

Implementasi Algoritma *K-Means Clustering* dan *Support vector machine* Untuk Mengetahui Faktor Pemicu Diabetes Melitus Di Puskesmas Lappae

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Mendapatkan Gelar Sarjana Komputer
(S.Kom) Program Studi Informatika



M. FAJAR PUTRA 10584112220

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH MAKASSAR

2024



UNIVERSITAS MUHAMMADIYAH MAKASSAR

FAKULTAS TEKNIK

GEDUNG MENARA IQRA LT. 3

Jl. Sultan Alauddin No. 259 Telp. (0411) 866 972 Fax (0411) 865 588 Makassar 90221

Website: www.unismuh.ac.id, e_mail: unismuh@gmail.com

Website: <http://teknik.unismuh.makassar.ac.id>

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

PENGESAHAN

Skripsi atas nama M. Fajar Putra dengan nomor induk Mahasiswa 105 84 11122 20, dinyatakan diterima dan disahkan oleh Panitia Ujian Tugas Akhir/Skripsi sesuai dengan Surat Keputusan Dekan Fakultas Teknik Universitas Muhammadiyah Makassar Nomor : 174/05/A.5-VI/45/2024, sebagai salah satu syarat guna memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar pada hari Sabtu tanggal 24 Agustus 2024.

Panitia Ujian :

1. Pengawas Umum

Makassar, -

19 Safar 1446 H

24 Agustus 2024 M

a. Rektor Universitas Muhammadiyah Makassar

Dr. Ir. H. Abd. Rakhim Nanda, ST., MT., IPU.

b. Dekan Fakultas Teknik Universitas Hasanuddin

Prof. Dr. Eng. Muhammad Isran Ramli, ST., MT.

2. Penguji

a. Ketua

Dr. Ir. Zahir Zainuddin, M.Sc.

b. Sekretaris

Desi Anggreani, S.Kom., MT.

3. Anggota

1. Rizki Yusliana Bakti ST., MT.

2. Lukman, S.Kom., M.T.

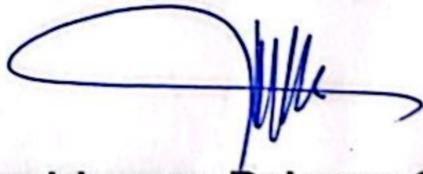
3. Muhyiddin A.M. Hayat, S.Kom., M.T.

Mengetahui :

Pembimbing I

Pembimbing II


Titin Wahyuni, S.Pd., M.T


Fahrir Irhamna Rahman S.Kom., MT

Dekan


Dr. Ir. Hj. Nurnawaty, ST., MT., IPM.
NBM 0795 108



UNIVERSITAS MUHAMMADIYAH MAKASSAR

FAKULTAS TEKNIK

GEDUNG MENARA IQRA LT. 3

Jl. Sultan Alauddin No. 259 Telp. (0411) 866 972 Fax (0411) 865 588 Makassar 90221

Website: www.unismuh.ac.id, e_mail: unismuh@gmail.com

Website: <http://teknik.unismuh.makassar.ac.id>

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan untuk memenuhi syarat ujian guna memperoleh gelar Sarjana Komputer (S.Kom) Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar.

Judul Skripsi : Implementasi Algoritma K-Means Clustering dan Support vector machine Untuk Mengetahui Faktor Pemicu Diabetes Melitus Di Puskesmas Lappae

Nama : M. Fajar Putra

Stambuk : 105 84 11122 20

Makassar, 30 Agustus 2024

Telah Diperiksa dan Disetujui
Oleh Dosen Pembimbing;

Pembimbing I

Pembimbing II

Titin Wahyuni S.Pd, M.T.

Fahrir Irhamna Rahman S.Kom., M.T.

Mengetahui,

Ketua Program Studi Informatika

Muhyiddin A.M.Hayat, S.Kom., MT.

NBM 1504 577



Abstract

The increasing number of Diabetes Mellitus cases in Indonesia is rising, necessitating preventive measures for early detection of the factors triggering Diabetes Mellitus. This research aims to implement the K-Means Clustering algorithm and Support Vector Machine (SVM) algorithm to classify Diabetes Mellitus risk at Lappae Health Center and measure the accuracy of SVM. The implementation of K-Means Clustering and SVM algorithms is used to classify patients with high and low Diabetes Mellitus risk at Lappae Health Center and determine their accuracy. The K-Means Clustering and SVM algorithms' method results in the high-risk group having 584 early detection high-risk patient data and the low-risk group having 456 low-risk patient data. The evaluation of the prediction results shows that the SVM model performs well with an accuracy of 93.75%. Precision, recall, and F1-score for both classes are also high, indicating that this model can predict diabetes risk accurately. The K-Means algorithm can be used to cluster patient data based on similar characteristics. SVM can be used for classification and prediction of Diabetes Mellitus risk with an accuracy of 93.75%, with 195 correct predictions out of 208 tested data.

Keywords : Algorithm Implementation, K-Means Algorithm,, SVM Algorithm, Lappae Health Center, Diabetes Mellitus



Abstrak

Peningkatan jumlah penderita Diabetes Melitus di Indonesia makin bertambah sehingga dibutuhkan langkah preventif dalam melakukan deteksi dini faktor pemicu Diabetes Melitus. Penelitian ini bertujuan untuk mengimplementasikan algoritma K-Means Clustering dan Algoritma Support vector machine untuk melakukan klasifikasi risiko Diabetes Melitus di Puskesmas Lappae dan mengukur tingkat akurasi SVM. mengimplementasikan algoritma K-Means Clustering dan algoritma SVM dalam mengklasifikasikan pasien dengan risiko Diabetes Melitus tinggi dan rendah di Puskesmas Lappae beserta tingkat akurasinya. metode algoritma k-means dan algoritma support vector machine Hasil Clustering menunjukkan kelompok beresiko tinggi dengan jumlah data yaitu 584 data pasien deteksi dini beresiko tinggi, dan kelompok beresiko rendah dengan jumlah data yaitu 456 data pasien. evaluasi hasil prediksi menunjukkan model SVM memiliki kinerja yang baik dengan akurasi sebesar 93.75%. Precision, recall, dan f1-score untuk kedua kelas juga tinggi, menunjukkan bahwa model ini mampu memprediksi risiko diabetes dengan baik. Algoritma K-Means dapat digunakan untuk mengelompokkan data pasien berdasarkan karakteristik yang mirip. SVM dapat digunakan untuk klasifikasi dan prediksi risiko Diabetes melitus dengan akurasi sebesar 93.75%. Dengan jumlah prediksi benar sebanyak 195 dari 208 data yang diujikan.

Kata Kunci : Implementasi Algoritma, Algoritma K-Means, Algoritma SVM, Puskesmas Lappae, Diabetes Melitus

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul ***“Implementasi Algoritma K-Means Clustering dan Support vector machine Untuk Mengetahui Faktor Pemicu Diabetes Melitus Di Puskesmas Lappae”*** Shalawat serta salam senantiasa tercurahkan kepada baginda Rasulullah SAW, yang telah memperbaiki akhlak dan budi pekerti manusia seperti yang kita rasakan sekarang ini.

Skripsi ini, yang merupakan salah satu persyaratan untuk mendapatkan gelar Sarjana (S1) di Universitas Muhammadiyah Makassar. Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna baik dari segi isi, bahasa, dan penulisannya meskipun penulis menerima bantuan dan bimbingan dari berbagai pihak selama proses penulisan. Dengan demikian, penulis sangat mengharapkan kritik dan saran yang bermanfaat untuk membantu mereka membuat kemajuan di masa depan.

Oleh karena itu, pada kesempatan kali ini penulis ingin mengucapkan banyak terima kasih kepada segenap pihak yang membantu khususnya kepada :

1. Kedua Orang Tua dan segenap Keluarga, yang selalu memberikan dukungan dan doa yang tiada henti-hentinya sehingga penulis dapat menyelesaikan studi ini
2. Bapak Prof. Dr.. H. Ambo Asse, M.Ag., sebagai Rektor Perguruan Tinggi Universitas Muhammadiyah Makassar.
3. Ibu Dr. Hj. Ir. Nurnawaty ST.,MT Selaku Dekan Fakultas Teknik Universitas Muhammadiyah Makassar.
4. Bapak Muhyidin A.M.Hayat S.Kom.,MT Selaku Ketua Prodi Informatika Fakultas Teknik Universitas Muhammadiyah

Makassar.

5. Ibu Titin Wahyuni S.Pd., M.T., Selaku Dosen Pembimbing I Dan Bapak Fahrim Irhamna Rachman, S.Kom., MT Selaku Dosen Pembimbing II yang senantiasa meluangkan waktunya, pikirannya untuk memberikan bimbingan serta mengarahkan penulis selama penyusunan skripsi ini .
6. Segenap Bapak/Ibu Dosen Prodi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar yang telah banyak memberikan bakat dan ilmu pengetahuan serta mendidik penulis selama proses belajar mengajar di Universitas Muhammadiyah Makassar.
7. Rekan-rekan mahasiswa utamanya dari Program studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar terima kasih atas dukungan dan kerjasamanya selama menempuh Pendidikan serta penyelesaian penyusunan skripsi ini.
8. Terakhir, terima kasih untuk diri sendiri, karena telah mampu berusaha keras dan berjuang sampai sejauh ini meski diterpa banyak problematika sepanjang penulisan tetapi mampu melewati semuanya dan menyelesaikan skripsi.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan dan keterbatasan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi kesempurnaan skripsi ini di masa yang akan datang. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan.

Billahi fi sabilil haq fastabiqul khairat Wassalamualaikum wr wb

Makassar , 26 Agustus 2024

M. Fajar Putra

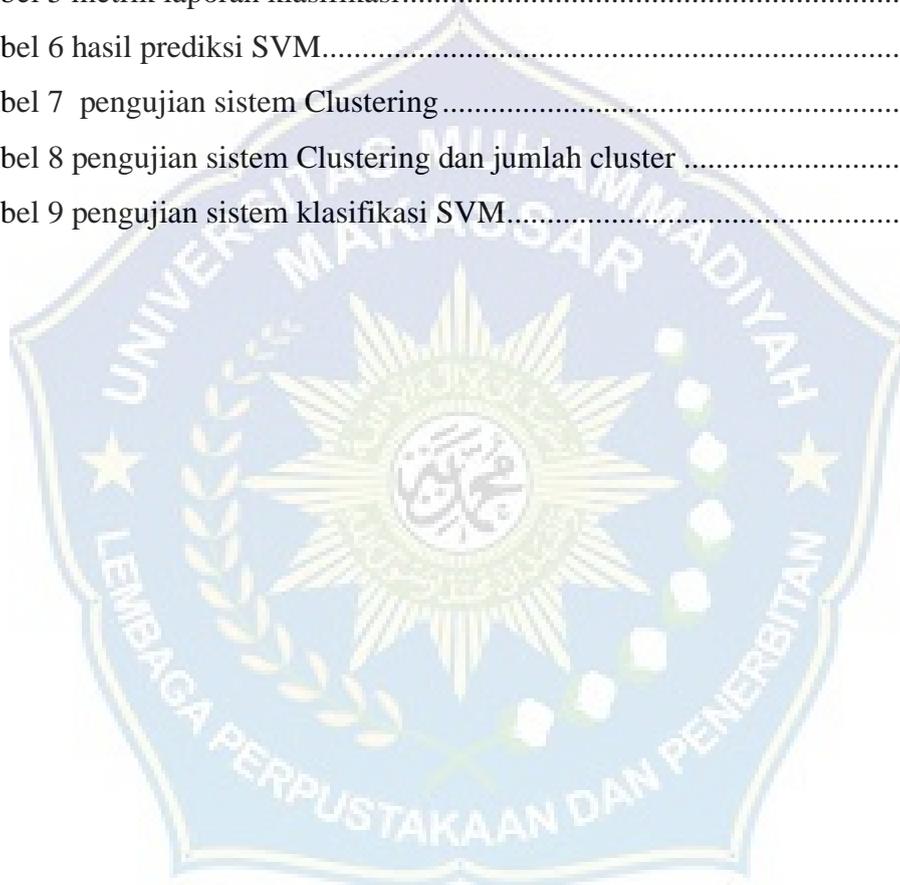
DAFTAR ISI

ABSTRAK.....	iii
KATA PENGANTAR.....	iii
DAFTAR ISI.....	v
DAFTAR TABEL.....	vii
DAFTAR GAMBAR.....	viii
DAFTAR LAMPIRAN.....	ix
DAFTAR ISTILAH.....	x
BAB I PENDAHULUAN.....	1
A. Latar Belakang.....	1
B. Rumusan Masalah.....	2
C. Tujuan Penelitian.....	3
D. Manfaat Penelitian.....	3
1. Aspek Teoritis.....	3
2. Aspek Praktis.....	3
E. Ruang Lingkup Penelitian.....	4
F. Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	6
A. Landasan Teori.....	6
1. Diabetes Melitus (DM).....	6
2. <i>K-Means Clustering</i>	6
3. <i>Support vector machine (SVM)</i>	8
4. Jupyter Notebook.....	8
5. Studi Kasus: Puskesmas Lappae.....	9
B. Penelitian Terkait.....	10
C. Kerangka Pikir.....	15
BAB III METODE PENELITIAN.....	16
A. Tempat dan Waktu Penelitian.....	16
B. Alat dan Bahan.....	16

C.	Perancangan Sistem.....	17
1.	Pengumpulan Data	17
2.	Preprocessing Data.....	17
3.	Implementasi Algoritma <i>K-MeansClustering</i>	17
4.	Implementasi Algoritma <i>Support vector machine</i>	18
5.	Analisis dan Interpretasi Hasil	19
6.	Penarikan Kesimpulan.....	19
D.	Teknik Pengujian Sistem	22
E.	Teknik Analisis Data.....	23
1.	Validasi Model:.....	23
2.	Visualisasi Hasil:.....	23
3.	Interpretasi Hasil:	23
4.	Penyusunan Laporan Hasil:.....	23
BAB IV PEMBAHASAN		24
A.	Pengumpulan Data	24
B.	Preprocessing Data.....	24
C.	Implementasi Algoritma K-Means.....	26
D.	Implemetasi Algoritma <i>SVM</i>	35
E.	Pengujian Sistem	41
BAB V KESIMPULAN DAN SARAN.....		44
A.	Kesimpulan	44
B.	Saran.....	45
DAFTAR PUSTAKA		46
LAMPIRAN		48

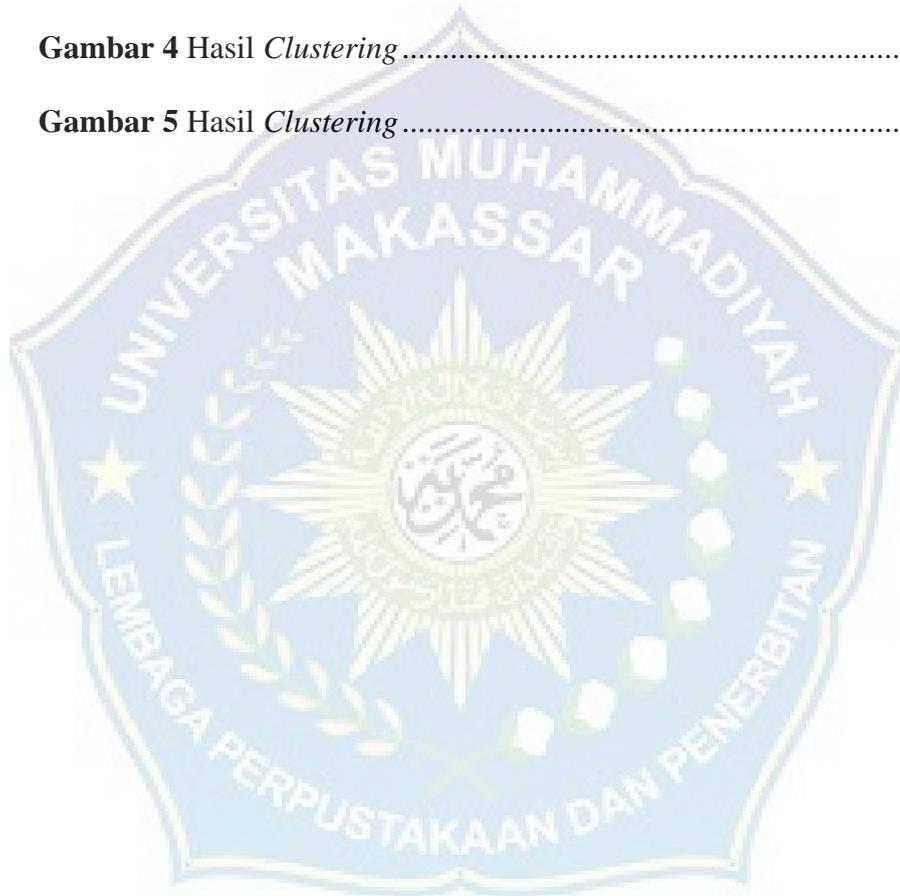
DAFTAR TABEL

Tabel 1. Pengumpulan data	24
Tabel 2. Preprocessing Data.....	25
Tabel 3. Data Normalization	25
Tabel 4. Hasil Clustering.....	33
Tabel 5 metrik laporan klasifikasi	37
Tabel 6 hasil prediksi SVM.....	40
Tabel 7 pengujian sistem Clustering	41
Tabel 8 pengujian sistem Clustering dan jumlah cluster	41
Tabel 9 pengujian sistem klasifikasi SVM.....	42



DAFTAR GAMBAR

Gambar 1 Kerangka Pikir.....	17
Gambar 2 Flowchart <i>Clustering</i> K-Means	22
Gambar 3 Flowchart Alur Penelitian.....	23
Gambar 4 Hasil <i>Clustering</i>	32
Gambar 5 Hasil <i>Clustering</i>	32



DAFTAR LAMPIRAN

Lampiran 1. Hasil Cluster	48
Lampiran 2. Prediksi SVM	57
Lampiran 3. Source Code.....	62
Lampiran 4 Scan Hasil Plagiasi PerBAB.....	65



DAFTAR ISTILAH

<i>Cluster</i>	Klaster biasanya didefinisikan sebagai kumpulan atau kelompok item dengan karakteristik yang sama atau berbeda. Kelompok atau kumpulan item tersebut membentuk sebuah klaster. Berikut ini adalah tiga definisi klaster yang terutama terkait dengan teknologi.
<i>Centroid</i>	Centroid adalah titik pusat objek atau area dalam ruang multidimensional. Dalam geometri, centroid adalah titik yang sesuai dengan pusat geometris suatu objek. Posisi pastinya di ruang angkasa dapat ditentukan dengan satu, dua, atau tiga koordinat, tergantung pada bentuk objek. Jika suatu bentuk memiliki sumbu simetri, maka centroid-nya akan selalu berada pada sumbu tersebut.
<i>Elbow Method</i>	Metode Elbow adalah teknik sederhana namun efektif untuk menentukan jumlah klaster (K) yang optimal dalam algoritma pengelompokan k-means. Metode ini menggunakan grafik siku untuk memetakan varians yang dijelaskan oleh jumlah klaster yang berbeda dan mengidentifikasi titik "elbow", di mana laju varians menurun tajam. Titik "elbow" ini menunjukkan jumlah klaster yang tepat untuk analisis atau pelatihan model.
<i>Silhouette</i>	Skor Silhouette adalah metrik yang digunakan untuk mengevaluasi seberapa baik hasil pengelompokan dalam pengelompokan data.
<i>precision</i>	Presisi mengukur seberapa sering prediksi untuk kelas positif benar.
<i>recall</i>	Recall mengukur seberapa baik model menemukan semua contoh positif

dalam kumpulan data

F1-score

F1 Score adalah harmonik mean dari recall dan precision, memberikan gambaran seimbang antara kedua metrik tersebut. Ini berguna ketika ada kebutuhan untuk meminimalkan False Positive dan False Negative secara seimbang.

SMOTE

Metode SMOTE ini adalah pengembangan dari metode oversampling, dimana cara kerja metode ini adalah dengan membangkitkan sampel baru yang berasal dari kelas minoritas untuk membuat proporsi data menjadi lebih seimbang dengan cara sampling ulang sampel kelas minoritas.



BAB I

PENDAHULUAN

A. Latar Belakang

Diabetes mellitus merupakan suatu kondisi jangka panjang yang ditunjukkan oleh kadar gula darah tinggi. Ketika pankreas tidak menghasilkan cukup insulin atau ketika tubuh tidak dapat menggunakan insulin secara efektif, kondisi ini dikenal sebagai diabetes mellitus. Salah satu efek paling umum dari diabetes mellitus yang tidak terkontrol adalah hiperglikemia atau peningkatan gula darah, yang secara bertahap dapat mengancam banyak sistem tubuh, terutama saraf dan pembuluh darah (WHO, 2019).

Diabetes mellitus (DM) merupakan penyakit kronis yang terjadi ketika tubuh tidak dapat menghasilkan cukup insulin (hormon yang mengontrol gula darah atau glukosa) karena gangguan pankreas, atau tubuh tidak dapat secara efisien memanfaatkan insulin yang diproduksi (Arania dkk, 2021). Di Indonesia, diabetes mellitus menjadi salah satu masalah kesehatan yang semakin meningkat. Puskesmas Lappae di Kabupaten Sinjai, Sulawesi Selatan, merupakan salah satu fasilitas kesehatan yang menghadapi banyak tantangan dalam menemukan dan mengelola pasien diabetes. Dengan jumlah penderita diabetes yang meningkat di wilayah ini, pendekatan yang lebih efektif diperlukan untuk mengidentifikasi dan mengelompokkan pasien berdasarkan faktor-faktor risiko yang mereka miliki.

Algoritma *K-Means* merupakan salah satu teknik kluster non-hirarki yang bertujuan untuk mengelompokkan variabel ke dalam kelas-kelas berdasarkan kemiripan. Teknik ini membagi data yang tersedia ke dalam satu atau lebih kelompok, sehingga data yang memiliki karakteristik serupa dikelompokkan dalam satu kluster (Wahyudi dkk., 2020). Dalam konteks diabetes mellitus, algoritma ini dapat digunakan untuk mengelompokkan pasien berdasarkan

faktor-faktor risiko seperti usia, indeks massa tubuh, dan kadar gula darah. Pengelompokan ini dapat membantu dalam memahami kelompok pasien yang memiliki risiko tinggi terkena diabetes mellitus.

Di sisi lain, classifier biner seperti *Support Vector Machine* digunakan untuk melakukan klasifikasi dengan membagi data menjadi dua kelas menggunakan hyperplane. *SVM* bekerja dengan mengubah ruang input asli menjadi ruang fitur yang lebih besar untuk memaksimalkan pemisahan antar kelas (Herwijayanti, 2019). Implementasi *SVM* dapat menghasilkan prediksi yang lebih akurat mengenai faktor-faktor yang meningkatkan risiko seseorang terkena diabetes mellitus.

Berdasarkan pernyataan di atas, penelitian ini bertujuan untuk menerapkan algoritma *K-Means Clustering* dan *SVM* guna mengetahui faktor-faktor pemicu diabetes mellitus di Puskesmas Lappae. Dengan mengidentifikasi faktor-faktor ini, diharapkan dapat memberikan wawasan yang lebih mendalam untuk upaya pencegahan dan pengelolaan penyakit diabetes mellitus yang lebih efektif. Penggunaan teknologi analisis data ini juga mendukung pengambilan keputusan berbasis data bagi tenaga kesehatan di Puskesmas Lappae.

B. Rumusan Masalah

Berdasarkan pada latar belakang masalah di atas, maka rumusan masalah pada penelitian ini yaitu:

1. Bagaimana cara mengimplementasikan algoritma *K-Means Clustering* untuk mengelompokkan pasien berdasarkan faktor-faktor risiko yang terkait dengan Diabetes Melitus
2. Bagaimana cara mengimplementasikan metode *Support vector machine* untuk melakukan klasifikasi risiko Diabetes Melitus serta mengukur tingkat akurasi *SVM* dalam melakukan klasifikasi risiko Diabetes Melitus di

Puskesmas Lappae.

C. Tujuan Penelitian

Berdasarkan pada latar belakang masalah, rumusan masalah di atas, maka tujuan yang ingin dicapai dari penelitian ini yaitu :

1. Mengimplementasikan algoritma *K-Means Clustering* dalam mengelompokkan pasien dengan risiko Diabetes Melitus tinggi dan rendah di Puskesmas Lappae
2. Mengimplementasikan algoritma *SVM* dalam mengklasifikasikan pasien dengan risiko Diabetes Melitus tinggi dan rendah di Puskesmas Lappae beserta tingkat akurasi.

D. Manfaat Penelitian

Berdasarkan pada uraian latar belakang , rumusan masalah dan tujuan penelitian di atas, maka manfaat penelitian ini dibagi menjadi beberapa aspek sebagai berikut :

1. Aspek Teoritis

Mengimplementasikan penggunaan algoritma *Clustering K-Means* dan *Support vector machine* dalam analisis data medis. Selain itu, memberikan referensi tambahan untuk penelitian-penelitian selanjutnya yang berminat mengembangkan metode analisis data khususnya yang berkaitan dengan penyakit diabetes melitus.

2. Aspek Praktis

A. Bagi Peneliti

- 1) Peneliti akan mendapatkan pengalaman praktis dalam menerapkan algoritma *Clustering K-Means* dan *Support vector machine* di dunia nyata.

2) Peneliti akan memperoleh keterampilan yang lebih baik dalam preprocessing data, analisis data, dan evaluasi model prediktif serta dapat mengubah metodologi mereka untuk memenuhi kebutuhan dan masalah yang muncul selama penelitian.

tambahan untuk penelitian-penelitian selanjutnya yang berminat mengembangkan metode analisis data khususnya yang berkaitan dengan penyakit diabetes melitus.

B. Bagi Universitas

1) Universitas dapat membantu meningkatkan keterampilan dan pengetahuan dalam analisis data dan klasifikasi menggunakan algoritma *K-Means Clustering* Dan *Support vector machine (SVM)*

2) Memberikan kontribusi pada pengetahuan dan pemahaman tentang analisis data, klasifikasi, yang merupakan aset bagi universitas untuk mendukung pengajaran, penelitian, dan layanan masyarakat.

E. Ruang Lingkup Penelitian

Dari analisis rumusan masalah di atas dapat dirumuskan beberapa batasan masalah yaitu :

1. Lokasi Penelitian di Puskesmas Lappae.
2. Rekam medis pasien termasuk variabel: usia, jenis kelamin, berat badan, hasil dan status indeks massa tubuh (BMI),

kadar gula darah (gula darah puasa dan gula darah sewaktu), tekanan darah, status tekanan darah.

3. Data pasien yang dikumpulkan dari periode tertentu untuk relevansi dan aktualitas yang terbatas pada data yang tersedia di Puskesmas Lappae

4. *K-Means Clustering*: Untuk mengelompokkan pasien berdasarkan faktor risiko DM.

5. *Support vector machine (SVM)*: Untuk klasifikasi dan prediksi risiko

DM

6. Analisis Hasil: Mengidentifikasi faktor risiko utama berdasarkan hasil *Clustering* dan prediksi *SVM*.

F. Sistematika Penulisan

Untuk memberikan gambaran umum dari seluruh penulisan ini, Adapun sistematika penulisan yaitu:

BAB I PENDAHULUAN

Bab ini menerangkan secara singkat dan jelas mengenai latar belakang penulisan penelitian tugas akhir, rumusan masalah, tujuan dan manfaat, Batasan permasalahan, metodologi yang digunakan dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini membahas tentang teori – teori yang melandasi penulis dalam melaksanakan skripsi.

BAB III METODE PENELITIAN

Membahas tentang metode penelitian dan alat yang digunakan untuk pembuatan sistem.

BAB IV PEMBAHASAN

Membahas tentang pengumpulan data, preprocessing data, serta pembahasan hasil dari penerapan metode dan pengujian sistem.

BAB V KESIMPULAN

Membahas tentang penarikan kesimpulan dari hasil pembahasan serta saran penelitian.

BAB II

TINJAUAN PUSTAKA

A. Landasan Teori

1. Diabetes Melitus (DM)

Diabetes Melitus (DM) adalah penyakit kronis yang terjadi ketika pankreas tidak menghasilkan cukup insulin atau ketika tubuh tidak dapat menggunakan insulin secara efektif. Insulin adalah hormon yang mengatur kadar gula darah. DM dibagi menjadi beberapa tipe, yaitu Diabetes Tipe 1, Diabetes Tipe 2, dan Diabetes Gestasional. Diabetes Tipe 2 adalah yang paling umum dan biasanya terkait dengan gaya hidup, seperti obesitas dan kurangnya aktivitas fisik. Diabetes melitus adalah kondisi jangka panjang yang ditunjukkan oleh kadar gula darah tinggi. Kondisi ini telah berkembang menjadi salah satu masalah kesehatan paling signifikan di banyak negara, termasuk Indonesia. Diabetes mellitus adalah penyakit metabolik menahun (kronis) yang ditandai dengan kadar gula darah melebihi batas normal (Kementrian Kesehatan Republik Indonesia, 2020).

2. *K-Means Clustering*

Metode *K-Means* merupakan sebuah metode yang melakukan proses clustering untuk setiap judul skripsi yang ada. Penelitian yang dilakukan menciptakan informasi baru yaitu dengan adanya *Clustering* data dari judul skripsi berdasarkan bidang dari judul itu sendiri dimana setiap hasilnya dapat dilihat di masing-masing cluster datanya. (Zakir, A. 2022). Prosesnya melibatkan langkah-langkah berikut

- a. Menentukan jumlah cluster (k).
- b. Memilih k titik sebagai *centroid* awal.
- c. Mengelompokkan setiap data ke *centroid* terdekat.
- d. Menghitung ulang posisi *centroid* berdasarkan anggota cluster.
- e. Mengulangi proses hingga posisi *centroid* stabil.

Rumus untuk menghitung jarak yang digunakan adalah Euclidean Distance.

Rumus tersebut adalah (Mathworld Wolfram, 2016):

$$d(X_j, C_j) = \sqrt{\sum_{j=1}^n (X_j - C_j)^2}$$

Rumus K-Means Cluster

Keterangan:

d = jarak;

j = banyaknya data;

c = *centroid* ;

x = data

K-Means efektif untuk menemukan kelompok dalam data yang memiliki karakteristik yang mirip satu sama lain. Algoritma ini sangat berguna dalam analisis data kesehatan untuk mengidentifikasi kelompok pasien berdasarkan risiko kesehatan mereka .

3. Support vector machine (SVM)

Support Vector Machine adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. *SVM* bekerja dengan mencari hyperplane optimal yang dapat memisahkan data ke dalam dua kelas yang berbeda. *SVM* menggunakan konsep margin maksimal untuk memisahkan kelas secara optimal, dengan tujuan meminimalkan kesalahan klasifikasi. *SVM* menggunakan konsep margin maksimal untuk memisahkan kelas secara optimal, dengan tujuan meminimalkan kesalahan klasifikasi. *SVM* merupakan metode yang bekerja berdasarkan prinsip Structural Risk Minimization (SRM). SRM digunakan untuk memaksimalkan margin dan meminimalkan batas atas risiko yang diharapkan dari risiko (Hasanah, S. 2018).

$$f(Xd) = \sum_{i=1}^{ns} \alpha_i Y_i \vec{x}_i \vec{x}_d + b$$

Rumus Klasifikasi SVM

Dimana:

ns = Jumlah support vector

α_i = Nilai bobot setiap titik data

y_i = Kelas data

\vec{x}_i = Variabel support vector

\vec{x}_d = Data yang akan diklasifikasikan b = Nilai error atau bias

4. Jupyter Notebook

Jupyter Notebook adalah aplikasi web open-source yang memungkinkan pengguna membuat dan berbagi dokumen yang berisi kode langsung, visualisasi, dan narasi teks. Ini adalah alat yang sangat berguna untuk pengembangan dan dokumentasi proyek ilmiah, analisis data, machine learning, dan banyak lagi (Kluyfer T. Dkk, 2016).

5. Studi Kasus: Puskesmas Lappae

Puskesmas Lappae adalah salah satu fasilitas kesehatan yang menyediakan layanan primer kepada masyarakat. Penelitian ini berfokus pada penerapan algoritma *K-Means Clustering* dan *SVM* untuk mengidentifikasi faktor pemicu Diabetes Melitus di Puskesmas Lappae. Dengan menggunakan data rekam medis pasien, algoritma *K-Means* digunakan untuk mengelompokkan pasien berdasarkan faktor risiko mereka, sementara *SVM* akan digunakan untuk mengklasifikasikan risiko DM dan membuat prediksi.



B. Penelitian Terkait

1) Cindi Wulandari, Tri Hasanah Bimastari Aviani, Rian Saputra (2024)

Pada penelitian yang dilakukan Cindi Wulandari, Tri Hasanah Bimastari Aviani, Rian Saputra dengan judul “Penerapan Algoritma Support vector machine Untuk Prediksi Tingkat Kelulusan Siswa SMA” Tujuan dari penelitian ini adalah untuk membantu pihak sekolah dalam memprediksi kelulusan siswa. Penelitian ini bertujuan memberikan alat yang dapat memprediksi tingkat kelulusan dengan akurat sehingga pihak sekolah dapat mengambil langkah yang tepat dalam meningkatkan kualitas pendidikan dan motivasi siswa .

Metode yang digunakan dalam penelitian ini adalah metode data mining, khususnya menggunakan algoritma Support vector machine (SVM). Tahapan penelitian dimulai dengan pengumpulan data nilai siswa tahun ajaran 2022/2023 dari SMA N 1 Kota Lubuklinggau. Data yang dikumpulkan kemudian digunakan dalam pengembangan sistem prediksi kelulusan. Sistem ini dikembangkan menggunakan aplikasi RapidMiner yang memungkinkan analisis dan prediksi berbasis data mining. Setelah sistem dikembangkan, tahap berikutnya adalah analisis sistem untuk memastikan bahwa sistem yang dibangun sesuai dengan tujuan penelitian dan dapat memprediksi kelulusan siswa dengan akurat. Pengujian sistem dilakukan menggunakan teknik validasi seperti split validation, cross validation, dan confusion matrix untuk mengukur tingkat akurasi prediksi. Dalam teknik split validation, data dibagi menjadi dua bagian, satu untuk melatih model dan satu lagi untuk menguji model. Cross validation digunakan sebagai metode statistik untuk mengevaluasi dan membandingkan algoritma pembelajaran dengan membagi data menjadi beberapa segmen. Pengujian dengan confusion matrix menghitung nilai precision, recall, dan accuracy dari model prediksi. Melalui tahapan ini, peneliti dapat memastikan bahwa sistem prediksi kelulusan yang dibangun menggunakan algoritma SVM memiliki tingkat akurasi yang tinggi dalam memprediksi kelulusan siswa.

Penelitian ini menyimpulkan bahwa penerapan algoritma Support vector machine mampu memberikan prediksi tingkat kelulusan siswa dengan akurasi yang tinggi. Hasil pengujian menunjukkan tingkat akurasi sebesar 98.81% untuk siswa kelas XII, 96.49% untuk siswa kelas XI, dan 98.25% untuk siswa kelas X. Dengan sistem ini, pengolahan data siswa tidak perlu dilakukan secara manual, sehingga sangat membantu pihak sekolah, masyarakat, dan peneliti lain yang membutuhkan informasi terkait tingkat kelulusan siswa .

2) Fauziah, S. N. (2023)

Pada penelitian yang dilakukan Fauziah, S. N. Dengan judul “KOMPARASI ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI DATA PADA

DIABETES PREDICTION DATASET” Tujuan penelitian ini adalah untuk mengetahui bagaimana algoritma Naive Bayes dan Support vector machine dapat membandingkan nilai akurasi, ketepatan, dan recall pada dataset prediktor diabetes pada split percentage 60%, 70%, dan 80%.

Metode Data dikumpulkan melalui observasi langsung dan studi literatur. Studi literatur mencakup teori, hasil, dan bahan penelitian dari jurnal nasional dan internasional, sedangkan observasi langsung mencari data yang sesuai untuk penelitian. Analisis data dilakukan dengan memilah dan mengelompokkan data sesuai dengan kriteria tertentu. Kemudian, algoritma Support vector machine dan Naive Bayes dibandingkan, dan kemudian data dikelompokkan berdasarkan persentase nilai dari kedua algoritma tersebut. Pada langkah implementasi solusi, dataset yang dihasilkan dari observasi data diproses menggunakan alat WEKA. Ini mengubah data numerik menjadi nominal sehingga aplikasi dapat membacanya. Untuk memproses data, Explorer dan Classify Rule menggunakan algoritma Naive Bayes dan SVM sebagai classifier. Ini menghasilkan kesimpulan item, ketepatan khusus kelas, dan matriks kekacauan.

Dari penelitian ini dapat ditarik kesimpulan bahwa Pengelompokkan objek yang didapat dari hasil analisis merupakan hasil akhir dari penelitian. Hasil yang diperoleh adalah nilai keakuratan dari hasil uji coba penerapan algoritma Naive Bayes dan Support vector machine terhadap diabetes prediction dataset. Yang kemudian disimpan sesuai kriteria masing-masing analisis

3) Fristi Riandari, Hengki Tamando Sihotang, Tarisa Tarigan, Muhammad Rafli (2022).

Pada penelitian yang dilakukan Fristi Riandari, Hengki Tamando Sihotang, Tarisa Tarigan, dan Muhammad Rafli. Dengan judul “Classification of Book Types Using the Support vector machine Method” Penelitian ini bertujuan untuk membuat model yang dapat mengklasifikasikan jenis buku berdasarkan beberapa kategori dan menganalisis hasil akurasi dari metode Support vector machine .

Penelitian ini dimulai dengan pengumpulan data berupa dataset buku yang diperoleh dari perpustakaan. Tahapan selanjutnya adalah pra-pemrosesan data yang meliputi segmentasi teks, case folding, tokenisasi, penghapusan stopword, dan stemming. Setelah itu, dilakukan ekstraksi fitur untuk mengeksplorasi informasi potensial dan merepresentasikan kata sebagai vektor fitur. Data kemudian dipisahkan menjadi data pelatihan dan data

uji. Proses klasifikasi dilakukan menggunakan metode SVM multiclass untuk mendapatkan hasil akhir pemodelan. Hasil klasifikasi yang diperoleh kemudian dievaluasi untuk mendapatkan nilai akurasi dan dianalisis apakah model klasifikasi yang dihasilkan layak untuk diimplementasikan.

Kesimpulan dari penelitian ini adalah Metode Support vector machine dapat digunakan untuk mengklasifikasikan jenis buku dengan melakukan tahap pra-pemrosesan teks, ekstraksi fitur, dan diikuti dengan implementasi metode SVM. Untuk penelitian selanjutnya, disarankan untuk membandingkan beberapa metode dalam mengklasifikasikan jenis buku untuk menentukan tingkat akurasi dari

metode yang digunakan serta melakukan pengembangan sistem lebih lanjut.

4) Much. afif masykur mughni, Tresna Maulana Fahrudin, Made Kamisutara (2021)

Pada penelitian yang dilakukan Much. afif masykur mughni, Tresna Maulana Fahrudin, Made Kamisutara. Dengan judul “Classification of Toddler Nutritional Status Based on Anthropometric Index and Feature Discrimination using Support vector machine Hyperparameter Tuning” Penelitian ini bertujuan untuk mengklasifikasikan status gizi balita berdasarkan indeks antropometri untuk membantu mengontrol pertumbuhan dan perkembangan balita. Metode yang digunakan adalah Support vector machine dengan tuning hyperparameter untuk membangun model klasifikasi status gizi balita.

Metode penelitian yang digunakan dalam studi ini mencakup beberapa tahapan. Pertama, data mengenai status gizi balita dikumpulkan dari berbagai sumber terpercaya. Setelah data terkumpul, dilakukan pemrosesan awal menggunakan Fisher's Discriminant Ratio untuk memilih fitur-fitur penting, seperti berat badan (BB) dan tinggi badan (BH), yang relevan dengan klasifikasi status gizi. Selanjutnya, data dibagi menjadi dua set, yaitu data pelatihan dan data pengujian dengan rasio 70:30. Proses klasifikasi menggunakan algoritma Support vector machine dengan penerapan hyperparameter tuning untuk meningkatkan kinerja model. Parameter yang disesuaikan mencakup Cost, Gamma, dan Kernel (RBF). Setelah model klasifikasi terbentuk, evaluasi dilakukan dengan mengukur akurasi, presisi, dan recall untuk memastikan kualitas dan keandalan model yang dihasilkan. Hasil dari metode ini menunjukkan efektivitas SVM dalam mengklasifikasikan status gizi balita dengan tingkat akurasi yang tinggi

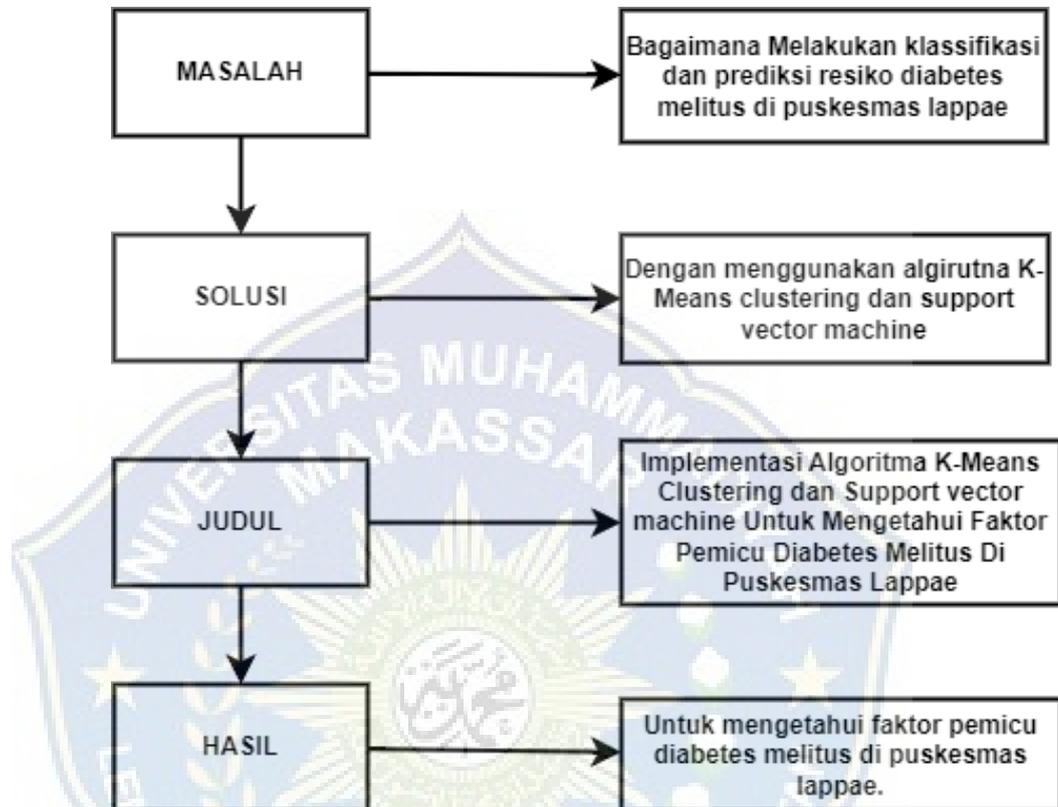
Penelitian ini menyimpulkan bahwa metode SVM dengan hyperparameter tuning dapat digunakan untuk mengklasifikasikan status gizi balita dengan tingkat akurasi yang signifikan. Model klasifikasi yang dibangun mencapai akurasi 97% dengan parameter Cost = 100, Gamma = 0.01, dan Kernel = RBF, menunjukkan peningkatan akurasi sebesar 13% dibandingkan dengan model SVM tanpa tuning.

5) Amanda Febrianti (2020)

Pada penelitian yang dilakukan Amanda Febrianti dengan judul “Penerapan metode K-Means dan Support vector machine dalam identifikasi Api dan citra warna digital” Penelitian ini bertujuan untuk mendeteksi api pada tahap Ignition dan Growth (menurut NFPA) pada citra digital hutan dan bangunan yang diambil di siang dan malam hari. Selain itu, penelitian ini juga bertujuan untuk menguji pengaruh gabungan fitur ruang warna RGB dan YCbCr dalam pendeteksian api menggunakan metode ekstraksi K-Means Clustering dan metode klasifikasi Support vector machine di Matlab. Metode penelitian yang digunakan dalam tugas akhir ini meliputi dua tahap utama: metode pengumpulan data dan metode simulasi. Data dikumpulkan melalui studi pustaka dari buku, jurnal, dan sumber elektronik lainnya yang relevan dengan objek penelitian. Penelitian ini menggunakan metode simulasi untuk mendeteksi api pada citra digital berdasarkan ciri warna dengan menggunakan metode K-Means Clustering untuk segmentasi gambar dan Support vector machine untuk klasifikasi. Tahapan metode simulasi meliputi formulasi masalah, pembuatan model konseptual, pengumpulan data input, pemodelan, simulasi, verifikasi dan validasi, eksperimen, dan analisis output.

Dari penelitian ini dapat ditarik kesimpulan bahwa Penelitian ini berhasil mengidentifikasi api pada tahap Ignition dan Growth menggunakan gabungan fitur ruang warna RGB dan YCbCr. Metode K- Means Clustering dan Support vector machine efektif digunakan untuk segmentasi dan klasifikasi citra api, menghasilkan akurasi yang memadai dalam pendeteksian api pada citra digital hutan dan bangunan yang diambil pada berbagai kondisi pencahayaan.

C. Kerangka Pikir



Gambar 1. Kerangka pikir

BAB III

METODE PENELITIAN

A. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Puskesmas Lappae Kecamatan Tellulimpoe, Kabupaten Sinjai, Adapun pelaksanaan penelitian ini dilakukan pada bulan Mei 2024 sampai semua proses pengumpulan data selesai.

B. Alat dan Bahan

Adapun alat dan bahan yang akan digunakan dalam penelitian ini adalah sebagai berikut :

1. Kebutuhan Hardware (Perangkat Keras)

- a. Laptop Asus Tuf
- b. HP

2. Kebutuhan Software (Perangkat Lunak)

- a. Excel
- b. Visual Studio Code
- c. Python (jupyter notebook)
- d. Google collab

C. Perancangan Sistem

Perancangan sistem adalah bagian penting dari pembangunan sistem karena menguraikan bagaimana suatu sistem dibangun dari tahap perencanaan hingga tahap pembuatan fungsi-fungsi yang diperlukan untuk mengoperasikannya. Tujuan dari perancangan sistem adalah untuk menentukan apakah sistem yang akan dibangun akan menghasilkan hasil yang diinginkan.

1. Pengumpulan Data

Mengumpulkan data pasien dari Puskesmas Lappa yang mencakup berbagai variabel seperti usia, jenis kelamin, berat badan, tinggi badan, indeks massa tubuh (BMI), riwayat keluarga, tingkat gula darah, tekanan darah

2. Preprocessing Data

1. *Data Cleaning dan Feature Selection*: Menghapus atau memperbaiki data yang hilang, tidak valid, atau anomali, serta Memilih fitur-fitur yang paling relevan untuk analisis, seperti usia, BMI, kadar gula darah
2. *Data Normalization*: Menormalkan nilai-nilai dalam dataset agar memiliki skala yang serupa.

3. Implementasi Algoritma *K-Means Clustering*

Menggunakan algoritma *K-Means Clustering* untuk mengelompokkan pasien berdasarkan karakteristik yang relevan. Menentukan jumlah cluster (k) yang optimal menggunakan metode seperti *Elbow Method* atau *Silhouette*

Score. Menganalisis hasil *Clustering* untuk mengidentifikasi kelompok-kelompok pasien dengan risiko DM yang berbeda.

1. Penentuan Jumlah Cluster: Menentukan jumlah cluster (k) yang optimal menggunakan metode seperti *Elbow Method* atau *Silhouette Score*.
2. Pengelompokan Data: Mengelompokkan pasien ke dalam k cluster berdasarkan kesamaan karakteristik.
3. Evaluasi *Clustering*: Menggunakan metrik seperti *Sum of Squared Errors (SSE)* atau *Silhouette Coefficient* untuk mengevaluasi kualitas pengelompokan.
4. Interpretasi Cluster: Menganalisis setiap cluster untuk memahami ciri-ciri pasien dalam setiap kelompok.

4. Implementasi Algoritma *Support vector machine*

Melatih model *SVM* menggunakan data pelatihan untuk membangun model prediktif yang dapat mengklasifikasikan risiko DM. Menggunakan kernel yang sesuai (linear, polynomial, atau RBF) untuk mendapatkan hasil terbaik.

1. Pembagian Data: Memisahkan data menjadi *training set* dan *testing set*.
2. *Training Model*: Melatih model *SVM* menggunakan trainingset.
3. *Hyperparameter Tuning*: Menyesuaikan parameter model (seperti C , γ , dan kernel) untuk meningkatkan kinerja model.

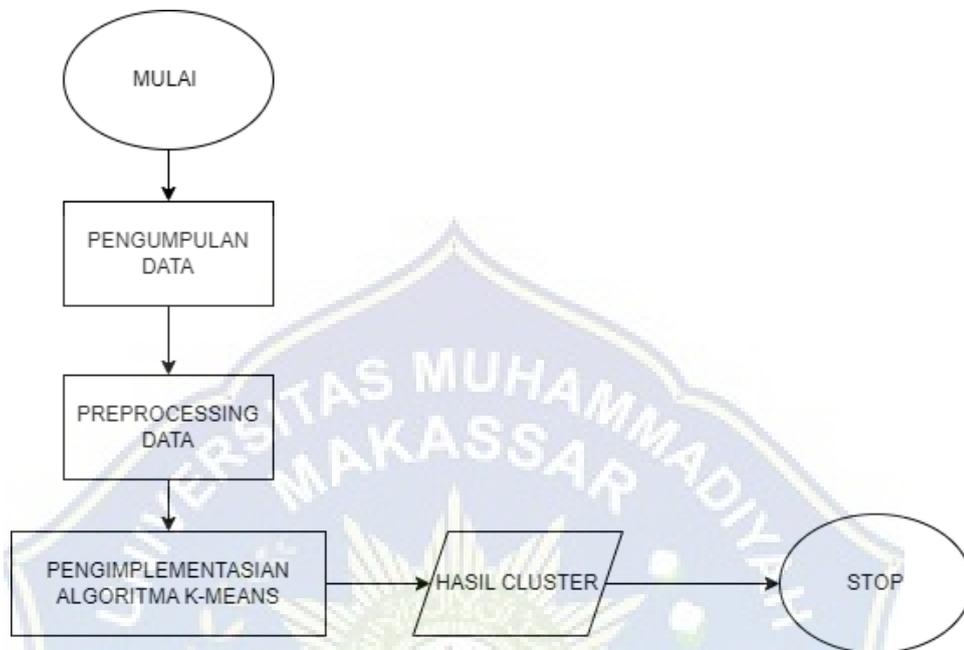
5. Analisis dan Interpretasi Hasil

Menganalisis hasil *Clustering* dan prediksi untuk mengidentifikasi faktor-faktor utama yang memicu Diabetes Melitus. Membandingkan hasil prediksi *SVM* dengan metode tradisional atau algoritma lain untuk menilai keakuratan dan efektivitasnya.

6. Penarikan Kesimpulan

Setelah pengujian selesai, kesimpulan dapat ditarik berdasarkan hasil evaluasi. Dengan perancangan sistem ini, diharapkan implementasi algoritma *K-Means Clustering* dan *SVM* dapat membantu Puskesmas Lappae dalam mengidentifikasi dan mengelola faktor risiko Diabetes Melitus dengan lebih efektif.

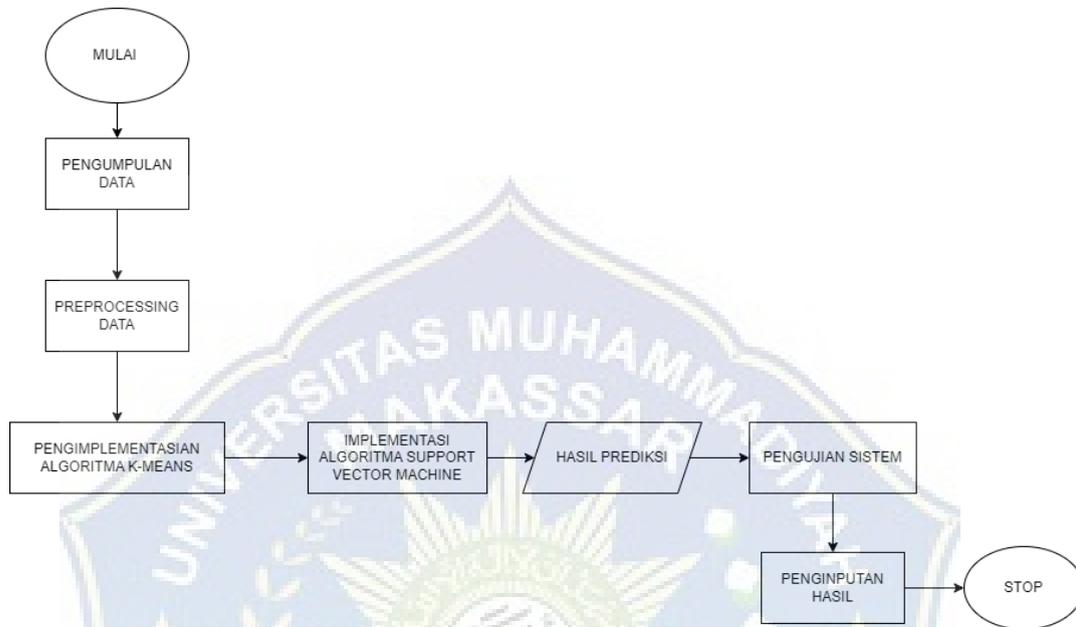




Gambar 2 Flowchart Clustering K-Means

Flowchart di atas menggambarkan Proses *Clustering* melibatkan beberapa langkah utama. Pertama, tentukan jumlah cluster yang diinginkan (k). Setelah itu, tentukan posisi awal dari *centroid* untuk masing-masing cluster. Selanjutnya, periksa apakah nilai *centroid* telah berubah. Jika nilai *centroid* berubah, hitung jarak setiap data dari *centroid*. Jika tidak, proses *Clustering* selesai. Langkah berikutnya adalah mengelompokkan data berdasarkan jarak terdekat ke *centroid* yang baru dihitung. Proses ini diulangi hingga nilai *centroid* tidak lagi berubah, menandakan bahwa cluster telah stabil

Dalam perancangan sistem atau diagram system yang akan dibuat yaitu sebagai berikut :



Gambar 3 Flowchart Alur Penelitian

Pada diagram di atas dapat dijelaskan bahwa alur proses analisis data yang melibatkan beberapa tahap utama. Dimulai dengan pengumpulan data dari berbagai sumber, data tersebut kemudian diproses melalui tahap preprocessing yang mencakup pembersihan, pengisian nilai yang hilang, dan normalisasi. Selanjutnya, data yang telah diproses diterapkan pada algoritma *K-Means Clustering* untuk mengelompokkan data berdasarkan kesamaan karakteristiknya. Hasil *Clustering* ini dianalisis untuk memahami pola yang terbentuk dalam setiap kluster. Tahap berikutnya adalah implementasi *Support vector machine (SVM)*, yang menggunakan hasil *Clustering* untuk membangun model klasifikasi. Setelah model *SVM* dibangun, dilakukan evaluasi dan validasi untuk memastikan kinerja dan kehandalan model. Proses ini diakhiri dengan menghasilkan output berupa laporan, visualisasi data, atau prediksi yang dihasilkan oleh model, yang kemudian dapat digunakan untuk pengambilan keputusan atau analisis lebih lanjut.

D. Teknik Pengujian Sistem

Implementasi *K-Means Clustering* dimulai dengan menentukan jumlah cluster (k) yang optimal menggunakan metode seperti *Elbow Method* atau *Silhouette Score*. Hasil *Clustering* dievaluasi dengan mengukur seberapa baik data dikelompokkan menggunakan metrik seperti SSE (*Sum of Squared Errors*) atau *Silhouette Coefficient*, dan cluster yang dihasilkan dianalisis untuk memahami karakteristik kelompok pasien berdasarkan faktor risiko yang teridentifikasi. Selanjutnya, algoritma *SVM* diimplementasikan untuk klasifikasi risiko DM dengan melatih model menggunakan training set, menyesuaikan parameter model untuk mendapatkan kinerja terbaik, dan menguji model pada testing set. Kinerja model dievaluasi menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1-score*, serta validasi silang dilakukan untuk memastikan model tidak *overfitting* dan memiliki generalisasi yang baik.

Hasil pengujian dan analisis kemudian divisualisasikan dalam bentuk grafik dan tabel yang mudah dipahami, seperti diagram sebaran untuk hasil *Clustering*. Laporan hasil analisis disusun untuk memberikan wawasan kepada tenaga kesehatan di Puskesmas Lappae, mencakup deskripsi hasil *Clustering* dan interpretasi setiap cluster serta analisis hasil prediksi *SVM* dan identifikasi faktor risiko utama yang mempengaruhi Diabetes Melitus. Akhirnya, hasil analisis diimplementasikan dalam bentuk dashboard interaktif bagi tenaga kesehatan untuk mengakses hasil analisis dan laporan secara *real-time*, sehingga membantu dalam pengambilan keputusan klinis. Tahapan pengujian teknik ini diharapkan dapat memberikan hasil yang akurat dan bermanfaat dalam mengidentifikasi faktor pemicu Diabetes Melitus, membantu Puskesmas Lappae dalam pengelolaan dan pencegahan penyakit tersebut secara lebih efektif.

E. Teknik Analisis Data

Analisis data merupakan serangkaian metode yang digunakan untuk menganalisis data, dan tujuan dari analisis data adalah untuk mengubah data mentah menjadi wawasan atau informasi yang bermanfaat untuk proses pengambilan keputusan bisnis. Wawasan yang ditemukan dalam data terdiri dari fakta, angka, dan metrik. Analisis data kualitatif dan kuantitatif adalah dua metode utama; keduanya dapat digunakan secara terpisah atau bersama-sama. Tujuannya jelas: membantu perusahaan membuat keputusan berbasis data.

(Latifatunnisa, H, 2022) Proses analisis data dilakukan pada penelitian ini adalah sebagai berikut:

1. Validasi Model:

Memastikan bahwa model tidak *overfitting* dan memiliki generalisasi yang baik.

2. Visualisasi Hasil:

Menggunakan grafik dan tabel untuk memvisualisasikan hasil *Clustering* dan prediksi *SVM*.

3. Interpretasi Hasil:

Menganalisis dan menginterpretasi hasil *Clustering* dan prediksi *SVM* untuk mengidentifikasi faktor-faktor pemicu Diabetes Melitus.

4. Penyusunan Laporan Hasil:

Menyusun laporan hasil analisis yang mencakup deskripsi hasil *Clustering*, interpretasi setiap cluster, analisis hasil prediksi *SVM*, dan rekomendasi untuk tindakan selanjutnya.

BAB IV PEMBAHASAN

A. Pengumpulan Data

Data yang dikumpulkan mencakup total 1040 data penderita diabetes melitus, yang diambil dari catatan medis dan laporan yang disimpan oleh Puskesmas Lappae, yang bersumber dari data prolanis (program lansia sehat), data posbindu (posko binaan dusun), serta data kunjungan penderita diabetes melitus di puskesmas lappae untuk periode tahun 2023 sampai 2024, sehingga didapat data sebagai berikut:

Table 1. Pengumpulan data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Nama	JK	Usia	Provinsi	Kab/Kota	Kecamatan	Desa/RT	Alamat	RT	Kode Pos	Puskesmas	Status	Hasil	Tgl	Status	Gula Darah	Status	Hasil	Tgl	Status
7087011: HASNI	female	52	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	KAMBUNO RW: 1 RT: 2			1070489	LAPPAE	HIPERTENS	150	100	HIPERGLIKE	285	OBESITAS	33.1556555		OBESITAS S
7087111: JUMRIAH	female	61	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	SUMPANG ALE RW: 1 R			1070489	LAPPAE	HIPERTENS	160	80	HIPERGLIKE	405	NORMAL	18.7180423		NORMAL
7085201: NURJIANNA	female	58	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	BORONG AMPIRIE RW:			1070489	LAPPAE	HIPERTENS	140	80	HIPERGLIKE	219	NORMAL	20.5498866		OBESITAS S
7087111: HUMRAH	female	57	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	BORONG AMPIRIE RW:			1070489	LAPPAE	PRE-HIPERT	120	70	HIPERGLIKE	272	NORMAL	23.2834185		OBESITAS S
307711: ROS	female	63	SULAWESI	KAB. SINJAI	TELLU LIMF MATTIBOW	LAPPA LURAE RW: RT:			1070489	LAPPAE	HIPERTENS	150	90	HIPERGLIKE	213	OBESITAS	55.2044753		OBESITAS S
708410: ENCKA	female	53	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN TORIBI RW: 2 R			1070489	LAPPAE	HIPERTENS	170	100	HIPERGLIKE	200	OBESITAS	56.7475803		OBESITAS S
7087111: SULKAWAT	female	52	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL			1070489	LAPPAE	PRE-HIPERT	120	80	HIPERGLIKE	422	BB LEBIH	25.2926006		OBESITAS S
7087111: SALMAWAT	female	43	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN TORIBI RW: 2 R			1070489	LAPPAE	PRE-HIPERT	120	70	HIPERGLIKE	278	OBESITAS	28.8415318		OBESITAS S
708410: HANIJUN	female	65	SULAWESI	KAB. SINJAI	TELLU LIMF MASSAILE	URANGAH RW: 1 RT: 11			1070489	LAPPAE	HIPERTENS	200	110	HIPERGLIKE	310				OBESITAS S
7087111: HAMMING	female	58	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN ATTIRONGE RW:			1070489	LAPPAE	HIPERTENS	140	70	HIPERGLIKE	217	NORMAL	24.4560634		OBESITAS S
7087111: NAWIRA	female	78	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN ATTIRONGE RW:			1070489	LAPPAE	HIPERTENS	170	100	HIPERGLIKE	274	NORMAL	21.8787158		NORMAL
7085901: NAWISA	female	47	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	ATTIRONGE RW: 1 R			1070489	LAPPAE	HIPERTENS	160	100	HIPERGLIKE	206	NORMAL	23.3725761		OBESITAS S
7087011: JUME	female	69	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL			1070489	LAPPAE	PRE-HIPERT	120	60	HIPERGLIKE	367	NORMAL	23.8330046		OBESITAS S
7087111: ENDANG	female	58	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN ATTIRONGE RW:			1070489	LAPPAE	HIPERTENS	140	80	HIPERGLIKE	217	NORMAL	24.8888888		OBESITAS S
7083111: RAJJA	male	73	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL			1070489	LAPPAE	PRE-HIPERT	130	60	HIPERGLIKE	320	NORMAL	24.9046178		NORMAL
708111: BEDDU	male	61	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	ATTIRONGE RW: 1 RT: 1			1070489	LAPPAE	HIPERTENS	140	80	HIPERGLIKE	209	NORMAL	21.2084487		NORMAL
708150: TEKKO K	male	55	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL			1070489	LAPPAE	PRE-HIPERT	120	60	HIPERGLIKE	403	NORMAL	23.1472551		NORMAL
708180: TAHIR	male	28	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL			1070489	LAPPAE	PRE-HIPERT	120	60	HIPERGLIKE	347	NORMAL	24.6548323		OBESITAS S
7087111: ATIMA	female	76	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN ATTIRONGE RW:			1070489	LAPPAE	HIPERTENS	150	90	HIPERGLIKE	229	NORMAL	21.6206477		NORMAL
207311: ABD. RASYI	male	66	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN ATTIRONGE RW:			1070489	LAPPAE	HIPERTENS	170	100	HIPERGLIKE	218	NORMAL	21.09375		NORMAL
708650: SUHRIAH	female	46	SULAWESI	KAB. SINJAI	TELLU LIMF MASSAILE	URANGAH RW: 1 RT: 11			1070489	LAPPAE	HIPERTENS	140	70	HIPERGLIKE	337	BB LEBIH	26.2222222		OBESITAS S
7087111: ROSNI	female	63	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	ATTIRONGE RW: 1 RT: 2			1070489	LAPPAE	HIPERTENS	180	100	HIPERGLIKE	267	BB LEBIH	26.6666666		OBESITAS S
7084111: ROSNI	female	52	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL			1070489	LAPPAE	PRE-HIPERT	120	70	HIPERGLIKE	340	NORMAL	24.5575723		OBESITAS S
708301: BABA	male	58	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL			1070489	LAPPAE	NORMAL	90	60	HIPERGLIKE	368	NORMAL	22.1002895		NORMAL
7087011: HASFIANI	female	36	SULAWESI	KAB. SINJAI	TELLU LIMF MASSAILE	DUSUN LAPPA JENE RW:			1070489	LAPPAE	PRE-HIPERT	130	80	HIPERGLIKE	212	OBESITAS	30.3819444		OBESITAS S
708610: HARTATI	female	51	SULAWESI	KAB. SINJAI	TELLU LIMF MASSAILE	BODDI RW: 1 RT: 1 KEL			1070489	LAPPAE	HIPERTENS	160	90	HIPERGLIKE	236	OBESITAS	29.0942823		OBESITAS S
7083111: MARSUKI	male	59	SULAWESI	KAB. SINJAI	TELLU LIMF MASSAILE	DUSUN LAPPAANNI RW:			1070489	LAPPAE	PRE-HIPERT	120	80	HIPERGLIKE	372	NORMAL	20.9321201		NORMAL
7084501: NURFATIMAH	female	36	SULAWESI	KAB. SINJAI	TELLU LIMF MASSAILE	BODDI RW: 1 RT: 3 KEL			1070489	LAPPAE	NORMAL	110	60	HIPERGLIKE	225	NORMAL	22.0740997		OBESITAS S
7086911: HASE	female	70	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	BORONG AMPIRIE RW:			1070489	LAPPAE	PRE-HIPERT	130	80	HIPERGLIKE	266	NORMAL	24.8356464		OBESITAS S
7087111: HAMSINA	female	83	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	SUMPANG ALE RW: 1 R			1070489	LAPPAE	HIPERTENS	150	80	HIPERGLIKE	285	NORMAL	22.0541902		OBESITAS S

B. Preprocessing Data

1. Data Cleaning Dan Feature Selection: Menghapus atau memperbaiki data yang hilang, tidak valid, atau anomali. Serta Memilih fitur-fitur yang paling relevan untuk analisis, seperti usia, BMI, kadar gula darah, Lingkar Perut, Tekanan Darah (Sistol), Tekanan Darah (Diastol), Gula Darah Sewaktu, dan status kolesterol.

Tabel 2. Preprocessing Data

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Nama Pr	JK	Usia	Provinsi	Kabkota	Kecamatan	Desa/rt	Alamat	rt	Kode Pos	Puskesmas	Status	Hasil	Ti	Status	Jula Darah	Status	Hasil	Status
3701: HASNI	female	52	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	KAMBUNO RW: 1 RT: 2	1070489 LAPPAE	HIPERTENS	150	100	HIPERGLIKE	285	OBESITAS	53.155655	OBESITAS	5	101	
3711: JUMRIA	female	61	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	SUMPANG ALE RW: 1 R	1070489 LAPPAE	HIPERTENS	160	80	HIPERGLIKE	405	NORMAL	18.7180423	NORMAL	79		
3520: NURJANNA	female	58	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	BORONG AMPIRIE RW:	1070489 LAPPAE	HIPERTENS	140	80	HIPERGLIKE	219	NORMAL	20.5498866	OBESITAS	5	91	
3711: HUMRAH	female	57	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	BORONG AMPIRIE RW:	1070489 LAPPAE	PRE-HIPERT	120	70	HIPERGLIKE	272	NORMAL	23.2834185	OBESITAS	5	81	
7711: ROS	female	63	SULAWESI	KAB. SINJAI	TELLU LIMF MATTIROW	LAPPA LURAE RW: RT:	1070489 LAPPAE	HIPERTENS	150	90	HIPERGLIKE	213	OBESITAS	35.2044753	OBESITAS	5	109	
3410: ENGA	female	53	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN TORIBI RW: 2 R	1070489 LAPPAE	HIPERTENS	170	100	HIPERGLIKE	200	OBESITAS	36.7475803	OBESITAS	5	91	
3711: SUKMAWA	female	52	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL	1070489 LAPPAE	PRE-HIPERT	120	80	HIPERGLIKE	422	BB LEBIH	25.2926006	OBESITAS	5	86	
3711: SALMAWAT	female	43	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN TORIBI RW: 2 R	1070489 LAPPAE	PRE-HIPERT	120	70	HIPERGLIKE	278	OBESITAS	28.8415318	OBESITAS	5	106	
3410: HANJUUN	female	65	SULAWESI	KAB. SINJAI	TELLU LIMF MASSALE	URANGAH RW: 1 RT: 1 I	1070489 LAPPAE	HIPERTENS	200	110	HIPERGLIKE	310						
3711: HAMMING	female	58	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN ATTIRONGE RW	1070489 LAPPAE	HIPERTENS	140	70	HIPERGLIKE	217	NORMAL	24.4566034	OBESITAS	5	92	
3711: NAWIRA	female	78	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN ATTIRONGE RW	1070489 LAPPAE	HIPERTENS	170	100	HIPERGLIKE	274	NORMAL	21.8771581	NORMAL	75		
3590: NAISYA	female	48	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	ATTIRONGE RW: 1 R	1070489 LAPPAE	HIPERTENS	160	100	HIPERGLIKE	206	NORMAL	23.3725761	OBESITAS	5	83	
3701: JUME	female	69	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL	1070489 LAPPAE	PRE-HIPERT	120	60	HIPERGLIKE	367	NORMAL	23.8330046	OBESITAS	5	82	
3711: ENDANG	female	58	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN ATTIRONGE RW	1070489 LAPPAE	HIPERTENS	140	80	HIPERGLIKE	217	NORMAL	24.8888888	OBESITAS	5	91	
3311: RAJJA	male	73	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL	1070489 LAPPAE	PRE-HIPERT	130	60	HIPERGLIKE	320	NORMAL	24.3046178	NORMAL	84		
3311: BEDDU	male	61	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	ATTIRONGE RW: 1 RT: 1	1070489 LAPPAE	HIPERTENS	140	80	HIPERGLIKE	209	NORMAL	21.2084487	NORMAL	74		
3150: TEKKO K	male	55	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL	1070489 LAPPAE	PRE-HIPERT	120	60	HIPERGLIKE	403	NORMAL	23.1472551	NORMAL	80		
3180: TAHIR	male	28	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL	1070489 LAPPAE	PRE-HIPERT	120	60	HIPERGLIKE	347	NORMAL	24.6548323	OBESITAS	5	91	
3711: ATIMA	female	76	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN ATTIRONGE RW	1070489 LAPPAE	HIPERTENS	150	90	HIPERGLIKE	229	NORMAL	21.6206471	NORMAL	78		
7311: ABD. RASYI	male	66	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	DUSUN ATTIRONGE RW	1070489 LAPPAE	HIPERTENS	170	100	HIPERGLIKE	218	NORMAL	21.09375	NORMAL	84		
3650: SUHRIAH	female	66	SULAWESI	KAB. SINJAI	TELLU LIMF MASSALE	URANGAH RW: 1 RT: 1	1070489 LAPPAE	HIPERTENS	140	70	HIPERGLIKE	337	BB LEBIH	26.2222222	OBESITAS	5	93	
3711: ROSNI	female	63	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	ATTIRONGE RW: 1 RT: 2	1070489 LAPPAE	HIPERTENS	180	100	HIPERGLIKE	267	BB LEBIH	26.6666666	OBESITAS	5	107	
3411: ROSNI	female	52	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL	1070489 LAPPAE	PRE-HIPERT	120	70	HIPERGLIKE	340	NORMAL	24.5577523	OBESITAS	5	84	
3301: BABA	male	58	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	TORIBI RW: 2 RT: 3 KEL	1070489 LAPPAE	NORMAL	90	60	HIPERGLIKE	368	NORMAL	22.1002895	NORMAL	78		
3701: HASFIANI	female	36	SULAWESI	KAB. SINJAI	TELLU LIMF MASSALE	DUSUN LAPPA JENE RW:	1070489 LAPPAE	PRE-HIPERT	130	80	HIPERGLIKE	212	OBESITAS	30.3819444	OBESITAS	5	86	
3610: HARTATI	female	51	SULAWESI	KAB. SINJAI	TELLU LIMF MASSALE	BODDI RW: 1 RT: 1 KEL	1070489 LAPPAE	HIPERTENS	160	90	HIPERGLIKE	236	OBESITAS	29.0942823	OBESITAS	5	106	
3311: MARSUKI	male	59	SULAWESI	KAB. SINJAI	TELLU LIMF MASSALE	DUSUN LAPPAANNU RW	1070489 LAPPAE	PRE-HIPERT	120	80	HIPERGLIKE	372	NORMAL	20.9321201	NORMAL	75		
3450: NURFATIM	female	36	SULAWESI	KAB. SINJAI	TELLU LIMF MASSALE	BODDI RW: 1 RT: 3 KEL	1070489 LAPPAE	NORMAL	110	60	HIPERGLIKE	225	NORMAL	22.0740999	OBESITAS	5	85	
3691: HASE	female	70	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	BORONG AMPIRIE RW:	1070489 LAPPAE	PRE-HIPERT	130	80	HIPERGLIKE	266	NORMAL	24.8356464	OBESITAS	5	91	
3711: HAMSINA	female	83	SULAWESI	KAB. SINJAI	TELLU LIMF KALOBBA	SUMPANG ALE RW: 1 R	1070489 LAPPAE	HIPERTENS	150	80	HIPERGLIKE	285	NORMAL	22.0541902	OBESITAS	5	84	

2. Data Normalization: Menormalkan nilai-nilai dalam dataset agar memiliki skala yang serupa, yaitu dengan mengubah kategori menjadi angka dengan memberikan nilai unik untuk setiap kategori. Misalnya, "Normal" menjadi 0, "prehiperglikemik" menjadi 1, "hiperglikemik" menjadi 2, dan variabel variabel lain juga dilakukan encoding.

Tabel 3. Data Normalization

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Nama Peserta Detek	NIK KELU	Usia	status	Tekanan Di	Kanan Di	Stus	Gula Darah Sew	Jula Darah	Status	Hasil	JMT	Status	Lingkar Perul	Lingki	Ik Darah	Tajam Penglihatan	Tajam Pender	
1	SULWANDI	0	45	2	150	80	2	282	0	24.26666667	1	83	1	1	0	0		
1	Arsyad	0	51	1	130	70	0	124	2	25.099501596	1	87	1	0	0	0		
1	ILHAM MUSTAFA	0	48	2	170	100	2	531	0	24.854832347	1	99	0	1	0	0		
1	ABO	1	59	2	150	80	2	264	0	21.19109462	1	87	1	1	0	0		
1	JUMRIA	1	61	2	160	80	2	405	0	18.718042367	0	79	1	0	0	1		
7	FARIDAWATI	1	47	1	130	70	2	217	1	14.177693762	0	80	1	0	0	1		
1	ANIS SUGIARTI	1	34	1	120	70	2	200	3	27.92666997	1	85	1	0	0	0		
1	RUKAYA	1	54	0	130	80	1	152	0	22.506925208	1	92	0	1	0	0		
0	HARTIA	1	45	1	120	80	2	236	0	23.7322594	1	82	0	0	0	0		
1	JAMAL	0	59	1	120	70	2	217	1	14.177693762	0	83	0	0	0	0		
2	MUSDALIFAH	1	55	2	200	130	2	341	3	28.57428371	1	100	0	0	0	0		
3	AMBO UDE	0	45	0	120	80	1	199	0	23.30905307	1	82	0	1	0	0		
4	ANSAR	0	63	0	130	80	0	112	3	63.775510204	1	87	0	1	1	1		
5	SALMAWATI	1	43	1	120	70	2	278	3	28.941531806	1	106	1	0	0	0		
6	Syamsuddin	0	59	0	120	80	1	173	2	25.33333333	1	89	0	1	0	0		
7	KONTI	0	74	2	190	100	2	262	0	23.942457623	1	81	1	0	0	1		
8	NAWIRA	1	78	2	170	100	2	274	0	21.878715815	0	75	0	1	1	0		
9	RAHMAN	0	57	2	180	100	2	210	0	19.756467309	1	92	0	1	0	0		
0	SABIR	0	52	1	130	90	0	94	0	24.22453287	1	92	0	0	0	0		
1	RAMLI	0	51	2	190	110	1	162	0	19.33729923	1	97	0	1	0	0		
2	RAJJA	0	73	1	130	60	2	320	0	24.304617877	0	84	1	1	1	1		
3	BEDDU	0	61	2	140	80	2	209	0	21.208448753	0	74	1	1	1	1		
4	TEKKO K	0	55	1	120	60	2	403	0	23.147255121	0	80	1	1	0	0		
5	TAHIR	0	28	1	120	60	2	347	0	24.654832347	1	91	1	0	0	0		
6	ATIMA	1	78	2	150	90	2	229	0	21.620647119	0	78	0	1	1	1		
7	ABD. RASYID.5	0	66	2	170	100	2	218	0	21.09375	0	84	1	1	1	1		
8	LATIF	0	64	2	160	100	2	379	3	28.969428008	1	102	1	0	0	1		
9	Jamaluddin	0	47	1	130	80	0	112	0	23.55555556	1	83	1	0	0	0		
0	HERWANWA	1	49	2	140	80	2	288	0	19.239158399	0	71	0	0	0	0		
1	A.Lingkar	0	59	2	180	100	2	466	0	21.641274238	1	85	1	1	0	0		
2	ABD. HALIM	0	65	0	110	70	0	124	0	19.227687871	1	87	0	1	1	1		
3	BABA	0	58	0	90	60	2	368	0	22.10028959	0	78	0	1	0	0		
4	HASFIANI	1	36	1	130	80	2	212	3	30.381944444	1	86	1	0	0	0		
5	ABDUL MAJID	0	48	2	170	90	2	203	0	24.9167874	1	98	1	1	0	0		
6	MARSUKI	0	59	1	120	80	2	372	0	20.932120125	0	75	0	1	0	0		
7	NURFATMAH	1	36	0	110	60	2	225	0	22.074099723	1	85	1	0	0	0		
8	LANTO	0	64	2	190	110	2	212	0	21.874228119	1	94	1	1	1	1		
9	MUH. DIRGANTARA																	

C. Implementasi Algoritma K-Means

K-means merupakan algoritma yang mengklusterisasikan berdasarkan partisi dan melakukan clustering melalui proses iterasi berkelanjutan sampai dengan bertemu kondisi akhir, proses iterasi berhenti dan hasil clustering adalah output. (Wang, Y., Li, D., & Wang, Y. (2019). *Clustering* merupakan metode analisis data yang bertujuan untuk mengelompokkan data ke dalam beberapa kluster atau kelompok berdasarkan kesamaan karakteristiknya. Pada penelitian ini, metode *K-Means* digunakan untuk melakukan *Clustering* pada data peserta deteksi dini diabetes dari Puskesmas Lappae.

Proses *Clustering* dengan *K-Means* melibatkan beberapa langkah utama:

1. **Memilih Kolom Numerik yang Relevan:** Kolom-kolom yang digunakan untuk *Clustering* dipilih berdasarkan relevansinya terhadap analisis diabetes.
2. **Pra-Pemrosesan Data:** Data pada kolom-kolom terpilih dibersihkan dan distandarkan.
3. **Penerapan Algoritma K-Means:** Algoritma *K-Means* diterapkan dengan menentukan jumlah kluster yang diinginkan.

```
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Memuat file Excel
jalur_file = 'Data DM puskesmas lappae - Copy.xlsx'
df = pd.read_excel(jalur_file, sheet_name='Sheet1')

# Memilih kolom numerik yang relevan untuk klustering
kolom_numerik = [
    'Usia',
    'Hasil - Tekanan Darah (Sistol)',
    'Hasil - Tekanan Darah (Diastol)',
```

```

    'Hasil - Gula Darah Sewaktu',
    'Hasil - IMT',
    'Hasil - Lingkar Perut'
]

# Membersihkan kolom numerik dengan mengubah string dengan koma
menjadi float
for kolom in kolom_numerik:
    df[kolom] = df[kolom].replace(',', '.',
regex=True).astype(float)

# Praproses data dengan menstandarkan
skaler = StandardScaler()
data_distandarkan = skaler.fit_transform(df[kolom_numerik])

# Menerapkan K-Means klustering dengan 2 kluster
kmeans = KMeans(n_clusters=2, random_state=42)
kmeans.fit(data_distandarkan)
df['Kluster'] = kmeans.labels_

# Plot kluster
plt.figure(figsize=(10, 6))
plt.scatter(data_distandarkan[:, 0], data_distandarkan[:, 1],
c=df['Kluster'], cmap='viridis')
plt.xlabel('Usia (distandarkan)')
plt.ylabel('Hasil - Tekanan Darah (Sistol) (distandarkan)')
plt.title('Klustering K-Means (2 Kluster)')
plt.colorbar(label='Kluster')
plt.show()

# Menyimpan dataframe dengan label kluster
df.to_excel('Hasil_Kluster.xlsx', index=False)

# Menampilkan beberapa baris pertama dari dataframe dengan label
kluster
print(df.head())

```

1. Implementasi metode

a. Memilih Kolom Numerik yang Relevan

```

# Memilih kolom numerik yang relevan untuk klustering
kolom_numerik = [

```

```

'Usia',
'Hasil - Tekanan Darah (Sistol)',
'Hasil - Tekanan Darah (Diastol)',
'Hasil - Gula Darah Sewaktu',
'Hasil - IMT',
'Hasil - Lingkar Perut'

```

b. **Pra-Pemrosesan Data:** Data dibersihkan dengan mengubah string yang mengandung koma menjadi float, kemudian distandarkan menggunakan StandardScaler.

```

# Membersihkan kolom numerik dengan mengubah string dengan koma
menjadi float
for kolom in kolom_numerik:
    df[kolom] = df[kolom].replace(',', '.').
    regex=True).astype(float)

# Praproses data dengan menstandarkan
skaler = StandardScaler()
data_distandarkan = skaler.fit_transform(df[kolom_numerik])

```

c. **Penerapan Algoritma K-Means:** Algoritma *K-Means* diterapkan dengan menentukan jumlah kluster sebanyak 2. Hasil *Clustering* divisualisasikan menggunakan plot scatter.

```

# Praproses data dengan menstandarkan
skaler = StandardScaler()
data_distandarkan = skaler.fit_transform(df[kolom_numerik])

# Menerapkan K-Means klustering dengan 2 kluster
kmeans = KMeans(n_clusters=2, random_state=42)
kmeans.fit(data_distandarkan)
df['Kluster'] = kmeans.labels_

# Plot kluster
plt.figure(figsize=(10, 6))
plt.scatter(data_distandarkan[:, 0], data_distandarkan[:, 1],
    c=df['Kluster'], cmap='viridis')
plt.xlabel('Usia (distandarkan)')
plt.ylabel('Hasil - Tekanan Darah (Sistol) (distandarkan)')
plt.title('Klustering K-Means (2 Kluster)')
plt.colorbar(label='Kluster')

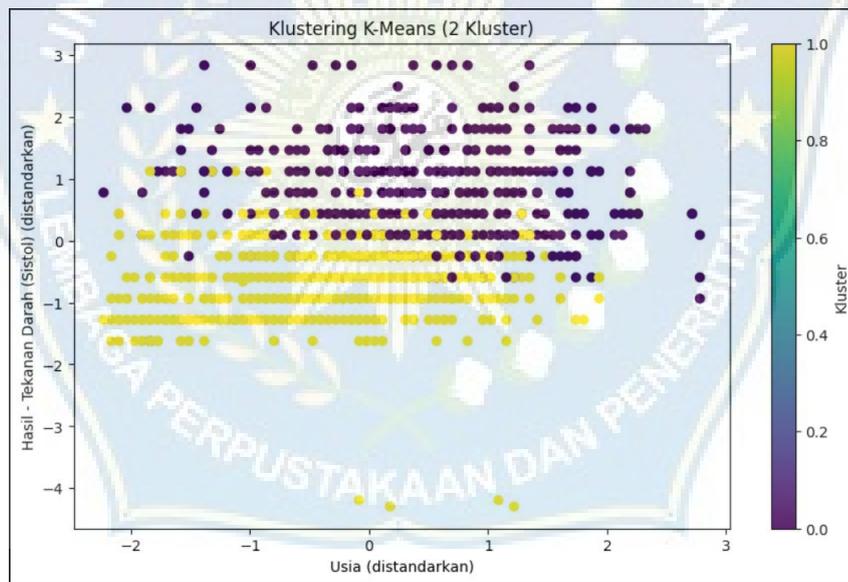
```

```
plt.show()

# Menyimpan dataframe dengan label kluster
df.to_excel('Hasil_Kluster.xlsx', index=False)

# Menampilkan beberapa baris pertama dari dataframe dengan label
kluster
print(df.head())
```

Hasil *Clustering* menunjukkan dua kelompok utama pada data peserta deteksi dini diabetes. Visualisasi scatter plot membantu dalam memahami distribusi data dan pembagian kluster. Hasil *Clustering* ini dapat digunakan untuk analisis lebih lanjut dalam menentukan karakteristik kelompok yang berisiko tinggi dan rendah terhadap diabetes.



Gambar 4 Hasil Clustering

Pada gambar diatas tiap dot berwarna kuning mewakili kluster karakteristik kelompok yang berisiko tinggi, sedangkan dot berwarna ungu mewakili kluster karakteristik kelompok yang berisiko rendah.

```

...   Nama Peserta Deteksi Dini  JENIS KELAMIN, MALE/FEMALE  Usia  \
0           SULWANDI                0  45.0
1           Arsyad                  0  51.0
2           ILHAM MUSTAFA            0  48.0
3           ABO                      1  59.0
4           JUMRIA                   1  61.0

Status - Tekanan Darah, NORMAL, PRE-HIPERTENSI, HIPERTENSI  \
0                                                                2
1                                                                1
2                                                                2
3                                                                2
4                                                                2

Hasil - Tekanan Darah (Sistol) Hasil - Tekanan Darah (Diastol) \
0           150.0           80.0
1           130.0           70.0
2           170.0           100.0
3           150.0           80.0
4           160.0           80.0

Status - Gula Darah Sewaktu, NORMAL, PRE-HIPERGLIKEMIK, HIPERGLIKEMIK  \
0                                                                2
1                                                                0
2                                                                2
...
1           0           1
2           0           0
3           0           0
4           1           0

```

Gambar 5 hasil cluster

Setelah proses *Clustering* dengan *K-Means* selesai dilakukan, hasil *Clustering* disimpan dalam file Excel dengan nama Hasil_Kluster.xlsx. Langkah selanjutnya adalah memuat file ini untuk melakukan analisis lebih lanjut dan menampilkan hasil *Clustering*.

```
df=pd.read_excel("Hasil_Kluster.xlsx",sheet_name="Sheet1")
df
```

File Excel yang berisi hasil *Clustering* dimuat menggunakan fungsi `pd.read_excel` dari library `pandas`. Fungsi ini memungkinkan kita untuk membaca file Excel dan menyimpannya dalam bentuk `DataFrame` `df`. `DataFrame` adalah struktur data dua dimensi yang digunakan untuk menyimpan data dalam format tabel, mirip dengan spreadsheet.

Re-Preprocessing

Dalam analisis data, tidak semua kolom yang terdapat dalam dataset selalu relevan untuk tujuan analisis tertentu. Beberapa kolom mungkin tidak memiliki kontribusi signifikan atau mungkin mengandung data yang tidak diperlukan. Oleh karena itu, langkah awal yang penting adalah menghapus kolom-kolom yang tidak relevan untuk fokus pada data yang penting. Pada penelitian ini, kolom-kolom yang tidak diperlukan untuk analisis diabetes diidentifikasi dan dihapus. Kolom-kolom ini meliputi:

- Status - Sadanis: Status pemeriksaan sadanis (pemeriksaan payudara klinis).
- Hasil - Sadanis kanan: Hasil pemeriksaan sadanis pada payudara kanan.
- Hasil - Sadanis kiri: Hasil pemeriksaan sadanis pada payudara kiri.
- Unnamed: 16: Kolom tambahan yang tidak memiliki nama dan biasanya merupakan artefak dari proses ekspor data.

Kolom-kolom ini dihapus karena tidak relevan dengan analisis *Clustering* dan prediksi diabetes yang dilakukan.

```
# Kolom-kolom yang akan dihapus
kolom_yang_dihapus = ['Status - Sadanis', 'Hasil - Sadanis kanan',
                      'Hasil - Sadanis kiri', 'Unnamed: 16']

# Menghapus kolom-kolom tersebut
df = df.drop(columns=kolom_yang_dihapus)
```

Setelah mengidentifikasi kolom-kolom yang tidak diperlukan, langkah selanjutnya adalah menghapus kolom-kolom tersebut dari DataFrame `df`. Fungsi `drop` dari `pandas` digunakan untuk menghapus kolom-kolom tersebut. Menghapus kolom-kolom yang tidak diperlukan merupakan langkah penting dalam pra-pemrosesan data. Pada penelitian ini, kolom-kolom seperti `Status - Sadanis`, `Hasil - Sadanis kanan`, `Hasil - Sadanis kiri`, dan `Unnamed: 16` dihapus

karena tidak relevan dengan tujuan analisis diabetes. Langkah ini membantu dalam fokus pada data yang penting, meningkatkan kinerja analisis, dan menghindari kebisingan dalam hasil analisis.

Setelah dilakukan re-preprocessing, data hasil *Clustering* dapat ditampilkan dengan perintah `df`. Variabel `df` adalah sebuah DataFrame yang berisi data hasil *Clustering*. DataFrame ini dapat digunakan untuk analisis lebih lanjut atau visualisasi data. Berikut adalah output dari hasil *Clustering* menggunakan Algoritma *K-Means*:



Tabel 4. Hasil Clustering

NO	JENIS KELAMIN	Usi a	Status - Tekanan Darah,	Hasil - Tekana n Darah (Sistol)	Hasil - Tekana n Darah (Diasto l)	Status - Gula Darah Sewaktu,	Hasil - Gula Darh Sewa ktu	Status - IMT,	Hasil - IMT	Status - Lingkar Perut,	Hasil - Lingk ar Perut	KOLESTR OL	Tajam Penglihat an, NORMAL/ GANGGU AN	Tajam Pendengar an, NORMAL/ GANGGU AN	Klust er
1	0	45	2	150	80	2	282	0	24,26667	1	89	1	1	0	0
2	0	51	1	130	70	0	124	2	25,0995	1	87	1	0	0	1
3	0	48	2	170	100	2	531	0	24,65483	1	99	0	1	0	0
4	1	59	2	150	80	2	264	0	23,19109	1	87	1	1	0	0
5	1	61	2	160	80	2	405	0	18,71804	0	79	1	0	1	0
...
1038	1	63	2	130	90	2	312	0	19,53125	1	81	1	0	1	0
1039	1	60	2	120	80	2	1109	2	25,91513	1	90	1	0	1	0
1038	1	66	2	110	80	2	262	0	23,94246	1	81	1	0	1	1
1039	1	49	2	140	90	2	204	2	25,29938	1	93	1	0	0	0
1040	1	38	2	140	80	2	232	2	26,89618	1	99	1	0	0	0

Hasil *Clustering* menunjukkan dua kelompok utama pada data peserta deteksi dini diabetes yaitu kelompok beresiko tinggi dengan jumlah data yaitu 584 data pasien deteksi dini beresiko tinggi, dan kelompok beresiko rendah dengan jumlah data yaitu 456 data pasien deteksi dini beresiko rendah. Analisis ini membantu mengidentifikasi karakteristik peserta terhadap diabetes. Ini sesuai dengan penjelasan Narasumber yaitu Hj. A. Haerul Baria S.Kep, Ns, MM.Kes, Selaku Kepala Puskesmas Lappae. Beliau menuturkan bahwa “Diabetes melitus merupakan penyakit yang dapat dipicu oleh hiperglikemia yaitu kondisi dimana gula darah sewaktu tinggi, selanjutnya yang berpengaruh juga yaitu tekanan darah atau hipertensi, kegemukan, dan yang berpengaruh juga yaitu keturunan dari penderita DM, semua yang disebut tadi itu faktor resiko tinggi semua.diluar dari itu faktor resiko rendah” Tuturnya.



D. IMPLEMETASI ALGORITMA SVM

PENDAHULUAN

Support vector machine adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. *SVM* bekerja dengan menemukan hyperplane terbaik yang memisahkan kelas-kelas dalam data. Pada penelitian ini, *SVM* digunakan untuk memprediksi diabetes berdasarkan data peserta deteksi dini. Konsep *SVM* bermula dari masalah klasifikasi dua kelas sehingga membutuhkan trainingset positif dan negatif. *SVM* berusaha menemukan hyperplane (pemisah) terbaik untuk memisahkan ke dalam dua kelas dan memaksimalkan margin antara dua kelas tersebut. (Pratama, A. et al 2017).

METODOLOGI

Proses prediksi dengan *SVM* melibatkan beberapa langkah utama:

1. Oversampling dengan SMOTE: Untuk mengatasi ketidakseimbangan kelas pada data, digunakan teknik oversampling dengan SMOTE (Synthetic Minority Over-sampling Technique).
2. Normalisasi Data: Data dinormalisasi untuk memastikan bahwa semua fitur memiliki skala yang sama.
3. Pelatihan Model *SVM*: Model *SVM* dilatih menggunakan data yang telah dinormalisasi.
4. Evaluasi Model: Model dievaluasi menggunakan metrik akurasi dan laporan klasifikasi.

IMPLEMENTASI

1. Oversampling dengan SMOTE:

SMOTE digunakan untuk menghasilkan sampel sintesis dari kelas minoritas, sehingga data menjadi lebih seimbang.

```

from sklearn.preprocessing import StandardScaler

# Terapkan SMOTE untuk oversampling kelas minoritas
smote = SMOTE(random_state=42)
x_train, y_train = smote.fit_resample(x_train, y_train)

```

2. Normalisasi Data: Data dinormalisasi menggunakan StandardScaler

```

# Normalisasi data (opsional, terutama untuk kernel RBF)
scaler = StandardScaler()
x_train = scaler.fit_transform(x_train)
x_test = scaler.transform(x_test)

```

3. Pelatihan Model SVM: Model SVM dilatih menggunakan kernel RBF (Radial Basis Function).

```

from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report

# Inisialisasi model SVM dengan kernel RBF
model = SVC(kernel='rbf', random_state=42)

# Latih model menggunakan data pelatihan
model.fit(x_train, y_train)

```

4. Evaluasi Model: Model dievaluasi dengan menghitung akurasi dan membuat laporan klasifikasi.

```

# Prediksi pada data uji
y_pred = model.predict(x_test)

# Evaluasi model
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred, zero_division=1)

```

```
# Cetak hasil
print("Accuracy:", accuracy)
print("Classification Report:\n", report)
```

HASIL PREDIKSI

Hasil prediksi menunjukkan bahwa model *SVM* dapat memisahkan kelas-kelas dengan baik. Akurasi dan laporan klasifikasi memberikan gambaran mengenai kinerja model. Hasil ini penting untuk mengidentifikasi peserta yang berisiko tinggi terkena diabetes, sehingga dapat dilakukan intervensi yang tepat.

Tabel 5 metrik laporan klasifikasi

Accuracy :	0.9375			
Classification Report :				
	precision	recall	f1-score	support
0	0.88	1.00	0.93	92
1	1.00	0.89	0.94	116
accuracy			0.94	208
macro avg	0.94	0.94	0.94	208
weighted avg	0.95	0.94	0.94	208

Penjelasan mengenai metrik-metrik dalam laporan klasifikasi:

Model SVM memiliki akurasi sebesar 93.75%, yang menunjukkan bahwa model ini mampu memprediksi dengan benar 93.75% dari data uji.

Precision: Precision adalah rasio prediksi benar dari kelas positif terhadap total prediksi untuk kelas tersebut. Precision tinggi menunjukkan bahwa model jarang memberikan prediksi positif yang salah (false positive).

- Precision untuk kelas 0: 0.88
- Precision untuk kelas 1: 1.00

Recall: Recall adalah rasio prediksi benar dari kelas positif terhadap total data aktual untuk kelas tersebut. Recall tinggi menunjukkan bahwa model mampu mengidentifikasi sebagian besar contoh positif dengan benar.

- Recall untuk kelas 0: 1.00
- Recall untuk kelas 1: 0.89

F1-Score: F1-score adalah rata-rata harmonis dari precision dan recall. F1-score digunakan untuk memberikan gambaran yang seimbang antara precision dan recall.

- F1-score untuk kelas 0: 0.93
- F1-score untuk kelas 1: 0.94

Support: Support menunjukkan jumlah data aktual yang ada untuk setiap kelas dalam data uji.

- Support untuk kelas 0: 92
- Support untuk kelas 1: 116

Accuracy: Akurasi keseluruhan untuk model, yang telah dibahas sebelumnya, adalah 0.94 atau 94%.

Macro Avg: Rata-rata precision, recall, dan f1-score dihitung secara rata-rata sederhana (tidak berbobot) untuk semua kelas.

- Precision: 0.94
- Recall: 0.94
- F1-Score: 0.94

Weighted Avg: Rata-rata precision, recall, dan f1-score dihitung dengan memberikan bobot sesuai dengan jumlah data aktual (support) untuk setiap kelas.

- Precision: 0.95
- Recall: 0.94

- F1-Score: 0.94

Selanjutnya adalah Penginputan Hasil, Model diinput kedalam file excel

```
# Dictionary untuk pemetaan label numerik ke label teks
label_map = {1: 'Diabetes Kronis', 0: 'Diabetes Ringan'}

# Gabungkan data uji, hasil prediksi, dan kolom yang diperlukan
results_df = pd.DataFrame({
    'Nama Peserta Deteksi Dini': nama_test,
    'JENIS KELAMIN, MALE/FEMALE': jenis_kelasmin_test,
    'Usia': usia_test,
    'Label Asli': y_test,
    'Prediksi': y_pred
})
# Mengganti nilai label numerik dengan label teks menggunakan map
results_df['Label Asli'] = results_df['Label Asli'].map(label_map)
results_df['Prediksi'] = results_df['Prediksi'].map(label_map)

# Ambil kolom-kolom yang perlu diekspor ke Excel
export_df = results_df[['Nama Peserta Deteksi Dini', 'JENIS KELAMIN,
MALE/FEMALE', 'Usia', 'Label Asli', 'Prediksi']]

# Simpan ke file Excel
file_path = 'Hasil Prediksi SVM.xlsx'
export_df.to_excel(file_path, index=False)

print(f"Hasil prediksi telah diekspor ke file: {file_path}")
```

Hasil prediksi telah diekspor ke file: Hasil Prediksi SVM.xlsx

Tabel 6 hasil prediksi SVM

NO	JENIS KELAMIN, MALE/FEMALE	Usia	Label Asli	Prediksi
1	1	71	Diabetes Ringan	Diabetes Ringan
2	1	54	Diabetes Kronis	Diabetes Kronis
3	0	59	Diabetes Kronis	Diabetes Kronis
4	0	58	Diabetes Kronis	Diabetes Kronis
5	1	58	Diabetes Ringan	Diabetes Ringan
...
203	1	49	Diabetes Kronis	Diabetes Kronis
204	1	34	Diabetes Kronis	Diabetes Kronis
205	1	35	Diabetes Kronis	Diabetes Kronis
206	0	47	Diabetes Ringan	Diabetes Ringan
207	1	57	Diabetes Kronis	Diabetes Kronis
208	1	46	Diabetes Ringan	Diabetes Ringan

Dari hasil prediksi dan penginputan hasil prediksi dapat disimpulkan bahwa evaluasi hasil prediksi menunjukkan model *SVM* memiliki kinerja yang baik dengan akurasi sebesar 93.75%. Precision, recall, dan f1-score untuk kedua kelas juga tinggi, menunjukkan bahwa model ini mampu memprediksi risiko diabetes dengan baik. Laporan klasifikasi memberikan rincian mendalam tentang kinerja model untuk setiap kelas, yang membantu dalam memahami kekuatan dan kelemahan model.

E. PENGUJIAN SISTEM

1. Pengujian Sistem *Clustering K-Means*

Pengujian Sistem *Clustering K-Means* dilakukan dengan melihat hasil kluster pada kolom kluster dimana nilai 0 itu merepresentasikan resiko rendah deteksi dini diabetes melitus, sedangkan nilai 1 merepresentasikan resiko tinggi deteksi dini diabetes melitus.

Tabel 7 pengujian sistem *Clustering*

MIN, MAL	Usia	RMAL, PR	kanan Dar	anan Dar	MAL, PRE	ula Darah	BB KURAN	Hasil - IMT	NORMAL	Lingkar PL	NORMA	an, NORM	ran, NORM	Kluster
0	45	2	150	80	2	282	0	24,26667	1	89	1	1	0	0
0	51	1	130	70	0	124	2	25,0995	1	87	1	0	0	1
0	48	2	170	100	2	531	0	24,65483	1	99	0	1	0	0
1	59	2	150	80	2	264	0	23,19109	1	87	1	1	0	0
1	61	2	160	80	2	405	0	18,71804	0	79	1	0	1	0
1	47	1	120	70	2	217	1	14,17769	0	80	1	0	1	1
1	34	1	120	70	2	200	3	27,92667	1	85	1	0	0	1
1	54	0	130	80	1	152	0	22,50693	1	92	0	1	0	0
1	45	1	130	80	2	236	0	23,73324	1	82	0	0	0	1
0	59	1	120	70	2	217	1	14,17769	0	83	0	0	0	1
1	55	2	200	130	2	341	3	28,57143	1	100	0	0	0	0
0	45	0	120	80	1	199	0	23,30905	1	82	0	1	0	1
0	63	0	130	80	0	112	3	63,77551	1	87	0	1	1	0
1	43	1	120	70	2	278	3	28,84153	1	106	1	0	0	1
0	59	0	120	80	1	173	2	25,33333	1	89	0	1	0	1
0	74	2	190	100	2	262	0	23,94246	1	81	1	0	1	0
1	78	2	170	100	2	274	0	21,87872	0	75	0	1	1	0
0	57	2	180	100	2	210	0	19,75647	1	92	0	1	0	0
0	52	1	130	90	0	94	0	24,22145	1	92	0	0	0	0
0	51	2	190	110	1	162	0	19,33373	1	97	0	1	0	0
0	73	1	130	60	2	320	0	24,30462	0	84	1	1	1	1
0	61	2	140	80	2	209	0	21,20845	0	74	1	1	1	1
0	55	1	120	60	2	403	0	23,14726	0	80	1	1	0	1
0	28	1	120	60	2	347	0	24,65483	1	91	1	0	0	1
1	76	2	150	90	2	229	0	21,62065	0	78	0	1	1	0
0	66	2	170	100	2	218	0	21,09375	0	84	1	1	1	0
0	64	2	160	100	2	379	3	28,96943	1	102	1	0	1	0
0	47	1	130	80	0	112	0	23,55556	1	83	1	0	0	1
1	49	2	140	80	2	288	0	19,23916	0	71	0	0	0	1
0	59	2	180	100	2	466	0	21,64127	1	85	1	1	0	0

Selanjutnya mencari jumlah data kluster beresiko tinggi dan beresiko rendah dari total 1040 data pasien deteksi dini.

Tabel 8 pengujian sistem *Clustering* dan jumlah cluster

1	53	2	200	100	2	310	0	21,62065	1	84	1	1	0	0
1	63	2	130	90	2	312	0	19,53125	1	81	1	0	1	0
1	60	2	120	80	2	1109	2	25,91513	1	90	1	0	1	0
1	66	2	110	80	2	262	0	23,94246	1	81	1	0	1	1
1	49	2	140	90	2	204	2	25,29938	1	93	1	0	0	0
1	38	2	140	80	2	232	2	26,89618	1	99	1	0	0	0
											jumlah kluster resiko tinggi :		584	
											jumlah kluster resiko rendah :		(1040-584)=456	

Hasil *Clustering* menunjukkan dua kelompok utama pada data peserta deteksi dini diabetes yaitu kelompok beresiko tinggi dengan jumlah data yaitu 584 data pasien

deteksi dini beresiko tinggi, dan kelompok beresiko rendah dengan jumlah data yaitu 456 data pasien deteksi dini beresiko rendah. Analisis ini membantu mengidentifikasi karakteristik peserta terhadap diabetes.

Hal ini sesuai dengan penjelasan Narasumber yaitu Hj. A. Haerul Baria S.Kep, Ns, MM.Kes, Selaku Kepala Puskesmas Lappae. Beliau menuturkan bahwa “Diabetes melitus merupakan penyakit yang dapat dipicu oleh hiperglikemia yaitu kondisi dimana gula darah sewaktu tinggi, selanjutnya yang berpengaruh juga yaitu tekanan darah atau hipertensi, kegemukan, dan yang berpengaruh juga yaitu keturunan dari penderita DM, semua yang disebut tadi itu faktor resiko tinggi semua.diluar dari itu faktor resiko rendah” Tuturnya

2. Pengujian Sistem Klasifikasi Dan Prediksi Dengan SVM

Pengujian Sistem klasifikasi dan prediksi SVM Dilakukan dengan cara manual yaitu menjumlah semua prediksi yang benar. Prediksi yang benar bernilai 1, sedangkan prediksi yang salah bernilai 0

Tabel 9 pengujian sistem klasifikasi SVM

NO	JENIS KELAMIN	Usia	Label Asli	Prediksi	Hasil Akurasi
1	1	71	Diabetes Ringan	Diabetes Ringan	1
2	1	54	Diabetes Kronis	Diabetes Kronis	1
3	0	59	Diabetes Kronis	Diabetes Kronis	1
4	0	58	Diabetes Kronis	Diabetes Kronis	1
5	1	58	Diabetes Ringan	Diabetes Ringan	1
6	0	39	Diabetes Ringan	Diabetes Ringan	1
7	1	57	Diabetes Kronis	Diabetes Kronis	1
8	1	48	Diabetes Kronis	Diabetes Kronis	1
9	0	47	Diabetes Kronis	Diabetes Kronis	1

10	1	50	Diabetes Kronis	Diabetes Kronis	1
...
199	0	42	Diabetes Kronis	Diabetes Kronis	1
200	1	38	Diabetes Kronis	Diabetes Kronis	1
201	1	46	Diabetes Kronis	Diabetes Kronis	1
202	1	23	Diabetes Ringan	Diabetes Ringan	1
203	1	49	Diabetes Kronis	Diabetes Kronis	1
204	1	34	Diabetes Kronis	Diabetes Kronis	1
205	1	35	Diabetes Kronis	Diabetes Kronis	1
206	0	47	Diabetes Ringan	Diabetes Ringan	1
207	1	57	Diabetes Kronis	Diabetes Kronis	1
208	1	46	Diabetes Ringan	Diabetes Ringan	1
Jumlah prediksi benar:					195
persentase prediksi benar :					0,9375

dari hasil prediksi dan penginputan hasil prediksi dapat disimpulkan bahwa evaluasi hasil prediksi menunjukkan model *SVM* memiliki kinerja yang baik dengan akurasi sebesar 93.75%. Dengan jumlah prediksi benar sebanyak 195 dari 208 data yang diujikan, sehingga dapat dipersentasekan $(195:208) \times 100 \% = 93,75 \%$, nilai ini sesuai dengan nilai hasil prediksi yang telah dibahas sebelumnya pada bagian pembahasan hasil prediksi implementasi algoritma *SVM*.

BAB V KESIMPULAN DAN SARAN

A. Kesimpulan

1. Implementasi Algoritma K-Means dapat digunakan untuk mengelompokkan data pasien berdasarkan karakteristik yang mirip. Proses ini melibatkan pemilihan kolom numerik yang relevan, pembersihan data, dan standarisasi menggunakan StandardScaler. *Clustering* dilakukan dengan menentukan jumlah kluster yang diinginkan, dalam hal ini sebanyak dua kluster yaitu kelompok beresiko tinggi dengan jumlah data yaitu 584 data pasien deteksi dini beresiko tinggi, dan kelompok beresiko rendah dengan jumlah data yaitu 456 data pasien deteksi dini beresiko rendah. Hasil dari *Clustering* ini diperkuat dengan penjelasan narasumber
2. Implementasi Algoritma *Support vector machine* dapat digunakan untuk klasifikasi dan prediksi risiko Diabetes Melitus . Model *SVM* dilatih menggunakan training set dan diuji pada testing set. Untuk akurasi Evaluasi kinerja model dilakukan menggunakan metrik seperti akurasi, precision, recall, dan F1-score . Dari hasil prediksi dan penginputan hasil prediksi dapat disimpulkan bahwa evaluasi hasil prediksi menunjukkan model *SVM* memiliki kinerja yang baik dengan akurasi sebesar 93.75%. Dengan jumlah prediksi benar sebanyak 195 dari 208 data yang diujikan, sehingga dapat dipersentasekan $(195:208) \times 100 \% = 93,75 \%$, nilai ini sesuai dengan nilai hasil prediksi yang telah dibahas sebelumnya pada bagian pembahasan hasil prediksi implementasi algoritma *SVM*.

Pengujian hasil *Clustering* dan prediksi *SVM* dilakukan dengan menggunakan berbagai metrik evaluasi. Hasil analisis divisualisasikan dalam bentuk grafik dan tabel untuk memudahkan pemahaman.

B. Saran

Mendorong penelitian lanjutan untuk mengeksplorasi metode baru dalam analisis dan prediksi risiko diabetes melitus jika di masa mendatang terdapat metode yang lebih efektif dan efisien dalam melakukan klasifikasi dan prediksi pada studi kasus ini.



DAFTAR PUSTAKA

- American Heart Association. (2016). Understanding blood pressure readings. Retrieved from /Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp
- Arania, R., Triwahyuni, T., Esfandiari, F., & Nugraha, F. R. (2021). Hubungan antara usia, jenis kelamin, dan tingkat pendidikan dengan kejadian diabetes mellitus di Klinik Mardi Waluyo Lampung Tengah. *Jurnal Medika Malahayati*, 5(3), 146-153.
- Fahrudin, T. M., & Kamisutara, M. (2021). Classification of Toddler Nutritional Status Based on Anthropometric Index and Feature Discrimination using *Support vector machine* Hyperparameter Tuning. *Ijconsist Journals*, 2(02), 60-65.
- Fauziah, S. N. (2023). *Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Klasifikasi Data Pada Diabetes Prediction Dataset* (Doctoral Dissertation, Universitas Siliwangi).
- Febrianti, A. (2020). Penerapan metode *K-Means Clustering* dan *Support vector machine SVM* dalam Identifikasi Api pada Citra Warna digital (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).
- Hasanah, S. (2018). *Propensity Score Matching Menggunakan Support Vector Machine Pada Kasus Diabetes Melitus (DM) Tipe 2* (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).
- Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(1), 306-312.
- Kemendes RI. Infodatin (2020) *Diabetes Melitus Pusat Data dan Informasi Kementerian Kesehatan RI*. Jakarta: Kementerian Kesehatan Republik Indonesia; 2020.
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., ... & Willing, C. (2016). Jupyter Notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87-90).
- Kusumah, Elsandy Wirahadi, Chandranegara, Didih Rizki, & Nuryasin, Ilyas. (2023). Pengembangan Sistem Informasi Geografis Pemetaan daerah Prioritas Tujuan Vaksin Berbasis Website. *Jurnal Impresi Indonesia*, 2(3), 236–245

- Latifatunnisa, H. (2022). Teknik Analisis Data: Jenis, Contoh, dan Langkah-langkahnya. Retrieved from Revoupedia: <https://revou. co/panduan-teknis/teknik-analisis-data>.
- Pratama, A., Wihandika, R. C., & Ratnawati, D. E. (2017). Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(4), 1704–1708
- Riandari, F., Sihotang, H. T., Tarigan, T. and Rafli, M. . (2022) “Classification of Book Types Using the *Support vector machine* Method”, *Jurnal Mantik*, 6(1), pp. 43-49. Available at: <http://iocscience.org/ejournal/index.php/mantik/article/view/2132> (Accessed: 26May2024).
- Wahyudi, M., Masitha, Saragih, R., & Solikhun. (2020). Data Mining: Penerapan Algoritma *K-Means Clustering* dan *K-Medoids Clustering* Oleh Mochamad Wahyudi, Masi.
- Wang, Y., Li, D., & Wang, Y. (2019). Realization of remote sensing image segmentation based on K-means clustering. *IOP Conference Series: Materials Science and Engineering*, 490(1), 072008.
- World Health Organization. (2019). *Diagnosis and classification of diabetes mellitus* (pp. 1–49). Geneva: WHO.
- Wulandari, C., Aviani, T. H. B., & Saputra, R. (2024). Penerapan Algoritma *Support vector machine* Untuk Prediksi Tingkat Kelulusan Siswa SMA. *Resolusi: Rekayasa Teknik Informatika dan Informasi*, 4(4), 397- 407.
- Zakir, A. (2022). Implementasi Algoritma *K-Means* Untuk *Clustering* Judul Skripsi Universitas Harapan Medan. *Jurnal Media Informatika*, 4(1), 40-47.

LAMPIRAN

Lampiran 1. Hasil Cluster

JENIS KELAMIN	Usia	Status Tekanan Darah	Hasil Tekanan Darah (Sistol)	Hasil Tekanan Darah (Diastol)	Status - Gula Darah Sewaktu	Hasil - Gula Darah Sewaktu	Status IMT	Hasil IMT	Status Lingkar Perut	Hasil Lingkar Perut	Tajam Penglihatan	Tajam Pendengaran	Kluster
0	45	2	150	80	2	282	0	24,26666667	1	89	1	0	0
0	51	1	130	70	0	124	2	25,0995016	1	87	0	0	1
0	48	2	170	100	2	531	0	24,65483235	1	99	1	0	0
1	59	2	150	80	2	264	0	23,19109462	1	87	1	0	0
1	61	2	160	80	2	405	0	18,71804237	0	79	0	1	0
1	47	1	120	70	2	217	1	14,17769376	0	80	0	1	1
1	34	1	120	70	2	200	3	27,92666997	1	85	0	0	1
1	54	0	130	80	1	152	0	22,50692521	1	92	1	0	0
1	45	1	130	80	2	236	0	23,7332384	1	82	0	0	1
0	59	1	120	70	2	217	1	14,17769376	0	83	0	0	1
1	55	2	200	130	2	341	3	28,57142857	1	100	0	0	0
0	45	0	120	80	1	199	0	23,30905307	1	82	1	0	1
0	63	0	130	80	0	112	3	63,7755102	1	87	1	1	0
1	43	1	120	70	2	278	3	28,84153181	1	106	0	0	1
0	59	0	120	80	1	173	2	25,33333333	1	89	1	0	1
0	74	2	190	100	2	262	0	23,94245762	1	81	0	1	0
1	78	2	170	100	2	274	0	21,87871582	0	75	1	1	0
0	57	2	180	100	2	210	0	19,75646731	1	92	1	0	0
0	52	1	130	90	0	94	0	24,22145329	1	92	0	0	0
0	51	2	190	110	1	162	0	19,33372992	1	97	1	0	0

0	73	1	130	60	2	320	0	24,30461788	0	84	1	1	1
0	61	2	140	80	2	209	0	21,20844875	0	74	1	1	1
0	55	1	120	60	2	403	0	23,14725512	0	80	1	0	1
0	28	1	120	60	2	347	0	24,65483235	1	91	0	0	1
1	76	2	150	90	2	229	0	21,62064772	0	78	1	1	0
0	66	2	170	100	2	218	0	21,09375	0	84	1	1	0
0	64	2	160	100	2	379	3	28,96942801	1	102	0	1	0
0	47	1	130	80	0	112	0	23,55555556	1	83	0	0	1
1	49	2	140	80	2	288	0	19,2391584	0	71	0	0	1
0	59	2	180	100	2	466	0	21,64127424	1	85	1	0	0
0	65	0	110	70	0	124	0	19,22768787	1	87	1	1	1
0	58	0	90	60	2	368	0	22,10028959	0	78	1	0	1
1	36	1	130	80	2	212	3	30,38194444	1	86	0	0	1
0	48	2	170	90	2	203	0	24,9107674	1	98	1	0	0
0	59	1	120	80	2	372	0	20,93212013	0	75	1	0	1
1	36	0	110	60	2	225	0	22,07409972	1	85	0	0	1
0	64	2	190	110	2	212	0	21,87242282	1	94	1	1	0
0	30	1	130	70	2	216	2	26,03749399	1	91	0	0	1
0	49	2	150	100	2	253	0	23,50780533	1	99	0	0	0
0	52	1	130	80	2	398	3	28,7890625	1	104	0	0	0
0	48	0	100	70	2	238	0	23,8330046	0	80	0	0	1
0	48	2	110	70	2	210	3	27,91551883	1	110	0	0	1
1	44	2	140	80	2	285	0	23831224602	0	79	0	0	1
0	58	2	220	120	2	328	2	26,35907202	1	100	1	0	0
0	49	1	130	70	0	126	2	25,40281608	1	93	0	0	1
1	66	2	200	100	2	211	0	18,59690844	0	68	1	1	0

1	66	2	150	70	2	201	0	23,80540166	0	86	0	1	0
0	51	0	100	70	2	239	3	31,20256367	1	99	1	0	1
1	45	1	120	80	2	211	1	17,5279431	0	71	1	0	1
0	51	2	170	80	1	143	0	25,78125	1	90	1	0	0
0	45	0	100	70	0	127	0	19,1	1	82	1	0	1
0	49	2	150	100	0	108	0	22	1	87	0	0	0
0	54	2	150	80	2	1109	2	25,91512796	1	90	0	0	0
0	91	2	150	80	2	297	3	27,58110047	1	105	1	1	0
0	33	1	120	70	2	278	3	28,84153181	1	106	0	1	1
0	60	0	110	70	0	113	0	19,33372992	1	87	1	1	1
1	58	2	150	80	2	211	1	12,5743027	0	70	0	0	1
0	52	2	190	80	2	228	0	23,78121284	1	91	1	0	0
1	47	0	100	60	0	104	0	19,33372992	1	97	1	0	1
1	48	2	170	80	2	244	1	16,55172414	0	60	1	0	1
0	47	2	200	100	2	302	3	36,69906765	1	105	1	0	0
0	46	1	130	80	1	189	3	30,8328938	1	103	0	0	0
1	29	0	110	70	2	250	3	29,7731569	1	88	0	0	1
1	64	0	100	70	2	218	2	26,39714625	1	95	0	1	1
0	60	1	120	70	2	245	3	70,0585245	1	88	1	1	1
1	53	1	130	80	2	332	3	27,35884546	1	111	1	0	0
0	52	2	140	90	2	292	2	25,39021852	1	100	1	0	0
0	48	1	130	80	2	332	3	27,35884546	1	111	1	0	0
0	66	2	150	80	2	297	3	27,58110047	1	105	0	1	0
0	84	2	150	70	2	276	0	21,2	0	85	0	1	0
0	78	2	140	80	2	335	3	33,15565553	1	89	0	0	0
0	79	0	110	70	0	106	0	25,78125	1	83	1	1	1

1	26	1	110	80	2	211	0	22,89281998	1	81	0	0	1
1	22	0	110	80	2	236	3	33,15565553	1	89	0	0	1
0	77	2	150	80	0	130	0	23,30905307	1	95	1	1	0
1	37	2	140	80	2	200	3	33,15565553	1	89	0	0	1
0	75	2	200	110	2	310	3	28,84153181	1	84	0	1	0
1	53	1	130	80	2	257	0	19,70553242	0	80	0	0	1
1	70	1	130	80	0	102	3	27,2	1	98	1	1	0
0	67	2	170	100	2	245	0	22,10884354	1	93	1	1	0
0	39	0	100	70	2	216	0	20,04988019	0	67	0	0	1
0	68	2	200	100	0	135	0	20,42941836	1	91	1	1	0
1	30	2	190	100	2	211	0	18,59690844	0	68	0	0	0
0	48	1	120	70	2	201	0	23,80540166	0	86	1	0	1
0	60	1	120	80	2	422	2	25,29260067	1	86	0	1	1
0	31	2	140	80	2	211	1	17,5279431	0	71	0	0	1
0	54	1	130	80	2	212	3	30,38194444	1	86	1	0	1
0	34	1	120	70	2	222	3	31,22945431	1	98	1	0	1
0	48	1	15	80	2	211	0	23,72528616	0	89	0	0	1
0	57	0	100	70	1	168	0	23,30905307	1	87	1	0	1
0	33	1	120	70	2	201	0	23,80540166	0	86	0	0	1
0	55	1	120	80	2	204	3	31,46267228	1	98	1	0	0
0	36	1	130	80	2	434	0	22,2136785	0	88	0	0	1
1	40	2	190	100	2	262	0	23,94245762	1	81	0	0	0
0	55	2	220	120	0	180	3	26	1	89	1	0	0
0	24	1	130	80	2	201	0	18,54934602	0	64	0	0	1
1	45	2	150	70	2	346	1	15,3787005	0	70	0	0	1
0	68	1	130	80	2	301	0	18,54934602	0	64	1	1	1

0	55	1	130	90	0	124	0	21,6	1	84	1	0	1
1	33	0	110	70	2	227	0	22,76943835	0	76	0	0	1
1	37	1	130	70	2	216	2	26,03749399	1	91	0	0	1
1	59	2	180	100	2	258	3	33,56351993	1	106	1	0	0
...
0	21	1	120	80	2	201	0	21,35930625	0	80	0	0	1
1	37	1	130	70	2	216	2	26,03749399	1	91	0	0	1
1	45	2	120	90	2	213	0	20,81165453	0	79	0	0	1
0	39	1	130	80	2	201	0	21,35930625	0	80	1	0	1
1	21	2	170	80	2	244	1	18,31153389	0	60	0	0	1
1	59	2	150	90	2	297	3	27,58110047	1	105	0	0	0
1	31	2	150	80	2	297	3	27,58110047	1	105	0	0	0
1	54	2	140	90	2	201	0	21,484375	0	80	0	0	0
0	47	2	150	90	2	284	0	22,22222222	0	78	0	0	0
1	57	2	180	100	2	268	0	20,54419284	0	72	0	0	0
0	46	2	110	70	2	212	0	23,55555556	1	83	0	0	1
1	50	2	170	100	2	399	3	29,24210864	1	101	1	0	0
1	59	1	120	70	2	211	1	16,17122473	0	79	0	0	1
1	59	2	170	100	2	206	0	21,77777778	0	74	1	0	0
0	37	2	140	80	2	237	2	25,33333333	0	88	1	0	1
1	58	2	160	90	2	399	3	29,24210864	1	101	1	0	0
1	68	2	140	80	2	200	0	24,94009487	1	97	0	1	0
1	26	1	130	80	2	332	3	27,35884546	1	111	1	0	0
0	17	2	140	80	2	267	0	23,30905307	0	86	0	0	1
1	76	1	120	70	2	245	3	27,339912	0	88	0	1	0

1	46	2	170	80	2	236	0	23,7332384	1	82	0	0	0
0	44	2	150	80	2	300	3	66,21315193	1	89	0	0	0
0	66	2	130	70	2	380	0	19,6282647	1	82	1	1	1
1	40	2	140	80	2	493	3	27,46806784	1	102	1	0	0
1	24	2	130	80	2	267	1	17,60474825	0	70	0	0	1
1	64	2	140	80	2	202	3	33,78378378	1	104	1	1	0
1	42	2	150	80	2	201	0	22,82688094	0	80	0	0	1
0	41	2	140	90	2	211	0	20,81165453	0	79	0	0	1
0	21	2	200	100	2	258	0	24,45606342	1	91	0	0	0
0	17	2	150	80	2	395	2	25,15315023	1	100	1	0	1
0	37	2	150	100	2	253	0	23,50780533	1	99	0	0	0
1	34	2	140	80	2	208	3	29,59183674	1	97	0	0	1
1	16	2	110	80	2	240	3	30,25193414	1	98	1	0	1
1	50	2	150	100	2	203	0	24,9107674	1	98	1	0	0
0	67	2	140	80	2	211	3	208,3333333	0	79	0	1	0
1	45	2	110	70	2	201	0	20,504934	0	80	1	0	1
0	54	2	130	90	2	206	0	24,88888889	0	74	0	0	1
0	50	2	150	80	2	201	3	562,1301775	0	80	0	0	1
1	26	2	140	80	2	211	0	21,05170826	0	79	0	0	1
1	19	2	100	70	2	490	2	25,29937595	1	96	0	0	1
1	40	2	110	70	2	600	1	18,22222222	1	82	1	0	1
0	37	2	140	80	2	237	3	44,20057918	1	132	0	0	0
1	15	2	160	90	2	326	3	31,32906055	1	107	1	0	0
1	63	2	160	90	2	216	2	26,03749399	1	91	0	1	0
1	39	2	150	60	2	413	0	23,61275089	0	87	0	0	1
1	34	2	140	50	2	390	0	24,97398543	1	88	1	0	1

1	54	1	120	70	2	318	0	22,43230252	1	86	1	0	1
0	63	2	130	80	2	345	0	23,73995617	1	82	1	1	1
1	59	2	140	80	2	301	1	17,96875	1	86	0	0	0
1	34	2	220	100	2	220	3	30,03917571	1	87	0	0	0
0	27	2	200	90	2	200	1	18,3950831	0	78	0	0	0
1	34	2	200	100	2	200	2	25,88757396	0	90	1	0	0
0	37	2	130	80	2	201	0	20,02884153	1	81	1	0	1
1	45	2	140	80	2	321	0	24,14151925	1	93	0	0	0
1	44	2	170	100	2	219	0	21,00073046	0	79	1	0	0
0	42	2	160	80	2	219	0	22,65625	0	88	1	0	0
1	23	2	110	70	2	210	3	27,91551883	1	110	0	0	1
0	21	2	140	80	2	276	0	21,2	0	85	0	0	1
1	32	2	120	70	2	395	0	20,613131	1	86	1	0	1
1	23	2	170	100	2	272	2	25,77777778	1	92	1	0	0
1	56	0	110	70	2	255	0	24,03460984	1	85	0	0	1
1	51	2	140	80	2	517	1	18,08021039	1	85	0	0	0
1	25	2	170	80	2	257	0	19,70553242	0	80	0	0	1
1	61	1	130	70	2	200	0	24,94009487	1	97	1	1	0
1	61	1	120	60	2	356	0	18,76524676	1	92	0	1	1
0	54	2	120	70	2	216	0	20,04988019	0	67	0	0	1
1	20	2	100	70	2	328	0	21,33333333	1	81	0	0	1
1	54	2	150	80	2	211	0	18,59690844	0	68	0	0	1
1	48	2	130	80	2	395	0	20,613131	1	86	1	0	1
1	50	2	140	80	2	227	0	22,76943835	0	76	0	0	1
1	27	2	100	70	2	216	2	26,03749399	1	91	0	0	1
1	46	2	150	80	2	222	3	31,22945431	1	98	1	0	0

1	50	2	140	80	2	211	1	16,17122473	0	79	1	0	1
1	83	2	160	80	2	200	0	24,94009487	1	97	0	1	0
1	67	2	150	80	2	236	0	23,7332384	1	82	0	1	0
1	51	2	150	80	2	339	3	29,24210864	1	101	1	0	0
1	52	2	170	80	2	300	1	18,31425598	0	77	0	0	0
1	54	2	180	100	2	222	3	31,22945431	1	98	1	0	0
1	41	2	150	80	2	333	0	21,75019668	1	88	1	0	0
1	60	2	170	80	2	345	0	23,73995617	1	82	1	1	0
1	56	2	190	100	2	211	0	18,59690844	0	68	0	0	0
1	76	2	180	100	2	262	0	23,94245762	1	81	0	1	0
1	70	0	110	70	2	227	0	22,76943835	0	76	0	1	1
0	73	2	130	86	2	243	3	27,76709813	1	94	0	1	0
1	55	2	150	90	2	284	0	22,22222222	0	78	1	0	0
1	55	2	120	70	2	395	0	20,613131	1	86	1	0	1
1	51	2	140	70	2	282	0	20,23950076	1	86	0	0	1
1	57	0	110	70	2	204	2	25,33308313	1	94	0	0	1
0	58	2	140	50	2	245	3	70,0585245	1	88	1	0	1
1	60	2	170	100	2	345	0	23,73995617	1	82	1	1	0
1	52	2	140	80	2	206	3	30,53090072	1	94	0	0	0
1	55	2	160	90	2	450	0	21,51694459	1	91	0	0	0
1	66	2	130	81	2	380	0	19,6282647	1	82	1	1	0
1	61	2	140	80	2	255	0	23,19109462	1	82	0	1	0
1	62	2	180	80	2	321	1	17,77777778	0	76	0	1	0
1	53	2	200	100	2	310	0	21,62064772	1	84	1	0	0
1	63	2	130	90	2	312	0	19,53125	1	81	0	1	0
1	60	2	120	80	2	1109	2	25,91512796	1	90	0	1	0

1	66	2	110	80	2	262	0	23,94245762	1	81	0	1	1
1	49	2	140	90	2	204	2	25,29937595	1	93	0	0	0
1	38	2	140	80	2	232	2	26,89618074	1	99	0	0	0



Lampiran 2. Prediksi SVM

NO	JENIS KELAMIN	Usia	Label Asli	Prediksi	Hasil Akurasi
1	1	71	Diabetes Ringan	Diabetes Ringan	1
2	1	54	Diabetes Kronis	Diabetes Kronis	1
3	0	59	Diabetes Kronis	Diabetes Kronis	1
4	0	58	Diabetes Kronis	Diabetes Kronis	1
5	1	58	Diabetes Ringan	Diabetes Ringan	1
6	0	39	Diabetes Ringan	Diabetes Ringan	1
7	1	57	Diabetes Kronis	Diabetes Kronis	1
8	1	48	Diabetes Kronis	Diabetes Kronis	1
9	0	47	Diabetes Kronis	Diabetes Kronis	1
10	1	50	Diabetes Kronis	Diabetes Kronis	1
11	1	42	Diabetes Kronis	Diabetes Kronis	1
12	0	67	Diabetes Kronis	Diabetes Kronis	1
13	0	53	Diabetes Ringan	Diabetes Ringan	1
14	1	27	Diabetes Kronis	Diabetes Kronis	1
15	0	59	Diabetes Ringan	Diabetes Ringan	1
16	0	75	Diabetes Ringan	Diabetes Ringan	1
17	1	42	Diabetes Kronis	Diabetes Kronis	1
18	1	58	Diabetes Ringan	Diabetes Ringan	1
19	1	47	Diabetes Ringan	Diabetes Ringan	1
20	0	69	Diabetes Kronis	Diabetes Kronis	1

21	0	27	Diabetes Kronis	Diabetes Ringan	0
22	1	44	Diabetes Ringan	Diabetes Ringan	1
23	1	52	Diabetes Kronis	Diabetes Kronis	1
24	1	54	Diabetes Kronis	Diabetes Kronis	1
25	0	34	Diabetes Kronis	Diabetes Kronis	1
26	1	43	Diabetes Ringan	Diabetes Ringan	1
27	0	38	Diabetes Kronis	Diabetes Kronis	1
28	0	54	Diabetes Kronis	Diabetes Ringan	0
29	1	60	Diabetes Kronis	Diabetes Kronis	1
30	1	33	Diabetes Kronis	Diabetes Kronis	1
31	0	68	Diabetes Ringan	Diabetes Ringan	1
32	0	49	Diabetes Ringan	Diabetes Ringan	1
33	1	54	Diabetes Kronis	Diabetes Kronis	1
34	1	34	Diabetes Kronis	Diabetes Ringan	0
35	0	37	Diabetes Kronis	Diabetes Kronis	1
36	1	64	Diabetes Ringan	Diabetes Ringan	1
37	1	35	Diabetes Kronis	Diabetes Kronis	1
38	1	35	Diabetes Kronis	Diabetes Kronis	1
39	1	41	Diabetes Kronis	Diabetes Kronis	1
40	1	48	Diabetes Kronis	Diabetes Kronis	1
41	0	54	Diabetes Kronis	Diabetes Kronis	1
42	1	42	Diabetes Kronis	Diabetes Kronis	1
43	1	84	Diabetes Ringan	Diabetes Ringan	1

44	1	37	Diabetes Kronis	Diabetes Kronis	1
45	1	51	Diabetes Ringan	Diabetes Ringan	1
46	1	36	Diabetes Kronis	Diabetes Kronis	1
47	1	53	Diabetes Kronis	Diabetes Kronis	1
48	0	21	Diabetes Kronis	Diabetes Kronis	1
49	0	47	Diabetes Ringan	Diabetes Ringan	1
50	1	62	Diabetes Ringan	Diabetes Ringan	1
51	1	54	Diabetes Ringan	Diabetes Ringan	1
52	0	51	Diabetes Ringan	Diabetes Ringan	1
53	1	47	Diabetes Ringan	Diabetes Ringan	1
54	0	37	Diabetes Ringan	Diabetes Ringan	1
55	0	37	Diabetes Ringan	Diabetes Ringan	1
56	1	52	Diabetes Ringan	Diabetes Ringan	1
57	1	30	Diabetes Kronis	Diabetes Kronis	1
58	0	66	Diabetes Kronis	Diabetes Ringan	0
59	1	55	Diabetes Ringan	Diabetes Ringan	1
60	0	28	Diabetes Kronis	Diabetes Kronis	1
61	0	64	Diabetes Ringan	Diabetes Ringan	1
62	1	47	Diabetes Kronis	Diabetes Kronis	1
63	0	78	Diabetes Ringan	Diabetes Ringan	1
64	1	38	Diabetes Ringan	Diabetes Ringan	1
65	1	36	Diabetes Kronis	Diabetes Kronis	1
66	1	55	Diabetes Kronis	Diabetes Kronis	1

67	1	64	Diabetes Ringan	Diabetes Ringan	1
68	1	50	Diabetes Kronis	Diabetes Ringan	0
69	1	40	Diabetes Ringan	Diabetes Ringan	1
70	0	40	Diabetes Kronis	Diabetes Kronis	1
71	1	17	Diabetes Kronis	Diabetes Kronis	1
72	1	50	Diabetes Ringan	Diabetes Ringan	1
73	1	67	Diabetes Kronis	Diabetes Ringan	0
74	1	67	Diabetes Ringan	Diabetes Ringan	1
75	1	37	Diabetes Ringan	Diabetes Ringan	1
...
174	1	50	Diabetes Kronis	Diabetes Kronis	1
175	1	34	Diabetes Ringan	Diabetes Ringan	1
176	1	33	Diabetes Kronis	Diabetes Kronis	1
177	1	36	Diabetes Ringan	Diabetes Ringan	1
178	1	37	Diabetes Kronis	Diabetes Kronis	1
179	1	59	Diabetes Ringan	Diabetes Ringan	1
180	0	60	Diabetes Kronis	Diabetes Kronis	1
181	1	63	Diabetes Ringan	Diabetes Ringan	1
182	0	29	Diabetes Ringan	Diabetes Ringan	1
183	0	66	Diabetes Ringan	Diabetes Ringan	1
184	1	26	Diabetes Kronis	Diabetes Kronis	1
185	0	37	Diabetes Kronis	Diabetes Kronis	1
186	0	49	Diabetes Kronis	Diabetes Kronis	1

187	1	48	Diabetes Kronis	Diabetes Ringan	0
188	0	35	Diabetes Ringan	Diabetes Ringan	1
189	0	58	Diabetes Kronis	Diabetes Kronis	1
190	0	68	Diabetes Kronis	Diabetes Kronis	1
191	1	20	Diabetes Kronis	Diabetes Kronis	1
192	1	77	Diabetes Kronis	Diabetes Kronis	1
193	1	26	Diabetes Ringan	Diabetes Ringan	1
194	0	42	Diabetes Kronis	Diabetes Kronis	1
195	0	22	Diabetes Ringan	Diabetes Ringan	1
196	1	62	Diabetes Kronis	Diabetes Kronis	1
197	1	58	Diabetes Ringan	Diabetes Ringan	1
198	0	23	Diabetes Kronis	Diabetes Kronis	1
199	0	42	Diabetes Kronis	Diabetes Kronis	1
200	1	38	Diabetes Kronis	Diabetes Kronis	1
201	1	46	Diabetes Kronis	Diabetes Kronis	1
202	1	23	Diabetes Ringan	Diabetes Ringan	1
203	1	49	Diabetes Kronis	Diabetes Kronis	1
204	1	34	Diabetes Kronis	Diabetes Kronis	1
205	1	35	Diabetes Kronis	Diabetes Kronis	1
206	0	47	Diabetes Ringan	Diabetes Ringan	1
207	1	57	Diabetes Kronis	Diabetes Kronis	1
208	1	46	Diabetes Ringan	Diabetes Ringan	1
Jumlah prediksi benar:					195
persentase prediksi benar :					0,9375

Lampiran 3. Source Code

```
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Memuat file Excel
jalur_file = 'Data DM puskesmas lappae - Copy.xlsx'
df = pd.read_excel(jalur_file, sheet_name='Sheet1')

# Memilih kolom numerik yang relevan untuk klustering
kolom_numerik = [
    'Usia',
    'Hasil - Tekanan Darah (Sistol)',
    'Hasil - Tekanan Darah (Diastol)',
    'Hasil - Gula Darah Sewaktu',
    'Hasil - IMT',
    'Hasil - Lingkar Perut'
]

# Membersihkan kolom numerik dengan mengubah string dengan koma
menjadi float
for kolom in kolom_numerik:
    df[kolom] = df[kolom].replace(',', '.').
    regex=True).astype(float)

# Praproses data dengan menstandarkan
skaler = StandardScaler()
data_distandarkan = skaler.fit_transform(df[kolom_numerik])

# Menerapkan K-Means klustering dengan 2 kluster
kmeans = KMeans(n_clusters=2, random_state=42)
kmeans.fit(data_distandarkan)
df['Kluster'] = kmeans.labels_

# Plot kluster
plt.figure(figsize=(10, 6))
plt.scatter(data_distandarkan[:, 0], data_distandarkan[:, 1],
            c=df['Kluster'], cmap='viridis')
plt.xlabel('Usia (distandarkan)')
plt.ylabel('Hasil - Tekanan Darah (Sistol) (distandarkan)')
plt.title('Klustering K-Means (2 Kluster)')
plt.colorbar(label='Kluster')
plt.show()
```

```

# Menyimpan dataframe dengan label kluster
df.to_excel('Hasil_Kluster.xlsx', index=False)

# Menampilkan beberapa baris pertama dari dataframe dengan label
kluster
print(df.head())

df=pd.read_excel("Hasil_Kluster.xlsx",sheet_name="Sheet1")
df

# Kolom-kolom yang akan dihapus
kolom_yang_dihapus = ['Status - Sadanis', 'Hasil - Sadanis kanan',
                      'Hasil - Sadanis kiri', 'Unnamed: 16']

# Menghapus kolom-kolom tersebut
df = df.drop(columns=kolom_yang_dihapus)

df
x=df.iloc[:,1:15]
y=df.iloc[:,15]

x
y

from sklearn.preprocessing import StandardScaler

# Terapkan SMOTE untuk oversampling kelas minoritas
smote = SMOTE(random_state=42)
x_train, y_train = smote.fit_resample(x_train, y_train)

# Normalisasi data (opsional, terutama untuk kernel RBF)
scaler = StandardScaler()
x_train = scaler.fit_transform(x_train)
x_test = scaler.transform(x_test)

from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report

# Inialisasi model SVM dengan kernel RBF
model = SVC(kernel='rbf', random_state=42)

```

```

# Latih model menggunakan data pelatihan
model.fit(x_train, y_train)

# Prediksi pada data uji
y_pred = model.predict(x_test)

# Evaluasi model
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred, zero_division=1)

# Cetak hasil
print("Accuracy:", accuracy)
print("Classification Report:\n", report)

# Dictionary untuk pemetaan label numerik ke label teks
label_map = {1: 'Diabetes Kronis', 0: 'Diabetes Ringan'}

# Gabungkan data uji, hasil prediksi, dan kolom yang diperlukan
results_df = pd.DataFrame({
    'Nama Peserta Deteksi Dini': nama_test,
    'JENIS KELAMIN, MALE/FEMALE': jenis_kelamin_test,
    'Usia': usia_test,
    'Label Asli': y_test,
    'Prediksi': y_pred
})

# Mengganti nilai label numerik dengan label teks menggunakan map
results_df['Label Asli'] = results_df['Label Asli'].map(label_map)
results_df['Prediksi'] = results_df['Prediksi'].map(label_map)

# Ambil kolom-kolom yang perlu diekspor ke Excel
export_df = results_df[['Nama Peserta Deteksi Dini', 'JENIS
KELAMIN, MALE/FEMALE', 'Usia', 'Label Asli', 'Prediksi']]

# Simpan ke file Excel
file_path = 'Hasil Prediksi SVM.xlsx'
export_df.to_excel(file_path, index=False)

print(f"Hasil prediksi telah diekspor ke file: {file_path}")

```

Lampiran 4 Scan Hasil Plagiasi PerBAB



MAJELIS PENDIDIKAN TINGGI PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH MAKASSAR
UPT PERPUSTAKAAN DAN PENERBITAN

Alamat kantor: Jl.Sultan Alauddin NO.259 Makassar 90221 Tlp.(0411) 866972,881593, Fax.(0411) 865588

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

SURAT KETERANGAN BEBAS PLAGIAT

UPT Perpustakaan dan Penerbitan Universitas Muhammadiyah Makassar,
Menerangkan bahwa mahasiswa yang tersebut namanya di bawah ini:

Nama : M.Fajar Putra
Nim : 105841112220
Program Studi : Teknik Informatika

Dengan nilai:

No	Bab	Nilai	Ambang Batas
1	Bab 1	9 %	10 %
2	Bab 2	21 %	25 %
3	Bab 3	7 %	10 %
4	Bab 4	0 %	10 %
5	Bab 5	4 %	5 %

Dinyatakan telah lulus cek plagiat yang diadakan oleh UPT- Perpustakaan dan Penerbitan Universitas Muhammadiyah Makassar Menggunakan Aplikasi Turnitin.

Demikian surat keterangan ini diberikan kepada yang bersangkutan untuk dipergunakan seperlunya.

Makassar, 06 Agustus 2024

Mengetahui,

Kepala UPT- Perpustakaan dan Penerbitan,



Nursikah, Ham., M.I.P
NPM. 964 591

Jl. Sultan Alauddin no 259 makassar 90222
Telepon (0411)866972,881 593,fax (0411)865 588
Website: www.library.unismuh.ac.id
E-mail : perpustakaan@unismuh.ac.id

1

M.Fajar Putra 105841112220

Bab I

by Tahap Tutup



Submission date: 06-Aug-2024 12:35PM (UTC+0700)

Submission ID: 2428031594

File name: BAB_1_81.docx (13.76K)

Word count: 869

Character count: 5883

ORIGINALITY REPORT

9% SIMILARITY INDEX
10% INTERNET SOURCES
8% PUBLICATIONS
9% STUDENT PAPERS

PRIMARY SOURCES



1	bpkpenabur.or.id Internet Source	2%
2	rama.unimal.ac.id Internet Source	2%
3	Submitted to Poltekkes Kemenkes Malang Student Paper	2%
4	cupdf.com Internet Source	2%
5	eprints.walisongo.ac.id Internet Source	2%

Exclude quotes 0%
Exclude bibliography 0%

Exclude matches



M.Fajar Putra 105841112220

Bab II

by Tahap Tutup

Submission date: 06-Aug-2024 12:36PM (UTC+0700)

Submission ID: 2428032024

File name: BAB_2_60.docx (45.66K)

Word count: 1734

Character count: 11607

M.Fajar Putra 105841112220 Bab II

ORIGINALITY REPORT

21%
SIMILARITY INDEX

21%
INTERNET SOURCES

6%
PUBLICATIONS

2%
STUDENT PAPERS

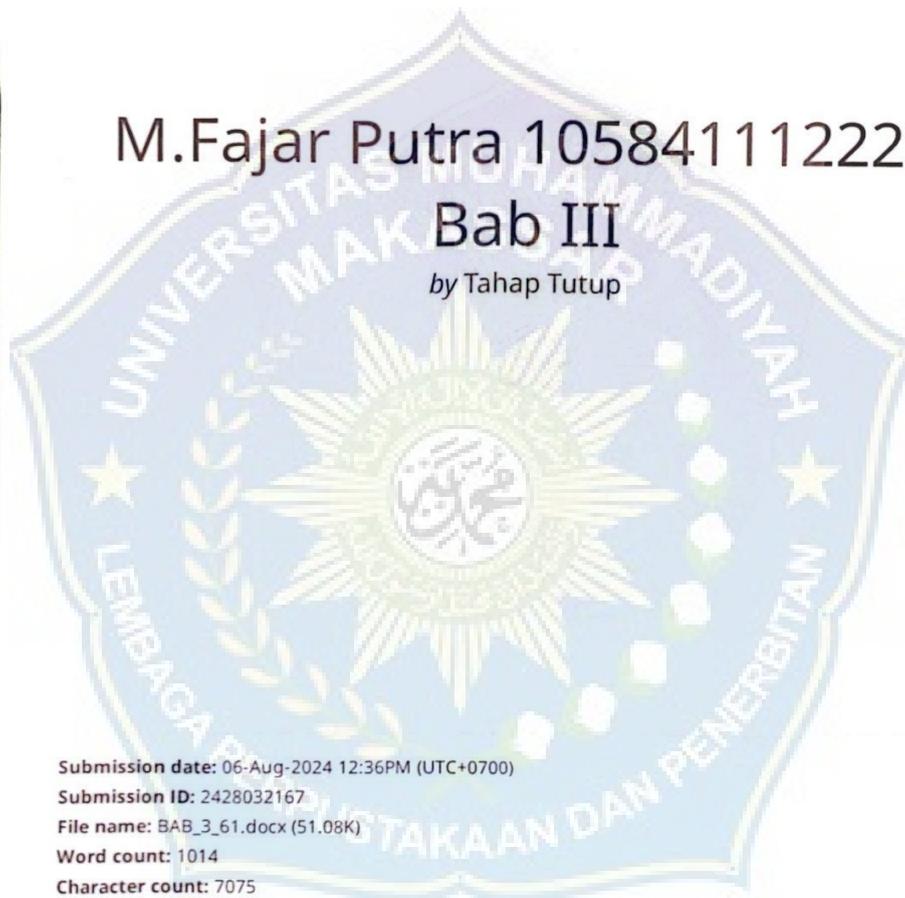
PRIMARY SOURCES

1	ejournal.uin-suska.ac.id Internet Source		3%
2	www.jurnal.iaii.or.id Internet Source		3%
3	pdfs.semanticscholar.org Internet Source		3%
4	repositori.unsil.ac.id Internet Source		2%
5	www.researchgate.net Internet Source		2%
6	www.scilit.net Internet Source		2%
7	repo.poltekkes-medan.ac.id Internet Source		2%
8	ojs.unud.ac.id Internet Source		2%
9	ejournal.poltektegale.ac.id Internet Source		2%

M.Fajar Putra 105841112220

Bab III

by Tahap Tutup



Submission date: 06-Aug-2024 12:36PM (UTC+0700)

Submission ID: 2428032167

File name: BAB_3_61.docx (51.08K)

Word count: 1014

Character count: 7075

ORIGINALITY REPORT

7%	7%	0%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	digilibadmin.unismuh.ac.id Internet Source	7%
----------	--	-----------



Exclude quotes Off Exclude matches 2%

Exclude bibliography Off



M.Fajar Putra 105841112220

Bab IV

by Tahap Tutup

Submission date: 06-Aug-2024 12:37PM (UTC+0700)

Submission ID: 2428032345

File name: BAB_4_51.docx (880.82K)

Word count: 2753

Character count: 17313

ORIGINALITY REPORT

0%

SIMILARITY INDEX

2%

INTERNET SOURCES

2%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES



turnitin

Exclude quotes

Exclude matches

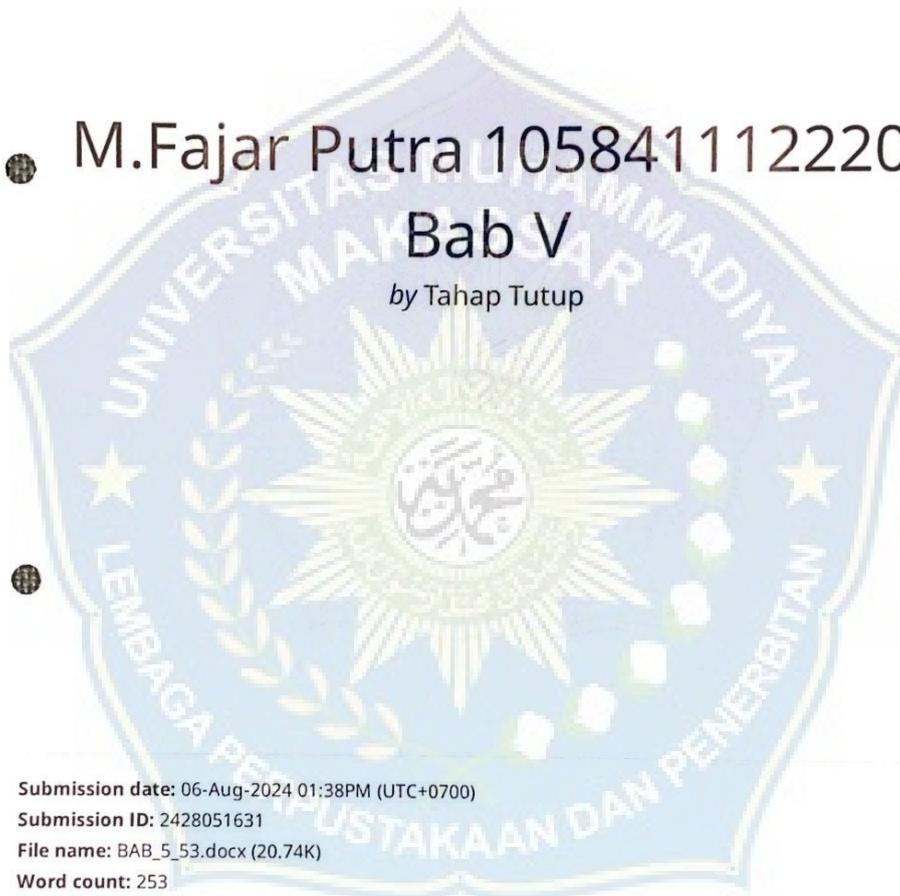
Exclude bibliography



● M.Fajar Putra 105841112220

Bab V

by Tahap Tutup



● Submission date: 06-Aug-2024 01:38PM (UTC+0700)

Submission ID: 2428051631

File name: BAB_5_53.docx (20.74K)

Word count: 253

Character count: 1640

M.Fajar Putra 105841112220 Bab V

ORIGINALITY REPORT

4%

SIMILARITY INDEX

4%

INTERNET SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1

eprints.poltekkesjogja.ac.id
Internet Source



4%

Exclude quotes

Off

Exclude matches

Exclude bibliography

Off

