

**PERBANDINGAN METODE *PRINCIPAL COMPONENT ANALYSIS*
(PCA) DAN *SEQUENTIAL FORWARD SELECTION* (SFS) DENGAN
SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI
DIABETES**

Tugas akhir

Untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S-1 Jurusan Teknik Elektro



Oleh:

SAPUTRA

F1B019129

JURUSAN TEKNIK ELEKTRO

FAKULTAS TEKNIK

UNIVERSITAS MATARAM

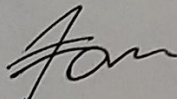
TAHUN 2023

Tugas Akhir

PERBANDINGAN METODE *PRINCIPAL COMPONENT ANALYSIS* (PCA)
DAN *SEQUENTIAL FORWARD SELECTION* (SFS) DENGAN *SUPPORT*
VECTOR MACHINE (SVM) DALAM KLASIFIKASI DIABETES

Telah diperiksa dan disetujui oleh Tim Pembimbing

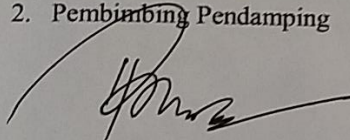
1. Pembimbing Utama



L. Ahmad S. Irfan Akbar, ST., M.Eng
NIP. 19830310 200912 1 004

Tanggal: 6/09/23

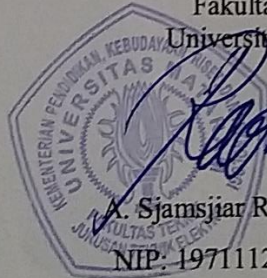
2. Pembimbing Pendamping

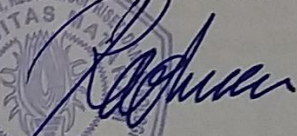


Cipta Ramadhani, ST., M.Eng
NIP. 19850616 201903 1 008

Tanggal: 6/09/23

Mengetahui
Ketua Jurusan Teknik Elektro
Fakultas Teknik
Universitas Mataram




A. Sjamsijar Rachman, ST., MT
NIP. 19711124 199903 1 004

Tugas Akhir

**PERBANDINGAN METODE *PRINCIPAL COMPONENT ANALYSIS* (PCA)
DAN *SEQUENTIAL FORWARD SELECTION* (SFS) DENGAN *SUPPORT VECTOR MACHINE* (SVM) DALAM KLASIFIKASI DIABETES**

Oleh:

Šaputra

F1B019129

Telah dipertahankan di depan Dewan Penguji

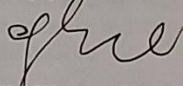
Pada Tanggal 30 Agustus 2023

dan dinyatakan telah memenuhi syarat mencapai derajat Sarjana S-1

Jurusan Teknik Elektro

Susunan Tim Penguji

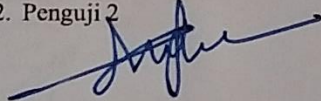
1. Penguji 1



Giri Wahyu Wiriasto, ST., MT.
NIP. 19820904 201012 1 001

Tanggal: 6-9-2023

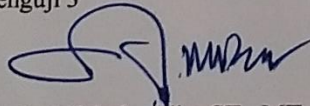
2. Penguji 2



Made Sutha Yadnya, ST., MT.
NIP. 19720117 199903 1 001

Tanggal: 07-2023
09

3. Penguji 3



Dr. Ir. Misbahuddin, ST., MT., IPU.
NIP. 19681005 199703 1 001

Tanggal: 6-9-2023

Mataram, 7 September 2023

Dekan Fakultas Teknik
Universitas Mataram



Muhamad Syamsu Iqbal, ST., MT., Ph.D.
NIP. 19720222 199903 1 002

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama :
NIM :
Program Studi :
Perguruan Tinggi :

Dengan ini menyatakan bahwa tugas akhir yang telah saya buat dengan judul: “Perbandingan Metode *Principal Component Analysis* (PCA) dan *Sequential Forward Selection* (SFS) Dengan *Support Vector Machine* (SVM) Dalam Klasifikasi Diabetes” adalah asli (orisinal) atau tidak plagiat (menjiplak) dan belum pernah diterbitkan/dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya tanpa ada paksaan dari pihak manapun juga. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tugas akhir yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Universitas Mataram dicabut/dibatalkan.

Dibuat di :
Pada tanggal :
Yang menyatakan

Saputra

PRAKATA

Dalam kesempatan ini, penulis ingin mengucapkan rasa syukur dan terima kasih atas dukungan dan bantuan dari berbagai pihak yang telah membantu dalam penyelesaian penelitian ini. Penelitian ini merupakan hasil dari dedikasi dan kerja keras penulis dalam mengeksplorasi dan memahami penelitian ini yang berjudul “Perbandingan Metode *Principal Component Analysis* (PCA) dan *Sequential Forward Selection* (SFS) Dengan *Support Vector Machine* (SVM) Dalam Klasifikasi Diabetes” dengan lebih mendalam.

Penelitian ini bertujuan untuk melihat perbandingan kinerja *Principal Component Analysis* dan *Sequential Forward Selection* dengan *Support Vector Machine* dalam klasifikasi diabetes, serta membuat suatu website yang dapat digunakan untuk melakukan klasifikasi diabetes. Penulis berharap hasil penelitian ini dapat memberikan kontribusi positif dalam pengembangan ilmu pengetahuan di bidang *machine learning*. Selama proses penelitian, penulis menghadapi berbagai tantangan dan kesulitan, namun berkat dukungan dan bimbingan dari berbagai pihak, penulis dapat mengatasi setiap hambatan tersebut.

Meskipun penelitian ini telah selesai, penulis menyadari bahwa penelitian ilmiah adalah sebuah proses berkesinambungan. Oleh karena itu, penulis terbuka untuk menerima masukan dan saran dari berbagai pihak guna meningkatkan kualitas penelitian ini di masa mendatang.

Akhir kata, penulis berharap hasil dari penelitian ini dapat memberikan manfaat bagi seluruh masyarakat. Semoga penelitian ini dapat menjadi pijakan untuk penelitian lebih lanjut di bidang yang sama, dan berkontribusi dalam upaya memajukan ilmu pengetahuan dan teknologi.

Mataram, 25 Juli 2023

Penulis

UCAPAN TERIMA KASIH

Tugas Akhir ini dapat diselesaikan berkat bimbingan, dukungan dan bantuan dari berbagai pihak, oleh karena itu pada kesempatan kali ini penulis ingin mengucapkan rasa terimakasih yang tulus kepada:

1. Keluarga penulis, terutama kedua orang tua yaitu Bapak Putu Sumatra dan Ibu Siti Hamidah serta keempat kakak penulis yaitu Agus Darmawan, Medi Sumardana, Nyoman Budi Trisutrisna dan Ketut Ade Kurniawan.
2. Bapak Muhamad Syamsu Iqbal, ST., MT., Ph.D., selaku Dekan Fakultas Teknik Universitas Mataram.
3. Bapak A. Sjamsjiar Rachman, S.T., M.T. selaku Ketua Jurusan Teknik Elektro Universitas Mataram.
4. Bapak L. Ahmad S. Irfan Akbar, ST., M.Eng. dan Bapak Cipta Ramadhani, ST., M.Eng. selaku Dosen Pembimbing yang telah memberikan bimbingan dan arahan selama penyusunan Tugas akhir ini dilakukan sehingga dapat diselesaikan.
5. Bapak Giri Wahyu Wiriasto, S.T., M.T., Bapak Made Sutha Yadnya, S.T., M.T. dan Bapak Dr. Ir. Misbahuddin, ST., MT., IPU. selaku Tim Dosen Penguji yang telah memberikan saran, kritik dan arahan untuk meningkatkan kualitas Tugas Akhir ini.
6. Jajaran Dosen, Staff dan Karyawan Teknik Elektro Universitas Mataram yang telah banyak memberikan ilmu yang bermanfaat selama masa perkuliahan.
7. Teman-teman seperjuangan Teknik Elektro 2019 yang telah banyak membantu dan memotivasi penulis dalam menyelesaikan Tugas Akhir ini.
8. A1B021081 sebagai perempuan yang terus memberikan dukungan dengan tulus untuk berjuang menyelesaikan Tugas Akhir ini.
9. Semua pihak lain yang tidak dapat disebutkan satu – persatu yang telah membantu baik dalam bentuk saran, kritik dan motivasi.

Semoga Tuhan Yang Maha Esa memberikan imbalan yang setimpal atas bantuan yang telah diberikan kepada penulis.

DAFTAR ISI

COVER.....	i
LEMBAR PENGESAHAN PEMBIMBING.....	ii
LEMBAR PENGESAHAN PENGUJI.....	iii
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR	iv
PRAKATA.....	v
UCAPAN TERIMA KASIH	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
ABSTRAK.....	xiii
ABSTRACT.....	xiv
BAB I.....	1
PENDAHULUAN	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan Laporan	3
BAB II.....	5
TINJAUAN PUSTAKA DAN DASAR TEORI	5
2.1 Tinjauan Pustaka	5
2.2 Dasar Teori.....	6
2.2.1 Diabetes	6
2.2.2 Machine Learning.....	7
2.2.3 <i>Support Vector Machine</i> (SVM)	8
2.2.4 <i>Principal Component Analysis</i> (PCA).....	9
2.2.5 <i>Sequential Forward Selection</i>	9
2.2.6 Performance Evaluation Measure.....	10
2.2.7 <i>Confusion Matrix</i>	11
2.2.8 <i>Python</i>	12
2.2.9 <i>Streamlit</i>	12
2.2.10 Visual Studio Code.....	13

BAB III	14
METODOLOGI PENELITIAN	14
3.1 Alur Penelitian	14
3.2 Studi Literatur	14
3.3 Pengumpulan <i>Dataset</i>	15
3.4 Analisis Kebutuhan	16
3.5 <i>Preprocessing Data</i>	17
3.6 Pelatihan Model	17
3.6.1 Ekstraksi Fitur PCA.....	17
3.6.2 Seleksi Fitur SFS	19
3.6.3 Latih Model SVM.....	20
3.7 Evaluasi Model.....	22
3.7.1 Uji Model SVM.....	22
3.7.2 Perbandingan Kedua Model	22
3.8 Perancangan Website Klasifikasi Diabetes	22
BAB IV	31
HASIL DAN PEMBAHASAN	31
4.1 Hasil <i>Data Cleaning</i>	31
4.2 Pembagian Dataset	31
4.3 <i>Preprocessing Data</i>	32
4.4 Model PCA-SVM	33
4.4.1 Reduksi Dimensi Dengan PCA	33
4.4.2 Evaluasi Model PCA-SVM	35
4.5 Model SFS-SVM	36
4.5.1 Seleksi Fitur Dengan SFS.....	36
4.5.2 Evaluasi Model SFS-SVM	38
4.6 Perbandingan Model PCA-SVM Dengan SFS-SVM.....	39
4.7 Hasil Perancangan Website.....	43
4.7.1 Klasifikasi Individu	44
4.7.2 Klasifikasi <i>Dataset</i>	46
BAB V	50
KESIMPULAN DAN SARAN	50
5.1 Kesimpulan	50
5.2 Saran.....	50

DAFTAR PUSTAKA	51
LAMPIRAN.....	54

DAFTAR TABEL

Tabel 2.1 Tabel <i>Confusion Matrix</i>	12
Tabel 3.1 Bentuk <i>Dataset</i>	15
Tabel 4.1 Hasil Pembagian <i>Dataset</i>	32
Tabel 4.2 Contoh Data Latih Dari Hasil <i>Scaling</i>	33
Tabel 4.3 Contoh Data Uji Dari Hasil <i>Scaling</i>	33
Tabel 4.4 Contoh Data Latih Hasil Reduksi Dengan PCA	34
Tabel 4.5 Contoh Data Uji Hasil Reduksi Dengan PCA	34
Tabel 4.6 Performa Model PCA-SVM Dengan Jumlah PC Yang Berbeda	35
Tabel 4.7 Contoh Data Latih Hasil Seleksi Fitur	36
Tabel 4.8 Contoh Data Uji Hasil Seleksi Fitur	37
Tabel 4.9 Performa Model SFS-SVM Dengan Jumlah Fitur Yang Berbeda	38

DAFTAR GAMBAR

Gambar 2.1 Skema <i>Artificial Intelligence</i> dan <i>Machine Learning</i>	7
Gambar 2.2 <i>Hyperplane</i> yang memisahkan dua <i>class</i>	9
Gambar 3.1 Alur penelitian	14
Gambar 3.2 Alur Reduksi Dimensi PCA.....	18
Gambar 3.3 Alur Seleksi Fitur SFS	19
Gambar 3.4 Alur Pelatihan SVM.....	21
Gambar 3.5 Alur Klasifikasi Diabetes Secara Individu.....	23
Gambar 3.6 <i>Activity Diagram</i> Klasifikasi Diabetes Secara Individu	24
Gambar 3.7 <i>Design UI</i> Halaman Web Klasifikasi Diabetes Individu	25
Gambar 3.8 <i>Design UI</i> Hasil Klasifikasi Diabetes Individu.....	26
Gambar 3.9 Diagram Alir Klasifikasi Dengan <i>Dataset</i>	27
Gambar 3.10 <i>Activity Diagram</i> Klasifikasi Dengan <i>Dataset</i>	28
Gambar 3.11 <i>Design UI</i> Halaman Web Klasifikasi Diabetes dengan <i>Dataset</i>	29
Gambar 3.12 <i>Design UI</i> Hasil Klasifikasi Diabetes dengan <i>Dataset</i>	30
Gambar 4.1 Jumlah Nilai 0 Pada Setiap Fitur	31
Gambar 4.2 Pengujian Model PCA-SVM.	35
Gambar 4.3 Korelasi Antar Fitur dan Label.	37
Gambar 4.4 Grafik Perbandingan Performa Model Dengan 1 Fitur.....	39
Gambar 4.5 Grafik Perbandingan Performa Model Dengan 2 Fitur.....	40
Gambar 4.6 Grafik Perbandingan Performa Model Dengan 3 Fitur.....	40
Gambar 4.7 Grafik Perbandingan Performa Model Dengan 4 Fitur.....	41
Gambar 4.8 Grafik Perbandingan Performa Model Dengan 5 Fitur.....	41
Gambar 4.9 Grafik Perbandingan Performa Model Dengan 6 Fitur.....	42
Gambar 4.10 Grafik Perbandingan Performa Model Dengan 7 Fitur.....	42
Gambar 4.11 Grafik Perbandingan Performa Model Dengan 8 Fitur.....	43
Gambar 4.12 Halaman Website Klasifikasi Diabetes Pada Individu	44
Gambar 4.13 Contoh <i>error</i> dari website klasifikasi individu	45
Gambar 4.14 Halaman Website Klasifikasi Diabetes Pada Individu	45
Gambar 4.15 Halaman Website Hasil Klasifikasi Diabetes Pada Individu	46
Gambar 4.16 Contoh Hasil <i>expand</i> dari <i>dataframe</i>	46
Gambar 4.17 Tampilan Menu Klasifikasi Dengan <i>Dataset</i>	47
Gambar 4.18 Tampilan <i>Error</i> Unggah <i>Dataset</i> Dengan Nilai <i>Null</i>	47

Gambar 4.19 Tampilan <i>Error</i> Unggah <i>Dataset</i> Tidak Sesuai Kriteria.....	48
Gambar 4.20 Tampilan Website Usai Berhasil Mengunggah <i>Dataset</i>	48
Gambar 4.21 Tampilan Website Setelah Melakukan Klasifikasi	49
Gambar 4.22 File Csv Hasil Klasifikasi Diabetes	49

ABSTRAK

Diabetes merupakan salah satu penyakit kronis dan dapat menyebabkan berbagai komplikasi jangka panjang jika tidak ditangani dengan baik, sehingga untuk mencegah hal tersebut diperlukan sebuah model machine learning yang dapat memberikan prediksi terhadap diabetes dengan akurasi yang tinggi. Tujuan dari penelitian ini adalah melihat pengaruh pengurangan jumlah dimensi dari fitur terhadap performa model, melihat perbandingan model yang menggunakan teknik *sequential forward selection* (SFS) dan *principal component analysis* (PCA) dengan menggunakan algoritma *support vector machine* (SVM), dan membuat website yang dapat digunakan untuk prediksi diabetes. Pada penelitian ini menggunakan data *Pima Indian Dataset*, dataset akan melalui tahap *data preprocessing* lalu akan dibuat dua buah model yaitu PCA-SVM dan SFS-SVM untuk dilihat pengaruh pengurangan jumlah dimensi dari fitur terhadap performa model. Perancangan website dilakukan untuk mengembangkan website prediksi diabetes dengan menggunakan model terbaik. Hasil penelitian menunjukkan bahwa pengurangan jumlah dimensi pada fitur menggunakan SFS dan PCA meningkatkan performa model. Didapatkan metode SFS-SVM dengan performa terbaik pada fitur hasil seleksi berjumlah 6, sedangkan dari metode PCA-SVM pada *principal component* berjumlah 4. Model SFS-SVM lebih baik dalam mengklasifikasikan individu yang tidak terkena diabetes, sedangkan model PCA-SVM lebih baik dalam mengklasifikasikan individu yang terkena diabetes.

Kata kunci: *Principal Component Analysis, Support Vector Machine, Sequential Forward Selection, Diabetes, Machine learning.*

ABSTRACT

Diabetes is a chronic disease and can cause various long-term complications if not handled properly, so to prevent this, a machine learning model is needed that can predict diabetes with high accuracy. The purpose of this study is to see the effect of reducing the number of dimensions of features on model performance, to see a comparison of models using sequential forward selection (SFS) techniques and principal component analysis (PCA) using support vector machine (SVM) algorithms, and to create a website that can used for diabetes prediction. In this study using the Pima Indian Dataset, the dataset will go through the data preprocessing stage and then two models will be made, namely PCA-SVM and SFS-SVM to see the effect of reducing the number of dimensions of features on model performance. Website design is carried out to develop a diabetes prediction website using the best model. The results show that reducing the number of dimensions in features using SFS and PCA improves model performance. It was found that the SFS-SVM method with the best performance on the feature selection results was 6, while the PCA-SVM method on the principal component was 4. The SFS-SVM model is better at classifying individuals who are not affected by diabetes, while the PCA-SVM model is better at classifying individuals affected by diabetes.

Keywords: *Principal Component Analysis, Support Vector Machine, Sequential Forward Selection, Diabetes, Machine learning.*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes merupakan salah satu penyakit kronis yang memberikan dampak signifikan terhadap kesehatan dan kualitas hidup individu. Komplikasi jangka panjang biasanya berkembang secara bertahap saat diabetes tidak ditangani dengan baik, beberapa diantaranya adalah gangguan pada mata, kerusakan ginjal dan saraf, hingga penyakit kardiovaskular. Untuk mencegah hal tersebut diperlukan sebuah model *machine learning* yang dapat mengklasifikasikan diabetes dengan akurasi yang tinggi.

Penelitian pernah dilakukan oleh Smith et al. (1988) untuk memprediksi seseorang akan menderita diabetes atau tidak dalam 5 tahun kedepan dengan menggunakan *ADAP Learning* dan dihasilkan sebuah model yang memiliki performa yang cukup baik. Penelitian tersebut menggunakan *Pima Indian Dataset* dan belum mengimplementasikan teknik reduksi dimensi. *Dataset* dengan dimensi yang tinggi dapat mengakibatkan dampak yang buruk bagi model yang dibuat, seperti kompleksitas dari model menjadi tinggi, proses training yang lama, dan juga dapat menyebabkan *overfitting*.

Principal Component Analysis (PCA) adalah salah satu teknik reduksi dimensi yang bekerja dengan cara menstransformasi linear fitur dari suatu data sehingga terbentuk sistem koordinat baru dengan mempertahankan variansi maksimum. *Support Vector Machine* (SVM) merupakan salah satu algoritma *machine learning* yang bekerja dengan cara mencari *hyperplane* (garis pemisah) terbaik yang dapat memisahkan kelas dari data. Penelitian tentang kombinasi PCA-SVM pernah dilakukan oleh Godara dan Aron (2021) untuk mendeteksi sarkasme pada twitter dan didapatkan hasil bahwa model yang menggunakan kombinasi PCA-SVM memiliki performa yang lebih baik daripada model yang hanya menggunakan SVM.

Sequential Forward Selection (SFS) adalah salah satu algoritma pencarian, algoritma ini merupakan algoritma yang sederhana dalam konteks pengurangan dimensi. Penelitian tentang kombinasi SFS-SVM juga pernah dilakukan oleh Kabir et al. (2017) untuk mengklasifikasikan diabetes dan didapatkan hasil bahwa model kombinasi SFS dengan SVM lebih bagus daripada model yang mengkombinasikan SFS dengan algoritma *machine learning* lainnya, pada penelitian ini juga dihasilkan bahwa model SFS-SVM dengan data yang fiturnya direduksi menghasilkan performa lebih

baik daripada model dengan data yang fiturnya tidak direduksi (hanya menggunakan SVM).

Berdasarkan permasalahan yang telah dipaparkan diatas, maka pada penelitian ini akan dibandingkan kedua jenis kombinasi metode yaitu PCA-SVM dan SFS-SVM dalam performanya untuk mengklasifikasikan diabetes.

1.2 Rumusan Masalah

Untuk mendeteksi diabetes dibutuhkan sebuah model *machine learning* dengan akurasi yang tinggi. Untuk mendapatkan akurasi yang tinggi dari model dibutuhkan teknik ekstraksi fitur dan algoritma *machine learning* yang tepat. Berdasarkan penelitian yang dilakukan oleh Godara dan Aron (2021) tentang PCA-SVM dan penelitian yang dilakukan oleh Kabir et al. (2017) tentang SFS-SVM, menunjukkan bahwa teknik ekstraksi fitur dapat meningkatkan performa model *machine learning*. Penelitian ini menjawab pertanyaan, bagaimana pengaruh pengurangan dimensi terhadap performa dari model dalam mengklasifikasikan diabetes?, dan bagaimana perbandingan performa model klasifikasi diabetes PCA-SVM dan SFS-SVM?

1.3 Batasan Masalah

Penelitian ini memiliki batasan masalah untuk memberikan sebuah lingkup penelitian yang dilakukan, adapun batasan-batasan masalahnya adalah sebagai berikut:

1. *Dataset* yang digunakan pada penelitian kali ini adalah *dataset* Pima Indian yang diperoleh dari *National Institute of Diabetes and Digestive and Kidney Disease*. *Dataset* ini diperoleh melalui situs *Kaggle*. *Dataset* yang diperoleh hanya memiliki 2 label, yaitu 0 (tidak diabetes) dan 1 (positif diabetes).
2. Penelitian ini berfokus pada klasifikasi diabetes berdasarkan fitur-fitur yang terdapat pada *dataset*. Variabel terikat pada *dataset* akan menjadi target.
3. Penelitian ini akan menggunakan metode PCA untuk reduksi dimensi data dan SVM sebagai algoritma yang digunakan, selanjutnya akan dibandingkan dengan teknik seleksi fitur SFS yang juga akan dikombinasikan dengan SVM.
4. Evaluasi performa model akan dilakukan menggunakan metrik akurasi, presisi, recall, *specificity* dan F1-score.
5. Bahasa pemrograman yang digunakan pada penelitian kali ini adalah *python*.
6. Kode editor yang digunakan pada penelitian kali ini adalah *Visual Studio Code*.

1.4 Tujuan Penelitian

1. Untuk melihat pengaruh pengurangan dimensi dari fitur terhadap performa model klasifikasi diabetes.
2. Untuk membandingkan performa model klasifikasi diabetes yang menggunakan kombinasi PCA-SVM dengan model yang menggunakan kombinasi SFS-SVM.
3. Untuk membuat website yang dapat digunakan untuk melakukan klasifikasi terhadap penyakit diabetes.

1.5 Manfaat Penelitian

1. Menunjukkan pengaruh pengurangan dimensi dari fitur terhadap performa model klasifikasi diabetes.
2. Menunjukkan perbandingan PCA-SVM dan SFS-SVM dalam melakukan klasifikasi penyakit diabetes.
3. Membuat website yang dapat digunakan untuk melakukan klasifikasi penyakit diabetes.

1.6 Sistematika Penulisan Laporan

Adapun sistematika penulisan dari Laporan Tugas Akhir ini terbagi menjadi beberapa bab, sesuai dengan ketentuan yang ditetapkan di Fakultas Teknik, Universitas Mataram adalah sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini berisi gambaran umum dan penjelasan mengenai latar belakang pemilihan judul penelitian, perumusan masalah yang dimana merupakan tulisan singkat berisi pertanyaan tentang topik penelitian yang nantinya akan dijawab oleh penulis sehingga penelitian yang dilakukan memiliki suatu kesimpulan dari hasil analisis dan visualisasi data penelitian yang dilakukan, selain itu terdapat pula batasan masalah yang merupakan batas-batas dari topik penelitian yang sedang dikaji atau diteliti, serta terdapat juga tujuan penelitian dan manfaat penelitian yang merupakan keuntungan yang didapat atau diperoleh oleh berbagai pihak dari penelitian yang dilakukan.

BAB II TINJAUAN PUSTAKA DAN DASAR TEORI

Pada bab ini berisi mengenai landasan teori dan tinjauan pustaka, yang dimana dalam tinjauan pustaka mengulas beberapa penelitian sebelumnya yang sejenis mengenai klasifikasi penyakit diabetes dan *machine learning*, sedangkan pada dasar

teori membahas mengenai teori-teori yang berkaitan dengan topik penelitian yang didapatkan melalui beberapa sumber-sumber seperti jurnal, buku, dsb.

BAB III METODE PENELITIAN

Pada bab ini menjelaskan tentang langkah-langkah dalam pelaksanaan penelitian yang dimana terdiri dari alur penelitian yang membahas mengenai gambaran umum dalam bentuk diagram terkait alur penelitian yang dilakukan dan uraian metodologi yang membahas tahapan yang dilakukan pada alur penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memuat hasil dan pembahasan dari penelitian yang telah dilakukan berupa tahapan dalam membuat model seperti *pre-processing* data, reduksi dimensi menggunakan PCA, seleksi fitur menggunakan SFS, pengimplementasian algoritma SVM, perbandingan kombinasi metode PCA-SVM dan SFS-SVM dalam melakukan klasifikasi diabetes dan proses pengembangan perangkat lunak.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan yang didapatkan dari keseluruhan proses penelitian dan saran dari penulis dengan harapan dapat bermanfaat untuk penelitian selanjutnya pada topik terkait.

BAB II TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Penelitian terdahulu adalah penelitian-penelitian yang telah dilakukan sebelumnya yang kemudian dijadikan bahan acuan dalam penelitian penulis. Berikut adalah penelitian-penelitian yang relevan dengan penelitian penulis:

Mohan dan Jain (2020) melakukan penelitian tentang algoritma *Support Vector Machine* (SVM) untuk memprediksi diabetes. Penelitian yang berjudul “*Performance Analysis of Support Vector Machine in Diabetes Prediction*” tersebut menggunakan *dataset Pima Indian Diabetes* yang diperoleh melalui *kaggle*. *Dataset* pertama-tama di-import lalu dibagi menjadi data latih dan data uji, setelah itu dilakukan pelatihan dan pengujian pada model SVM. Pada penelitian tersebut membandingkan performa setiap kernel dalam SVM yaitu *linear*, *polynomial*, *RBF*, dan *Sigmoid*, adapun akurasi tertinggi didapatkan oleh model SVM dengan kernel *RBF*. Adapun akurasi dari setiap model yang dibuat dari penelitian tersebut yaitu model SVM dengan kernel *linear* diperoleh akurasi 77%, model SVM dengan kernel *polynomial* diperoleh akurasi 80%, model SVM dengan kernel *RBF* diperoleh akurasi 82%, dan model SVM dengan kernel *sigmoid* diperoleh akurasi 69%. Pada penelitian tersebut hanya berfokus pada nilai akurasi, tidak mengukur nilai presisi, *recall*, *specificity*, dan *F1-Score*.

Godara dan Aron (2021) melakukan penelitian tentang kombinasi teknik reduksi dimensi yaitu *Principal Component Analysis* (PCA) dengan *Support Vector Machine* (SVM) dan dibandingkan dengan beberapa algoritma yang tidak dikombinasikan dengan PCA yaitu *Naïve Bayes*, *K-Nearest Neighbor* dan *Support Vector Machine*. Penelitian yang berjudul “*Support Vector Machine Classifier with Principal Component Analysis and K Mean for Sarcasm Detection*” tersebut membahas tentang pendeteksi sarkasme pada twitter dan *dataset* pada penelitian tersebut dikumpulkan dengan menggunakan *Tweepy API*. Setelah *dataset* dikumpulkan lalu dilakukan tahap *preprocessing* dengan cara menghapus nilai yang hilang dan berlebihan (*redundant*). Setelah melalui tahap *preprocessing* dilakukan ekstraksi fitur dan reduksi fitur yang menggunakan PCA sehingga dimensi dari *dataset* menjadi lebih kecil. Tahap akhir pada penelitian tersebut dilakukan klasifikasi menggunakan beberapa algoritma yaitu *Naive Bayes*, *K-Nearest*

Neighbor, *Support Vector Machine*, *Principal Component Analysis* dan *Support Vector Machine*, *Principal Component Analysis* dengan *Support Vector Machine* dan *K-means*. Dari beberapa algoritma yang digunakan tersebut, kombinasi dari *K-mean*, PCA dan SVM menghasilkan performa yang paling tinggi dengan parameter evaluasi *accuracy*, *recall*, dan *precision*.

Kabir et al. (2017) melakukan penelitian tentang klasifikasi diabetes menggunakan beberapa algoritma *machine learning* yaitu *Support Vector Machine*, *C4.5* dan *K-Nearest Neighbor*. Penelitian yang berjudul “*Developing Diabetes Disease Classification Model Using Sequential Forward Selection Algorithm*” tersebut menggunakan *dataset Pima Indian Diabetes* yang diperoleh melalui *kaggle*. Penelitian ini membandingkan performa algoritma-algoritma tersebut pada saat fitur asli digunakan dan pada saat setelah fitur diseleksi menggunakan *Sequential Forward Selection* (SFS). Hasil dari penelitian tersebut adalah performa model yang menggunakan teknik seleksi fitur menghasilkan performa yang jauh lebih tinggi dibandingkan model yang tidak menggunakan teknik seleksi fitur. Adapun dari beberapa algoritma yang digunakan, didapatkan performa terbaik oleh algoritma *Support Vector Machine* (SVM) dengan parameter evaluasi *accuracy*, *specificity*, *sensitivity*.

Berdasarkan penelitian-penelitian yang telah dilakukan diatas, penulis mengangkat topik penelitian tentang “Perbandingan Metode *Principal Component Analysis* (PCA) dan *Sequential Forward Selection* (SFS) Dengan *Support Vector Machine* (SVM) Dalam Klasifikasi Diabetes”. Penulis akan membandingkan dua kombinasi metode yaitu PCA-SVM dan SFS-SVM dalam performanya untuk klasifikasi diabetes. Adapun nilai yang akan dilihat adalah akurasi, presisi, *recall*, *f1-score*, dan *specificity*. Setelah dilakukan perbandingan, penulis akan menggunakan model terbaik untuk diintegrasikan ke aplikasi berbasis website sehingga dapat digunakan untuk melakukan klasifikasi diabetes.

2.2 Dasar Teori

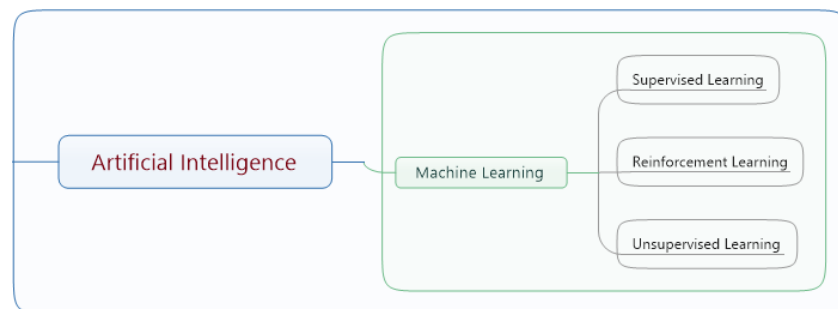
2.2.1 Diabetes

Diabetes melitus atau dapat disebut juga kencing manis adalah penyakit yang berupa kumpulan gejala yang timbul pada seseorang yang disebabkan oleh peningkatan kadar gula darah karena kekurangan insulin ataupun resistensi insulin. Dalam

perjalanannya, diabetes dapat menyebabkan sejumlah komplikasi, baik yang bersifat akut maupun kronis, jika tidak dikelola dengan baik. Diabetes merupakan salah satu penyakit kronis yang tidak dapat sembuh sepenuhnya, tetapi dapat diatur atau dikelola. Ini berarti bahwa setelah seseorang didiagnosis dengan diabetes, mereka akan menghadapinya sepanjang hidup (Toharin et al., 2015).

2.2.2 Machine Learning

Machine learning adalah aplikasi komputer dan algoritma matematika yang bekerja dengan cara mempelajari pola dari data dan menghasilkan prediksi untuk data yang belum diketahui (masa depan). Adapun proses pembelajaran dari *machine learning* terdiri dari dua tahap, yaitu proses pelatihan (*training*) dan pengujian (*testing*), yang bertujuan untuk mendapatkan kecerdasan buatan.



Gambar 2.1 Skema *Artificial Intelligence* dan *Machine Learning*.

Supervised Learning menggunakan metode klasifikasi di mana seluruh dataset memiliki label yang telah diberikan, untuk tujuan mengklasifikasikan kelas yang tidak memiliki label. *Unsupervised Learning*, yang juga disebut *clustering*, tidak memerlukan label dalam dataset dan berfokus pada mengidentifikasi pola atau kelompok dalam data yang tidak terdefinisi sebelumnya. Sedangkan *Reinforcement Learning* berada di antara kedua pendekatan tersebut, teknik ini bekerja dalam lingkungan yang dinamis dan mempunyai tujuan untuk mencapai hasil tertentu tanpa menerima panduan eksplisit dari komputer mengenai apakah tujuan tersebut telah tercapai.

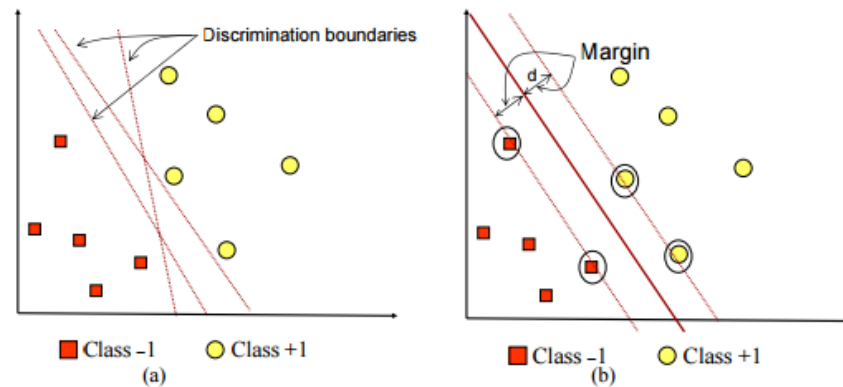
Metode *supervised learning* mendasarkan pada *dataset* yang dilengkapi dengan label. *Dataset* ini digunakan untuk mengidentifikasi karakteristik distribusi perilaku atau fitur dalam berbagai jenis aplikasi, membentuk model yang merepresentasikan perilaku data. *Supervised learning* kemudian dibagi menjadi dua masalah utama, yaitu klasifikasi dan regresi. Klasifikasi melibatkan variabel *output* dalam bentuk kategori, seperti jenis warna atau status kesehatan (sakit atau sehat). Sementara itu, masalah

regresi melibatkan variabel *output* berupa nilai riil, seperti nilai mata uang atau berat suatu objek. *Supervised learning* memiliki beberapa algoritma populer seperti *Back-propagation*, *Linear regression*, *Random Forest*, *Support Vector Machines*, *Naive Bayesian*, *Metode Rocchio*, *Decision Tree*, *k-Nearest Neighbor*, *Neural Network*, *Logistic Regression*, dan *Neural Network*. Kemudian beberapa algoritma untuk klasifikasi pun disebutkan dalam seperti *Support Vector Machines (SVM)*, *Normal Bayesian Classifier (NBC)*, *K-Nearest Neighbor (KNN)*, *Trees Gradient Boosted (GBT)*, *Random Trees (RT)*, dan *Artificial Neural Networks (ANN)* (Roihan et al., 2020).

2.2.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu algoritma *machine learning* yang bekerja dengan prinsip *Structural Risk Minimization (SRM)* yang bertujuan untuk mencari *hyperplane* yang digunakan untuk memisahkan dua buah kelas data pada *input space*. Metode ini memanfaatkan hipotesis yang berbentuk fungsi linier dalam dimensi yang tinggi, diimplementasikan dengan memperoleh pengetahuan dari teori pembelajaran statistik. Tingkat akurasi model yang dihasilkan melalui proses SVM sangat dipengaruhi oleh fungsi kernel dan parameter yang digunakan dalam proses perhitungan.

Pada Gambar 2.2 terlihat sejumlah pola data yang termasuk dalam dua kelas: +1 dan -1. Pola yang termasuk dalam kelas -1 ditunjukkan dengan simbol kotak berwarna merah, sedangkan pola yang termasuk dalam kelas +1 ditunjukkan dengan simbol lingkaran berwarna kuning. Tantangan dalam masalah klasifikasi ini adalah menemukan garis *hyperplane* yang dapat memisahkan kedua kelompok ini. Garis solid pada Gambar 2.2 menunjukkan *hyperplane* terbaik, yaitu yang berada di tengah-tengah kedua kelas. Titik-titik merah dan kuning yang berada dalam lingkaran hitam merupakan vektor pendukung (*support vector*) (Monika dan Furqon, 2018).



Gambar 2.2 *Hyperplane* yang memisahkan dua *class*.

2.2.4 *Principal Component Analysis (PCA)*

Principal Component Analysis (PCA) dapat disebut juga transformasi *KarhunenLoeve* merupakan metode yang digunakan untuk mengurangi dimensi dari suatu data, PCA bekerja dengan cara tranformasi linear fitur dari suatu data sehingga terbentuk sistem koordinat baru dengan mempertahankan variansi maksimum. PCA dapat digunakan untuk mengurangi dimensi suatu data tanpa mengurangi karakteristik dari data tersebut secara signifikan. Metode ini mengubah variabel asli yang saling berkorelasi menjadi beberapa variabel baru yang lebih kecil dan tidak berkorelasi lagi. Prinsip dasar dari algoritma PCA adalah mengurangi dimensi dari sekumpulan data tetapi tetap mempertahankan sebanyak mungkin variasi yang terdapat dalam data tersebut. Secara matematis, PCA mengubah variabel yang saling berhubungan menjadi bentuk yang independen atau tidak memiliki korelasi.

Principal Component adalah hasil dari transformasi linier proyeksi dari data variabel. *Principal Component* ini tidak saling berkorelasi satu sama lain dan diurutkan sedemikian rupa sehingga *Principal Component* pertama mengandung sebanyak mungkin variasi dari *dataset*. *Principal Component* kedua mengandung variasi yang tidak terdapat dalam *Principal Component* pertama, dan begitu seterusnya (Firliana, 2015).

2.2.5 *Sequential Forward Selection*

Metode *Sequential Forward Selection (SFS)* atau seleksi maju adalah salah satu algoritma pencarian, algoritma ini merupakan algoritma yang sederhana dalam konteks pengurangan dimensi. SFS didasarkan pada model regresi linear. Teknik ini bertujuan

untuk mengurangi dimensi dari *dataset* dengan menghapus atribut-atribut yang tidak relevan. Dalam metode SFS, proses pemodelan dimulai dari model kosong, lalu satu per satu atribut dimasukkan ke dalam model hingga kriteria tertentu terpenuhi (Nugroho dan Wibowo, 2017).

2.2.6 Performance Evaluation Measure

Performance Evaluation Measure (PEM) atau pengukuran evaluasi performa adalah suatu kumpulan nilai yang digunakan untuk mengukur performa suatu sistem. PEM dalam banyak kasus digunakan dalam data latih, yang bertujuan untuk mengevaluasi model yang sudah dibuat. Terdapat berbagai perhitungan yang digunakan untuk memperoleh nilai PEM, yang umumnya diimplementasikan sebagai kombinasi atau dalam beberapa kasus secara parsial (Imron, 2019). Beberapa perhitungan dalam PEM antara lain:

- *Precision.*

Precision adalah ukuran tingkat keakuratan antara permintaan yang diajukan oleh pengguna dan jawaban yang diberikan oleh sistem.

Rumus *precision*

$$\frac{TP}{FP+TP} \dots \dots \dots (1)$$

Keterangan:

-TP = *True Positive*

-FP = *False Positive*

- *Accuration.*

Accuration adalah ukuran perbandingan antara informasi yang benar dijawab oleh sistem dengan keseluruhan dari informasi.

Rumus *accuration*

$$\frac{TN+TP}{FN+FP+TN+TP} \dots \dots \dots (2)$$

Keterangan:

-TN = *True Negative*

-TP = *True Positive*

-FN = *False Negative*

-FP = *False Positive*

- *Recall.*

Recall adalah ukuran tingkat ketepatan antara informasi yang relevan yang telah dipanggil sebelumnya dengan informasi yang sama yang sebenarnya ada.

Rumus *recall*

$$\frac{TP}{FN+TP} \dots\dots\dots (3)$$

Keterangan:

-TP = *True Positive*

-FN = *False Negative*

- *F1-Score.*

F-1 Score adalah ukuran perbandingan rata-rata *precision* dan *recall* yang dibobotkan (Sihombing dan Arsani, 2021).

Rumus *F1-Score*

$$2 \frac{x \text{ Precision } x \text{ Recall}}{\text{Precision} + \text{Recall}} \dots\dots\dots (5)$$

- *Specifiticy.*

Specificity mengukur berapa banyak klasifikasi yang benar negatif dibandingkan dengan total *instance* yang sebenarnya negatif (Astuti et al., 2022).

Rumus *specificity*

$$\frac{TN}{FP+TN} \dots\dots\dots (5)$$

Keterangan:

-TN = *True Negative*

-FP = *False Positive*

2.2.7 *Confusion Matrix*

Confusion matrix merupakan suatu metode yang umum digunakan dalam analisis *data mining* untuk menghitung akurasi. Matriks ini direpresentasikan dalam bentuk tabel yang mengindikasikan jumlah data uji yang berhasil diprediksikan dengan benar dan yang salah. *Confusion matrix* dapat dilihat pada Tabel 2.1.

Tabel 2.1 Tabel *Confusion Matrix*

<i>Correct Classification</i>	<i>Classified as</i>	
	<i>Actual “+”</i>	<i>Actual “-”</i>
<i>Predicted “+”</i>	TP (<i>True Positive</i>)	FP (<i>False Positive</i>)
<i>Predicted “-”</i>	FN (<i>False Negative</i>)	TN (<i>True Negative</i>)

Penjelasan tabel *Confusion Matrix* diatas, yaitu:

1. *True Positive* (TP) merupakan jumlah data yang memiliki nilai positif sebenarnya dan juga diprediksi positif.
2. *False Positive* (FP) menggambarkan jumlah data yang pada kenyataannya memiliki nilai negatif, namun diprediksi sebagai positif.
3. *False Negative* (FN) mencakup jumlah data yang memiliki nilai positif sebenarnya, tetapi diprediksi sebagai negatif.
4. *True Negative* (TN) merupakan jumlah data yang mempunyai nilai negatif sebenarnya dan juga diprediksi sebagai negatif (Lufiana, 2021).

2.2.8 Python

Python adalah bahasa pemrograman yang disediakan secara bebas atau *freeware*, yang berarti Anda dapat menyalin dan mendistribusikannya tanpa batasan. Ini termasuk *source code*, alat *debugging*, dan *profiler* yang berguna untuk analisis kinerja. *Python* juga menyediakan antarmuka yang luas, termasuk antarmuka sistem, GUI (*Graphical User Interface*), dan basis data. Bahasa pemrograman ini dapat dijalankan di berbagai sistem operasi, seperti *UNIX*, *PCs* (termasuk *DOS*, *Windows*, *OS/2*), *Macintosh*, dan lainnya. Di banyak distribusi sistem operasi Linux, *Python* menjadi bagian standar yang disertakan. *Python* didistribusikan dengan lisensi *Open Source* yang diakui oleh *Open Source Initiatives* (OSI). Ini berarti bahwa *Python* dapat digunakan secara bebas dan gratis, bahkan dalam pengembangan produk komersial. Organisasi *Python Software Foundation* (PSF) bertanggung jawab atas perlindungan hak kekayaan intelektual *Python*. Mereka juga mengawasi konferensi *PyCon* dan mendanai berbagai proyek di dalam komunitas *Python* (Clinton dan Sengkey, 2019).

2.2.9 Streamlit

Streamlit adalah sebuah *framework* atau kerangka kerja web yang dirancang khusus untuk menghubungkan model dan visualisasi dengan mudah menggunakan bahasa pemrograman *Python*. Meskipun simpel dan ringkas, *Streamlit* memiliki

tampilan yang menarik dan ramah untuk pengguna. Di dalamnya sudah tersedia berbagai widget bawaan untuk menerima input dari pengguna, seperti pengunggahan gambar, penggeser, *input* teks, serta elemen HTML yang umum seperti kotak centang dan tombol radio. Saat pengguna berinteraksi dengan aplikasi *Streamlit*, skrip *Python* akan dijalankan ulang dari atas ke bawah. Konsep ini perlu diperhatikan ketika mengelola status dan interaksi dalam aplikasi (Widi Hastomo et al., 2022).

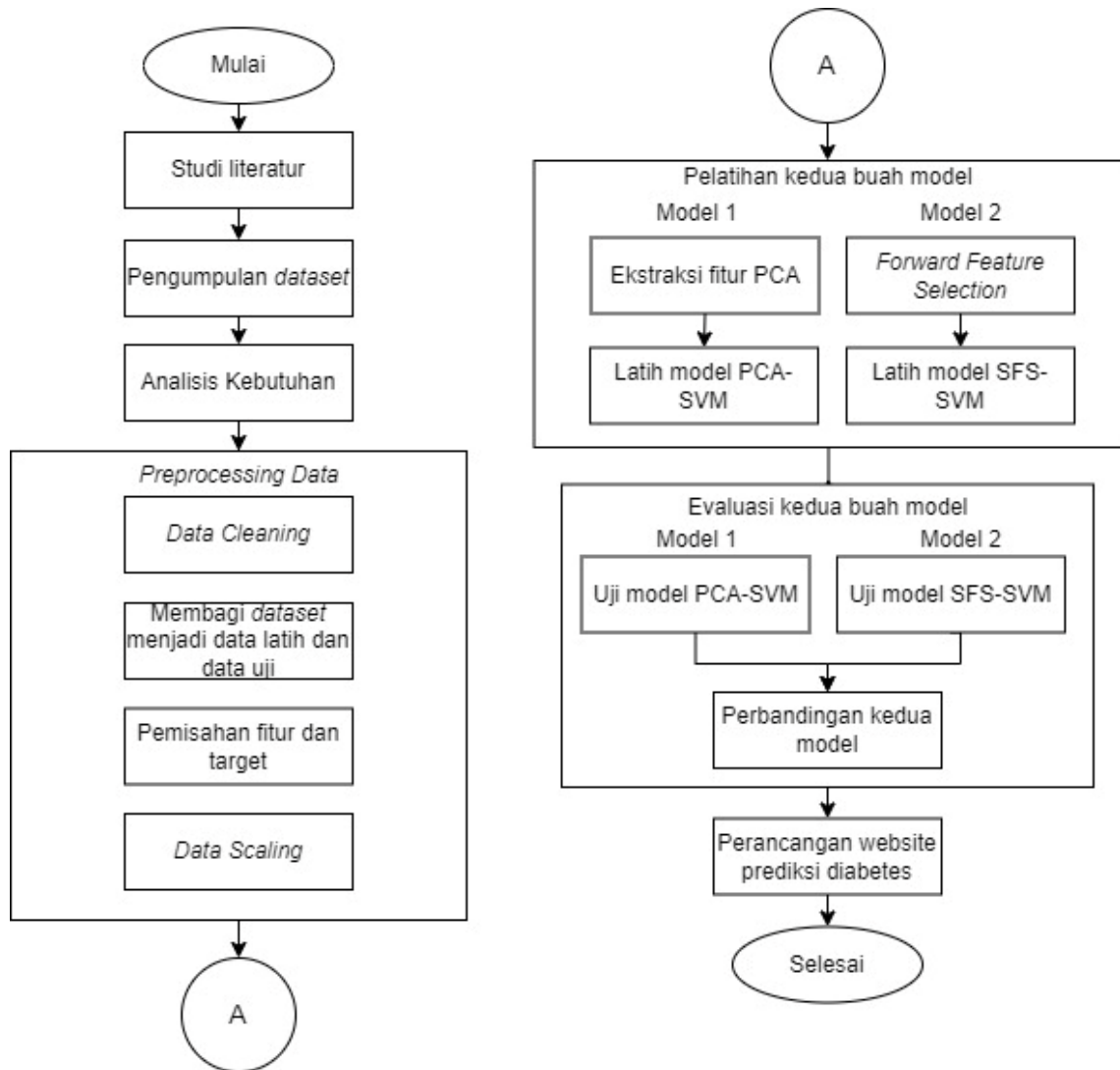
2.2.10 Visual Studio Code

Visual Studio Code adalah sebuah perangkat lunak editor kode sumber yang dirancang oleh *Microsoft* untuk digunakan pada sistem operasi *Windows*, *Linux*, dan *macOS*. Perangkat ini memiliki sejumlah fitur termasuk kemampuan untuk melakukan debugging, integrasi dengan *Git* dan *GitHub*, penyorotan sintaksis, alat bantu penyelesaian kode yang cerdas, potongan kode (*snippet*), serta kemampuan untuk melakukan perbaikan kode (*refactoring*). Editor ini memiliki tingkat kustomisasi yang tinggi, memungkinkan pengguna untuk mengatur tema tampilan, pintasan keyboard, preferensi, dan juga memasang ekstensi untuk menambahkan fungsi-fungsi tambahan (Agustini dan Kurniawan, 2019).

BAB III METODOLOGI PENELITIAN

3.1 Alur Penelitian

Penelitian ini dilakukan dari awal hingga akhir sesuai dengan alur penelitian yang digambarkan dalam Gambar 3.1. Alur penelitian digunakan oleh penulis dalam pelaksanaan penelitian ini agar hasil yang dicapai tidak menyimpang dari tujuan yang telah ditetapkan sebelumnya. Alur penelitian dapat dilihat dalam Gambar 3.1 berikut.



Gambar 3.1 Alur penelitian

3.2 Studi Literatur

Studi literatur tentang “Perbandingan Metode *Principal Component Analysis* (PCA) dan *Sequential Forward Selection* (SFS) Dengan *Support Vector Machine* (SVM) Dalam Klasifikasi Diabetes” akan melibatkan pencarian, seleksi, dan penelaahan

tentang literatur terkait dengan metode tersebut serta aplikasinya dalam klasifikasi diabetes. Studi literatur yang dilakukan didapatkan melalui E-Book, Jurnal dan lain-lain.

3.3 Pengumpulan *Dataset*

Pada penelitian ini *dataset* yang digunakan memiliki 8 fitur atau variabel bebas dan 1 variabel terikat (label/target) yang memiliki 768 baris, dengan 268 sampel didiagnosis penyakit diabetes dan 500 sampel didiagnosis sehat atau tidak diabetes. Adapun *dataset* pada penelitian ini didapatkan melalui *kaggle* dan bisa diakses melalui link berikut: <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>. Dalam *dataset* ini, sampel diambil dari populasi perempuan *Pima Indian* yang berada di dekat Phoenix, Arizona. Populasi tersebut telah diteliti secara terus menerus sejak tahun 1965 oleh *National Institute of Diabetes and Digestive and Kidney Diseases* karena tingginya angka kejadian diabetes (Mucholladin et al., 2021). Adapun bentuk dari *dataset* tersebut adalah sebagai berikut.

Tabel 3.1 Bentuk *Dataset*

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age	Out
6.00	148.00	72.00	35.00	0.00	33.60	0.63	50.00	1.00
1.00	85.00	66.00	29.00	0.00	26.60	0.35	31.00	0.00
8.00	183.00	64.00	0.00	0.00	23.30	0.67	32.00	1.00
1.00	89.00	66.00	23.00	94.00	28.10	0.17	21.00	0.00
0.00	137.00	40.00	35.00	168.00	43.10	2.29	33.00	1.00
5.00	116.00	74.00	0.00	0.00	25.60	0.20	30.00	0.00
3.00	78.00	50.00	32.00	88.00	31.00	0.25	26.00	1.00
10.00	115.00	0.00	0.00	0.00	35.30	0.13	29.00	0.00
2.00	197.00	70.00	45.00	543.00	30.50	0.16	53.00	1.00
8.00	125.00	96.00	0.00	0.00	0.00	0.23	54.00	1.00

Adapun penjelasan dari masing-masing fitur dalam *dataset* tersebut adalah sebagai berikut:

- Pregnancies* (Pre) jumlah kehamilan dari individu.
- Glucose* (Glu) adalah konsentrasi glukosa plasma 2 jam dalam tes toleransi glukosa oral dari individu.
- BloodPressure* (BP) adalah tekanan darah diastolik (mm Hg) dari individu.
- SkinThickness* (ST) adalah ketebalan lipatan kulit trisep (mm) dari individu.

- Insulin* adalah insulin serum 2 jam (mu U/ml) dari individu.
- BMI* adalah indeks massa tubuh (berat dalam kg/(tinggi dalam m)²) dari individu.
- Diabetes pedigree function* (DBF) adalah indikator riwayat diabetes dalam keluarga dari individu.
- Age* adalah umur dari individu.
- Outcome* (Out) adalah target label yang merupakan label apakah individu memiliki penyakit diabetes (0=tidak, 1=ya).

3.4 Analisis Kebutuhan

Adapun berbagai kebutuhan yang mendukung untuk dilakukannya penelitian ini mencakup *hardware*, *software*, dan *library* dari *Python* yang dimana akan dirincikan sebagai berikut:

Hardware:

- 1 buah laptop

Software:

- Sistem operasi *windows* 10.
- *Microsoft Excel* untuk menampilkan *dataset* mentah.
- *Microsoft Word* untuk membuat laporan penelitian.
- Aplikasi *Draw.io* untuk membuat diagram
- Aplikasi *Pencil* untuk membuat desain *user interface*.
- *Visual Studio Code* untuk kode editor dalam melakukan perancangan model dan perancangan sistem menggunakan *streamlit*.
- *Python* sebagai bahasa pemrograman dalam *machine learning*.

Library Python:

- *Pandas* untuk memuat, menyimpan, mengubah dan menggabungkan data dalam bentuk file excel.
- *Numpy* untuk melakukan operasi matriks, vektor dan aljabar.
- *Matplotlib* membuat visualisasi data dalam bentuk *plot*.
- *Scikit-learn* untuk melakukan *scaling data*, pembagian *dataset*, membuat *confusion matrix* dan untuk melakukan seleksi fitur menggunakan *Sequential Forward Selection*.
- *Seaborn* untuk visualisasi dari korelasi data dalam bentuk *heatmap*.

- *Pickle* untuk menyimpan dan membaca model dalam bentuk file.
- *Sys* untuk menangani eksepsi pada website.
- *Datetime* untuk melakukan operasi tanggal dan waktu.
- *Streamlit* untuk pengembangan website *machine learning*.

3.5 Preprocessing Data

Untuk mendapatkan hasil yang maksimal dalam pembuatan model nantinya diperlukan *preprocessing* terhadap data agar cocok digunakan dalam algoritma yang digunakan. Adapun tahap *preprocessing* yang dilakukan adalah:

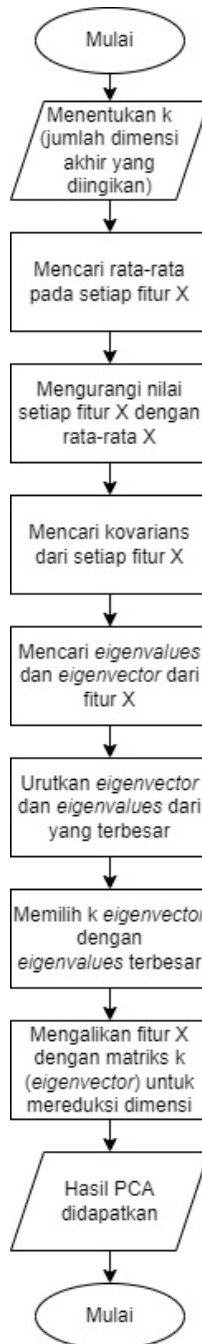
- *Data cleaning* dilakukan untuk menghilangkan data yang memiliki nilai *null* untuk menjadi *integer*. Untuk memastikan data yang digunakan memiliki kualitas yang baik.
- Pembagian *dataset* dilakukan dalam skala 80:20, 80% untuk data latih dan 20% untuk data uji. Data latih diperlukan untuk melatih model SVM nantinya agar model dapat mengenali pola dari data. Sedangkan data uji diperlukan untuk mengukur kinerja model dari data yang belum pernah dilihat oleh model.
- Pemisahan antara fitur dan target dilakukan agar target tidak ikut serta dalam proses *scaling* data.
- Melakukan *scaling* data untuk memastikan skala yang seragam antara fitur-fitur yang digunakan.

3.6 Pelatihan Model

Pada tahap ini, setelah melalui tahap *preprocessing* akan dibuat dua buah model *machine learning* dengan metode yang berbeda yaitu PCA-SVM dan SFS-SVM. PCA-SVM akan menggunakan metode reduksi dimensi PCA dan metode SFS-SVM akan menggunakan metode seleksi fitur SFS, lalu akan dilatih dengan algoritma SVM.

3.6.1 Ekstraksi Fitur PCA

Sebelum melalui tahap pelatihan PCA-SVM, PCA akan mentransformasikan fitur asli menjadi sejumlah komponen utama yang merupakan kombinasi linear dari variabel asli. Komponen utama ini diurutkan berdasarkan tingkat varian yang dijelaskan. Adapun alur dari proses ekstraksi fitur PCA adalah sebagai berikut.

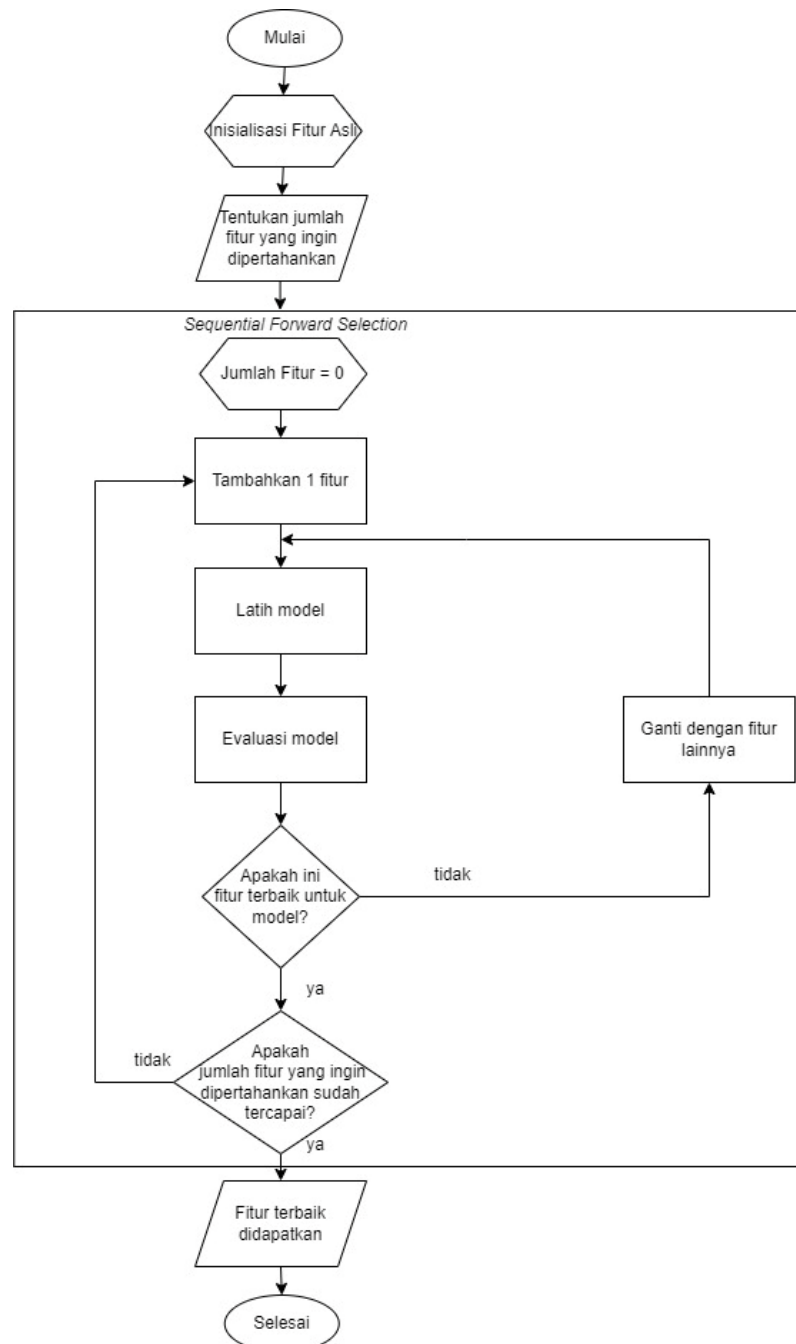


Gambar 3.2 Alur Reduksi Dimensi PCA

Gambar diatas merupakan langkah-langkah untuk melakukan ekstrasi fitur dengan menggunakan PCA. Pada PCA ditentukan jumlah komponen (*principal component*) akhir yang akan dipertahankan. Lalu masuk ke tahapan mencari rata-rata dari setiap fitur X, mencari kovarians dari setiap fitur X, mencari *eigenvalues* dan *eigenvector* fitur X, mengurutkan *eigenvector* dengan *eigenvalues* dari yang terbesar ke yang terkecil, lalu memilih k *eigenvector* teratas dari hasil pengurutan, dan yang terakhir adalah mengalikan fitur X dengan jumlah dari *eigenvector* yang dipilih.

3.6.2 Seleksi Fitur SFS

Sebelum melalui tahap pelatihan SFS-SVM, SFS akan menseleksi fitur untuk mengurangi dimensi dari fitur dengan memilih fitur yang memiliki bobot terbesar. *Sequential Forward Selection* akan melakukan kombinasi-kombinasi dari fitur untuk mendapatkan jumlah fitur yang menghasilkan performa model yang terbaik.



Gambar 3.3 Alur Seleksi Fitur SFS

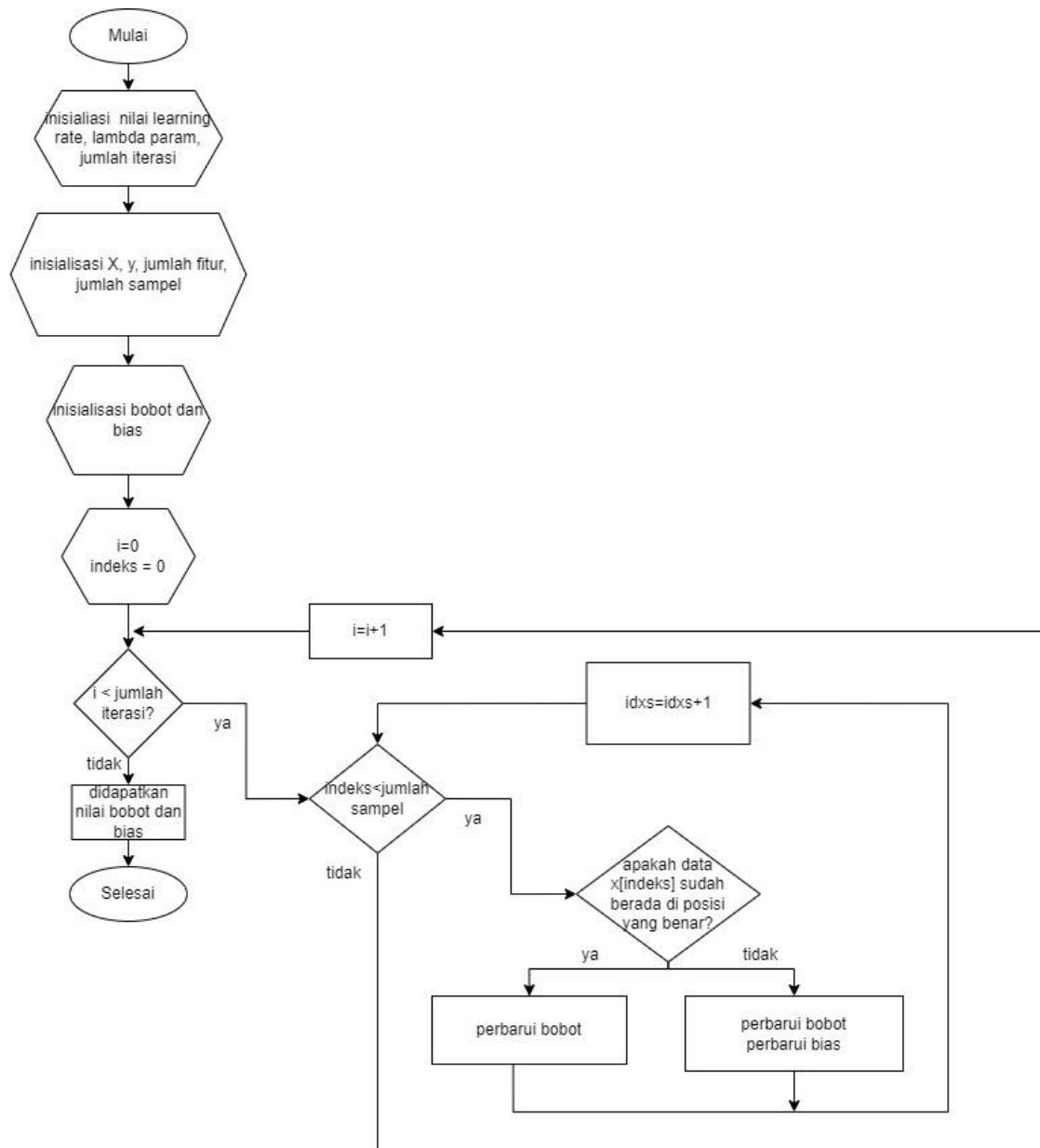
Gambar diatas merupakan alur dari proses seleksi fitur menggunakan *sequential forward selection* (SFS). Pertama-tama dari fitur asli yang digunakan

(*dataset*) ditentukan jumlah dari fitur yang ingin dipertahankan, Setelah menentukan jumlah fitur yang ingin dipertahankan, model akan memulai proses seleksi fitur.

Model akan menambahkan 1 fitur terlebih dahulu dan melakukan evaluasi sampai menemukan fitur yang terbaik (fitur dengan evaluasi model tertinggi). Setelah ditemukan 1 fitur, akan dilakukan pengecekan kondisi apakah sudah mencapai jumlah fitur yang diinginkan, jika belum maka akan ditambahkan 1 fitur lagi, lalu proses pencarian fitur terbaik dilakukan kembali. Jika sudah mencapai jumlah fitur yang ditentukan, artinya jumlah fitur yang ingin dipertahankan atau dipilih sudah tercapai. Setelah proses seleksi berakhir, didapatkan data dengan hasil seleksi fitur dengan fitur-fitur terbaik.

3.6.3 Latih Model SVM

Setelah data dibagi menjadi dua subset, subset data latih akan digunakan untuk melatih model SVM. Pada tahap ini, SVM akan mempelajari pola dari data pelatihan untuk membedakan antara pasien diabetes dan non-diabetes, terdapat dua buah model yang dilatih pada tahap ini yaitu model dengan metode PCA-SVM dan SFS-SVM yang sudah melalui tahap reduksi fitur.



Gambar 3.4 Alur Pelatihan SVM

Gambar diatas merupakan alur dari proses pelatihan SVM. Dari alur diatas dapat dilihat pada tahap awal, pendeklarasian nilai *learning rate*, *lambda param*, dan jumlah iterasi dilakukan untuk melakukan proses pelatihan SVM dalam menemukan garis pemisah antar kelas yang terbaik. Selanjutnya ada proses pendeklarasian fitur x, y, jumlah fitur dan jumlah sampel. Fitur x (data latih) yang dideklarasikan merupakan fitur yang berasal dari hasil proses PCA ataupun SFS. Sedangkan y adalah target dari sampel (diabetes atau tidak diabetes). Untuk jumlah fitur dan jumlah sampel diambil dari fitur yang sudah melalui tahap ekstraksi fitur PCA ataupun SFS. Selanjutnya memasuki proses inisialisasi bobot dan bias dilanjutkan dengan proses pencarian nilai bobot dan

bias yang tepat sehingga membentuk garis pemisah atau *decision bouondary* terbaik yang dapat memisahkan 2 kelas (diabetes dan tidak diabetes).

3.7 Evaluasi Model

Setelah melakukan pelatihan model dengan algoritma SVM, selanjutnya adalah tahap evaluasi dari model. Tahap ini dilakukan dengan cara mengukur kinerja dari model yang sudah dibuat menggunakan data uji. Setelah dilihat performanya, kedua buah model akan dibandingkan untuk dilihat kombinasi metode mana yang lebih baik.

3.7.1 Uji Model SVM

Pada tahap ini akan menguji kedua buah model yaitu PCA-SVM dan SFS-SVM dalam melakukan klasifikasi pada data yang belum pernah dilihat oleh model. Setelah mempersiapkan data pengujian, digunakan kedua model yang telah dilatih pada subset pelatihan untuk mengklasifikasi kelas dari data pengujian. SVM akan menggunakan pola yang telah dipelajari selama pelatihan untuk mengklasifikasikan data menjadi kelas yang sesuai pada tahap uji, yaitu melakukan klasifikasi individu memiliki diabetes atau tidak.

3.7.2 Perbandingan Kedua Model

Model yang dibuat dengan metode PCA-SVM akan dibandingkan dengan model SFS-SVM. Nilai ukur yang dibandingkan dari model yang dibuat adalah akurasi, presisi, *recall*, *specificity* dan *F1-Score* yang dihitung melalui *confusion matrix*. Pada tahap perbandingan ini akan memberikan informasi tentang seberapa baik kedua model yang sudah dibuat.

3.8 Perancangan Website Klasifikasi Diabetes

Dalam membangun sistem aplikasi yang dapat digunakan untuk diabetes, perancangan sitem diperlukan untuk mengetahui gambaran dari alur kerja dari sebuah sistem. Akan dibuat dua buah menu yang dapat dipilih oleh user untuk dapat melakukan klasifikasi diabetes, yaitu menu klasifikasi individu dan klasifikasi dataset. Berikut akan dibahas alur kerja dari sebuah sistem dan interaksinya dengan user yang akan digambarkan melalui beberapa diagram UML dan *design user interface* (UI) berikut.

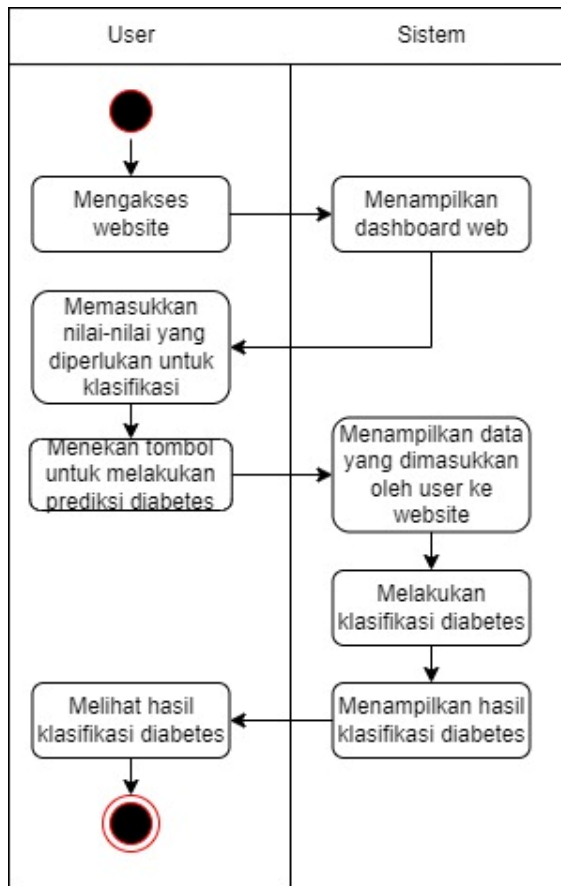
1. Diagram Alur Menu Klasifikasi Individu



Gambar 3.5 Alur Klasifikasi Diabetes Secara Individu

Gambar diatas merupakan alur dari klasifikasi diabetes secara individu. Pertama-tama setelah user memasuki halaman web, user akan melihat *dashboard* dari website, lalu user akan diminta untuk memasukkan nilai-nilai yang diperlukan sebelum melakukan klasifikasi diabetes seperti nama, kadar gula darah, insulin dan lain-lain. Setelah user memasukkan nilai-nilai tersebut, user dapat menekan tombol “Prediksi Diabetes” untuk melakukan klasifikasi terhadap data dari user tersebut. Setelah menekan tombol tersebut, website akan menampilkan data yang sudah dimasukkan sebelumnya oleh user, lalu website akan melakukan klasifikasi terhadap data tersebut dan akan menampilkan hasil dari klasifikasi yang dilakukan oleh website ke halaman web.

2. Activity Diagram Menu Klasifikasi Individu



Gambar 3.6 Activity Diagram Klasifikasi Diabetes Secara Individu

Gambar diatas merupakan *activity diagram* dari klasifikasi diabetes secara individu. Setelah user mengakses website, sistem akan menampilkan dashboard web yang akan dilihat oleh user. Setelah user dapat melihat halaman web, user dapat memasukkan nilai-nilai yang diperlukan untuk klasifikasi, seperti nama, kadar gula darah, insulin, dan lain-lain. Jika sudah memasukkan nilai-nilai yang diperlukan, user dapat menekan tombol “Prediksi Diabetes” untuk melakukan klasifikasi. Selanjutnya sistem akan menampilkan data-data yang sudah dimasukkan oleh user tersebut ke halaman web dan sistem akan melakukan klasifikasi terhadap data tersebut. Setelah melakukan klasifikasi, sistem akan menampilkan hasil klasifikasi tersebut ke halaman web. Terakhir, user dapat melihat hasil klasifikasi diabetes yang sudah dilakukan oleh sistem, hasil klasifikasi disajikan dalam bentuk tabel yang disatukan dengan data-data yang dimasukkan oleh user sebelumnya.

3. Design User Interface Menu Klasifikasi Individu

Design User Interface (UI) dari website ini bertujuan untuk membantu dan merancang aplikasi dengan tujuan untuk menciptakan pengalaman user yang baik dalam melakukan klasifikasi terhadap penyakit diabetes melitus. Pada penelitian ini penulis menggunakan aplikasi *pencil* untuk membuat *design* UI aplikasi web.

X	Klasifikasi Diabetes Pada Individu
Klasifikasi Diabetes	Masukkan Jumlah Kehamilan
Pilihan klasifikasi	Masukkan Kadar Gula Darah
<input type="radio"/> Klasifikasi Individu	Masukkan Tekanan Darah
<input checked="" type="radio"/> Klasifikasi Dengan Dataset	Masukkan Ketebalan Kulit
	Masukkan Insulin
	Masukkan Massa Tubuh
	Masukkan Diabetes Pedigree Function
	Masukkan Tahun Lahir
	Prediksi Diabetes

Gambar 3.7 *Design* UI Halaman Web Klasifikasi Diabetes Individu

Gambar diatas merupakan *design* UI halaman awal dari klasifikasi diabetes. Pada halaman awal pada website ini, merupakan menu untuk melakukan klasifikasi diabetes pada individu, user dapat memasukan nilai-nilai yang sesuai pada kolom yang disediakan, lalu user dapat melakukan klasifikasi apakah individu tersebut memiliki diabetes atau tidak memiliki diabetes.

X

Klasifikasi Diabetes

Pilihan klasifikasi

☒ Klasifikasi Individu

☐ Klasifikasi Dengan Dataset

Hasil Klasifikasi Diabetes Pada Individu

Name	Pregnanci	Glucosa	BP	SkinThick	Insulin	BMI	DBF	Age	Target
Uta	0	120	70	30	100	30	0.2	21	0

Gambar 3.8 *Design* UI Hasil Klasifikasi Diabetes Individu

Gambar diatas merupakan *design* UI hasil klasifikasi diabetes individu. Setelah memasukkan nilai-nilai yang diperlukan untuk melakukan klasifikasi diabetes dan menekan tombol “Prediksi Diabetes”, model akan melakukan proses klasifikasi terhadap data tersebut. Setelah model mendapatkan hasil klasifikasi, website akan menampilkan hasil klasifikasi berupa tabel yang menampilkan data dari individu yang sudah dimasukkan sebelumnya serta hasil klasifikasi. Hasil klasifikasi akan terletak di paling kanan dari tabel yang ditampilkan.

4. Diagram Alur Klasifikasi Dengan *Dataset*



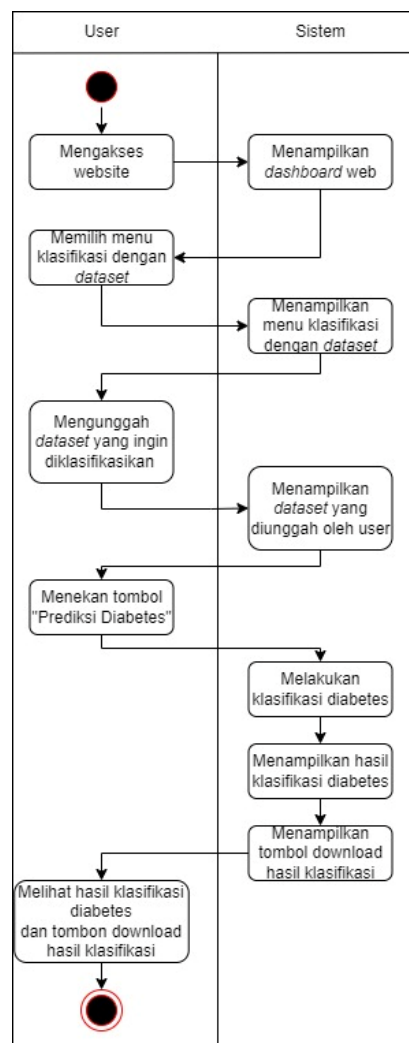
Gambar 3.9 Diagram Alir Klasifikasi Dengan *Dataset*

Gambar diatas merupakan diagram alir dari menu klasifikasi diabetes dengan *dataset*. Pertama user akan melihat *dashboard* web dan user memilih menu klasifikasi diabetes dengan *dataset*. Setelah memilih menu tersebut user akan diminta untuk mengunggah file *dataset* yang akan diklasifikasi oleh website, file harus berekstensi .csv, .xls, dan .xlsx. Setelah user menunggah *dataset* sistem akan menampilkan *dataset* yang sudah diupload dalam bentuk *dataframe*. *Dataset* yang diupload harus memiliki seluruh nilai yang terdapat pada setiap individu, seperti nama, kadar gula darah, insulin dan lain-lain. Jika tidak maka seperti klasifikasi pada individu, akan terjadi error dan

user akan diminta untuk mengunggah ulang *dataset* dengan kriteria yang sesuai untuk dapat melakukan klasifikasi.

Setelah sistem menampilkan *dataset* yang diunggah, maka user dapat menekan tombol “Prediksi Diabetes” untuk melakukan klasifikasi terhadap *dataset* tersebut. Setelah menekan tombol tersebut sistem akan melakukan klasifikasi terhadap *dataset* yang diunggah, lalu setelah mendapatkan hasil klasifikasi, sistem akan menampilkan hasil klasifikasi ke halaman web. User juga dapat melihat tombol untuk mengunduh file hasil klasifikasi yang dilakukan oleh sistem, file ini akan berbentuk file csv.

5. Activity Diagram Klasifikasi Dengan Dataset



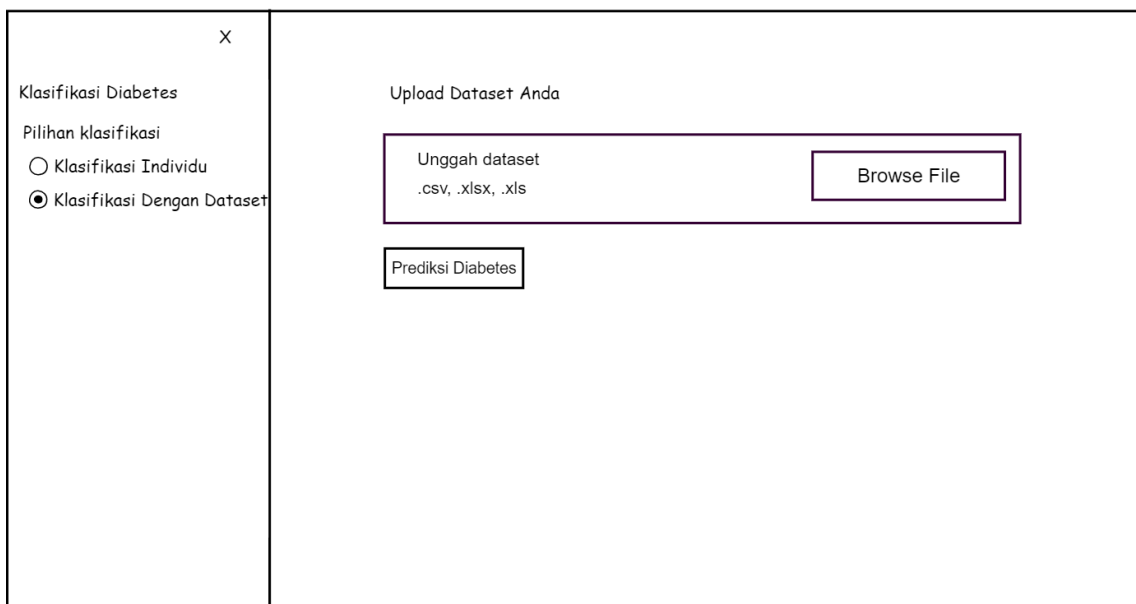
Gambar 3.10 Activity Diagram Klasifikasi Dengan Dataset

Gambar diatas merupakan *activity diagram* dari menu klasifikasi diabetes dengan *dataset*. Dari *activity diagram* tersebut dapat dilihat peran user dan peran sistem dalam website. User pertama-tama akan mengakses website klasifikasi diabetes, lalu

sistem akan menampilkan *dashboard* web untuk dilihat oleh user. Setelah user dapat melihat *dashboard* dari web, user dapat memilih menu klasifikasi dengan *dataset* untuk melakukan klasifikasi dengan *dataset* lalu sistem akan menampilkan menu tersebut. Setelah itu user dapat mengunggah file dari *dataset* yang akan diklasifikasikan oleh website. Setelah mengunggah file, sistem akan menampilkan file tersebut.

Setelah user melihat file yang ditampilkan oleh website, user dapat menekan tombol “Prediksi Diabetes” untuk melakukan klasifikasi terhadap *dataset* yang diunggah. Setelah itu sistem akan melakukan klasifikasi terhadap *dataset* tersebut dan akan menampilkan hasil klasifikasinya ke halaman web. Sistem juga menampilkan tombol untuk mengunduh file hasil klasifikasi yang dilakukan oleh sistem. User dapat melihat hasil klasifikasi pada web, dan dapat juga mengunduh hasil klasifikasi tersebut.

6. Design User Interface Menu Klasifikasi Dataset



The image shows a web interface design for 'Klasifikasi Diabetes'. It features a sidebar on the left with a close button (X) and two radio button options: 'Klasifikasi Individu' and 'Klasifikasi Dengan Dataset', with the latter being selected. The main content area is titled 'Upload Dataset Anda' and contains a box for uploading a dataset, specifying supported formats (.csv, .xlsx, .xls) and a 'Browse File' button. Below this is a 'Prediksi Diabetes' button.

Gambar 3.11 Design UI Halaman Web Klasifikasi Diabetes dengan *Dataset*

Gambar diatas merupakan *design* UI dari halaman web klasifikasi diabetes menggunakan *dataset*. Pada halaman ini, user dapat mengunggah *dataset* yang ingin dilakukan klasifikasi. User hanya dapat mengunggah file dengan ekstensi .csv, .xlsx, .xls. Setelah megunggah *dataset*, user dapat menekan tombol “Prediksi Diabetes” untuk melakukan klasifikasi.

X

Klasifikasi Diabetes

Pilihan klasifikasi

☐ Klasifikasi Individu
 ☒ Klasifikasi Dengan Dataset

Upload Dataset Anda

Unggah dataset

.csv, .xlsx, .xls

Browse File

Prediksi Diabetes

Name	Pregna	Glucosi	BloodPr	SkinThi	Insulin	BMI	DBF	Age	Target
A	0	170	70	30	100	37.2	0.2	25	1
B	1	160	80	40	110	36.5	0.5	22	0
C	2	150	70	30	120	35.3	0.3	23	1
D	3	180	80	40	100	38.4	0.4	22	0
D	4	140	70	30	110	32.2	0.5	27	1

Unduh file CSV

Gambar 3.12 *Design UI Hasil Klasifikasi Diabetes dengan Dataset*

Gambar diatas merupakan *design UI* dari hasil klasifikasi diabetes menggunakan *dataset*. Setelah menekan tombol “Prediksi Diabetes” model akan melakukan klasifikasi kepada tiap sampel yang terdapat pada dataset tersebut. Setelah model melakukan klasifikasi, website akan menampilkan hasil klasifikasi untuk dilihat user dalam bentuk tabel. Website akan menampilkan *dataset* yang diunggah oleh user disertakan dengan hasil klasifikasi yang terletak di paling kanan dalam tabel yang ditampilkan. Pada menu ini user juga dapat mengunduh hasil klasifikasi dalam bentuk csv pada tombol “Unduh file CSV”.

BAB IV HASIL DAN PEMBAHASAN

4.1 Hasil *Data Cleaning*

Dataset yang sudah di-*import* ke dalam *jupyter notebook*, selanjutnya akan melalui tahap penghapusan pada sampel yang memiliki nilai yang hilang pada fitur-fitur tertentu. Penghapusan sampel ini diperlukan untuk mencegah pengaruh buruk kepada model yang dikarenakan oleh nilai fitur yang hilang tersebut. Untuk menghapus sampel tersebut, pertama-tama dilakukan perhitungan jumlah nilai yang hilang pada fitur-fitur tersebut.

```
Jumlah nilai 0 pada kolom Pregnancies: 111
Jumlah nilai 0 pada kolom Glucose: 5
Jumlah nilai 0 pada kolom BloodPressure: 35
Jumlah nilai 0 pada kolom SkinThickness: 227
Jumlah nilai 0 pada kolom Insulin: 374
Jumlah nilai 0 pada kolom BMI: 11
Jumlah nilai 0 pada kolom DiabetesPedigreeFunction: 0
Jumlah nilai 0 pada kolom Age: 0
Jumlah nilai 0 pada kolom Outcome: 500
```

Gambar 4.1 Jumlah Nilai 0 Pada Setiap Fitur

Setelah melakukan perhitungan jumlah nilai 0, akan dipilih sampel yang tidak memiliki nilai 0 pada fitur *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, dan *BMI*. Dikarenakan nilai 0 pada fitur-fitur tersebut adalah kemungkinan nilai yang hilang atau nilai yang tidak diukur. Setelah melalui tahap ini, sampel yang tersisa adalah 392 sampel.

4.2 Pembagian *Dataset*

Dengan sampel yang tersisa (392 sampel), *dataset* akan dibagi menjadi data latih dan data uji. Data latih akan digunakan untuk melatih model SVM, sedangkan data uji akan digunakan untuk mengukur performa SVM. Data yang dibagi ini juga akan disimpan menjadi file CSV agar dapat digunakan secara terus-menerus (tidak dibagi ulang). *Dataset* dibagi menjadi dengan perbandingan 80% untuk data latih, dan 20% untuk data uji dari total 392 sampel. Setelah dibagi *dataset* akan disimpan menjadi file csv dan akan di-*import* kembali ke *jupyter notebook*.

Tabel 4.1 Hasil Pembagian *Dataset*

Data Latih	Data Uji
313 Sampel	79 Sampel

Tabel 4.1 merupakan hasil dari pembagian *dataset*, dapat dilihat sampel total pada data latih adalah 313 sampel. Sedangkan pada data uji adalah 79 sampel. Data latih dan data uji yang digunakan pada penelitian ini dapat dilihat pada lampiran I.

4.3 *Preprocessing Data*

Data latih dan data uji selanjutnya akan memasuki tahap *preprocessing* data. Adapun tahapan *preprocessing* disini terbagi menjadi 2 yaitu:

1. Memisahkan fitur dan target.

Pemisahan fitur dan target diperlukan agar segala proses yang akan dilakukan terhadap fitur seperti *scaling* data, tahap reduksi dimensi dan lain-lain tidak mempengaruhi atau dipengaruhi variabel target/label.

2. *Scaling* data

Scaling dilakukan untuk setiap data yang akan dilatih ataupun diuji, karena pada saat tahap pelatihan, jika data tidak dilakukan *scaling* maka akan berpengaruh buruk terhadap performa model. Adapun persamaan yang digunakan untuk mendapatkan nilai hasil *scaling* adalah sebagai berikut.

$$X = \frac{X-\mu}{\sigma} \dots \dots \dots (6)$$

X adalah setiap sampel yang terdapat pada data, μ adalah rata-rata dari data, dan σ adalah standar deviasi. Kemudian *scaler* akan di latih pada data latih sehingga akan mendapatkan nilai standar deviasi, dan rata-rata dari data latih. Selanjutnya data latih akan di-*transform* untuk mendapatkan data yang di-*scale*. Begitu pula dengan data uji, akan di-*tranform* untuk mendapatkan data yang di-*scale*, berdasarkan nilai σ dan μ yang didapatkan dari data latih. Berikut adalah hasil dari tahap *scaling* data.

Tabel 4.2 Contoh Data Latih Dari Hasil *Scaling*

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age
-1.03	1.86	1.51	-0.29	-0.59	0.47	-0.60	0.38
-0.07	-1.37	1.20	-1.22	-0.79	-0.82	-0.63	-0.86
1.52	0.10	1.36	0.64	-0.43	0.75	-0.50	1.72
-0.39	-0.88	-0.21	-1.50	-0.93	-1.73	0.37	-0.48
-0.07	-1.41	0.89	0.18	-0.75	0.14	2.26	-0.38
-1.03	-0.10	0.26	-1.04	-0.81	-0.39	-0.69	-0.48
-0.39	1.11	0.26	0.55	2.38	0.88	-1.13	-0.10
-0.07	-0.69	-2.10	-0.85	-0.55	-0.35	-0.35	-0.48
0.25	-1.27	1.51	-0.57	-0.87	0.89	-1.06	-0.57
-0.71	-0.92	0.89	-0.38	0.18	0.25	-0.84	1.15

Tabel 4.3 Contoh Data Uji Dari Hasil *Scaling*

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age
1.20	-0.69	0.26	1.02	-0.46	0.57	-0.93	1.34
-1.03	0.13	0.73	0.74	0.43	0.44	0.83	-0.76
0.25	-1.24	-1.00	-0.66	-0.93	-0.77	-0.63	-0.29
-1.03	1.27	0.41	2.51	-0.50	2.85	0.70	-0.57
-1.03	1.90	1.36	1.39	2.97	1.44	-0.87	-0.48
-0.07	-0.88	-1.16	0.46	-0.37	-1.21	1.24	0.76
-0.71	-1.70	-1.78	-1.04	-0.70	-1.83	-0.58	-0.86
0.56	-1.21	-0.21	-0.10	-0.75	-0.43	-0.46	-0.67
2.16	1.24	-0.21	-0.57	-0.23	-1.10	-0.57	1.53
0.25	-1.05	-0.06	0.27	-0.60	-0.02	-0.22	-0.86

Setelah melalui tahap *preprocessing* dibuat 2 buah model yaitu model PCA-SVM dan model SFS-SVM.

4.4 Model PCA-SVM

4.4.1 Reduksi Dimensi Dengan PCA

Model PCA-SVM akan dibuat sebanyak 8 model, dengan masing-masing akan mempertahankan jumlah *principal component* yang berbeda. Untuk melihat bagaimana

bentuk data setelah proses PCA, dapat dilihat pada Tabel 4.3 dan 4.4, dipaparkan bagaimana bentuk data setelah reduksi dimensi dari PCA ke 2 dimensi (artinya mempertahankan 2 *principal component*).

Tabel 4.4 Contoh Data Latih Hasil Reduksi Dengan PCA

PC 1	PC 2
-0.88	0.28
1.67	-0.52
-2.05	-0.84
2.26	-1.06
0.35	0.94
1.41	-0.04
-1.62	0.59
1.90	-0.40
0.50	0.36
-0.09	-0.23

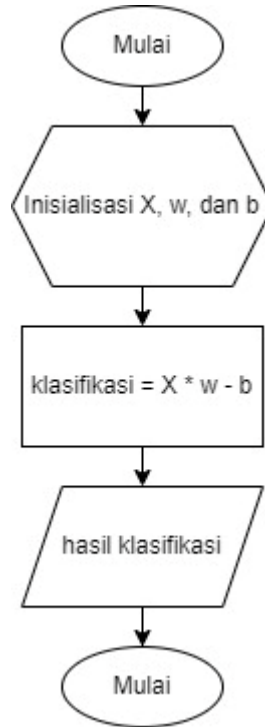
Dapat dilihat pada tabel diatas merupakan hasil reduksi dimensi data latih, dari fitur yang semula terdapat 8 fitur, hasil perkaliannya dengan *eigenvector* menghasilkan 2 fitur baru yang disebut PC1 dan PC2.

Tabel 4.5 Contoh Data Uji Hasil Reduksi Dengan PCA

PC 1	PC 2
-1.09	-0.48
-0.32	1.58
1.83	-0.73
-1.97	3.53
-2.70	1.77
0.72	-0.47
3.30	-0.73
1.19	-0.22
-1.05	-2.83
0.89	0.45

Dapat dilihat pada tabel diatas merupakan hasil reduksi dimensi data uji, dari fitur yang semula terdapat 8 fitur, hasil perkaliannya dengan *eigenvector* menghasilkan 2 fitur baru yang disebut PC1 dan PC2.

4..4.2 Evaluasi Model PCA-SVM



Gambar 4.2 Pengujian Model PCA-SVM.

Setelah melalui proses pelatihan dengan menggunakan data latih yang bertujuan untuk mendapatkan vektor bobot dan nilai bias. Nilai-nilai tersebut digunakan untuk mendapatkan garis pemisah terbaik antara 2 kelas. Pada tahap pengujian model PCA-SVM ini, data uji yang sudah melalui proses ekstraksi fitur PCA digunakan. Setelah mendeklarasikan data uji, masuk ke tahap kalkulasi matriks dari data uji (X) dengan vektor bobot (w), lalu dikurangi dengan nilai b (bias). Nilai hasil kalkulasi akan disimpan pada sebuah variabel. Lalu setelah melakukan kalkulasi, hasil klasifikasi dapat dilihat.

Dibuat 8 model yang dimana masing-masing model mempertahankan jumlah *principal component* yang berbeda, dilakukan evaluasi model dengan parameter akurasi, presisi, *recall*, *specificity* dan *F1-Score*. Berikut adalah performa dari model-model PCA-SVM dengan jumlah *principal component* yang berbeda.

Tabel 4.6 Performa Model PCA-SVM Dengan Jumlah PC Yang Berbeda

Jumlah <i>principal component</i>	Akurasi (%)	Presisi (%)	<i>Recall</i> (%)	<i>Specificity</i> (%)	<i>F1-Score</i> (%)
1	72.15	80.77	77.78	60.00	79.25
2	72.15	80.77	77.78	60.00	79.25
3	74.68	86.54	77.59	66.67	81.82
4	75.95	88.46	77.97	70.00	82.88

Jumlah <i>principal component</i>	Akurasi (%)	Presisi (%)	<i>Recall</i> (%)	<i>Specificity</i> (%)	F1-Score (%)
5	75.95	88.46	77.97	70.00	82.88
6	75.95	84.62	80.00	66.67	82.24
7	75.95	86.54	78.95	68.18	82.57
8	74.68	84.62	78.57	65.22	81.48

Berdasarkan tabel diatas dapat dilihat bagaimana performa dari model PCA-SVM jika jumlah dimensinya direduksi. Jika jumlah dimensi direduksi atau jumlah *principal component* yang dipertahankan dikurangi, performa model semakin meningkat. Performa tertinggi model terletak pada jumlah *principal component* 4 dengan akurasi 75.95%, presisi 88.46%, *recall* 77.97%, *specificity* 70.00%, dan F1-Score 82.88%.

4.5 Model SFS-SVM

4.5.1 Seleksi Fitur Dengan SFS

Model SFS-SVM akan dibuat sebanyak 8 model, dengan masing-masing akan mempertahankan jumlah fitur yang berbeda. SFS akan memilih fitur berdasarkan pengaruh fitur tersebut terhadap performa dari model. Untuk melihat bagaimana bentuk data latih dan data uji setelah dilakukannya seleksi fitur dapat dilihat Tabel dan Tabel. Dipaparkan bagaimana bentuk dari data setelah seleksi fitur menjadi 2 dimensi.

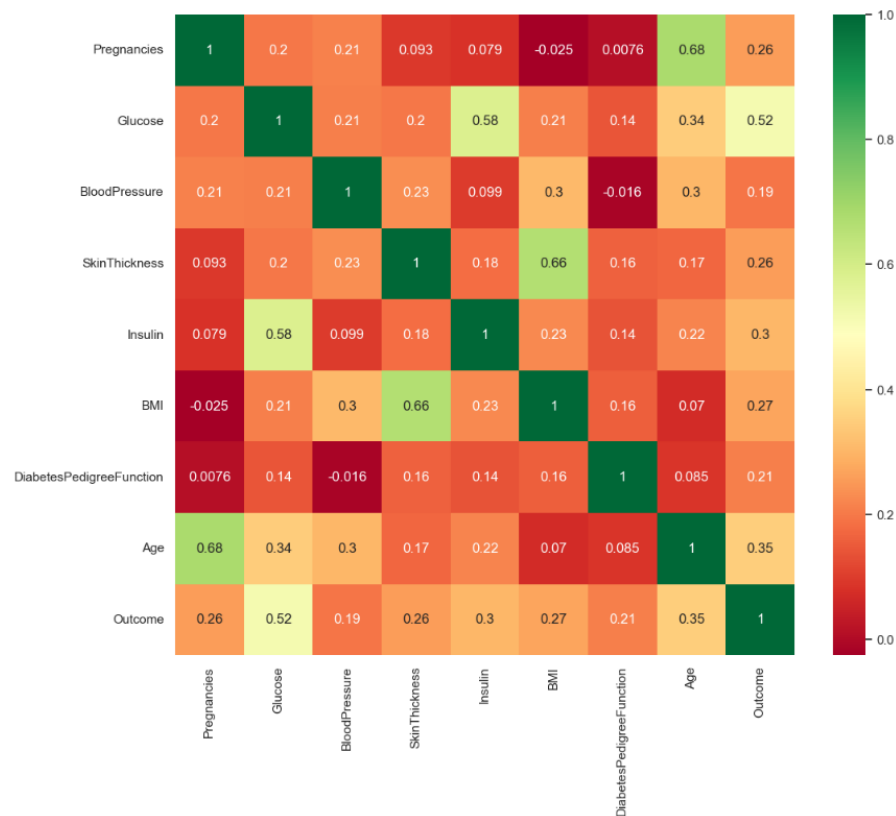
Tabel 4.7 Contoh Data Latih Hasil Seleksi Fitur

Glucose	Age
1.86	0.38
-1.37	-0.86
0.10	1.72
-0.88	-0.48
-1.41	-0.38
-0.10	-0.48
1.11	-0.10
-0.69	-0.48
-1.27	-0.57
-0.92	1.15

Tabel 4.8 Contoh Data Uji Hasil Seleksi Fitur

Glucose	Age
-0.69	1.34
0.13	-0.76
-1.24	-0.29
1.27	-0.57
1.90	-0.48
-0.88	0.76
-1.70	-0.86
-1.21	-0.67
1.24	1.53
-1.05	-0.86

Dapat dilihat pada tabel-tabel diatas bentuk data setelah diseleksi menjadi 2 fitur, fitur yang dipertahankan adalah *Glucose* dan *Age*. Hasil seleksi dari SFS ini juga dapat dipastikan dengan melihat bobot dari setiap fitur terhadap target yang dapat dilihat pada gambar berikut.



Gambar 4.3 Korelasi Antar Fitur dan Label.

Dari gambar diatas dapat dilihat korelasi antar fitur dan juga korelasi antar fitur dengan target. Berdasarkan proses seleksi fitur sebelumnya, dipilih hanya 2 fitur yaitu *Glucose* dan *Age*. Hal itu dapat diperkuat dengan dilihatnya korelasi fitur dengan target yaitu *Outcome*, fitur *Age* dan *Glucose* memiliki korelasi tertinggi dengan *Outcome* diantara fitur fitur lain.

4.5.2 Evaluasi Model SFS-SVM

Alur pengujian dari model SFS-SVM sama dengan proses pengujian model PCA-SVM yang dapat dilihat pada Gambar 4.4. Setelah mendapatkan vektor bobot (w) dan nilai bias (b) dari proses pelatihan model SFS-SVM, dilakukan pengujian model SFS-SVM dengan menggunakan nilai-nilai tersebut. Vektor w dan nilai b digunakan untuk membuat garis pemisah antara 2 kelas (diabetes dan tidak diabetes). Proses pengujian menggunakan data uji yang sudah melalui proses seleksi fitur menggunakan SFS.

Dibuat 8 model yang dimana masing-masing model mempertahankan jumlah fitur yang berbeda, dilakukan evaluasi model dengan parameter akurasi, presisi, *recall*, *specificity* dan *F1-Score*. Berikut adalah performa dari model-model SFS-SVM dengan jumlah fitur yang berbeda.

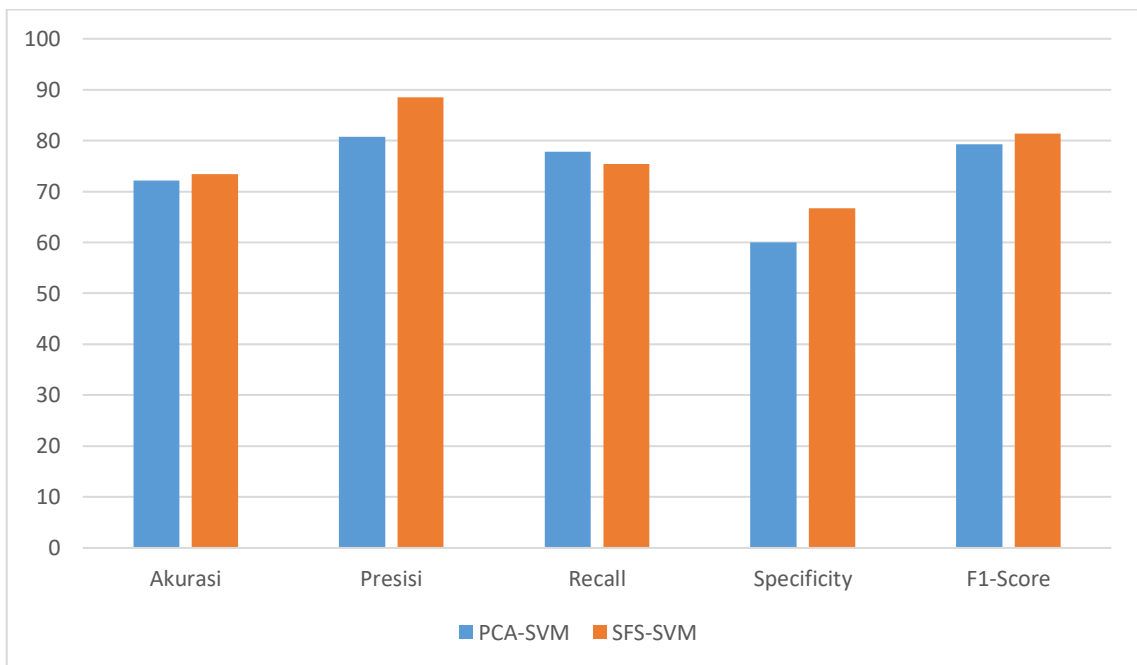
Tabel 4.9 Performa Model SFS-SVM Dengan Jumlah Fitur Yang Berbeda

Jumlah fitur hasil seleksi	Akurasi (%)	Presisi (%)	<i>Recall</i> (%)	<i>Specificity</i> (%)	<i>F1-Score</i> (%)
1	73.42	88.46	75.41	66.67	81.42
2	75.95	90.38	77.05	72.22	83.19
3	75.95	90.38	77.05	72.22	83.19
4	75.95	88.46	77.97	70.00	82.88
5	73.42	86.54	76.27	65.00	81.08
6	77.22	86.54	80.36	69.57	83.33
7	73.42	84.62	77.19	63.64	80.73
8	74.68	84.62	78.57	65.22	81.48

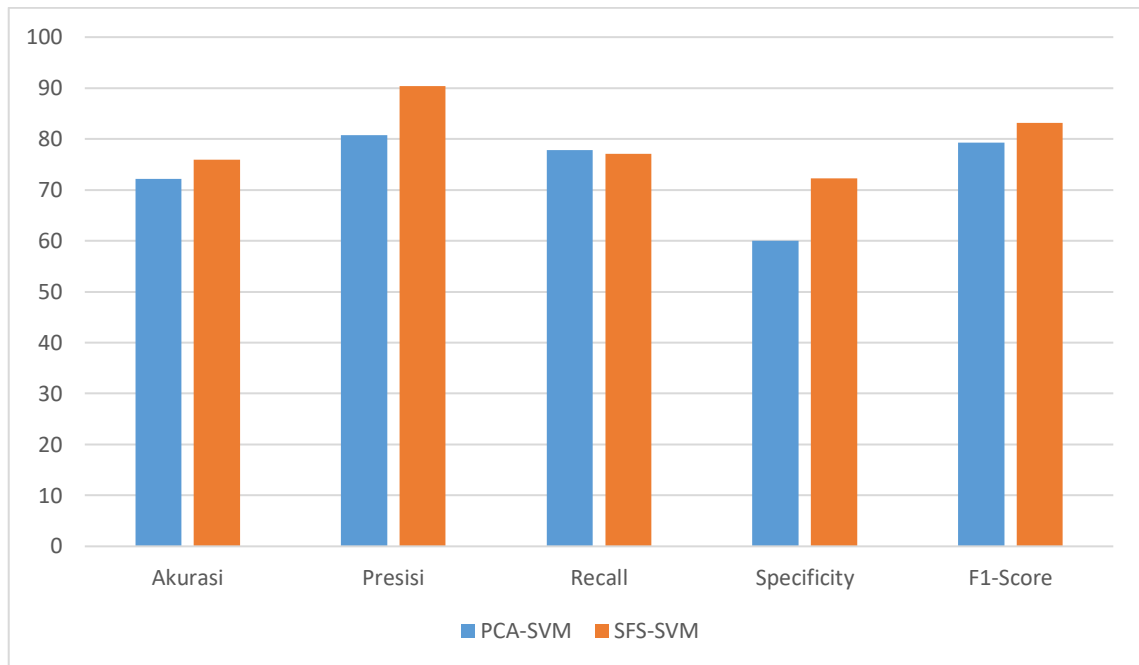
Berdasarkan tabel diatas dapat dilihat bagaimana performa dari model SFS-SVM jika jumlah dimensinya direduksi. Jika jumlah dimensi direduksi atau jumlah fitur yang dipertahankan dikurangi, performa model semakin meningkat. Performa tertinggi model terletak pada jumlah fitur hasil seleksi 6 dengan akurasi 77.22%, presisi 86.54%, *recall* 80.36%, *specificity* 69.57%, dan *F1-Score* 83.33%.

4.6 Perbandingan Model PCA-SVM Dengan SFS-SVM

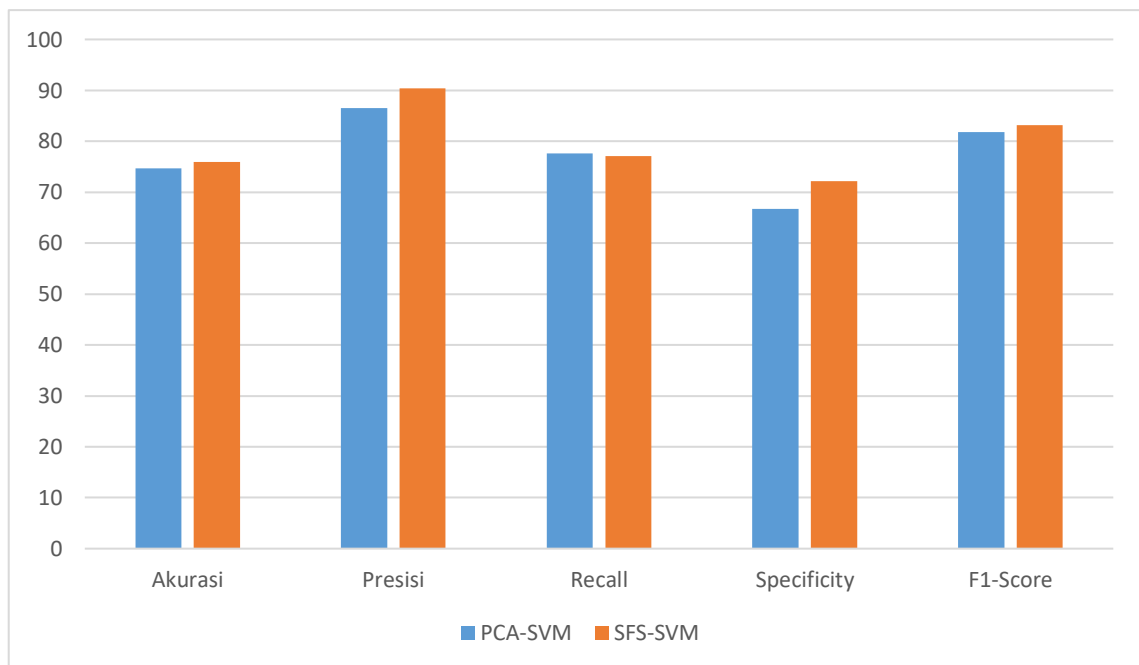
Pada tahap ini, kedua model yaitu model PCA-SVM yang menggunakan metode reduksi dimensi *Principal Component Analysis* (PCA) untuk ekstraksi fitur, dan model SFS-SVM yang menggunakan metode seleksi fitur *Sequential Forward Selection* (SFS) untuk seleksi fitur akan dilihat perbandingan performa yang dihasilkan dari kedua model yang sudah dibuat tersebut. Perbandingan akan ditinjau berdasarkan jumlah fitur atau dimensi dari masing-masing model, adapun perbandingannya dapat dilihat dari grafik-grafik berikut berikut.



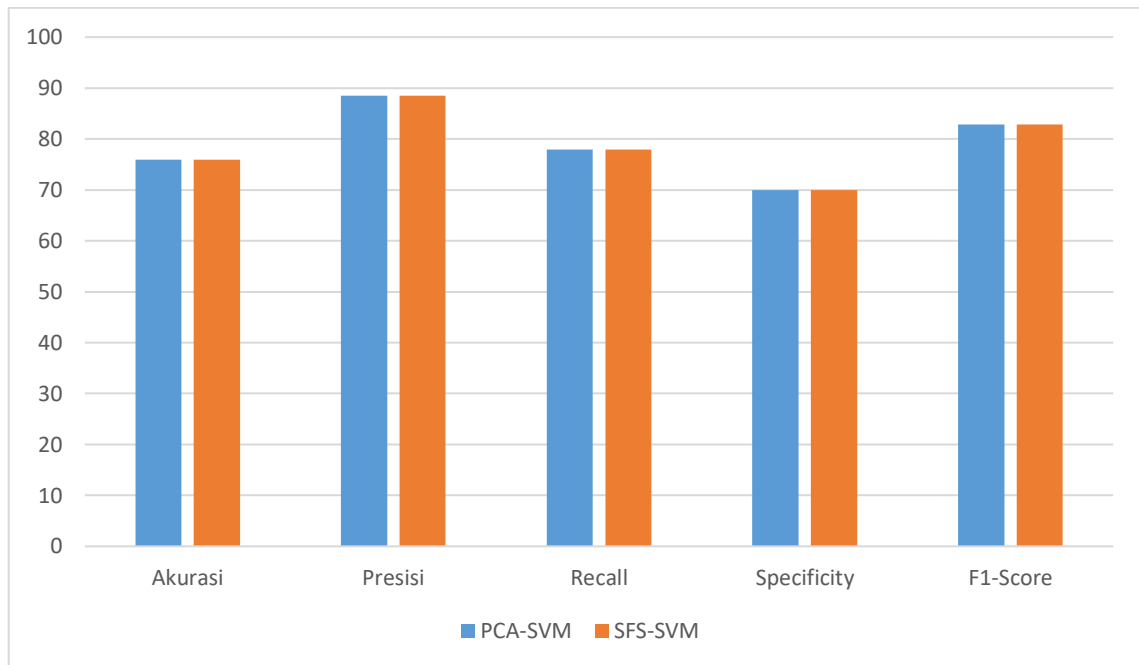
Gambar 4.4 Grafik Perbandingan Performa Model Dengan 1 Fitur



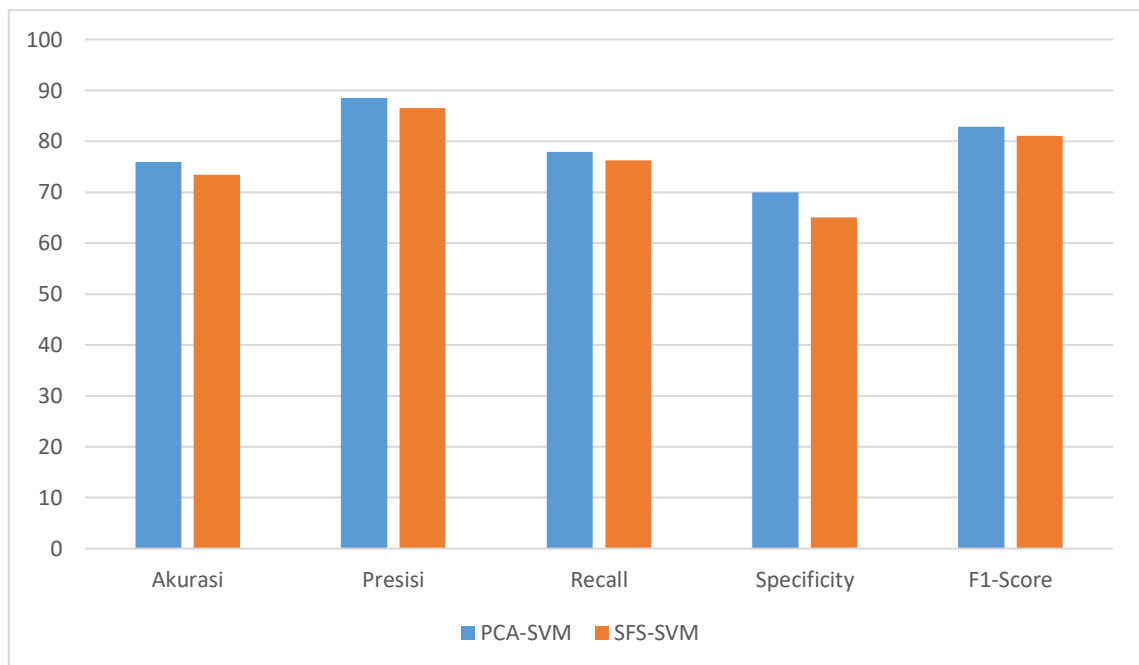
Gambar 4.5 Grafik Perbandingan Performa Model Dengan 2 Fitur



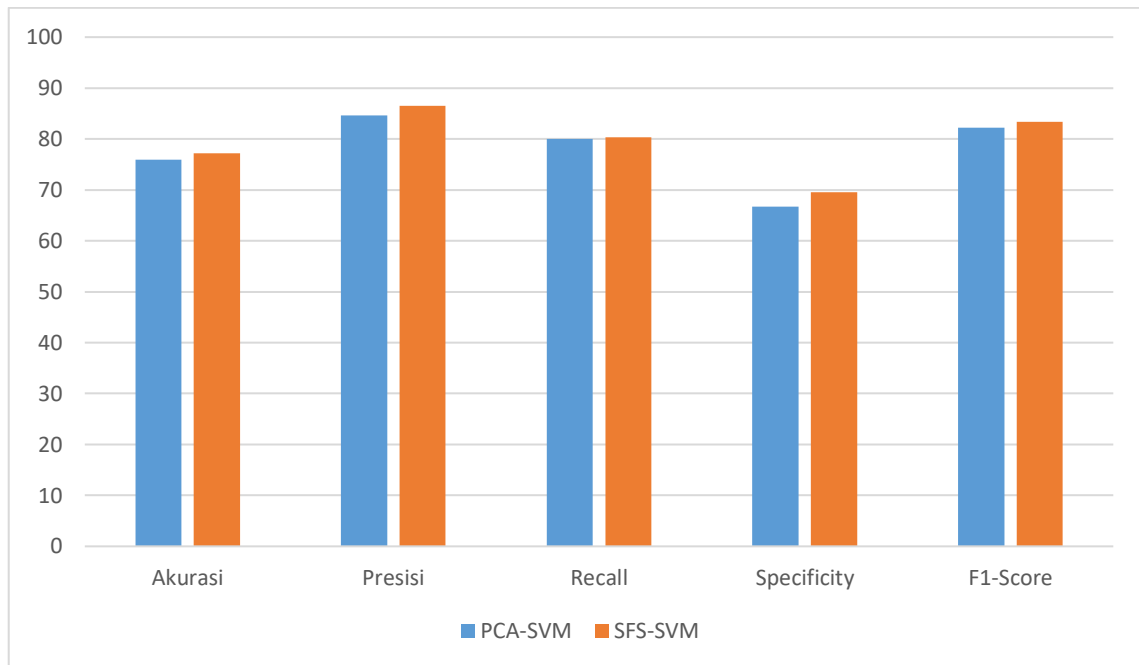
Gambar 4.6 Grafik Perbandingan Performa Model Dengan 3 Fitur



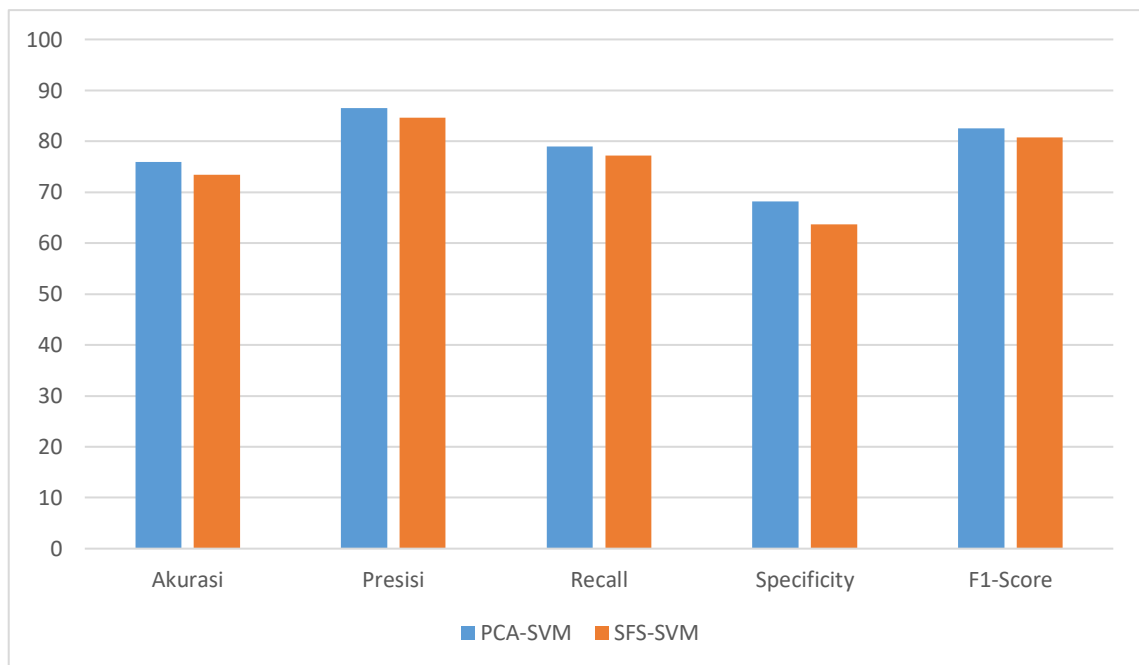
Gambar 4.7 Grafik Perbandingan Performa Model Dengan 4 Fitur



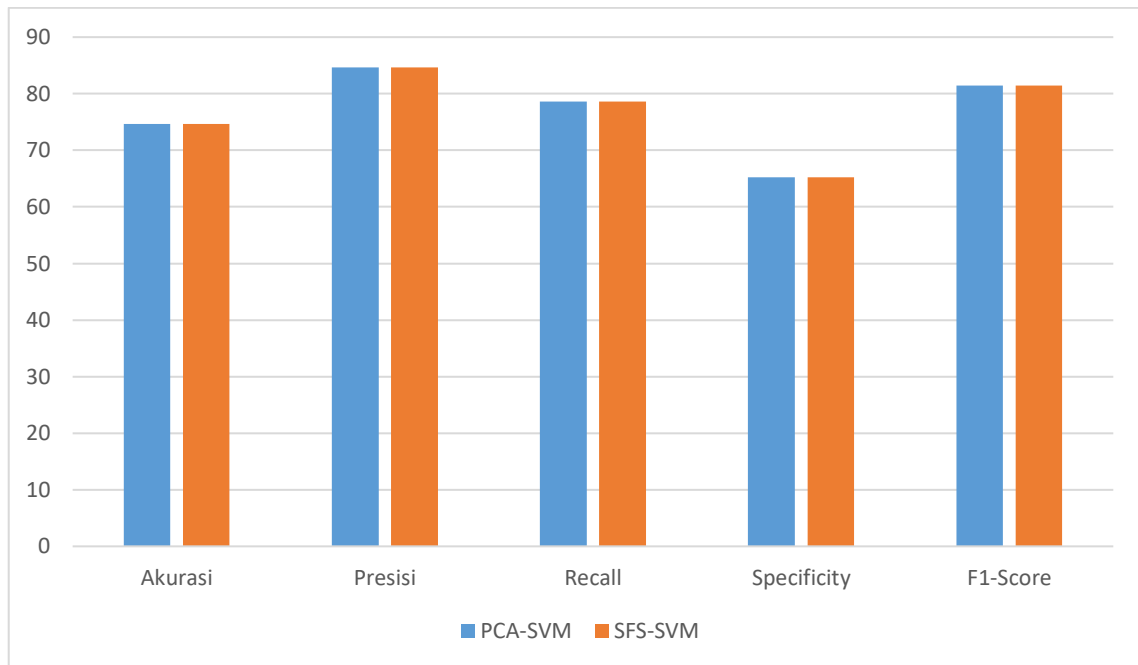
Gambar 4.8 Grafik Perbandingan Performa Model Dengan 5 Fitur



Gambar 4.9 Grafik Perbandingan Performa Model Dengan 6 Fitur



Gambar 4.10 Grafik Perbandingan Performa Model Dengan 7 Fitur



Gambar 4.11 Grafik Perbandingan Performa Model Dengan 8 Fitur

Berdasarkan grafik-grafik diatas dapat dilihat perbandingan performa model PCA-SVM dan SFS-SVM berdasarkan jumlah fitur. Didapatkan bahwa performa model meningkat jika fiturnya dikurangi dengan menggunakan teknik reduksi fitur baik PCA ataupun SFS. Performa metode PCA-SVM terbaik didapatkan pada *principal component* berjumlah 4, sedangkan performa metode SFS-SVM terbaik didapatkan pada hasil seleksi fitur berjumlah 6.

Didapatkan metode SFS-SVM mencapai performa terbaik dengan akurasi 77.22%, presisi 86.54%, *recall* 80.36%, *specificity* 69.57%, dan *F1-Score* 83.33%, sedangkan performa metode PCA-SVM mencapai performa terbaik dengan akurasi 75.95%, presisi 88.46%, *recall* 77.97%, *specificity* 70.00%, dan *F1-Score* 82.88%. Berdasarkan hasil tersebut dapat diartikan bahwa model SFS-SVM lebih baik dalam mengklasifikasikan kelas negatif, sedangkan model PCA-SVM lebih baik dalam mengklasifikasikan kelas positif.

4.7 Hasil Perancangan Website

Setelah melalui tahap pembuatan model dan melakukan perbandingan antara kedua buah model yang sudah dibuat, dibuat sebuah website untuk melakukan klasifikasi diabetes menggunakan model PCA-SVM yang memiliki performa terbaik. Website ini dapat mengklasifikasikan diabetes secara individu (hanya untuk 1 orang), dan dapat mengklasifikasikan diabetes dengan menggunakan *dataset*

(mengklasifikasikan lebih dari 1 orang secara sekaligus). Kedua fitur tersebut disajikan dalam bentuk menu yang dapat dipilih oleh user. Adapun penyajian hasil perancangan website tersebut akan dibahas berdasarkan menu-menu tersebut.

4.7.1 Klasifikasi Individu

Pada menu klasifikasi diabetes pada individu, akan dilakukan sebuah klasifikasi diabetes hanya pada 1 orang / 1 sampel. User akan diminta untuk memasukkan nilai-nilai yang dibutuhkan untuk melakukan klasifikasi diabetes.



The screenshot shows a web application interface for 'Klasifikasi Diabetes'. On the left, a sidebar titled 'Klasifikasi Diabetes' contains a section 'Pilihan Klasifikasi' with two radio buttons: 'Klasifikasi Individu' (selected) and 'Klasifikasi dengan Dataset'. Below this is a blue box with text: 'Anda dapat memilih opsi untuk melakukan klasifikasi diabetes dengan pilihan diatas'. The main content area is titled 'Klasifikasi Diabetes pada Individu' and contains several input fields with labels and increment/decrement buttons: 'Masukan nama' (text input), 'Masukan jumlah kehamilan' (number input, value 0), 'Masukan kadar gula darah' (number input, value 0), 'Masukan tekanan darah' (number input, value 0), 'Masukan ketebalan kulit' (number input, value 0), 'Masukan insulin' (number input, value 0), and 'Masukan massa tubuh' (number input, value 0.00).

Gambar 4.12 Halaman Website Klasifikasi Diabetes Pada Individu

Gambar diatas merupakan halaman website klasifikasi diabetes pada menu klasifikasi individu. Pada halaman ini user akan diminta untuk memasukkan nilai-nilai seperti nama, jumlah kehamilan (jika wanita), kadar gula darah, tekanan darah dan lain-lain. Setiap nilai harus diisi sesuai dengan ketentuan yang diminta oleh website. Sebagai contoh seluruh kolom kecuali jumlah kehamilan tidak boleh kosong atau nol, jika tidak sesuai ketentuan maka tombol klasifikasi diabetes jika ditekan akan memunculkan *error message* pada website.

Klasifikasi Diabetes

Pilihan Klasifikasi

☒ Klasifikasi Individu

☐ Klasifikasi dengan Dataset

Anda dapat memilih opsi untuk melakukan klasifikasi diabetes dengan pilihan diatas

Masukkan kadar gula darah: 120

Masukkan tekanan darah: 70

Masukkan ketebalan kulit: 30

Masukkan insulin: 0

Masukkan massa tubuh: 32.00

Masukkan diabetes pedigree function: 0.300

Masukkan tahun lahir: 2002

Prediksi Diabetes

Isi nilai yang diperlukan!

Gambar 4.13 Contoh *error* dari website klasifikasi individu

Gambar diatas merupakan contoh *error* yang ditampilkan oleh website jika user tidak memasukkan nilai-nilai sesuai kriteria. Dapat dilihat pada kolom insulin nilai yang diinputkan adalah 0, maka website akan menganggap bahwa kolom ini belum terisi, dan user harus mengisinya terlebih dahulu sebelum melakukan klasifikasi diabetes.

Klasifikasi Diabetes

Pilihan Klasifikasi

☒ Klasifikasi Individu

☐ Klasifikasi dengan Dataset

Anda dapat memilih opsi untuk melakukan klasifikasi diabetes dengan pilihan diatas

Masukkan kadar gula darah: 120

Masukkan tekanan darah: 70

Masukkan ketebalan kulit: 30

Masukkan insulin: 90

Masukkan massa tubuh: 32.00

Masukkan diabetes pedigree function: 0.300

Masukkan tahun lahir: 2002

Prediksi Diabetes

Gambar 4.14 Halaman Website Klasifikasi Diabetes Pada Individu

Gambar diatas merupakan contoh pengisian nilai-nilai dalam website, setelah memasukkan nilai-nilai tersebut sesuai dengan kriteria yang sudah ditentukan, maka user dapat menekan tombol “Prediksi Diabetes” untuk melakukan klasifikasi terhadap data yang diisi.

Klasifikasi Diabetes

Pilihan Klasifikasi

☒ Klasifikasi Individu
 ☐ Klasifikasi dengan Dataset

Anda dapat memilih opsi untuk melakukan klasifikasi diabetes dengan pilihan diatas

32.00

Masukkan diabetes pedigree function

0.300

Masukkan tahun lahir

2002

Prediksi Diabetes

Data Individu

	Name	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFu
0	Saputra	0	120	70	30	90	32	

Hasil Prediksi

	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Prediksi
0 0	70	30	90	32	0.3	21	Tidak Diabetes

Gambar 4.15 Halaman Website Hasil Klasifikasi Diabetes Pada Individu

Gambar diatas merupakan halaman website hasil klasifikasi diabetes pada individu, dapat dilihat setelah menekan tombol “Prediksi Diabetes”, sistem akan menampilkan data yang sudah dimasukkan oleh user sebelumnya. Selain itu, user juga dapat melihat hasil klasifikasi yang dilakukan oleh sistem, hasil klasifikasi terletak dipaling kanan dalam *dataframe* yang ditampilkan. Data yang ditampilkan berupa *dataframe*, *dataframe* tersebut dapat digeser untuk mengetahui keseluruhannya, atau dapat menekan tombol *expand* yang berada disamping *dataframe*.

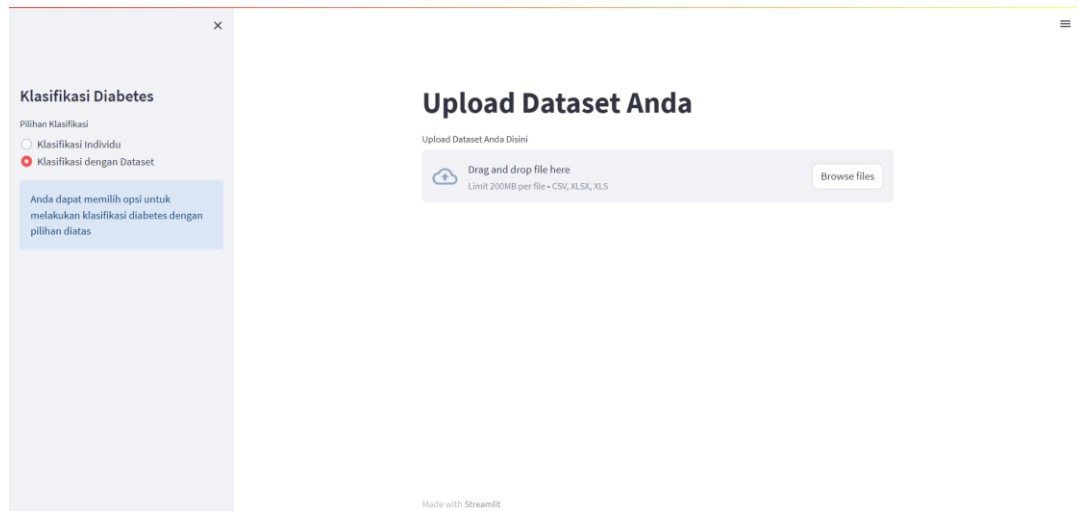
	Name	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Prediksi
0	Saputra	0	100	70	30	100	30	0.2	21	Tidak Diabetes

Gambar 4.16 Contoh Hasil *expand* dari *dataframe*

Gambar diatas merupakan hasil *expand* dari *dataframe* yang ditampilkan oleh website sehingga user dapat melihat keseluruhan kolom dari *dataframe* tanpa menggeser-geser *dataframe* tersebut.

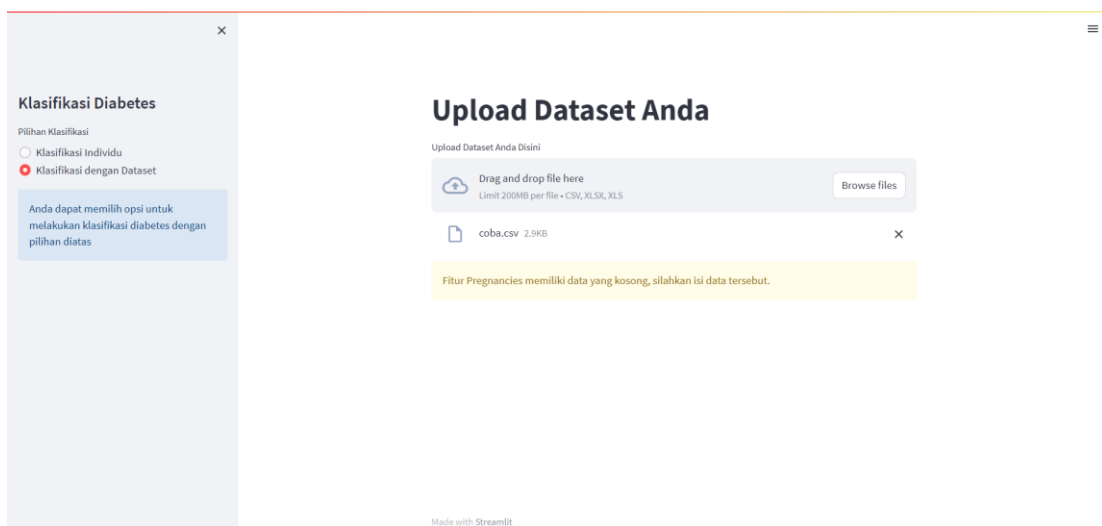
4.7.2 Klasifikasi *Dataset*

Pada menu klasifikasi diabetes dengan *dataset*, akan dilakukan sebuah klasifikasi diabetes lebih dari 1 orang atau lebih dari 1 sampel. User akan diminta untuk mengunggah *dataset* yang dibutuhkan untuk melakukan klasifikasi diabetes. *Dataset* yang diunggah harus sesuai dengan ekstensi file yang ditentukan oleh website (.csv, .xls, .xlsx).



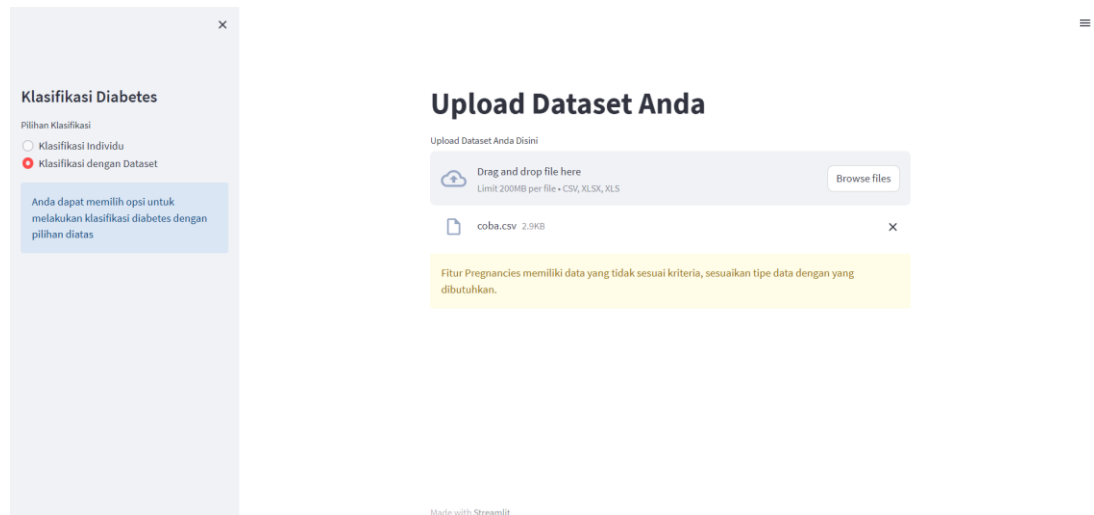
Gambar 4.17 Tampilan Menu Klasifikasi Dengan *Dataset*

Gambar diatas merupakan tampilan menu klasifikasi diabetes dengan *dataset*, dapat dilihat user diminta untuk mengunggah file dari *dataset* yang akan diklasifikasi. File yang diupload harus berekstensi .csv, .xls, .xlsx. File yang diupload tidak boleh memiliki nilai yang hilang (null), dan harus sesuai dengan kriteria, sebagai contoh pada fitur *Pregnancies* harus bernilai *integer*.



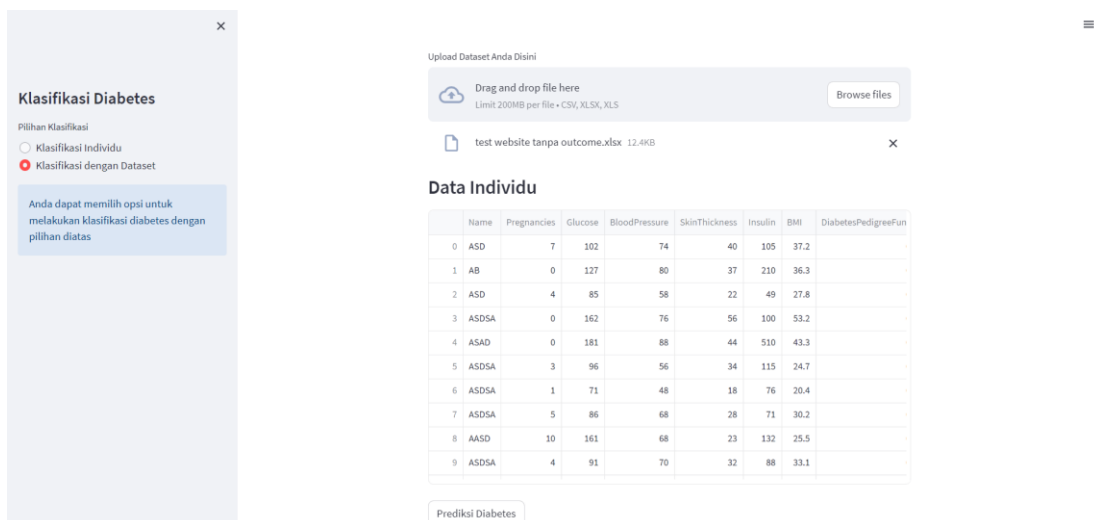
Gambar 4.18 Tampilan *Error Unggah Dataset Dengan Nilai Null*

Gambar diatas merupakan tampilan menu klasifikasi diabetes dengan *dataset* jika user mengunggah *dataset* yang memiliki nilai yang hilang atau *null*. User akan diminta untuk mengunggah *dataset* kembali yang sesuai dengan kriteria yang sudah ditentukan.



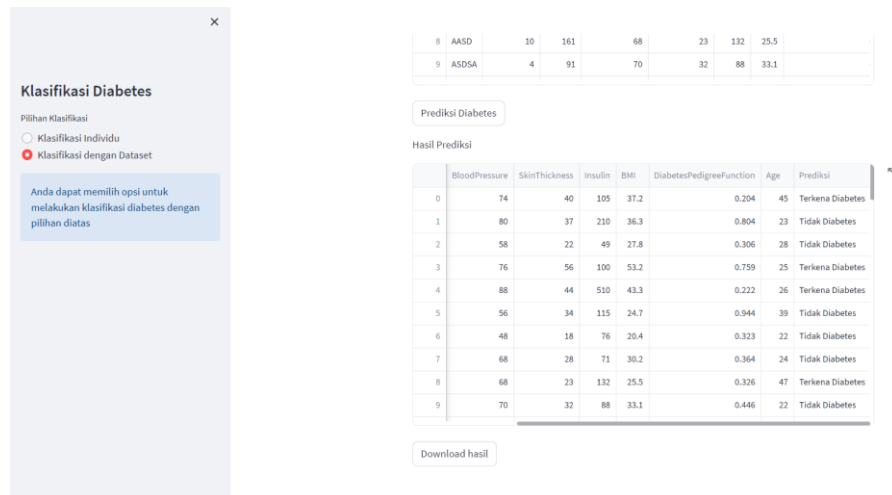
Gambar 4.19 Tampilan *Error Unggah Dataset* Tidak Sesuai Kriteria

Gambar diatas merupakan tampilan menu klasifikasi diabetes dengan *dataset* jika user menunggah *dataset* yang memiliki nilai yang tidak sesuai dengan tipe data yang dibutuhkan, sebagai contoh pada error diatas, pada fitur *Pregnancies* kemungkinan terdapat nilai selain tipe data *integer* pada fitur tersebut.



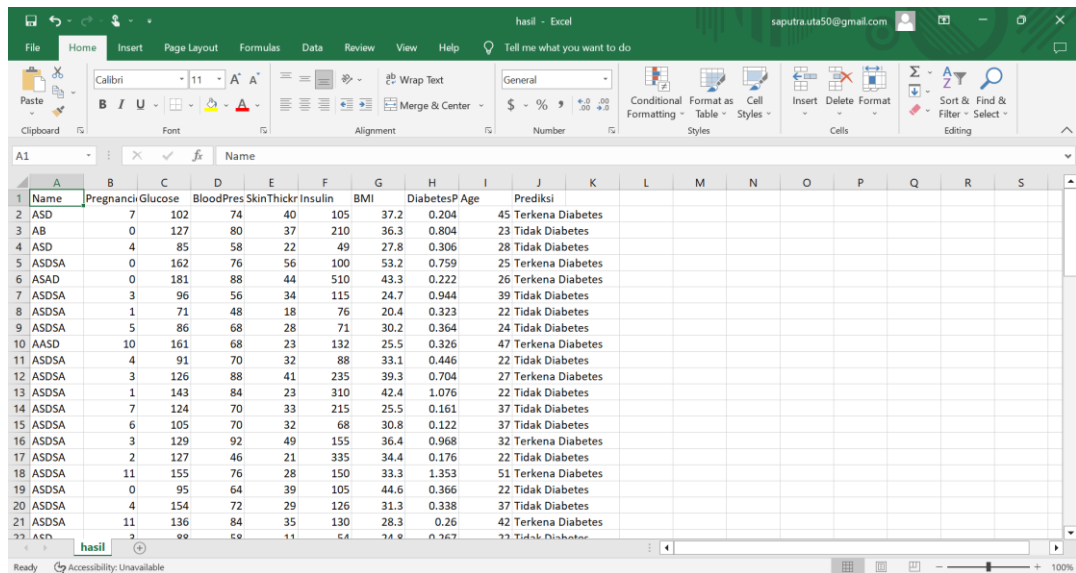
Gambar 4.20 Tampilan Website Usai Berhasil Mengunggah *Dataset*

Gambar diatas merupakan tampilan website klasifikasi diabetes menggunakan *dataset* usai mengunggah *dataset* yang sesuai dengan kriteria. User dapat melihat data yang sudah diunggah dalam bentuk *dataframe*, user juga dapat melakukan *expand* untuk melihat *dataset*. Setelah dapat melihat tampilan seperti ini, user dapat menekan tombol “Prediksi Diabetes” untuk melakukan klasifikasi terhadap *dataset* tersebut.



Gambar 4.21 Tampilan Website Setelah Melakukan Klasifikasi

Gambar diatas merupakan tampilan website klasifikasi diabetes dengan menu klasifikasi menggunakan *dataset* setelah melakukan klasifikasi. Dapat dilihat hasil klasifikasi yang dilakukan oleh sistem ditampilkan ke halaman web, user dapat melihat hasil klasifikasi terletak dipaling kanan dari *dataframe* yang ditampilkan oleh sistem. User juga dapat mengunduh file hasil klasifikasi dalam bentuk csv dengan menekan tombol “Download Hasil”.



Gambar 4.22 File Csv Hasil Klasifikasi Diabetes

Gambar diatas merupakan file csv yang sudah diunduh melalui website klasifikasi diabetes dengan menekan tombol “Download Hasil”, file yang diunduh akan tersimpan dengan nama “hasil.csv”.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang sudah dilakukan diperoleh beberapa kesimpulan yaitu sebagai berikut:

1. Percobaan menggunakan 313 data latih dan 79 data uji. Didapatkan performa model meningkat jika fiturnya dikurangi dengan menggunakan teknik reduksi fitur baik PCA ataupun SFS. Performa metode PCA-SVM terbaik didapatkan pada *principal component* berjumlah 4, sedangkan performa metode SFS-SVM terbaik didapatkan pada hasil seleksi fitur berjumlah 6.
2. Didapatkan hasil yang cukup baik dari kedua buah metode yang digunakan untuk membuat model. Metode SFS-SVM mencapai performa terbaik dengan akurasi 77.22%, presisi 86.54%, *recall* 80.36%, *specificity* 69.57%, dan *F1-Score* 83.33%, sedangkan performa metode PCA-SVM mencapai performa terbaik dengan akurasi 75.95%, presisi 88.46%, *recall* 77.97%, *specificity* 70.00%, dan *F1-Score* 82.88%.
3. Didapatkan perbandingan bahwa model SFS-SVM lebih baik dalam mengklasifikasikan individu yang tidak terkena diabetes, sedangkan model PCA-SVM lebih baik dalam mengklasifikasikan individu yang terkena diabetes.

5.2 Saran

Dengan dilakukannya penelitian ini, masih terdapat beberapa kekurangan dan potensi yang dapat dikembangkan dari website ini sebagai referensi yang dapat digunakan untuk penelitian selanjutnya, adapun saran yang dapat disampaikan oleh penulis adalah sebagai berikut:

1. Menggunakan *dataset* yang memiliki korelasi yang tinggi antara fitur dengan fitur-fitur lainnya. Sehingga mendapatkan model yang lebih baik.
2. Pada penelitian selanjutnya dapat menggunakan algoritma *machine learning* yang lain untuk dikombinasikan dengan teknik reduksi dimensi sehingga dapat diperoleh nilai akurasi, presisi, *recall*, *specificity*, dan *F1-Score* yang lebih baik.
3. Menambahkan lebih banyak fitur pada website yang dibangun.

DAFTAR PUSTAKA

- Agustini., & Kurniawan W.J. (2019). Sistem *E-Learning* Do'a dan Iqro' dalam Peningkatan Proses Pembelajaran pada TK Amal Ikhlas. *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi (JMApTeKsi)*. *Jurnal Mahasiswa*, 1(3), 155.
- Astuti, Y.P., Alexander, Nathaniel., Subhiyakto, E.R., & Kartikadarma, Etika (2022). Analisis Sentimen Menggunakan Metode *Naïve Bayes* Untuk Mengetahui Respon Masyarakat Terhadap Vaksinasi. Semarang. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, Vol.4, No.02, 182.
- Clinton, R. M. R., & Sengkey, R. (2019). Purwarupa Sistem Daftar Pelanggaran Lalulintas Berbasis Mini-Komputer *Raspberry Pi*. *Jurnal Teknik Elektro Dan Komputer*, 8(3), 183.
- Firliana, Rina., Wulaningrum, Resti., & Sasongko, Wisnu. (2015). Implementasi *Principal Component Analysis* (PCA) Untuk Pengenalan Wajah Manusia. *Nusantara of Engineering*, Vol. 2, No.1, 66.
- Godara, J., & Aron, R. (2021). *Support Vector Machine Classifier with Principal Component Analysis and K Mean for Sarcasm Detection*. *7th International Conference on Advanced Computing and Communication Systems (ICACCS)*. doi: 10.1109/ICACCS51430.2021.9442033.
- Imron, Ali. (2019). Analisis Sentimen Terhadap Tempat Wisata Di Kabupaten Rembang Menggunakan Metode *Naive Bayes Classifier*. Yogyakarta. Universitas Islam Indonesia. *DSpace Repository*.
- Kabir, E., Shahid, M., & Rokibul, M. (2017). *Developing Diabetes Disease Classification Model using Sequential Forward Selection Algorithm*. *International Journal of Computer Applications*, 180(5), 1–6. <https://doi.org/10.5120/ijca2017916018>
- Lufiana, Fanni. (2021). *Klasifikasi Pendertia Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine* [Skripsi, Universitas Lampung]. *Digital Repository UNILA*.
- Monika, I. P., & Furqon, M. T. (2018). Penerapan Metode *Support Vector Machine* (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. Malang.

- Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3165-3166.
- Mucholladin, A. W., Bachtiar, F. A., & Furqon, M. T. (2021). Klasifikasi Penyakit Diabetes Menggunakan Metode *Support Vector Machine*. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(2), 624.
- Mohan, N., & Jain, V. (2020). *Performance Analysis of Support Vector Machine in Diabetes Prediction. Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020*. <https://doi.org/10.1109/ICECA49313.2020.9297411>
- Nugroho, M. F., & Wibowo, S. (2017). Fitur Seleksi *Forward Selection* Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma *Naïve Bayes*. *Jurnal Informatika Upgris*, 3(1), 66. <https://doi.org/10.26877/jiu.v3i1.1669>.
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan *Machine Learning* dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 76. <https://doi.org/10.31294/ijcit.v5i1.7951>.
- Sihombing, P. R., & Arsani, A. M. (2021). *Comparison Of Machine Learning Methods in Classifying Poverty in Indonesia in 2018. Jurnal Teknik Informatika (Jutif)*, 2(1), 54. <https://doi.org/10.20884/1.jutif.2021.2.1.52>
- Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., & Johannes, R. S. (1988). *Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. Proceedings - Annual Symposium on Computer Applications in Medical Care*, 261–265.
- Sulistiana. (2020). *Optimasi Support Vector Machine (SVM) Menggunakan Grid Search dan Unigram Guna Meningkatkan Akurasi Review Pada Situs E-Commerce* [Skripsi, Universitas Negeri Semarang]. UNNES Repository.
- Ramadani, S., Ayu, N. Z. S., Nurhayati, N., Azzahra, F., & Windarto, A. P. (2020). Analisis *Data Mining Naïve Bayes* Klasifikasi Pada Kelayakan Penerima PKH. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 4(1), 375.
- Toharin, S. N. R., CAHYATI, S., M Kes, W. H., & Kes, Z. M. H. (2015). Hubungan Modifikasi Gaya Hidup Dan Kepatuhan Konsumsi Obat Antidiabetik Dengan

Kadar Gula Darah Pada Penderita Diabetes Melitus Tipe 2 Di RS Qim Batang Tahun 2013. *Unnes Journal of Public Health*, 4(2), 154.

Widi Hastomo, Nur Aini, Adhitio Satyo Bayangkari Karno, & L.M. Rasdi Rere. (2022). Metode Pembelajaran Mesin untuk Memprediksi Emisi *Manure Management*. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 11(2), 134. <https://doi.org/10.22146/jnteti.v11i2.2586>

LAMPIRAN

1. Lampiran I

Data latih yang digunakan dalam penelitian:

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age	Target
0.00	180.00	90.00	26.00	90.00	36.50	0.31	35.00	1.00
3.00	81.00	86.00	16.00	66.00	27.50	0.31	22.00	0.00
8.00	126.00	88.00	36.00	108.00	38.50	0.35	49.00	0.00
2.00	96.00	68.00	13.00	49.00	21.10	0.65	26.00	0.00
3.00	80.00	82.00	31.00	70.00	34.20	1.29	27.00	1.00
0.00	120.00	74.00	18.00	63.00	30.50	0.29	26.00	0.00
2.00	157.00	74.00	35.00	440.00	39.40	0.13	30.00	0.00
3.00	102.00	44.00	20.00	94.00	30.80	0.40	26.00	0.00
4.00	84.00	90.00	23.00	56.00	39.50	0.16	25.00	0.00
1.00	95.00	82.00	25.00	180.00	35.00	0.23	43.00	1.00
2.00	108.00	62.00	10.00	278.00	25.30	0.88	22.00	0.00
8.00	186.00	90.00	35.00	225.00	34.50	0.42	37.00	1.00
2.00	121.00	70.00	32.00	95.00	39.10	0.89	23.00	0.00
8.00	74.00	70.00	40.00	49.00	35.30	0.71	39.00	0.00
5.00	126.00	78.00	27.00	22.00	29.60	0.44	40.00	0.00
0.00	173.00	78.00	32.00	265.00	46.50	1.16	58.00	0.00
1.00	143.00	74.00	22.00	61.00	26.20	0.26	21.00	0.00
1.00	109.00	56.00	21.00	135.00	25.20	0.83	23.00	0.00
5.00	109.00	62.00	41.00	129.00	35.80	0.51	25.00	1.00
6.00	129.00	90.00	7.00	326.00	19.60	0.58	60.00	0.00
4.00	103.00	60.00	33.00	192.00	24.00	0.97	33.00	0.00
0.00	93.00	60.00	25.00	92.00	28.70	0.53	22.00	0.00
3.00	99.00	80.00	11.00	64.00	19.30	0.28	30.00	0.00
1.00	103.00	30.00	38.00	83.00	43.30	0.18	33.00	0.00
9.00	171.00	110.00	24.00	240.00	45.40	0.72	54.00	1.00
1.00	119.00	44.00	47.00	63.00	35.50	0.28	25.00	0.00
10.00	68.00	106.00	23.00	49.00	35.50	0.29	47.00	0.00
8.00	167.00	106.00	46.00	231.00	37.60	0.17	43.00	1.00
2.00	146.00	76.00	35.00	194.00	38.20	0.33	29.00	0.00
12.00	88.00	74.00	40.00	54.00	35.30	0.38	48.00	0.00
1.00	84.00	64.00	23.00	115.00	36.90	0.47	28.00	0.00
2.00	99.00	60.00	17.00	160.00	36.60	0.45	21.00	0.00
2.00	95.00	54.00	14.00	88.00	26.10	0.75	22.00	0.00
1.00	117.00	88.00	24.00	145.00	34.50	0.40	40.00	1.00
0.00	114.00	80.00	34.00	285.00	44.20	0.17	27.00	0.00
1.00	181.00	78.00	42.00	293.00	40.00	1.26	22.00	1.00
2.00	129.00	74.00	26.00	205.00	33.20	0.59	25.00	0.00
0.00	135.00	68.00	42.00	250.00	42.30	0.37	24.00	1.00
2.00	124.00	68.00	28.00	205.00	32.90	0.88	30.00	1.00
1.00	87.00	78.00	27.00	32.00	34.60	0.10	22.00	0.00
4.00	111.00	72.00	47.00	207.00	37.10	1.39	56.00	1.00

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age	Target
2.00	68.00	62.00	13.00	15.00	20.10	0.26	23.00	0.00
7.00	150.00	66.00	42.00	342.00	34.70	0.72	42.00	0.00
5.00	96.00	74.00	18.00	67.00	33.60	1.00	43.00	0.00
1.00	149.00	68.00	29.00	127.00	29.30	0.35	42.00	1.00
1.00	90.00	62.00	12.00	43.00	27.20	0.58	24.00	0.00
2.00	100.00	54.00	28.00	105.00	37.80	0.50	24.00	0.00
1.00	109.00	60.00	8.00	182.00	25.40	0.95	21.00	0.00
0.00	124.00	56.00	13.00	105.00	21.80	0.45	21.00	0.00
0.00	128.00	68.00	19.00	180.00	30.50	1.39	25.00	1.00
6.00	165.00	68.00	26.00	168.00	33.60	0.63	49.00	0.00
2.00	93.00	64.00	32.00	160.00	38.00	0.67	23.00	1.00
4.00	147.00	74.00	25.00	293.00	34.90	0.39	30.00	0.00
0.00	165.00	90.00	33.00	680.00	52.30	0.43	23.00	0.00
1.00	89.00	24.00	19.00	25.00	27.80	0.56	21.00	0.00
0.00	135.00	94.00	46.00	145.00	40.60	0.28	26.00	0.00
0.00	78.00	88.00	29.00	40.00	36.90	0.43	21.00	0.00
5.00	77.00	82.00	41.00	42.00	35.80	0.16	35.00	0.00
8.00	124.00	76.00	24.00	600.00	28.70	0.69	52.00	1.00
2.00	100.00	68.00	25.00	71.00	38.50	0.32	26.00	0.00
1.00	88.00	62.00	24.00	44.00	29.90	0.42	23.00	0.00
3.00	173.00	84.00	33.00	474.00	35.70	0.26	22.00	1.00
0.00	180.00	78.00	63.00	14.00	59.40	2.42	25.00	1.00
1.00	100.00	74.00	12.00	46.00	19.50	0.15	28.00	0.00
1.00	118.00	58.00	36.00	94.00	33.30	0.26	23.00	0.00
0.00	117.00	80.00	31.00	53.00	45.20	0.09	24.00	0.00
7.00	187.00	50.00	33.00	392.00	33.90	0.83	34.00	1.00
0.00	165.00	76.00	43.00	255.00	47.90	0.26	26.00	0.00
0.00	118.00	84.00	47.00	230.00	45.80	0.55	31.00	1.00
4.00	131.00	68.00	21.00	166.00	33.10	0.16	28.00	0.00
4.00	90.00	88.00	47.00	54.00	37.70	0.36	29.00	0.00
11.00	143.00	94.00	33.00	146.00	36.60	0.25	51.00	1.00
2.00	127.00	58.00	24.00	275.00	27.70	1.60	25.00	0.00
12.00	151.00	70.00	40.00	271.00	41.80	0.74	38.00	1.00
3.00	111.00	90.00	12.00	78.00	28.40	0.50	29.00	0.00
1.00	181.00	64.00	30.00	180.00	34.10	0.33	38.00	1.00
10.00	148.00	84.00	48.00	237.00	37.60	1.00	51.00	1.00
0.00	102.00	78.00	40.00	90.00	34.50	0.24	24.00	0.00
0.00	138.00	60.00	35.00	167.00	34.60	0.53	21.00	1.00
9.00	145.00	80.00	46.00	130.00	37.90	0.64	40.00	1.00
1.00	128.00	82.00	17.00	183.00	27.50	0.12	22.00	0.00
1.00	128.00	98.00	41.00	58.00	32.00	1.32	33.00	1.00
8.00	126.00	74.00	38.00	75.00	25.90	0.16	39.00	0.00
2.00	144.00	58.00	33.00	135.00	31.60	0.42	25.00	1.00
3.00	99.00	54.00	19.00	86.00	25.60	0.15	24.00	0.00
9.00	123.00	70.00	44.00	94.00	33.10	0.37	40.00	0.00
4.00	154.00	62.00	31.00	284.00	32.80	0.24	23.00	0.00

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age	Target
1.00	193.00	50.00	16.00	375.00	25.90	0.66	24.00	0.00
4.00	148.00	60.00	27.00	318.00	30.90	0.15	29.00	1.00
6.00	151.00	62.00	31.00	120.00	35.50	0.69	28.00	0.00
2.00	98.00	60.00	17.00	120.00	34.70	0.20	22.00	0.00
2.00	99.00	70.00	16.00	44.00	20.40	0.24	27.00	0.00
2.00	83.00	66.00	23.00	50.00	32.20	0.50	22.00	0.00
2.00	108.00	52.00	26.00	63.00	32.50	0.32	22.00	0.00
8.00	176.00	90.00	34.00	300.00	33.70	0.47	58.00	1.00
15.00	136.00	70.00	32.00	110.00	37.10	0.15	43.00	1.00
2.00	146.00	70.00	38.00	360.00	28.00	0.34	29.00	1.00
12.00	92.00	62.00	7.00	258.00	27.60	0.93	44.00	1.00
5.00	105.00	72.00	29.00	325.00	36.90	0.16	28.00	0.00
1.00	117.00	60.00	23.00	106.00	33.80	0.47	27.00	0.00
0.00	134.00	58.00	20.00	291.00	26.40	0.35	21.00	0.00
3.00	106.00	54.00	21.00	158.00	30.90	0.29	24.00	0.00
1.00	92.00	62.00	25.00	41.00	19.50	0.48	25.00	0.00
1.00	121.00	78.00	39.00	74.00	39.00	0.26	28.00	0.00
3.00	191.00	68.00	15.00	130.00	30.90	0.30	34.00	0.00
14.00	100.00	78.00	25.00	184.00	36.60	0.41	46.00	1.00
0.00	104.00	64.00	37.00	64.00	33.60	0.51	22.00	1.00
0.00	129.00	110.00	46.00	130.00	67.10	0.32	26.00	1.00
1.00	109.00	58.00	18.00	116.00	28.50	0.22	22.00	0.00
3.00	100.00	68.00	23.00	81.00	31.60	0.95	28.00	0.00
0.00	179.00	50.00	36.00	159.00	37.80	0.46	22.00	1.00
1.00	90.00	62.00	18.00	59.00	25.10	1.27	25.00	0.00
3.00	158.00	64.00	13.00	387.00	31.20	0.30	24.00	0.00
3.00	111.00	58.00	31.00	44.00	29.50	0.43	22.00	0.00
1.00	133.00	102.00	28.00	140.00	32.80	0.23	45.00	1.00
1.00	119.00	54.00	13.00	50.00	22.30	0.21	24.00	0.00
3.00	128.00	72.00	25.00	190.00	32.40	0.55	27.00	1.00
1.00	80.00	74.00	11.00	60.00	30.00	0.53	22.00	0.00
3.00	173.00	78.00	39.00	185.00	33.80	0.97	31.00	1.00
2.00	100.00	70.00	52.00	57.00	40.50	0.68	25.00	0.00
4.00	123.00	80.00	15.00	176.00	32.00	0.44	34.00	0.00
10.00	125.00	70.00	26.00	115.00	31.10	0.21	41.00	1.00
1.00	120.00	80.00	48.00	200.00	38.90	1.16	41.00	0.00
0.00	84.00	64.00	22.00	66.00	35.80	0.55	21.00	0.00
1.00	164.00	82.00	43.00	67.00	32.80	0.34	50.00	0.00
6.00	134.00	80.00	37.00	370.00	46.20	0.24	46.00	1.00
5.00	189.00	64.00	33.00	325.00	31.20	0.58	29.00	1.00
1.00	106.00	70.00	28.00	135.00	34.20	0.14	22.00	0.00
1.00	86.00	66.00	52.00	65.00	41.30	0.92	29.00	0.00
6.00	119.00	50.00	22.00	176.00	27.10	1.32	33.00	1.00
2.00	122.00	76.00	27.00	200.00	35.90	0.48	26.00	0.00
2.00	122.00	60.00	18.00	106.00	29.80	0.72	22.00	0.00
2.00	106.00	56.00	27.00	165.00	29.00	0.43	22.00	0.00

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age	Target
5.00	187.00	76.00	27.00	207.00	43.60	1.03	53.00	1.00
1.00	144.00	82.00	46.00	180.00	46.10	0.34	46.00	1.00
2.00	102.00	86.00	36.00	120.00	45.50	0.13	23.00	1.00
3.00	158.00	70.00	30.00	328.00	35.50	0.34	35.00	1.00
1.00	119.00	86.00	39.00	220.00	45.60	0.81	29.00	1.00
0.00	106.00	70.00	37.00	148.00	39.40	0.61	22.00	0.00
1.00	100.00	66.00	29.00	196.00	32.00	0.44	42.00	0.00
5.00	144.00	82.00	26.00	285.00	32.00	0.45	58.00	1.00
1.00	81.00	74.00	41.00	57.00	46.30	1.10	32.00	0.00
3.00	78.00	50.00	32.00	88.00	31.00	0.25	26.00	1.00
5.00	136.00	84.00	41.00	88.00	35.00	0.29	35.00	1.00
7.00	181.00	84.00	21.00	192.00	35.90	0.59	51.00	1.00
0.00	95.00	80.00	45.00	92.00	36.50	0.33	26.00	0.00
0.00	91.00	68.00	32.00	210.00	39.90	0.38	25.00	0.00
0.00	100.00	88.00	60.00	110.00	46.80	0.96	31.00	0.00
9.00	124.00	70.00	33.00	402.00	35.40	0.28	34.00	0.00
0.00	94.00	70.00	27.00	115.00	43.50	0.35	21.00	0.00
4.00	117.00	64.00	27.00	120.00	33.20	0.23	24.00	0.00
1.00	114.00	66.00	36.00	200.00	38.10	0.29	21.00	0.00
0.00	139.00	62.00	17.00	210.00	22.10	0.21	21.00	0.00
0.00	102.00	86.00	17.00	105.00	29.30	0.70	27.00	0.00
2.00	155.00	52.00	27.00	540.00	38.70	0.24	25.00	1.00
2.00	108.00	62.00	32.00	56.00	25.20	0.13	21.00	0.00
2.00	99.00	52.00	15.00	94.00	24.60	0.64	21.00	0.00
7.00	168.00	88.00	42.00	321.00	38.20	0.79	40.00	1.00
1.00	95.00	66.00	13.00	38.00	19.60	0.33	25.00	0.00
1.00	79.00	60.00	42.00	48.00	43.50	0.68	23.00	0.00
2.00	94.00	76.00	18.00	66.00	31.60	0.65	23.00	0.00
7.00	142.00	60.00	33.00	190.00	28.80	0.69	61.00	0.00
1.00	116.00	78.00	29.00	180.00	36.10	0.50	25.00	0.00
1.00	139.00	62.00	41.00	480.00	40.70	0.54	21.00	0.00
7.00	133.00	88.00	15.00	155.00	32.40	0.26	37.00	0.00
3.00	173.00	82.00	48.00	465.00	38.40	2.14	25.00	1.00
3.00	84.00	68.00	30.00	106.00	31.90	0.59	25.00	0.00
1.00	122.00	90.00	51.00	220.00	49.70	0.33	31.00	1.00
0.00	126.00	86.00	27.00	120.00	27.40	0.52	21.00	0.00
0.00	100.00	70.00	26.00	50.00	30.80	0.60	21.00	0.00
6.00	99.00	60.00	19.00	54.00	26.90	0.50	32.00	0.00
8.00	179.00	72.00	42.00	130.00	32.70	0.72	36.00	1.00
2.00	105.00	80.00	45.00	191.00	33.70	0.71	29.00	1.00
2.00	105.00	58.00	40.00	94.00	34.90	0.23	25.00	0.00
2.00	106.00	64.00	35.00	119.00	30.50	1.40	34.00	0.00
3.00	115.00	66.00	39.00	140.00	38.10	0.15	28.00	0.00
1.00	97.00	66.00	15.00	140.00	23.20	0.49	22.00	0.00
2.00	110.00	74.00	29.00	125.00	32.40	0.70	27.00	0.00
2.00	141.00	58.00	34.00	128.00	25.40	0.70	24.00	0.00

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age	Target
1.00	130.00	70.00	13.00	105.00	25.90	0.47	22.00	0.00
1.00	128.00	48.00	45.00	194.00	40.50	0.61	24.00	1.00
0.00	152.00	82.00	39.00	272.00	41.50	0.27	27.00	0.00
2.00	68.00	70.00	32.00	66.00	25.00	0.19	25.00	0.00
3.00	89.00	74.00	16.00	85.00	30.40	0.55	38.00	0.00
4.00	109.00	64.00	44.00	99.00	34.80	0.91	26.00	1.00
8.00	155.00	62.00	26.00	495.00	34.00	0.54	46.00	1.00
0.00	97.00	64.00	36.00	100.00	36.80	0.60	25.00	0.00
1.00	99.00	72.00	30.00	18.00	38.60	0.41	21.00	0.00
1.00	143.00	86.00	30.00	330.00	30.10	0.89	23.00	0.00
1.00	196.00	76.00	36.00	249.00	36.50	0.88	29.00	1.00
6.00	144.00	72.00	27.00	228.00	33.90	0.26	40.00	0.00
2.00	82.00	52.00	22.00	115.00	28.50	1.70	25.00	0.00
1.00	172.00	68.00	49.00	579.00	42.40	0.70	28.00	1.00
7.00	195.00	70.00	33.00	145.00	25.10	0.16	55.00	1.00
7.00	97.00	76.00	32.00	91.00	40.90	0.87	32.00	1.00
1.00	189.00	60.00	23.00	846.00	30.10	0.40	59.00	1.00
2.00	117.00	90.00	19.00	71.00	25.20	0.31	21.00	0.00
2.00	56.00	56.00	28.00	45.00	24.20	0.33	22.00	0.00
1.00	96.00	64.00	27.00	87.00	33.20	0.29	21.00	0.00
1.00	100.00	72.00	12.00	70.00	25.30	0.66	28.00	0.00
5.00	123.00	74.00	40.00	77.00	34.10	0.27	28.00	0.00
1.00	119.00	88.00	41.00	170.00	45.30	0.51	26.00	0.00
1.00	95.00	74.00	21.00	73.00	25.90	0.67	36.00	0.00
6.00	103.00	72.00	32.00	190.00	37.70	0.32	55.00	0.00
0.00	102.00	64.00	46.00	78.00	40.60	0.50	21.00	0.00
4.00	173.00	70.00	14.00	168.00	29.70	0.36	33.00	1.00
4.00	144.00	58.00	28.00	140.00	29.50	0.29	37.00	0.00
5.00	139.00	64.00	35.00	140.00	28.60	0.41	26.00	0.00
3.00	124.00	80.00	33.00	130.00	33.20	0.31	26.00	0.00
1.00	131.00	64.00	14.00	415.00	23.70	0.39	21.00	0.00
10.00	101.00	76.00	48.00	180.00	32.90	0.17	63.00	0.00
0.00	105.00	64.00	41.00	142.00	41.50	0.17	22.00	0.00
2.00	101.00	58.00	17.00	265.00	24.20	0.61	23.00	0.00
1.00	82.00	64.00	13.00	95.00	21.20	0.42	23.00	0.00
2.00	84.00	50.00	23.00	76.00	30.40	0.97	21.00	0.00
0.00	198.00	66.00	32.00	274.00	41.30	0.50	28.00	1.00
1.00	157.00	72.00	21.00	168.00	25.60	0.12	24.00	0.00
1.00	91.00	54.00	25.00	100.00	25.20	0.23	23.00	0.00
5.00	108.00	72.00	43.00	75.00	36.10	0.26	33.00	0.00
3.00	170.00	64.00	37.00	225.00	34.50	0.36	30.00	1.00
5.00	155.00	84.00	44.00	545.00	38.70	0.62	34.00	0.00
1.00	71.00	78.00	50.00	45.00	33.20	0.42	21.00	0.00
1.00	115.00	70.00	30.00	96.00	34.60	0.53	32.00	1.00
2.00	125.00	60.00	20.00	140.00	33.80	0.09	31.00	0.00
6.00	108.00	44.00	20.00	130.00	24.00	0.81	35.00	0.00

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age	Target
3.00	158.00	76.00	36.00	245.00	31.60	0.85	28.00	1.00
3.00	113.00	50.00	10.00	85.00	29.50	0.63	25.00	0.00
2.00	142.00	82.00	18.00	64.00	24.70	0.76	21.00	0.00
3.00	169.00	74.00	19.00	125.00	29.90	0.27	31.00	1.00
11.00	138.00	74.00	26.00	144.00	36.10	0.56	50.00	1.00
8.00	151.00	78.00	32.00	210.00	42.90	0.52	36.00	1.00
1.00	88.00	78.00	29.00	76.00	32.00	0.37	29.00	0.00
2.00	101.00	58.00	35.00	90.00	21.80	0.16	22.00	0.00
9.00	120.00	72.00	22.00	56.00	20.80	0.73	48.00	0.00
1.00	112.00	72.00	30.00	176.00	34.40	0.53	25.00	0.00
3.00	129.00	64.00	29.00	115.00	26.40	0.22	28.00	1.00
0.00	121.00	66.00	30.00	165.00	34.30	0.20	33.00	1.00
7.00	129.00	68.00	49.00	125.00	38.50	0.44	43.00	1.00
2.00	83.00	65.00	28.00	66.00	36.80	0.63	24.00	0.00
13.00	153.00	88.00	37.00	140.00	40.60	1.17	39.00	0.00
7.00	142.00	90.00	24.00	480.00	30.40	0.13	43.00	1.00
2.00	112.00	78.00	50.00	140.00	39.40	0.18	24.00	0.00
3.00	116.00	74.00	15.00	105.00	26.30	0.11	24.00	0.00
4.00	116.00	72.00	12.00	87.00	22.10	0.46	37.00	0.00
13.00	152.00	90.00	33.00	29.00	26.80	0.73	43.00	1.00
6.00	93.00	50.00	30.00	64.00	28.70	0.36	23.00	0.00
4.00	184.00	78.00	39.00	277.00	37.00	0.26	31.00	1.00
1.00	87.00	60.00	37.00	75.00	37.20	0.51	22.00	0.00
7.00	150.00	78.00	29.00	126.00	35.20	0.69	54.00	1.00
1.00	130.00	60.00	23.00	170.00	28.60	0.69	21.00	0.00
6.00	125.00	68.00	30.00	120.00	30.00	0.46	32.00	0.00
7.00	160.00	54.00	32.00	175.00	30.50	0.59	39.00	1.00
0.00	137.00	40.00	35.00	168.00	43.10	2.29	33.00	1.00
0.00	95.00	85.00	25.00	36.00	37.40	0.25	24.00	1.00
6.00	104.00	74.00	18.00	156.00	29.90	0.72	41.00	1.00
2.00	123.00	48.00	32.00	165.00	42.10	0.52	26.00	0.00
7.00	83.00	78.00	26.00	71.00	29.30	0.77	36.00	0.00
3.00	163.00	70.00	18.00	105.00	31.60	0.27	28.00	1.00
5.00	139.00	80.00	35.00	160.00	31.60	0.36	25.00	1.00
1.00	89.00	76.00	34.00	37.00	31.20	0.19	23.00	0.00
9.00	145.00	88.00	34.00	165.00	30.30	0.77	53.00	1.00
6.00	154.00	78.00	41.00	140.00	46.10	0.57	27.00	0.00
13.00	145.00	82.00	19.00	110.00	22.20	0.25	57.00	0.00
3.00	174.00	58.00	22.00	194.00	32.90	0.59	36.00	1.00
8.00	181.00	68.00	36.00	495.00	30.10	0.62	60.00	1.00
1.00	77.00	56.00	30.00	56.00	33.30	1.25	24.00	0.00
2.00	87.00	58.00	16.00	52.00	32.70	0.17	25.00	0.00
3.00	103.00	72.00	30.00	152.00	27.60	0.73	27.00	0.00
3.00	187.00	70.00	22.00	200.00	36.40	0.41	36.00	1.00
7.00	136.00	74.00	26.00	135.00	26.00	0.65	51.00	0.00
0.00	188.00	82.00	14.00	185.00	32.00	0.68	22.00	1.00

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age	Target
2.00	174.00	88.00	37.00	120.00	44.50	0.65	24.00	1.00
2.00	112.00	86.00	42.00	160.00	38.40	0.25	28.00	0.00
7.00	114.00	76.00	17.00	110.00	23.80	0.47	31.00	0.00
4.00	99.00	76.00	15.00	51.00	23.20	0.22	21.00	0.00
3.00	171.00	72.00	33.00	135.00	33.30	0.20	24.00	1.00
3.00	99.00	62.00	19.00	74.00	21.80	0.28	26.00	0.00
0.00	98.00	82.00	15.00	84.00	25.20	0.30	22.00	0.00
2.00	88.00	58.00	26.00	16.00	28.40	0.77	22.00	0.00
2.00	122.00	52.00	43.00	158.00	36.20	0.82	28.00	0.00
9.00	152.00	78.00	34.00	171.00	34.20	0.89	33.00	1.00
3.00	120.00	70.00	30.00	135.00	42.90	0.45	30.00	0.00
3.00	83.00	58.00	31.00	18.00	34.30	0.34	25.00	0.00
6.00	98.00	58.00	33.00	190.00	34.00	0.43	43.00	0.00
9.00	134.00	74.00	33.00	60.00	25.90	0.46	81.00	0.00
4.00	125.00	70.00	18.00	122.00	28.90	1.14	45.00	1.00
1.00	101.00	50.00	15.00	36.00	24.20	0.53	26.00	0.00
1.00	111.00	62.00	13.00	182.00	24.00	0.14	23.00	0.00
4.00	129.00	60.00	12.00	231.00	27.50	0.53	31.00	0.00
8.00	100.00	74.00	40.00	215.00	39.40	0.66	43.00	1.00
0.00	93.00	100.00	39.00	72.00	43.40	1.02	35.00	0.00
6.00	92.00	62.00	32.00	126.00	32.00	0.09	46.00	0.00
1.00	112.00	80.00	45.00	132.00	34.80	0.22	24.00	0.00
7.00	94.00	64.00	25.00	79.00	33.30	0.74	41.00	0.00
1.00	109.00	38.00	18.00	120.00	23.10	0.41	26.00	0.00
2.00	120.00	76.00	37.00	105.00	39.70	0.22	29.00	0.00
0.00	104.00	64.00	23.00	116.00	27.80	0.45	23.00	0.00
0.00	137.00	68.00	14.00	148.00	24.80	0.14	21.00	0.00
4.00	127.00	88.00	11.00	155.00	34.50	0.60	28.00	0.00
8.00	196.00	76.00	29.00	280.00	37.50	0.61	57.00	1.00
5.00	117.00	86.00	30.00	105.00	39.10	0.25	42.00	0.00
0.00	119.00	64.00	18.00	92.00	34.90	0.73	23.00	0.00
1.00	167.00	74.00	17.00	144.00	23.40	0.45	33.00	1.00
2.00	94.00	68.00	18.00	76.00	26.00	0.56	21.00	0.00
2.00	90.00	80.00	14.00	55.00	24.40	0.25	24.00	0.00
5.00	121.00	72.00	23.00	112.00	26.20	0.25	30.00	0.00
1.00	103.00	80.00	11.00	82.00	19.40	0.49	22.00	0.00
12.00	140.00	82.00	43.00	325.00	39.20	0.53	58.00	1.00
1.00	136.00	74.00	50.00	204.00	37.40	0.40	24.00	0.00
0.00	177.00	60.00	29.00	478.00	34.60	1.07	21.00	1.00
0.00	117.00	66.00	31.00	188.00	30.80	0.49	22.00	0.00
3.00	123.00	100.00	35.00	240.00	57.30	0.88	22.00	0.00
9.00	112.00	82.00	32.00	175.00	34.20	0.26	36.00	1.00

Data uji yang digunakan dalam penelitian:

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age	Target
7.00	102.00	74.00	40.00	105.00	37.20	0.20	45.00	0.00
0.00	127.00	80.00	37.00	210.00	36.30	0.80	23.00	0.00
4.00	85.00	58.00	22.00	49.00	27.80	0.31	28.00	0.00
0.00	162.00	76.00	56.00	100.00	53.20	0.76	25.00	1.00
0.00	181.00	88.00	44.00	510.00	43.30	0.22	26.00	1.00
3.00	96.00	56.00	34.00	115.00	24.70	0.94	39.00	0.00
1.00	71.00	48.00	18.00	76.00	20.40	0.32	22.00	0.00
5.00	86.00	68.00	28.00	71.00	30.20	0.36	24.00	0.00
10.00	161.00	68.00	23.00	132.00	25.50	0.33	47.00	1.00
4.00	91.00	70.00	32.00	88.00	33.10	0.45	22.00	0.00
3.00	126.00	88.00	41.00	235.00	39.30	0.70	27.00	0.00
1.00	143.00	84.00	23.00	310.00	42.40	1.08	22.00	0.00
7.00	124.00	70.00	33.00	215.00	25.50	0.16	37.00	0.00
6.00	105.00	70.00	32.00	68.00	30.80	0.12	37.00	0.00
3.00	129.00	92.00	49.00	155.00	36.40	0.97	32.00	1.00
2.00	127.00	46.00	21.00	335.00	34.40	0.18	22.00	0.00
11.00	155.00	76.00	28.00	150.00	33.30	1.35	51.00	1.00
0.00	95.00	64.00	39.00	105.00	44.60	0.37	22.00	0.00
4.00	154.00	72.00	29.00	126.00	31.30	0.34	37.00	0.00
11.00	136.00	84.00	35.00	130.00	28.30	0.26	42.00	1.00
3.00	88.00	58.00	11.00	54.00	24.80	0.27	22.00	0.00
1.00	128.00	88.00	39.00	110.00	36.50	1.06	37.00	1.00
9.00	154.00	78.00	30.00	100.00	30.90	0.16	45.00	0.00
1.00	89.00	66.00	23.00	94.00	28.10	0.17	21.00	0.00
0.00	107.00	62.00	30.00	74.00	36.60	0.76	25.00	1.00
1.00	79.00	80.00	25.00	37.00	25.40	0.58	22.00	0.00
17.00	163.00	72.00	41.00	114.00	40.90	0.82	47.00	1.00
4.00	129.00	86.00	20.00	270.00	35.10	0.23	23.00	0.00
0.00	140.00	65.00	26.00	130.00	42.60	0.43	24.00	1.00
2.00	88.00	74.00	19.00	53.00	29.00	0.23	22.00	0.00
2.00	155.00	74.00	17.00	96.00	26.60	0.43	27.00	1.00
6.00	154.00	74.00	32.00	193.00	29.30	0.84	39.00	0.00
1.00	87.00	68.00	34.00	77.00	37.60	0.40	24.00	0.00
1.00	140.00	74.00	26.00	180.00	24.10	0.83	23.00	0.00
8.00	109.00	76.00	39.00	114.00	27.90	0.64	31.00	1.00
11.00	120.00	80.00	37.00	150.00	42.30	0.79	48.00	1.00
4.00	96.00	56.00	17.00	49.00	20.80	0.34	26.00	0.00
1.00	88.00	30.00	42.00	99.00	55.00	0.50	26.00	1.00
1.00	153.00	82.00	42.00	485.00	40.60	0.69	23.00	0.00
7.00	187.00	68.00	39.00	304.00	37.70	0.25	41.00	1.00
9.00	156.00	86.00	28.00	155.00	34.30	1.19	42.00	1.00
0.00	126.00	84.00	29.00	215.00	30.70	0.52	24.00	0.00
1.00	108.00	60.00	46.00	178.00	35.50	0.42	24.00	0.00
2.00	128.00	78.00	37.00	182.00	43.30	1.22	31.00	1.00
2.00	112.00	68.00	22.00	94.00	34.10	0.32	26.00	0.00

Pre	Glu	BP	ST	Insulin	BMI	DBF	Age	Target
5.00	158.00	84.00	41.00	210.00	39.40	0.40	29.00	1.00
1.00	126.00	56.00	29.00	152.00	28.70	0.80	21.00	0.00
3.00	176.00	86.00	27.00	156.00	33.30	1.15	52.00	1.00
1.00	100.00	66.00	15.00	56.00	23.60	0.67	26.00	0.00
1.00	125.00	70.00	24.00	110.00	24.30	0.22	25.00	0.00
7.00	81.00	78.00	40.00	48.00	46.70	0.26	42.00	0.00
2.00	107.00	74.00	30.00	100.00	33.60	0.40	23.00	0.00
4.00	146.00	85.00	27.00	100.00	28.90	0.19	27.00	0.00
3.00	130.00	78.00	23.00	79.00	28.40	0.32	34.00	1.00
12.00	100.00	84.00	33.00	105.00	30.00	0.49	46.00	0.00
10.00	129.00	76.00	28.00	122.00	35.90	0.28	39.00	0.00
1.00	122.00	64.00	32.00	156.00	35.10	0.69	30.00	1.00
1.00	95.00	60.00	18.00	58.00	23.90	0.26	22.00	0.00
5.00	88.00	66.00	21.00	23.00	24.40	0.34	30.00	0.00
2.00	100.00	66.00	20.00	90.00	32.90	0.87	28.00	1.00
3.00	107.00	62.00	13.00	48.00	22.90	0.68	23.00	1.00
1.00	107.00	72.00	30.00	82.00	30.80	0.82	24.00	0.00
6.00	123.00	72.00	45.00	230.00	33.60	0.73	34.00	0.00
1.00	125.00	50.00	40.00	167.00	33.30	0.96	28.00	1.00
1.00	97.00	64.00	19.00	82.00	18.20	0.30	21.00	0.00
5.00	99.00	54.00	28.00	83.00	34.00	0.50	30.00	0.00
2.00	81.00	72.00	15.00	76.00	30.10	0.55	25.00	0.00
1.00	139.00	46.00	19.00	83.00	28.70	0.65	22.00	0.00
0.00	84.00	82.00	31.00	125.00	38.20	0.23	23.00	0.00
5.00	166.00	72.00	19.00	175.00	25.80	0.59	51.00	1.00
1.00	81.00	72.00	18.00	40.00	26.60	0.28	24.00	0.00
2.00	75.00	64.00	24.00	55.00	29.70	0.37	33.00	0.00
0.00	74.00	52.00	10.00	36.00	27.80	0.27	22.00	0.00
4.00	197.00	70.00	39.00	744.00	36.70	2.33	31.00	0.00
4.00	110.00	76.00	20.00	100.00	28.40	0.12	27.00	0.00
6.00	134.00	70.00	23.00	130.00	35.40	0.54	29.00	1.00
3.00	74.00	68.00	28.00	45.00	29.70	0.29	23.00	0.00
2.00	197.00	70.00	45.00	543.00	30.50	0.16	53.00	1.00
3.00	180.00	64.00	25.00	70.00	34.00	0.27	26.00	0.00