



Analisis Kluster Pasien Diabetes Menggunakan Algoritma K-Means Berdasarkan Usia, Kadar Glukosa, dan Tekanan Darah

Muhammad Rizqi Ramadhani¹, Rezan Naufal Hermawan², Ihsan Fajrian³, Daffa Aulia Rachmat⁴, Sumanto⁵,
Andi Diah Kuswanto⁶

^{1,2,3,4,5,6} Informatika, Universitas Bina Sarana Informatika

[1m.rizqiramadhan0896@gmail.com](mailto:m.rizqiramadhan0896@gmail.com), [2rezanhermawan11@gmail.com](mailto:rezanhermawan11@gmail.com), [3ihsanfajrian03@gmail.com](mailto:ihsanfajrian03@gmail.com),

[4daffaauliyarachmat12@gmail.com](mailto:daffaauliyarachmat12@gmail.com), [5sumanto@bsi.ac.id](mailto:sumanto@bsi.ac.id), [6andi.ahk@bsi.ac.id](mailto:andi.ahk@bsi.ac.id)

Abstract

Diabetes mellitus is a chronic disease caused by a disorder in the production or utilization of insulin, resulting in high blood sugar levels. This disease can lead to serious complications, such as heart, kidney, and nerve damage. The number of diabetes sufferers continues to rise, including in Indonesia, influenced by factors such as genetics, unhealthy lifestyles, and poor diets. To detect the risk of diabetes earlier, data mining technology can be utilized. This research uses the K-Means Clustering algorithm to analyze health data such as blood glucose levels, blood pressure, and age. The algorithm groups individuals into several clusters based on similarities in health characteristics to identify high-risk groups for diabetes. The analysis revealed that blood glucose levels were the most significant factor influencing the likelihood of diabetes, followed by age and blood pressure. The classification of patients using K-Means achieved an accuracy of approximately 85%, which is a substantial improvement over traditional methods. These results can help healthcare professionals better identify at-risk individuals and develop more targeted prevention strategies, such as personalized diet plans and lifestyle changes. Furthermore, the findings can be applied in hospitals or clinics to assist doctors in early detection and management of diabetes. By integrating this clustering approach into healthcare systems, it is possible to proactively address the growing diabetes epidemic, particularly in regions like Indonesia, where the prevalence is high. This approach provides an efficient solution for managing big data in health care and supports efforts to systematically combat diabetes.

Keywords: *Diabetes Mellitus, Data Mining, K-Means Clustering, Health Risks, Blood Glucose, Early Detection, Data Analysis.*

Abstrak

Diabetes mellitus adalah penyakit kronis yang disebabkan oleh gangguan dalam produksi atau pemanfaatan insulin, yang mengakibatkan tingginya kadar gula darah. Penyakit ini dapat menyebabkan komplikasi serius, seperti kerusakan jantung, ginjal, dan saraf. Jumlah penderita diabetes terus meningkat, termasuk di Indonesia, yang dipengaruhi oleh faktor-faktor seperti genetika, gaya hidup tidak sehat, dan pola makan buruk. Untuk mendeteksi risiko diabetes lebih dini, teknologi data mining dapat dimanfaatkan. Penelitian ini menggunakan algoritma K-Means Clustering untuk menganalisis data kesehatan seperti kadar glukosa darah, tekanan darah, dan usia. Algoritma ini mengelompokkan individu ke dalam beberapa kluster berdasarkan kesamaan karakteristik kesehatan untuk mengidentifikasi kelompok yang berisiko tinggi terhadap diabetes. Analisis ini mengungkapkan bahwa kadar glukosa darah merupakan faktor yang paling signifikan mempengaruhi kemungkinan terkena diabetes, diikuti oleh usia dan tekanan darah. Klasifikasi pasien menggunakan K-Means mencapai akurasi sekitar 85%, yang merupakan peningkatan signifikan dibandingkan dengan metode tradisional. Hasil ini dapat membantu profesional kesehatan untuk lebih baik dalam mengidentifikasi individu yang berisiko dan mengembangkan strategi pencegahan yang lebih tepat sasaran, seperti rencana diet yang dipersonalisasi dan perubahan gaya hidup. Selain itu, temuan ini dapat diterapkan di rumah sakit atau klinik untuk membantu dokter dalam deteksi dini dan manajemen diabetes. Dengan mengintegrasikan pendekatan klusterisasi ini ke dalam sistem perawatan kesehatan, kita dapat secara proaktif menangani epidemi diabetes yang terus berkembang, terutama di daerah seperti Indonesia, di mana prevalensinya tinggi. Pendekatan ini menyediakan solusi yang efisien untuk mengelola data besar di bidang kesehatan dan mendukung upaya untuk memerangi diabetes secara sistematis.

Kata kunci: *Diabetes Melitus, Data Mining, K-Means Clustering, Risiko Kesehatan, Glukosa Darah, Deteksi Dini, Analisis Data.*



1. Pendahuluan

Pada tahun 2015, International Diabetes Federation melaporkan bahwa jumlah penderita Diabetes Mellitus (DM) mencapai 415 juta orang, dengan 98% di antaranya mengidap DM Tipe 2. Hal ini menunjukkan bahwa masyarakat lebih rentan terkena DM tipe 2. Diabetes Mellitus sendiri adalah suatu kondisi yang ditandai oleh hiperglikemia, yaitu tingginya kadar glukosa dalam darah, serta gangguan pada proses metabolisme akibat kekurangan insulin. [1]

Diabetes Mellitus dibagi menjadi dua tipe menurut klasifikasinya. DM Tipe 1 adalah jenis diabetes yang ditandai dengan kadar insulin yang berada di bawah tingkat normal. Sementara itu, DM Tipe 2 terjadi karena tubuh gagal memanfaatkan insulin secara efektif, yang mengarah pada peningkatan berat badan dan penurunan aktivitas fisik. Berbeda dengan itu, diabetes gestasional adalah jenis diabetes yang pertama kali terdeteksi selama masa kehamilan dan disebut sebagai hiperglikemia [2] Ruis menyatakan bahwa berbagai komplikasi dapat muncul akibat Diabetes Mellitus yang tidak dikelola dengan baik. Selain itu, DM juga merupakan salah satu faktor penyebab Gangguan Fungsi Kognitif (GFK) [3].

Menurut World Health Organization (WHO), pada tahun 2030 diperkirakan jumlah penderita Diabetes Mellitus akan meningkat menjadi minimal 366 juta orang. Selain itu, hasil survei WHO menunjukkan bahwa Indonesia termasuk dalam empat negara dengan jumlah penduduk terbanyak yang menderita DM, bersama dengan China, Amerika Serikat, dan India [4]

Menurut Kemenkes, pencegahan Diabetes Mellitus Tipe 2 (DMT2) dapat dilakukan dengan mengetahui faktor risiko yang ada. Ada dua jenis faktor risiko DMT2, yaitu faktor risiko yang dapat diubah dan faktor risiko yang tidak dapat diubah. Faktor yang dapat diubah meliputi gaya hidup, seperti pola makan, kualitas istirahat, tingkat aktivitas fisik, dan manajemen stres. Sementara itu, faktor yang tidak dapat diubah antara lain usia dan faktor genetik [2]

Tujuan dari pengkajian ini adalah untuk mengetahui, memahami, dan mengidentifikasi faktor risiko yang berhubungan dengan kejadian penyakit Diabetes Mellitus (DM). Mengetahui faktor risiko yang dapat meningkatkan kasus DM sangat penting, karena dengan pemahaman tentang faktor-faktor tersebut, pencegahan terhadap penyakit ini dapat dilakukan dengan lebih efektif.

Data mining adalah proses untuk mengekstrak informasi penting dari suatu data. Proses ini melibatkan teknik-teknik kompleks seperti kecerdasan buatan (artificial intelligence), teknik statistik, matematika, machine learning, dan lainnya. Teknik-teknik tersebut

digunakan untuk mengidentifikasi dan mengekstrak informasi yang berguna dari database besar. Data mining telah berkembang menjadi disiplin ilmu yang berada dalam domain kecerdasan buatan (AI) dan rekayasa pengetahuan (KE). Meskipun berakar pada machine learning dan statistika, data mining juga meluas ke berbagai bidang lain, seperti biologi, lingkungan, keuangan, jaringan, dan sebagainya [4].

Menurut Sucipto, klasterisasi dalam data mining adalah suatu teknik yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok atau klaster yang memiliki karakteristik yang mirip satu sama lain. Tujuan dari klasterisasi adalah untuk menemukan struktur atau pola yang tersembunyi dalam data tanpa adanya label atau informasi sebelumnya mengenai kategori data tersebut. Dengan teknik ini, data yang serupa akan dikelompokkan bersama, sementara data yang berbeda akan berada dalam kelompok yang terpisah [5] Klasterisasi adalah suatu teknik untuk mengelompokkan data ke dalam suatu klaster tertentu, yang memungkinkan data dalam klaster tersebut memiliki kesamaan satu sama lain dan memiliki perbedaan yang jelas dengan data pada klaster lainnya. Sedangkan menurut Marzuki, klasterisasi adalah proses pengelompokan data berdasarkan kesamaan atribut atau karakteristik yang dimiliki, sehingga data yang berada dalam satu klaster memiliki kemiripan yang tinggi dan berbeda dengan data di klaster lainnya [6] Marzuki berpendapat bahwa klasterisasi adalah suatu kumpulan objek atau data yang memiliki kesamaan di antara mereka, sementara data yang tidak memiliki kesamaan dimasukkan ke dalam klaster yang berbeda. Klasterisasi merupakan proses pengelompokan objek atau data ke dalam grup yang anggotanya memiliki kesamaan tertentu. Berdasarkan pengertian ini, dapat disimpulkan bahwa klasterisasi adalah metode untuk mengelompokkan data berdasarkan kesamaan tertentu.

Hal ini berbeda dengan klasifikasi, yang merupakan proses pengelompokan data baru berdasarkan kelompok atau klasifikasi yang sudah ada. Sedangkan dalam klasterisasi, data baru akan dikelompokkan berdasarkan atribut dengan karakteristik yang sama. Oleh karena itu, klasterisasi lebih cocok digunakan dalam penelitian yang memerlukan pengelompokan data berdasarkan pola atau kesamaan yang ditemukan dalam data itu sendiri [7].

Algoritma K-Means adalah metode yang digunakan untuk mengelompokkan objek atau data berdasarkan karakteristik yang diukur, dengan tujuan agar objek-objek yang memiliki kemiripan satu sama lain dikelompokkan dalam klaster yang sama. K-Means merupakan teknik clustering yang termasuk dalam kategori unsupervised learning, yang berarti tidak

memerlukan fase pelatihan (learning) atau pelabelan data pada setiap kelompok.

Metode clustering ini mempartisi data ke dalam sejumlah kelompok atau klaster, sehingga data yang memiliki karakteristik yang serupa akan dikelompokkan dalam satu klaster yang sama. Dengan kata lain, K-Means mencari pola atau struktur dalam data tanpa informasi sebelumnya mengenai kategori atau label yang harus diterapkan pada data tersebut [8].

2. Metode Penelitian

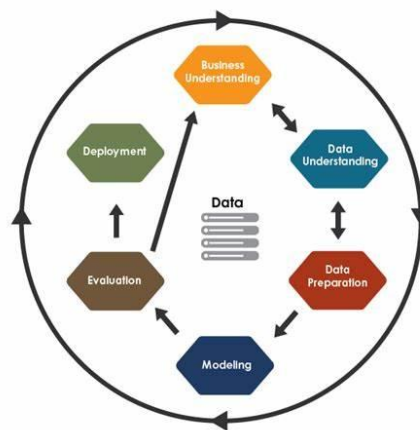
2.1 CRISP-DM

CRISP-DM (Cross Industry Standard Process for Data Mining) adalah metodologi yang digunakan untuk mengolah masalah dalam data mining dengan menerapkan enam tahapan utama. Tahapan-tahapan tersebut adalah sebagai berikut:

1. **Business Understanding:** Tahap pertama ini bertujuan untuk memahami tujuan bisnis atau permasalahan yang ingin diselesaikan. Dalam tahap ini, fokusnya adalah untuk mengonversi tujuan bisnis menjadi tujuan data mining yang jelas dan terukur.
2. **Data Understanding:** Tahap ini melibatkan eksplorasi data untuk memahami lebih dalam tentang karakteristik data yang tersedia. Ini termasuk mengumpulkan data, melakukan analisis awal, dan mencari informasi yang relevan untuk proses selanjutnya.
3. **Data Preparation:** Tahap ini meliputi pemrosesan data agar siap digunakan untuk modeling. Langkah-langkah yang dilakukan termasuk pembersihan data, pengolahan data yang hilang, transformasi data, dan pemilihan data yang relevan.
4. **Modeling:** Pada tahap ini, berbagai teknik atau algoritma data mining diterapkan untuk membangun model yang dapat digunakan untuk menganalisis data dan membuat prediksi atau keputusan.
5. **Evaluation:** Setelah model dibangun, tahap ini bertujuan untuk mengevaluasi hasilnya untuk memastikan bahwa model yang dibuat benar-benar dapat menyelesaikan masalah yang diajukan di tahap Business Understanding. Evaluasi ini melibatkan pengujian kinerja model dan analisis apakah tujuan awal tercapai.
6. **Deployment:** Tahap terakhir adalah implementasi model yang telah dievaluasi dan diperbaiki. Ini bisa melibatkan penyebaran model ke dalam lingkungan bisnis atau aplikasi yang lebih luas agar bisa

digunakan dalam pengambilan keputusan sehari-hari.

Metode CRISP-DM digunakan secara luas karena fleksibilitasnya dan kemampuannya untuk diterapkan pada berbagai industri dan masalah data mining.



Gambar 1. Metode CRISP-DM

2.2 Pemahaman Bisnis (Business Understanding)

Menyusun tujuan dan kebutuhan proyek secara rinci dalam konteks bisnis, mengubah tujuan serta batasan yang ada menjadi perumusan masalah yang dapat diselesaikan dengan data mining, serta merancang strategi untuk mencapai tujuan yang telah ditetapkan.

2.3 Pemahaman Data (Data Understanding)

Data Understanding adalah tahap "berkenalan" dengan data sebelum menganalisisnya. Intinya kita melakukan tiga hal: 1. Mengumpulkan data dari berbagai sumber yang relevan, 2. Mengeksplorasi data untuk melihat pola, tren, dan karakteristiknya, 3. Mengecek kualitas data apakah layak pakai atau ada yang rusak/hilang. Tujuannya untuk memastikan kita benar-benar paham yang akan diolah, sehingga bisa memilih metode analisis yang tepat dan menghasilkan kesimpulan yang akurat.

2.4 Persiapan Data (Data Preparation)

Data Preparation adalah tahap mempersiapkan dan membersihkan data agar siap digunakan untuk analisis atau pemodelan. Dalam tahap ini, kita melakukan beberapa aktivitas penting yaitu: 1. Pembersihan Data Menghilangkan atau memperbaiki data yang rusak, duplikat, atau tidak konsisten. 2. Transformasi Data Mengubah format data sesuai dengan kebutuhan analisis. 3. Integrasi Data Menggabungkan data dari berbagai sumber menjadi satu dataset yang utuh. 4. Seleksi Data Memilih kolom atau baris data yang relevan dengan tujuan analisis yang tidak perlu.

2.5 Pemodelan (Modeling)

Tahap di mana kita membangun model atau algoritma untuk menyelesaikan masalah bisnis yang sudah diidentifikasi sebelumnya. Seperti 1. Memilih dan implementasi model. 2. Proses pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan kebutuhan teknik data mining.

2.6 Evaluation (Evaluasi)

Tahap ini bertujuan untuk menilai apakah model yang telah dibuat efektif dan dapat menyelesaikan masalah bisnis yang diinginkan, yaitu:

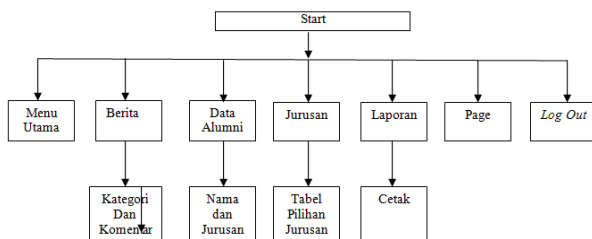
1. Mengevaluasi satu atau lebih model yang dikembangkan pada fase pemodelan untuk menilai kualitas dan efektivitasnya sebelum dipublikasikan.
2. Menentukan apakah model tersebut memenuhi tujuan yang telah ditetapkan.
3. Menilai apakah masalah atau penelitian yang ada telah ditangani dengan baik oleh model.
4. Mengambil keputusan terkait penggunaan hasil dari data mining dalam konteks bisnis.

2.7 Deployment (Penerapan Hasil)

Adalah tahap terakhir di mana kita menempatkan model yang sudah jadi ke dalam sistem operasional perusahaan agar bisa digunakan secara nyata untuk membantu bisnis sehari-hari, seperti menggunakan model yang dihasilkan, penyebaran pembuatan laporan sederhana.

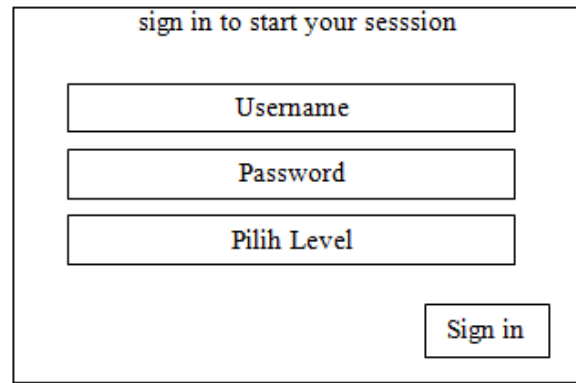
3. Hasil dan Pembahasan

Berikut adalah perancangan struktu *menu* dari sistem yang dirancang agar memudahkan pada tahap implementasi aplikasi pendataan alumni SMK Negeri 1 Tandun. Struktur dapat dilihat sebagai berikut:



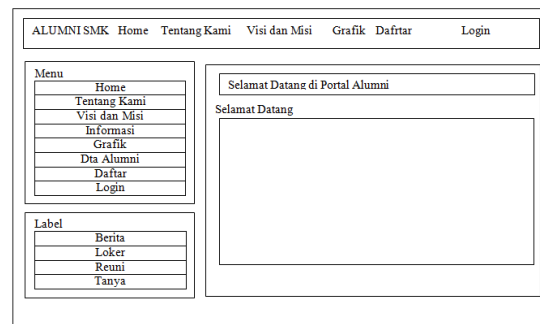
Gambar 4. Struktur Menu Sistem

Untuk masuk kedalam halaman *administrator* atau *user*, pertama kali akan ditampilkan halaman *login* untuk memasukkan *username* dan *password*.



Gambar 5. Desain Login

Desain menu pada halaman *user* yaitu menu halaman utama, tentang sekolah, visi misi, informasi, grafik, daftar dan *login*. Berikut ini adalah rancangan desainnya:



Gambar 6. Desain Halaman Utama User

4. Kesimpulan

Penelitian ini berhasil mengungkap bahwa K-Means adalah alat yang sangat berguna untuk mengklasifikasikan pasien diabetes berdasarkan tingkat risiko mereka. Dengan memanfaatkan informasi seperti usia, kadar glukosa dalam darah, dan tekanan darah, pasien dapat dibagi menjadi tiga kelompok: risiko rendah, risiko sedang, dan risiko tinggi. Data menunjukkan bahwa kadar glukosa dalam darah adalah faktor utama yang mempengaruhi kemungkinan terkena diabetes, diikuti oleh usia serta tekanan darah. Pasien yang memiliki kadar glukosa tinggi, usia yang lebih tua, dan tekanan darah tinggi termasuk dalam kategori risiko tinggi dan memerlukan perhatian medis yang lebih segera.

Pendekatan K-Means terbukti sangat efisien dan dapat diterapkan oleh profesional kesehatan sebagai alat untuk mendeteksi risiko diabetes secara lebih awal. Dengan demikian, dokter atau klinik dapat memberikan rekomendasi atau langkah pencegahan yang lebih tepat seperti perubahan pola makan atau penerapan gaya hidup sehat. Metode K-Means sebaiknya dimanfaatkan oleh fasilitas kesehatan untuk mendeteksi risiko diabetes lebih awal. Penambahan variabel seperti pola

makan dan aktivitas fisik akan meningkatkan akurasi. Edukasi kepada masyarakat dan integrasi hasil ke sistem digital kesehatan juga penting, serta perlu dilakukan evaluasi rutin agar tetap akurat.

Daftar Rujukan

- [1] Haerani, E. (2023). Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means (Hasil Check Similarity).
- [2] Anggraini, R. (2022). Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means. UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU, 9(6), 1840-1849.
- [3] L. Rahmawati, S. Widya Siwi, and E. Suryani, "Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret)," *J. Teknol. Inf. ITSmart* vol. 3, no. 2, p. 66, 2016, doi: 10.20961/its.v3i2.654.
- [4] M. A. Sembiring, "Penerapan Metode Algoritma K - Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (Dbd)," *J. Sci. Soc. Res.* vol. 4, no. 3, p. 336, 2021, doi: 10.54314/jssr.v4i3.712.
- [5] Praja, A., Lubis, C., & Herwindiati, D. E. (2017). Deteksi Penyakit Diabetes Dengan Metode Fuzzy C-Means Clustering Dan K-Means Clustering. *Computatio*, 1(1), 223887.
- [6] Solehah, H. F. (2021). ANALISIS DAN PERBANDINGAN KELOMPOK PASIEN COVID-19 BERDASARKAN KOMORBIDITAS MENGGUNAKAN K-MEANS CLUSTERING (Doctoral dissertation, Universitas Pembangunan Nasional Veteran Jakarta).
- [7] Prasatya, A., Siregar, R. R. A., & Arianto, R. (2020). Penerapan Metode K-Means dan C4. 5 Untuk prediksi penderita diabetes.
- [8] Gustiane, I. T., Martanto, M., & Suprapti, T. (2024). Clustering Hasil Cek Darah Diabetes Lansia Menggunakan Metode K-Means di Posbindu Kp. Lebakjero Desa Ciharang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 2125-2129.
- [9] Yosmar, R., Almasty, D., & Rahma, F. (2018). Survei risiko penyakit diabetes melitus terhadap masyarakat Kota Padang. *Jurnal sains farmasi & klinis*, 5(2).
- [10] Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE-Journal of Business and Audit Information Systems*, 4(1).
- [11] Nasution, F., Andilala, A., & Siregar, A. A. (2021). Faktor risiko kejadian diabetes mellitus. *Jurnal Ilmu Kesehatan*, 9(2), 94-102.
- [12] Lathifah, N. L. (2017). Hubungan durasi penyakit dan kadar gula darah dengan keluhan subyektif penderita diabetes melitus. *Jurnal berkala epidemiologi*, 5(2), 231-239.
- [13] Wiratama, M. A., & Pradnya, W. M. (2022). Optimasi algoritma data mining menggunakan backward elimination untuk klasifikasi penyakit diabetes. *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, 11(1), 1-12.
- [14] Mawarni, A. C., Rusdah, R., Hin, L. L., & Anubhakti, D. (2023). DETEKSI DINI GEJALA AWAL PENYAKIT DIABETES MENGGUNAKAN ALGORITMA RANDOM FOREST. *IDEALIS: InDonEsiA journal Information System*, 6(2), 165-171.
- [15] Chandra, K., & Prasetyo, J. S. (2024, September). Prediksi Penyakit Jantung Koroner Menggunakan Metode K-NN dan Regresi Logistik Berdasarkan Kerangka Kerja CRISP-DM. In *Prosiding Seminar Nasional Universitas Ma Chung (Informatika & Sistem Informasi; Bahasa dan Seni; Farmasi)* (Vol. 4, pp. 241-248).