepat dan dapat diandalkan. Pemilihan dataset ini didasarkan pada pertimbangan bahwa data tersebut cukup lengkap dan mencakup atribut- atribut yang penting dan relevan untuk tujuan penelitian, yaitu untuk memprediksi risiko diabetes dengan akurasi tinggi. Dengan jumlah 768 entri data, dataset ini dianggap cukup representatif dan memadai untuk digunakan dalam proses pelatihan serta pengujian model prediksi tanpa mengorbankan efisiensi komputasi.

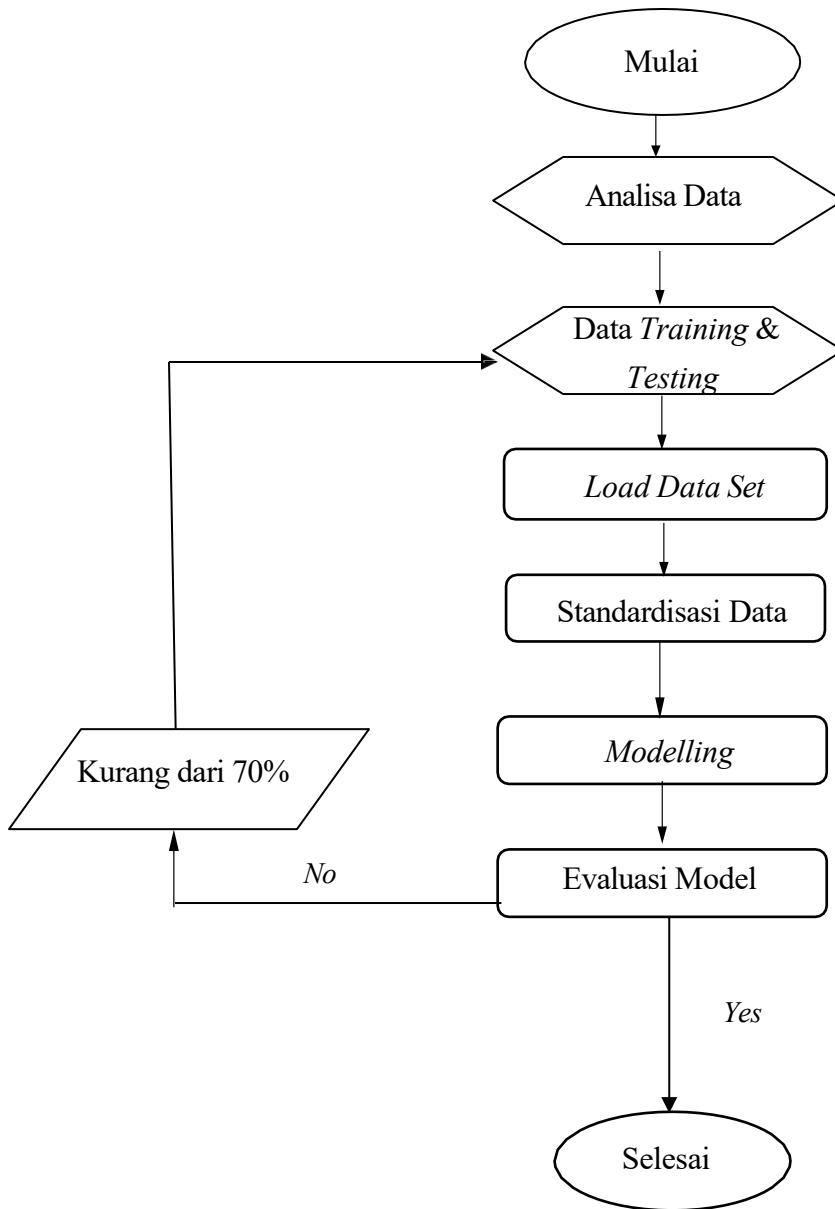
Secara keseluruhan, proses pengumpulan dan pengolahan data dilakukan dengan hati- hati untuk memastikan kualitas data yang digunakan dalam penelitian ini dapat mendukung tujuan utama penelitian, yaitu mengembangkan model prediksi yang efektif yang dapat membantu dalam pencegahan dan pengelolaan diabetes.

### **3.2. Perencanaan Blok Diagram**

Tahapan ini menggambarkan suatu pendekatan yang dirancang secara sistematis dan terperinci untuk menjelaskan setiap prosedur yang terlibat dalam proses pengembangan program yang bertujuan untuk memprediksi kemungkinan seseorang menderita penyakit diabetes. Dalam proses ini, langkah-langkah yang dilakukan mencakup pengumpulan data yang relevan dan memiliki signifikansi tinggi terhadap faktor-faktor risiko diabetes, seperti data klinis, riwayat kesehatan pasien, serta parameter fisiologis yang berhubungan dengan kadar glukosa dalam tubuh.

Data yang telah dikumpulkan kemudian melalui serangkaian tahap pengolahan, termasuk pembersihan data, normalisasi, serta analisis awal untuk memastikan kualitas dan keakuratan informasi yang digunakan dalam sistem deteksi diabetes. Dengan pendekatan yang sistematis ini, penelitian bertujuan untuk menghasilkan suatu model prediksi yang dapat digunakan secara efektif dalam membantu tenaga medis atau individu dalam mengidentifikasi risiko diabetes secara lebih dini, sehingga langkah-langkah pencegahan atau pengelolaan penyakit dapat dilakukan dengan lebih optimal.

Semua langkah-langkah yang terlibat dalam proses ini, sebagaimana digambarkan dalam Gambar 1 pada *flowchart* penelitian, menunjukkan urutan dan keterkaitan setiap tahap yang diperlukan untuk mencapai tujuan prediksi yang akurat.



**Gambar 3.2.** *Flowchart Sistem*

## BAB 4

### PEMBAHASAN DAN ANALISIS

#### 4.1. Analisis Data

##### 4.1.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan memanfaatkan dataset yang diperoleh dari platform Kaggle. Dataset tersebut tersedia dalam format file *Comma-Separated Values* (CSV), merupakan salah satu format data terstruktur yang umum digunakan dalam penelitian. Data yang digunakan dalam penelitian ini mencakup informasi mengenai individu yang didiagnosis menderita diabetes sebanyak 268 orang, serta individu yang tidak menderita diabetes sebanyak 500 orang. Dengan demikian, total jumlah data yang tersedia untuk dianalisis dalam penelitian ini adalah sebanyak 768 entri. 768 terdiri dari 9 atribut terdiri dari *Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Body Mass Index, Diabetes Pedigree Function, Age* dan *Outcome*. Dataset ini dipilih karena relevansinya dengan topik penelitian dan kelengkapannya dalam menyediakan variabel-variabel yang dibutuhkan untuk menjawab tujuan penelitian.

**Tabel 4.1.1.** Jenis Keterangan Yang Digunakan

Keterangan	label	Jumlah
Terkena	1	268
Tidak terkena	0	500

#### 4.1.2. Pelabelan Atribut Data

Dalam penelitian ini, proses pelabelan data dilakukan secara manual dengan menerapkan sistem klasifikasi berbasis dua kategori utama, yaitu label 0 dan label 1, yang digunakan untuk membedakan kondisi kesehatan setiap pasien dalam dataset. Label 0 diberikan kepada pasien yang berdasarkan data medisnya tidak menunjukkan adanya indikasi atau diagnosis diabetes, Sementara itu, label 1 diberikan kepada pasien yang telah didiagnosis menderita diabetes berdasarkan hasil pemeriksaan medis yang tersedia. Informasi mengenai teks dan label yang telah diberikan dapat dilihat pada Tabel 1, yang berisi rincian data sesuai dengan kategori yang ditetapkan.

**Tabel 4.1.2.** Jenis prediksi beserta Labelnya

Keterangan	Label
Tidak terkena diabetes	0
Terkena diabetes	1

#### 4.1.3. *Preprocessing* Data

Tahap *preprocessing* data merupakan langkah awal yang sangat penting dalam proses pengolahan data, khususnya dalam pengklasifikasian, karena bertujuan untuk meningkatkan kualitas data sehingga dapat menghasilkan hasil analisis yang lebih akurat dan dapat diandalkan. Proses ini mencakup serangkaian langkah yang dilakukan untuk membersihkan, menghilangkan, serta mengubah sumber data agar lebih optimal dan siap digunakan dalam tahapan berikutnya ( Wiratama & Aziz, 2024 ).

Dalam penelitian ini, tahap *preprocessing* data terdiri dari dua langkah utama, yaitu pembersihan data (*data cleaning*) dan transformasi data (*data transformation*). Pembersihan data dilakukan dengan tujuan untuk menghapus atau mengeliminasi data yang tidak relevan atau tidak diperlukan, seperti data kosong, data duplikat, serta data yang dianggap tidak signifikan dalam proses analisis. Proses pembersihan ini dilakukan secara manual untuk memastikan bahwa hanya data yang benar-benar relevan dan memiliki nilai informasi yang tinggi yang akan digunakan dalam tahap selanjutnya.

Setelah proses pembersihan data selesai, langkah berikutnya adalah transformasi data, di mana data diubah dari tipe kategori menjadi tipe *binary* untuk mempermudah pengolahan dan analisis lebih lanjut. Transformasi ini dilakukan agar data lebih mudah diinterpretasikan oleh model klasifikasi yang digunakan, sehingga meningkatkan efisiensi serta akurasi dalam pengklasifikasian data. Dengan demikian, tahap *preprocessing* data berperan krusial dalam memastikan bahwa data yang digunakan memiliki kualitas terbaik, sehingga mendukung hasil yang lebih optimal dalam proses analisis dan klasifikasi.

#### 4.1.4. Klasifikasi Data *Training & Testing*

Klasifikasi adalah suatu proses yang bertujuan untuk membangun sebuah model atau fungsi yang mampu menjelaskan serta membedakan antara berbagai kelas data tertentu dengan mengacu pada tujuan yang telah ditentukan sebelumnya. Dalam proses ini, data yang tersedia dianalisis dan dipelajari guna mengidentifikasi pola-pola yang dapat digunakan untuk mengelompokkan objek atau entitas ke dalam kelas-kelas yang sesuai.

Dengan demikian, klasifikasi berperan dalam memperkirakan atau memprediksi kategori dari suatu objek berdasarkan karakteristik yang dimilikinya, sehingga memungkinkan pengambilan keputusan yang lebih akurat dan sistematis dalam berbagai bidang, seperti data mining, pembelajaran mesin, dan kecerdasan buatan. Umumnya, proses klasifikasi melibatkan dua tahap utama, yaitu:

a. Proses Pelatihan

Pada tahap pelatihan (*training*), digunakan sekumpulan data yang telah memiliki label atau kelas yang diketahui sebelumnya. Data ini digunakan untuk membangun model atau fungsi yang dapat memetakan hubungan antara fitur-fitur data dengan kelas yang menjadi target.

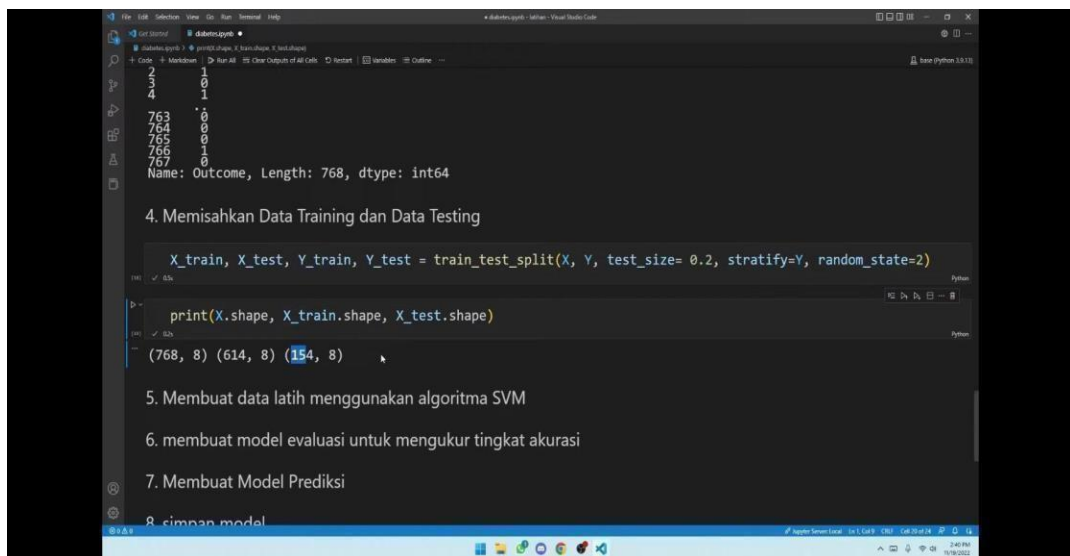
b. Proses Pengujian

Setelah model atau fungsi selesai dibangun pada tahap pelatihan, dilakukan tahap pengujian (*testing*) untuk mengevaluasi keakuratan model yang telah dibuat. Pada tahap ini, digunakan data yang dikenal sebagai *testing set*, yaitu sekumpulan data yang memiliki label yang diketahui, tetapi tidak digunakan selama tahap pelatihan. Tujuannya adalah untuk mengukur sejauh mana model mampu memprediksi kelas data dengan benar.

**Tabel 4.1.4.** Data *Training* dan Data *Testing*

Pembagian	Jumlah
Data latih	614
Data uji	154
Jumlah	768

Data yang di sajikan sebanyak 768 data. Data tersebut akan dibagi menjadi Data *training* dan Data *Testing*. Data yang dihasilkan tercantum dalam tabel di bawah ini. Dapat dilihat dari tabel 2 terdapat 2 kelompok yaitu Data latih (*training*) dan Data uji (*testing*). Kelompok data latih 80% atau sebanyak 614 data sedangkan data uji 20% atau sebanyak 154 data yang dihasilkan.



**Gambar 4.1.4.** Proses Pemisahan Data *Training* dan *Testing*

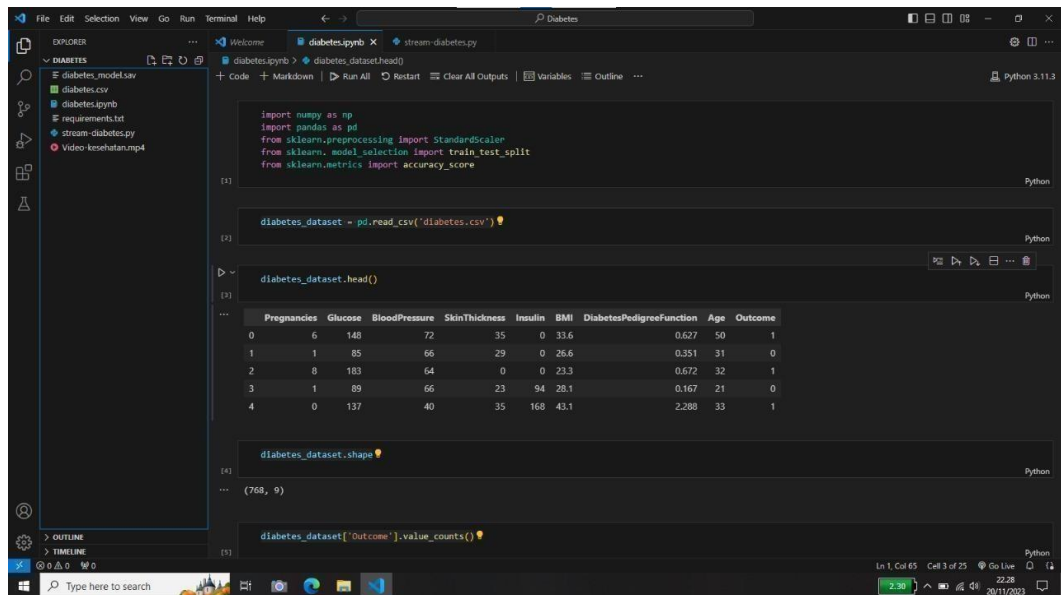
## 4.2. Load Data Set

Dalam proses pengembangan aplikasi DIATECT, yang dirancang sebagai sistem prediksi penyakit diabetes berbasis web menggunakan algoritma Support Vector Machine (SVM) dengan bahasa pemrograman Python, terdapat beberapa tahapan utama yang harus dilakukan, salah satunya adalah proses load dataset. Tahapan ini memiliki peran yang sangat penting karena memastikan bahwa data yang digunakan dalam pelatihan dan pengujian model memiliki kualitas yang baik serta sesuai dengan kebutuhan sistem (M. Ardiansyah, 2021).

Dataset yang digunakan dalam penelitian ini merupakan dataset diabetes yang tersedia dalam format CSV atau Excel. Dataset ini kemudian dipersiapkan dan dibaca menggunakan pustaka Python, seperti *pandas*, yang memungkinkan pengguna untuk melakukan manipulasi serta analisis data dengan lebih efisien.

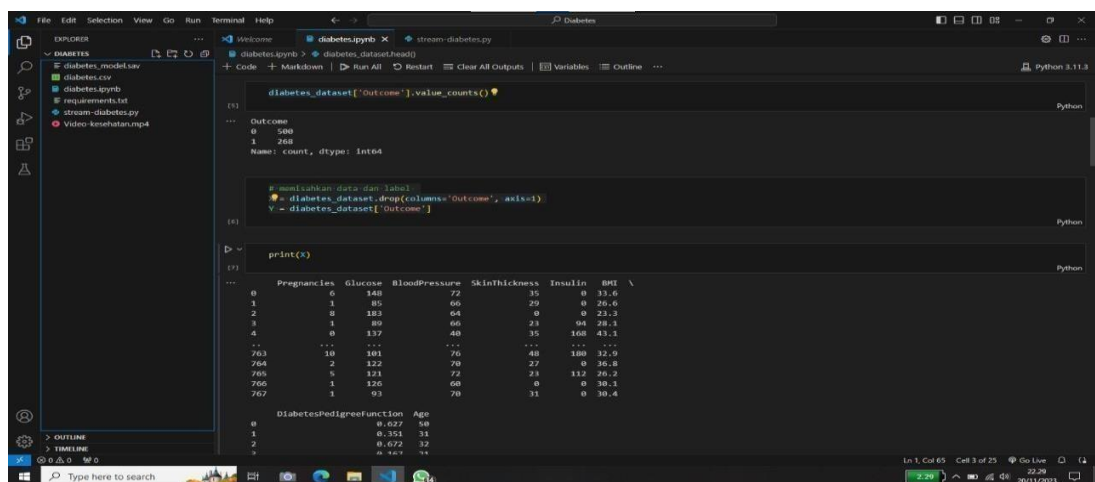
Setelah proses ekstraksi fitur selesai, data yang telah diproses kemudian dimuat ke dalam struktur yang sesuai untuk keperluan pelatihan dan pengujian model SVM. Pada tahap ini, dataset dibagi menjadi dua bagian utama, yaitu data latih (*training set*) yang digunakan untuk membangun model, serta data uji (*testing set*) yang digunakan untuk mengukur performa dan keakuratan model dalam melakukan prediksi.

Langkah berikutnya adalah membangun model SVM, di mana algoritma ini diterapkan untuk melakukan klasifikasi berdasarkan pola yang ditemukan dalam dataset. Model yang telah dilatih dengan data latih akan dievaluasi menggunakan data uji guna menentukan tingkat akurasi serta efektivitasnya dalam melakukan prediksi diabetes. Setelah diperoleh model dengan performa yang optimal, model tersebut kemudian diintegrasikan ke dalam aplikasi web DIATECT.



*Gambar 4.2. Proses Load Data Set*

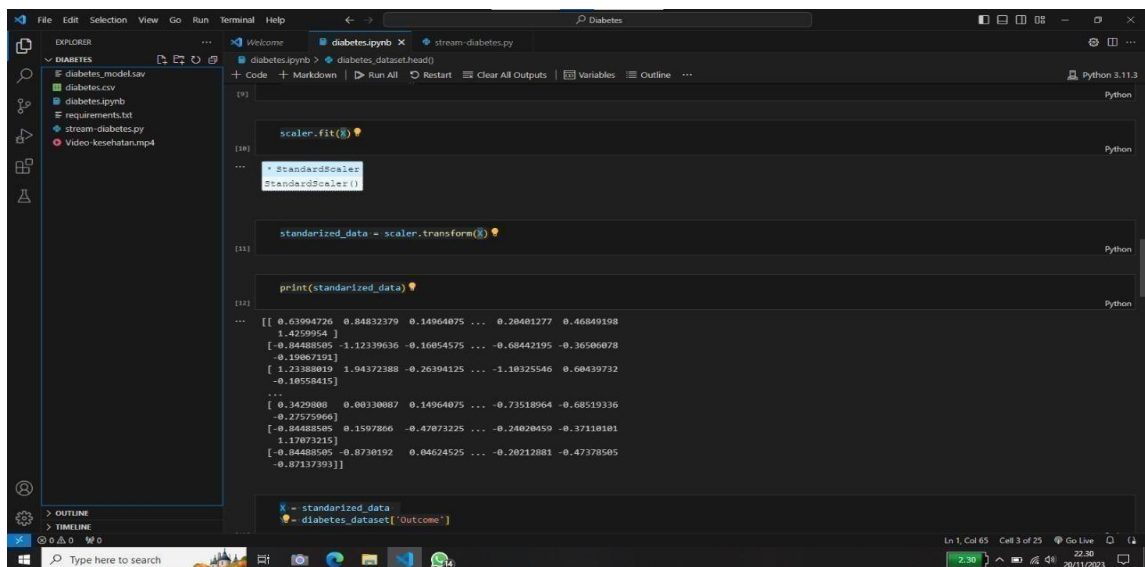
Setelah dataset berhasil dimuat, data tersebut diproses lebih lanjut untuk mengekstrak berbagai atribut penting yang dapat digunakan dalam prediksi diabetes. Beberapa atribut yang diambil meliputi "Usia", "Tekanan Darah", "Indeks Massa Tubuh (BMI)", serta label "Diabetes" menunjukkan apakah seorang individu mengidap diabetes atau tidak. Integrasi model ke dalam aplikasi web memungkinkan pengguna untuk memasukkan data pasien secara real-time dan memperoleh prediksi apakah pasien tersebut berisiko mengidap diabetes atau tidak.



*Gambar 4.2. Proses Baris Data Set*

### 4.3. Standardisasi Data

Dalam pengembangan aplikasi DIATECT untuk memprediksi diabetes menggunakan algoritma Support Vector Machine ( SVM ) berbasis web, proses standardisasi data dilakukan dengan menghitung z-skor pada setiap fitur dalam dataset. Dengan cara ini, setiap nilai dalam dataset dikonversi ke skala yang memiliki rata-rata 0 dan simpangan baku 1, sehingga distribusinya menjadi lebih seragam.



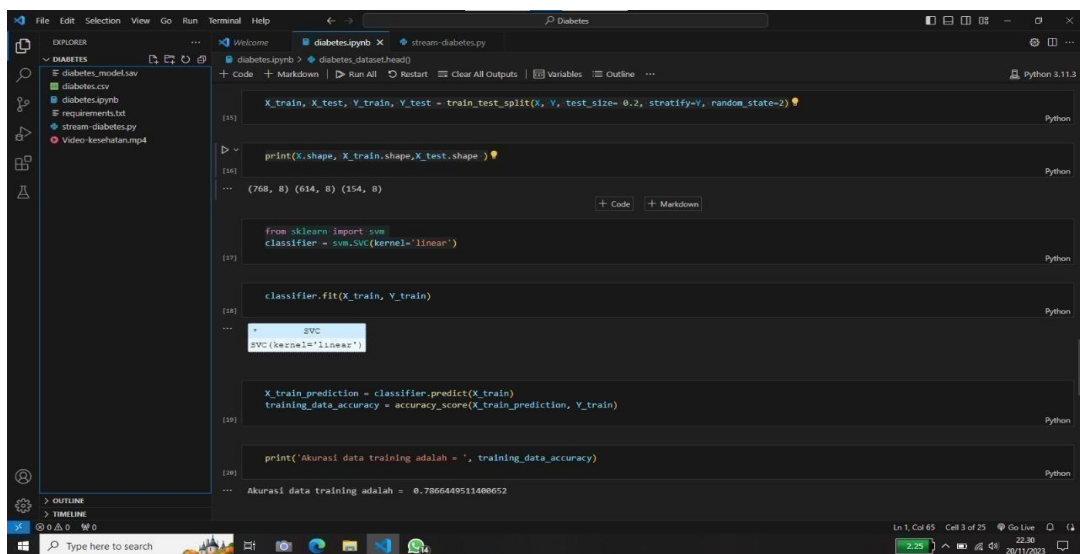
```
diabetes.py  
diabetes_dataset.head()  
scaler.fit(X)  
StandardScaler  
StandardScaler()  
standardized_data = scaler.transform(X)  
print(standardized_data)  
[[ 0.63984726  0.84832379  0.14964875 ... 0.20481277  0.46849198  
  1.4259954  
 [-0.84488505 -1.12339636 -0.16054575 ... -0.68442195 -0.36506078  
  0.15867191  
 [ 1.22388819  1.94372388 -0.26394125 ... -1.10325546  0.68439732  
 -0.19558415  
 ...  
 [ 0.3429588  0.08330887  0.14964875 ... -0.73518964 -0.68519336  
 -0.22572966  
 [-0.84488505  0.1597866 -0.47073225 ... -0.24829459 -0.37118181  
  1.17673215  
 [-0.84488505 -0.8738192  0.04624525 ... -0.20212881 -0.47378505  
 -0.87137393]]
```

*Gambar 4.3. Proses Standardisasi Data*

Standardisasi model lebih mudah mengenali pola dalam data, yang pada akhirnya meningkatkan akurasi dan stabilitas prediksi, ini sangat penting karena algoritma SVM sensitif terhadap perbedaan skala pada fitur input. Jika nilai antar fitur memiliki rentang yang terlalu jauh, model bisa mengalami bias dan menghasilkan prediksi yang kurang akurat. Dengan menyamakan skala semua fitur, model dapat bekerja lebih efisien, mengurangi kemungkinan bias, serta meningkatkan konsistensi hasil prediksi.

#### 4.4. Pembagian Data

Pada pengembangan aplikasi DIATECT untuk memprediksi risiko diabetes, langkah awal yang diambil adalah membagi dataset menjadi dua segmen utama, yaitu set data untuk pelatihan (*training set*) dan set data untuk pengujian (Rahmasari, 2024). Pembagian ini dimaksudkan untuk melatih model dalam mengenali pola-pola spesifik dalam data melalui penggunaan data latih, sekaligus menguji performa model dalam memprediksi data baru yang belum pernah diproses sebelumnya dengan menggunakan data uji.



```
File Edit Selection View Go Run Terminal Help
Diabetes
diabetes.py:nb X stream-diabetes.py
diabetes.py:nb > diabetes_dataset.head()
Python
[153] X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=2)
Python
[154] print(X.shape, X_train.shape, X_test.shape)
Python
.... (768, 8) (614, 8) (154, 8)
Python
[157] from sklearn import svm
classifier = svm.SVC(kernel='linear')
Python
[158] classifier.fit(X_train, Y_train)
Python
.... SVC
SVC(kernel='linear')
Python
[159] X_train_prediction = classifier.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)
Python
[160] print('Akurasi data training adalah = ', training_data_accuracy)
Python
.... Akurasi data training adalah = 0.7866449511400652
Python 3.11.3
Ln 1, Col 65 Cell 3 of 25 Go Live 22:30 26/11/2023
```

Gambar 4.4. Proses Pembagian Data

Untuk mengevaluasi kinerja model, beberapa metrik digunakan, termasuk akurasi, presisi, *recall*, dan F1-skor. Metrik-metrik ini digunakan untuk mengevaluasi kemampuan model untuk melakukan klasifikasi dengan benar serta keseimbangan antara kemampuan model untuk menemukan hasil positif dan negatif terkena diabetes. Model diuji menggunakan data uji yang telah dipisahkan dari data pelatihan untuk mengevaluasi performa sistem dalam menghasilkan akurasi yang akurat.

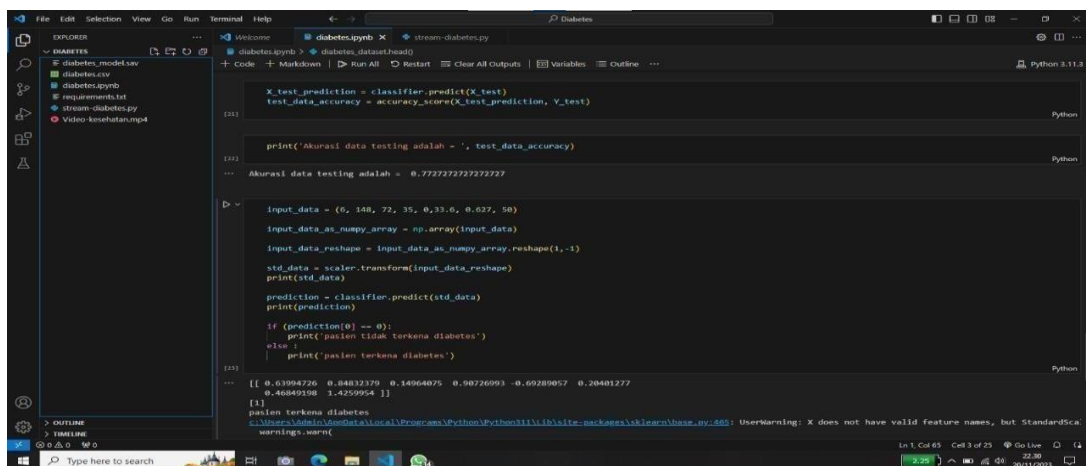
#### 4.5. Modelling

Source code yang digunakan dalam pengembangan aplikasi ini difokuskan pada pembuatan antarmuka web menggunakan Streamlit, yang memungkinkan pengguna melakukan prediksi diabetes secara interaktif. Penelitian ini menggunakan algoritma Support Vector Classification (SVC), yang merupakan komponen dari metode Support Vector Machine (SVM). SVC adalah algoritma klasifikasi yang berfungsi dengan mencari *hyperplane* optimal, yaitu garis atau bidang pemisah terbaik di ruang fitur yang berupaya untuk memaksimalkan jarak antara dua kelas.

Beberapa keunggulan SVC, termasuk kemampuan untuk menangani data dengan dimensi tinggi, stabilitas dalam klasifikasi biner, dan fleksibilitas dalam penerapan berbagai jenis fungsi kernel. SVC juga mampu menciptakan model yang tepat dan efisien dalam mendeteksi pola-pola rumit pada data medis, sehingga sangat cocok untuk digunakan dalam kasus klasifikasi risiko penyakit seperti diabetes. Sebelum model diintegrasikan ke dalam aplikasi, proses pelatihan dilakukan menggunakan dataset yang relevan dan representatif.

Selama tahap pelatihan, model SVM menganalisis berbagai fitur dalam dataset, menyesuaikan parameter-parameter penting, dan mengoptimalkan kinerjanya melalui teknik penyetelan hyperparameter tuning. Setelah model mencapai tingkat akurasi yang optimal, hasil pelatihannya disimpan dalam sebuah file bernama 'diabetes\_model.sav'. File ini berisi seluruh informasi serta parameter yang telah dipelajari model, sehingga dapat digunakan kembali tanpa perlu melakukan pelatihan ulang.

Dalam implementasinya, aplikasi web ini membaca file model yang telah disimpan menggunakan library pandas dan memanfaatkannya untuk melakukan proses prediksi berdasarkan input yang diberikan oleh pengguna. Setelah pengguna memasukkan nilai-nilai yang sesuai, sistem akan mengolah data tersebut melalui model SVM dan memberikan hasil prediksi yang ditampilkan secara jelas di antarmuka aplikasi. Selain menampilkan hasil prediksi, aplikasi ini juga dapat memberikan pesan informatif yang membantu pengguna memahami hasil yang diperoleh.



```
File Edit Selection View Go Run Terminal Help
Python
diabetes.pyrb x stream-diabetes.py
diabetes.pyrb diabetes_data.csv
diabetes_data.csv
diabetes.pyrb
requirements.txt
stream-diabetes.py
Video-kesehatan.mpd

X_test_prediction = classifier.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)

print('Akurasi data testing adalah = ', test_data_accuracy)
... Akurasi data testing adalah = 0.7772727272727272

input_data = (6, 148, 72, 35, 0.33, 0, 0.027, 50)
input_data_as_numpy_array = np.array(input_data)
input_data_reshape = input_data_as_numpy_array.reshape(1,-1)
std_data = scaler.transform(input_data_reshape)
print(std_data)

prediction = classifier.predict(std_data)
print(prediction)

if (prediction[0] == 0):
    print('pasien tidak terkena diabetes')
else:
    print('pasien terkena diabetes')

... [[ 0.63994726  0.84832379  0.14964075  0.90726093 -0.69289057  0.20401277
  0.46849198  1.4259954 ]]
[1]
pasien terkena diabetes
c:\users\dimas\appdata\local\programs\python\python111\h\site-packages\sklearn\base.py:866: UserWarning: X does not have valid feature names, but StandardSca
warnings.warn(

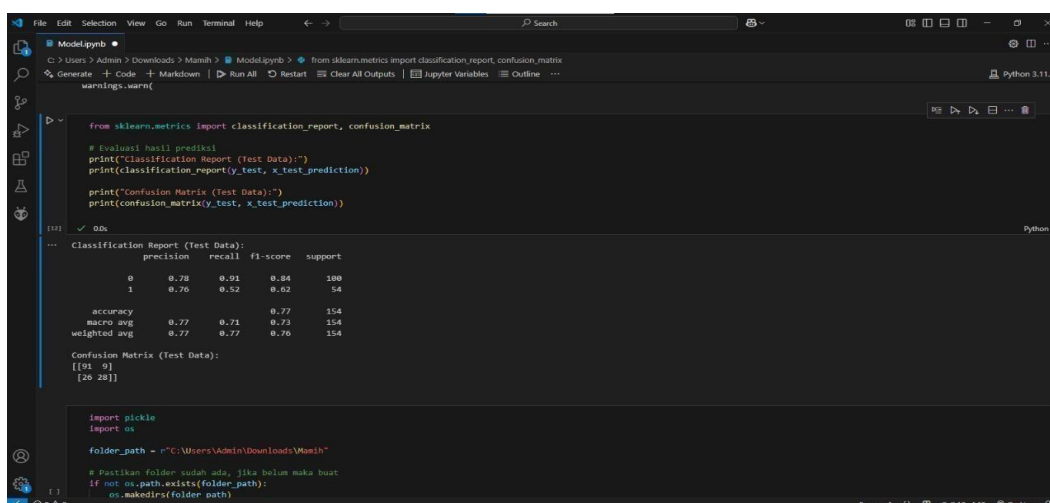
Ln 1, Col 65  Cell 3 of 25  Go Live  22:08  20/11/2023
```

*Gambar 4.5. Modelling*

Kualitas dataset yang digunakan dalam fase pelatihan model SVM sangat menentukan keberhasilan dan kehandalan aplikasi ini. Oleh sebab itu, pemilihan dataset yang merepresentasikan, pengolahan data yang teliti, serta validasi model yang baik adalah faktor penting dalam memastikan bahwa sistem ini mampu memberikan prediksi yang tepat dan dapat dipercaya.

## 4.6. Akurasi

Dalam pengembangan aplikasi pendeteksi diabetes berbasis Support Vector Machine (SVM), akurasi menjadi salah satu aspek penting yang digunakan untuk menilai seberapa baik model dapat mengklasifikasikan data dengan benar. Akurasi ini menunjukkan kemampuan model dalam mengenali pola pada data baru yang belum pernah diproses sebelumnya. Berdasarkan pengujian menggunakan data testing, model ini mencapai tingkat akurasi 77%, yang berarti bahwa dari seluruh data yang diuji, 77% prediksi yang dihasilkan oleh model sesuai dengan label aslinya, sementara 23% lainnya keliru.



```
from sklearn.metrics import classification_report, confusion_matrix

# Evaluasi hasil prediksi
print(classification_report(y_test, x_test_prediction))
print(classification_report(y_test, x_test_prediction))

print("Confusion Matrix (Test Data):")
print(confusion_matrix(y_test, x_test_prediction))
```

```
Classification Report (Test Data):
precision    recall  f1-score   support

     0       0.78      0.91      0.84      100
     1       0.70      0.52      0.62       54

 accuracy          0.77      0.71      0.73      154
 macro avg          0.77      0.77      0.76      154
weighted avg          0.77      0.77      0.76      154

Confusion Matrix (Test Data):
[[ 91  9]
 [20 28]]
```

```
import pickle
import os

folder_path = r"C:\Users\Adnin\Downloads\Masih"

# Pastikan folder sudah ada, jika belum maka buat
if not os.path.exists(folder_path):
    os.makedirs(folder_path)
```

*Gambar 4.6. Akurasi Data*

Dalam penelitian ini, model klasifikasi yang dibuat menunjukkan tingkat akurasi 77% dalam memprediksi potensi seseorang mengidap diabetes, yang dalam perspektif pengklasifikasian medis awal, bisa dianggap cukup efisien. Walaupun angka tersebut belum tergolong sangat tinggi, tetap menunjukkan bahwa model mampu mengidentifikasi pola-pola penting dari parameter input yang kompleks, seperti tingkat glukosa, tekanan darah, insulin, Indeks Masa Tubuh, hingga usia pasien. Kualitas ini jauh lebih unggul dibandingkan dengan baseline acak yang hanya mencapai sekitar 50%, menunjukkan bahwa model tidak sembarangan dalam melakukan prediksi, melainkan berdasarkan pola pembelajaran dari data yang ada.

#### 4.7. Performa Aplikasi

Aplikasi ini dirancang menggunakan Streamlit sebagai platform utama untuk mengembangkan antarmuka pengguna (UI) yang interaktif dan intuitif. Tujuan utama dari aplikasi ini adalah memberikan kemudahan bagi pengguna dalam memprediksi kemungkinan terkena diabetes berdasarkan model yang telah dilatih sebelumnya menggunakan algoritma Support Vector Machine (SVM). Saat pertama kali dijalankan, aplikasi melakukan proses inisialisasi, di mana model SVM yang telah melalui tahap pelatihan sebelumnya dibaca dari file 'diabetes\_model.sav'. Model ini berisi parameter-parameter penting yang telah dipelajari dari dataset pelatihan, sehingga dapat digunakan untuk menganalisis input yang diberikan oleh pengguna.

Meskipun proses pemuatan model dari file dapat memengaruhi waktu respons aplikasi, khususnya jika ukuran model cukup besar, namun dalam kebanyakan kasus, waktu loading model tetap dalam rentang yang dapat diterima dan tidak menyebabkan gangguan yang signifikan terhadap pengalaman pengguna. Aplikasi ini memiliki antarmuka yang dirancang sederhana dan ramah pengguna, dengan berbagai kolom input yang memungkinkan pengguna memasukkan sejumlah parameter medis yang relevan dalam proses prediksi diabetes. Parameter yang diminta meliputi:

- a. Jumlah kehamilan (*Pregnancies*)
- b. Kadar glukosa dalam darah (*Glucose Level*)
- c. Tekanan darah (*Blood Pressure*)
- d. Ketebalan lipatan kulit (*Skin Thickness*)
- e. Jumlah insulin dalam darah (*Insulin Level*)
- f. Indeks Massa Tubuh (BMI)
- g. Fungsi keturunan diabetes (*Diabetes Pedigree Function*)
- h. Usia pasien

Input tersebut harus diisi oleh pengguna agar sistem dapat melakukan perhitungan dengan model SVM yang telah tersedia. Proses prediksi tidak berjalan secara otomatis, melainkan dilakukan setelah pengguna menekan tombol "Test Prediksi Diabetes". Tombol ini berfungsi untuk memberikan kontrol penuh kepada pengguna, sehingga mereka dapat memastikan bahwa semua data telah diinput dengan benar sebelum sistem melakukan analisis.

Ketika tombol ditekan, model SVM akan mulai bekerja dengan menganalisis data yang dimasukkan, kemudian melakukan klasifikasi berdasarkan pola yang telah dipelajari selama proses pelatihan model. Setelah prediksi selesai dilakukan, aplikasi akan menampilkan hasilnya langsung pada antarmuka pengguna. Aplikasi menggunakan indikator visual berupa warna latar belakang yang berbeda untuk membedakan hasil prediksi:

1. Warna hijau digunakan untuk menunjukkan bahwa prediksi sesuai dengan kondisi sebenarnya.
2. Warna merah menandakan bahwa prediksi yang diberikan oleh model tidak sesuai dengan kondisi sebenarnya.

Dengan penggunaan indikator warna ini, pengguna dapat lebih cepat memahami hasil prediksi, tanpa harus membaca teks secara mendetail. Secara keseluruhan, aplikasi ini dirancang dengan kesederhanaan dan efisiensi sebagai prioritas utama. Dengan menggunakan Streamlit, antarmuka tetap responsif dan ringan, sehingga pengguna dapat berinteraksi dengan mudah. Meskipun proses prediksi bergantung pada model SVM yang telah dilatih sebelumnya, waktu eksekusinya relatif cepat dan tidak membebani sistem.

Hasil performa aplikasi diatect yang dicantumkan melibatkan tiga metrik evaluasi utama:

1. Presisi (*Precision*):

Presisi mengukur seberapa banyak dari hasil positif yang diberikan oleh model yang sebenarnya relevan. Dalam konteks aplikasi ini, presisi mencapai 77%. Artinya, dari semua hasil positif yang diidentifikasi oleh aplikasi, 77% di antaranya benar-benar relevan atau akurat ( Rahmasari, 2024 ).

2. *Recall* (Sensitivitas atau *True Positive Rate*):

*Recall* mengukur seberapa banyak dari total kasus positif yang diidentifikasi oleh model. Dalam hasil performa ini, *recall* mencapai 71%. Ini berarti bahwa aplikasi berhasil mengidentifikasi 71% dari keseluruhan kasus positif yang sebenarnya (Wiratama & Aziz, 2024).

3. F1-skor:

F1-skor adalah metrik yang menyatukan presisi dan *recall* untuk memberikan gambaran holistik tentang kinerja model ( Fardeen, 2025 ). F1-skor mencapai 77% pada aplikasi ini. F1-skor yang tinggi menunjukkan keseimbangan yang baik antara presisi dan *recall*, menunjukkan bahwa model memiliki kemampuan baik dalam mengidentifikasi kelas positif dan menghindari kesalahan klasifikasi.

Dengan angka-angka ini, dapat disimpulkan bahwa aplikasi memiliki kinerja yang kuat dalam melakukan klasifikasi atau deteksi, dengan kemampuan tinggi dalam mengenali kasus positif dan memberikan hasil yang akurat. Namun, tetap penting untuk mempertimbangkan konteks spesifik dari aplikasi ini, dan evaluasi lebih lanjut dapat dilakukan untuk memahami kecocokan model dengan kebutuhan dan tujuan penggunaannya.

```

diabetes.py > from sklearn.metrics import classification_report, confusion_matrix
practis_1\diabetes.py:10: print("=====")
print(confusion_matrix(Y_test, Y_test_prediction))

# Menampilkan classification report (precision, recall, f1-score)
print("=====")
print("Laporan klasifikasi:")
print(classification_report(Y_test, Y_test_prediction))
else:
    print("Error: Jumlah sampel tidak konsisten antara Y_test dan Y_test_prediction")

# ... (kode selanjutnya tetap sama)

```

```

Evaluasi Model:
=====
Panjang_Y_test: 154
Panjang_Y_test_prediction: 154

Confusion Matrix:
=====
[[ 91  0]
 [ 26 28]]

Laporan klasifikasi:
=====
              precision    recall  f1-score   support

0               0.78         0.91         0.84         100
1               0.76         0.52         0.62          54

accuracy         0.77         0.77         0.77         154
macro avg        0.77         0.71         0.73         154
weighted avg     0.77         0.77         0.76         154

```

*Gambar 4.7. Perfroma Aplikasi*

#### 4.8. Tampilan Aplikasi

Pada tampilan utama aplikasi, terdapat delapan kolom input yang dirancang untuk menerima berbagai parameter kesehatan yang digunakan dalam proses prediksi diabetes. Setiap kolom ini memungkinkan pengguna untuk memasukkan data pribadi mereka sesuai dengan informasi medis yang relevan.

*Gambar 4.8. Tampilan Dashboard Aplikasi*

# Data Mining Prediksi Diabetes

input nilai pregnancies  
2

input nilai glucose  
3

input nilai blood pressure  
147

input nilai skinthickness  
2

input nilai insulin  
6.34

input nilai bmi  
65

input nilai diabetes pedigree function  
0

input nilai Age  
50

test prediksi diabetes

pasien terkena diabetes

**Gambar 4.8.** Tampilan Dashboard Hasil Prediksi

Berdasarkan hasil pengujian sistem prediksi menggunakan metode data mining, dengan nilai input seperti kehamilan, glukosa, tekanan darah, ketebalan kulit, insulin, indeks massa tubuh, riwayat diabetes keluarga, dan usia, sistem memprediksi bahwa pasien tersebut terkena diabetes.

Untuk delapan kolom input tersebut, yaitu:

1. Kolom Input Jumlah Kehamilan (*Pregnancies*)
  - a. Digunakan untuk memasukkan jumlah kehamilan yang pernah dialami oleh pengguna (khusus untuk perempuan).
  - b. Faktor ini penting karena riwayat kehamilan dapat berpengaruh terhadap risiko diabetes gestasional.
2. Kolom Input Nilai Glukosa (*Glucose Level*)
  - a. Pengguna diminta untuk memasukkan kadar glukosa dalam darah mereka.
  - b. Glukosa yang tinggi dapat menjadi indikator utama adanya risiko diabetes.
3. Kolom Input Nilai Tekanan Darah (*Blood Pressure*)
  - a. Pada kolom ini, pengguna dapat memasukkan nilai tekanan darah mereka dalam satuan mmHg.
  - b. Tekanan darah tinggi dapat menjadi faktor risiko tambahan bagi penderita diabetes.
4. Kolom Input Ketebalan Lipatan Kulit (*Skin Thickness*)
  - a. Digunakan untuk menginput ketebalan lipatan kulit dalam satuan milimeter.
  - b. Parameter ini digunakan sebagai indikator tingkat lemak tubuh yang berhubungan dengan risiko diabetes.
5. Kolom Input Nilai Insulin (*Insulin Level*)
  - a. Pengguna diminta untuk memasukkan jumlah insulin dalam darah yang diukur dalam satuan  $\mu\text{U}/\text{mL}$ .
  - b. Insulin berperan penting dalam metabolisme glukosa, dan kadar insulin yang tidak normal dapat menjadi tanda resistensi insulin atau diabetes tipe 2.

6. Kolom Input Indeks Massa Tubuh (*Body Mass Index / BMI*)
  - a. Digunakan untuk memasukkan nilai BMI, yang merupakan ukuran standar untuk menentukan kategori berat badan seseorang berdasarkan tinggi dan berat badan.
  - b. Nilai BMI yang lebih tinggi dapat meningkatkan risiko diabetes.
7. Kolom Input Diabetes Pedigree Function
  - a. Parameter ini menunjukkan kemungkinan seseorang terkena diabetes berdasarkan riwayat keluarga dan faktor keturunan.
  - b. Makin tinggi nilainya, makin besar kemungkinan seseorang memiliki risiko genetik terhadap diabetes.
8. Kolom Input Usia (*Age*)
  - a. Pengguna diminta untuk memasukkan usia mereka dalam tahun.
  - b. Faktor usia berperan penting karena makin tua seseorang, makin besar risikonya terkena diabetes tipe 2.

Dengan adanya delapan kolom input ini, pengguna dapat memasukkan data kesehatan mereka secara akurat, yang kemudian akan diproses oleh model prediksi berbasis Support Vector Machine (SVM) untuk menentukan apakah mereka berisiko mengalami diabetes atau tidak. Tampilan ini dirancang agar mudah digunakan, dengan setiap kolom input dilengkapi dengan label yang jelas untuk membantu pengguna memasukkan informasi dengan benar.

## **BAB 5**

### **KESIMPULAN DAN SARAN**

#### **5.1. Kesimpulan**

Setelah menjalani berbagai tahap penelitian dan perancangan, aplikasi Diatect diperoleh hasil berikut :

##### **1) Aplikasi Diatect**

Aplikasi ini dirancang sebagai sistem berbasis web yang memungkinkan siapa saja untuk menggunakannya. DIATECT bekerja dengan mengolah data kesehatan pengguna, seperti kadar glukosa darah, tekanan darah, indeks massa tubuh (BMI), serta faktor keturunan, untuk memprediksi kemungkinan seseorang memiliki diabetes atau tidak. Antarmuka aplikasi dibuat sederhana dan user-friendly. Dengan akses berbasis web, aplikasi ini dapat digunakan kapan saja dan di mana saja tanpa perlu menginstal aplikasi tambahan, sehingga memberikan kemudahan bagi pengguna.

##### **2) Kinerja Model SVM dalam Memprediksi Diabetes**

Setelah melalui berbagai tahapan pengujian, aplikasi ini menunjukkan performa yang cukup baik dalam mendeteksi risiko diabetes.

- a. Akurasi model mencapai 77%, yang berarti dari semua prediksi yang dibuat, 77% di antaranya benar dan sesuai dengan kondisi kesehatan pengguna.
- b. Presisi model sebesar 77%, yang menunjukkan bahwa jika aplikasi menyatakan seseorang memiliki diabetes, kemungkinan besar prediksi tersebut benar.
- c. *Recall* model mencapai 71%, yang berarti aplikasi ini dapat mendeteksi hampir semua orang yang benar-benar memiliki diabetes dengan cukup baik.
- d. F1-skor sebesar 77% mengindikasikan bahwa model ini memiliki keseimbangan yang baik antara presisi dan *recall*, sehingga prediksinya cukup dapat diandalkan.

### **3) Manfaat Aplikasi**

Aplikasi ini dapat membantu individu dalam memantau kesehatannya sendiri, sehingga mereka bisa mengetahui apakah mereka memiliki risiko diabetes dan perlu berkonsultasi lebih lanjut dengan dokter. Menjangkau lebih banyak pengguna karena berbasis web, sehingga siapa pun yang memiliki koneksi internet dapat menggunakannya tanpa terbatas oleh lokasi geografis. Pada proses pengembangan dan implementasi aplikasi Diatetect, terdapat beberapa tantangan yang memengaruhi hasil penelitian, yaitu:

#### **a) Kinerja Algoritma yang Masih Bisa Ditingkatkan**

Algoritma Support Vector Machine (SVM) yang digunakan memang memiliki performa cukup baik, tetapi masih bisa ditingkatkan dengan pengoptimalan parameter agar lebih akurat. Akurasi model yang diperoleh hanya 77%, yang berarti masih ada kemungkinan kesalahan prediksi sebesar 23%, sehingga perlu dilakukan evaluasi lebih lanjut. Penggunaan kernel dan parameter regulasi dalam SVM masih memerlukan penyesuaian lebih lanjut untuk mendapatkan hasil prediksi yang lebih baik.

#### **b) Keterbatasan dalam Implementasi Aplikasi Web**

Aplikasi berbasis web yang dikembangkan masih dalam tahap awal dan hanya menyediakan fitur prediksi diabetes, tanpa adanya fitur tambahan seperti riwayat pemeriksaan pengguna, rekomendasi pola hidup sehat, atau konsultasi dengan dokter. Responsivitas dan kecepatan sistem masih perlu ditingkatkan, terutama ketika memproses input data dalam jumlah besar.

#### **c) Potensi Bias dalam Data dan Model**

Kondisi di mana model terlalu menyesuaikan diri terhadap data latih dan gagal menggeneralisasi dengan baik terhadap data baru lebih mungkin terjadi dengan model Support Vector Machine (SVM) yang digunakan. Risiko ini meningkat jika dataset yang digunakan tidak cukup beragam atau tidak mewakili karakteristik populasi secara keseluruhan.

## 5.2. Saran

Berdasarkan kendala yang telah diidentifikasi dalam penelitian ini, terdapat beberapa saran yang dapat membantu meningkatkan kualitas dan efektivitas aplikasi DIATECT ke depannya:

### 1. Meningkatkan Kualitas dan Variasi Dataset

- a. Menggunakan dataset yang lebih besar dan lebih beragam, termasuk data dari berbagai kelompok usia, latar belakang etnis, serta gaya hidup untuk meningkatkan akurasi prediksi.
- b. Mengumpulkan data primer dengan melakukan survei atau kerja sama dengan rumah sakit dan klinik agar data yang digunakan lebih relevan dengan kondisi di Indonesia.
- c. Mempertimbangkan penambahan variabel lain, seperti pola makan, tingkat aktivitas fisik, riwayat keluarga, dan kondisi kesehatan lainnya yang dapat memengaruhi risiko diabetes.

### 2. Meningkatkan Akurasi Model Prediksi

- a. Mengumpulkan data primer dengan melakukan survei atau kerja sama dengan rumah sakit dan klinik agar data yang digunakan lebih relevan dengan kondisi di Indonesia.
- b. Mempertimbangkan penambahan variabel lain, seperti pola makan, tingkat aktivitas fisik, riwayat keluarga, dan kondisi kesehatan lainnya yang dapat memengaruhi risiko diabetes.

### 3. Pengembangan dan Penyempurnaan Aplikasi Web

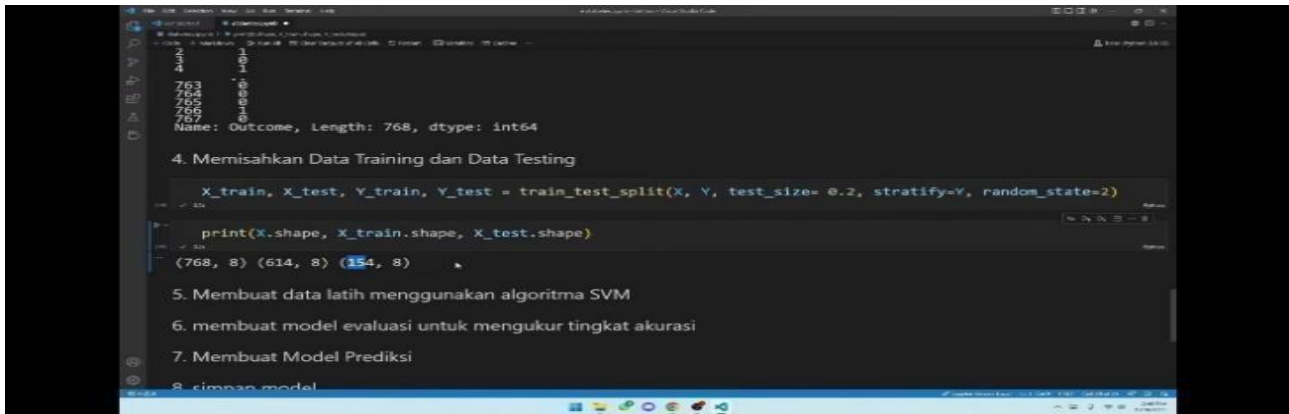
- a. Mengimplementasikan dan penyempurnaan rancangan menjadi sebuah web atau suatu aplikasi berbentuk android maupun ios yang dapat dijangkau masyarakat luas.
- b. Mengintegrasikan fitur rekomendasi pola hidup sehat, seperti panduan pola makan dan olahraga.
- c. Menambahkan fitur tambahan seperti riwayat pemeriksaan pengguna, sehingga pengguna dapat melihat perubahan risiko mereka dari waktu ke waktu.

## DAFTAR PUSTAKA

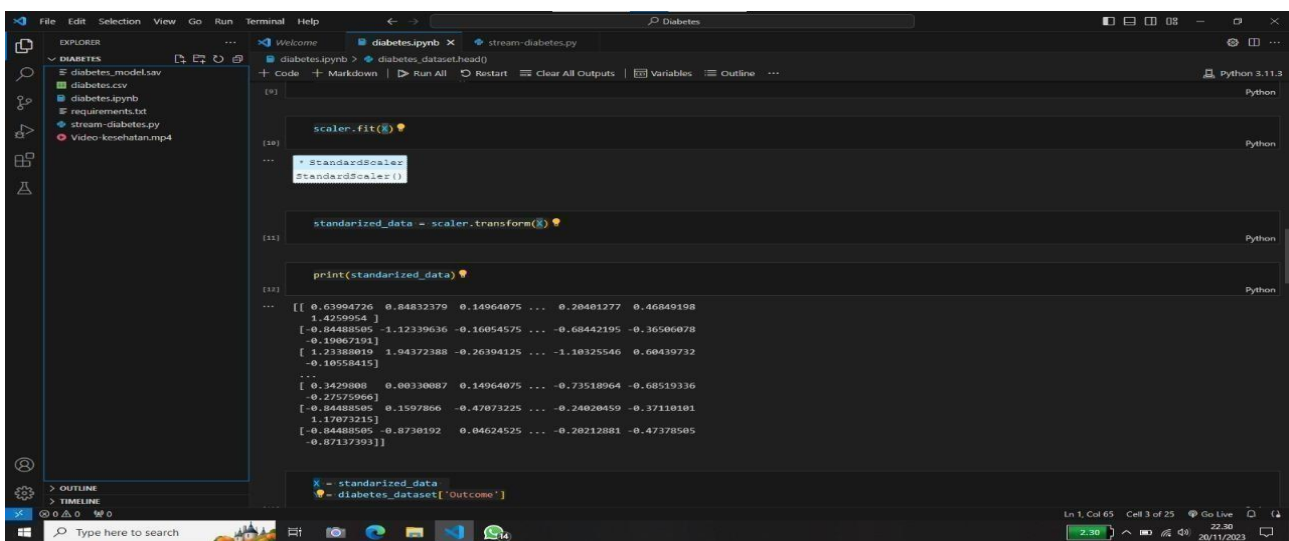
- A. Abimayu. ( 2023 ). Klasifikasi Penderita Penyakit Diabetes Berbasis Data Rekam Medis Klasifikasi Penderita Penyakit Diabetes Berbasis Data Rekam Medis. (Skripsi). Universitas Islam Negeri Maulana Malik Ibrahim, Malang.
- King H, Aubert RE, Herman WH. ( 1998 ). “ Global Burden of Diabetes, 1995–2025: Prevalence, numerical estimates, and projections “ Diakses pada 31 Juli 2025, dari <https://diabetesjournals.org/care/article/21/9/1414/22253/Global-Burden-of-Diabetes-1995-2025-Prevalence>.
- F. Amira Mumtaz, R. Maulana, and A. S. Budi. ( 2020 ). Sistem Pendeteksi Penyakit Diabetes Melitus berdasarkan Kondisi Urin dan Gas Buang Pernapasan menggunakan K-Nearest Neighbor berbasis Arduino. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 4 ( 8 ), 2628-2636.
- Fardeen, R. Eka Putra. ( 2025 ). Perbandingan Analisis Sentimen Untuk Prediksi Kepuasan Pelanggan Kedai Kopi Di Kofind Menggunakan Algoritma SVM Dan Naive Bayes. JINACS, 6 ( 4 ), 2220 – 2686.
- F. R. Lumbanraja. ( 2022 ). Implementasi Support Vector Machine (SVM) Untuk Klasifikasi Penderita Diabetes Mellitus. Jurnal Komputasi, 10 (1), 75-83.
- F. Rosyking Lumbanraja. ( 2023 ). Penerapan Metode Support Vector Machine (SVM) dalam Klasifikasi Penderita Diabetes Mellitus Jurnal Pepadun. 4 ( 1 ), 76-83.
- H. Amir. ( 2024 ). Klasifikasi Kematangan Buah Alpukat Mentega Berdasarkan Fitur Warna Menggunakan Support Vector Machine (SVM). IFTECH, 6 ( 1 ), 108-120.
- H. Ichwanul Muslim Karo Karo. ( 2022 ). Klasifikasi Penderita Diabetes Menggunakan Algoritma Machine Learning dan Z-Score” Jurnal Teknologi Terpadu, 8 ( 2 ), 94-99.
- J. Wira and G. Putra, ( 2020 ). Pengenalan Konsep Pembelajaran Mesin dan Deep Learning Edisi 1.4. Tokyo, Jepang : Researchgate.
- K. Khoirudin Mohammad Burhan Hanif. ( 2020 ). Sistem Aplikasi Prediksi Penyakit Diabetes Menggunakan Fitures Selection Korelasi Pearson Dan Klasifikasi Naive Bayes. Jurnal Pengembangan Rekayasa dan Teknologi, 4 ( 2 ),199-205.

- K. Nugrogo. (2023). Perilaku Makan Pada Individu Obesitas Dengan DM Tipe 2 Selama Pandemi Covid-19 Di Wilayah Kerja Puskesmas Sidorejo Lor, Kota Slatiga. *Care Jurnal Ilmiah Kesehatan*, 11 ( 2 ), 326 – 339.
- M. Ardiansyah, A. Sunyoto, and E. T. Luthfi. (2021) “Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4.5 untuk Klasifikasi Diabetes,” *Edumatic: Jurnal Pendidikan Informatika*, 5 ( 2 ), 147 – 156.
- M. D. Purbolaksono. ( 2021 ). Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5 ( 2 ) , 393 - 399.
- M. I Gunawan, D. Sugiarto, Mardianto. ( 2020 ). Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Search pada Algoritma Logistic Regression. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 6 ( 3 ), 280-284.
- Muttaqin, Wahyu Wijaya Widiyanto. ( 2023). *Pengenalan Data Mining. Indonesia : Yayasan Kita Menulis.*
- N. Pratiwi and Y. Setyawan. (2021). Analisis Akurasi Dari Perbedaan Fungsi Kernel Dan Cost Pada Support Vector Machine Studi Kasus Klasifikasi Curah Hujan Di Jakarta. *Journal of Fundamental Mathematics and Applications (JFMA)*, 4 ( 2 ), 203–212.
- Rahmasari, F., Rifany, M., Priharyanto, T., & Kurniawan, R. ( 2024 ). Perbandingan Algoritma Machine Learning dalam Klasifikasi Status Banjir di Sumatera Utara. *Seminar Nasional Sains Data. Senada*, 4 (1), 307-318.
- Wiratama, Y., & Aziz, R. A. (2024). Perbandingan Prediksi Penyakit Stunting Balita Menggunakan Algoritma Support Vektor Machine dan Random Forest. *Building of Informatics, Technology and Science (BITS)*, 6 (2), 1159–1168.

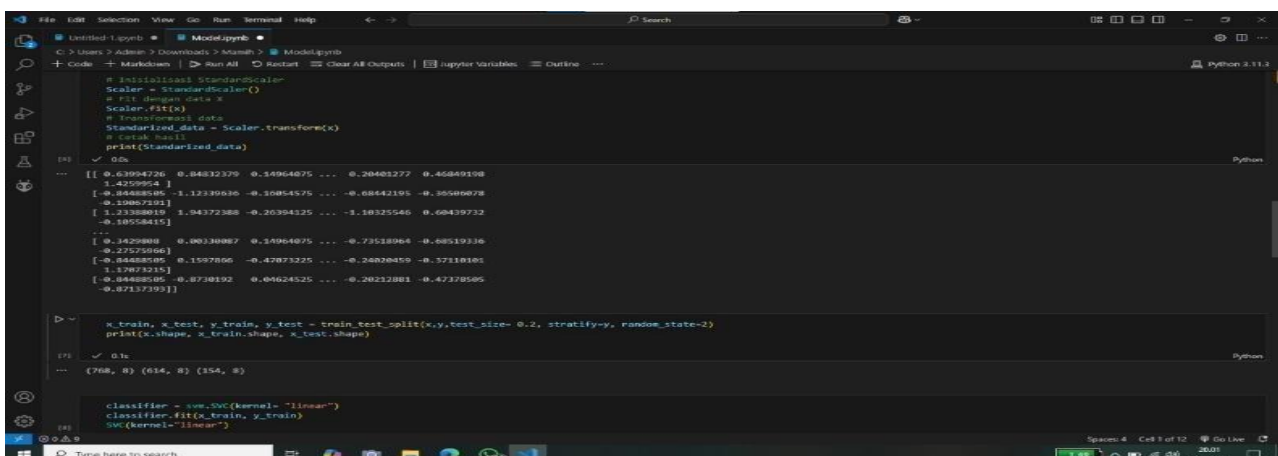
## LAMPIRAN



*Proses Data Training Dan Testing*



*Proses Standardisasi Data*



*Proses Standardisasi Data*

```

import numpy as np
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

diabetes_dataset = pd.read_csv('diabetes.csv')

diabetes_dataset.head()

diabetes_dataset.shape

diabetes_dataset['Outcome'].value_counts()

```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreefunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

*Proses Load Data Set*

```

X_test_prediction = classifier.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)

print('Akurasi data testing adalah - ', test_data_accuracy)

Akurasi data testing adalah = 0.7727272727272727

input_data = (6, 148, 72, 35, 0, 33.6, 0.627, 50)
input_data_as_numpy_array = np.array(input_data)

input_data_reshape = input_data_as_numpy_array.reshape(1,-1)
std_data = scaler.transform(input_data_reshape)
print(std_data)

prediction = classifier.predict(std_data)
print(prediction)

if (prediction[0] == 0):
    print('pasien tidak terkena diabetes')
else:
    print('pasien terkena diabetes')

[[ 0.63994726  0.84832379  0.14964075  0.90726993 -0.69289057  0.20401277
  0.46849198  1.4259954 ]]
[1]
pasien terkena diabetes
Warning: X does not have valid feature names, but StandardScaler

```

*Proses Akurasi Data*

```

X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=2)

print(X.shape, X_train.shape, X_test.shape)

(768, 8) (614, 8) (154, 8)

from sklearn import svm
classifier = svm.SVC(kernel='linear')

classifier.fit(X_train, Y_train)

classifier.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)

print('Akurasi data training adalah - ', training_data_accuracy)

Akurasi data training adalah = 0.786489511480052

```

*Proses Pelabelan Data*

The screenshot shows a Python IDE with a file explorer on the left containing files like 'diabetes\_diagnosis.ipynb', 'diabetes\_model.sav', 'diabetes.csv', 'diabetes.ipynb', 'requirements.txt', 'stream-diabetes.py', and 'Video-kesehatan.mp4'. The main editor displays Python code that imports 'classification\_report' and 'confusion\_matrix' from 'sklearn.metrics'. It prints the confusion matrix and a classification report for 'Y\_test' and 'Y\_test\_prediction'. The code includes a check for consistent sample sizes and an error message if they are not.

```
from sklearn.metrics import classification_report, confusion_matrix
print("\nConfusion Matrix:")
print("-----")
print(confusion_matrix(Y_test, Y_test_prediction))

# Menampilkan classification report (precision, recall, f1-score)
print("\nLaporan Klasifikasi:")
print("-----")
print(classification_report(Y_test, Y_test_prediction))
else:
    print("Error: Jumlah sampel tidak konsisten antara Y_test dan Y_test_prediction")
# ... (Kode setelahnya tetap sama)
```

The output of the code is as follows:

```
Evaluasi Model:
-----
Panjang Y_test: 154
Panjang Y_test_prediction: 154
Confusion Matrix:
-----
[[91  9]
 [26 28]]

Laporan Klasifikasi:
-----
              precision    recall  f1-score   support

     0       0.78      0.91      0.84       100
     1       0.76      0.52      0.62         54

 accuracy          0.77
 macro avg          0.77      0.71      0.73       154
 weighted avg          0.77      0.77      0.76       154
```

*Proses Modeling*

## SOURCECODE SISTEM

```
Import numpy as np Import pandas as pd
From sklearn.preprocessing import StandardScaler From sklearn.model_selection import
train_test_split From sklearn import svm
From sklearn.metrics import accuracy_score Diabetes_dataset = pd.read_csv( '
diabetes.csv ' ) Diabetes_dataset.head() Diabetes_dataset.shape (768, 9)
Diabetes_dataset[ ' outcome ' ] - value_counts()
    0    500
    1    268

Name : Outcome, dtype : int04 # memisahkan data dan label
X = diabetes_dataset.drop (columns= ' outcome ' , axis = 1 )
Y = diabetes_dataset [ ' outcome ' |
Print(x) Print(Y)
Name : outcome, Length : 768, dtype: int64 Scaler = StandardScaler()
Scaler.fit(x) StandardScaler()
Standardized_data = scaler . transform(X) Print( standardized_data) X_train, x_test,
Y_train_test = train_test_split(X,Y,test_size= 0.2, stratify=Y, random_state=2)
Print(X,shape, X_train.shape, x_test.shape (768,8) (614,8) (154, 8)
Classifier = svm.SVC(kernel= ' linear' ) Classifier.fit(X_train, Y_train)
SVC(kernel='linear') X_train_prediction(X_train)
Training_data_accuracy = accuracy_score(X_train_prediction, Y_train) Print('Akurasi data
training adalah = ' , training_data_accuracy) Akurasi data training adalah = 0,7866
X_test_prediction = classifier.prediction(X_test) Test_data_accuracy =
accuracy_score(X_test_prediction,Y_test) Print('Akurasi data testing adalah = ' ,
test_data_accuracy) Akurasi data testing adalah = 0.7727
```

```

Input_data = (6, 168,72,35,0,33.6,0.627,50)
Input_data_as_numpy_array = np.array(input_data) Input_data_reshape =
input_data_as_numpy_array.reshape(1,-1) Std_data= scaler.transform(input_data_reshape)
Print(std_data)
Prediction = classifier.predetection(std_data) Print(prediction)
If (prediction[0] == 0)
Print(' pasien tidak terkena diabetes ')
Else :
Print(' pasien terkena diabetes ') Import pickle
File name = ' diabetes_model.sav '
Pickle dump(classifier, open(filename, 'wb'))
import pickle import streamlit as st # membaca model
diabetes_model = pickle.load(open('diabetes_model.sav', 'rb'))
#judulweb
st.title('Data Mining Prediksi Diabetes')
pregnancies = st.text_input ('input nilai pregnancies') glucose = st.text_input ('input nilai
glucose') bloodpressure = st.text_input ('input nilai blood pressure') skinthickness =
st.text_input ('input nilai skinthickness')
insulin = st.text_input ('input nilai insulin') bmi = st.text_input ('input nilai bmi')
Diabetespedigreefunction = st.text_input ('input nilai diabetes pedigree function' ) Age =
st.text_input('input nilai Age')
# code untuk prediksi diab_diagnosis = "
# membuat tombol prediksi
if st.button('test prediksi diabetes'): diab_prediction = diabetes_model.predict([[pregnancies,
glucose, bloodpressure, skinthickness, insulin, bmi, Diabetespedigreefunction, Age]])
if(diab_prediction[0] == 1): diab_diagnosis = 'pasien terkena diabetes'else : diab_diagnosis = '
pasien tidak terkena diabetes st.success(diab_diagnosis)

```