

Perancangan Aplikasi Prediksi Resiko Diabetes Menggunakan Algoritma Support Vector Machine (SVM) Berbasis Streamlit

Dimas Aditya Fajar Nugroho^{1*}, Sri Hadiani²

^{1,2}Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri, Jakarta, Indonesia

MEDINFTECH is licensed under a Creative Commons 4.0 International License.



ARTICLE HISTORY

Received: 00 February 00
 Final Revision: 00 March 00
 Accepted: 00 April 00
 Online Publication: 00 March 00

KEYWORDS

Diabetes, Data Mining, Support Vector Machine (SVM), CRISP-DM, Streamlit, Prediksi Risiko

CORRESPONDING AUTHOR

aditdimas130@gmail.com
 DOI
 10.37034/medinftech.v1i1.1

ABSTRACT

Diabetes mellitus merupakan salah satu penyakit kronis yang banyak diderita masyarakat Indonesia dan membutuhkan deteksi dini untuk mencegah komplikasi. Penelitian ini bertujuan untuk membangun sebuah aplikasi berbasis web yang mampu memprediksi risiko diabetes secara otomatis berdasarkan data kesehatan pasien. Aplikasi ini dibangun menggunakan algoritma klasifikasi Support Vector Machine (SVM) yang diimplementasikan melalui framework Streamlit. Proses pengembangan sistem mengikuti tahapan metode CRISP-DM (Cross-Industry Standard Process for Data Mining) yang mencakup pemahaman masalah, pengolahan data, pembuatan model, evaluasi, dan implementasi. Dataset yang digunakan adalah Pima Indians Diabetes Dataset, yang terdiri dari 768 data pasien dengan 8 atribut kesehatan. Model dilatih dengan SVM dan dievaluasi menggunakan metrik confusion matrix dengan hasil akurasi sebesar 80.86%, precision 77.38%, recall 63.81%, dan f1-score 69.94%. Hasil dari penelitian ini adalah sebuah aplikasi interaktif yang dapat digunakan oleh masyarakat umum untuk mendeteksi risiko diabetes, menampilkan visualisasi hasil prediksi, serta menyediakan informasi edukatif dan evaluasi model. Diharapkan aplikasi ini dapat menjadi alat bantu dalam meningkatkan kesadaran akan pentingnya deteksi dini diabetes secara non-klinis.

1. Introduction

Diabetes melitus (DM) adalah penyakit atau gangguan metabolisme tubuh yang kronis dan multietiologi yang ditandai dengan peningkatan kadar gula dalam darah disertai dengan gangguan metabolisme lipid dan protein akibat dari tidak mampunya insulin dalam melakukan fungsinya. Menurut World Health Organization (WHO) angka kejadian DM di dunia meningkat dari 108 juta pada 1980 menjadi 422 juta pada 2014 (World Health Organization, 2018). Pada Riset Kesehatan Dasar tahun 2018 di Indonesia memperlihatkan peningkatan 6,9% tahun 2013 menjadi 10,9 % tahun 2018[1].

Pada tahun 2021, diperkirakan 537 juta orang menderita diabetes, dan jumlah ini diprediksi mencapai 643 juta pada tahun 2030, dan 783 juta pada tahun 2045. Selain itu, 541 juta orang diperkirakan mengalami gangguan toleransi glukosa pada tahun 2021. Diperkirakan lebih dari 6,7 juta orang berusia 20-79 tahun akan meninggal karena penyebab terkait diabetes pada tahun 2021. Jumlah anak-anak dan remaja (yaitu hingga usia 19 tahun) yang hidup dengan

diabetes meningkat setiap tahunnya. Pada tahun 2021, lebih dari 1,2 juta anak dan remaja menderita diabetes tipe 1[2].

Salah satu jenis penyakit tidak menular yang harus diwaspadai adalah diabetes mellitus. Diabetes mellitus merupakan silent killer yang tetap asimtomatik pada tahap awal dan tidak terdeteksi sampai terjadi kejadian vaskular seperti angina, infark miokard atau stroke. [3].

Diabetes melitus atau yang biasa masyarakat pada umumnya menyebutnya dengan penyakit kencing manis merupakan penyakit menahun yang dapat diderita seumur hidup. Diabetes memiliki 2 tipe yakni diabetes melitus tipe 1 yang merupakan hasil dari reaksi autoimun terhadap protein sel pulau pankreas, kemudian diabetes tipe 2 yang mana disebabkan oleh kombinasi faktor genetik yang berhubungan dengan gangguan sekresi insulin, resistensi insulin dan faktor lingkungan seperti obesitas, makan berlebihan, kurang makan, olahraga dan stres, serta penuaan[4].

Diabetes melitus (DM) adalah penyakit metabolik yang ditandai dengan hiperglikemia karena kurangnya sekresi insulin, kerja insulin, atau keduanya.

International Diabetes Federation (IDF) melaporkan bahwa 371 juta orang antara usia 20 dan 79 menderita DM di seluruh dunia. Risiko menderita akan meningkat seiring bertambahnya usia dan usia >40 tahun lebih rentan terkena intoleransi glukosa. Sekitar 80% penderita DM tipe 2 terbukti mengalami obesitas dan risiko diabetes meningkat secara progresif yang ditunjukkan oleh indeks massa tubuh[5].

Diabetes melitus adalah keadaan hiperglikemia kronik disertai berbagai kelainan metabolik akibat gangguan hormonal yang menimbulkan berbagai komplikasi kronik pada mata, ginjal, saraf, dan pembuluh. Perubahan vaskular di ekstremitas bawah dapat mengakibatkan terjadinya arteriosklerosis sehingga terjadi komplikasi yang mengenai kaki[6].

Dalam rangka meminimalisir komplikasi yang ditimbulkan dari penyakit diabetes, membuat campuran tangan medis menjadi kebutuhan yang mendesak baik secara nasional maupun internasional. Keputusan medis harus dibuat berdasarkan pengetahuan dan logika klinis. Metode prediktif yang efektif melibatkan analisa data data klinis pasien diabetes yang dikumpulkan dalam dataset, dengan menerapkan teknik klasifikasi Machine Learning untuk mengidentifikasi pola. Penggunaan algoritma Machine Learning seperti Support Vector Machine (SVM) telah memberikan wawasan baru dan kontribusi pada pengembangan perencanaan di masa depan.

2. Research Method

Pada penelitian ini penulis menggunakan metodologi CRISP-DM (Cross Industry Standard Process for Data Mining) sebagai pemecah masalah yang umum untuk bisnis dan penelitian. Metodologi ini terdiri dari enam tahapan sebagai berikut yaitu Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment. Proses 6 tahapan dalam metodologi ini dapat dijelaskan sebagai berikut.



1. Business Understanding (Pemahaman Bisnis)

Pada tahapan awal ini bertujuan untuk memahami secara menyeluruh bahwa tujuan bisnis dari penelitian ini yaitu untuk merancang sebuah aplikasi berbasis web dengan kegunaan untuk prediksi resiko terkena penyakit diabetes yang dapat membantu dalam mendeteksi dini kemungkinan terkena diabetes berdasarkan data

kesehatan, permasalahan yang diangkat adalah bagaimana memanfaatkan metode data mining, khususnya algoritma klasifikasi SVM (Support Vector Machine), untuk menghasilkan system yang lebih akurat dan gampang digunakan.

2. Data understanding (Pemahaman Data)

Pada tahap ini, memahami dataset yang digunakan, yaitu catatan data Pima-India-Diabetes. Catatan data ini berisi 768 data pasien dengan delapan atribut utama: kehamilan, glukosa, tekanan darah, ketebalan kulit, insulin, BMI, dan fungsi pohon batang diabetes. Level ini mencakup analisis struktur data, distribusi nilai, dan identifikasi nilai yang tidak wajar.

3. Data Preparation (Persiapan Data)

Tahapan ini mencakup proses pembersihan dan transformasi data agar siap digunakan untuk pelatihan model. Data disiapkan menjadi data latih dan data uji. Tujuannya adalah untuk menyatukan data agar dapat digunakan dengan sesuai dalam pelatihan model.

4. Modelling

Tahapan ini merupakan proses pelatihan model klasifikasi menggunakan algoritma Support Vector Machine (SVM). Algoritma ini dipilih karena mampu memisahkan dua kelas (diabetes dan tidak diabetes) secara optimal dengan menggunakan hyperplane. Model dilatih menggunakan data latih dan disimpan dalam bentuk file.sav agar dapat digunakan kembali pada tahap implementasi.

5. Evaluation

Setelah model dilatih, dilakukan evaluasi menggunakan data uji untuk mengukur performa model. Evaluasi dilakukan dengan menghitung confusion matrix dan metrik evaluasi seperti akurasi, precision, recall, dan f1-score. Evaluasi ini penting untuk memastikan model bekerja dengan baik dan dapat diandalkan.

6. Deployment

Tahap akhir dari metodologi ini adalah implementasi model ke dalam aplikasi berbasis web menggunakan framework Streamlit. Aplikasi memungkinkan pengguna untuk memasukkan data kesehatan, mendapatkan hasil prediksi risiko diabetes secara langsung, melihat visualisasi berupa speedometer risiko, serta mengakses halaman edukasi dan evaluasi model.

3. Result and Discussion

Desain

Desain Pemodelan Sistem

Tabel IV.2 Format Penyimpanan Data dan Model

No	Komponen	Format/File
1	Dataset Latih	diabetes.csv
2	Model SVM	diabetes model with scaling.sav
3	Scaler	scaler.sav

Sumber: Penelitian

Pada tabel IV.2 menunjukkan komponen utama yang disimpan dan format berkas yang digunakan untuk proyek pengembangan model prediksi diabetes. Pertama, set data pelatihan disimpan dalam berkas diabetes.csv. Berkas ini berisi data mentah seperti kadar glukosa, tekanan darah, BMI, dan label yang menunjukkan status diabetes. Data ini digunakan untuk melatih model. Selanjutnya, model SVM yang telah dilatih disimpan dalam berkas diabetes_model_with_scaling.sav. Hal ini memungkinkan model Support Vector Machine (SVM) untuk digunakan kembali tanpa perlu dilatih ulang. Ketiga, Scaler disimpan dalam file scaler.sav, yang berisi objek normalisasi atau standarisasi data seperti StandardScaler atau MinMaxScaler dari scikit-learn. Ini penting untuk memastikan konsistensi preprocessing data saat model digunakan dalam prediksi. Penyimpanan ketiga komponen ini secara sistematis memudahkan penerapan, reproduksi, dan pemeliharaan sistem prediksi diabetes.

Desain Pemodelan Data

Tabel IV. 3 Desain Pemodelan Data

N o	Nama Atribut	Tipe Data	Range / Format	Keterangan
1	Pregnancies	Integer	0 – 20	Jumlah kehamilan sebelumnya
2	Glucose	Integer	0 – 200+	Kadar glukosa darah (mg/dL)
3	BloodPressure	Integer	0 – 140	Tekanan darah diastolik (mmHg)
4	SkinThickness	Integer	0 – 99	Ketebalan lipatan kulit triceps (mm)
5	Insulin	Integer	0 – 850	Kadar insulin dalam darah (mu U/ml)
6	BMI	Float	0.0 – 70.0	Indeks massa tubuh (kg/m ²)
7	DiabetesPedigreeFunction	Float	0.000 – 2.500	Skor riwayat keluarga (genetik)
8	Age	Integer	18 – 100	Usia pasien dalam tahun
9	Outcome	Integer (Label)	0 / 1	Output prediksi: 0 = Tidak diabetes, 1 = Diabetes

Sumber: Penelitian 2025

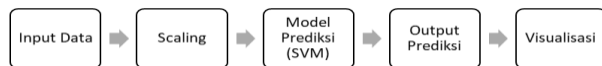
Pada tabel IV.3 memastikan model *Support Vector Machine* (SVM) yang digunakan dalam sistem dapat

memproses data secara optimal, setiap atribut dalam set data ini telah melalui proses standardisasi. Proses standardisasi ini krusial untuk menghilangkan perbedaan skala antar atribut, sehingga model dapat berkinerja lebih baik dan lebih akurat. Dengan mempertimbangkan relevansi klinis dan dampaknya terhadap kinerja model dalam deteksi risiko diabetes dini, fitur-fitur ini dipilih untuk meningkatkan akurasi prediksi dan memberikan wawasan yang lebih mendalam tentang faktor-faktor kesehatan penting.

Atribut pertama adalah kehamilan, yaitu jumlah kehamilan yang pernah dialami pasien, terutama bagi wanita. Glucose attribute menunjukkan kadar glukosa dalam darah, yang juga bertipe integer, dengan rentang nilai dari 0 hingga 200 mg/dL; atribut ini merupakan salah satu indikator utama untuk diagnosis diabetes. Tekanan darah diastolik pasien ditunjukkan dalam rentang 0–140 mmHg dalam atribut ketiga, darah tekanan. Ketebalan kulit triceps, ukuran yang mengukur ketebalan kulit, diikuti oleh SkinThickness, yang bertipe integer dan memiliki nilai antara 0 dan 99 mm. Atribut Insulin menunjukkan kadar insulin dalam darah, yang sangat terkait dengan resistensi insulin pada penderita diabetes. Nilai mulai dari 0 hingga 850 mu U/ml. Selain itu, indeks massa tubuh, juga dikenal sebagai BMI, adalah ukuran yang menunjukkan tingkat obesitas seseorang, diukur dalam kilogram per meter persegi dan memiliki nilai antara 0,0 dan 70 persen, dan bertipe data float. Atribut ketujuh adalah Fungsi Diabetes Pedigree, yang merupakan skor yang menunjukkan faktor genetik atau keturunan diabetes pada pasien. Nilai berkisar antara 0.000 dan 2.500. Usia pasien dalam tahun, yang merupakan integer dari 18 hingga 100 tahun, adalah atribut ke-delapan. Terakhir, hasil adalah label atau tujuan output sistem prediksi. Pasien yang tidak memiliki diabetes memiliki nilai 0 untuk atribut ini, sedangkan pasien yang memiliki diabetes memiliki nilai 1.

Software Architecture

Untuk memastikan pemisahan tanggung jawab yang jelas antara komponen-komponen utama aplikasi, pendekatan modular diadopsi. Implementasi antarmuka pengguna pada antarmuka pengguna (UI) dibangun menggunakan kerangka kerja Streamlit, yang menawarkan fitur-fitur interaktif berbasis Python yang memungkinkan visualisasi data dinamis dan kemudahan interaksi dengan sistem. Model *Support Vector Machine* (SVM) yang didukung oleh mekanisme pemrosesan data yang efisien merupakan langkah pertama dalam proses prediktif antarmuka pengguna (backend). Langkah prapemrosesan data ditangani oleh alur data yang dirancang khusus, yang memantau masukan pengguna dan dinormalisasi dengan scaler. Karena masing-masing komponen dapat ditingkatkan atau diganti secara independen, arsitektur modular ini memfasilitasi pengembangan lebih lanjut dan meningkatkan kemudahan pemeliharaan kode. Sistem dapat digambarkan sebagai berikut:



Gambar IV.1 Diagram, Arsitektur Sistem

User Interface

a. Menentukan Library yang Akan Digunakan

Langkah pertama dalam mengembangkan sistem prediksi risiko diabetes adalah mengidentifikasi dan mengimpor pustaka yang diperlukan ke dalam Jupyter Notebook. Pustaka Numpy membantu komputasi ilmiah seperti statistik, manipulasi array multidimensi, dan operasi numerik yang diperlukan untuk pemrosesan data.

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
import pickle

# Load data
data = pd.read_csv("diabetes.csv")
x = data.drop("Outcome", axis=1)
y = data["Outcome"]

# Scaling
  
```

Gambar IV. 2 Diagram, Arsitektur Sistem

b. Pemuatan Dataset dan Pemahaman Data

Dataset yang digunakan dalam penelitian ini adalah berkas diabetes.csv, yang berisi data kesehatan pasien dengan beberapa parameter seperti jumlah kehamilan, kadar glukosa, tekanan darah diastolik, ketebalan lipatan kulit, kadar insulin, indeks massa tubuh (IMT), dan skor riwayat. Data ini diintegrasikan ke dalam pengembangan model prediksi diabetes. Data tersebut dimuat sebagai kerangka data ke dalam lingkungan pemrograman Python menggunakan pustaka pandas untuk memudahkan analisis, transformasi, dan pelatihan model. Tampilan awal dataset, yang menampilkan beberapa baris data pertama dan struktur kolomnya, ditampilkan pada Gambar IV. 3. Sebelum membuat model, peneliti dapat melakukan pemrosesan tambahan, seperti pembersihan data, normalisasi, dan pemilihan fitur, dengan memahami format dan isi dataset secara menyeluruh.

```

tes.py  diabetes_scaled_model.ipynb  diabetes.csv
diabetes.csv > data
1 Pregnancies,Glucose,BloodPressure,SkinThickness,Insulin,Outcome
2 6,148,72,35,0,33.6,0.627,50,1
3 1,85,66,29,0,26.6,0.351,31,0
4 8,183,64,0,0,23.3,0.672,32,1
5 1,89,66,23,94,28.1,0.167,21,0
6 0,137,40,35,168,43.1,2.288,33,1
7 5,116,74,0,0,25.6,0.201,30,0
8 3,78,50,32,88,31,0.248,26,1
9 10,115,0,0,0,35.3,0.134,29,0
10 2,197,70,45,543,30.5,0.158,53,1
11 8,125,96,0,0,0,0.232,54,1
12 4,110,92,0,0,37.6,0.191,30,0
13 10,168,74,0,0,38,0.537,34,1
  
```

Gambar IV. 3 Pemuatan Dataset dan Menampilkan Dataset

Dalam gambar IV. 3, penggunaan kode-kode yang ada berfungsi untuk membaca dataset yang terdapat di dalam file lokal computer dan yang lainnya yang dimana berfungsi untuk menampilkan dataset pasien.

c. Halaman Input Prediksi Diabetes

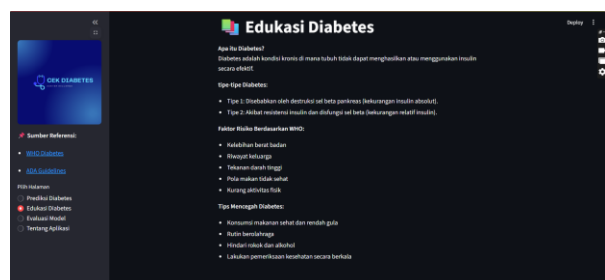
Halaman ini merupakan inti dari aplikasi, dimana pengguna diminta untuk mengisi sejumlah data kesehatan pribadi. Input yang diminta meliputi:

Gambar IV. 4 Halaman Input Prediksi Diabetes

Pengguna dapat memulai analisis data dengan mengeklik tombol "Prediksi Diabetes" setelah memasukkan semua data. Setelah memproses data yang diberikan, sistem akan menampilkan hasil prediksi dalam bentuk teks deskriptif beserta tingkat risiko diabetes, yang dapat diklasifikasikan sebagai tinggi, sedang, atau rendah. Hasil tersebut juga akan disertai visualisasi berupa grafik pengukur untuk menunjukkan tingkat risiko secara grafis. Ringkasan data masukan akan disajikan dalam bentuk tabel, sehingga informasi lebih mudah dipahami dan diinterpretasikan.

d. Halaman Edukasi

Pada halaman ini berfungsi sebagai sarana informasi bagi pengguna untuk mengenal lebih dalam mengenai penyakit diabetes. Dimana informasi yang ditampilkan meliputi:



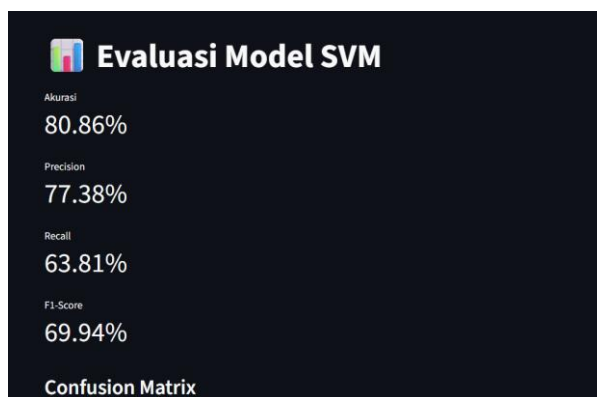
Gambar IV. 5 Halaman Input Prediksi Diabetes

Gambar IV.5 menunjukkan tampilan halaman Input Prediksi Diabetes dalam sistem atau aplikasi yang berkaitan dengan edukasi diabetes. Pengguna dapat menemukan informasi edukasi tentang apa itu diabetes, jenis-jenisnya (Tipe 1 dan Tipe 2), dan faktor risiko

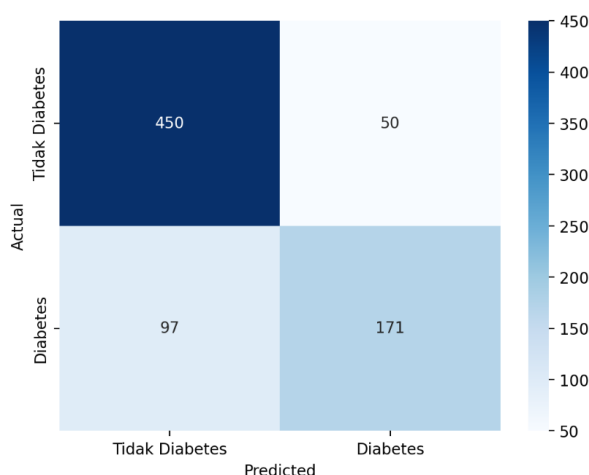
yang dapat menyebabkannya. Selain itu, dijelaskan pula sejumlah metrik medis termasuk ketebalan lipatan kulit, tekanan darah, glukosa darah, dan lainnya yang digunakan dalam proses prediksi. Halaman ini juga memberikan tips tentang cara mencegah diabetes, termasuk menjalani gaya hidup sehat yang mencakup olahraga teratur, mengonsumsi makanan seimbang, dan menjaga berat badan ideal. Dengan mempertimbangkan semua hal tersebut, halaman ini berfungsi sebagai sumber belajar sekaligus tempat bagi pengguna untuk memasukkan informasi guna memperkirakan risiko mereka terkena diabetes menggunakan sistem yang ditawarkan.

e. Halaman Evaluasi Model

Halaman ini menunjukkan kinerja model prediktif sistem yang digunakan; Support Vector Machine (SVM) yang digunakan dievaluasi dengan data uji. Metrik evaluasi yang ditampilkan meliputi akurasi, precision, recal, dan F-1 score.



Gambar IV. 6 Halaman Evaluasi Model



Gambar IV. 7 Halaman Evaluasi Model

Gambar IV.6 dan IV.7 menunjukkan hasil evaluasi model Support Vector Machine (SVM) yang digunakan untuk memprediksi diabetes. Nilai kinerjanya, termasuk akurasi 80,86%, presisi 77,38%, recall 63,81%, dan skor F1 69,94%, menunjukkan bahwa model tersebut

cukup baik dalam mengklasifikasikan data. Namun, masih terdapat ruang untuk perbaikan, terutama dalam hal recall, yang menunjukkan adanya ruang untuk perbaikan.

Di sisi lain, matriks kebingungan hasil prediksi model ditunjukkan pada Gambar IV.7. Matriks tersebut menunjukkan bahwa model berhasil mengklasifikasikan 450 data "Tidak Ada Diabetes" dan 171 data "Diabetes" dengan benar; namun, model tersebut salah mengklasifikasikan 50 data "Tidak Ada Diabetes" yang diprediksi sebagai "Diabetes" (positif palsu) dan 97 data "Tidak Ada Diabetes" yang diprediksi sebagai "Tidak Ada Diabetes" (negatif palsu). Matriks ini juga memberikan gambaran umum yang lebih rinci tentang jenis kesalahan yang dibuat.

Sebelum masuk ke tahap berikutnya, berikut ini digunakan confusion matrix untuk memberikan informasi bagaimana nilai tersebut akurasi, precision, recal, F1-score didapatkan, untuk perhitungan yang digunakan pada nilai-nilai di atas, berikut adalah keterangan dari nilai confusion matrix:

Tabel IV. 4 Confusion Matrix

True Positive (TP)	False Positive (FP)
False Negative (FN)	True Negative (TN)

Sumber: Penelitian, 2025

Untuk menghitung nilai precision, recall, dan akurasi, dapat menggunakan perhitungan confusion matrix. Berikut adalah rumus dan perhitungannya:

1) Perhitungan Akurasi

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Akurasi = \frac{171 + 450}{171 + 450 + 50 + 97} = \frac{621}{768} \approx 80.86$$

2) Precision

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{171}{171 + 50} = \frac{171}{221} \approx 77.38\%$$

3) Recall

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{171}{171 + 97} = \frac{171}{268} \approx 63.81\%$$

4) F-1 Score

$$F1 - Score = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall}$$

$$= \frac{2 \cdot (0.7738 \times 0.6381)}{0.7738 + 0.6381} = \frac{2 \times 0.4939}{1.4119} \approx 69.94\%$$

f. Halaman Tentang Aplikasi

Halaman ini disusun sebagai sarana untuk memberikan informasi kepada pengguna mengenai latar belakang dan struktur sistem yang dikembangkan,

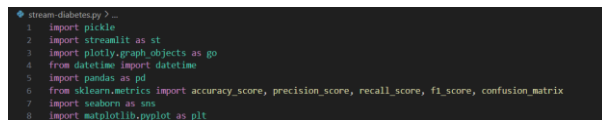
yaitu:

1) Tujuan Pengembangan Sistem

Sistem ini bertujuan untuk membantu pengguna, terutama masyarakat umum dan tenaga kesehatan, mendeteksi risiko diabetes dengan cepat dan mudah. Diharapkan sistem ini dapat membantu pengguna membuat keputusan awal sebelum berkonsultasi dengan tenaga medis. Oleh karena itu, sistem ini membantu pencegahan dan pengendalian diabetes dengan meningkatkan kesadaran dan pemahaman tentang faktor risiko yang terkait.

2) Teknologi yang Digunakan

Sistem ini dibangun menggunakan beberapa teknologi, yaitu:



```

1 import pickle
2 import streamlit as st
3 import plotly.graph_objects as go
4 from datetime import datetime
5 import pandas as pd
6 from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
7 import seaborn as sns
8 import matplotlib.pyplot as plt

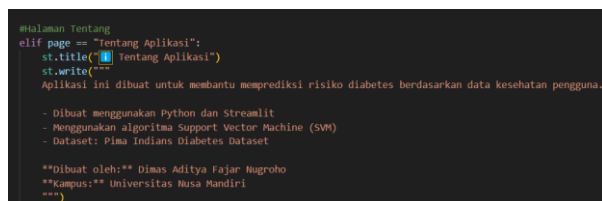
```

Gambar IV. 7 Halaman Evaluasi Model

Potongan kode program yang menunjukkan teknologi dan pustaka yang digunakan dalam pembuatan sistem prediksi diabetes ditunjukkan pada Gambar IV.7. Sistem dibangun menggunakan bahasa pemrograman Python, seperti yang ditunjukkan pada gambar. Ada beberapa pustaka populer yang digunakan, seperti pickle untuk menyimpan dan memuat model pembelajaran mesin, pandas untuk pengolahan data, dan datetime untuk pengelolaan waktu. Selain itu, pustaka plotly.graph_objects digunakan untuk menampilkan data interaktif, sedangkan scikit-learn digunakan untuk mengevaluasi kinerja model dengan modul seperti accuracy_score, precision_score, recall_score, f1_score, dan confusion_matrix. Pustaka matplotlib.pyplot juga digunakan untuk membantu visualisasi. Berbagai teknologi ini menunjukkan bahwa sistem dikembangkan dengan cara ilmiah dan memanfaatkan pustaka pembelajaran mesin, yang telah digunakan secara luas dalam bidang data science.

3) Identitas Pembuat Aplikasi dan Institusi

Aplikasi ini dikembangkan sebagai bagian tugas akhir oleh:



```

#Halaman Tentang
elif page == "Tentang Aplikasi":
    st.title("📄 Tentang Aplikasi")
    st.write("""
    Aplikasi ini dibuat untuk membantu memprediksi risiko diabetes berdasarkan data kesehatan pengguna.

    - Dibuat menggunakan Python dan Streamlit
    - Menggunakan algoritma Support Vector Machine (SVM)
    - Dataset: Pima Indians Diabetes Dataset

    ""Dibuat oleh:"" Dimas Aditya Fajar Nugroho
    ""Kampus:"" Universitas Nusa Mandiri
    """"

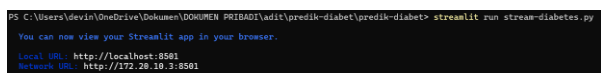
```

Gambar IV. 8 Identitas Pembuat Aplikasi dan Institusi

Halaman ini dirancang secara sederhana namun informatif dan jelas, sehingga pengguna dapat dengan mudah memahami identitas pengembang sistem, sumber data yang digunakan, serta teknologi dan perangkat yang digunakan selama proses pengembangan. Dengan memberikan pemahaman kepada pengguna tentang kredibilitas dan keandalan sistem, desain ini menunjukkan komitmen terhadap transparansi dan akuntabilitas akademis. Oleh karena itu, pengguna akan merasa lebih percaya diri dan yakin saat menggunakan sistem ini untuk mendeteksi risiko diabetes.

Code Generation

1. Streamlit Untuk Antarmuka Web



```

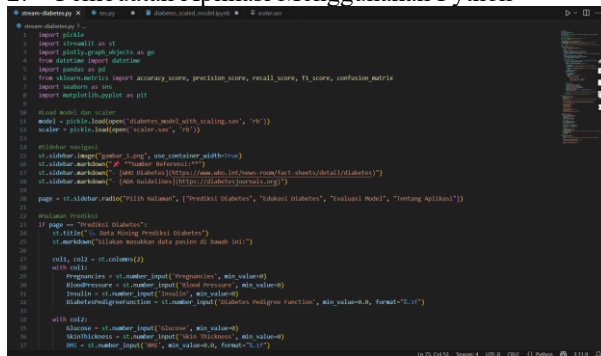
PS C:\Users\devin\OneDrive\Dokumen\PRIBADI\adit\predik-diabet\predik-diabet> streamlit run stream-diabetes.py
You can now view your Streamlit app in your browser.
Local URL: http://localhost:8501
Network URL: http://172.28.10.3:8501

```

Gambar IV. 9 Aplikasi dijalankan Menggunakan Streamlit (localhost)

Aplikasi prediksi diabetes berbasis web yang dikembangkan dengan kerangka kerja Streamlit telah dijelaskan di atas. Perintah "streamlit run stream-diabetes.py" dimasukkan ke terminal direktur proyek. Alamat ini secara otomatis membuat aplikasi di peramban menggunakan alamat jaringan yang tersedia, atau http://localhost:8501. Sebelumnya, aplikasi ini dikembangkan menggunakan Visual Studio Code sebagai penyunting kode, serupa dengan yang dilakukan saat membuat model deteksi diabetes. Penulisan dan pengembangan baris-baris kode Python digunakan untuk membuat aplikasi pengguna berbasis web menggunakan Visual Studio Code. Model pembelajaran yang telah dikembangkan dalam format Pickle (.pkl) dari Jupyter Notebook kemudian diintegrasikan ke dalam aplikasi Streamlit untuk melakukan prediksi interaktif berdasarkan masukan pengguna.

2. Pembuatan Aplikasi Menggunakan Python



```

# main-diabetes.py
1 import pickle
2 import streamlit as st
3 import plotly.graph_objects as go
4 from datetime import datetime
5 import pandas as pd
6 from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
7 import seaborn as sns
8 import matplotlib.pyplot as plt

9 # Load model dan scaler
10 model = pickle.load(open('diabetes_model_with_scaling.sav', 'rb'))
11 scaler = pickle.load(open('scaler.sav', 'rb'))

12 # Sidebar navigasi
13 st.sidebar.page_link('page1.py', label="Prediksi Diabetes")
14 st.sidebar.page_link('page2.py', label="Tentang Aplikasi")
15 st.sidebar.page_link('page3.py', label="Evaluasi Model")

16 # Judul dan deskripsi
17 st.title("🏠 Home")
18 st.subheader("Prediksi Risiko Diabetes")
19 st.write("""
20 Aplikasi ini dibuat untuk membantu memprediksi risiko diabetes berdasarkan data kesehatan pengguna.
21 """)

22 # Input form
23 st.form(
24     form=st.form("Form Input Data"),
25     key="form_input_data",
26     on_submit=lambda: predict_diabetes(st.session_state["form_input_data"]),
27     on_cancel=lambda: st.session_state["form_input_data"] = {}
28 )

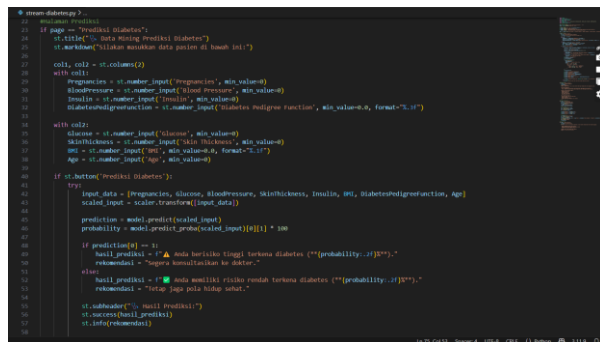
29 # Fungsi prediksi
30 def predict_diabetes(form_data):
31     # Mengambil data input
32     X_test = pd.DataFrame({
33         'Pregnancies': form_data['Pregnancies'],
34         'Glucose': form_data['Glucose'],
35         'BloodPressure': form_data['BloodPressure'],
36         'SkinThickness': form_data['SkinThickness'],
37         'Insulin': form_data['Insulin'],
38         'BMI': form_data['BMI'],
39         'DiabetesPedigreeFunction': form_data['DiabetesPedigreeFunction'],
40         'Age': form_data['Age']
41     })

42     # Melakukan prediksi
43     y_pred = model.predict(X_test)

44     # Mengembalikan hasil prediksi
45     return y_pred

```

Gambar IV. 11 Kode di dalam File Stream-diabetes



Gambar IV. 12 Kode di dalam File Stream-diabetes

Cuplikan kode dalam berkas bernama stream-diabetes.py, yang merupakan inti dari aplikasi web prediksi diabetes, ditunjukkan pada Gambar IV.11 dan IV.12. Berkas ini dibuat menggunakan bahasa pemrograman Python dan memiliki ekstensi.py, yang menunjukkan bahwa berkas ini merupakan berkas skrip Python. Fungsi berkas ini adalah untuk menghubungkan model pembelajaran mesin yang telah dibor sebelumnya dengan antarmuka pengguna yang dikembangkan menggunakan Streamlit.

Testing

1. Evaluasi Model dengan Metrik Klasifikasi

Tabel IV.5 Atribut Dataset Tahun 2024

Metrik	Nilai
Accuracy	79.79%
Precision	80.00%
Recall	76.20%
F1-Score	78.06%

Sumber: Penelitian Software Python

Pada tabel IV.5 nilai diperoleh dari model yang telah dilatih dengan database Pima Indians Diabetes dan diuji pada keseluruhan data atau testing, tergantung pada pembagian database.

2.

3. Confusion Matrix



Gambar IV. 12 Kode di dalam File Stream-diabetes

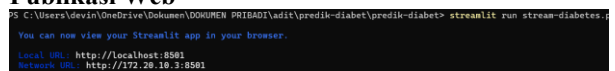
Terlihat dalam kode yang ditampilkan bahwa model melakukan prediksi.prediksi(X_test), lalu bandingkan hasil prediksi dengan label sebenarnya (y_test) untuk menghitung metrik evaluasi seperti akurasi, ketepatan, recall, dan skor F1. Selanjutnya, confusion matrix dibuat dengan menggunakan fungsi confusion_matrix() dari pustaka sklearn.metrics. Matriks ini kemudian

divisualisasikan menggunakan pustaka plotly.graph_objects, yang memungkinkan melihat matriks dalam bentuk heatmap interaktif. Setiap kelas ("Diabetes" dan "Tidak Diabetes") memiliki warna dan nilai yang menunjukkan jumlah prediksi yang benar dan salah. Ini memudahkan pengguna untuk menilai kinerja model secara visual dan mudah dipahami.

Support

Bagian ini menjelaskan dukungan teknis terhadap sistem, baik dalam bentuk publikasi aplikasi maupun perangkat keras dan lunak yang digunakan dalam pengembangannya. Hal ini sangat penting agar sistem dapat dijalankan secara optimal serta juga dapat dipahami oleh pihak yang ingin mengakses atau mereplikasi sistem.

Publikasi Web



Gambar IV. 13 Streamlit run stream-diabetes.py CMD

Sistem aplikasi prediksi diabetes ini akan dipublikasikan secara online menggunakan layanan Streamlit Cloud. Pertama, membuat akun GitHub dan kemudian mengunggah file proyek lengkap, yang terdiri dari kode program, model yang telah diekspor, dan aset pendukung lainnya, ke dalam repositori publik GitHub. Setelah proses unggah selesai, Streamlit Cloud dapat dikonfigurasi untuk menarik proyek tertentu. Oleh karena itu, aplikasi dapat dijalankan dan digunakan secara online melalui URL hosting resmi yang disediakan oleh Streamlit. Ini membuat pengguna dapat mengakses dan menggunakan sistem ini kapan saja tanpa perlu melakukan instalasi lokal.

Spesifikasi Hardware dan Software

Tabel IV.6 Spesifikasi Hardware

Komponen	Spesifikasi
Prosesor	Intel Core i5-8250U 1.6GHz up to 3.4GHz
RAM	8 GB DDR4
Penyimpanan	512 GB SSD
VGA	Intel UHD Graphics 620
Sistem Operasi	Windows 10 Pro 64-bit

Sumber: Penelitian

Tabel IV.7 Spesifikasi Software

Perangkat Lunak	Keterangan
Bahasa Pemrograman	Python 3.11.9
IDE	Visual Studio Code
Framework	Streamlit 1.x
Library ML	Scikit-learn, Pandas, NumPy
Visualisasi	Matplotlib, Seaborn
Environment Manager	Anaconda / pip
Sistem Operasi	Windows 10 Pro

Sumber:Penelitian

4. Conclusion

Aplikasi prediksi risiko diabetes yang telah dikembangkan berbasis web menggunakan framework Streamlit. Tampilan utama aplikasi didesain dengan warna biru-putih yang bersih dan profesional, disesuaikan dengan tema kesehatan. Aplikasi ini memiliki struktur navigasi yang ditempatkan di sidebar, memudahkan pengguna untuk berpindah antar halaman seperti halaman prediksi, edukasi diabetes, dan evaluasi model. Fitur utama yang disediakan meliputi:

1. Form input data pasien yang terdiri dari 8 parameter kesehatan (Glucose, BMI, Age, dll).
2. Output hasil prediksi berupa status risiko diabetes serta persentase probabilitas dalam bentuk visualisasi speedometer.
3. Ringkasan data input pengguna dan waktu prediksi dilakukan.
4. Halaman edukasi yang memberikan informasi mengenai diabetes dan pentingnya deteksi dini.
5. Halaman evaluasi model, yang menampilkan hasil evaluasi akurasi, precision, recall, dan f1-score dari model SVM yang digunakan.

Aplikasi ini telah berhasil mengintegrasikan model klasifikasi SVM ke dalam antarmuka yang user-friendly dan edukatif. Dengan struktur yang sederhana namun informatif, pengguna dari berbagai latar belakang dapat menggunakan aplikasi ini dengan mudah dan cepat.

References

- [1] F. A. Fatmona, D. R. Permana, and A. Sakurawati, "Gambaran Tingkat Pengetahuan Masyarakat tentang Pencegahan Diabetes Melitus Tipe 2 di Puskesmas Perawatan Siko," *MAHESA: Malahayati Health Student Journal*, vol. 3, no. 12, pp. 4166–4178, Dec. 2023, doi: 10.33024/mahesa.v3i12.12581.
- [2] R. Risty Wardhani, "PENGARUH AEROBIC EXERCISE TERHADAP KADAR GULA DARAH SEWAKTU DAN KUALITAS HIDUP PADA LANSIA DENGAN KONDISI DIABETES MELITUS TIPE II," no. 1, 2025.
- [3] E. Ristiyowati, P. Studi Ilmu Keperawatan, and S. Dian Husada Mojokerto, "OPTIMALISASI KADAR GULA DALAM DARAH PADA PENDERITA DIABETES MELLITUS MELALUI HIPNOTERAPI," 2023. [Online]. Available: <http://e-journal.lppmdianhusada.ac.id/index.php/PIPK>
- [4] J. Biologi et al., "Diabetes Melitus: Review Etiologi." [Online]. Available: <http://journal.uin-alauddin.ac.id/index.php/psb>
- [5] A. Riset et al., "FAKUMI MEDICAL JOURNAL Hubungan Usia, Jenis Kelamin dan Indeks Massa Tubuh (IMT) dengan Kadar HbA1c di Rumah Sakit Ibnu Sina Makassar."
- [6] E. Indriyani, T. Kesuma Dewi, and P. DIII Keperawatan Akper Dharma Wacana Metro, "PENERAPAN SENAM KAKI DIABETES MELITUS TERHADAP KADAR GLUKOSA DARAH PADA PENDERITA DIABETES MELLITUS DI PUSKESMAS YOSOMULYO THE APPLICATION OF DIABETES MELLITUS FOOT EXERCISE TO BLOOD GLUCOSE LEVELS IN DIABETES MELLITUS PATIENTS AT PUSKESMAS YOSOMULYO," *Jurnal Cendikia Muda*, vol. 3, no. 2, 2023.
- [7] N. D. Hendrawan, A. Saivul Affandi, and R. F. Fadhlirifat, "Implementasi Support Vector Machine dalam Deteksi Diabetes Melalui Indikator Kesehatan," pp. 664–671, 2023, [Online]. Available: <https://jurnal.usahidsolo.ac.id/>
- [8] Adih Adih, Wahyu Aji Dwi Pangestu, Muhamad Fauzi Akbar, Purnamasari Purnamasari, Farlin Wabula, and Ines Heidiani Ikasari, "Literature Review : Penggunaan Machine Learning Berbasis SVM untuk Klasifikasi Penyakit Diabetes," *Merkurius: Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 3, no. 1, pp. 156–168, Jan. 2025, doi: 10.61132/mercurius.v3i1.616.
- [9] T. Septiana, M. A. Muda, D. Budiyo, M. Pratama, and W. P. Jaya, "Analisis Penggunaan Support Vector Machine pada Deteksi Dini Penyakit Diabetes Melitus," *Jurnal Penelitian Inovatif*, vol. 4, no. 3, pp. 1631–1640, Aug. 2024, doi: 10.54082/jupin.643.
- [10] S. Suliman, "Implementasi Data Mining Terhadap Prestasi Belajar Mahasiswa Berdasarkan Pergaulan dan Sosial Ekonomi Dengan Algoritma K-Means Clustering," *SIMKOM*, vol. 6, no. 1, pp. 1–11, Jan. 2021, doi: 10.51717/simkom.v6i1.48.
- [11] S. Putri Yuanita et al., "Science, Technology and Management Journal Identifikasi Teknik Data Mining Metode Asosiasi: Systematic Literature Review Info Artikel," *Science Technology and Management Journal*, vol. 4, no. 2, p. 57, 2024, doi: 10.53416/stmj.v4i2.
- [12] A. Rahman Kadafi, R. Eko Indrajid, dan Muh Fauzi, and S. Nusa Mandiri Jakarta, "PENERAPAN KONSEP BUSINESS INTELLIGENCE UNTUK MERANCANG STRATEGI MARKETING PADA SEKOLAH ISLAM TERPADU NURUL FIKRI."
- [13] M. A. Wiratama and W. M. Pradnya, "Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 1, p. 1, Apr. 2022, doi: 10.23887/janapati.v11i1.45282.