

**IMPLEMENTASI ALGORITMA KNN DAN METRIK
MANHATTAN UNTUK MENGETAHUI FAKTOR RISIKO
PENYAKIT DIABETES MELITUS**

SKRIPSI

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



M. FADLIANSYAH
105841112320

**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 91222)
 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. Fadliansyah dengan nomor induk Mahasiswa 105841112320, dinyatakan diterima dan disahkan oleh Panitia Ujian Tugas Akhir/Skripsi sesuai dengan Surat Keputusan Dekan Fakultas Teknik Universitas Muhammadiyah Makassar Nomor : 0008/SK-Y/55202/091004/2024, sebagai salah satu syarat guna memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar pada hari Sabtu tanggal Senin 26 Agustus 2024.

Panitia Ujian : Makassar, 21 Safar 1446 H
 26 Agustus 2024 M

1. Pengawas Umum

- a. Rektor Universitas Muhammadiyah Makassar
 Dr. Ir. H. Abd. Rakhim Nanda, S.T., M.T., I.P.U.
- 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 : Titin Wahyuni, S.Pd., M.T.

3. Anggota

- 1. Rizki Yusliana Bakti, S.T., M.T.
- 2. Lukman Anas, S.Kom., MT.
- 3. Lukman, S.Kom., MT.

Mengetahui :

Pembimbing I

Pembimbing II

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

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



Dekan
Dr. Ir. Hj. Nurawaty, ST., MT., IPM.
 NBM: 795 108



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 KNN DAN METRIC MANHATTAN
UNTUK MENGETAHUI FAKTOR RESIKO DIABETES MELITUS**

Nama : M. Fadliansyah
Stambuk : 105 8411123 20

Makassar, 26 Agustus 2024

Telah Diperiksa dan Disetujui
Oleh Dosen Pembimbing;

Pembimbing I

Pembimbing II

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

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

Mengetahui,
Ketua Program Studi Informatika

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

NBM : 1504 577

ABSTRAK

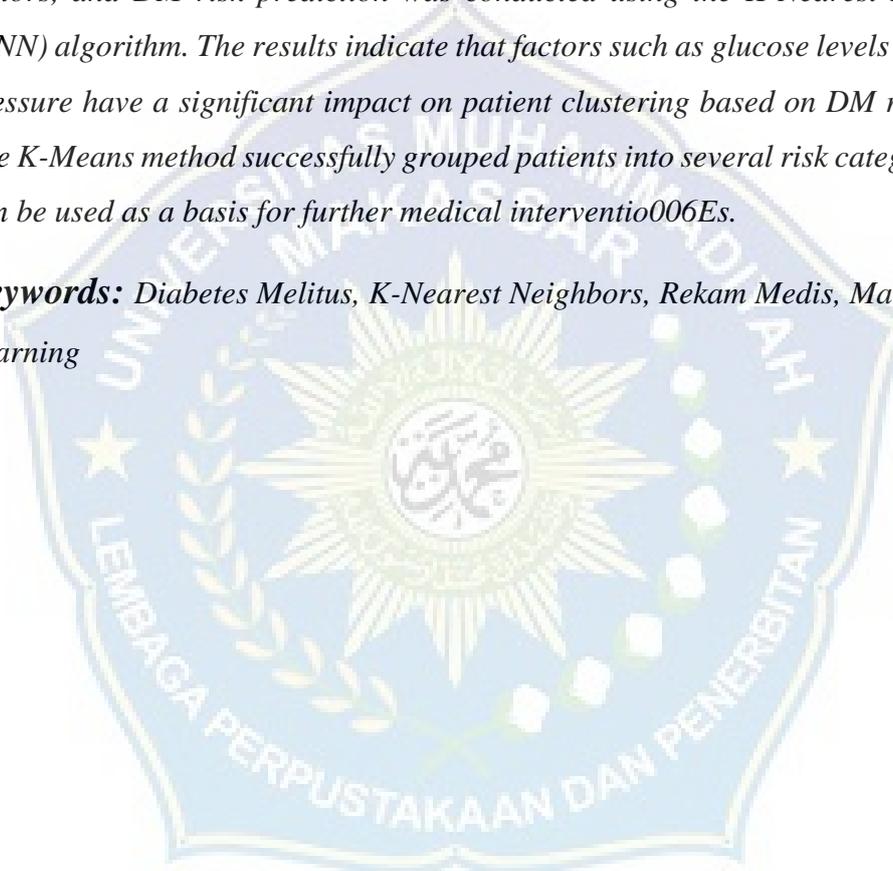
Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor risiko utama diabetes melitus (DM) dan mengelompokkan pasien berdasarkan risiko DM menggunakan metode K-Means. Data yang digunakan diperoleh dari rekam medis pasien di Puskesmas Lappae, yang mencakup variabel usia, jenis kelamin, berat badan, BMI, kadar glukosa darah puasa dan acak, serta tekanan darah. Pengelompokan dilakukan dengan mempertimbangkan faktor-faktor risiko tersebut, dan prediksi risiko DM dilakukan menggunakan algoritma K-Nearest Neighbors (KNN). Hasil penelitian menunjukkan bahwa beberapa faktor seperti kadar glukosa dan tekanan darah memiliki pengaruh signifikan dalam pengelompokan pasien berdasarkan tingkat risiko DM. Metode K-Means berhasil mengelompokkan pasien menjadi beberapa kategori risiko yang dapat digunakan sebagai dasar untuk intervensi medis lebih lanjut.

Kata Kunci: Diabetes Melitus, K-Nearest Neighbors, Rekam Medis, Machine Learning

ABSTRACT

This study aims to identify the main risk factors for diabetes mellitus (DM) and cluster patients based on DM risk using the K-Means method. The data used were obtained from patient medical records at Puskesmas Lappae, which include variables such as age, gender, weight, BMI, fasting and random blood glucose levels, and blood pressure. Clustering was performed by considering these risk factors, and DM risk prediction was conducted using the K-Nearest Neighbors (KNN) algorithm. The results indicate that factors such as glucose levels and blood pressure have a significant impact on patient clustering based on DM risk levels. The K-Means method successfully grouped patients into several risk categories that can be used as a basis for further medical interventio006Es.

Keywords: *Diabetes Melitus, K-Nearest Neighbors, Rekam Medis, Machine Learning*



KATA PENGANTAR

Assalamualaikum Wr. Wb,

Segala puji bagi Allah SWT atas limpahan rahmat, kesehatan, dan kekuatannya, sehingga peneliti dapat menyelesaikan proposal skripsi berjudul "IMPLEMENTASI ALGORITMA KNN DAN METRIK *MANHATTAN* UNTUK MENGETAHUI FAKTOR RISIKO PENYAKIT DIABETES MELITUS" sebagai salah satu persyaratan untuk penyusunan Skripsi Program Studi Informatika. Shalawat dan junjungan Nabi Muhammad SAW sebagai rahmatan lil alamin dan uswatun hasanah.

Kami juga mengucapkan banyak terima kasih kepada semua orang yang membantu dan memotivasi penyusunan proposal ini, terutama kepada:

1. Kedua Orang Tua, yang selalu mendoakan dan memberikan dukungan serta telah menyemangati penulis dan memberikan bantuan materi.
2. Aslam Muwahid, Rini Puspita, dan Astin, Keluarga yang langsung membantu penulis menyusun Skripsi dan memberi dukungan
3. Ibu Dr.Ir.Hj Nurnawati, S.T., M.T., I.P.M, selaku Dekan Fakultas Teknik.
4. Bapak Muh. Syafaat S Kuba, ST., M.T, selaku Wakil Dekan Fakultas Teknik.
5. Bapak Muhyiddin AM Hayat S.Kom., MT, selaku Ketua Prodi Informatika sekaligus Dosen Pembimbing 1 Proposal Skripsi.
6. Bapak Fahrir Irhamna Rahman S.Kom., M.T., selaku Dosen Pembimbing 2 Proposal Skripsi.
7. Dosen dan Staf Fakultas Teknik Universitas Muhammadiyah Makassar.
8. Teman-teman khususnya Angkatan 2020 Fakultas Teknik, Universitas Muhammadiyah Makassar, terima kasih atas dukungan dan doanya.

Demikian laporan proposal skripsi ini kami buat, dan kami menyadari bahwa laporan ini masih memiliki banyak kekurangan, jadi kami sangat mengharapkan kritik dan masukan dari pembaca untuk membantu memperbaikinya di masa depan.

Wassalamualaikum Wr.Wb.

Makassar, Juni 2024

M. Fadliansyah

DAFTAR ISI

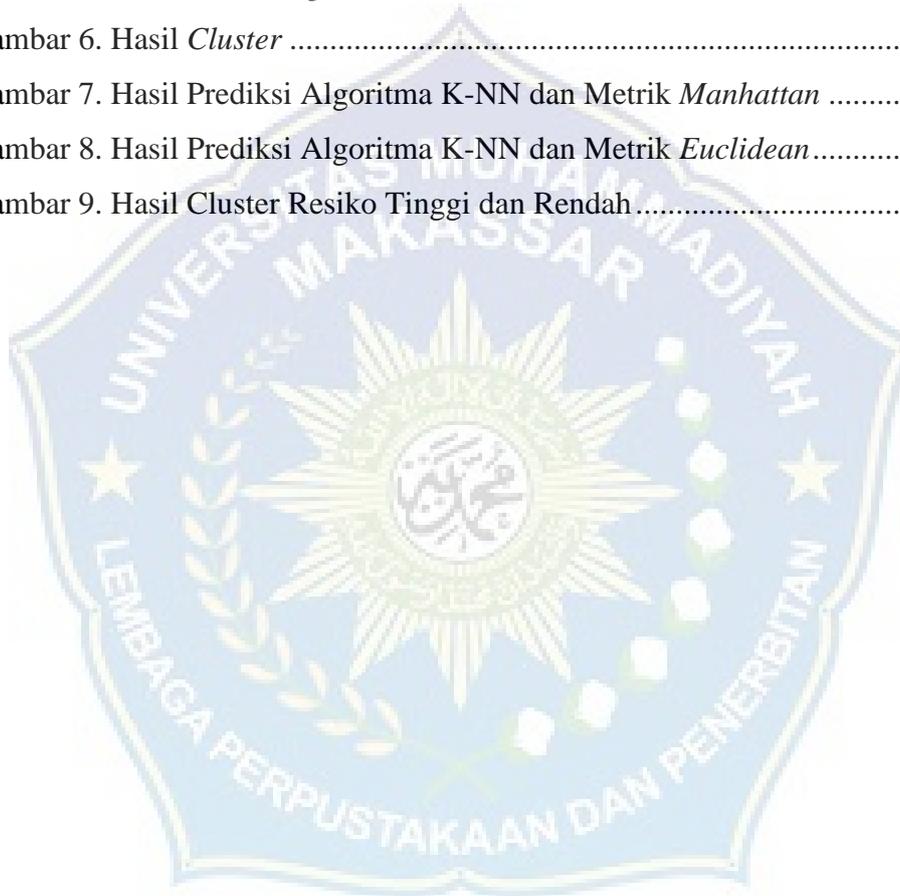
ABSTRAK.....	iii
ABSTRACT.....	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	ix
DAFTAR LAMPIRAN.....	x
DAFTAR ISTILAH.....	xi
BAB I PENDAHULUAN.....	1
A. Latar Belakang	1
B. Rumusan Masalah	2
C. Tujuan Penelitian.....	2
D. Manfaat Penelitian.....	2
E. Ruang Lingkup Penelitian	3
F. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
A. Landasan Teori.....	6
B. Penelitian Terkait.....	12
C. Kerangka Pikir	13
BAB III METODE PENELITIAN	14
A. Tempat dan Waktu Penelitian	14
B. Alat dan Bahan	14
C. Perancangan Sistem	15
D. Teknik Pengujian Sistem.....	16
E. Teknik Analisis Data	16
BAB IV HASIL DAN PEMBAHASAN	17
A. Pengambilan Data	17
B. Preprocessing	17
C. Implementasi Algoritma <i>K-Means</i>	18

D. Implementasi Metode K-NN.....	27
E. Pengujian Sistem.....	34
BAB V PENUTUP	37
A. KESIMPULAN.....	37
B. SARAN.....	38
DAFTAR PUSTAKA.....	39
LAMPIRAN.....	40



DAFTAR GAMBAR

Gambar 1. Klasifikasi Berdasarkan Tetangga Terdekat	8
Gambar 2. Tahapan <i>Knowledge Discovery in Database</i> (KDD).....	10
Gambar 3. Kerangka Pikir	13
Gambar 4. Flowchart Teknik Perancangan Sistem.....	15
Gambar 5. Hasil <i>Clustering</i>	22
Gambar 6. Hasil <i>Cluster</i>	23
Gambar 7. Hasil Prediksi Algoritma K-NN dan Metrik <i>Manhattan</i>	28
Gambar 8. Hasil Prediksi Algoritma K-NN dan Metrik <i>Euclidean</i>	28
Gambar 9. Hasil Cluster Resiko Tinggi dan Rendah.....	33



DAFTAR TABEL

Tabel 1. Penelitian Terkait	11
Tabel 2. Data Pasien Diabetes Mellitus	17
Tabel 3. Dataset Pasien Diabetes Mellitus	18
Tabel 4. Hasil <i>Clustering</i>	25
Tabel 5. Hasil Prediksi K-NN dan Metrik Manhattan	30
Tabel 6. Hasil Prediksi K-NN dan Metrik Euclidean	30
Tabel 7. Pengujian Sistem <i>Clustering</i>	32
Tabel 8. Pengujian Sistem Klasifikasi K-NN	34



DAFTAR LAMPIRAN

Lampiran 1. Data Mentah	46
Lampiran 2. <i>Preprocessing</i>	47
Lampiran 3. Hasil <i>Clustering</i>	48
Lampiran 4. Hasil Prediksi K-NN Metric Manhattan.....	49



DAFTAR ISTILAH

- KNN*** KNN adalah algoritma pembelajaran mesin yang sederhana namun efektif. Algoritma ini mengklasifikasikan data baru dengan cara mencari k data terdekat dari data baru tersebut dalam ruang fitur. Kelas dari data baru kemudian ditentukan berdasarkan mayoritas kelas dari k tetangga terdekatnya.
- Flowchart*** *Flowchart* adalah diagram alir yang menggunakan simbol-simbol standar untuk menggambarkan langkah-langkah, keputusan, dan aliran data dalam suatu proses atau algoritma.
- Machine learning*** *Machine learning* adalah cabang ilmu komputer yang memungkinkan sistem komputer untuk belajar dari data tanpa perlu diprogram secara eksplisit untuk setiap tugas. Dengan menggunakan algoritma khusus, komputer dapat mengidentifikasi pola dalam data dan menggunakan pola tersebut untuk membuat prediksi, mengambil keputusan, atau melakukan tindakan tertentu.
- Data Mining*** *Data mining* adalah teknik untuk menemukan pengetahuan yang tersembunyi dalam data. Proses ini melibatkan penggunaan algoritma dan model statistik untuk menganalisis kumpulan data yang besar dan kompleks, sehingga dapat mengidentifikasi pola, tren, dan hubungan yang signifikan.
- Classification*** Klasifikasi adalah salah satu tugas utama dalam pembelajaran mesin. Model klasifikasi digunakan untuk memisahkan data menjadi kelompok-kelompok yang berbeda berdasarkan karakteristik yang telah dipelajari dari data pelatihan..
- Scikit-learn*** *Scikit-learn* merupakan *toolkit* yang lengkap untuk melakukan analisis data dan membangun model pembelajaran mesin. Dengan menggunakan *Scikit-learn*, kita dapat dengan mudah menerapkan berbagai teknik pembelajaran mesin, mulai dari klasifikasi, regresi, hingga pengelompokan data.
- Supervised learning*** Model "diajarkan" dengan data yang sudah diberi label, sehingga model dapat belajar untuk membuat prediksi atau klasifikasi.
- Manhattan Distance*** Jarak *Manhattan* adalah jarak antara dua titik yang dihitung sebagai jumlah absolut perbedaan koordinat mereka. Ini

seperti menghitung jarak antara dua titik dengan hanya berjalan di sepanjang blok-blok kota, tanpa memotong diagonal.

Euclidean Distance Jarak *Euclidean* adalah cara untuk menghitung jarak garis lurus antara dua titik dalam ruang..

KDD *Knowledge Discovery in Databases* adalah proses menemukan pola dan informasi berarti dari data yang besar dan kompleks.



BAB I

PENDAHULUAN

A. Latar Belakang

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 (Sela, 2023). Diabetes Mellitus adalah salah satu penyakit yang masih mengancam kesehatan masyarakat di dunia, ini dibuktikan dengan tahun 2012 sekitar 1,5 juta orang di dunia kehilangan nyawa mereka karena diabetes mellitus, umumnya terjadi di negara-negara berpenghasilan menengah ke bawah, sebagian besar menyebar di sebagian Asia Tenggara (Agustiningrum & Kusbaryanto, 2019). Semua orang di Indonesia harus waspada terhadap diabetes melitus, yang merupakan salah satu penyakit tidak menular dan disebabkan oleh pankreas yang tidak dapat memproduksi hormon yang mengatur gula darah atau glukosa. Pendekatan data mining memungkinkan untuk mengidentifikasi informasi yang sebelumnya tidak diketahui. Beberapa faktor penyebab DM yaitu, Keturunan, Berat badan berlebihan, kurang aktivitas fisik, Riwayat penyakit jantung, Hipertensi, Diet tidak seimbang.

Orang-orang di masyarakat yang sangat sibuk, yang dapat mengganggu metabolisme tubuh, sering mengalami penyakit diabetes. Penyakit masyarakat seperti diabetes mellitus juga dipengaruhi oleh gaya hidup yang tidak teratur. Oleh karena itu, sangat penting bagi orang-orang untuk memahami betapa pentingnya menjaga pola hidup sehat dan berolahraga secara rutin untuk mengurangi risiko terkena penyakit. Mereka juga harus memperhatikan betapa pentingnya untuk menghindari konsumsi makanan cepat saji. Maka, mengetahui faktor-faktor ini dapat membantu Anda mengambil langkah-langkah pencegahan dan mengelola risiko diabetes melitus. Diperlukan juga sistem yang dapat membantu mengidentifikasi penyakit diabetes mellitus sehingga masyarakat dapat menangani penyakit ini dengan tepat.

K-Nearest Neighbor (K-NN), yang termasuk dalam kelompok pembelajaran instance, adalah salah satu algoritma dengan teknik pembelajaran lambat yang menggunakan metode klasifikasi terhadap objek berdasarkan data yang jaraknya paling dekat dengan objek tersebut. (Farrel Nur Rilwanu *et al.*, 2022). Ini merupakan algoritma klasifikasi berbasis supervised, yang berarti data yang dimasukkan ke dalam algoritma diberi label dan digunakan untuk mengklasifikasikan data berdasarkan kesamaannya. KNN bekerja dengan menghitung jarak antara satu data uji dengan semua data pelatihan, lalu menemukan K tetangga terdekatnya. Kemudian, berdasarkan mayoritas kategori dari tetangga-tetangga tersebut, data uji diklasifikasikan.

Algoritma *K-Nearest Neighbor* (K-NN) memiliki Salah satu kelebihan K-NN adalah melakukan pelatihan dengan cepat, sederhana, mudah dipelajari, dan efektif jika data pelatihan besar. Kekurangan K-NN termasuk nilai k yang bias, komputasi yang kompleks, dan kemungkinan tertipu dengan atribut yang tidak relevan. (Iffah'da & Anita Desiani, 2022).

Metrik Manhattan digunakan dalam algoritma *K-Nearest Neighbors* (KNN) untuk mengukur kesamaan antara data dengan menghitung jarak antara titik data dalam ruang fitur. Manhattan Distance (juga dikenal sebagai *taxicab distance* atau *city block distance*) dihitung dengan menjumlahkan selisih absolut koordinat kartesian antara dua titik. Kelebihan metrik ini meliputi kesederhanaan dan ketahanan terhadap *outlier*. Jarak Manhattan lebih disukai ketika kita memiliki data dengan dimensi tinggi karena mengatasi beberapa masalah yang muncul dalam ruang berdimensi tinggi. Namun, metrik ini tidak memperhitungkan arah dan kurang cocok untuk data kontinu. Pemilihan metrik jarak harus mempertimbangkan karakteristik data dan tujuan analisis kita.

Penelitian ini bertujuan untuk menerapkan metode *K-Nearest Neighbors* (K-NN) dan Manhattan *Distance* dalam mengidentifikasi faktor-faktor pemicu diabetes melitus dan untuk mengkaji akurasi model KNN dengan metrik jarak Manhattan dalam mengidentifikasi faktor-faktor yang memicu diabetes melitus. Dengan memahami faktor-faktor ini, kita dapat mengambil langkah-langkah pencegahan dan manajemen kondisi yang lebih baik. Dengan metode ini,

diharapkan dapat membantu dalam menganalisis data kesehatan dan menentukan variabel-variabel yang memiliki pengaruh signifikan terhadap perkembangan diabetes melitus.

Dengan mengimplementasikan algoritma K-NN menggunakan metrik jarak Manhattan, penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam bidang kesehatan, khususnya dalam upaya pencegahan dan penanganan penyakit diabetes mellitus.

B. Rumusan Masalah

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

1. Bagaimana cara mengimplementasikan algoritma K-NN dengan metrik jarak *Manhattan* untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap Diabetes Melitus?
2. Bagaimana akurasi model K-NN dengan metrik jarak *Manhattan* dalam mengidentifikasi faktor-faktor yang memicu Diabetes Melitus?

C. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan penelitian pada penelitian ini adalah:

1. Mengidentifikasi faktor-faktor yang paling signifikan dalam memicu diabetes melitus berdasarkan hasil dari model KNN dengan metrik jarak *Manhattan*.
2. Untuk mengetahui akurasi model K-NN dengan metrik jarak *Manhattan* dalam mengidentifikasi faktor-faktor yang memicu Diabetes Melitus

D. Manfaat Penelitian

Penelitian tentang “Implementasi Algoritma K-NN dan Metrik Manhattan Untuk Mengetahui Faktor Pemicu Diabetes Melitus” diharapkan dapat bermanfaat baik secara teoritis maupun secara praktis :

1. Secara Teoritis

Manfaat penelitian ini secara teoritis diharapkan agar dapat memberikan pemahaman mengenai penerapan Algoritma K-NN khususnya untuk mengetahui faktor pemicu diabetes melitus.

2. Secara Praktis

a. Bagi Peneliti

- 1) Peneliti dapat mengembangkan keahlian dalam penggunaan algoritma *K-Nearest Neighbors* (K-NN) dan perhitungan jarak *Manhattan*, serta dalam analisis data kesehatan.
- 2) Hasil penelitian dapat dipublikasikan di jurnal-jurnal ilmiah bereputasi, yang membantu peneliti mendapatkan pengakuan dan memperluas jaringan profesional mereka.

b. Bagi Universitas

Penelitian yang berhasil dan dipublikasikan dapat meningkatkan reputasi universitas sebagai institusi yang berkontribusi pada pengetahuan ilmiah dan penelitian inovatif.

E. Ruang Lingkup Penelitian

1. Lokasi Penelitian: Puskesmas Lappae.
2. Rekam Medis Pasien: Variabel yang termasuk dalam penelitian adalah: Usia, Jenis Kelamin, Berat Badan, Hasil dan status indeks massa tubuh (BMI), Kadar gula darah (gula darah puasa dan gula darah sewaktu), Tekanan darah, dan Status tekanan darah
3. Data Pasien: Data dikumpulkan dari periode tertentu untuk memastikan relevansi dan aktualitas, terbatas pada data yang tersedia di Puskesmas Lappae.
4. *K-Means Clustering*: Digunakan untuk mengelompokkan pasien berdasarkan faktor risiko diabetes melitus (DM).
5. *K-Nearest Neighbors* (KNN): Digunakan untuk klasifikasi dan prediksi risiko DM.
6. Pengumpulan Data: Mengumpulkan data dari rekam medis pasien di Puskesmas Lappae.
7. *Preprocessing* Data: Meliputi pembersihan, normalisasi, dan pembagian data menjadi set pelatihan dan pengujian.

8. Analisis Hasil: Mengidentifikasi faktor risiko utama berdasarkan hasil clustering dan prediksi KNN.

F. Sistematika Penulisan

Secara garis besar penulisan ini terbagi menjadi beberapa bagian yang tersusun sebagai berikut:

BAB I PENDAHULUAN

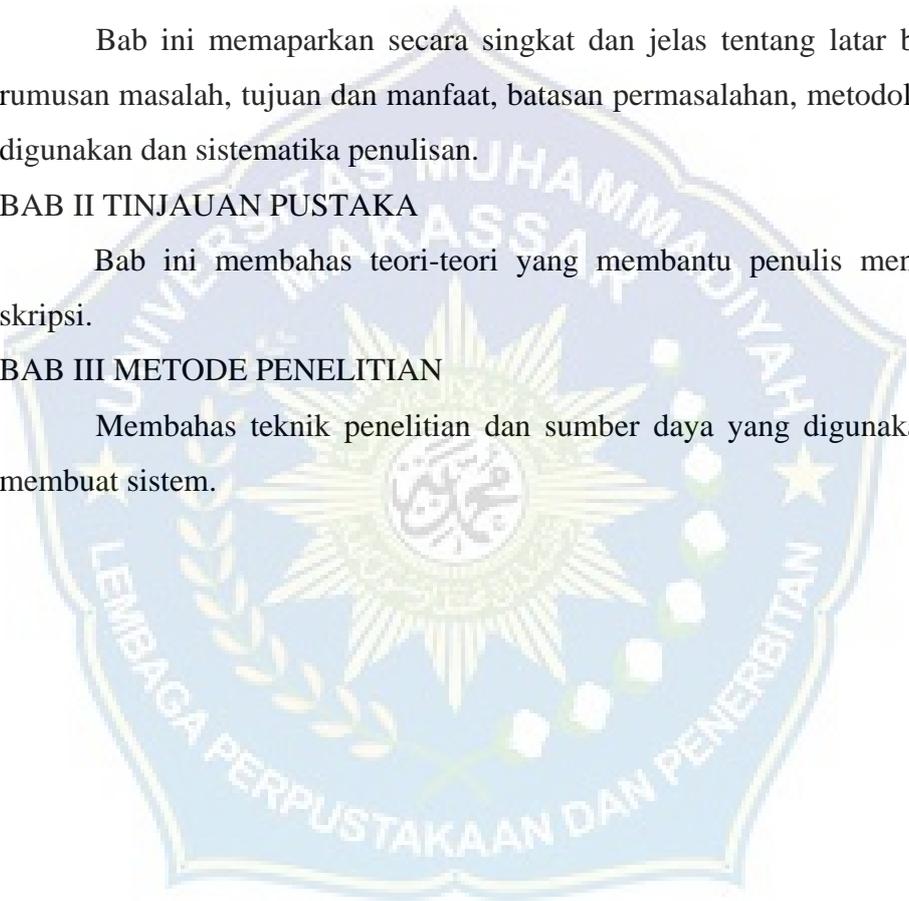
Bab ini memaparkan secara singkat dan jelas tentang latar belakang, rumusan masalah, tujuan dan manfaat, batasan permasalahan, metodologi yang digunakan dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini membahas teori-teori yang membantu penulis menjalankan skripsi.

BAB III METODE PENELITIAN

Membahas teknik penelitian dan sumber daya yang digunakan untuk membuat sistem.



BAB II

TINJAUAN PUSTAKA

A. Landasan Teori

1. Diabetes Melitus

Diabetes Mellitus adalah salah satu masalah kesehatan yang umum saat ini. Penyakit ini disebabkan oleh tingginya kadar gula darah.(Rahayu *et al.*, 2022). Diabetes melitus merupakan penyakit kronis yang memerlukan perawatan jangka panjang untuk mengurangi risiko komplikasi. Salah satu faktor yang mempengaruhi fluktuasi tekanan darah adalah kadar gula darah. Salah satu faktor risiko hipertensi adalah hiperglikemia. Sindrom metabolik, yang terdiri dari hipertensi, dislipidemia, obesitas, disfungsi endotel, dan faktor pro trombotik, sering kali disertai dengan hiperglikemia. Semua kondisi tersebut dapat memicu dan memperburuk komplikasi kardiovaskular(Julianti, 2021)

Diabetes Mellitus (DM), yang di Indonesia lebih dikenal sebagai kencing manis, adalah suatu gangguan yang disebabkan oleh peningkatan gula darah akibat penurunan sekresi insulin atau ketidakseimbangan dalam pemanfaatan insulin oleh tubuh. Hal ini menyebabkan kadar glukosa atau gula dalam darah menjadi tidak terkontrol. Diabetes tipe 2 adalah penyakit hiperglikemia atau gangguan metabolik yang ditandai dengan peningkatan kadar gula darah akibat penurunan sekresi insulin dalam penelitian (Apriyani & Kurniati, 2020)

Beberapa faktor risiko, seperti asam urat serum tinggi, merokok, depresi, penyakit kardiovaskular, dislipidemia, hipertensi, penuaan, etnis, riwayat keluarga diabetes, ketidakaktifan fisik, dan obesitas, diperkirakan menyebabkan diabetes.(Widiasari *et al.*, 2021)

Di seluruh dunia, diabetes adalah penyakit yang sangat umum. Pada tahun 2021, jumlah penderita diabetes di Indonesia mencapai hampir 20 juta orang, dan diperkirakan akan meningkat menjadi lebih dari 23 juta orang pada tahun 2030. Ketika tubuh tidak dapat membuat atau menggunakan insulin

dengan baik, kadar gula darah meningkat, itulah yang menyebabkan diabetes. Diabetes terbagi menjadi diabetes tipe 1, diabetes tipe 2, dan diabetes gestasional. Diabetes tipe 1 muncul ketika sistem kekebalan tubuh menyerang dan menghancurkan sel beta pankreas, yang bertanggung jawab untuk memproduksi insulin. Diabetes tipe 2 muncul ketika tubuh tidak dapat menggunakan insulin dengan baik, yang menyebabkan pankreas secara bertahap mengurangi produksi insulinnya. Sementara itu, diabetes gestasional terjadi pada wanita hamil yang sebelumnya tidak memiliki riwayat diabetes. Dalam penelitian (Abdianto Nggego et al., 2023)

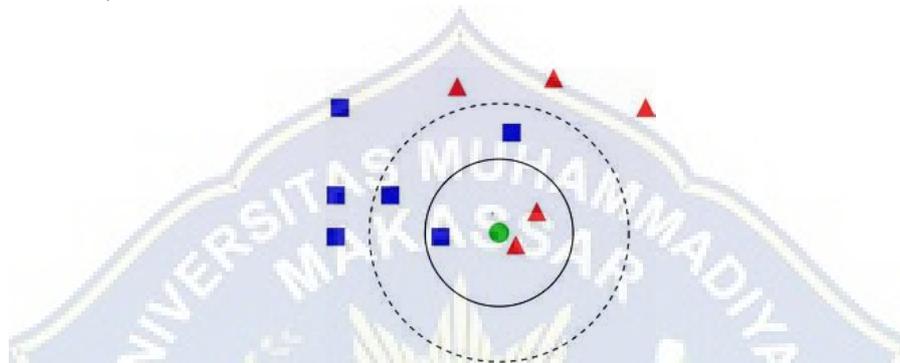
Diabetes melitus adalah sekelompok penyakit metabolik yang dicirikan oleh hiperglikemia yang disebabkan oleh gangguan sekresi insulin, kerja insulin, atau keduanya. Kondisi hiperglikemia kronis pada diabetes menyebabkan kerusakan, disfungsi, dan kegagalan berbagai organ, terutama mata, ginjal, saraf, jantung, dan pembuluh darah Dalam penelitian (Junus et al., 2023)

2. Risiko

Pada penelitian (Susilowati & Waskita, 2019), Risiko penyakit dapat muncul karena berbagai faktor, termasuk kejadian alam, operasional, perilaku manusia, kebijakan politik, perkembangan teknologi, kondisi pekerjaan, situasi keuangan, aspek hukum, dan pengelolaan dalam sistem kesehatan. Risiko berkaitan dengan ketidakpastian, yang muncul akibat kurangnya atau tidak adanya informasi yang cukup mengenai apa yang akan terjadi. Ketidakpastian ini bisa menghasilkan dampak yang menguntungkan atau merugikan. Ketidakpastian yang berpotensi menghasilkan keuntungan disebut sebagai peluang (*Opportunity*), sedangkan ketidakpastian yang berpotensi menimbulkan kerugian disebut sebagai risiko (*Risk*). Secara umum, risiko dapat diartikan sebagai kondisi yang dihadapi oleh seseorang atau perusahaan di mana ada kemungkinan terjadinya kerugian.

3. K – NN

K-Nearest Neighbor (K-NN) adalah salah satu algoritma data mining yang sering digunakan dalam penelitian, terutama untuk klasifikasi objek dalam data. Algoritma ini bekerja dengan mencari sekelompok k data terdekat dari kumpulan data. K-NN menghitung jarak terdekat antara data training dan data test untuk menentukan klasifikasi data. (Margolang et al., 2022).



Gambar 1. Klasifikasi Berdasarkan Tetangga Terdekat

Dalam penelitian (Suarisman et al., 2023), K-Nearest Neighbor (K-NN) adalah algoritma klasifikasi yang digunakan untuk mengklasifikasikan data berdasarkan klasifikasi yang sudah diketahui sebelumnya. Algoritma ini termasuk dalam kategori pembelajaran berbasis contoh (instance-based learning) dan merupakan teknik pembelajaran malas (lazy learning). Metode K-NN mencari K objek terdekat dalam data pelatihan yang memiliki kesamaan dengan objek dalam data uji. Dengan menghitung jarak antara data pelatihan (x) dan data uji (y), metode ini memperhitungkan kesamaan bobot dari berbagai atribut yang ada. Rumus jarak metrik Manhattan digunakan untuk menentukan jarak ini.

Algoritma K-Nearest Neighbor (K-NN) adalah metode klasifikasi yang mengklasifikasikan data baru berdasarkan data yang sudah diklasifikasikan sebelumnya. Sebagai bagian dari pembelajaran terawasi (supervised learning), algoritma ini mengklasifikasikan instance query baru berdasarkan mayoritas kedekatan dari kategori yang ada dalam K-NN.

Algoritma K-NN bekerja dengan mengukur jarak terpendek dari sampel uji ke sampel latih untuk menentukan K tetangga terdekatnya. Setelah mengumpulkan K tetangga terdekat tersebut, mayoritas dari tetangga ini digunakan untuk membuat prediksi terhadap sampel eksperimen. Kedekatan atau jauhnya tetangga biasanya dihitung dengan menghitung jarak dari Manhattan.(Admojo & Ahsanawati, 2020)

4. Manhattan

Dalam penelitian (Pratama Artana et al., 2024), Jarak Manhattan adalah metode untuk mengukur jarak antara dua titik dalam ruang geometri.

Menurut (Wahyu Pribadi et al., 2022), Manhattan Distance digunakan untuk menghitung total perbedaan absolut antara koordinat dari dua objek. Rumus dari manhattan adalah sebagai berikut :

$$d(x,y) = \sum_{i=1}^n |x_i - y_i|$$

Keterangan :

d = Jarak antara x dan y

x = Data pusat klaster

y = Data pada atribut

i = Setiap data

n = Jumlah data

x_i = Data pada pusat klaster ke i

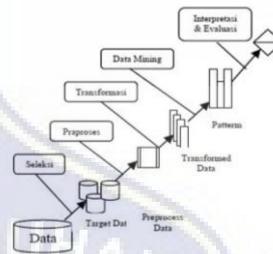
y_i = Data pada setiap data ke i

5. Data Mining

Tidak mengherankan bahwa pengolahan data semakin berkembang dengan cepat, meskipun kemajuan dalam teknologi informasi yang memungkinkan pengumpulan sejumlah besar data telah menyebabkan kondisi yang disebut sebagai "kaya data tetapi kekurangan informasi", di mana data yang terkumpul tidak dapat digunakan untuk aplikasi yang bermanfaat. Bahkan, kumpulan data sering dibiarkan begitu saja sehingga disebut sebagai "kuburan data" atau "tombe data". (Marlina & Bakri, 2021).

Tujuan dari data mining dan Knowledge Discovery in Databases (KDD) adalah sama: menggunakan data yang tersedia dalam basis data untuk

mengolahnya dan menghasilkan informasi baru yang berguna. Beberapa istilah lain yang serupa dengan data mining termasuk penemuan pengetahuan dalam basis data, ekstraksi pengetahuan, analisis pola atau data, dan Menurut beberapa orang, data mining adalah bagian penting dari proses penemuan pengetahuan dari data.



Gambar 2. Tahapan Knowledge Discovery in Database (KDD)

6. Machine Learning

Machine learning adalah cabang dari kecerdasan buatan (Artificial Intelligence) dan ilmu komputer yang berfokus pada pemanfaatan data dan algoritma untuk meniru cara manusia belajar, sehingga akurasi dapat meningkat secara bertahap. Algoritma machine learning yang lebih baik akan menghasilkan keputusan yang lebih optimal. (Faiza et al., 2022)

7. Supervised Learn

Algoritma supervised learning memanfaatkan data input berlabel untuk mempelajari fungsi yang dapat menghasilkan output yang tepat ketika dihadapkan pada data baru tanpa label. Algoritma K-NN (K-Nearest Neighbors) berasumsi bahwa objek yang mirip akan berada dalam jarak yang dekat satu sama lain. Dengan kata lain, objek-objek yang serupa cenderung berdekatan. (Kristiawan et al., 2020)

8. Classification

Dalam penelitian (Setio et al., 2020) Klasifikasi adalah teknik dalam data mining yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas yang telah ditetapkan. Sebagai metode pembelajaran terawasi (supervised learning), Klasifikasi membutuhkan data pelatihan yang dilabelkan untuk menghasilkan aturan untuk mengklasifikasikan data uji ke dalam kelas atau kategori yang telah ditentukan.

Dalam Penelitian (Abdurrahman, 2022) Klasifikasi adalah teknik analisis data yang bertujuan untuk mengembangkan model yang dapat mengelompokkan data ke dalam kategori tertentu. Salah satu algoritma yang dapat digunakan untuk klasifikasi adalah Naïve Bayes Classifier. Algoritma ini memanfaatkan teorema Bayes untuk memprediksi kemungkinan masa depan berdasarkan data pengalaman sebelumnya.

9. Scikit-Learn

Maksud untuk machine learning menggunakan Python. Pustaka ini menyediakan berbagai alat untuk machine learning dan pemodelan statistik, termasuk berbagai metode klasifikasi, regresi, dan algoritma pengelompokan seperti *Support Vector Machines* (SVM), *Random Forests*, *Gradient Boosting*, *k-Means*, dan *K-Nearest Neighbors* (KNN). *Scikit-learn* adalah modul yang didistribusikan di bawah *lisensi 3-Clause BSD* dan memungkinkan pemrograman machine learning berbasis python. (Nafisah Nurul Hakim, 2020).

Ketika menggunakan bahasa pemrograman Python untuk menerapkan metode pembelajaran yang diawasi, peneliti sering memanfaatkan perpustakaan dan kerangka kerja yang umum digunakan seperti *scikit-learn*, *TensorFlow*, dan Keras. Perangkat-perangkat ini mempermudah peneliti dalam menerapkan berbagai algoritma pembelajaran yang diawasi. Para peneliti menjelaskan langkah-langkah yang terlibat dalam proses pembelajaran, termasuk tahap pemrosesan data, pemilihan fitur, pelatihan model, dan evaluasi kinerja (Mestika et al., 2022).

10. Flowchart

Menurut (Rosaly & Prasetyo, 2020), dalam penelitiannya, Flowchart, juga dikenal sebagai diagram alir, adalah jenis diagram yang menggambarkan algoritma atau urutan langkah instruksi dalam suatu sistem. Analisis sistem menggunakan flowchart sebagai dokumentasi untuk menjelaskan secara logis bagaimana sistem yang akan dibangun berfungsi kepada programmer. Dengan demikian, flowchart membantu dalam memberikan solusi terhadap potensi masalah yang mungkin terjadi selama

pengembangan sistem. Secara dasar, flowchart menggunakan simbol-simbol yang mewakili berbagai proses. Untuk menghubungkan satu proses ke proses berikutnya, digunakan garis penghubung.

B. Penelitian Terkait

Tabel 1. Penelitian Terkait

Peneliti	Judul/Kasus	Metode/Algoritma	Hasil
(Rahayu et al., 2022)	Komparasi Jarak Euclidean dan Manhattan Pada Algoritma K-Nearest Neighbor Dalam Mendeteksi Penyakit Diabetes Mellitus	Algoritma K-NN dan Manhattan	Hasil uji coba menunjukkan bahwa nilai K terbaik pada confusion matrix dengan jarak Euclidean adalah $K=7$
(Setiawan, 2022)	Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris	Algoritma K-NN dan Manhattan	Jika menggunakan $k = 5$ dalam metode KNN, dengan variasi random sample dari 1 hingga 1000 dan data uji sebesar 20%, diperoleh akurasi rata-rata sebesar 96,6667%. Selanjutnya, dilakukan studi simulasi untuk mengamati hasilnya ketika variabel

			respons atau klasifikasi diubah.
(Dwi Fasnuari et al., 2022)	Penerapan algoritma k-nearest neighbor (k-nn) untuk klasifikasi penyakit diabetes melitus studi kasus : warga desa jatitengah	Algoritma K-NN	Dari hasil percobaan untuk meningkatkan akurasi Algoritma KNN, dapat disimpulkan bahwa model KNN terbaik diperoleh dengan membagi data testing sebanyak 20% dan menggunakan nilai K=9.

C. Kerangka Pikir



Gambar 3. Diagram Kerangka Pikir

BAB III

METODE PENELITIAN

A. Tempat dan Waktu Penelitian

1. Tempat Penelitian

Tempat penelitian merupakan lokasi dan objek yang akan digunakan pada suatu penelitian. Lokasi pengambilan data pasien pada penelitian ini dilakukan di Puskesmas Lappae Kab. Sinjai.

2. Waktu Penelitian

Jadwal penelitian yang akan dilaksanakan dimulai pada bulan Juli sampai semua proses pengumpulan data selesai. Berikut ini adalah tabel jadwal penelitian:

B. Alat dan Bahan

1. Kebutuhan Hardware (perangkat keras)

- a. Laptop Asus Vivobook
- b. RAM 16 GB

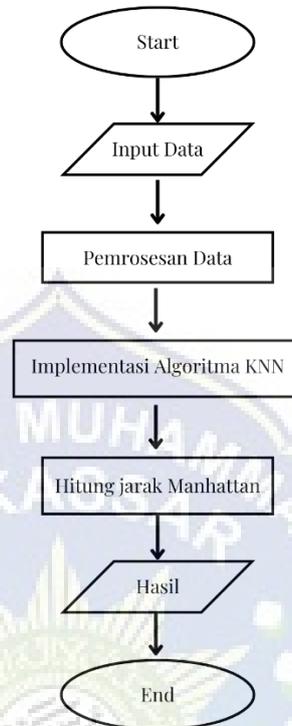
2. Kebutuhan Software (Perangkat Lunak)

- a. SKLearn
- b. Microsoft Excel
- c. Python

3. Bahan

Data mengenai usia, tekanan darah, kadar gula darah, dan BMI

C. Perancangan Sistem



Gambar 4. Flowchat Algoritma K - NN

Penjelasan Gambar 2. Flowchat Algoritma K – NN

1. Mulai : Flowchart dimulai dengan awal, yang menunjukkan dimulainya proses.
2. Input Data : Proses pengumpulan data pasien penyakit Diabetes melitus
3. Pemrosesan Data : Langkah-langkah untuk membersihkan dan mempersiapkan data termasuk menangani nilai yang hilang, normalisasi, dan menghapus atribut yang tidak relevan.
4. Implementasi Algoritma K – NN : Melaksanakan algoritma K - NN pada data testing set untuk menentukan faktor penyebab diabetes mellitus berdasarkan ketetanggaan terdekat.
5. Hitung jarak *Manhattan* : Menghitung jarak antara data uji dan semua data pelatihan menggunakan metrik Manhattan.
6. Hasil : Tentukan label kelas mayoritas dari K tetangga terdekat sebagai prediksi untuk titik data uji.

7. Selesai : Tanda berakhir proses flowchart

D. Teknik Pengujian

Metode untuk menguji sistem dalam implementasi algoritma K-NN dengan metrik Manhattan untuk mengetahui faktor pemicu penyakit diabetes melitus melibatkan beberapa langkah utama. Pertama, data uji yang telah dipisahkan diambil dan digunakan untuk memprediksi hasil dengan model KNN yang telah dilatih. Kemudian, evaluasi Prediksi dibuat dengan menghitung akurasi, presisi, recall, dan skor F1 untuk menilai kinerja model. Selanjutnya, analisis kesalahan dilakukan untuk memahami di mana model bekerja dengan baik atau kurang baik, dengan fokus pada pola atau anomali dalam prediksi yang salah. Hasil prediksi divalidasi dengan data tambahan atau metode lain untuk memastikan bahwa faktor-faktor yang diidentifikasi benar-benar terkait dengan diabetes melitus. Berdasarkan analisis kesalahan dan validasi temuan, model disesuaikan jika diperlukan untuk meningkatkan akurasi dan relevansi. Tahapan ini memastikan bahwa model yang digunakan adalah akurat, dapat diandalkan, dan relevan dengan tujuan penelitian.

E. Teknik Analisis Data

Dalam konteks studi untuk mengetahui faktor pemicu diabetes melitus menggunakan algoritma KNN dengan jarak Manhattan, proses analisis data melibatkan beberapa langkah. Langkah-langkah tersebut mencakup pengumpulan data pasien, pra-pemrosesan data, pembagian data menjadi set data latih dan uji, pelatihan model KNN, validasi menggunakan data uji, evaluasi performa model, hingga analisis hasil evaluasi untuk menentukan keandalan model dalam mengidentifikasi faktor-faktor pemicu diabetes melitus pada tahap awal penelitian.

(KNN) bersih, konsisten, dan siap untuk digunakan. Proses ini dimulai dengan mengumpulkan informasi.

Tabel 3. Dataset Pasien Penyakit Diabetes Mellitus

NO	Usia	Status - Tekanan Darah	Hasil - Tekanan Darah (Sistol)	Hasil - Tekanan Darah (Diastol)	Status - Gula Darah Sewaktu	Hasil - Gula Darah Sewaktu	Hasil - IMT	Hasil - Lingkar Perut
1	52	HIPERTE NSI	150	100	HIPERGLIKE MIK	285	33.155655 533	101
2	60	PRE- HIPERTE NSI	120	80	HIPERGLIKE MIK	422	25.292600 674	86
3	60	PRE- HIPERTE NSI	120	70	HIPERGLIKE MIK	278	28.841531 806	106
...
1040	47	HIPERTE NSI	140	90	HIPERGLIKE MIK	249	27.025809 648	95
1041	38	HIPERTE NSI 30	140	80	HIPERGLIKE MIK	232	26.896180 742	99

C. Implementasi Algoritma K – Means

Klustering K – Means digunakan untuk memproses dan menganalisis data pasien dari Puskesmas Lappae dengan tujuan mengelompokkan pasien berdasarkan faktor risiko diabetes. Langkah-langkahnya mencakup memuat data dari file Excel, membersihkan dan menstandarkan data numerik, serta menerapkan algoritma K-Means untuk mengelompokkan pasien menjadi dua kluster. Hasil klustering kemudian divisualisasikan dalam plot scatter untuk analisis visual dan disimpan kembali ke dalam file Excel untuk referensi lebih lanjut. Proses ini membantu dalam mengidentifikasi pola dan kelompok pasien yang memiliki karakteristik serupa terkait risiko diabetes.

```

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 mengonversi string yang mengandung tanda koma menjadi angka desimal (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 Alogritma K-Means

Algoritma K-Means digunakan dengan menetapkan dua kluster. Hasil klustering kemudian divisualisasikan dengan plot scatter.

```

# 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)

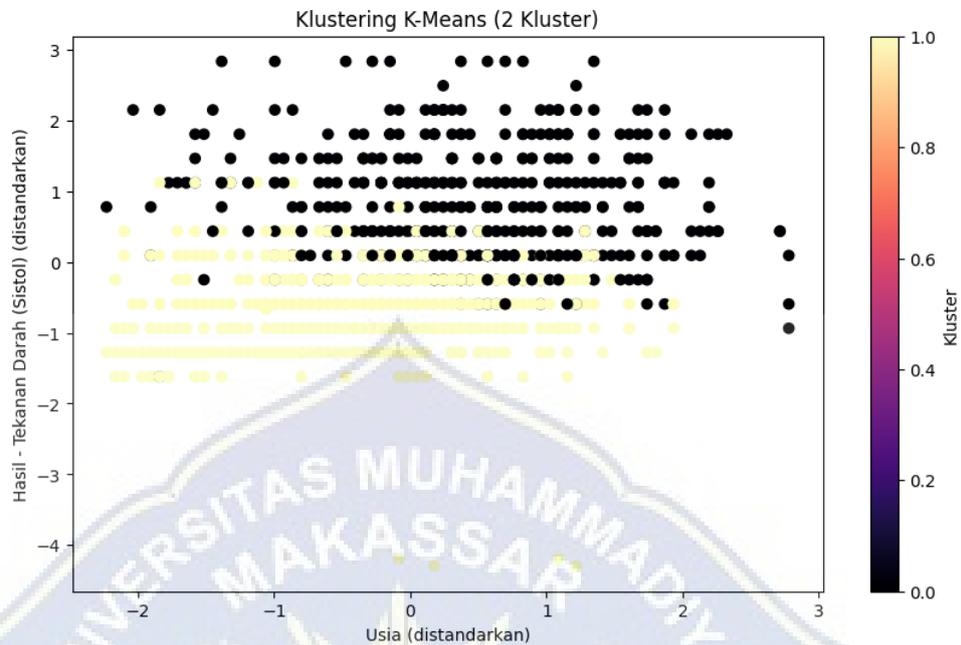
# Menampilkan beberapa baris pertama dari
dataframe dengan label kluster
print(df.head())

```

2. Hasil Clustering

Hasil visualisasi klustering menggunakan algoritma K-Means dengan 2 kluster pada data yang telah distandarkan dari Puskesmas Lappae. Sumbu horizontal merepresentasikan usia pasien yang telah distandarkan, sedangkan sumbu vertikal merepresentasikan tekanan darah sistol yang juga telah distandarkan. Setiap titik dalam plot ini mewakili satu pasien.

Gambar 5. Hasil Clustering



Gambar tersebut menunjukkan dua kluster hasil K-Means, yaitu kluster kuning dan hitam, yang mengelompokkan peserta berdasarkan usia dan tekanan darah sistol yang sudah distandarkan. Kluster kuning cenderung merepresentasikan peserta dengan nilai usia dan tekanan darah sistol lebih rendah, sementara kluster hitam menunjukkan peserta dengan nilai lebih tinggi. Visualisasi ini memudahkan identifikasi kelompok berdasarkan pola kesehatan tertentu yang relevan dengan risiko penyakit. Standarisasi ini dilakukan agar data memiliki skala yang seragam, biasanya dengan nilai rata-rata 0 dan standar deviasi 1, sehingga muncul nilai-nilai seperti -4, -3, ataupun 3. Dalam konteks ini, nilai negatif pada sumbu x dan y menunjukkan usia atau tekanan darah yang lebih rendah dari rata-rata, sedangkan nilai positif menunjukkan yang lebih tinggi dari rata-rata. Warna titik-titik pada grafik menggambarkan kluster yang berbeda, dengan legenda warna di sebelah kanan mengindikasikan perbedaan kluster dari 0 (kuning) hingga 1 (hitam).

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
3	2
4	2

Hasil - Gula Darah Sewaktu	\
0	282.0
1	124.0
2	531.0
3	264.0
4	405.0

Status - IMT, NORMAL/BB KURANG/BB LEBIH/ OBESITAS	Hasil - IMT	\
0	0	24.266667
1	2	25.099502
2	0	24.654832
3	0	23.191095
4	0	18.718042

Status - Lingkar Perut, NORMAL/OBESITAS SENTRAL	Hasil - Lingkar Perut	\
0	1	89.0
1	1	87.0
2	1	99.0
3	1	87.0
4	0	79.0

KOLESTROL NORMAL/TINGGI	Status - Sadanis	Hasil - Sadanis kanan	\
0	1	NaN	NaN
1	1	NaN	NaN
2	0	NaN	NaN
3	1	NaN	NaN
4	1	NaN	NaN

Hasil - Sadanis kiri	Unnamed: 16	Tajam Penglihatan, NORMAL/GANGGUAN	\
0	NaN	NaN	1
1	NaN	NaN	0
2	NaN	NaN	1
3	NaN	NaN	1
4	NaN	NaN	0

Tajam Pendengaran, NORMAL/GANGGUAN	Kluster
0	0
1	0
2	0
3	0
4	1

Gambar 6. Hasil Cluster

Gambar tersebut menampilkan hasil clustering dari data deteksi dini kesehatan beberapa peserta berdasarkan berbagai indikator seperti jenis kelamin, usia, tekanan darah, kadar gula darah, IMT, lingkar perut, kolesterol, serta pemeriksaan sadanis, penglihatan, dan pendengaran. Data ini dikelompokkan ke dalam tiga kluster yang

menggambarkan kategori kesehatan peserta. Setiap indikator memiliki status tertentu, seperti tekanan darah (normal, pre-hipertensi, hipertensi) dan gula darah sewaktu (normal, pre-hiperglikemik, hiperglikemik), yang digunakan dalam proses clustering.

3. Re-Preprocessing

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

- Status - Sadanis: Status dari pemeriksaan payudara klinis.
- Hasil - Sadanis kanan: Hasil dari pemeriksaan payudara kanan.
- Hasil - Sadanis kiri: Hasil dari pemeriksaan payudara kiri.
- Unnamed: 16: Kolom tambahan tanpa nama, biasanya merupakan artefak dari proses ekspor data

Kolom-kolom tersebut dihapus karena tidak relevan dengan prediksi diabetes yang sedang 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 berikutnya adalah menghapus kolom-kolom tersebut dari DataFrame `df`. Fungsi `drop` dari pandas digunakan untuk menghapus

kolom-kolom tersebut. Menghapus kolom-kolom yang tidak relevan merupakan bagian penting dari pra-pemrosesan data. Dalam penelitian ini, kolom seperti Status - Sadanis, Hasil - Sadanis kanan, Hasil - Sadanis kiri, dan Unnamed: 16 dihapus karena tidak berkaitan dengan analisis diabetes. Langkah ini membantu memfokuskan pada data yang relevan, meningkatkan efisiensi analisis, dan mengurangi gangguan dalam hasil analisis.

Setelah melakukan pra-pemrosesan ulang, data hasil *Clustering* dapat ditampilkan menggunakan perintah ``df``. Variabel ``df`` adalah *DataFrame* yang berisi data hasil *Clustering* dan dapat digunakan untuk analisis lebih lanjut atau visualisasi data. Berikut adalah output dari hasil *Clustering* dengan Algoritma *K-Means*:

Tabel 4. Hasil Clustering

Nama Peserta Deteksi Dini	JENIS KELAMIN, MALE/FE MALE	Usia	Hasil Teka Tekanan Darah (Sistol)	Hasil Tekanan Darah (Diastol)	Hasil Gula Darah Sewaktu	Hasil - IMT	Hasil Lingkar Perut	Status_Diabetes	Kluster
SULWANDI	0	45	150	80	282	24.26 667	89	Diabetes	0
Arsyad	0	51	130	70	124	25.09 95	87	Tidak Diketahui	1
ILHAM MUSTAFA	0	48	170	100	531	24.65 483	99	Diabetes	0
...
SALMA	1	38	140	80	232	26.89 618	99	Diabetes	0

Pasien dengan risiko rendah dikelompokkan dalam Kluster 0, sementara pasien dengan risiko tinggi berada di Kluster 1. Pengelompokan ini memberikan gambaran tentang distribusi faktor risiko di antara kelompok yang berbeda, membantu dalam mengidentifikasi pola dan kecenderungan kesehatan di antara peserta.

D. Implementasi Metode K – NN metric Manhattan

Tahapan ini akan dilakukan proses data mining dengan menggunakan algoritma K-Nearest Neighbor (KNN) dengan dataset yang telah disiapkan berdasarkan atribut-atribut tertentu seperti atribut seperti Usia, Tekanan Darah (Sistol), Tekanan Darah (Diastol), Gula Darah Sewaktu, IMT, Lingkar Perut.

1. *Oversampling* dengan SMOTE : SMOTE diterapkan untuk membuat sampel sintetis dari kelas minoritas di data pelatihan (*x_train* dan *y_train*) untuk mengatasi ketidakseimbangan kelas. *fit_resample* melakukan *oversampling* dan mengembalikan data yang telah di-*resample*.

```
# 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 : *StandardScaler* digunakan untuk normalisasi fitur. *fit_transform* diterapkan pada data pelatihan untuk menyesuaikan skala dan mentransformasi data pelatihan, sementara *transform* diterapkan pada data pengujian untuk menggunakan skala yang sama.

```
# Normalisasi data
scaler = StandardScaler()
x_train = scaler.fit_transform(x_train)
x_test = scaler.transform(x_test)
```

3. Membuat Model K-Nearest Neighbors dengan Metric Manhattan : Ini membuat model K-Nearest Neighbors dengan *n_neighbors=5*, artinya model akan mempertimbangkan 5 tetangga terdekat untuk menentukan kelas. *metric='manhattan'* berarti jarak Manhattan (atau jarak L1) digunakan untuk menghitung jarak antara titik-titik data.

```
# Create a K-Nearest Neighbors model with
Manhattan distance metric
model =
KNeighborsClassifier(n_neighbors=5,metric='manh
attan')
```

1. Menentukan parameter K = 5
2. Perhitungan *Manhattan Distance* pada masing-masing objek terhadap data sampel yang diberikan

3. Rumus Jarak Manhattan = $\sum_{i=1}^n |x_i - y_i|$

4. Data 1:

45 2 150 80 2 282 0 24.26666667 1 89

Data 2:

51 1 130 70 0 124 2 25.0995016 1 87

Fitur untuk kedua data adalah:

- Umur
- Status - Tekanan Darah
- Hasil - Tekanan Darah (Sistol)
- Hasil - Tekanan Darah (Sistol)
- Status - Gula Darah Sewaktu
- Hasil - Gula Darah Sewaktu
- Status IMT
- Hasil - IMT
- Status - Lingkar Perut
- Hasil - Lingkar Perut

Untuk menghitung jarak Manhattan, kita hitung selisih absolut antara nilai-nilai yang bersesuaian:

- Umur: $|45 - 51| = 6$
- Status - Tekanan Darah : $|2 - 1| = 1$
- Hasil - Tekanan Darah (Sistol) : $|150 - 130| = 20$
- Hasil - Tekanan Darah (Sistol) : $|80 - 70| = 10$

- Status - Gula Darah Sewaktu : $|2 - 0| = 2$
- Hasil - Gula Darah Sewaktu : $|282 - 124| = 158$
- Status - IMT : $|0 - 2| = 2$
- Hasil - IMT : $|24.26666667 - 25.0995016| = 0.83283493$
- Status - Lingkar Perut : $|1 - 1| = 0$
- Hasil - Lingkar Perus : $|89 - 87| = 2$

Kemudian, jumlahkan semua selisih absolut tersebut:

$$\text{Jarak Manhattan} = 6 + 1 + 20 + 10 + 2 + 158 + 2 + 0.83283493 + 0 + 2 = 201.83283493$$

$$\text{Jarak Manhattan} = 6 + 1 + 20 + 10 + 2 + 158 + 2 + 0.83283493 + 0 + 2 = 201.83$$

Jadi, jarak Manhattan antara Data 1 dan Data 2 adalah sekitar 201.83

4. Latih model : *fit* digunakan untuk melatih model *K-Nearest Neighbors* pada data pelatihan (*x_train* dan *y_train*), sehingga model dapat mempelajari pola dalam data tersebut.

```
model.fit(x_train, y_train)
```

5. Evaluasi Model : Melakukan prediksi pada data pengujian, mengevaluasi akurasi model, dan mencetak laporan klasifikasi yang menunjukkan kinerja model dalam mengklasifikasikan data.

```
from sklearn.metrics import accuracy_score, classification_report
```

```
# Evaluate the model on the test data
```

```
y_pred = model.predict(x_test)
```

```
# Calculate and print the accuracy
```

```
accuracy = model.score(x_test, y_test)
```

```
print(f"Accuracy: {accuracy:.2f}")
```

```
# 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

```

Accuracy: 0.90
Accuracy: 0.9038461538461539
Classification Report:

```

	precision	recall	f1-score	support
0	0.84	0.97	0.90	92
1	0.97	0.85	0.91	116
accuracy			0.90	208
macro avg	0.91	0.91	0.90	208
weighted avg	0.91	0.90	0.90	208

Gambar 7. Hasil Prediksi Algoritma K-NN dan Metrik Manhattan

Hasil prediksi pada gambar menunjukkan kinerja yang sangat baik dengan akurasi 90%, yang berarti model benar dalam 90% dari prediksi yang dibuat. Laporan klasifikasi menunjukkan precision yang tinggi untuk kedua kelas, yakni 0.84 untuk kelas 0 dan 0.97 untuk kelas 1, serta recall yang solid, yaitu 0.97 untuk kelas 0 dan 0.85 untuk kelas 1. F1-Score, yang mengukur keseimbangan antara precision dan recall, juga memuaskan dengan nilai 0.90 untuk kelas 0 dan 0.91 untuk kelas 1. Baik macro average maupun weighted average menunjukkan nilai 0.90 untuk precision, recall, dan F1-Score, mengindikasikan bahwa model ini konsisten dalam performanya di seluruh kelas meski terdapat ketidakseimbangan jumlah contoh di masing-masing kelas.

```

Accuracy: 0.89
Accuracy: 0.8894230769230769
Classification Report:

```

	precision	recall	f1-score	support
0	0.83	0.95	0.88	92
1	0.95	0.84	0.89	116
accuracy			0.89	208
macro avg	0.89	0.90	0.89	208
weighted avg	0.90	0.89	0.89	208

Gambar 8. Hasil Prdiksi Algoritma K-NN dan Metrik Euclidean

Model ini, menggunakan metrik Euclidean, mencapai akurasi 0.89, menunjukkan bahwa 89% prediksi model adalah benar. Precision untuk kelas 0 adalah 0.83 dengan recall 0.95, sedangkan precision untuk kelas 1 adalah 0.95 dengan recall 0.84. F1-score untuk kedua kelas adalah 0.88 dan 0.89, masing-masing, mencerminkan keseimbangan yang baik antara precision dan recall. Rata-rata makro dan tertimbang untuk precision, recall, dan f1-score konsisten dengan akurasi keseluruhan, menandakan bahwa model ini memberikan performa yang solid dan seimbang dalam mengklasifikasikan kedua kelas, meskipun ada perubahan dari metrik Euclidean.

Selanjutnya penginputan hasil, hasil di input kedalam file excel

```

# Dictionary untuk pemetaan label numerik ke label
teks

label_map = {1: 'Diabetes Kronis', 0: 'Diabetes
Ringan'}

# Gabungkan da ta 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 KNN.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 KNN.xlsx

Tabel 5. Hasil Prediksi K-NN dan Metrik Manhattan

NO	NAMA PASIEN	JENIS KELAMIN	USIA	LABEL ASLI	PREDIKSI
0	JUNAEDI	1	71	DIABETES RINGAN	DIABETES RINGAN
1	TATI	1	54	DIABETES KRONIS	DIABETES KRONIS
2	JAMALUDDIN B	0	59	DIABETES KRONIS	DIABETES KRONIS
...
206	ROSTINA	1	57	DIABETES KRONIS	DIABETES KRONIS
207	MARNI	1	46	DIABETES RINGAN	DIABETES RINGAN

Gambar tersebut menampilkan tabel hasil deteksi dini diabetes dari 208 peserta yang mencakup beberapa kolom, yaitu nama peserta, jenis kelamin (1 untuk perempuan, 0 untuk laki-laki), usia, label asli kondisi diabetes (seperti "Diabetes Ringan" atau "Diabetes Kronis"), dan hasil prediksi model. Tabel ini bertujuan untuk membandingkan label asli dengan prediksi model untuk mengidentifikasi seberapa akurat model dalam memprediksi tingkat keparahan diabetes. Sebagai contoh, peserta bernama Junaedah (perempuan, usia 71 tahun) memiliki label asli "Diabetes Ringan," dan model juga memprediksi hal yang sama. Tabel ini menunjukkan bagaimana model memprediksi kondisi diabetes peserta dan seberapa sering prediksinya sesuai dengan kondisi asli.

Tabel 6. Hasil Prediksi K-NN dan Metrik Eculidean

NO	NAMA PASIEN	JENIS KELAMIN	USIA	LABEL ASLI	PREDIKSI
0	JUNAEDI	1	71	DIABETES RINGAN	DIABETES RINGAN
1	TATI	1	54	DIABETES KRONIS	DIABETES KRONIS
2	JAMALUDDIN B	0	59	DIABETES KRONIS	DIABETES KRONIS
...
206	ROSTINA	1	57	DIABETES KRONIS	DIABETES KRONIS
207	MARNI	1	46	DIABETES RINGAN	DIABETES RINGAN

Tabel hasil prediksi diabetes tersebut memberikan gambaran perbandingan antara kondisi diabetes yang sebenarnya (label asli) dengan hasil prediksi yang diberikan oleh model machine learning. Tabel ini menampilkan data dari sejumlah pasien, termasuk identitas, jenis kelamin, usia, kondisi diabetes sebenarnya, dan prediksi model. Dengan membandingkan kolom label asli dan prediksi, kita dapat menilai akurasi model dalam mendeteksi diabetes. Informasi

ini sangat berguna untuk mengevaluasi kinerja model, mengambil keputusan, dan memperbaiki model di masa depan.

E. Hasil Pengujian Sistem

1. Pengujian sistem *Clustering*

Analisis clustering yang telah dilakukan menggunakan algoritma K-Means berhasil mengelompokkan peserta menjadi beberapa kelompok berdasarkan kesamaan karakteristik kesehatan mereka. Algoritma ini bekerja dengan cara membagi data menjadi K kelompok yang tidak tumpang tindih, di mana K ditentukan sebelumnya. Setiap individu kemudian ditugaskan ke kelompok terdekat berdasarkan centroid-nya. Hasil clustering menunjukkan bahwa peserta dalam satu kelompok cenderung memiliki profil kesehatan yang serupa, seperti rentang usia, tekanan darah, gula darah, dan indeks massa tubuh yang mirip. Dengan demikian, clustering ini dapat membantu mengidentifikasi kelompok-kelompok tertentu yang memiliki karakteristik kesehatan yang unik, sehingga dapat menjadi dasar untuk memberikan rekomendasi perawatan yang lebih personal.

Tabel 7. Hasil Pengujian Sistem Clustering

Nama Peserta Deteksi Dini	JENIS KELAMI N, MALE/FE MALE	Usia	Hasil Teka nan Dara h (Sistol)	Hasil Teka nan Dara h (Dias tol)	Hasil Gula Dara h Sewa ktu	Hasil IMT	Hasil Ling kar Peru t	Status_Di abetes	Klu ster
SULW ANDI	0	45	150	80	282	24.26 667	89	Diabetes	0
Arsyad	0	51	130	70	124	25.09 95	87	Tidak Diketahui	1
ILHAM MUST AFA	0	48	170	100	531	24.65 483	99	Diabetes	0
...

SALM	1	38	140	80	232	26.89	99	Diabetes	0
A						618			

Berikutnya dengan mencari jumlah data kluster beresiko tinggi dan beresiko rendah dari total 1 039 data pasien deteksi dini

```

      Kluster
1      1
5      1
6      1
8      1
9      1
...    ...
1024   1
1025   1
1026   1
1027   1
1037   1
[584 rows x 1 columns]
      Kluster
0      0
2      0
3      0
4      0
7      0
...    ...
1034   0
1035   0
1036   0
1038   0
1039   0
[456 rows x 1 columns]

```

Gambar 9. Hasil Cluster Resiko Tinggi dan Rendah

Berdasarkan hasil clustering, peserta deteksi dini diabetes dapat dikelompokkan menjadi dua kategori utama: risiko tinggi (584 pasien) dan risiko rendah (456 pasien). Temuan ini konsisten dengan penjelasan Kepala Puskesmas Lappae mengenai faktor risiko diabetes, seperti kadar gula darah tinggi, tekanan darah tinggi, obesitas, dan riwayat keluarga. Pengelompokkan ini memungkinkan kita untuk mengidentifikasi karakteristik unik dari masing-masing kelompok dan merancang intervensi yang lebih tepat sasaran untuk mencegah dan mengelola diabetes.

2. Pengujian Sistem Klasifikasi dan Prediksi K-NN

Pengujian sistem klasifikasi dan prediksi K-NN dilakukan dengan menggunakan file Excel "Hasil Prediksi KNN Manhattan.xlsx", yang berisi data hasil prediksi model. Proses ini dimulai dengan membaca data dari sheet "Sheet1" dan memfilter baris-baris di mana kolom 'Hasil Prediksi (Numerik)' menunjukkan nilai 1, yang berarti prediksi tersebut benar. Setelah memfilter, jumlah total data yang memuat prediksi benar dihitung untuk memberikan indikasi seberapa banyak prediksi yang akurat dilakukan oleh model. Hasil akhir yang ditampilkan adalah DataFrame yang menunjukkan hanya kolom 'Hasil Prediksi (Numerik)' dari data yang telah difilter, memungkinkan analisis lebih lanjut mengenai kelompok data yang diprediksi dengan benar.

Tabel 8. Pengujian Sistem Klasifikasi K-NN

NO	NAMA PASIEN	JENIS KELAMIN	USIA	LABEL ASLI	PREDIKSI	HASIL PREDIKSI (NUMERIK)
1	JUNAEDI	1	71	DIABETES RINGAN	DIABETES RINGAN	1
2	TATI	1	54	DIABETES KRONIS	DIABETES KRONIS	1
3	JAMALUDDIN B	0	59	DIABETES KRONIS	DIABETES KRONIS	1
...
206	ROSTINA	1	57	DIABETES KRONIS	DIABETES KRONIS	1
207	MARNI	1	46	DIABETES RINGAN	DIABETES RINGAN	1

Berdasarkan hasil evaluasi prediksi dan input yang diberikan, dapat disimpulkan bahwa model K-NN menunjukkan performa yang sangat baik dengan akurasi mencapai 90%. Dari total 208 data yang diuji, 188 di antaranya terprediksi dengan benar, sehingga persentasenya adalah $(188:208) \times 100\% = 90\%$. Nilai ini konsisten dengan hasil yang telah dibahas sebelumnya dalam bagian pembahasan mengenai implementasi algoritma K-NN.

BAB V

KESIMPULAN DAN SARAN

A. KESIMPULAN

Berdasarkan hasil penelitian tersebut, maka dapat disimpulkan sebagai berikut :

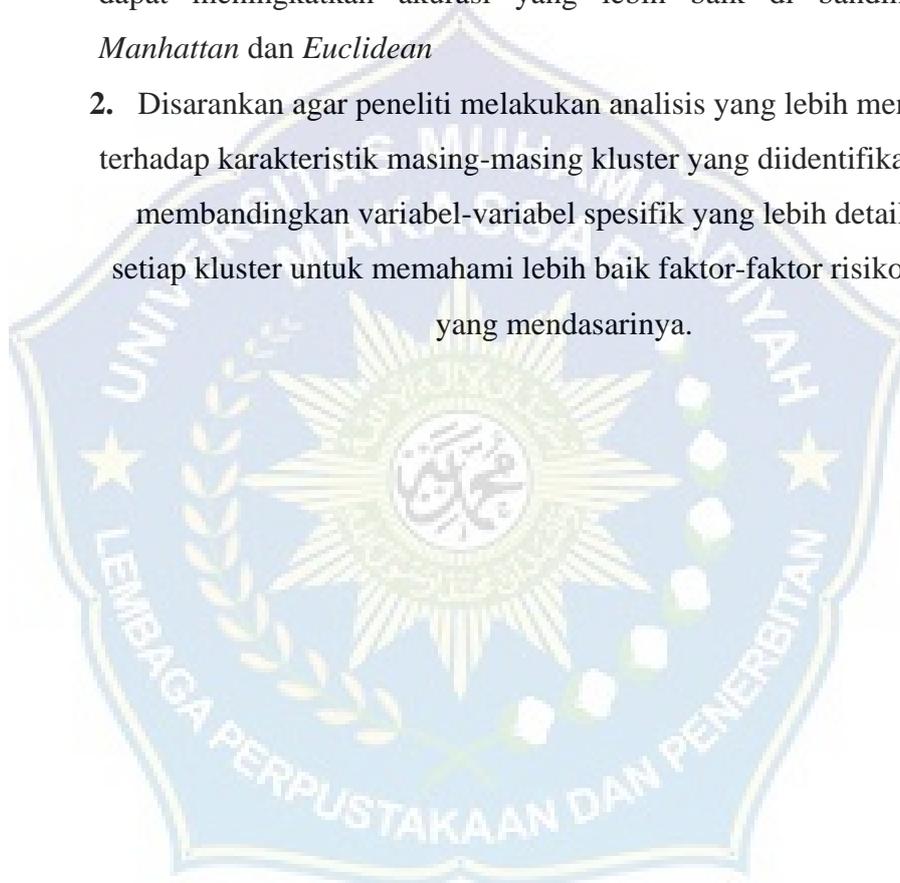
1. Penelitian ini berhasil mengidentifikasi dan mengelompokkan faktor-faktor yang mempengaruhi risiko diabetes mellitus melalui dua pendekatan utama. Hasil analisis menggunakan algoritma K-Means menunjukkan bahwa data pasien dapat dikelompokkan menjadi dua kategori utama, yaitu risiko tinggi dan risiko rendah. Kluster risiko tinggi, yang terdiri dari 584 pasien, dan kluster risiko rendah dengan 456 pasien, mengindikasikan adanya perbedaan signifikan dalam karakteristik kesehatan antara kedua kelompok. Peserta dalam kluster risiko tinggi cenderung memiliki kadar gula darah tinggi, tekanan darah tinggi, obesitas, dan riwayat keluarga diabetes, yang sejalan dengan faktor-faktor risiko yang diidentifikasi.
2. Model K-Nearest Neighbors (K-NN) menunjukkan performa yang sangat baik dengan akurasi mencapai 90%. Dari 208 data yang diuji, 188 data berhasil diprediksi dengan benar. Model ini mengkonfirmasi bahwa faktor-faktor seperti usia, BMI, kadar gula darah, lingkar perut, tekanan darah (sistol dan diastol), serta kolesterol adalah indikator penting dalam menentukan risiko diabetes. Dengan menggunakan model K-NN dan teknik clustering, penelitian ini mampu memberikan alat yang berguna bagi tenaga medis dan pembuat kebijakan untuk memprioritaskan individu yang berisiko tinggi dan merancang program pencegahan yang lebih terarah.

Secara keseluruhan, hasil penelitian ini memberikan wawasan berharga mengenai faktor-faktor penyebab diabetes mellitus dan menunjukkan efektivitas teknik analisis data dalam mengidentifikasi individu dengan risiko tinggi. Model K-NN dan hasil clustering dapat digunakan untuk meningkatkan strategi pencegahan dan pengelolaan diabetes, memastikan bahwa intervensi

dapat diarahkan secara efektif kepada mereka yang paling membutuhkan perhatian medis dan pencegahan lebih awal.

B. SARAN

1. Sebaiknya peneliti mengeksplorasi metrik jarak lain seperti *Minkowski* atau *Mahalanobis* untuk mengidentifikasi apakah ada metrik lain yang dapat meningkatkan akurasi yang lebih baik di banding metrik *Manhattan* dan *Euclidean*
2. Disarankan agar peneliti melakukan analisis yang lebih mendalam terhadap karakteristik masing-masing kluster yang diidentifikasi, seperti membandingkan variabel-variabel spesifik yang lebih detail dalam setiap kluster untuk memahami lebih baik faktor-faktor risiko diabetes yang mendasarinya.



DAFTAR PUSTAKA

- Abdianto Nggego, D., Taufik Bau, R. R., & Patawaran, N. (2023). Fa-Knn: Hybrid Algoritma Untuk Klasifikasi Penyakit Diabetes Melitus. *MJRICT : Musamus Journal Of Research Information and Communication Technology*, 5(2), 71–80. <https://ejournal.unmus.ac.id/index.php/mjriict>
- Abdurrahman, G. (2022). Klasifikasi Penyakit Diabetes Melitus Menggunakan Adaboost Classifier. *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, 7(1), 59–66. <https://doi.org/10.32528/justindo.v7i1.4949>
- Admojo, F. T., & Ahsanawati. (2020). Klasifikasi Aroma Alkohol Menggunakan Metode KNN. *Indonesian Journal of Data and Science*, 1(2), 34–38. <https://doi.org/10.33096/ijodas.v1i2.12>
- Agustiningrum, R., & Kusbaryanto, K. (2019). Efektifitas Diabetes Self Management Education Terhadap Self Care Penderita Diabetes Mellitus: A Literature Review. *Jurnal Keperawatan Respati Yogyakarta*, 6(2), 558. <https://doi.org/10.35842/jkry.v6i2.309>
- Apriyani, H., & Kurniati, K. (2020). Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. *Journal of Information Technology Ampera*, 1(3), 133–143. <https://doi.org/10.51519/journalita.volume1.issue3.year2020.page133-143>
- Dwi Fasnuari, H. A., Yuana, H., & Chulkamdi, M. T. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus. *Antivirus : Jurnal Ilmiah Teknik Informatika*, 16(2), 133–142. <https://doi.org/10.35457/antivirus.v16i2.2445>
- Faiza, I. M., Gunawan, G., & Andriani, W. (2022). Tinjauan Pustaka Sistematis: Penerapan Metode Machine Learning untuk Deteksi Bencana Banjir. *Jurnal Minfo Polgan*, 11(2), 59–63. <https://doi.org/10.33395/jmp.v11i2.11657>
- Farrel Nur Rilwanu, M., Taufikurachman, H., Faris Huwaidi, dan, Perangkat Lunak, R., & Daerah Cibiru, K. (2022). Penerapan Algoritma K-Nearest

- Neighbor untuk Mendeteksi Diabetes Berbasis Web Application. *Journal of Software Engineering, Information and Communication Technology*, 3(1), 145–152. <https://doi.org/10.13140/RG.2.2.14361.34400>
- Iffah'da, A. N., & Anita Desiani. (2022). Implementasi Algoritma K-Nearest Neighbor (K-NN) dan Single Layer Perceptron (SLP) Dalam Prediksi Penyakit Sirosis Biliari Primer. *Jurnal Ilmiah Informatika*, 7(1), 65–74. <https://doi.org/10.35316/jimi.v7i1.65-74>
- Julianti, I. M. D. (2021). Hubungan antara kadar gula darah dengan tekanan darah pada pasien diabetes melitus tipe II. *Jurnal Penelitian Kedokteran*, 3(2), 1–7.
- Junus, C. Z. V., Tarno, T., & Kartikasari, P. (2023). Klasifikasi Menggunakan Metode Support Vector Machine Dan Random Forest Untuk Deteksi Awal Risiko Diabetes Melitus. *Jurnal Gaussian*, 11(3), 386–396. <https://doi.org/10.14710/j.gauss.11.3.386-396>
- Kristiawan, K., Somali, D. D., Linggan jaya, T. A., & Widjaja, A. (2020). Deteksi Buah Menggunakan Supervised Learning dan Ekstraksi Fitur untuk Pemeriksa Harga. *Jurnal Teknik Informatika Dan Sistem Informasi*, 6(3), 541–548. <https://doi.org/10.28932/jutisi.v6i3.3029>
- Margolang, K. F., Siregar, M. M., Riyadi, S., & Situmorang, Z. (2022). Analisa Distance Metric Algoritma K-Nearest Neighbor Pada Klasifikasi Kredit Macet. *Journal of Information System Research (JOSH)*, 3(2), 118–124. <https://doi.org/10.47065/josh.v3i2.1262>
- Marlina, D., & Bakri, M. (2021). Penerapan Data Mining Untuk Memprediksi Transaksi Nasabah Dengan Algoritma C4.5. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(1), 23–28.
- Mestika, J. C., Selan, M. O., & Qadafi, M. I. (2022). Menjelajahi Teknik-Teknik Supervised Learning untuk Pemodelan Prediktif Menggunakan Python. *BIIKMA : Buletin Ilmiah Ilmu Komputer Dan Multimedia*, 99(99), 216–219.
- Nafisah Nurul Hakim. (2020). Implementasi Machine Learning pada Sistem

- Prediksi Kejadian dan Lokasi Patah Rel Kereta Api di Indonesia. *Jurnal Sistem Cerdas*, 3(1), 25–35. <https://doi.org/10.37396/jsc.v3i1.58>
- Pratama Artana, P. N., Prakarsa Mandyartha, E., & Hanindia Prami S, M. (2024). Penerapan Data Mining Pada Algoritma Hierarchical Clustering Tentang Pengelolaan Mitra Perjalanan Wisatawan Bali Backpacker. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2903–2909. <https://doi.org/10.36040/jati.v7i4.7284>
- Rahayu, A. E., Fauzan, A. C., & Harliana, H. (2022). Komparasi Jarak Euclidean dan Manhattan Pada Algoritma K-Nearest Neighbor Dalam Mendeteksi Penyakit Diabetes Mellitus. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(2), 413. <https://doi.org/10.30865/json.v4i2.5046>
- Rosaly, R., & Prasetyo, A. (2020). Flowchart Beserta Fungsi dan Simbol-Simbol. *Journal of Chemical Information and Modeling*, 2(3), 5–7.
- Sela, A. (2023). Resource allocations in the best-of-k ($k=2, 3$) contests. *Journal of Economics/ Zeitschrift Fur Nationalokonomie*, 139(3), 235–260. <https://doi.org/10.1007/s00712-023-00827-w>
- Setiawan, A. (2022). Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris. *Jurnal Sains Dan Edukasi Sains*, 5(1), 28–37. <https://doi.org/10.24246/juses.v5i1p28-37>
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Suarisman, A., Nazir, A., Syafria, F., & Afriyanti, L. (2023). Perbandingan Jarak Metrik pada Klasifikasi Jamur Beracun Menggunakan Algoritma K-Nearest Neighbor (K-NN). *Journal of Computer System and Informatics (JoSYC)*, 5(1), 10–19. <https://doi.org/10.47065/josyc.v5i1.4511>
- Susilowati, A. A., & Waskita, K. N. (2019). Pengaruh Pola Makan Terhadap

Potensi Resiko Penyakit Diabetes Melitus. *Jurnal Mandala Pharmacon Indonesia*, 5(01), 43–47. <https://doi.org/10.35311/jmpi.v5i01.43>

Wahyu Pribadi, W., Yunus, A., & Wiguna, A. S. (2022). Perbandingan Metode K-Means Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid-19 Di Kabupaten Malang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 493–500. <https://doi.org/10.36040/jati.v6i2.4808>

Widiasari, K. R., Wijaya, I. M. K., & Suputra, P. A. (2021). Diabetes Melitus Tipe 2: Faktor Risiko, Diagnosis, Dan Tatalaksana. *Ganesha Medicine*, 1(2), 114. <https://doi.org/10.23887/gm.v1i2.40006>





Lampiran 1. Data Mentah

	A	B	C	D	E	F	G	H	I
1	Usia	Status - Tekanan	akan Di	kanan Da	Status - Gula Dara	Gula Daral	Status - H	Hasil - IMT	L - Lingka
2	52	HIPERTENSI	150	100	HIPERGLIKEMIK	285	OBESITAS	33.155655533	101
3	60	PRE-HIPERTENSI	120	80	HIPERGLIKEMIK	422	BB LEBIH	25.292600674	86
4	60	PRE-HIPERTENSI	120	70	HIPERGLIKEMIK	278	OBESITAS	28.841531806	106
5	75	HIPERTENSI	200	110	HIPERGLIKEMIK	310	OBESITAS	28.841531806	84
6	61	HIPERTENSI	160	80	HIPERGLIKEMIK	405	NORMAL	18.718042367	79
9	58	HIPERTENSI	140	80	HIPERGLIKEMIK	219	NORMAL	20.549886621	91
10	57	PRE-HIPERTENSI	120	70	HIPERGLIKEMIK	272	NORMAL	23.283418554	81
11	59	PRE-HIPERTENSI	120	70	HIPERGLIKEMIK	217	BB KURANG	14.177693762	83
12	63	HIPERTENSI	150	90	HIPERGLIKEMIK	213	OBESITAS	35.204475309	109
13	53	HIPERTENSI	170	100	HIPERGLIKEMIK	200	OBESITAS	36.747580353	91
14	52	PRE-HIPERTENSI	120	80	HIPERGLIKEMIK	422	BB LEBIH	25.292600674	86
15	43	PRE-HIPERTENSI	120	70	HIPERGLIKEMIK	278	OBESITAS	28.841531806	106
16	65	HIPERTENSI	200	110	HIPERGLIKEMIK	310	OBESITAS	28.841531806	84
17	58	HIPERTENSI	140	70	HIPERGLIKEMIK	217	NORMAL	24.456063417	92
18	78	HIPERTENSI	170	100	HIPERGLIKEMIK	274	NORMAL	21.878715815	75
19	47	HIPERTENSI	160	100	HIPERGLIKEMIK	206	NORMAL	23.372576177	83
20	69	PRE-HIPERTENSI	120	60	HIPERGLIKEMIK	367	NORMAL	23.833004602	82
21	58	HIPERTENSI	140	80	HIPERGLIKEMIK	217	NORMAL	24.888888889	91
22	73	PRE-HIPERTENSI	130	60	HIPERGLIKEMIK	320	NORMAL	24.304617877	84
23	61	HIPERTENSI	140	80	HIPERGLIKEMIK	209	NORMAL	21.208448753	74
24	55	PRE-HIPERTENSI	120	60	HIPERGLIKEMIK	403	NORMAL	23.147255121	80
25	28	PRE-HIPERTENSI	120	60	HIPERGLIKEMIK	347	NORMAL	24.654832347	91
26	76	HIPERTENSI	150	90	HIPERGLIKEMIK	229	NORMAL	21.620647719	78
27	66	HIPERTENSI	170	100	HIPERGLIKEMIK	218	NORMAL	21.09375	84
28	46	HIPERTENSI	140	70	HIPERGLIKEMIK	337	BB LEBIH	26.222222222	93
31	63	HIPERTENSI	180	100	HIPERGLIKEMIK	267	BB LEBIH	26.666666667	107
32	52	PRE-HIPERTENSI	120	70	HIPERGLIKEMIK	340	NORMAL	24.557752341	84
33	58	NORMAL	90	60	HIPERGLIKEMIK	368	NORMAL	22.10028959	78
34	36	PRE-HIPERTENSI	130	80	HIPERGLIKEMIK	212	OBESITAS	30.381944444	86
35	51	HIPERTENSI	160	90	HIPERGLIKEMIK	236	OBESITAS	29.094282341	106
36	59	PRE-HIPERTENSI	120	80	HIPERGLIKEMIK	372	NORMAL	20.932120125	75
37	36	NORMAL	110	60	HIPERGLIKEMIK	225	NORMAL	22.074099723	85
38	70	PRE-HIPERTENSI	130	80	HIPERGLIKEMIK	266	NORMAL	24.835646457	91
39	83	HIPERTENSI	150	80	HIPERGLIKEMIK	285	NORMAL	22.054190296	84
40	46	HIPERTENSI	140	70	HIPERGLIKEMIK	433	OBESITAS	27.141582391	103
41	49	PRE-HIPERTENSI	120	60	HIPERGLIKEMIK	309	NORMAL	22.942130454	83
42	48	NORMAL	100	70	HIPERGLIKEMIK	238	NORMAL	23.833004602	80
44	44	HIPERTENSI	140	80	HIPERGLIKEMIK	285	NORMAL	23.831224602	79
46	50	HIPERTENSI	140	80	HIPERGLIKEMIK	267	NORMAL	23.833004602	86
47	66	HIPERTENSI	200	100	HIPERGLIKEMIK	211	NORMAL	18.596908442	68
48	66	HIPERTENSI	150	70	HIPERGLIKEMIK	201	NORMAL	23.805401662	86

	A	B	C	D	E	F	G	H	I
1000	23	HIPERTENSI	170	100	HIPERGLIKEMIK	272	BB LEBIH	25.777777778	92
1001	51	HIPERTENSI	140	80	HIPERGLIKEMIK	211	NORMAL	22.892819979	81
1002	45	HIPERTENSI	130	70	HIPERGLIKEMIK	380	NORMAL	19.628264701	82
1003	25	HIPERTENSI	170	80	HIPERGLIKEMIK	257	NORMAL	19.705532421	80
1004	52	HIPERTENSI	130	80	HIPERGLIKEMIK	345	NORMAL	23.739956172	82
1005	51	HIPERTENSI	140	80	HIPERGLIKEMIK	206	OBESITAS	30.530900723	94
1006	54	HIPERTENSI	120	70	HIPERGLIKEMIK	216	NORMAL	20.04988019	67
1007	20	HIPERTENSI	100	70	HIPERGLIKEMIK	328	NORMAL	21.333333333	81
1008	54	HIPERTENSI	150	80	HIPERGLIKEMIK	211	NORMAL	18.596908442	68
1009	52	HIPERTENSI	150	80	HIPERGLIKEMIK	236	NORMAL	23.7332384	82
1010	50	HIPERTENSI	140	80	HIPERGLIKEMIK	227	NORMAL	22.769438354	76
1011	27	HIPERTENSI	100	70	HIPERGLIKEMIK	216	BB LEBIH	26.037493991	91
1012	92	HIPERTENSI	110	70	HIPERGLIKEMIK	212	NORMAL	23.555555556	83
1013	50	HIPERTENSI	140	80	HIPERGLIKEMIK	211	BB KURANG	16.171224732	79
1014	64	HIPERTENSI	140	80	HIPERGLIKEMIK	202	OBESITAS	33.783783784	104
1015	54	PRE-HIPERTENSI	120	70	HIPERGLIKEMIK	318	NORMAL	22.432302516	86
1016	66	HIPERTENSI	130	90	HIPERGLIKEMIK	393	BB LEBIH	26.953125	103
1017	52	HIPERTENSI	170	80	HIPERGLIKEMIK	300	BB KURANG	18.314255983	77
1018	59	HIPERTENSI	140	80	HIPERGLIKEMIK	301	BB KURANG	17.96875	86
1019	41	HIPERTENSI	150	80	HIPERGLIKEMIK	333	NORMAL	21.750196677	88
1020	45	HIPERTENSI	140	80	HIPERGLIKEMIK	321	NORMAL	24.141519251	93
1021	56	HIPERTENSI	190	100	HIPERGLIKEMIK	211	NORMAL	18.596908442	68
1022	51	HIPERTENSI	140	80	HIPERGLIKEMIK	517	BB KURANG	18.080210388	85
1023	70	NORMAL	110	70	HIPERGLIKEMIK	227	NORMAL	22.769438354	76
1024	61	PRE-HIPERTENSI	120	60	HIPERGLIKEMIK	356	NORMAL	18.765246763	92
1025	55	HIPERTENSI	150	90	HIPERGLIKEMIK	284	NORMAL	22.222222222	78
1026	59	HIPERTENSI	140	90	HIPERGLIKEMIK	247	NORMAL	21.049817902	92
1027	56	PRE-HIPERTENSI	120	80	HIPERGLIKEMIK	206	NORMAL	19.227687871	83
1028	57	NORMAL	110	70	HIPERGLIKEMIK	204	BB LEBIH	25.33308313	94
1029	51	HIPERTENSI	140	70	HIPERGLIKEMIK	282	NORMAL	20.239500759	86
1030	60	HIPERTENSI	170	100	HIPERGLIKEMIK	345	NORMAL	23.739956172	82
1031	61	PRE-HIPERTENSI	130	70	HIPERGLIKEMIK	200	NORMAL	24.94009487	97
1032	52	HIPERTENSI	160	90	HIPERGLIKEMIK	450	NORMAL	21.516944594	91
1033	67	HIPERTENSI	190	100	HIPERGLIKEMIK	6148	NORMAL	20.932120125	91
1034	63	HIPERTENSI	130	90	HIPERGLIKEMIK	312	NORMAL	19.53125	81
1035	62	HIPERTENSI	180	80	HIPERGLIKEMIK	321	BB KURANG	17.777777778	76
1036	45	HIPERTENSI	150	100	HIPERGLIKEMIK	404	NORMAL	19.051973784	83
1037	71	NORMAL	100	70	HIPERGLIKEMIK	256	OBESITAS	28.377100131	101
1038	56	NORMAL	110	70	HIPERGLIKEMIK	255	NORMAL	24.034609838	85
1039	53	HIPERTENSI	200	100	HIPERGLIKEMIK	310	NORMAL	21.620647719	84
1040	47	HIPERTENSI	140	90	HIPERGLIKEMIK	249	OBESITAS	27.025809648	95
1041	38	HIPERTENSI 30	140	80	HIPERGLIKEMIK	232	BB LEBIH	26.896180742	99

Gambar 10. Data Mentah Pasien Diabetes Melitus

Lampiran 2. Preprocessing

1	A	B	C	D	E	F	G	H
2	Nama Peserta Deteksi	AMIN, M	Usia	h, NORMAL, PRE-H	ekanan Da	kanan Dal	tu, NORMAL, PRE-HIPE	Gula Darah
3	SULWANDI	0	45	2	150	80	2	282
4	Arsyad	0	51	1	130	70	0	124
5	ILHAM MUSTAFA	0	48	2	170	100	2	531
6	ABO	1	59	2	150	80	2	264
7	JUMRIA	1	61	2	160	80	2	405
8	FARIDAWATI	1	47	1	120	70	2	217
9	ANIS SUGIARTI	1	34	1	120	70	2	200
10	RUKAYA	1	54	0	130	80	1	152
11	HARTIA	1	45	1	130	80	2	236
12	JAMAL	0	59	1	120	70	2	217
13	MUSDALIFAH	1	55	2	200	130	2	341
14	AMBO UDE	0	45	0	120	80	1	199
15	ANSAR	0	63	0	130	80	0	112
16	SALMAWATI	1	43	1	120	70	2	278
17	Syamsuddn	0	59	0	120	80	1	173
18	KONTI	0	74	2	190	100	2	262
19	NAWIRA	1	78	2	170	100	2	274
20	RAHMAN	0	57	2	180	100	2	210
21	SABIR	0	52	1	130	90	0	94
22	RAMLI	0	51	2	190	110	1	162
23	RAJA	0	73	1	130	60	2	320
24	BEDDU	0	61	2	140	80	2	209
25	TEKKO K	0	55	1	120	60	2	403
26	TAHIR	0	28	1	120	60	2	347
27	ATIMA	1	76	2	150	90	2	229
28	ABD. RASYID. S	0	66	2	170	100	2	218
29	LATIF	0	64	2	160	100	2	379
30	Jamaluddin	0	47	1	130	80	0	112
31	HERWANA	1	49	2	140	80	2	288
32	A. Ungkar	0	59	2	180	100	2	466
33	ABD. HALIM	0	65	0	110	70	0	124
34	BABA	0	58	0	90	60	2	368
35	HASFIANI	1	36	1	130	80	2	212
36	ABDUL MAJID	0	48	2	170	90	2	203
37	MARSUKI	0	59	1	120	80	2	372
38	NURFATIMAH	1	36	0	110	60	2	225
39	LANTO	0	64	2	190	110	2	212
40	MUH. DIRGANTARA	0	30	1	130	70	2	216
41	SYUKRI	0	49	2	150	100	2	253
42	MUHAMMA	0	52	1	130	80	2	398
43	SYAKIR	0	48	0	100	70	2	238
44	ANWAR. P	0	48	2	110	70	2	210
45	NURAEDA	1	44	2	140	80	2	285
46	SUARDI	0	58	2	220	120	2	328
47	Jufri	0	49	1	130	70	0	126
48	RENGE	1	66	2	200	100	2	211
49	ABANG	1	66	2	150	70	2	201
50	NANDA BIN AHMAD	0	51	0	100	70	2	239
51	SINAR	1	45	1	120	80	2	211
52	MARSUKI	0	51	2	170	80	1	143
53	Rustan Efendi	0	45	0	100	70	0	127
54	Muhammad Hatta	0	49	2	150	100	0	108
55	AMIRUDDIN	0	54	2	150	80	2	1109
56	MANGNGARIBI	0	91	2	150	80	2	297
57	MUH. LUTFI MIKAIL	0	33	1	120	70	2	278
58	A. KAMARUDDIN	0	60	0	110	70	0	113
59	JASMIN	1	58	2	150	80	2	211
60	MUH. KASRIM	0	52	2	190	80	2	228
61	RAHMI	1	47	0	100	60	0	104
62	SUPARMA	1	40	2	170	80	2	244

	A	B	C	D	E	F	G	H
982	RAFARDAN	0	37	2	140	80	2	237
983	NURHAYATI	1	15	2	160	90	2	326
984	DARMA	1	63	2	160	90	2	216
985	HASMI	1	39	2	150	60	2	413
986	SUMARNI	1	34	2	140	50	2	390
987	DINAR	1	54	1	120	70	2	318
988	HAMKA	0	63	2	130	80	2	345
989	HASNA	1	59	2	140	80	2	301
990	SINA	1	34	2	220	100	2	220
991	M. ADNAN ARDHANI	0	27	2	200	90	2	200
992	SHOFIE SALSABILAH	1	34	2	200	100	2	200
993	AHMAD	0	37	2	130	80	2	201
994	Hj. HAERUL BARIA	1	45	2	140	80	2	321
995	NURSYIDAH	1	44	2	170	100	2	219
996	MADDA	0	42	2	160	80	2	219
997	FARIDA RIDWAN	1	23	2	110	70	2	210
998	SULWANDI	0	21	2	140	80	2	276
999	RAMLAH	1	32	2	120	70	2	395
1000	AZYILAH FIDYA AINUN	1	23	2	170	100	2	272
1001	Hj. ST. MAEMUNA	1	56	0	110	70	2	255
1002	INDRAWATI	1	51	2	140	80	2	517
1003	SRI WULAN DARI	1	25	2	170	80	2	257
1004	ISYA	1	61	1	130	70	2	200
1005	JUMRIA	1	61	1	120	60	2	356
1006	KADIR	0	54	2	120	70	2	216
1007	NURMIATI	1	20	2	100	70	2	328
1008	SUSI	1	54	2	150	80	2	211
1009	MARHUMA TALIBE	1	48	2	130	80	2	395
1010	RAHMA	1	50	2	140	80	2	227
1011	SURIANI	1	27	2	100	70	2	216
1012	MARNI	1	46	2	150	80	2	222
1013	HASMA	1	50	2	140	80	2	211
1014	MISNAH	1	83	2	160	80	2	200
1015	MISNAWATI	1	67	2	150	80	2	236
1016	MULIATI	1	51	2	150	80	2	339
1017	ERNAWATI	1	52	2	170	80	2	300
1018	NAHRIA	1	54	2	180	100	2	222
1019	HASNIATI	1	41	2	150	80	2	333
1020	NURAENI	1	60	2	170	80	2	345
1021	Hj. KURNIA	1	56	2	190	100	2	211
1022	NURBAYA	1	76	2	180	100	2	262
1023	JUNAEDAH	1	70	0	110	70	2	227
1024	NURDIN	0	73	2	130	86	2	243
1025	JUMIATI	1	55	2	150	90	2	284
1026	NURHAYATI	1	55	2	120	70	2	395
1027	NURHAYATI	1	51	2	140	70	2	282
1028	NURJANNAH	1	57	0	110	70	2	204
1029	ONNANG	0	58	2	140	50	2	245
1030	PETTA SITTI	1	60	2	170	100	2	345
1031	ROSMIATI	1	52	2	140	80	2	206
1032	ROSMIATI	1	55	2	160	90	2	450
1033	SAHRYA	1	66	2	130	81	2	380
1034	SAKKA	1	61	2	140	80	2	255
1035	SULLE	1	62	2	180	80	2	321
1036	SALMA	1	53	2	200	100	2	310
1037	SALMAH AMIR	1	63	2	130	90	2	312
1038	SUFIANAH	1	60	2	120	80	2	1109
1039	SUKAENA	1	66	2	110	80	2	262
1040	UFE	1	49	2	140	90	2	204
1041	SALMA	1	38	2	140	80	2	232

Gambar 11. Preprocessing Data

Lampiran 3. Hasil Clustering

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P				
1	eserta	Det	MIN, MAL	Usia	ORMAI, P	Prkan Dar	an Dar	ORMAI, P	Jula Dar	5/BB KURAN	Hasil- IMT	Status -	Lingkar Perut, NORMAL/OBESITAS SENTRAL	Lingkar	PDJ	NORMA	an,NORM	an,NORM	Kluster
1	SULWANDI	0	45	2	150	80	2	282	0	24.26867	1	89	1	1	0	0	0	0	1
2	Aryevy	0	51	1	130	70	0	124	2	25.0995	1	87	1	1	0	0	0	0	1
3	ILHAM MUR	0	48	2	170	100	2	531	0	24.65483	1	99	0	1	0	0	0	0	0
4	ABO	1	59	2	150	80	2	264	0	23.19109	1	87	1	1	0	0	0	0	0
5	JUMIRIA	1	61	2	160	80	2	405	0	18.71804	0	79	1	0	0	1	0	0	0
6	FARIDAWA	1	47	1	120	70	2	217	1	14.12769	0	80	1	0	1	0	1	0	1
7	ANIS SUGIA	1	34	1	120	70	2	200	3	27.92667	1	85	1	0	0	0	0	1	1
8	RUKAYA	1	54	0	130	80	1	152	0	22.50693	1	92	0	1	0	0	0	0	0
9	HARTIA	1	45	1	130	80	2	236	0	23.73324	1	82	0	0	0	0	0	0	1
10	JAMAL	0	59	1	120	70	2	217	1	14.17769	0	83	0	0	0	0	0	0	1
11	MUSDALIFA	1	55	2	200	150	2	341	3	28.57143	1	100	0	0	0	0	0	0	0
12	AMBO UDE	0	45	0	120	80	1	199	0	23.30905	1	82	0	1	0	0	0	0	1
13	ANSAR	0	63	0	130	80	0	112	3	63.77551	1	87	0	1	1	1	0	0	0
14	SALMAWAT	1	43	1	120	70	2	278	3	28.84153	1	106	1	0	0	0	0	0	1
15	Syamsuddin	0	59	0	120	80	1	173	2	23.33333	1	89	0	0	0	0	0	0	0
16	KONTI	0	74	2	190	100	2	262	0	23.94246	1	81	1	0	0	1	0	0	0
17	NAWIRA	1	78	2	170	100	2	274	0	21.87872	0	75	0	1	1	0	0	0	0
18	RAHMAN	0	57	2	180	100	2	210	0	19.75647	1	92	0	1	0	0	0	0	0
19	SABIR	0	52	1	130	90	0	94	0	24.22145	1	92	0	0	0	0	0	0	0
20	RAMLI	0	51	2	190	110	1	162	0	19.33373	1	97	0	0	1	0	0	0	0
21	RAJIA	0	73	1	130	60	2	320	0	24.30462	0	84	1	1	1	1	1	1	1
22	BEDDU	0	61	2	140	80	2	209	0	21.20845	0	74	1	1	1	1	1	1	1
23	TERKO K	0	55	1	120	60	2	405	0	23.14726	0	80	1	1	0	0	0	0	1
24	TAMIR	0	38	1	120	60	2	347	0	24.65483	1	91	0	0	0	0	0	0	1
25	ATIMA	1	76	2	150	90	2	229	0	21.62065	0	78	0	1	1	0	0	0	0
26	ABD. RASYID	0	66	2	170	100	2	218	0	21.09375	0	84	1	1	1	1	0	0	0
27	LATIF	0	44	2	160	100	2	379	3	28.96943	1	102	1	0	0	0	0	0	0
28	Imahudin	0	47	1	130	80	0	112	0	21.55856	1	85	1	0	0	0	0	0	0
29	HERWANJA	1	49	2	140	80	2	288	0	19.23916	0	71	0	0	0	0	0	0	1
30	A.Ungkar	0	59	2	180	100	2	466	0	21.64127	1	85	1	1	0	0	0	0	0
31	ABD. HALIM	0	65	0	110	70	0	124	0	19.22769	1	87	0	1	1	1	1	1	1
32	BAKUR	0	48	0	90	60	0	368	0	22.10029	0	78	0	0	0	0	0	0	0
33	HASFANI	1	36	1	130	80	2	212	3	30.38194	1	86	1	0	0	0	0	0	0
34	ABDUL MAJ	0	48	2	170	90	2	203	0	24.91077	1	98	1	1	0	0	0	0	0
35	MARSUKI	0	59	1	120	80	2	372	0	20.93212	0	75	0	1	0	0	0	0	1
36	NURFATIM	1	36	0	110	60	2	225	0	22.0741	1	85	1	0	0	0	0	0	1
37	LANTO	0	64	2	190	110	2	211	0	21.87142	1	94	1	1	0	0	0	0	0
38	MUH. DIRG	0	30	1	130	70	2	216	2	26.03749	1	91	1	0	0	0	0	0	0
39	SYUKRI	0	49	2	150	100	2	253	0	23.50781	1	99	1	0	0	0	0	0	0
40	MUHAMMAD	0	52	1	130	80	2	398	3	28.78906	1	104	1	0	0	0	0	0	0
41	SYAKR	0	48	0	120	70	2	239	0	23.8338	1	80	0	0	0	0	0	0	0
42	ANWAR P	0	48	2	110	70	2	210	5	27.91552	1	110	1	0	0	0	0	0	0
43	NURAEDA	1	44	2	140	80	2	285	0	23.88410	0	79	1	0	0	0	0	0	0
44	SUARDA	0	58	2	220	120	2	328	2	26.35907	1	100	1	1	0	0	0	0	0
45	JUFI	0	49	1	130	70	0	126	2	25.40282	1	93	1	0	0	0	0	0	1
46	RENGE	1	66	2	140	100	2	211	1	18.59691	0	68	0	0	0	0	0	0	0
47	ARANG	1	66	2	140	70	2	301	0	14.80642	0	86	1	0	0	0	0	0	0
973	DAFFAR	0	67	2	140	80	2	211	3	208.3333	0	79	1	0	0	0	0	0	0
974	MULIANI	1	45	2	110	70	2	201	0	20.50493	0	80	1	1	0	0	0	0	1
975	FAHARUDD	0	54	2	130	90	2	206	0	21.58889	1	74	1	0	0	0	0	0	0
976	NURUL FAH	0	50	2	150	80	2	201	3	562.1302	0	80	1	0	0	0	0	0	0
977	DINAR	1	26	2	140	80	2	211	0	21.05171	0	79	1	0	0	0	0	0	1
978	NUR VITA	1	19	2	100	70	2	490	2	25.29938	1	96	0	0	0	0	0	0	0
979	BAKUR	1	40	2	110	70	2	600	0	18.22222	1	82	0	0	0	0	0	0	0
980	RAFARDAN	0	37	2	140	80	2	257	3	44.20756	1	132	0	0	0	0	0	0	0
981	NURHAYATI	1	31	2	160	90	2	326	3	31.32906	1	107	0	0	0	0	0	0	0
982	DARMA	1	63	2	160	90	2	216	2	26.03749	1	91	1	0	0	0	0	0	0
983	HASMI	1	39	2	150	60	2	413	0	23.61275	0	87	1	0	0	0	0	0	1
984	SUMARNI	1	34	2	140	50	2	390	0	14.87389	1	88	1	1	0	0	0	0	0
985	DINAR	1	54	1	120	70	2	318	0	22.4323	1	86	1	1	0	0	0	0	1
986	HAMKA	0	63	2	130	80	2	345	0	23.73996	1	82	1	1	1	1	1	1	1
987	HASNA	1	59	2	140	80	2	301	1	17.96875	1	86	1	0	0	0	0	0	0
988	SINA	1	34	2	220	100	2	220	3	30.08913	1	87	1	0	0	0	0	0	0
989	M. ADNAN	0	27	2	200	90	2	200	1	18.39538	0	78	1	0	0	0	0	0	0
990	SHOFIE SAL	1	34	2	200	100	2	200	2	25.88757	0	90	1	1	0	0	0	0	0
991	AHMAD	0	37	2	130	80	2	201	0	20.02884	1	81	1	1	0	0	0	0	0
992	HJ. HANIBUL	1	45	2	140	80	2	331	0	24.14153	1	93	1	0	0	0	0	0	0
993	NURSYIDAH	1	44	2	170	100	2	219	0	21.00775	0	78	1	0	0	0	0	0	0
994	MADDA	0	42	2	160	80	2	219	0	22.65625	0	88	0	0	0	0	0	0	0
995	FARIDA RID	1	23	2	110	70	2	210	3	27.91552	1	110	1	0	0	0	0	0	0
996	SULWANDI	0	21	2	140	80	2	276	0	21.12	0	85	0	0	0	0	0	0	1
997	RAMLAH	1	32	2	120	70	2	395	0	20.61313	1	86	1	0	0	0	0	0	0
1000	AZYLAH FIC	1	23	2	170	100	2	272	2	25.77778	1	92	1	1	0	0	0	0	0
1001	HJ. ST. MAE	1	56	0	110	70	2	255	0	24.03461	1	85	1	0	0	0	0	0	0
1002	INDRAWAT	1	51	2	140	80	2	517	1	18.08021	1	85	1	0	0	0	0	0	0
1003	SRI WULAN	1	25	1	170	80	2	217	0	19.70553	0	80	1	0	0	0	0	0	0
1004	ISYA	1	61	1	130	70	2	200	0	24.94009	1	97	1	1	0	0	0	0	0
1005	JUMIRIA	1	61	1	120	60	2	356	0	18.76525	1	92	1	0	0	0	0	0	0
1006	KADIR	0	54	2	120	70	2	216	0	20.04988	0	67	1	0	0	0	0	0	0
1007	NURMIATI	1	20	2	100	70	2	328	0	21.33333	1	83	1	0	0	0	0	0	0
1008	SUSI	1	54	2	150	80	2	211	0	18.59691	0	68	1	0	0	0	0	0	0
1009	MARHUMA	1	48	2	130	80	2	395	0	20.61313	1	86	1	1	0	0	0	0	0
1010	RAHMA	1	50	2	140	80	2	227	0	22.76944	0	76	1	0	0	0	0	0	0
1011	SURIANI	1	27	2	100	70	2	216	2	26.03749	1	91	1	0	0	0	0	0	0
1012																			

Lampiran 4. Hasil Prediksi K-NN Metric Manhattan

	A	B	C	D	E	F	G
1	ma Peserta Deteksi	ADMIN, MAL	Usia	Label Asli	Prediksi		
2	JUNAEDAH	1	71	Diabetes R	Diabetes Ringan		
3	TATI	1	54	Diabetes K	Diabetes Kronis		
4	JAMALUDDIN. B	0	59	Diabetes K	Diabetes Kronis		
5	BABA	0	58	Diabetes K	Diabetes Kronis		
6	HASNI	1	58	Diabetes R	Diabetes Ringan		
7	NUR IZZATUL JANNAH	0	39	Diabetes R	Diabetes Kronis		
8	DAMRIA	1	57	Diabetes K	Diabetes Kronis		
9	SARSINA. M	1	48	Diabetes K	Diabetes Kronis		
10	Uttang	0	47	Diabetes K	Diabetes Kronis		
11	FATIMAH	1	50	Diabetes K	Diabetes Kronis		
12	SUNARTI	1	42	Diabetes K	Diabetes Kronis		
13	Colli	0	67	Diabetes K	Diabetes Kronis		
14	MUHYAR YUSUF	0	53	Diabetes R	Diabetes Ringan		
15	A. NUR AZNIZAH	1	27	Diabetes K	Diabetes Ringan		
16	PADU	0	59	Diabetes R	Diabetes Ringan		
17	KEDO	0	75	Diabetes R	Diabetes Ringan		
18	AMALIA	1	42	Diabetes K	Diabetes Kronis		
19	NURAENI	1	58	Diabetes R	Diabetes Ringan		
20	ATI	1	47	Diabetes R	Diabetes Ringan		
21	TAMRING	0	69	Diabetes K	Diabetes Kronis		
22	ALFIN HIDAYAT	0	17	Diabetes K	Diabetes Ringan		
23	ALIFATUNNISA	1	44	Diabetes R	Diabetes Ringan		
24	NADI	1	52	Diabetes K	Diabetes Kronis		
25	MARIATI	1	54	Diabetes K	Diabetes Kronis		
26	ISHAR IHWAN	0	34	Diabetes K	Diabetes Kronis		
27	IRMAWATI	1	43	Diabetes R	Diabetes Ringan		
28	SUARDI	0	38	Diabetes K	Diabetes Kronis		
29	BURHAN	0	54	Diabetes K	Diabetes Ringan		
30	BECCE	1	60	Diabetes K	Diabetes Kronis		
31	Andi Reskinawati	1	33	Diabetes K	Diabetes Kronis		
179	KASMAWATI	1	37	Diabetes K	Diabetes Kronis		
180	HASNA	1	59	Diabetes R	Diabetes Ringan		
181	A. KAMARUDDIN	0	60	Diabetes K	Diabetes Kronis		
182	MARWIYAH	1	63	Diabetes R	Diabetes Ringan		
183	JAMALUDDIN. B	0	29	Diabetes R	Diabetes Ringan		
184	ABD. RASYID. S	0	66	Diabetes R	Diabetes Ringan		
185	SUKMA	1	26	Diabetes K	Diabetes Kronis		
186	HAMZAH RIJAL	0	37	Diabetes K	Diabetes Kronis		
187	Jufri	0	49	Diabetes K	Diabetes Kronis		
188	NURAEDA	1	48	Diabetes K	Diabetes Kronis		
189	MUHAMMAD ALFARI	0	35	Diabetes R	Diabetes Ringan		
190	NASRUN	0	58	Diabetes K	Diabetes Kronis		
191	LATIF	0	68	Diabetes K	Diabetes Ringan		
192	ROSTINA	1	20	Diabetes K	Diabetes Ringan		
193	MINA	1	77	Diabetes K	Diabetes Kronis		
194	MURNIATI	1	26	Diabetes R	Diabetes Ringan		
195	EDI	0	42	Diabetes K	Diabetes Kronis		
196	MAS'UD NUR HAIKAL	0	19	Diabetes K	Diabetes Kronis		
197	BUATANG	1	62	Diabetes K	Diabetes Kronis		
198	HAMMING	1	58	Diabetes R	Diabetes Ringan		
199	ALIMUDDIN	0	23	Diabetes K	Diabetes Kronis		
200	Agus	0	42	Diabetes K	Diabetes Kronis		
201	ATI	1	38	Diabetes K	Diabetes Kronis		
202	Irma	1	46	Diabetes K	Diabetes Kronis		
203	AZYILAH FIDYA AINUN	1	23	Diabetes R	Diabetes Ringan		
204	MUNAWARAH	1	49	Diabetes K	Diabetes Kronis		
205	Esse	1	34	Diabetes K	Diabetes Kronis		
206	SATRIANI	1	35	Diabetes K	Diabetes Kronis		
207	HUSNUN NAZAR	0	47	Diabetes R	Diabetes Ringan		
208	ROSTINA	1	57	Diabetes K	Diabetes Kronis		
209	MARNI	1	46	Diabetes R	Diabetes Ringan		

Gambar 12. Hasil Prediksi KNN Metric Manhattan

Lampiran Turnitin

UNIVERSITAS MUHAMMADIYAH MAKASSAR
FAKULTAS TEKNIK

 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. Fadliansyah
Nim : 105841112320
Program Studi : Teknik Informatika

Dengan nilai:

No	Bab	Nilai	Ambang Batas
1	Bab 1	10 %	10 %
2	Bab 2	19 %	25 %
3	Bab 3	10 %	10 %
4	Bab 4	10 %	10 %
5	Bab 5	5 %	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, 21 Agustus 2024
Mengetahui,
Kepala UPT- Perpustakaan dan Penerbitan,


Nur Anik S. Hum., M.I.P.
NIBM. 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



BAB I M. Fadliansyah -
105841112320

by Tahap Tutup

Submission date: 21-Aug-2024 10:22AM (UTC+0700)

Submission ID: 2435338787

File name: BAB_I_FADLI.docx (21K)

Word count: 909

Character count: 6221

BAB I M. Fadliansyah - 105841112320

ORIGINALITY REPORT

10% SIMILARITY INDEX
11% INTERNET SOURCES
6% PUBLICATIONS
0% STUDENT PAPERS

PRIMARY SOURCES

- 1** Annisa Nurba Iffah'da, Anita Desiani. "Implementasi Algoritma K-Nearest Neighbor (K-NN) dan Single Layer Perceptron (SLP) Dalam Prediksi Penyakit Sirosis Biliari Primer", Jurnal Ilmiah Informatika, 2022
Publication **3%**
- 2** id.123dok.com
Internet Source **2%**
- 3** repository.uinsu.ac.id
Internet Source **2%**
- 4** adoc.pub
Internet Source **2%**
- 5** repository.uma.ac.id
Internet Source **2%**

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On

BAB II M. Fadliansyah -
105841112320

by Tahap Tutup

Submission date: 21-Aug-2024 10:25AM (UTC+0700)

Submission ID: 2435339843

File name: BAB_II_FADLI.docx (102.04K)

Word count: 1512

Character count: 10005

AB II M. Fadliansyah - 105841112320

ORIGINALITY REPORT

19%

SIMILARITY INDEX

20%

INTERNET SOURCES

9%

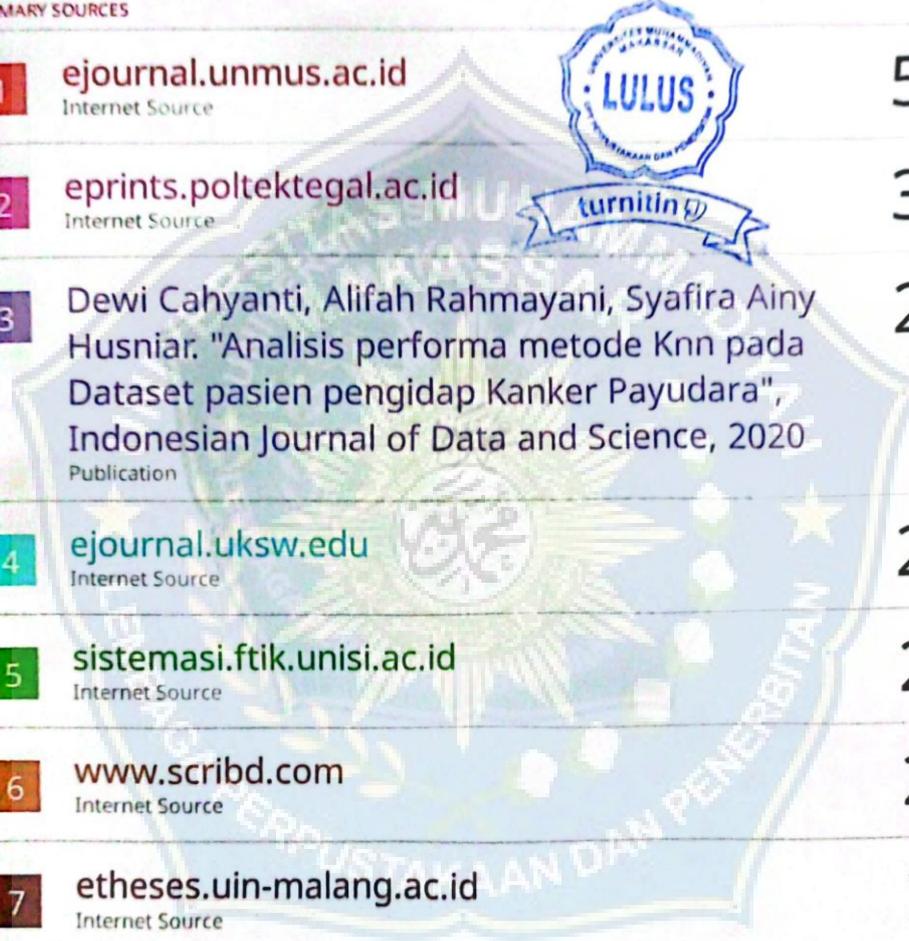
PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1	ejournal.unmus.ac.id Internet Source	5%
2	eprints.poltektegal.ac.id Internet Source	3%
3	Dewi Cahyanti, Alifah Rahmayani, Syafira Ainy Husniar. "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara", Indonesian Journal of Data and Science, 2020 Publication	2%
4	ejournal.uksw.edu Internet Source	2%
5	sistemasi.ftik.unisi.ac.id Internet Source	2%
6	www.scribd.com Internet Source	2%
7	etheses.uin-malang.ac.id Internet Source	2%
8	kc.umn.ac.id Internet Source	2%



BAB III M. Fadliansyah - 105841112320

by Tahap Tutup

Submission date: 21-Aug-2024 10:26AM (UTC+0700)

Submission ID: 2435340740

File name: BAB_III_FADLI.docx (38.17K)

Word count: 435

Character count: 2691

BAB III M. Fadliansyah - 105841112320

ORIGINALITY REPORT

10% SIMILARITY INDEX **6%** INTERNET SOURCES **7%** PUBLICATIONS **%** STUDENT PAPERS

PRIMARY SOURCES

- 1** Steven Joses, Stefanie Quinevera, Ricky Mardianto, Donata Yulvida, Ary Mazharuddin Shiddiqi. "Pendekatan Metode Ensemble Learning untuk Deteksi Serangan DDoS menggunakan Soft Voting Classifier", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2024
Publication 4%
- 2** pt.scribd.com
Internet Source 4%
- 3** repository.radenintan.ac.id
Internet Source 2%

Exclude quotes On Exclude matches < 2%
Exclude bibliography On



BAB IV M. Fadliansyah -
105841112320

by Tahap Tutup

Submission date: 21-Aug-2024 10:27AM (UTC+0700)

Submission ID: 2435341057

File name: BAB_IV_FADLI.docx (731.69K)

Word count: 2598

Character count: 16473

AB IV M. Fadliansyah - 105841112320

ORIGINALITY REPORT

10%

SIMILARITY INDEX

10%

INTERNET SOURCES

7%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1	digilibadmin.unismuh.ac.id Internet Source	2%
2	www.sea-stat.com Internet Source	1%
3	www.coursehero.com Internet Source	1%
4	www.geeksforgeeks.org Internet Source	1%
5	publishing.intelgr.com Internet Source	1%
6	www.codewithc.com Internet Source	1%
7	repository.its.ac.id Internet Source	1%
8	docplayer.info Internet Source	1%
9	www.analyticsvidhya.com Internet Source	1%

10	medium.com Internet Source	1%
11	Ajeng Setianingrum, Ayu Hindayanti, Dita Meilani Cahya, Dini Silvi Purnia. "Perbandingan Metode Algoritma K-NN & Metode Algoritma C45 Pada Analisa Kredit Macet (Studi Kasus PT Tungmung Textil Bintan)", EVOLUSI : Jurnal Sains dan Manajemen, 2021 Publication	<1%
12	Sunil Kumar. "Python for Accounting and Finance", Springer Science and Business Media LLC, 2024 Publication	<1%
13	id.123dok.com Internet Source	<1%
14	www.adaface.com Internet Source	<1%
15	www.kemkes.go.id Internet Source	<1%
16	Dedy Elisa Limbong. "Rekonsiliasi Data Keuangan dan Karakterisasi Satuan Kerja dengan Teknik Data Mining", Indonesian Treasury Review Jurnal Perbendaharaan Keuangan Negara dan Kebijakan Publik, 2016 Publication	<1%



BAB V M. Fadliansyah -
105841112320

by Tahap Tutup

Submission date: 21-Aug-2024 10:28AM (UTC+0700)

Submission ID: 2435341308

File name: BAB_V_FADLI.docx (18.5K)

Word count: 313

Character count: 2151

AB V M. Fadliansyah - 105841112320

ORIGINALITY REPORT

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

PRIMARY SOURCES

1	repository.ub.ac.id Internet Source	5%
---	--	----

Exclude quotes On
Exclude bibliography On

Exclude matches On

