

**PREDIKSI PENYAKIT DIABETES MENGGUNAKAN MACHINE
LEARNING**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh

HERLANGGA GUSTIAN

18.83.0178

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**PREDIKSI PENYAKIT DIABETES MENGGUNAKAN MACHINE
LEARNING**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh

HERLANGGA GUSTIAN

18.83.0178

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**PREDIKSI PENYAKIT DIABETES MENGGUNAKAN MACHINE
LEARNING**

yang disusun dan diajukan oleh

Herlangga Gustian

18.83.0178

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Agustus 2023

Dosen Pembimbing,



Banu Santoso, S.T., M.Eng
NIK. 190302327

HALAMAN PENGESAHAN

SKRIPSI

HERLANGGA GUSTIAN

yang disusun dan diajukan oleh

Herlangga Gustian

18.83.0178

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Agustus 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Muhammad Kopravi, S.Kom., M.Eng
NIK. 190302454

Jeki Kuswanto, M.Kom
NIK. 190302456

Banu Santoso, S.T., M.Eng
NIK. 190302327



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Agustus 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Herlangga Gustian
NIM : 18.83.0178

Menyatakan bahwa Skripsi dengan judul berikut:

Prediksi Penyakit Diabetes Menggunakan Machine Learning

Dosen Pembimbing : Banu Santoso, S.T., M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 22 Agustus 2023

Yang Menyatakan,

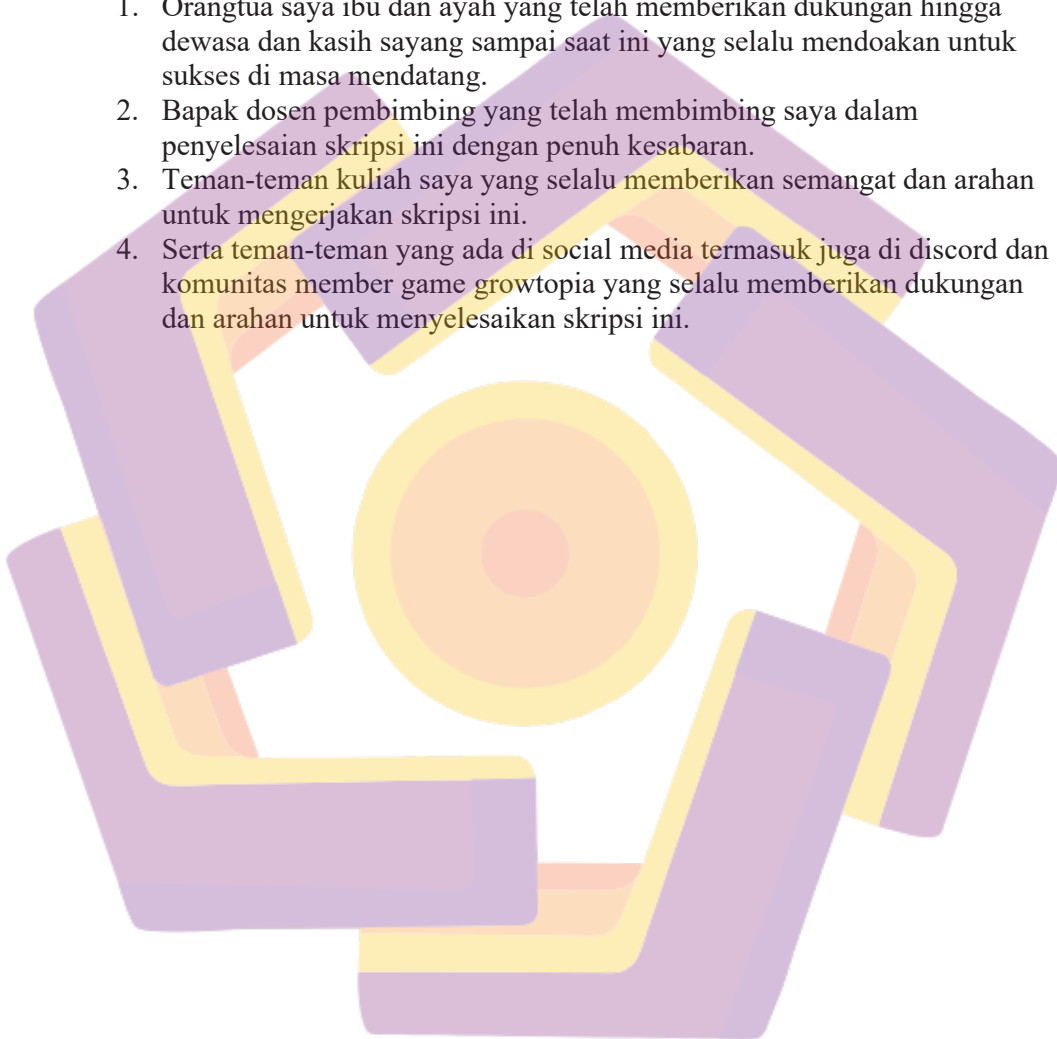


Herlangga Gustian

HALAMAN PERSEMBAHAN

Saya ucapkan syukur kepada Allah SWT atas selesainya naskah skripsi ini dengan baik dan lancar. Skripsi ini saya persembahkan untuk :

1. Orangtua saya ibu dan ayah yang telah memberikan dukungan hingga dewasa dan kasih sayang sampai saat ini yang selalu mendoakan untuk sukses di masa mendatang.
2. Bapak dosen pembimbing yang telah membimbing saya dalam penyelesaian skripsi ini dengan penuh kesabaran.
3. Teman-teman kuliah saya yang selalu memberikan semangat dan arahan untuk mengerjakan skripsi ini.
4. Serta teman-teman yang ada di social media termasuk juga di discord dan komunitas member game growtopia yang selalu memberikan dukungan dan arahan untuk menyelesaikan skripsi ini.



KATA PENGANTAR

Bismillahirrohmanirrohim.

Puji dan syukur saya panjatkan kepada Allah SWT, atas ridanya sehingga saya dapat menyelesaikan penyusunan skripsi ini. Adapun judul skripsi yang saya ajukan adalah prediksi penyakit diabetes dengan machine learning.

Skripsi ini diajukan untuk memenuhi syarat kelulusan mata kuliah Skripsi di Universitas Amikom Yogyakarta. Tidak dapat disangkal bahwa butuh usaha yang sangat keras dalam penyelesaian pengerjaan skripsi ini. Namun, skripsi ini tidak akan selesai tanpa orang-orang tercinta di sekeliling kita yang mendukung dan membantu. Terima kasih saya sampaikan kepada:

1. Prof, Dr. M. Suyanto, MM., Selaku Rektor Universitas Amikom Yogyakarta
2. Banu Santoso, S.T., M.Eng selaku Dosen pembimbing yang telah memberikan bimbingan dan berbagai pengalaman kepada saya.
3. Segenap Dosen Teknik Komputer yang telah mendidik dan memberikan ilmu selama kuliah dan seluruh staf yang selalu sabar melayani segala administrasi.
4. Semua pihak yang telah membantu dan tidak dapat satu persatu.

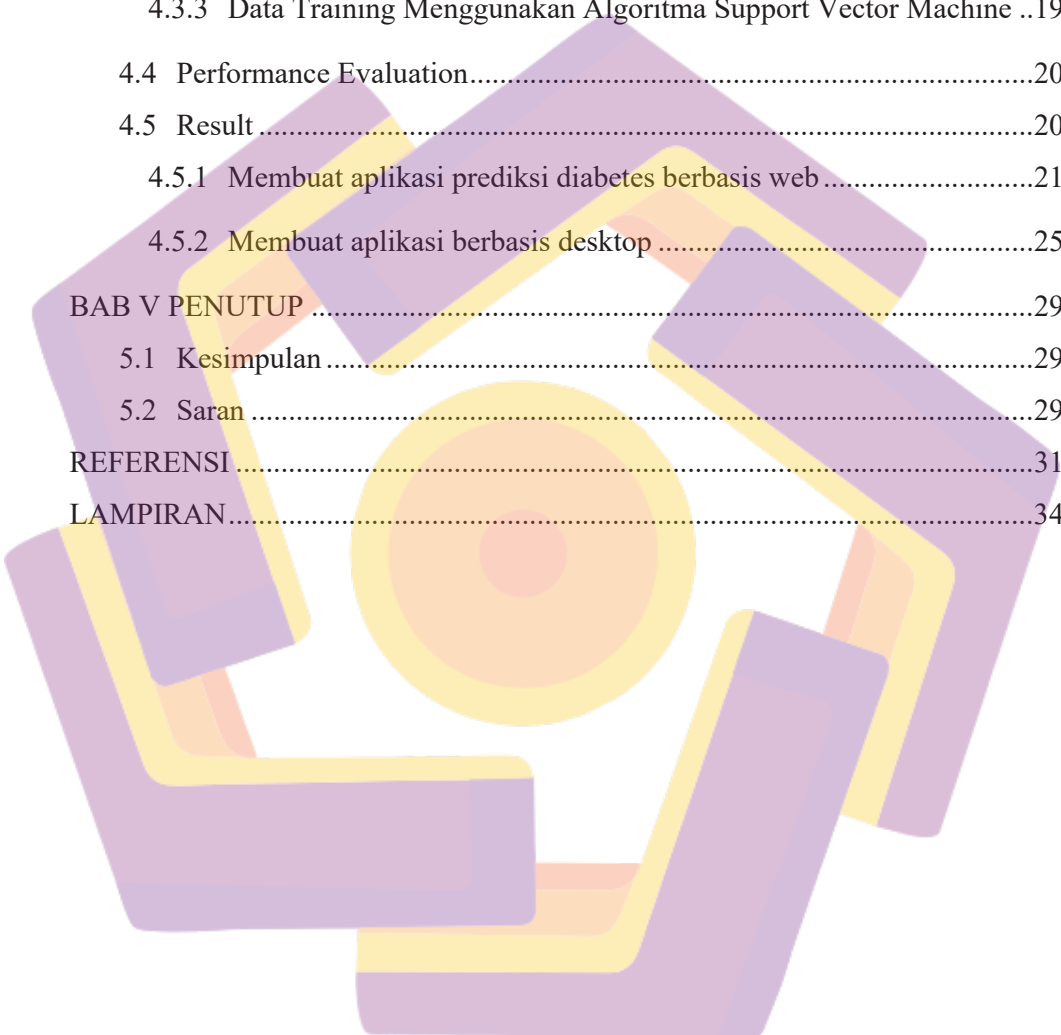
Semoga segala pertolongan dan kebaikan semuanya mendapat berkah dari Allah SWT. Dan akhirnya saya menyadari bahwa skripsi ini masih jauh dari kata sempurna, karena keterbatasan ilmu yang saya miliki. Untuk itu saya dengan kerendahan hati mengharapkan kritik dan saran yang sifatnya membangun dari semua pihak demi membangun laporan penelitian ini.

Yogyakarta, 22 Agustus 2023

Penulis

DAFTAR ISI

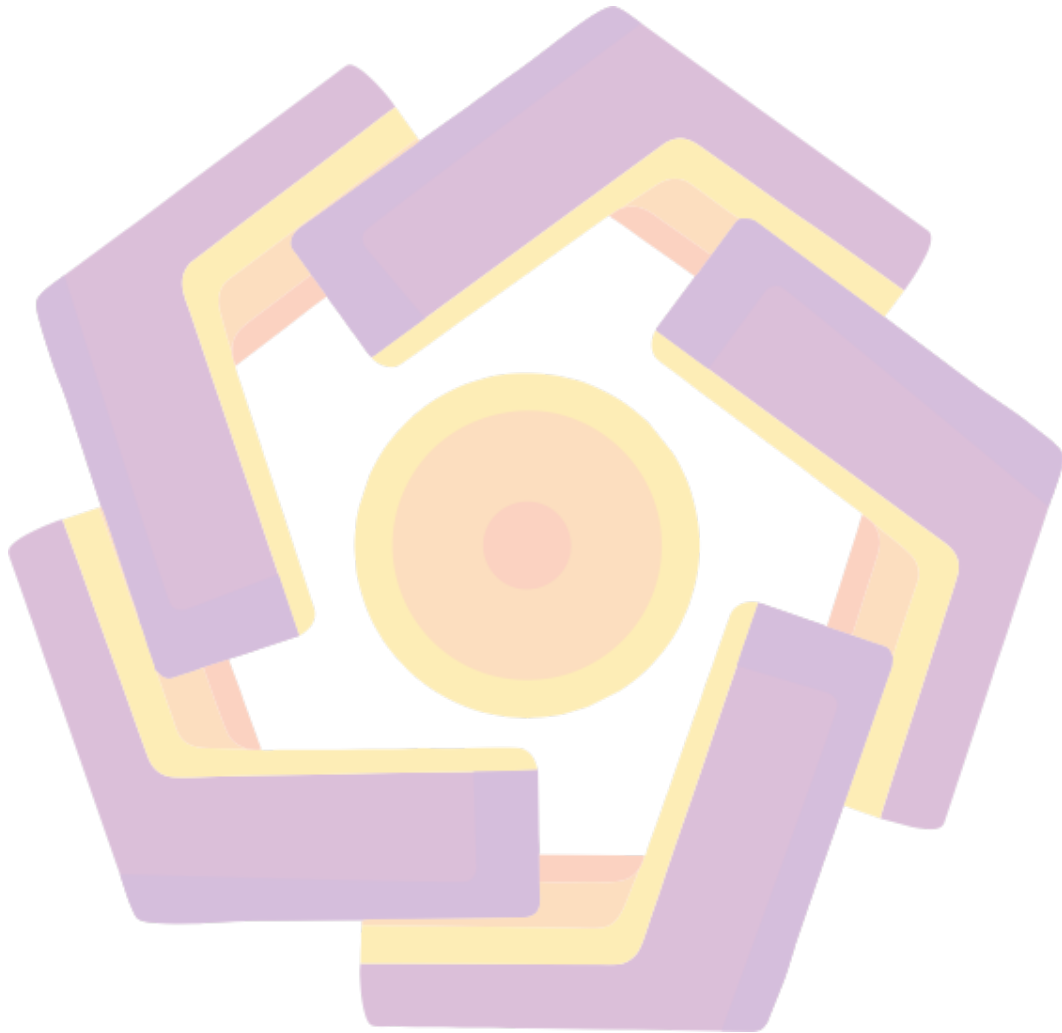
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
INTISARI	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Landasan Teori	7
BAB III METODE PENELITIAN	13
3.1 Objek Penelitian.....	13
3.2 Alur Penelitian	13
3.3 Alat dan Bahan.....	14
BAB IV HASIL DAN PEMBAHASAN	16
4.1 Dataset	16



4.2	Preprocessing Data	16
4.3	Classification Algorithms	19
4.3.1	Data Training menggunakan Algoritma Random Forest.....	19
4.3.2	Data Training menggunakan Algoritma Logistic Regression	19
4.3.3	Data Training Menggunakan Algoritma Support Vector Machine ..	19
4.4	Performance Evaluation.....	20
4.5	Result	20
4.5.1	Membuat aplikasi prediksi diabetes berbasis web	21
4.5.2	Membuat aplikasi berbasis desktop	25
BAB V	PENUTUP	29
5.1	Kesimpulan	29
5.2	Saran	29
REFERENSI	31
LAMPIRAN	34

DAFTAR TABEL

Tabel 2. 1 Perbandingan Metode	6
Tabel 4. 1 Hasil akurasi dari training data	20

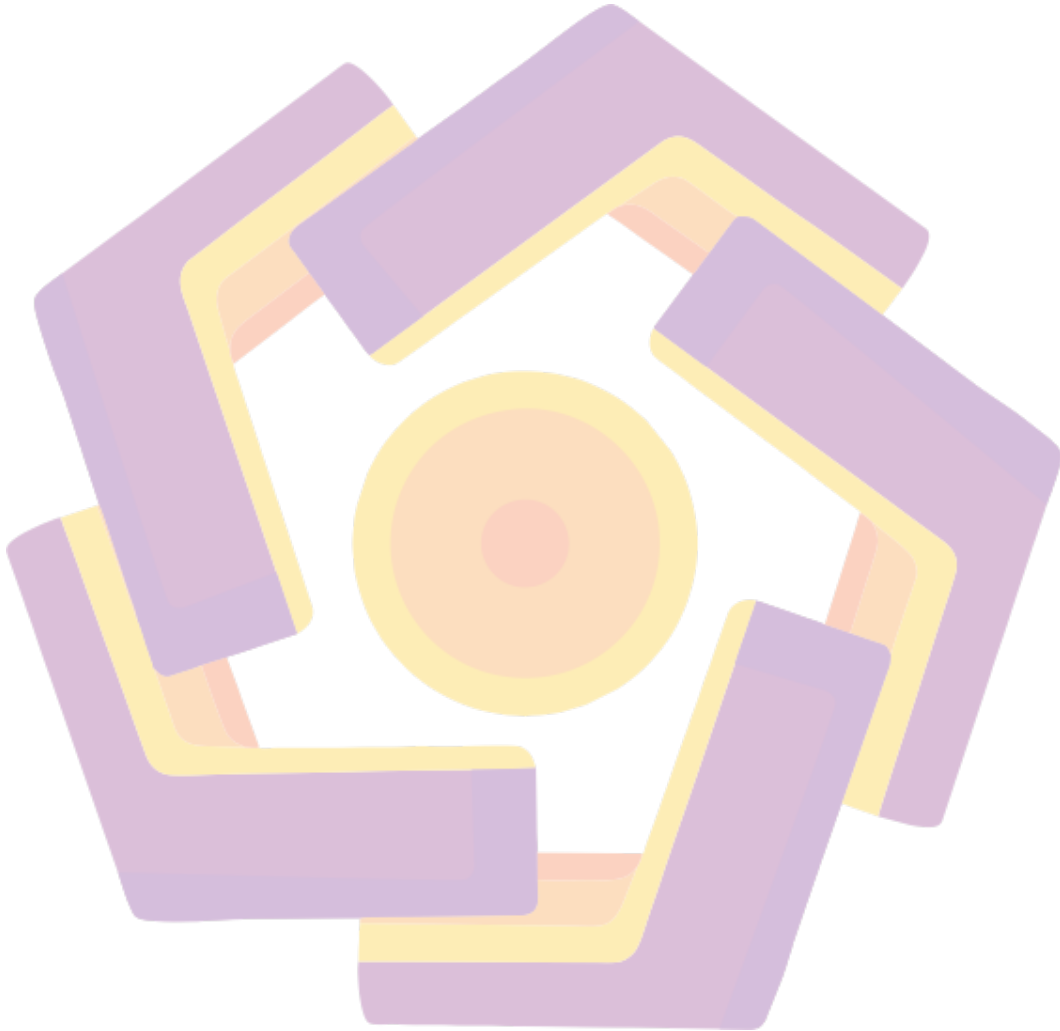


DAFTAR GAMBAR

Gambar 2. 2 Ilustrasi Gambaran Random Forest.....	11
Gambar 2. 3 Ilustrasi Gambaran Support Vector Machine.....	12
Gambar 3. 1 Alur Penelitian	14
Gambar 4. 1 Tampilan Isi Dataset Dengan Bentuk Chart	16
Gambar 4. 2 Proses Ubah Nilai 0 Menjadi Null	17
Gambar 4. 3 Mengecek Jumlah Nilai Null Yang Sudah Diubah	17
Gambar 4. 4 Mengisi Nilai Null dengan Median.....	18
Gambar 4. 5 Mengecek Jumlah Nilai Setelah Diisi Dengan Median	18
Gambar 4. 6 Proses Membagi Data Training dan Data Testing	18
Gambar 4. 7 Proses Training Data dengan Random Forest.....	19
Gambar 4. 8 Proses Training Data dengan Logistic Regression	19
Gambar 4. 9 Proses Training Data Dengan SVM.....	20
Gambar 4. 10 Hasil Matrix Heatmap Dari Logistic Regression	20
Gambar 4. 11 Proses Menyimpan File Dari Model Machine Learning.....	21
Gambar 4. 12 Proses Import Modul Untuk Aplikasi Web.....	21
Gambar 4. 13 Memasukan File Model Prediksi Ke Program	21
Gambar 4. 14 Membuat Inputan Data Ke Aplikasi	22
Gambar 4. 15 Proses Untuk Menjalankan Prediksi	22
Gambar 4. 16 Menjalankan File Program Prediksi Dari CMD	22
Gambar 4. 17 Tampilan Input Basis Web.....	23
Gambar 4. 18 Tampilan tabel dari 2 data pertama	23
Gambar 4. 19 Hasil Jika Pasien Positif Terkena Diabetes	24
Gambar 4. 20 Hasil jika pasien tidak terkena diabetes	24
Gambar 4. 21 Tampilan UI aplikasi desktop	25
Gambar 4. 22 Nama Object dan Class digunakan	26
Gambar 4. 23 Memasukan File Model Prediksi Ke Program	26
Gambar 4. 24 Fungsi Untuk Input Data dan Menjalankan Program	27
Gambar 4. 25 Tampilan Menjalankan Program Aplikasi Desktop	27
Gambar 4. 26 Hasil Jika Pasien Terkena Diabetes	28
Gambar 4. 27 Hasil Jika Pasien Tidak Terkena Diabetes	28

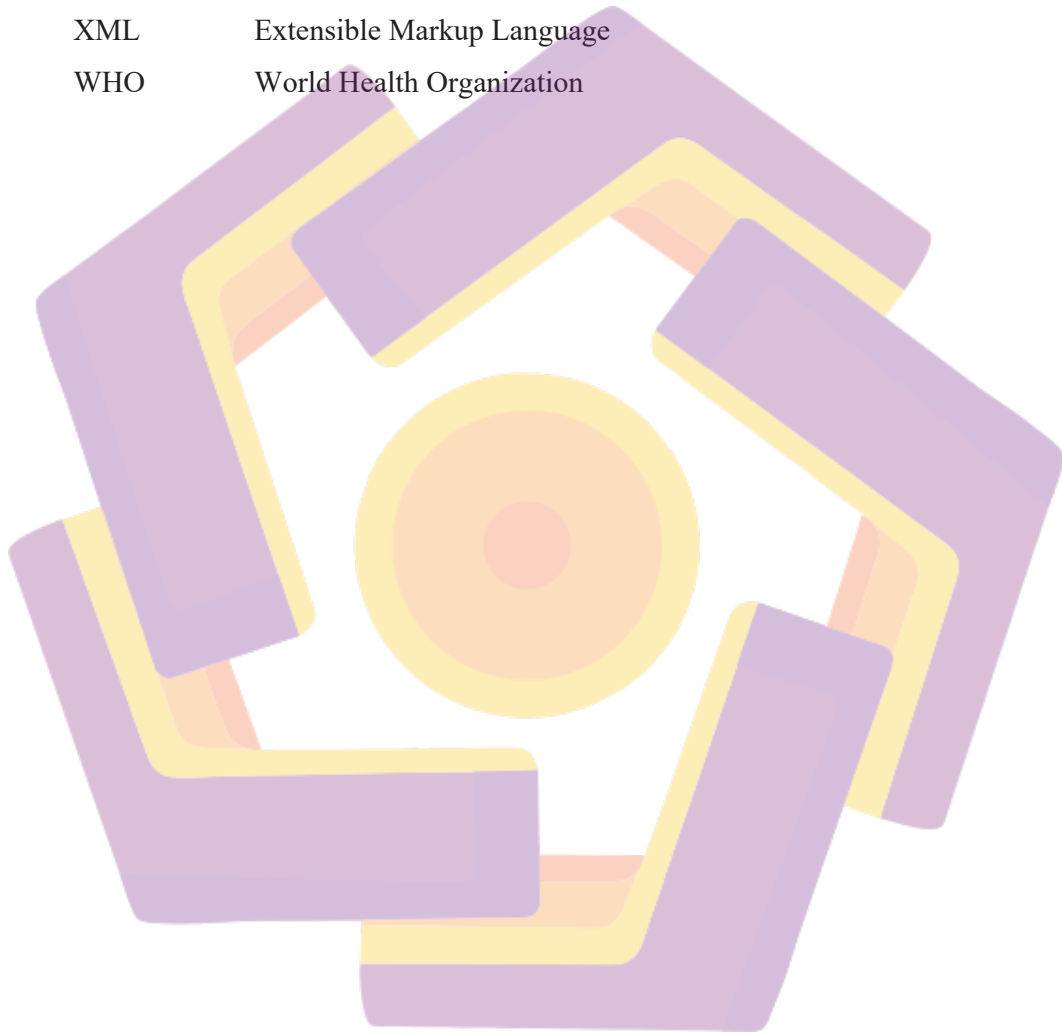
DAFTAR LAMPIRAN

Lampiran 1. Source Code Program	33
Lampiran 2. Source Code Aplikasi Desktop	35
Lampiran 3. Source Code Streamlit	35
Lampiran 4. Class yang digunakan membuat UI di Qt Designer	36



DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
IDF	International Diabetes Federation
UI	User Interface
XML	Extensible Markup Language
WHO	World Health Organization



INTISARI

Diabetes adalah salah satu penyakit yang bisa menyerang siapa saja yang menjadi masalah dalam dunia kesehatan, karena bisa menyerang anak-anak, remaja, hingga orang dewasa. Diabetes merupakan penyakit yang mengancam jiwa karena menyebabkan peningkatan gula darah, selain itu juga memiliki pertumbuhan tercepat yang telah mempengaruhi 422 juta orang di dunia berdasarkan laporan dari Organisasi Kesehatan Dunia (WHO), pada tahun 2018. Bahkan International Diabetes Federation (IDF) memperkirakan jumlah penderita diabetes di seluruh dunia akan mencapai 700 juta orang pada 2045. Namun untuk bisa membuat penyakit diabetes bisa didiagnosis bisa dilakukan menggunakan machine learning sebagai solusi untuk memecahkan permasalahan ini. Maksud dari penelitian ini adalah membuat model yang dapat memperkirakan kemungkinan terjadinya diabetes sejak awal dengan ketelitian yang maksimal. Pada penelitian ini menggunakan menggunakan Klasifikasi untuk teknik data mining yang menetapkan kategori dari sekumpulan data untuk mempermudah dalam memprediksi dan analisis yang akurat. Untuk algoritma yang digunakan ada tiga yaitu Logistic Regression, Support Vector Machine, dan Random Forest dalam percobaan mendeteksi penyakit diabetes. Untuk mengevaluasi dari performa ketiga algoritma ini menggunakan Accuracy. Akurasi dinilai melalui instance yang diklasifikasikan dengan benar dan salah. Pada penelitian ini algoritma tertinggi adalah Support Vector Machine dengan nilai 78% pada akurasinya.

Kata kunci: machine learning, diabetes, logistic regression, random forest, akurasi.

ABSTRACT

Diabetes is a disease that can affect anyone who is a problem in the world of health, because it can attack children, adolescents, and adults. Diabetes is a life-threatening disease because it causes an increase in blood sugar, besides that it also has the fastest growth which has affected 422 million people in the world based on a report from the World Health Organization (WHO), in 2018. Even the International Diabetes Federation (IDF) estimates the number of people with diabetes worldwide will reach 700 million people by 2045. But to be able to make diabetes can be diagnosed can be done using machine learning as a solution to solve this problem. The purpose of this study is to create a model that can estimate the likelihood of developing diabetes from the outset with maximum accuracy. In this study using Classification for data mining techniques that define categories from a set of data to facilitate accurate prediction and analysis. There are three algorithms used, namely Logistic Regression, Support Vector Machine, and Random Forest in an experiment to detect diabetes. To evaluate the performance of these three algorithms using Accuracy. Accuracy is assessed through correctly and falsely classified instances. In this research, the highest algorithm is Support Vector Machine with a value of 78% in accuracy.

Keyword: machine learning, diabetes, logistic regression, random forest, accuracy.