



The Implementation of the K-Means Clustering Algorithm Based on the Severity Level of Diabetes in Patients Using a Website Platform

Syaputri Maharani¹, Yumai Wendra ², Melladia³, Radiyan Rahim⁴

Email: syaputrimaharani20@gmail.com

¹ Universitas Nahdlatul Ulama Sumatera Barat

ABSTRACT

Diabetes mellitus is a chronic non-communicable disease with a steadily increasing prevalence worldwide, posing a significant public health challenge due to its potential for severe complications if not managed properly. In many healthcare facilities, including RSUD Pariaman, there is still no structured system to classify patients according to the severity of their condition, which hampers timely intervention and optimal resource allocation. This study aims to develop and implement a web-based system for clustering the severity levels of type 2 diabetes mellitus using the K-Means Clustering algorithm as a decision support tool for medical staff. A quantitative system development research design was applied, utilizing secondary medical records from January 2023 to December 2024, with five clinical variables: Hemoglobin A1c (HbA1c), Fasting Blood Glucose (GDP), Systolic Blood Pressure (TDS), Diastolic Blood Pressure (TDD), and Body Mass Index (BMI). The system was built using the CodeIgniter PHP framework, MySQL database, and Bootstrap-based interface, following the Knowledge Discovery in Database (KDD) process for data preprocessing. K-Means clustering was configured into three categories (mild, moderate, and severe). Validation using RapidMiner confirmed that the clustering results from the web-based system were consistent with the benchmark model, ensuring the correctness of the algorithm's implementation. The developed system enables real-time data processing, displays results in both tabular and graphical forms, and provides an intuitive interface for medical personnel, thus supporting clinical decision-making and improving healthcare service quality.

Keywords: Body Mass Index; Clustering; Diabetes Mellitus; K-Means; Medical Records

PENDAHULUAN

Diabetes mellitus merupakan salah satu penyakit tidak menular (PTM) yang menjadi masalah kesehatan utama baik secara global maupun nasional. Penyakit ini ditandai oleh kadar glukosa darah yang tinggi akibat gangguan sekresi insulin, kerja insulin, atau keduanya. Data Organisasi Kesehatan Dunia (World Health Organization/WHO) pada tahun 2021 mencatat sekitar 422 juta orang di dunia hidup dengan diabetes, dan jumlah tersebut diproyeksikan terus meningkat pada dekade mendatang seiring dengan perubahan gaya hidup, pola makan, dan penuaan populasi. Kondisi ini menjadi ancaman serius karena diabetes tidak hanya mengurangi kualitas hidup, tetapi juga meningkatkan risiko komplikasi yang mengancam jiwa seperti penyakit jantung koroner, stroke, nefropati, retinopati, neuropati, dan ulkus kaki diabetik yang berpotensi berujung amputasi (Alpian & Alfarizi, 2022).

Di Indonesia, Riset Kesehatan Dasar (Risksdas) menunjukkan adanya tren peningkatan prevalensi diabetes yang signifikan, terutama di wilayah perkotaan. Penyakit ini sering disebut



silent killer karena gejala awalnya kerap tidak disadari pasien, sehingga diagnosis sering terlambat dan komplikasi sudah berkembang. Dalam layanan kesehatan, rumah sakit sebagai garda terdepan memerlukan sistem pendukung yang mampu memberikan gambaran komprehensif tentang kondisi pasien secara cepat dan tepat(Harmaja et al., 2023).

Kondisi serupa terjadi di RSUD Pariaman, Sumatera Barat, di mana data rekam medis menunjukkan tren peningkatan jumlah kunjungan pasien diabetes dalam dua tahun terakhir. Meskipun rumah sakit telah memiliki sistem pencatatan rekam medis, sistem tersebut masih bersifat administratif dan belum memiliki kemampuan analisis untuk mengelompokkan pasien berdasarkan tingkat keparahan penyakit. Hal ini menghambat optimalisasi edukasi pasien yang seharusnya dipersonalisasi sesuai tingkat risiko, alokasi sumber daya medis seperti obat dan peralatan yang tepat sasaran, serta memperlambat pengambilan keputusan klinis karena dokter harus menganalisis data pasien secara manual. Permasalahan ini, baik secara empiris maupun teoretis, dapat diatasi melalui penerapan analisis berbasis *data mining*, khususnya metode pengelompokan (*clustering*), yang mampu mengubah data medis menjadi informasi yang bermakna dan langsung dapat digunakan dalam praktik klinis (Sari & Beti, 2023).

Berbagai penelitian telah membuktikan efektivitas *data mining* di bidang kesehatan. Salah satu metode yang sering digunakan untuk mengelompokkan data tanpa label adalah *K-Means Clustering*, yang bekerja dengan mempartisi data menjadi sejumlah klaster berdasarkan kesamaan karakteristik antar data. Penelitian oleh Nahjan et al., (2023) membuktikan bahwa *K-Means* efektif dalam segmentasi pasien rumah sakit berdasarkan parameter laboratorium, sehingga mempermudah penyusunan prioritas perawatan. Kurniawan et al., (2020) menerapkan metode ini untuk mengelompokkan pasien hipertensi berdasarkan tingkat risiko komplikasi dan hasilnya membantu tenaga medis dalam menentukan strategi intervensi. Penelitian oleh KHAIRANI, (2023) mengidentifikasi HbA1c sebagai indikator penting pengendalian glukosa darah jangka panjang, sementara penelitian lain menekankan pentingnya GDP, tekanan darah, dan indeks massa tubuh (IMT) dalam menentukan risiko komplikasi. Akan tetapi, sebagian besar penelitian ini masih terbatas pada analisis statistik atau grafik, tanpa integrasi ke dalam sistem informasi yang dapat diakses secara *real-time* oleh tenaga medis, serta sering kali menggunakan subset variabel yang kecil.

Berdasarkan tinjauan tersebut, terlihat adanya kesenjangan penelitian yang jelas. Belum ada penelitian yang secara khusus memfokuskan diri pada pengelompokan pasien diabetes berdasarkan tingkat keparahan di RSUD Pariaman. Meski algoritma *K-Means* telah terbukti efektif, sebagian besar implementasi masih berhenti pada tahap analisis di perangkat lunak seperti *RapidMiner* tanpa pengembangan sistem berbasis web yang terintegrasi dengan alur kerja rumah sakit (Asy Aria et al., 2023). Padahal, di RSUD Pariaman tersedia data klinis yang relatif lengkap meliputi HbA1c, GDP, tekanan darah sistolik (TDS), tekanan darah diastolik (TDD), dan IMT, yang berpotensi menghasilkan klasterisasi lebih akurat dan representatif. Hal ini membuka peluang untuk mengembangkan solusi yang tidak hanya menghasilkan model analisis, tetapi juga mempermudah akses dan visualisasi hasil bagi tenaga medis.

Berdasarkan analisis kesenjangan tersebut, penelitian ini dirancang untuk menjawab pertanyaan bagaimana algoritma *K-Means Clustering* dapat digunakan untuk mengelompokkan pasien diabetes di RSUD Pariaman berdasarkan tingkat keparahan penyakit, bagaimana karakteristik tiap klaster dapat dimanfaatkan untuk mendukung pengambilan keputusan klinis, dan bagaimana hasil klasterisasi tersebut dapat diimplementasikan dalam sistem berbasis web yang dapat diakses secara *real-time*. Tujuan penelitian adalah mengembangkan sistem berbasis web yang mengintegrasikan *K-Means Clustering* untuk mengelompokkan pasien diabetes berdasarkan tingkat keparahan penyakit, memvalidasi hasilnya menggunakan perangkat *data mining* (*RapidMiner*), serta menyajikan visualisasi data



yang memudahkan interpretasi oleh tenaga medis. Kebaruan penelitian ini terletak pada penerapan algoritma *K-Means* secara langsung pada dataset medis pasien diabetes RSUD Pariaman dengan variabel klinis yang lengkap, serta pengembangan sistem berbasis web yang memungkinkan hasil analisis digunakan secara praktis, mendukung proses klinis secara cepat, terintegrasi, dan berbasis data.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain pengembangan sistem (*system development research*), yang bertujuan membangun aplikasi berbasis web untuk mengelompokkan pasien diabetes di RSUD Pariaman berdasarkan tingkat keparahan penyakit menggunakan algoritma *K-Means Clustering*. Data yang digunakan merupakan data sekunder yang diambil dari rekam medis pasien periode Januari 2023 hingga Desember 2024, dengan variabel klinis yang relevan meliputi Hemoglobin A1c (HbA1c), Kadar Gula Darah Puasa (GDP), Tekanan Darah Sistolik (TDS), Tekanan Darah Diastolik (TDD), dan Indeks Massa Tubuh (IMT). Pengumpulan data dilakukan melalui observasi terhadap proses pencatatan pasien, wawancara dengan tenaga medis untuk menentukan variabel yang digunakan, serta dokumentasi data rekam medis yang tersedia di rumah sakit.

Tahapan penelitian mengacu pada alur *Knowledge Discovery in Database (KDD)*, dimulai dari seleksi data untuk menentukan atribut yang relevan, pembersihan data untuk menghilangkan entri yang tidak lengkap atau tidak valid, transformasi data menjadi format numerik yang siap dianalisis, penerapan *data mining* dengan algoritma *K-Means Clustering*, hingga evaluasi hasil klasterisasi. Proses *K-Means* dilakukan dengan menentukan jumlah klaster sebanyak tiga (ringan, sedang, dan berat), menginisialisasi centroid awal berdasarkan sampel data representatif, menghitung jarak setiap data pasien ke centroid menggunakan rumus *Euclidean Distance*, mengelompokkan data ke klaster dengan jarak terdekat, memperbarui centroid berdasarkan rata-rata anggota klaster, dan mengulangi proses hingga posisi centroid stabil.

Implementasi sistem dilakukan menggunakan *CodeIgniter* sebagai kerangka kerja PHP, *MySQL* sebagai basis data, *Bootstrap* untuk desain antarmuka, dan diuji pada server lokal menggunakan *XAMPP*. Pemodelan sistem dilakukan dengan *Unified Modeling Language (UML)* meliputi *use case diagram*, *class diagram*, dan *activity diagram* untuk memvisualisasikan kebutuhan fungsional dan alur kerja sistem. Sistem yang dibangun memungkinkan pengguna untuk memasukkan data pasien, menjalankan proses klasterisasi, dan menampilkan hasil pengelompokan dalam bentuk tabel dan grafik yang mudah dipahami. Validasi hasil dilakukan dengan membandingkan output klasterisasi dari sistem dengan hasil yang diperoleh melalui *RapidMiner*, sehingga dapat memastikan bahwa algoritma diimplementasikan dengan benar dan menghasilkan pengelompokan yang akurat.

HASIL DAN PEMBAHASAN

Analisa Proses



Tabel 1. Hasil Perhitungan K-Means

| NO | NAMA | JARAK C1 | JARAK C2 | JARAK C3 | CLUSTER |
|----|---------------|----------|----------|----------|---------|
| 1 | TINA | 12.24 | 21.36 | 94.57 | C1 |
| 2 | NURYETI | 157.1 | 119.31 | 0 | C3 |
| 3 | SYAFRIZAL | 0 | 40.75 | 125.11 | C1 |
| 4 | NETTY MARLINA | 40.75 | 0 | 87.47 | C2 |
| 5 | ALI YUSNA | 10.25 | 28.51 | 115.76 | C1 |
| 6 | AZWAR | 55.39 | 17.57 | 69.93 | C2 |
| 7 | SUDIRMAN | 6.84 | 31.11 | 116.71 | C1 |
| 8 | MARYULIS | 55.39 | 17.57 | 69.93 | C2 |
| 9 | ZAINAB | 103.07 | 65.47 | 22.48 | C3 |
| 10 | KURNIAWATI | 10.25 | 28.51 | 115.76 | C1 |
| 11 | YUSMAWATI | 48.94 | 11.24 | 76.59 | C2 |
| 12 | SUWARLI | 15.18 | 24.22 | 111.45 | C1 |
| 13 | SAFREDI | 15.18 | 24.22 | 111.45 | C1 |
| 14 | ARMAN | 48.92 | 11.22 | 76.6 | C2 |
| 15 | ARLAN SYAFRI | 48.94 | 11.24 | 76.59 | C2 |
| 16 | YASMA | 157.1 | 119.31 | 32.28 | C3 |
| 17 | AZWIRMAN | 103.07 | 65.47 | 22.48 | C3 |
| 18 | AYOM MASRIL | 32.29 | 5.75 | 93.15 | C2 |
| 19 | MARLINA | 6.23 | 32.08 | 119.17 | C1 |
| 20 | ASMAWARNI | 42.62 | 5 | 83.23 | C2 |
| 21 | ALI AMRA | 125.11 | 87.47 | 0 | C3 |
| 22 | ENDRIDAWATI | 55.39 | 17.57 | 69.93 | C2 |
| 23 | YUSMANIDAR | 10.25 | 28.51 | 115.76 | C1 |
| 24 | RAZALI | 15.18 | 24.22 | 111.45 | C1 |
| 25 | MARDAN | 125.11 | 87.47 | 0 | C3 |
| 26 | SAMINEM | 15.18 | 24.22 | 111.45 | C1 |
| 27 | MASNETI | 55.39 | 17.57 | 69.93 | C2 |
| 28 | BASRI | 10.25 | 28.51 | 115.76 | C1 |
| 29 | SUARNI | 15.18 | 24.22 | 111.45 | C1 |
| 30 | SYAFRINA | 5.39 | 33.95 | 121.22 | C1 |

Setelah menghitung jarak ke semua centroid, setiap pasien akan ditugaskan ke klaster yang centroid-nya memiliki jarak paling pendek dari pasien tersebut. Artinya, pasien tersebut paling "mirip" dengan centroid klaster tersebut.

Setelah seluruh pasien pada dataset ditugaskan ke klaster terdekat, sistem menghitung



ulang posisi centroid untuk setiap klaster dengan mengambil nilai rata-rata dari seluruh variabel klinis yang digunakan, yaitu HbA1c, GDP, TDS, TDD, dan IMT. Proses pembaruan ini menghasilkan koordinat baru untuk masing-masing centroid, di mana klaster 1 (kategori ringan) memiliki nilai (7.11, 160.57, 132.07, 83.57, 27.90), klaster 2 (kategori sedang) memiliki nilai (7.87, 186.7, 145.1, 92.0, 31.3), dan klaster 3 (kategori berat) memiliki nilai (9.27, 246.67, 177.5, 109.67, 35.68). Iterasi pembaruan ini terus dilakukan melalui perhitungan jarak, penugasan ulang anggota klaster, dan perhitungan ulang centroid hingga tercapai kondisi konvergensi. Kondisi tersebut tercapai apabila perubahan posisi centroid sangat kecil dibandingkan iterasi sebelumnya, tidak ada pasien yang berpindah klaster, atau jumlah iterasi maksimum yang telah ditetapkan tercapai. Begitu salah satu kondisi ini terpenuhi, algoritma *K-Means* berhenti dan setiap pasien secara final ditempatkan pada klaster yang sesuai (ringan, sedang, atau berat) berdasarkan karakteristik medisnya, sehingga hasil pengelompokan dapat langsung ditampilkan dan diinterpretasikan oleh sistem.

Implementasi Sistem

Tahap implementasi program mencakup pengujian berbagai modul yang telah dikembangkan. Antarmuka sistem terdiri dari beberapa halaman utama. Halaman Login digunakan untuk mengakses sistem dengan memasukkan username dan password yang valid. Setelah berhasil masuk, pengguna diarahkan ke Dashboard, yang menampilkan ringkasan jumlah pasien, pembagian berdasarkan jenis kelamin, jumlah variabel diagnosa, serta jumlah pasien yang telah didiagnosis. Selanjutnya terdapat halaman Kelola User, yang memungkinkan pengguna dengan hak akses tertentu untuk menambah, mengubah, atau menghapus data pengguna, serta Kelola Pasien, yang memiliki fungsi serupa tetapi untuk data pasien.

Selain itu, tersedia halaman Kelola Variabel Diagnosa, yang memfasilitasi pengelolaan variabel diagnostik, termasuk penentuan tipe data dan aturan numerik. Halaman Kelola Diagnosa digunakan untuk mengatur data diagnosa pasien berdasarkan variabel yang telah ditentukan. Inti dari sistem terdapat pada halaman Clustering, yang menampilkan hasil pengelompokan pasien menggunakan algoritma K-Means dalam bentuk tabel, serta halaman Visual Clustering yang menyajikan hasil tersebut dalam bentuk grafik kurva dua dimensi berdasarkan pasangan variabel yang dipilih. Seluruh komponen antarmuka ini dirancang untuk memudahkan pengguna dalam mengelola data, menjalankan proses klasterisasi, dan memahami hasil pengelompokan secara visual maupun numerik.

Proses Pengujian Model K-Means Menggunakan RapidMiner

Pengujian model *K-Means* menggunakan RapidMiner dilakukan untuk mengevaluasi efektivitas dan kualitas pengelompokan tingkat keparahan penyakit diabetes pada data pasien, sekaligus memperoleh konfigurasi klaster yang dapat dijadikan acuan perbandingan dengan hasil implementasi pada sistem berbasis website. Proses pengujian dimulai dengan membuat proyek baru pada RapidMiner melalui menu *File* kemudian memilih opsi *New Process*, dan dilanjutkan dengan memilih *Blank Process* sebagai kanvas kerja. Selanjutnya, operator *Read CSV* dicari dan ditambahkan untuk membaca data penelitian, kemudian file CSV yang berisi data pasien dimasukkan ke dalam operator tersebut.

Pada tahap *Format Your Columns*, seluruh kolom yang tidak relevan dengan proses klasterisasi dihapus, sehingga hanya tersisa lima variabel utama yaitu HbA1c, GDP, TDS, TDD, dan IMT. Setelah itu, operator *k-Means* ditambahkan dengan memilih *Segmentation k-Means* dari daftar operator. Parameter algoritma diatur dengan jumlah klaster (*k*) sebesar tiga dan *measure types* menggunakan *MixedMeasure* agar dapat menangani variasi tipe data. Operator *Read CSV* kemudian dihubungkan dengan operator *Clustering* untuk membentuk alur proses, dan pengujian dijalankan.

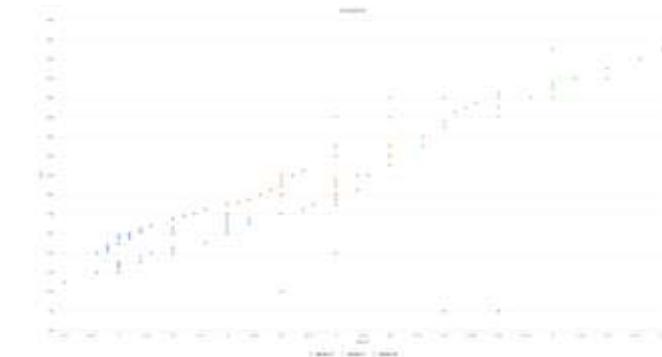
Hasil eksekusi proses menghasilkan *Cluster Model* dan *ExampleSet Clustering* yang



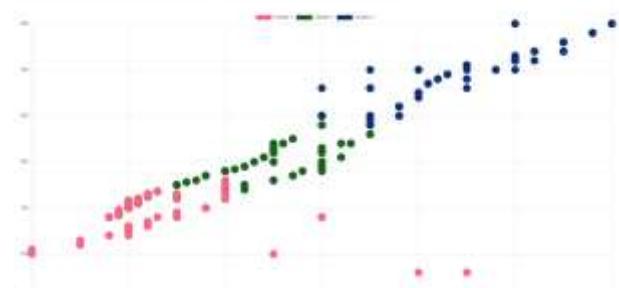
menampilkan pembagian data pasien ke dalam tiga klaster berdasarkan kesamaan karakteristik medisnya. Model ini menjadi acuan dalam menganalisis kinerja algoritma *K-Means* pada RapidMiner, serta sebagai tolok ukur untuk memvalidasi kesesuaian hasil klasterisasi yang dihasilkan oleh sistem berbasis website yang dikembangkan.

Perbandingan Hasil Pengujian Model K-Means Menggunakan RapidMiner dengan Sistem yang Dibangun

Secara spesifik, perbandingan akan dilakukan dengan membandingkan kurva XY dari metrik evaluasi klasterisasi yang dihasilkan oleh RapidMiner dengan kurva XY yang divisualisasikan oleh sistem yang dibangun, guna mengidentifikasi keselarasan dan perbedaan pola klasterisasi. Berikut perbandingan kurva XY hasil RapidMiner dengan visualisasi oleh sistem:

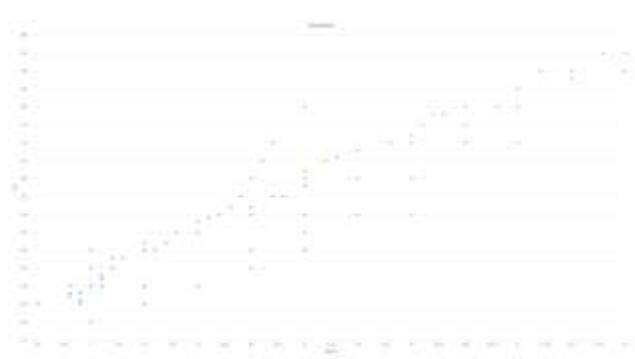


Gambar 1 Kurva HbA1c, GDP RapidMiner

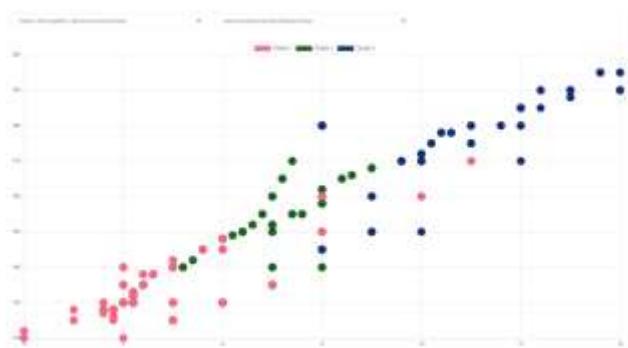


Gambar 2 Kurva HbA1c, GDP Sistem yang Dibangun

Kedua grafik secara konsisten menunjukkan pola pengelompokan data pasien ke dalam tiga klaster yang sangat serupa. Konsistensi ini merupakan indikator kuat terhadap validitas dan keandalan implementasi algoritma *K-Means* pada sistem website, merefleksikan kemampuan sistem untuk mereplikasi logika dan hasil klasterisasi dari perangkat lunak standar industri.

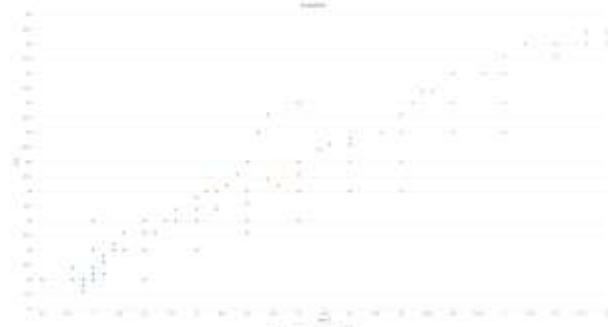


Gambar 3. Kurva HbA1c, TDS RapidMiner

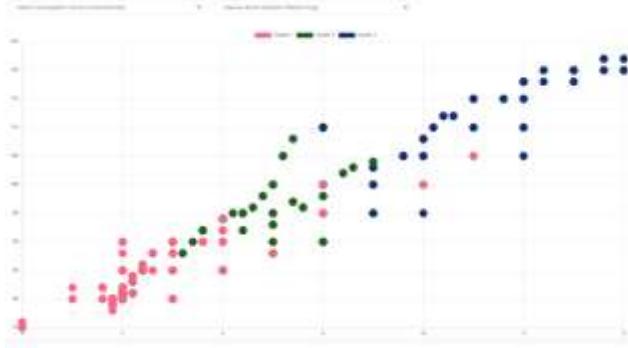


Gambar 4. Kurva HbA1c, TDS Sistem yang Dibangun

Kedua grafik secara konsisten menunjukkan pola pengelompokan data pasien ke dalam tiga klaster yang sangat serupa, sekali lagi menegaskan validitas dan keandalan implementasi algoritma K-Means pada sistem website yang dibangun. Kesamaan pola ini memvalidasi kemampuan sistem untuk mereplikasi logika dan hasil klasterisasi dari perangkat lunak analisis data standar.

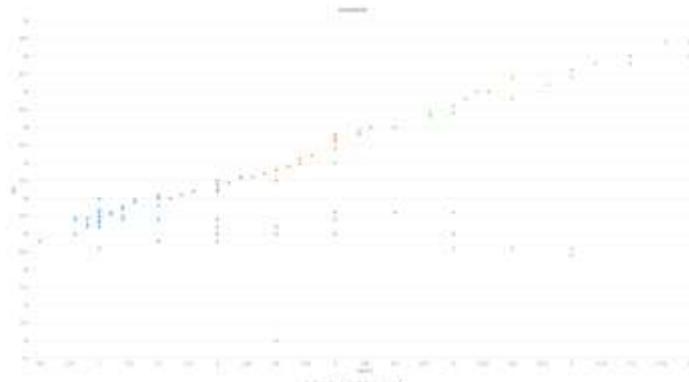


Gambar 5. Kurva HbA1c, TDD RapidMiner

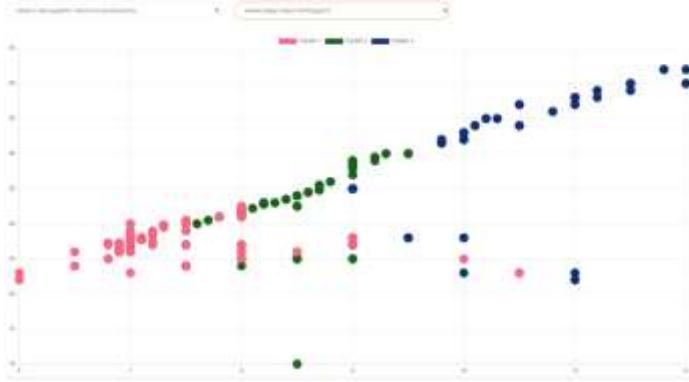


Gambar 6 Kurva HbA1c, TDD Sistem yang Dibangun

Kedua grafik secara konsisten menunjukkan pola pengelompokan data pasien ke dalam tiga klaster yang sangat serupa. Konsistensi ini menegaskan validitas dan keandalan implementasi algoritma K-Means pada sistem website Anda, membuktikan kemampuannya untuk mereplikasi logika dan hasil klasterisasi yang dihasilkan oleh perangkat lunak analisis data standar.

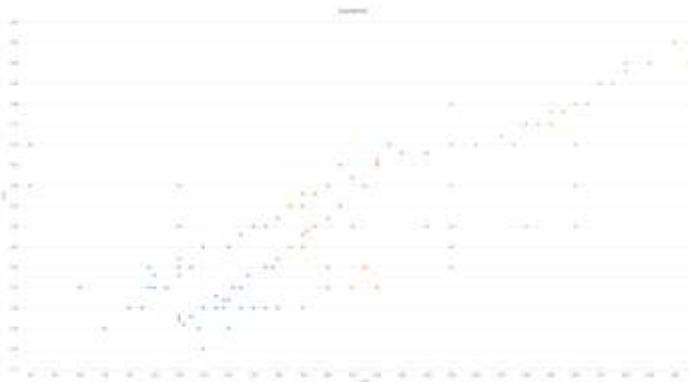


Gambar 7 Kurva HbA1c, IMT RapidMiner

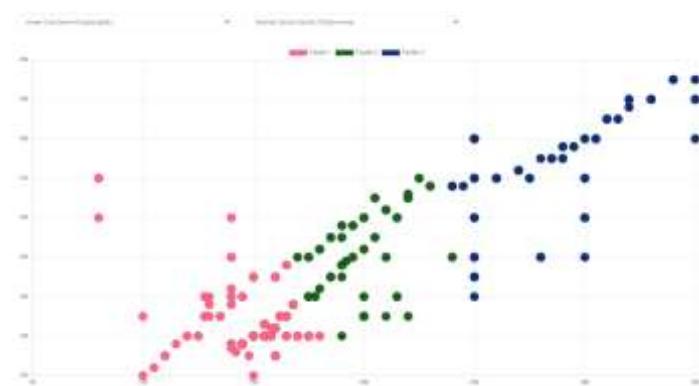


Gambar 8. Kurva HbA1c, IMT Sistem yang Dibangun

Kedua grafik secara konsisten menunjukkan pola pengelompokan data pasien ke dalam tiga klaster yang sangat serupa. Konsistensi ini memperkuat validitas dan keandalan implementasi algoritma K-Means pada sistem website, membuktikan kemampuannya untuk mereplikasi logika dan hasil klasterisasi yang dihasilkan oleh perangkat lunak analisis data standar.

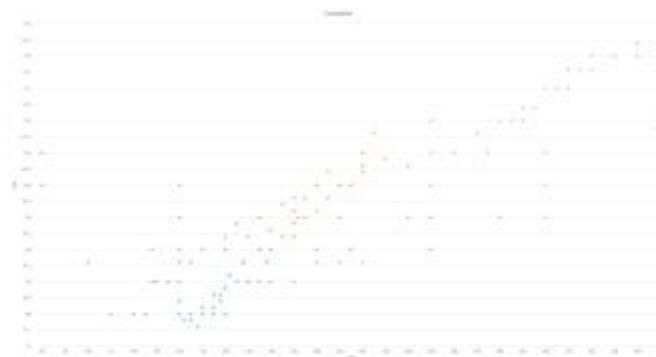


Gambar 9. Kurva GDP, TDS RapidMiner

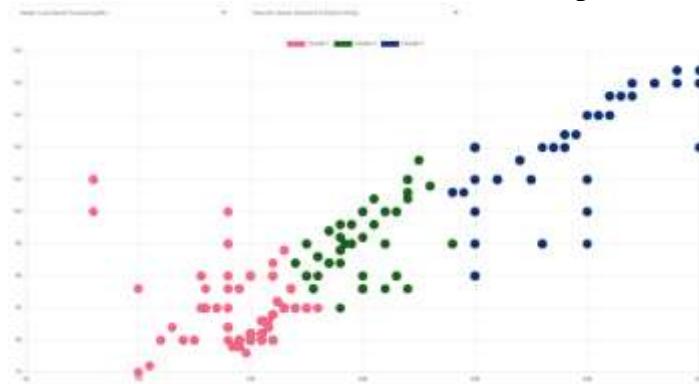


Gambar 10. Kurva GDP, TDS Sistem yang Dibangun

Kedua grafik secara konsisten menunjukkan pola pengelompokan data pasien ke dalam tiga klaster yang sangat serupa. Konsistensi ini menegaskan validitas dan keandalan implementasi algoritma K-Means pada sistem website, membuktikan kemampuannya untuk mereplikasi logika dan hasil klasterisasi yang dihasilkan oleh perangkat lunak analisis data standar.

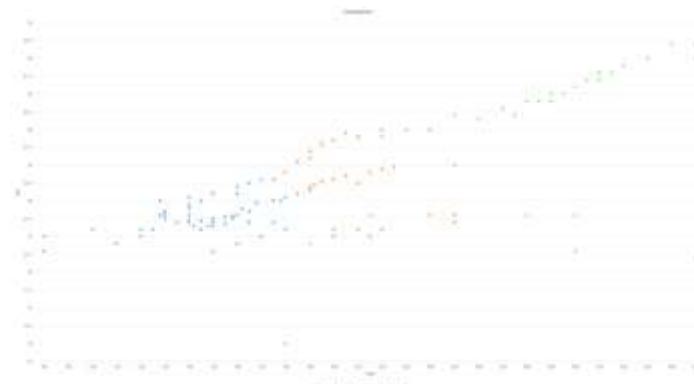


Gambar 11. Kurva GDP, TDD RapidMiner

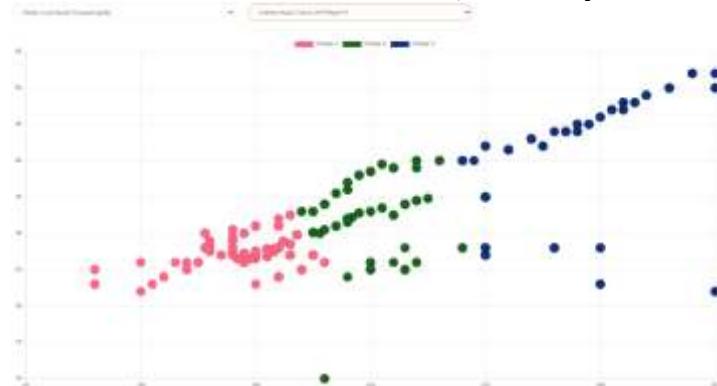


Gambar 12. Kurva GDP, TDD Sistem yang Dibangun

Terlihat adanya pola pengelompokan yang sangat serupa dan konsisten ke dalam tiga klaster pada kedua platform. Konsistensi ini menjadi bukti kuat akan validitas dan keandalan algoritma K-Means yang terimplementasi dalam sistem website, yang berhasil mereplikasi logika dan hasil klasterisasi dari perangkat lunak analisis data standar.

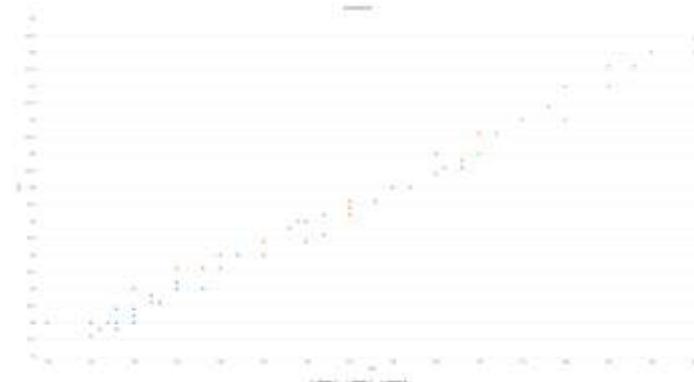


Gambar 13. Kurva GDP, IMT RapidMiner

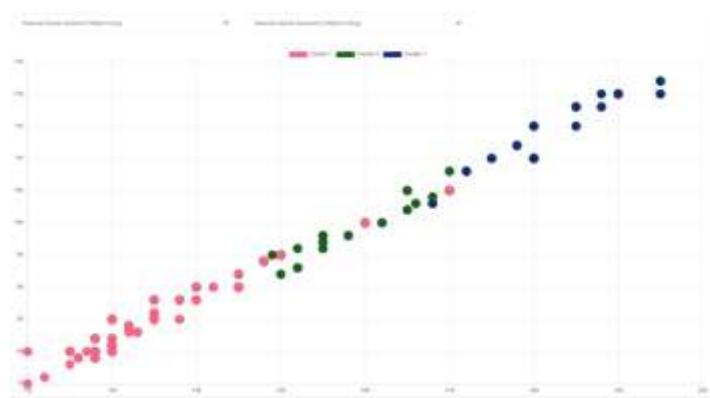


Gambar 14. Kurva GDP, IMT Sistem yang Dibangun

Terlihat adanya pola pengelompokan yang sangat serupa dan konsisten ke dalam tiga klaster pada kedua platform. Konsistensi ini menegaskan validitas dan keandalan implementasi algoritma K-Means pada sistem website Anda, membuktikan kemampuannya untuk mereplikasi logika dan hasil klasterisasi yang dihasilkan oleh perangkat lunak analisis data standar.

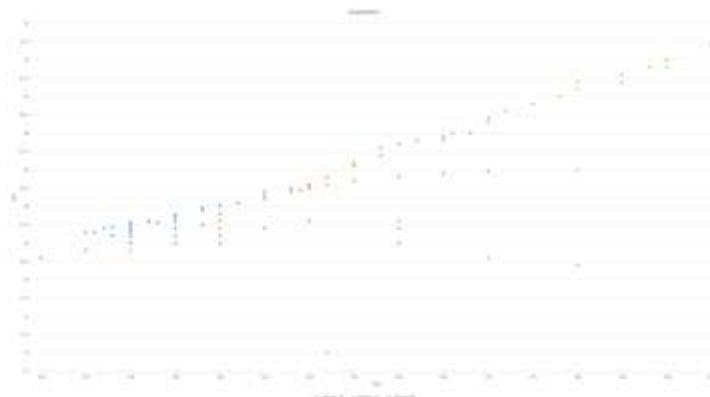


Gambar 15. Kurva TDS, TDD RapidMiner

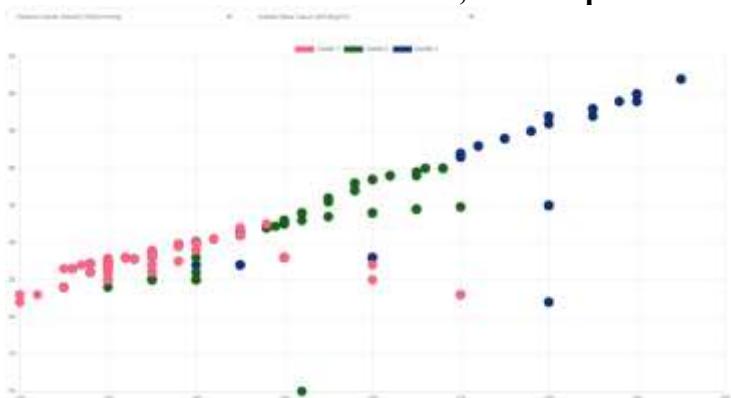


Gambar 16. Kurva TDS, TDD Sistem yang Dibangun

Terlihat adanya pola pengelompokan yang sangat serupa dan konsisten ke dalam tiga klaster pada kedua platform. Konsistensi ini menegaskan validitas dan keandalan implementasi algoritma K-Means pada sistem website penulis, membuktikan kemampuannya untuk mereplikasi logika dan hasil klasterisasi yang dihasilkan oleh perangkat lunak analisis data standar.

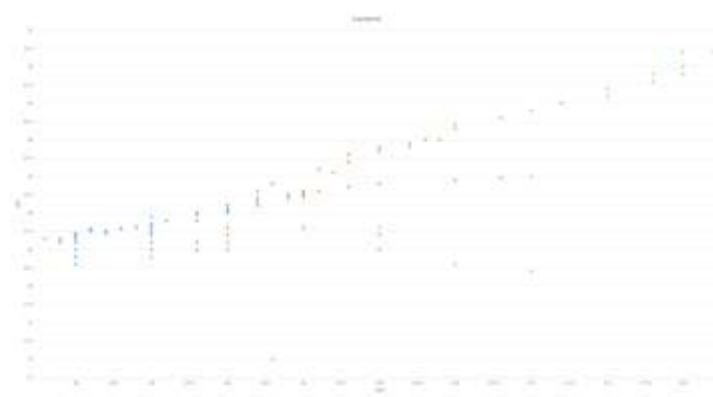


Gambar 17. Kurva TDS, IMT RapidMiner

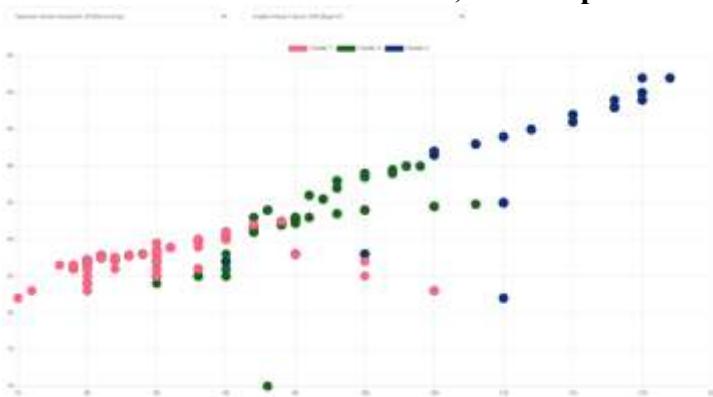


Gambar 18. Kurva TDS, IMT Sistem yang Dibangun

Terlihat adanya pola pengelompokan yang sangat serupa dan konsisten ke dalam tiga klaster pada kedua platform. Konsistensi ini menegaskan validitas dan keandalan implementasi algoritma K-Means pada sistem website penulis, membuktikan kemampuannya untuk mereplikasi logika dan hasil klasterisasi yang dihasilkan oleh perangkat lunak analisis data standar.



Gambar 19. Kurva TDD, IMT RapidMiner



Gambar 20. Kurva TDD, IMT Sistem yang Dibangun

Terlihat adanya pola pengelompokan yang sangat serupa dan konsisten ke dalam tiga klaster pada kedua platform. Konsistensi ini menegaskan validitas dan keandalan implementasi algoritma K-Means pada sistem website penulis, membuktikan kemampuannya untuk mereplikasi logika dan hasil klasterisasi yang dihasilkan oleh perangkat lunak analisis data standar.

Pembahasan

Berdasarkan serangkaian perbandingan visual melalui diagram scatter plot dari berbagai kombinasi variabel klinis (HbA1c, GDP, TDS, TDD, dan IMT) antara hasil klasterisasi yang diperoleh dari RapidMiner dan sistem berbasis website yang telah dibangun, dapat ditarik beberapa kesimpulan penting. Secara konsisten, penulis mengamati adanya pola pengelompokan data pasien ke dalam tiga klaster yang sangat serupa dan identik pada seluruh pasangan variabel yang diuji. Meskipun penamaan klaster mungkin bervariasi antara RapidMiner dan sistem, substansi pengelompokan data dan karakteristik setiap klaster (yaitu, kategori tingkat keparahan rendah, sedang, dan berat) tetap konsisten.

Keselarasan visual yang signifikan ini secara kuat memvalidasi akurasi dan konsistensi implementasi algoritma K-Means clustering pada sistem berbasis website penulis. Hal ini membuktikan bahwa logika dan proses klasterisasi yang tertanam dalam sistem tersebut mampu mereplikasi hasil yang divalidasi oleh RapidMiner sebagai perangkat lunak analisis data terkemuka. Dengan demikian, dapat disimpulkan bahwa sistem yang telah dibangun berhasil mengelompokkan pasien diabetes berdasarkan tingkat keparahan secara efektif dan akurat, menjadikan sistem ini andal dan sesuai untuk tujuan yang dirancang, yaitu membantu staf medis dalam mengidentifikasi dan mengelola pasien diabetes berdasarkan profil keparahan penyakit mereka.

KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma *K-Means Clustering* untuk mengelompokkan pasien diabetes di RSUD Pariaman berdasarkan tingkat keparahan penyakit, dengan memanfaatkan variabel klinis HbA1c, GDP, TDS, TDD, dan IMT. Proses klasterisasi menghasilkan tiga kategori utama, yaitu ringan, sedang, dan berat, yang dapat membantu tenaga medis dalam memprioritaskan penanganan, mengoptimalkan alokasi sumber daya, serta menyusun program edukasi pasien yang lebih terarah. Hasil pengujian menggunakan RapidMiner menunjukkan bahwa konfigurasi klaster yang dihasilkan memiliki konsistensi dengan implementasi pada sistem berbasis web, sehingga membuktikan akurasi dan keandalan algoritma yang diterapkan. Sistem berbasis website yang dibangun dengan framework *CodeIgniter* dan basis data *MySQL* mampu memproses data pasien secara *real-time*, menampilkan hasil klasterisasi dalam bentuk tabel maupun grafik, serta menyediakan antarmuka yang mudah digunakan oleh staf medis. Dengan demikian, sistem ini dapat berfungsi sebagai alat bantu pengambilan keputusan klinis yang praktis, cepat, dan terintegrasi dengan alur kerja rumah sakit. Ke depan, sistem ini dapat dikembangkan lebih lanjut dengan menambahkan variabel medis lain seperti riwayat komplikasi dan kadar kolesterol, mengintegrasikannya dengan Sistem Informasi Rumah Sakit (SIRS), serta menerapkan algoritma *clustering* atau *classification* lain untuk membandingkan performa dan meningkatkan akurasi hasil.

DAFTAR PUSTAKA

- Alpian, M., & Alfarizi, L. M. (2022). Diabetes Mellitus Tipe 2 (Dua) Dan Pengobatannya: Suatu Tinjauan Literatur. *Journal of Public Health and Medical Studies*, 1(1), 13–23. <https://scientium.co.id/journals/index.php/jphms/article/view/254>
- Asy Aria, T., Julkarnain, M., & Hamdani, F. (2023). Penerapan Algoritma K-Means Clustering Untuk Data Obat. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(1), 649–657. <https://doi.org/10.30865/klik.v4i1.1117>
- Harmaja, O. J., L., H. S., & S., H. (2023). Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Pulo Brayan. *Sains Dan Teknologi*, 5(1), 150–157.
- KHAIRANI, A. F. (2023). *KLASTERISASI PESERTA BPJS BERDASARKAN REKAM MEDIS MENGGUNAKAN ALGORITMA K-MEANS*.
- Kurniawan, H., Defit, S., & Sumijan. (2020). Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal. *Journal of Applied Computer Science and Technology*, 1(2), 80–89. <https://doi.org/10.52158/jacost.v1i2.102>
- Nahjan, M. R., Heryana, N., Voutama, A., Komputer, F. I., Karawang, U. S., & Miner, R. (2023). *IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL*. 7(1), 101–104.
- Sari, H. L., & Beti, I. Y. (2023). Penerapan Data Mining Dalam Pengelompokan Buku Yang Dipinjam Menggunakan Algoritma K-Means. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 3(6), 925–933. <https://doi.org/10.30865/klik.v3i6.826>

