Jurnal Teknik Informatika dan Teknologi Informasi
Volume. 5 Nomor. 3 Desember 2025
E-ISSN .: 2827-9387, P-ISSN .: 2827-9379, Hal 247-260
DOI: https://doi.org/10.55606/jutiti.v5i3.6281
Tersedia: https://journalshub.org/index.php/jutiti

Klastering Penyakit Diabetes Melitus dengan Algoritma K-Means berdasarkan Karakteristik Klinis

Audy Aulia Azzahra[1*], Fajar Yoga Adiansyah [2], Erlangga Rizki Ekaptra [3], Sumanto[4], Imam Budiawan[5], Roida Pakpahan[6]
[1-6]Program Studi Informatika, Fakultas Teknik dan Informatika , Universitas Bina Sarana Informatika, Indonesia
**Penulis Korespondensi: audyauliaaz@email.com*

Abstract. *Diabetes Mellitus is a complex and progressive chronic metabolic disorder that requires a personalized management strategy tailored to each individual's clinical, physiological, and lifestyle characteristics. Addressing this challenge, the present study aims to apply the K-Means algorithm to identify clustering patterns among diabetic patients using the Knowledge Discovery in Databases (KDD) framework. The dataset was obtained from the Kaggle repository, consisting of 769 patient medical records with key variables such as glucose levels, body mass index (BMI), blood pressure, age, and other metabolic parameters relevant to the diagnosis of Diabetes Mellitus. The research methodology includes several stages: data selection, preprocessing to handle missing values, duplication, and normalization to ensure the dataset is properly structured for analysis. The implementation of the K-Means algorithm was carried out using Orange Data Mining software to produce optimal clustering patterns. The analysis identified three primary clusters (C1, C2, C3) that demonstrated significant differences, particularly based on glucose levels as the dominant variable in cluster formation. The scatter plot visualization revealed clear separations among clusters, with high intra-cluster homogeneity and strong inter-cluster heterogeneity. These findings confirm the effectiveness of the K-Means algorithm as an unsupervised learning method capable of uncovering hidden patterns within clinical diabetes data. The results are expected to serve as a foundation for developing more adaptive and precise clinical decision support systems, assisting healthcare professionals in designing targeted management and intervention strategies aligned with each patient's risk profile.*

Keywords: *Clinical Decision Support; Clustering; Data Mining; Diabetes Mellitus; K-Means.*

Abstrak: Diabetes Melitus merupakan gangguan metabolik kronis yang kompleks dan bersifat progresif, yang menuntut strategi manajemen terpersonalisasi berdasarkan karakteristik klinis, fisiologis, dan gaya hidup masing-masing individu. Menjawab tantangan tersebut, penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam mengidentifikasi pola pengelompokan pasien diabetes melalui kerangka kerja *Knowledge Discovery in Databases* (KDD). Dataset yang digunakan bersumber dari repositori Kaggle, terdiri atas 769 catatan medis pasien dengan variabel utama seperti kadar glukosa, indeks massa tubuh (BMI), tekanan darah, usia, dan berbagai parameter metabolik lain yang relevan terhadap diagnosis Diabetes Melitus. Metodologi penelitian meliputi tahapan seleksi data, praproses untuk mengatasi *missing values*, duplikasi, serta normalisasi agar data berada dalam format yang siap dianalisis. Implementasi algoritma K-Means dilakukan menggunakan perangkat lunak *Orange Data Mining* untuk menghasilkan pola klaster yang optimal. Hasil analisis menghasilkan tiga kelompok utama (C1, C2, C3) yang menunjukkan perbedaan signifikan, terutama berdasarkan kadar glukosa sebagai variabel dominan dalam pembentukan klaster. Visualisasi *scatter plot* memperlihatkan pemisahan antar-klaster yang jelas, dengan tingkat homogenitas tinggi di dalam klaster dan heterogenitas kuat antar-klaster. Temuan ini menegaskan efektivitas algoritma K-Means sebagai metode *unsupervised learning* yang mampu mengungkap pola tersembunyi dalam data klinis pasien diabetes. Hasil penelitian ini diharapkan menjadi dasar pengembangan sistem pendukung keputusan klinis yang lebih adaptif dan presisi dalam membantu tenaga medis menentukan strategi pengelolaan dan intervensi yang sesuai dengan profil risiko tiap pasien.

Kata kunci: Dukungan Keputusan Klinis; Pengelompokan; Penambangan Data; Diabetes Melitus; K-Means.

1. LATAR BELAKANG

Diabetes Melitus adalah gangguan metabolik kronis dengan angka tingkat pengidap yang terus meningkat secara global, ditandai oleh kondisi hiperglikemia yang berlangsung secara persisten akibat gangguan sekresi insulin, resistensi insulin, atau kombinasi dari keduanya. Penyakit ini berkontribusi terhadap munculnya komplikasi makrovaskular (seperti

Naskah Masuk: 16 Juni 2025; Revisi: 05 Agustus 2025; Diterima: 23 September 2025; Tersedia: 27 November 2025.

kardiovaskular dan serebrovaskular) serta mikrovaskular (meliputi retinopati, nefropati, dan neuropati), yang secara signifikan meningkatkan tingkat morbiditas dan mortalitas.(Akbar Kombat Ginting, 2021) Pengelolaan DM yang optimal memerlukan pendekatan yang terpersonalisasi dengan mempertimbangkan profil risiko dan karakteristik klinis pasien yang beragam.(Cahyani & Basuki, 2019)

Dalam konteks tersebut, metode data mining, khususnya teknik *clustering*, menawarkan pendekatan *unsupervised learning* untuk mengungkap pola tersembunyi dan mengidentifikasi subkelompok pasien dengan kemiripan atribut klinis tanpa memerlukan label awal. Algoritma *K-Means*, yang merupakan salah satu metode *partitional clustering* paling luas digunakan, menunjukkan efektivitas tinggi dalam mengelompokkan data numerik berdimensi tinggi, seperti parameter klinis pasien diabetes (misalnya kadar glukosa, indeks massa tubuh, tekanan darah, dan profil lipid).(Method, 2024)

Penelitian ini bertujuan untuk menerapkan algoritma *K-Means* pada dataset karakteristik klinis pasien diabetes guna mengidentifikasi klaster pasien yang memiliki homogenitas tinggi di dalam klaster dan heterogenitas antar-klaster. Hasil dari proses klasterisasi diharapkan dapat memberikan pemahaman yang lebih mendalam mengenai stratifikasi pasien berdasarkan profil klinisnya, sehingga dapat mendukung perancangan strategi intervensi dan manajemen penyakit yang lebih tepat sasaran.(Basyir2025,)

2. KAJIAN TEORITIS

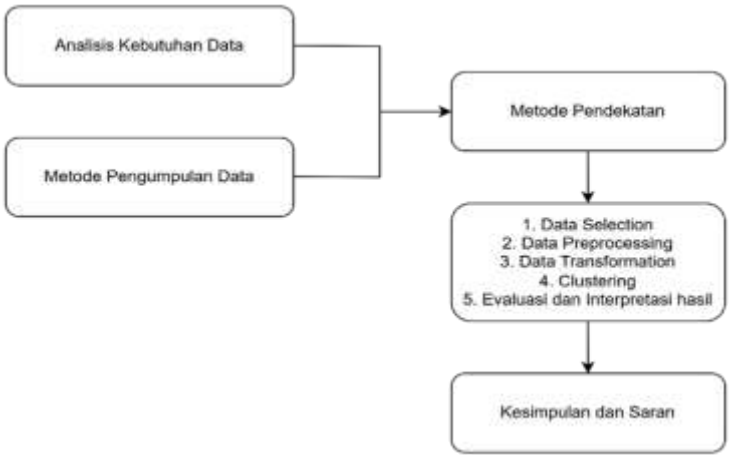
Data mining merupakan teknik esensial yang berfungsi untuk menggali informasi tersembunyi dari kumpulan data berskala besar melalui dua pendekatan utama, yaitu klastering dan klasifikasi(Elang et al., 2023). Klastering digunakan untuk mengelompokkan objek atau titik data ke dalam dua atau lebih kelompok, di mana setiap anggota dalam satu kelompok memiliki tingkat kemiripan yang lebih tinggi dibandingkan dengan anggota dari kelompok lain. Dalam konteks kesehatan, tujuan klastering adalah mengidentifikasi subkelompok pasien dengan karakteristik serta faktor risiko yang sejenis guna mendukung pengembangan strategi intervensi yang lebih spesifik dan personal (Maulida Nuzula Firdaus, 2023). Salah satu algoritma klastering yang paling umum digunakan adalah K-Means Clustering**,** yaitu metode non-hirarki yang mengelompokkan variabel berdasarkan tingkat kesamaan melalui proses iteratif hingga posisi centroid atau titik pusat tiap klaster mencapai kestabilan dalam ruang multidimensional (Syarat et al., 2024). Penentuan jumlah klaster optimal (*k*) pada algoritma ini dapat dilakukan dengan Metode Elbow**,** yang mengamati titik "siku" pada grafik ketika

penurunan nilai inertia mulai melambat secara signifikan, atau menggunakan Skor Silhouette, di mana nilai tertinggi menunjukkan jumlah klaster yang paling ideal (*SKRIPSI.Pdf*, n.d.). Algoritma K-Means Clustering telah banyak digunakan dalam klasifikasi penyakit Diabetes Melitus (DM)**,** yaitu gangguan metabolik kronis yang ditandai oleh peningkatan kadar gula darah (hiperglikemia) akibat disfungsi insulin(Syarat et al., 2024). Dalam salah satu penelitian yang menggunakan 768 data pasien dengan delapan atribut numerik, diperoleh tiga klaster optimal yang merepresentasikan profil risiko berbeda: klaster pertama mencakup pasien dengan kadar glukosa dan BMI tinggi (berkaitan dengan DM tipe 2), klaster kedua terdiri dari pasien dengan BMI rendah dan kadar glukosa normal (risiko rendah), sedangkan klaster ketiga mengelompokkan pasien dengan kadar insulin rendah dan usia relatif muda. Penelitian lain di Puskesmas Lappa**e** berhasil mengelompokkan 1040 data pasien menjadi dua klaster utama, yaitu kelompok berisiko tinggi (584 data) dan berisiko rendah (456 data), dengan faktor risiko utama berupa hiperglikemia, hipertensi, obesitas (BMI tinggi), dan riwayat keturunan. Sementara itu, analisis serupa di Puskesmas Mojokerto terhadap 1163 data pasien menghasilkan dua klaster optimal berdasarkan usia dan jenis kelamin: Klaster 1 terdiri atas 755 pasien perempuan berusia 20–80 tahun, sedangkan Klaster 2 berisi 404 pasien laki-laki berusia 40–90 tahun. Jenis diagnosis yang paling dominan di kedua klaster adalah Non-insulin-dependent diabetes mellitus with unspecified complications (E11.8) **atau** Diabetes Melitus tipe 2 (Elang et al., 2023)**.**

3. METODE PENELITIAN

Tahap Penelitian

Tahap penelitian ini meliputi rangkaian proses yang disusun secara sistematis guna merealisasikan objekif penelitian. Melalui penentuan tahapan yang terstruktur, proses penelitian dapat dilaksanakan secara terarah dan efisien. Setiap tahapan mencakup aktivitas mulai dari identifikasi permasalahan, pengumpulan dan pengolahan dataset, hingga analisis hasil dan penarikan kesimpulan serta pemberian saran. Adapun alur tahapan penelitian dalam studi ini disajikan sebagai berikut:



Gambar 1. Tahap Penelitian.

Berdasarkan Gambar 1, tahapan penelitian yang digunakan dalam studi ini terdiri atas beberapa langkah sistematis sebagai berikut:

a. Identifikasi Masalah

Tahap awal penelitian dimulai dengan proses identifikasi masalah, yang bertujuan untuk memahami secara mendalam isu atau permasalahan yang akan diselesaikan. Pada tahap ini, dilakukan pula studi literatur dengan menelaah berbagai sumber relevan seperti jurnal ilmiah, buku, maupun artikel, guna memperoleh dasar teoritis dan referensi yang mendukung dalam penyusunan solusi penelitian.

b. Pengumpulan Dataset

Tahap berikutnya adalah pengumpulan dataset yang akan digunakan sebagai bahan analisis. Proses ini mencakup pengumpulan data dari sumber external. Dataset yang digunakan dalam penelitian ini bersumber dari platform Kaggle.com.

c. Metode Pendekatan

Setelah dataset terkumpul, penelitian dilanjutkan dengan penerapan metode *Knowledge Discovery in Databases* (KDD). Metodologi KDD berperan sebagai kerangka ilmiah yang sistematis dalam proses *data mining*, yang terdiri atas beberapa tahapan penting untuk mengekstraksi pengetahuan dari basis data. Tahapan tersebut meliputi:

1) Data Selection

Merupakan proses pemilihan dan ekstraksi data spesifik dari kumpulan data mentah berdasarkan kriteria yang telah ditentukan sebelumnya, guna memastikan relevansi data terhadap tujuan penelitian.

2) Data Preprocessing

Tahap ini bertujuan untuk mengatasi permasalahan pada dataset, seperti *missing value*, duplikasi, serta inkonsistensi data. Proses ini juga mencakup konversi tipe

data dan normalisasi nilai atribut agar data berada dalam format yang seragam dan siap untuk dianalisis lebih lanjut.

3) Data Transformation

Proses ini mencakup konvensi format data untuk memenuhi kebutuhan analisis. Data yang tidak sesuai akan disesuaikan atau diubah menjadi format standar yang diterapkan dalam proses data mining.

4) Data Mining

Tahap ini merupakan inti dari proses KDD, di mana dilakukan pencarian pola atau karakteristik tertentu pada dataset. Informasi yang dihasilkan dari tahap ini menjadi dasar dalam memperoleh pengetahuan baru yang relevan dengan tujuan penelitian.

5) Interpretation

Tahap akhir KDD melibatkan interpretasi dan evaluasi hasil *data mining*. Data yang diperoleh diubah ke dalam bentuk model, tabel, atau visualisasi grafik untuk kemudian dievaluasi berdasarkan tingkat keakuratan dan relevansinya terhadap permasalahan penelitian.

d. Kesimpulan dan Saran

Berdasarkan penelitian yang dilakukan terciptalah hasil dari pengolahan data yang terjadi, serta saran yang dapat diberikan selama terjadinya proses penelitian.

Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis *data mining* dengan mengadopsi kerangka kerja *Knowledge Discovery in Databases* (KDD) sebagai metode utama dalam proses ekstraksi pengetahuan dari data. Pendekatan KDD dipilih karena mampu menyediakan tahapan analisis yang sistematis mulai dari pemilihan data hingga interpretasi hasil. Dalam tahap *data mining*, digunakan algoritma *K-Means Clustering* sebagai metode klasterisasi tanpa supervisi (*unsupervised learning*) untuk mengelompokkan data berdasarkan kesamaan karakteristik antar atribut (Nuryamin & Risyda, 2025)

Tinjauan Pustaka

Data Mining

Data mining atau penggalian data secara fundamental didefinisikan sebagai proses sistematis untuk mengekstraksi informasi, pola, dan pengetahuan yang bernilai dari kumpulan data berukuran besar atau basis data. Proses ini tidak sekadar berfokus pada pengumpulan data, tetapi lebih kepada usaha menemukan pola tersembunyi (*hidden patterns*) yang tidak mudah

terdeteksi melalui metode analisis konvensional. Tujuan utama dari *data mining* adalah mengungkap hubungan signifikan serta fakta-fakta baru yang dapat digunakan untuk mendukung proses pengambilan keputusan yang lebih akurat, baik dalam bidang bisnis, industri, maupun sektor khusus seperti kesehatan.(Syarat et al., 2024)

Sebagai bidang kajian interdisipliner, *data mining* merupakan hasil integrasi dari beberapa disiplin ilmu seperti statistik, kecerdasan buatan (*artificial intelligence*), *machine learning*, dan sistem basis data. Kolaborasi antarbidang ini memungkinkan pengembangan metode analisis canggih untuk menangani data dalam skala besar dan kompleks(Tri Gustiane et al., 2024). Salah satu teknik utama dalam *data mining* adalah *clustering* atau pengelompokan, yaitu pendekatan *unsupervised learning* yang bertujuan mengelompokkan objek berdasarkan kemiripan karakteristik tanpa memerlukan label data awal. Dalam konteks penelitian medis, terutama pada studi mengenai penyakit diabetes, metode *clustering* berperan penting dalam mengidentifikasi kelompok pasien dengan profil klinis yang serupa, sehingga dapat memberikan pemahaman lebih mendalam terhadap stratifikasi risiko dan mendukung penerapan strategi manajemen penyakit yang lebih terpersonalisasi.(Usia et al., 2025).

Clustering

Clustering, atau pengelompokan data, adalah sebuah teknik fundamental dalam data mining yang bertujuan untuk mengelompokkan sekumpulan objek atau data ke dalam beberapa kelompok (klaster) berdasarkan tingkat kesamaan karakteristik di antara mereka. Tujuan utamanya adalah untuk memaksimalkan kemiripan (homogenitas) objek dalam satu klaster yang sama, sementara meminimalkan kemiripan (heterogenitas) dengan objek di klaster lain. Berbeda dengan klasifikasi yang merupakan supervised learning (memerlukan label data sebelumnya), clustering termasuk dalam kategori unsupervised learning(Elang et al., 2023). Artinya, proses pengelompokan dilakukan tanpa adanya informasi atau kategori awal tentang bagaimana data seharusnya dikelompokkan, sehingga pola yang ditemukan murni berasal dari struktur data itu sendiri.(Feronika et al., 2025)

Dalam konteks analisis data, clustering digunakan untuk mengidentifikasi struktur atau pola tersembunyi hidden patternsdalam data, memahami distribusi data, dan menyederhanakan kumpulan data besar menjadi kelompok-kelompok yang lebih mudah dikelola dan diinterpretasikan. Beberapa algoritma clustering yang umum digunakan, seperti yang disebutkan dalam berbagai kajian, termasuk algoritma berbasis partisi seperti K-Means dan algoritma berbasis hierarki seperti Divisive Analysis (DIANA) (Gestavito et al., 2024). Penerapan clustering pada data medis, seperti data klinis pasien diabetes, memungkinkan peneliti untuk mengidentifikasi subkelompok pasien dengan profil risiko atau karakteristik

penyakit yang serupa, yang pada akhirnya dapat mendukung pengembangan strategi manajemen penyakit yang lebih tertarget.(C, 2019)

3.3.3 K-Means Clustering

Algoritma *K-Means Clustering* merupakan salah satu metode pengelompokan (*clustering*) yang paling populer dan banyak diterapkan dalam bidang *data mining*. Tujuan utama algoritma ini adalah membagi sekumpulan data ke dalam sejumlah kelompok (klaster) berdasarkan tingkat kemiripan karakteristik antar objek, sehingga data dengan karakteristik yang serupa tergabung dalam klaster yang sama. Metode ini termasuk dalam kategori *unsupervised learning*, di mana proses pengelompokan dilakukan tanpa memerlukan label atau kelas awal pada data. Pendekatan ini memungkinkan algoritma untuk menemukan pola dan struktur tersembunyi dalam data secara otomatis, tanpa intervensi atau pengetahuan awal dari peneliti.(Basyir, 2025)

Secara prinsip, *K-Means* bekerja melalui proses iteratif untuk mencapai hasil pengelompokan yang optimal. Tahapan awal melibatkan penentuan jumlah klaster (k) yang diinginkan, kemudian pemilihan k titik awal sebagai pusat klaster (*centroid*). Selanjutnya, setiap data akan dikaitkan dengan *centroid* terdekat berdasarkan ukuran jarak tertentu, umumnya menggunakan jarak *Euclidean*. Setelah semua data dikelompokkan, posisi *centroid* diperbarui berdasarkan rata-rata posisi anggota dalam masing-masing klaster. Proses ini berlangsung secara berulang hingga posisi *centroid* tidak lagi berubah secara signifikan atau telah mencapai kondisi konvergen. Karena kesederhanaan dan efisiensinya dalam menangani data berukuran besar, algoritma *K-Means* banyak digunakan untuk mengidentifikasi pola, termasuk dalam penelitian medis seperti pengelompokan pasien diabetes berdasarkan kesamaan profil klinis mereka.(Fernanda, S. I., Ratnawati, D. E., dan Adikara, 2017).

Orange

Berdasarkan hasil kajian literatur, *Orange* merupakan perangkat lunak berbasis visualisasi yang berfungsi sebagai alat bantu dalam penerapan berbagai algoritma *data mining*. Aplikasi ini menyediakan lingkungan analisis yang interaktif, sehingga pengguna dapat melakukan proses eksplorasi data, penerapan algoritma, serta evaluasi hasil secara terintegrasi (Cahyani & Basuki, 2019). Dalam konteks penelitian terkait klasifikasi penyakit diabetes, *Orange* digunakan sebagai platform untuk mengimplementasikan algoritma *Support Vector Machine* (SVM). Keunggulan utama aplikasi ini terletak pada kemampuannya menyederhanakan alur analisis melalui antarmuka visual yang intuitif, memungkinkan peneliti menjalankan seluruh tahapan mulai dari praproses data hingga evaluasi model tanpa perlu

melakukan pengkodean kompleks, sehingga mempercepat proses eksperimen dan meningkatkan efisiensi validasi hasil.

4. HASIL DAN PEMBAHASAN

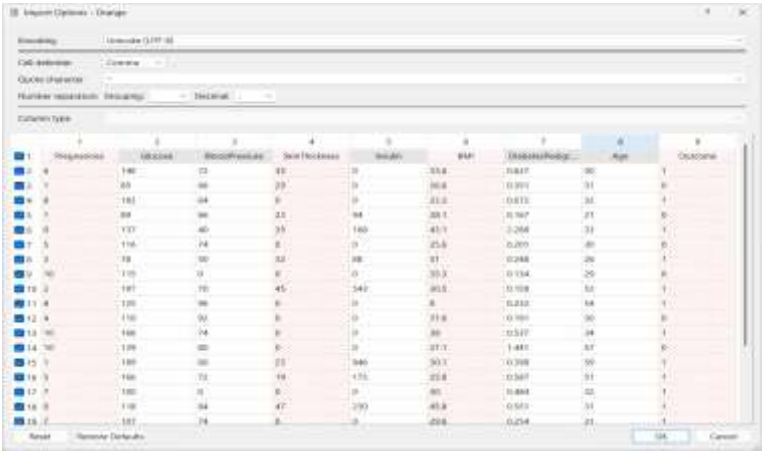
Dataset

Tahap awal dalam penelitian ini adalah proses pengumpulan dataset yang diperoleh dari berbagai sumber terpercaya. Pada penelitian ini, digunakan dataset publik yang diunduh dari situs Kaggle (kaggle.com) dengan jumlah total sebanyak 769 data. Adapun tautan sumber dataset tersebut adalah sebagai berikut: (https://www.kaggle.com/code/yazidivan1/diabetes-prediction-using-logistic-regression/input).

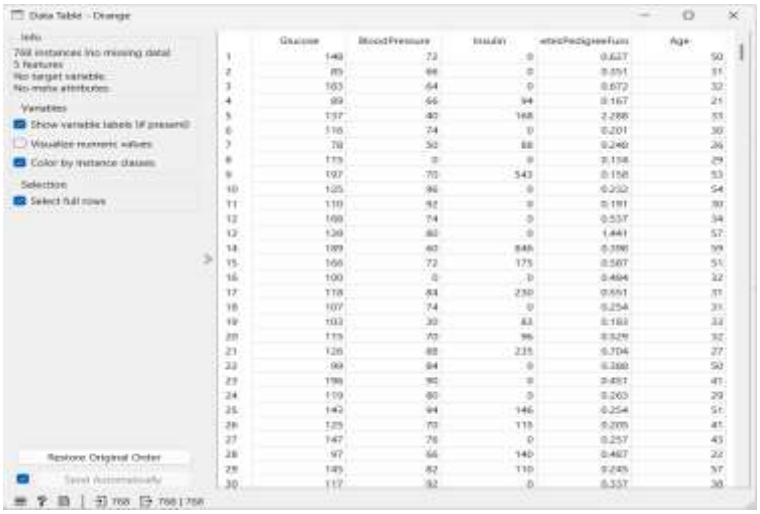
Data awal yang diperoleh terdiri atas 9 atribut atau kolom, yaitu Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI (Body Mass Index), Diabetes Pedigree Function, Age, Outcome. Setiap atribut tersebut berpotensi mempengaruhi hasil proses data mining, sehingga diperlukan penyesuaian dan pemilihan atribut yang relevan pada tahap data selection untuk memastikan kualitas dan akurasi analisis data.

Data Selection

Setelah dataset yang akan digunakan dalam penelitian berhasil dikumpulkan, tahap selanjutnya adalah data selection atau seleksi data, yakni proses pemilihan data berdasarkan atribut/kolom serta jumlah data yang akan digunakan dalam proses data mining.(C, 2019) Pada penelitian ini, dilakukan pemilihan data dari dataset yang telah diperoleh dengan kriteria sebagai berikut:



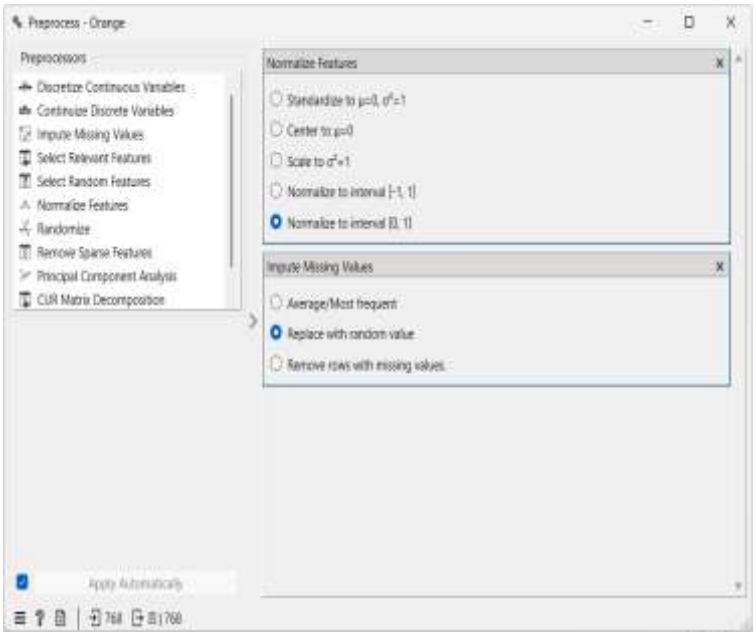
Gambar 2. Atribut Awal.



Gambar 3. Hasil Data Selection.

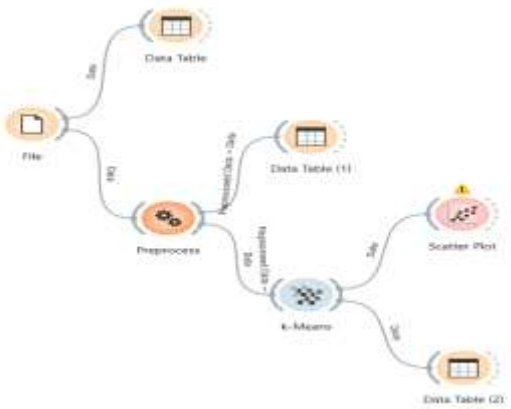
Data Preprocessing

Setelah tahap data selection selesai dilakukan, langkah berikutnya adalah data preprocessing, yaitu proses pengolahan data mentah menjadi data yang siap untuk digunakan dalam analisis. Pada tahap ini, dilakukan serangkaian prosedur seperti pembersihan data (data cleaning) untuk mengatasi missing values, menghapus data duplikat, serta melakukan penyesuaian lain yang diperlukan agar kualitas data memenuhi standar analisis yang optimal.



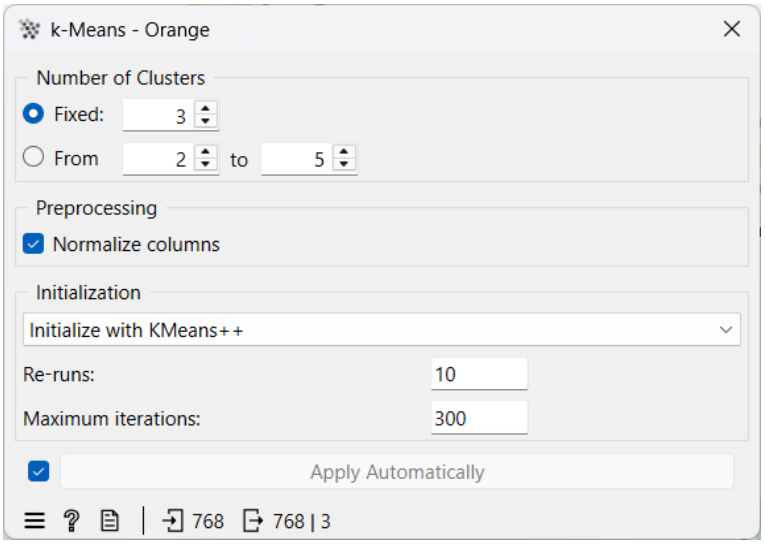
Gambar 4. Preprosessing.

Proses Klastering



Gambar 5. Proses Klastering data melalui Orange.

a. Pada proses awal memasukan dataset penderita diabetes dalam bentuk .csv, Setelah file berhasil diimpor, data tersebut dimasukkan ke dalam sebuah data table yang berfungsi sebagai representasi awal dari data mentah. Pada tahap ini, data disajikan dalam format tabular yang terdiri atas baris dan kolom sehingga memudahkan proses observasi dan analisis awal.
b. Data yang berasal dari *Data Table* selanjutnya dialirkan ke modul *Preprocess* untuk melalui tahap praproses (*preprocessing*), yang merupakan fase esensial dalam mempersiapkan data sebelum analisis lanjutan. Tahap ini bertujuan untuk membersihkan serta mentransformasi data mentah agar menjadi lebih representatif dan berkualitas bagi algoritma *machine learning*. Proses yang dilakukan mencakup pembersihan data, seperti penanganan nilai hilang melalui imputasi, koreksi kesalahan entri, serta deteksi dan penanganan *outlier*; transformasi data melalui normalisasi atau standarisasi fitur numerik agar tidak terjadi dominasi variabel tertentu dalam perhitungan jarak pada algoritma *k-Means*; serta rekayasa fitur melalui seleksi atau pembuatan fitur baru yang lebih relevan. Hasil akhir dari tahap ini berupa *Data Table (1)* yang berisi data terproses dan berlabel "*Preprocessed Data*".

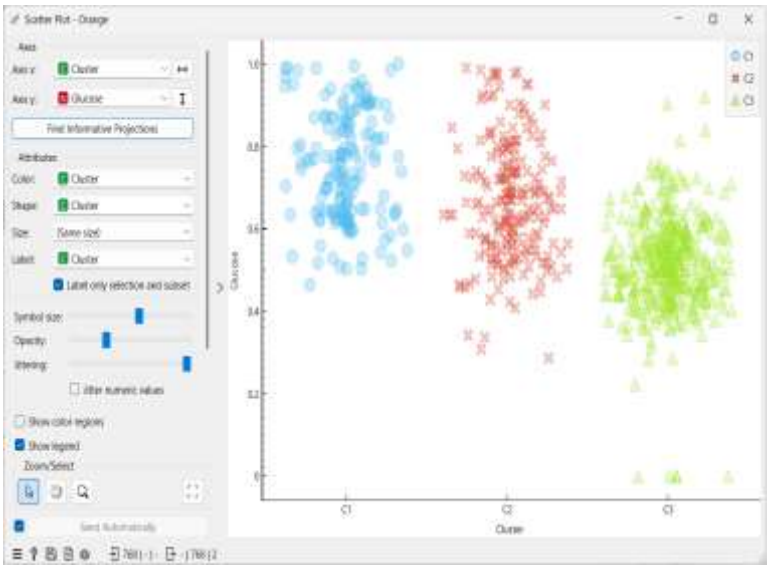


Gambar 6. K-Means.

c. Data yang telah melalui tahap praproses pada *Data Table (1)* selanjutnya digunakan sebagai masukan untuk algoritma *k-Means*, yang merupakan metode *unsupervised learning* dengan tujuan mengelompokkan dataset ke dalam sejumlah *k* klaster yang berbeda. Proses kerja algoritma ini meliputi penentuan jumlah klaster (*k*), inisialisasi acak titik pusat klaster (*centroid*), penugasan setiap data ke klaster dengan *centroid* terdekat berdasarkan jarak Euclidean, serta pembaruan posisi *centroid* hingga mencapai kondisi konvergensi. Hasil dari proses ini berupa *Data Table (2)*, yang memiliki struktur serupa dengan *Data Table (1)* namun dilengkapi dengan kolom tambahan berisi label klaster untuk setiap entri data, dan aliran data ini disebut sebagai "*Clustered Data*". Selain itu, diagram menunjukkan adanya *Data Table (3)* yang merepresentasikan penggabungan antara data praproses dan label klaster sebagai dataset akhir yang siap untuk divisualisasikan.

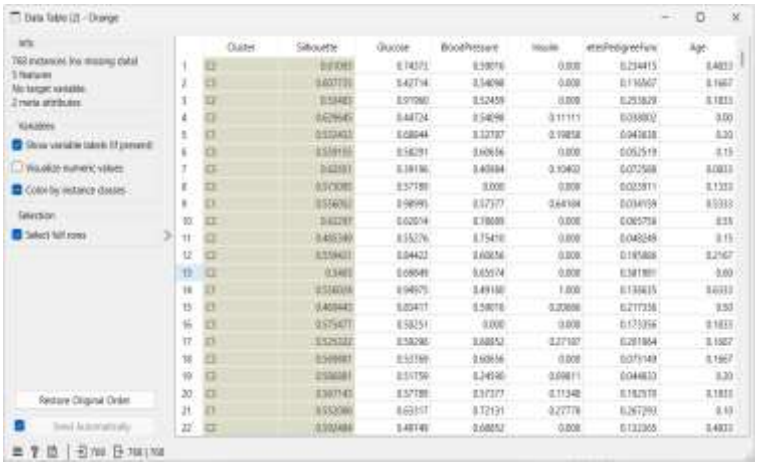
d. Data yang telah diberi label klaster pada *Data Table (2)* maupun *Data Table (3)* selanjutnya dimasukkan ke dalam modul *Scatter Plot* untuk tahap visualisasi, yang berfungsi menampilkan hasil analisis klaster secara intuitif dan informatif. Hasil dari proses ini berupa diagram pencar (*scatter plot*), di mana setiap titik mewakili satu observasi dari dataset. Posisi titik pada sumbu X dan Y ditentukan oleh dua variabel terpilih, sedangkan warna atau bentuk titik merepresentasikan keanggotaan klaster yang ditetapkan oleh algoritma *k-Means*. Melalui visualisasi ini, analis dapat dengan mudah mengamati pemisahan antar klaster, distribusi kepadatan data, serta pola-pola tersembunyi yang tidak dapat teridentifikasi hanya melalui tampilan data dalam bentuk tabel.

Hasil Klastering



Gambar 7. Hasil Klastering.

Pada gambar 7, visualisasi yang dihasilkan berupa *scatter plot* yang merepresentasikan hasil analisis klaster menggunakan algoritma *k-Means*. Pada grafik tersebut, sumbu horizontal (X) menunjukkan label klaster yang terbentuk (C0, C1, dan C2), sedangkan sumbu vertikal (Y) menampilkan nilai variabel *Glucose*. Setiap titik data diberi warna dan bentuk geometris berbeda lingkaran biru untuk C1, silang merah untuk C2, dan segitiga hijau untuk C3 guna membedakan masing-masing kelompok secara visual. Representasi ini memperlihatkan distribusi nilai *Glucose* di dalam tiap klaster, sehingga memfasilitasi pemahaman terhadap karakteristik dan pola dominan yang membedakan setiap segmen data hasil pengelompokan.



Gambar 8. Hasil Akhir.

5. KESIMPULAN DAN SARAN

Penelitian ini membuktikan efektivitas algoritma K-Means dalam mengelompokkan pasien diabetes berdasarkan karakteristik klinis melalui kerangka kerja Knowledge Discovery in Databases (KDD) dengan menggunakan dataset publik dari Kaggle yang berisi 769 rekam medis. Melalui tahapan seleksi data, praproses, dan clustering menggunakan perangkat lunak Orange, diperoleh tiga klaster pasien (C1, C2, dan C3) yang berbeda terutama berdasarkan variabel Glucose, menunjukkan kemampuan K-Means dalam mengidentifikasi pola stratifikasi alami pada data klinis. Untuk penelitian selanjutnya, disarankan dilakukan analisis lebih mendalam terhadap karakteristik setiap klaster, seperti menghitung nilai rata-rata atau median dari variabel klinis (BMI, usia, tekanan darah) agar diperoleh profil yang lebih komprehensif. Selain itu, perlu dilakukan perbandingan dengan algoritma clustering lain seperti DBSCAN atau Hierarchical Clustering guna mengevaluasi keandalan hasil, serta melibatkan validasi dari ahli medis agar klaster yang terbentuk tidak hanya signifikan secara statistik, tetapi juga relevan secara klinis dan berpotensi diterapkan dalam sistem pendukung keputusan klinis.

UCAPAN TERIMA KASIH

Terimakasih untuk semua pihak yang terlibat dalam penulisan jurnal ini.

DAFTAR REFERENSI

Albirruni, R. A., & Suwitho, S. (2023). Pengaruh kualitas layanan, harga, dan citra merek terhadap kepuasan pelanggan (studi pada kasus jasa pengiriman Anter Aja Surabaya). *Jurnal Ilmu dan Riset Manajemen (JIRM), 12*(10).

Ghozali, I. (2021). *Aplikasi analisis multivariate dengan program IBM SPSS 26*. Semarang: Badan Penerbit Universitas Diponegoro.

Hidayat, R. (2021). Pengaruh kualitas pelayanan dan harga terhadap kepuasan pelanggan pada lembaga kursus. *Jurnal Manajemen Pendidikan, 9*(1), 45-55. https://doi.org/10.52300/jmso.v1i1.2370

Kotler, P., & Armstrong, G. (2021). *Principles of marketing* (18th ed.). Pearson.

Kotler, P., & Keller, K. L. (2020). *Marketing management* (16th ed.). Pearson Education.

Oliver, R. L. (2019). *Satisfaction: A behavioral perspective on the consumer* (2nd ed.). New York: Routledge.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(1), 12-40.

Rahmania, N. C. (2022). Pengaruh kualitas pelayanan terhadap kepuasan peserta didik pada lembaga kursus dan pelatihan. *DIKLUS: Jurnal Pendidikan Luar Sekolah*. https://doi.org/10.21831/diklus.v6i1.39620

Rahmawati, Y., Widayati, C. C., & Perkasa, D. H. (2023). Pengaruh cita rasa, harga, dan kualitas pelayanan terhadap kepuasan konsumen (studi kasus pada resto Street Sushi cabang Meruya Jakarta Barat). *Jurnal Humaniora, Ekonomi Syariah dan Muamalah, 1*(3), 120-130. https://doi.org/10.38035/jhesm.v1i3.71

Setiawan, B. P., & Frianto, A. (2021). Pengaruh harga dan kualitas pelayanan terhadap kepuasan pelanggan (studi kasus perusahaan jasa ekspedisi Krian). *BIMA: Journal of Business and Innovation Management, 3*(3), 352-366. https://doi.org/10.33752/bima.v3i3.5493

Sugiyono. (2019). *Metode penelitian kualitatif, kuantitatif, dan R&D*. Bandung: Alfabeta.

Tugiyono, J. (2020). Kualitas pelayanan dan pengaruhnya terhadap kepuasan peserta kursus LKP Pramidia Bandung. *Jurnal TEDC, 14*(2), 134-144.

Wijaya, M. (2022). Analisis pengaruh kualitas pelayanan dan citra institusi terhadap kepuasan pelanggan. *Jurnal Bisnis/Ekonomi* (contoh studi institusi pendidikan).

Wulandari, A. (2022). Analisis pengaruh kualitas pelayanan dan harga terhadap kepuasan peserta kursus bahasa Inggris. *Jurnal Ekonomi dan Bisnis Pendidikan, 11*(2), 112-120.

Wulandari, Z. (2022). Pengaruh kualitas pelayanan terhadap kepuasan pelanggan pada JNE Express Karawang. *Jurnal Pendidikan dan Konseling (JPDK), 4*(6), 11496-11501. https://doi.org/10.31004/jpdk.v4i4.6390

Zeithaml, V. A., Bitner, M. J., & Gremler, D. D. (2020). *Services marketing: Integrating customer focus across the firm*. McGraw-Hill Education.