Terbit online pada laman web jurnal: **https://journal.upp.ac.id/index.php/rjti/index**

RJTI
Riau Jurnal Teknik Informatika

Vol. 4 No. 3 (2025) 386 – 395 | ISSN: 2829-2529 (Media Online)

Penerapan Algoritma *K-Means* Untuk Pengelompokan Risiko Penyakit Diabetes: Pendekatan Berbasis Data untuk Deteksi Dini

Nabila Yuniarti[1], Najah Nur Aliyah[2], Widiana Salsabilah[3], Ziyad Dante Al Najji[4], Bibit Sudarsono[5], Wisti Dwi Septiani[6]

[1,2,3,4,5,6] Program Studi Informatika, Fakultas Teknik Informatika, Universitas Bina Sarana Informatika

[1]nabilaybrt@gmail.com, [2]najahnuraliyah17@gmail.com, [3]widianasalsabilah12@gmail.com, [4]ziyaddantealnajji@gmail.com, [5]bibit.bbs@bsi.ac.id, [6]wisti.wst@bsi.ac.id

Abstract

Diabetes is a condition in which the pancreas is unable to produce insulin optimally, or when the body cannot use insulin effectively, thereby disrupting insulin distribution. In conducting this research, the researchers adopted two main approaches that formed the basis for the data collection and analysis process, namely a literature study conducted by searching for, evaluating, and reviewing various scientific journal articles and other reliable sources related to the research topic, and the application of the K-Means algorithm, which provides a more structured overview of the distribution of patient groups. In the analysis process, the RapidMiner application was used to facilitate data grouping and enable researchers to evaluate the performance of the applied K-Means algorithm. The dataset used contained 5,000 patient data and 9 health attributes, which were grouped using the Elbow method and validated with the Davies-Bouldin Index, with a value of 0.827. Overall, there were three main cluster s, each showing different health characteristics. The first cluster consisted of patients with low risk (45%), who had normal blood sugar levels and no diagnosed diabetes. The second cluster shows a group with medium risk (35%) who are beginning to show symptoms of pre-diabetes as well as increased blood sugar levels and several other risk factors. Meanwhile, the third cluster contains patients with high risk (20%) who have very high blood sugar levels, most of whom are already in the diabetes phase and face more serious conditions. These findings indicate that the cluster ing results not only describe variations in patients' health conditions but also have practical value in a medical context, such as helping healthcare professionals perform early detection, prioritize high-risk patients, and support more personalized and targeted intervention strategies.

Keywords: *Diabetes, K-Means, RapidMiner, Cluster ing, Diabetes Risk*

Abstrak

Diabetes merupakan kondisi ketika pankreas tidak mampu memproduksi insulin secara optimal, atau ketika tubuh tidak dapat menggunakan insulin dengan efektif, sehingga distribusi insulin menjadi terganggu. Dalam pelaksanaan penelitian ini, peneliti mengadopsi dua pendekatan utama yang menjadi landasan dalam proses pengumpulan dan analisis data, yaitu studi literatur yang dilakukan dengan mencari, mengevaluasi, dan mengkaji berbagai artikel jurnal ilmiah, sumber terpercaya lainnya yang berkaitan dengan topik penelitian, dan penerapan algoritma *K-Means* yang memberikan gambaran lebih terstruktur mengenai distribusi kelompok pasien. Dalam proses analisis, digunakan aplikasi *RapidMiner* untuk mempermudah pengelompokan data dan memungkinkan peneliti mengevaluasi kinerja algoritma *K-Means* yang diterapkan. Dataset yang digunakan berisi 5.000 data pasien dan 9 atribut kesehatan, yang dikelompokkan menggunakan metode *Elbow* dan validasi dengan *Davies-Bouldin Index*, dengan nilai 0,827. Secara keseluruhan, terdapat tiga *cluster* utama yang masing-masing menunjukkan karakteristik kesehatan berbeda. *Cluster* pertama terdiri dari pasien dengan risiko rendah (45%), yang memiliki kadar gula darah normal dan tidak ada yang terdiagnosis diabetes. *Cluster* kedua menunjukkan kelompok dengan risiko menengah (35%) yang mulai menunjukkan gejala pra-diabetes serta peningkatan kadar gula darah dan beberapa faktor risiko lain. Sedangkan *cluster* ketiga berisi pasien dengan risiko tinggi (20%) yang memiliki kadar gula darah sangat tinggi, di mana sebagian besar sudah berada dalam fase diabetes dan menghadapi kondisi yang lebih serius. Temuan ini menunjukkan bahwa hasil klasterisasi tidak hanya menggambarkan variasi kondisi kesehatan pasien, tetapi juga memiliki nilai praktis dalam konteks medis, seperti membantu tenaga kesehatan melakukan deteksi dini, memprioritaskan pasien berisiko tinggi, serta mendukung strategi intervensi yang lebih personal dan tepat sasaran.

Kata kunci: *Diabetes, K-Means, RapidMiner , Klasterisasi, Risiko Diabetes*

1. Pendahuluan

Diabetes merupakan kondisi ketika pankreas tidak mampu memproduksi insulin secara optimal, atau ketika tubuh tidak dapat menggunakan insulin dengan efektif, sehingga distribusi insulin menjadi terganggu [1]. Insulin merupakan salah satu hormon utama yang dihasilkan oleh organ pankreas dan memiliki peran yang sangat vital dalam menjaga kestabilan kadar glukosa (gula) di dalam aliran darah. Hormon ini membantu tubuh dalam mengatur proses penyerapan dan penyimpanan glukosa, sehingga kadar gula darah tetap berada dalam rentang normal [2]. Kadar glukosa atau gula darah yang tinggi menyebabkan diabetes, suatu gangguan metabolisme yang terjadi di dalam tubuh. Karena gula darah merupakan sumber energi utama bagi sel dan jaringan, maka gula darah sangat penting untuk kesehatan yang baik. Diabetes dapat menyebabkan sejumlah masalah jika tidak dikontrol dengan baik, termasuk obesitas, penyakit jantung, stroke, dan masalah pada mata, ginjal, dan saraf [3].

Berdasarkan data dari *International Diabetes Federation (IDF)*, pada tahun 2021 ada sekitar 536,6 juta orang di dunia yang menderita diabetes. Angka ini diperkirakan akan terus naik dan mencapai lebih dari 783,2 juta orang pada tahun 2045. Yang mengejutkan, sekitar 50% dari penderita diabetes sebenarnya tidak sadar bahwa mereka punya penyakit ini. Dari sisi medis, mendeteksi diabetes sejak dini, terutama saat belum muncul gejala, sangat penting. Dengan begitu, pengobatan bisa dimulai lebih cepat dan risiko komplikasi serius bisa dikurangi. Saat ini, diperkirakan hampir satu dari dua orang dewasa berusia 20 sampai 79 tahun yang mengidap diabetes tidak tahu kalau mereka mengidapnya, itu sekitar 44,7% atau 239,7 juta orang. Beberapa wilayah di dunia memiliki tingkat kasus diabetes yang tidak terdiagnosis cukup tinggi. Afrika menempati urutan teratas (53,6%), disusul wilayah Pasifik Barat (52,8%) dan Asia Tenggara (51,3%). Indonesia, yang termasuk dalam wilayah Asia Tenggara, menjadi salah satu dari tiga negara dengan jumlah penderita diabetes yang tidak terdiagnosis paling tinggi di dunia [4].

Beberapa penelitian sebelumnya telah menerapkan algoritma *K-Means* dalam analisis pola penyakit kronis seperti diabetes melitus tipe 2 (T2DM) dan hipertensi. mengelompokkan pasien T2DM dengan komplikasi kronis menjadi empat klaster berdasarkan 11 variabel klinis umum, dan setiap klaster menunjukkan perbedaan signifikan dalam pola klinis serta risiko komplikasi metabolik dan vaskular. Hasil tersebut membuktikan bahwa analisis *klaster* menggunakan *K-Means* dapat membantu mengidentifikasi *fenotipe* penyakit yang berbeda dan memiliki nilai potensial dalam pengambilan keputusan medis. Namun, sebagian besar penelitian tersebut masih berfokus pada hasil pengelompokan, tanpa mengaitkannya secara langsung dengan strategi intervensi atau prediksi risiko pasien. Oleh karena itu, penelitian ini menawarkan pendekatan yang lebih aplikatif, yaitu dengan mengelompokkan pasien berdasarkan tingkat risiko diabetes menggunakan algoritma *K-Means* untuk mendukung proses deteksi dini dan pengawasan medis yang lebih efektif [5].

Perkembangan teknologi informasi saat ini memberi pengaruh besar pada proses digitalisasi di berbagai bidang, termasuk layanan kesehatan. Penerapan rekam medis berbasis digital membuat penyimpanan serta pengelolaan data pasien yang jumlahnya banyak menjadi lebih terstruktur. Selain membantu mempercepat dan mempermudah pelayanan kesehatan, pemanfaatan sistem ini juga memberikan peluang baru untuk menggunakan data sebagai dasar pengambilan keputusan klinis yang lebih akurat dan terarah [6]. Dalam hal ini pemanfaatan algoritma *machine learning* menjadi semakin penting dalam pengolahan dan analisis data kesehatan. Machine learning adalah sistem yang dapat belajar membuat keputusan sendiri tanpa harus diprogram berulang kali oleh manusia sehingga komputer dapat menjadi lebih pintar dan belajar dari pengalamannya dengan data [7].

Berdasarkan pendekatan atau metode pembelajarannya, yaitu: *supervised learning, unsupervised learning, semi-supervised learning*, dan *reinforcement learning*. Setiap metode memiliki karakteristik serta pendekatan yang berbeda dalam mengenali pola, mengolah data, dan menghasilkan output yang akurat sesuai tujuan analisis. *Supervised learning*, sistem dilatih menggunakan himpunan data yang telah diberi label, yaitu data yang mencakup pasangan antara input dan output yang diharapkan. Melalui proses pelatihan ini, sistem mempelajari hubungan atau pola yang terkandung dalam data tersebut. Setelah pola berhasil dikenali, sistem akan menggunakan pola tersebut sebagai dasar untuk memprediksi atau mengklasifikasikan data baru yang belum pernah dilihat sebelumnya, algoritma ini terbagi menjadi dua kategori, yaitu kategori *Regresi* yang terdiri dari *Regresi Linear, Dicision Tree, Random Forest*, dan kategori Klasisifikasi yang terdiri dari Klasifikasi *Biner, Naive Bayes, Support Vector Machine,* dan *Neural Networks. Semi-Supervised Learning* adalah metode pembelajaran mesin yang menggabungkan teknik dari *supervised learnin*g dan *unsupervised learning*. Dalam pendekatan ini, data yang digunakan terdiri dari dua jenis, yaitu sebagian data yang sudah dilengkapi dengan label dan sebagian lainnya masih berupa data tanpa label, beberapa jenis dari algoritma ini, yaitu *Sel Training,* Label *Propagation*,dan *Generative Models*. *Reinforcement learning* merupakan salah satu pendekatan dalam *machine learning* yang fokus utamanya adalah mempelajari bagaimana suatu agen perangkat lunak *(software agent)* dapat mengambil keputusan atau tindakan tertentu dalam suatu lingkungan (*environment*) guna mencapai hasil yang optimal. Pendekatan ini dirancang untuk melatih agen agar mampu belajar dari pengalaman secara bertahap, melalui proses *trial* and *error*, dengan tujuan akhir memperoleh

DOI : https://doi.org/10.52158/rjti.v4i3.3518

urutan keputusan yang paling efektif demi memaksimalkan keberhasilan atau keuntungan dalam konteks dunia nyata, beberapa contoh dari algoritma ini, yaitu *Value Based Methods, Evolutionary Algorithms,* dan *Multi Agent Reinforcement Learning* [8].

Unsupervised Learning adalah, suatu teknik dalam pembelajaran mesin yang digunakan untuk menarik kesimpulan atau pola dari sebuah dataset, di mana data tersebut tidak memiliki label atau respons yang telah ditentukan sebelumnya. Dengan kata lain, teknik ini menganalisis input data tanpa adanya panduan output, untuk menemukan struktur tersembunyi, pengelompokan, atau hubungan di dalam data tersebut [9]. Beberapa algoritma yang termasuk dalam kategori ini antara lain *K-Means, Hierarchical Cluster ing, DBSCAN,* dan *Gaussian Mixture Model.* Dalam hal ini, peneliti menggunakan *unsupervised learning*, yaitu algoritma *K-Means* untuk mengelompokkan data pasien ke dalam beberapa klaster berdasarkan kesamaan karakteristik kesehatan mereka.

2. Metode Penelitian

Dalam pelaksanaan penelitian ini, peneliti mengadopsi dua pendekatan utama yang menjadi landasan dalam proses pengumpulan dan analisis data. Pendekatan-pendekatan tersebut dipilih secara cermat untuk memastikan bahwa hasil penelitian dapat diperoleh secara komprehensif dan sesuai dengan tujuan yang telah ditetapkan, yaitu:

2.1. Studi Literatur

Dilakukan dengan mencari, mengevaluasi, dan mengkaji berbagai artikel jurnal ilmiah, serta sumber terpercaya lainnya yang berkaitan dengan topik penelitian. Langkah ini bertujuan untuk memperoleh pemahaman mendalam mengenai konsep teoritis, metodologi, dan hasil-hasil penelitian sebelumnya terkait algoritma *K-Means* dan penerapannya dalam analisis data kesehatan, khususnya dalam mendeteksi penyakit diabetes.

Sebagai bagian dari proses ini, peneliti juga membandingkan beberapa temuan terdahulu untuk mengidentifikasi kesenjangan penelitian (*research gap*). Upaya ini dilakukan agar penelitian yang disusun tidak hanya mengulang studi sebelumnya, tetapi juga memberikan kontribusi baru dalam konteks pengelompokan risiko diabetes menggunakan pendekatan *unsupervised learning*. Dengan demikian, landasan teori yang digunakan menjadi lebih kuat dan relevan terhadap tujuan penelitian.

2.2. Penerapan Algoritma *K-Means*

Penelitian ini menggunakan pendekatan klasterisasi dengan algoritma *K-Means* untuk membantu mengidentifikasi kelompok pasien dengan tingkat risiko diabetes yang serupa, sehingga perawatan atau tindakan medis dapat difokuskan secara lebih tepat sasaran. Hasil klasterisasi memberikan gambaran yang lebih terstruktur mengenai distribusi kelompok pasien, yang diharapkan dapat mempermudah tenaga medis dalam mengenali individu dengan potensi risiko tinggi serta mendukung pengambilan keputusan terkait penanganan atau pemantauan secara lebih efisien. Dataset yang digunakan dalam penelitian ini memuat informasi medis seperti tekanan darah tinggi, penyakit jantung, riwayat merokok, standar berat badan, rata-rata kadar gula darah, kadar gula darah saat ini, status diabetes, serta data pribadi seperti jenis kelamin dan umur.

Data tersebut diperoleh dari *platform* dataset penelitian, yaitu diabetes dataset, yang terdiri dari 5.000 baris data dan dipilih karena memuat atribut-atribut yang relevan dengan kondisi penyakit diabetes. Dalam proses analisis, digunakan aplikasi *RapidMiner* untuk mempermudah pengelompokan data dan memungkinkan peneliti secara langsung mengevaluasi kinerja algoritma *K-Means* yang diterapkan [10]. Evaluasi model dapat dilakukan secara cepat dan efisien dengan bantuan alat ini. Fokus utama dari penelitian ini adalah untuk mengetahui sejauh mana algoritma *K-Means* berperan dalam mengelompokkan data pasien berdasarkan status diabetes mereka.

Hasil dari penelitian ini diharapkan dapat berkontribusi pada pengembangan teknologi deteksi dini diabetes, mendukung pengambilan keputusan medis, serta menjadi dasar bagi penelitian lanjutan dalam penerapan pembelajaran mesin untuk analisis risiko kesehatan yang lebih komprehensif. Untuk memastikan kualitas proses klasterisasi, peneliti juga menerapkan mekanisme validasi *internal*, seperti pemeriksaan distribusi data setiap klaster dan kesesuaian karakteristik antar kelompok. Pendekatan ini dilakukan agar pembentukan klaster tidak hanya bergantung pada hasil numerik, tetapi juga memenuhi logika klinis yang berkaitan dengan faktor risiko diabetes. Dengan cara ini, hasil analisis tidak hanya akurat secara komputasi, tetapi juga relevan secara praktis di lingkungan medis. Berikut atribut yang ada didalam dataset, yaitu:

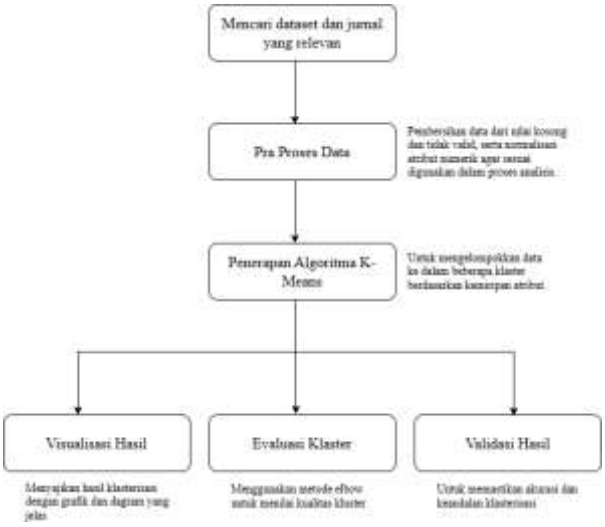
Tabel 1. Atribut Dataset

Atribut	Keterangan
Jenis Kelamin	Menyatakan jenis kelamin pasien, laki-laki atau wanita
Umur	Mengindikasi umur pasien dalam tahun
Tekanan Darah Tinggi	Menandakan apakah pasien tersebut seseorang yang menderita tekanan darah tinggi
Penyakit Jantung	Menunjukkan kondisi jantung
Riwayat Merokok	Keterangan kebiasaan merokok, apakah seseorang tidak pernah merokok, sudah berhenti, masih merokok, atau tidak tersedia informasinya.
Standar Berat Badan	Menggambarkan rasio berat badan terhadap tinggi badan untuk menilai status gizi pasien
Rata-Rata Kadar Gula Darah	Menyatakan kadar rata-rata glukosa dalam darah dalam periode waktu tertentu

DOI : https://doi.org/10.52158/rjti.v4i3.3518

Kadar Gula Darah	Menunjukkan jumlah gula dalam darah saat dilakukan tes, diukur dalam miligram per desiliter.
Status Diabetes	Menunjukkan apakah seseorang telah didiagnosis dengan diabetes

Berikut adalah serangkaian tahapan yang akan dilakukan dalam penelitian ini untuk memastikan proses analisis data berjalan secara efektif dan dapat menghasilkan temuan yang valid serta bermanfaat.



Gambar 1. Tahapan Penelitian

Adapun tahapan penelitian pada *Gambar 1* dapat dijelaskan sebagai berikut :

1. Mencari dataset dan jurnal yang relevan, peneliti mengumpulkan data yang dibutuhkan dari sumber yang terpercaya, seperti situs data publik atau jurnal ilmiah. Tujuannya adalah untuk mendapatkan bahan yang bisa digunakan dalam proses analisis. Tahap ini juga memastikan bahwa dataset memiliki kualitas yang baik, struktur jelas, serta atribut yang sesuai dengan kebutuhan analisis risiko diabetes.
2. Pra Proses Data, untuk mempersiapkan data agar sesuai untuk analisis [11]. Dalam tahap ini, data yang akan digunakan dibersihkan terlebih dahulu. Caranya adalah dengan menghapus data yang kosong atau keliru, jika ada kolom yang tidak diisi atau berisi angka yang salah. Setelah itu, angka-angka yang ada disesuaikan agar semuanya berada dalam skala yang sama. Ini penting supaya data bisa dibandingkan dengan adil dan hasil analisisnya tidak bias karena perbedaan ukuran atau satuan. Pada bagian ini, dilakukan juga pengecekan outlier untuk memastikan bahwa nilai ekstrem tidak mengganggu proses pengelompokan.
3. Penerapan Algoritma *K-Means*, setelah data bersih dan siap, dilakukan proses pengelompokan data menggunakan algoritma *K-Means*. Algoritma ini membagi data ke dalam beberapa kelompok (klaster) berdasarkan kesamaan sifat atau nilai dari data tersebut.
4. Visualisasi Hasil, pengelompokan ditampilkan dalam bentuk gambar seperti grafik atau diagram, agar lebih mudah dipahami. Visualisasi ini membantu melihat pola atau hubungan dalam data secara lebih jelas. Dalam hal ini tampilan visual juga digunakan untuk membandingkan hasil antar klaster dan memastikan apakah pola yang muncul sudah konsisten dengan karakteristik klinis.
5. Evaluasi Klaster, tahap ini diterapkan dengan mengelompokan data (*cluster* ing), yang dilakukan dengan bantuan metode *Elbow*, yaitu sebuah teknik yang digunakan untuk menentukan jumlah kelompok (klaster) yang paling tepat. Metode ini dipilih karena mampu memberikan titik acuan yang jelas dalam menentukan jumlah klaster optimal tanpa mengandalkan intuisi semata. Dalam metode ini juga, peneliti menghitung nilai *Sum of Squared Error (SSE*), yaitu total selisih antara data dan pusat kelompoknya, untuk berbagai jumlah klaster. Nilai SSE ini kemudian digambarkan dalam sebuah grafik. Biasanya, semakin banyak jumlah klaster, nilai SSE akan semakin kecil. Namun, pada titik tertentu, penurunan nilai SSE mulai melambat atau "menekuk" seperti bentuk siku (*Elbow*). Titik inilah yang dianggap sebagai jumlah klaster optimal, karena setelahnya penambahan klaster tidak memberikan peningkatan yang signifikan dalam kualitas pengelompokan. Pendekatan ini membantu peneliti menghindari penggunaan jumlah klaster yang terlalu sedikit atau terlalu banyak [12].
6. Validasi Hasil, Tahap ini bertujuan untuk memastikan bahwa hasil pengelompokan data benar-benar akurat dan dapat dipercaya. Selain mengevaluasi jumlah klaster, peneliti juga meninjau konsistensi karakteristik dalam setiap klaster untuk menghindari terjadinya tumpang tindih (*overlapping*) yang dapat menurunkan kualitas klasterisasi. Pada prinsipnya, data dengan karakteristik serupa seharusnya berada dalam kelompok yang sama, sementara data yang berbeda tidak seharusnya tercampur dalam satu klaster. Dalam proses validasi, jumlah klaster menjadi aspek penting yang diperhatikan. Jika jumlah klaster terlalu sedikit, data yang berbeda dapat terpaksa bergabung dalam satu kelompok. Namun, jika jumlahnya terlalu banyak, hasilnya justru menjadi rumit dan sulit diinterpretasikan. Oleh karena itu, peneliti mencoba beberapa opsi jumlah klaster dan mengevaluasi mana yang paling logis serta paling mampu merepresentasikan pola yang terdapat dalam dataset. Pendekatan ini membantu memastikan bahwa hasil klasterisasi yang diperoleh benar-benar mencerminkan struktur data secara optimal.

DOI : https://doi.org/10.52158/rjti.v4i3.3518

3. Hasil dan Pembahasan

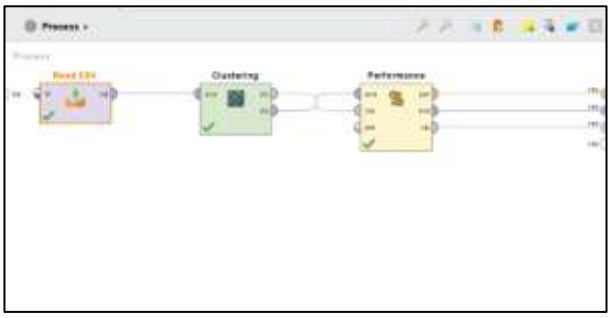
Pada penelitian ini, peneliti memanfaatkan perangkat lunak *RapidMiner* sebagai alat bantu utama untuk mempermudah seluruh proses pengolahan data. *RapidMiner* dimanfaatkan sebagai alat yang membantu pengguna menjalankan berbagai tahap analisis data, mulai dari membersihkan dan mengolah data, melakukan eksplorasi, sampai membangun model prediktif serta analisis statistik. Di dalam *RapidMiner* , algoritma *K-Means* bisa digunakan dengan sangat praktis melalui tampilan *drag-and-drop*, sehingga pengguna dapat menyusun alur kerja sendiri untuk melakukan proses analisis data [13]. Dalam penelitian ini, proses dimulai dari tahap awal, yaitu pengumpulan data dan pembersihan data (pra-proses data), kemudian dilanjutkan dengan penerapan algoritma *K-Means*, dan berakhir pada tahap validasi terhadap klaster yang terbentuk.

Data yang digunakan dalam penelitian ini bersumber dari file bernama diabetes.csv, yang diambil dari situs penyedia dataset untuk keperluan penelitian. Dataset tersebut berisi sekitar 5000 baris data (*record*), yang masing-masing memiliki 9 kolom informasi (atribut). Seluruh data ini kemudian dimasukkan secara manual oleh peneliti ke dalam platform *RapidMiner* untuk diproses lebih lanjut menggunakan algoritma *K-Means*, dengan tujuan untuk menemukan pola atau pengelompokan tertentu dalam data yang berkaitan dengan kondisi diabetes. Tahapan ini dilakukan agar setiap data yang masuk ke proses analisis berada dalam kondisi bersih dan layak digunakan, sehingga hasil klasterisasi yang diperoleh benar-benar mencerminkan keadaan sebenarnya dari pola kesehatan pasien.

Tabel 2. Dataset Diabetes

Jenis Kelamin	Umur	Tekanan Darah Tinggi	Penyakit Jantung	Riwayat Rokok	Standar Berat Badan	Rata-RataKadar Gula Darah	Kadar Gula Darah	Status diabetes
Female	80.0	0	1	never	25.19	6.6	140	0
Female	54.0	0	0	No Info	27.32	6.6	80	0
Male	28.0	0	0	never	27.32	5.7	158	0
Female	36.0	0	0	current	23.45	5.0	155	0
....	....	....	....	....	....	....	....	....
Female	12.0	0	0	never	17.9	4.0	200	0
Male	23.0	0	0	No Info	24.22	6.2	145	0
Female	18.0	0	0	No Info	27.32	6.0	130	0
Male	79.0	0	0	not current	25.77	6.0	130	1
Female	50.0	0	0	never	29.59	6.1	200	0

Sebelum memulai proses pengelompokan data dengan algoritma *K-Means*, langkah awal yang perlu dilakukan adalah menentukan titik pusat atau centroid dari data. Dalam tahap ini, *RapidMiner* Studio secara otomatis menghitung dan menetapkan nilai-nilai centroid tersebut tanpa perlu input manual [14]. Oleh karena itu, peneliti akan terlebih dahulu membuat design view di *RapidMiner* sebagai langkah awal untuk menjalankan proses *clustering* dan melihat nilai *centroid* yang dihasilkan. Penentuan centroid secara otomatis ini membantu memastikan bahwa proses inisialisasi berjalan konsisten, sehingga kemungkinan bias akibat penentuan titik awal *centroid* bisa diminimalkan.

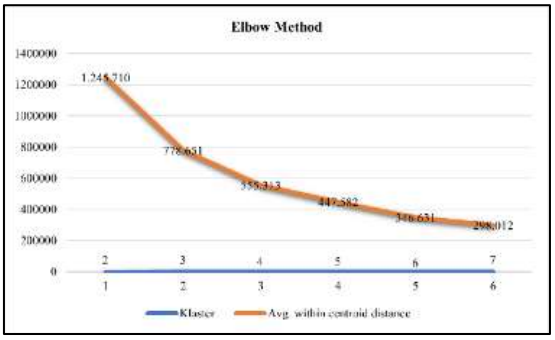


Gambar 2. *Design View*

Keterangan:

1. *Read CSV*: Pada tahap ini, peneliti mengimpor data menggunakan operator *Read CSV* di *RapidMiner* . Dalam proses tersebut, peneliti melakukan pemilihan atribut yang relevan, menetapkan jenis data pada masing-masing kolom (seperti numerik atau kategori), serta melakukan pembersihan awal terhadap data, seperti menghapus baris kosong atau data yang tidak diperlukan. Langkah ini bertujuan agar data yang digunakan sudah siap untuk dianalisis lebih lanjut dalam proses *clustering*.
2. *Cluster* ing (*K-Means*): Setelah data berhasil diimpor dan dipersiapkan, peneliti melanjutkan ke tahap *cluster* ing dengan menggunakan algoritma *K-Means*. Proses ini dilakukan dengan bantuan operator *Cluster* ing (*K-Means*) di *RapidMiner* . Pada tahap ini, peneliti menentukan jumlah klaster yang diinginkan. Disini peneliti menggunakan metode *Elbow* untuk menentukan jumlah klaster, mulai dari mencoba klaster 2 sampai 7. Metode *Elbow* digunakan karena mampu memberikan gambaran visual yang jelas mengenai titik optimal jumlah klaster, sehingga keputusan yang diambil tidak sekadar berdasarkan perkiraan, tetapi melalui pertimbangan pola penurunan nilai *WCSS (Within-Cluster Sum of Squares).*

DOI : https://doi.org/10.52158/rjti.v4i3.3518



Gambar 3. *Elbow Method*

Hasil dari metode *Elbow* menunjukkan bahwa nilai rata-rata jarak paling signifikan turun pada k = 3, dan penurunan mulai melandai setelah itu. Oleh karena itu, untuk proses pengelompokan data menggunakan algoritma *K-Means*, k = 3. Temuan ini menunjukkan bahwa struktur data cenderung terbagi ke tiga kelompok dominan, yang kemudian menjadi dasar dalam pemilihan jumlah klaster yang paling sesuai untuk dataset diabetes ini.

Gambar 4. Jumlah klaster

3. Performance: Setelah proses *clustering* selesai dilakukan, peneliti menggunakan operator *Performance* untuk mengevaluasi hasil pengelompokan data. Operator ini membantu dalam menilai kualitas *clustering* yang telah dilakukan, misalnya dengan memperhatikan kedekatan data dalam satu klaster atau penyebaran antar klaster.

Setelah semua proses selesai dijalankan, *RapidMiner* akan menampilkan beberapa output hasil, yaitu:

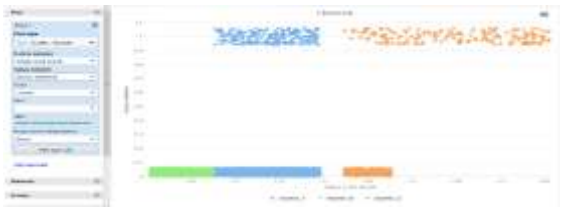
1. *ExampleSet*, menyajikan data asli yang telah diproses, dengan tambahan atribut baru yang menunjukkan hasil pengelompokan, penambahan atribut *cluster* ini memudahkan peneliti mengamati perubahan pada data serta menghubungkan setiap kelompok dengan karakteristik kesehatan tertentu.

Gambar 5. *ExampleSet*

2. *Davies Bouldin Index (DBI),* adalah salah satu teknik validasi yang umum dipakai untuk melihat seberapa efektif hasil pengelompokan data yang dihasilkan oleh algoritma *clustering* [15]. Berdasarkan hasil perhitungan, nilai Davies-Bouldin Index (DBI) yang diperoleh adalah 0,827. Nilai ini menunjukkan bahwa kualitas pemisahan antar-*cluster* sudah cukup baik. Semakin kecil nilai DBI, semakin optimal hasil pengelompokan karena setiap *cluster* memiliki jarak yang jelas satu sama lain dan anggota di dalamnya relatif homogen. Dengan nilai 0,827, dapat dikatakan bahwa model klasterisasi yang digunakan telah menghasilkan pemisahan *cluster* yang cukup stabil dan dapat dijadikan acuan untuk analisis lebih lanjut.

Gambar 6. Davies Bouldin

3. *Scatter*, adalah diagram statistik yang menggunakan sistem koordinat kartesius untuk memperlihatkan nilai dari dua variabel dalam satu kumpulan data, guna menunjukkan sejauh mana satu variabel dipengaruhi oleh variabel lainnya [16]. Berikut *scatter* diagram yang dihasilkan, yaitu:



Gambar 7. *Scatter* Diagram kelompok pasien berdasarkan kadar gula yang dimiliki

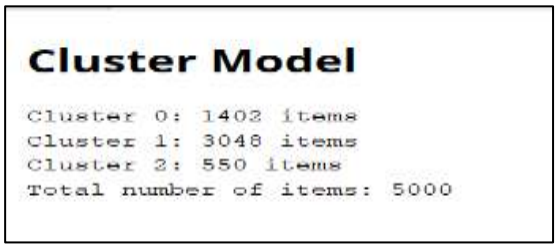
Keterangan: Berdasarkan visualisasi data menggunakan *scatter* diagram, terlihat adanya pola yang cukup jelas antara kadar gula darah pasien, status diabetes, dan hasil pengelompokan

DOI : https://doi.org/10.52158/rjti.v4i3.3518

berdasarkan algoritma *K-Means*. Setiap pasien dikelompokkan ke dalam salah satu dari tiga *cluster* , yakni *cluster* 0, *cluster* 1, dan *cluster* 2. Berdasarkan kesamaan karakteristik tertentu, salah satunya adalah kadar gula darah. Pada grafik tersebut, sumbu horizontal (X-axis) menunjukkan kadar gula darah setiap pasien, sedangkan sumbu vertikal (Y-axis) menunjukkan status diabetes, di mana nilai 1 berarti pasien terdiagnosis diabetes, dan 0 berarti tidak.

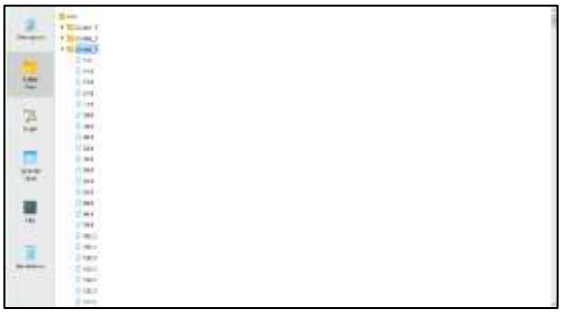
Warna titik-titik pada grafik menunjukkan *cluster* masing-masing pasien, sehingga pola distribusinya dapat dianalisis lebih mudah. *Cluster* 0 didominasi oleh pasien dengan kadar gula darah yang relatif rendah hingga sedang, dan sebagian besar dari mereka memiliki status diabetes negatif (0). Ini mengindikasikan bahwa kelompok ini cenderung memiliki risiko diabetes yang rendah. *Cluster* 1 berisi pasien dengan kadar gula darah dalam kisaran sedang hingga tinggi, dan sebagian besar dari mereka sudah terdiagnosis menderita diabetes (status 1). Oleh karena itu, *cluster* ini menggambarkan kelompok dengan risiko tinggi, bahkan sudah berada dalam fase penyakit diabetes. Sementara itu, *cluster* 2 mencakup pasien dengan kadar gula darah yang sangat tinggi, namun menariknya, sebagian besar dari mereka belum terdiagnosis diabetes (status 0). Hal ini bisa menunjukkan dua kemungkinan penting: Mereka berpotensi tinggi untuk mengembangkan diabetes dalam waktu dekat (pre-diabetes), atau ada keterlambatan dalam proses diagnosis. Dengan demikian, *cluster* 2 merupakan kelompok dengan risiko paling tinggi dan perlu mendapatkan perhatian medis lebih lanjut meskipun belum terdiagnosis secara resmi. Visualisasi ini membantu memperjelas perbedaan antar klaster, sehingga interpretasi terhadap risiko diabetes dapat dilakukan dengan lebih terarah dan dapat dipertanggungjawabkan secara analitis.

4. *Cluster* Model, yang memperlihatkan bahwa data terbagi ke dalam tiga kelompok klaster, yaitu:
   a. *Cluster* 0: Terdiri dari 1402 items
   b. *Cluster* 1: Terdiri dari 3048 items
   c. *Cluster* 2: Terdiri dari 550 items



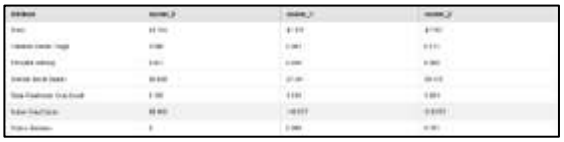
Gambar 8. *Cluster* Model

5. *Folder View,* menampilkan data dari setiap kelompok atau *cluster* akan diperlihatkan secara lengkap dan menyeluruh. Artinya, semua anggota yang termasuk dalam satu *cluster* akan ditampilkan satu per satu sesuai dengan hasil pengelompokan yang sudah dilakukan sebelumnya. Dengan cara ini, kita bisa melihat dengan jelas siapa saja yang tergabung dalam masing-masing *cluster* dan bagaimana karakteristik atau nilai-nilai mereka berdasarkan data yang tersedia. Tampilan ini juga mempermudah pengecekan ulang terhadap distribusi data pada tiap klaster, sehingga proses verifikasi dapat dilakukan lebih menyeluruh.



Gambar 9. Tampilan kelompok data *cluster* berdasarkan Folder View

6. *Centroid* Data, yang digunakan melalui tabel *centroid*. Berikut adalah data *centroid* yang dihasilkan:



Gambar 10. *Centroid Table*

Berdasarkan Gambar 10. Centroid Table, bisa disimpulkan bahwa data tiga *cluster* yang sudah didapatkan, memiliki tingkat risikonya masing-masing, yaitu:

a. *Cluster* 0: Kelompok ini dikategorikan sebagai klaster dengan tingkat risiko paling rendah terhadap diabetes. Rata-rata kadar gula darah individu dalam kelompok ini adalah 88,406 mg/dL, yang termasuk dalam kategori normal menurut standar medis. Nilai ini menunjukkan bahwa mayoritas anggota klaster memiliki kadar glukosa dalam darah yang masih dalam batas sehat dan belum menunjukkan tanda-tanda gangguan metabolisme gula. Selain itu, usia rata-rata individu dalam klaster ini berada di sekitar 40 tahun, yang tergolong usia produktif dan relatif muda dalam konteks risiko penyakit kronis. Prevalensi tekanan darah tinggi dan penyakit jantung dalam kelompok ini juga sangat rendah, yaitu masing-masing hanya sekitar 4,9% dan 2,1% dari total anggota klaster. Hal ini mengindikasikan bahwa gaya hidup dan kondisi kesehatan mereka secara umum masih tergolong baik dan tidak banyak faktor komorbiditas yang

DOI : https://doi.org/10.52158/rjti.v4i3.3518

berkontribusi terhadap risiko diabetes. Menariknya, nilai status diabetes pada klaster ini adalah nol, yang berarti tidak ditemukan satu pun individu yang terdiagnosis diabetes dalam kelompok ini. Dengan demikian, klaster 0 dapat dianggap sebagai kelompok paling sehat dan paling tidak rentan terhadap risiko diabetes di antara ketiga klaster yang dianalisis.

b. *Cluster* 1: Menunjukkan karakteristik yang mengarah pada risiko menengah terhadap diabetes. Rata-rata kadar gula darah dalam kelompok ini adalah sebesar 146,537 mg/dL, yang telah memasuki kategori pra-diabetes atau bahkan tahap awal dari diabetes. Angka ini menunjukkan bahwa kadar glukosa darah pada individu dalam klaster ini sudah mulai menunjukkan ketidakseimbangan dan berpotensi berkembang menjadi diabetes tipe 2 jika tidak segera ditangani. Dari segi demografi, usia rata-rata individu dalam klaster ini adalah sekitar 41,9 tahun, sedikit lebih tua dibandingkan klaster 0. Peningkatan usia sering dikaitkan dengan meningkatnya risiko penyakit kronis, termasuk diabetes. Selain itu, proporsi penderita tekanan darah tinggi dan penyakit jantung dalam klaster ini juga mengalami peningkatan dibanding klaster 0, masing-masing sebesar 8,1% dan 3,9%. Ini menunjukkan bahwa sebagian individu dalam kelompok ini mulai mengalami komplikasi kesehatan yang dapat memperburuk risiko diabetes. Nilai status diabetes pada klaster ini adalah 0,080, yang mengindikasikan bahwa sekitar 8% anggota klaster sudah mengalami diabetes. Meskipun bukan mayoritas, angka ini cukup signifikan sebagai sinyal peringatan dini untuk melakukan intervensi gaya hidup dan pemantauan medis secara rutin.

c. *Cluster* 2: Memiliki kelompok dengan tingkat risiko paling tinggi terhadap diabetes. Rata-rata kadar gula darah yang sangat tinggi, yaitu sebesar 218,073 mg/dL, secara jelas mengindikasikan bahwa kelompok ini telah masuk dalam kategori diabetes menurut standar diagnosis medis. Kondisi ini menunjukkan adanya gangguan metabolisme glukosa yang serius di antara mayoritas anggota klaster. Dari segi usia, usia rata-rata individu dalam klaster ini adalah 47,66 tahun, tertinggi dibandingkan dua klaster lainnya. Usia yang lebih lanjut seringkali berhubungan dengan peningkatan risiko penyakit kronis, termasuk diabetes, hipertensi, dan penyakit jantung. Hal ini diperkuat dengan tingginya prevalensi tekanan darah tinggi (13,1%) dan penyakit jantung (6,9%) di antara individu dalam kelompok ini, yang menunjukkan adanya faktor komorbiditas yang signifikan. Nilai status diabetes yang mencapai 0,351 berarti bahwa lebih dari sepertiga (sekitar 35,1%) individu dalam klaster ini telah secara medis terdiagnosis diabetes. Fakta ini menunjukkan bahwa klaster 2 merupakan kelompok dengan kondisi kesehatan yang paling rentan dan membutuhkan perhatian medis serta pengelolaan gaya hidup yang sangat serius.

Hasil analisis centroid ini memperlihatkan bahwa setiap klaster memiliki ciri khas yang membedakan satu sama lain, sehingga pembagian kelompok dapat digunakan sebagai dasar penyusunan strategi mitigasi risiko diabetes yang lebih spesifik. Informasi ini juga memberikan gambaran awal mengenai pola kesehatan populasi dalam dataset, terutama terkait faktor-faktor yang berpengaruh terhadap peningkatan kadar gula darah. Secara keseluruhan, terdapat tiga *cluster* utama yang masing-masing menunjukkan karakteristik kesehatan yang berbeda. *cluster* 0 terdiri dari sekitar 45% individu yang kondisinya masih tergolong sehat atau normal. Pada kelompok ini, kadar gula darah serta faktor risiko lain seperti tekanan darah dan indeks massa tubuh berada dalam batas aman, menunjukkan bahwa mereka belum menunjukkan tanda-tanda risiko diabetes yang signifikan. *cluster* 1, yang mencakup sekitar 35% dari populasi, menggambarkan kelompok dengan gejala awal atau tanda-tanda pra-diabetes.

Dalam kelompok ini, kadar gula darah mulai meningkat meskipun belum mencapai level diabetes penuh, dan beberapa faktor risiko mulai muncul. Kelompok ini memerlukan perhatian lebih lanjut dan pemantauan agar kondisi tidak memburuk. Sedangkan *cluster* 2 meliputi sekitar 20% individu yang menunjukkan risiko tinggi dengan kadar gula darah yang sudah cukup tinggi dan disertai faktor risiko lain seperti obesitas atau hipertensi. Kondisi mereka sudah lebih serius dan membutuhkan penanganan medis yang lebih intensif serta pengelolaan kesehatan yang ketat. Dengan adanya pemahaman ini, langkah pencegahan dan pengelolaan diabetes dapat dilakukan secara lebih terarah dan sesuai dengan kebutuhan masing-masing kelompok, sehingga upaya kesehatan dapat lebih efektif dan tepat sasaran.

4. Kesimpulan dan Saran

Penelitian ini membahas penerapan algoritma *K-Means*, yang termasuk dalam metode unsupervised learning, untuk mengelompokkan pasien berdasarkan risiko terhadap penyakit diabetes. Diabetes merupakan gangguan metabolik serius yang jika tidak ditangani sejak dini dapat memicu komplikasi seperti penyakit jantung, stroke, dan kerusakan organ lainnya. Dalam penelitian ini, digunakan dataset dari situs yang berisi 5000 data pasien dengan atribut seperti umur, jenis kelamin, kadar gula darah, tekanan darah tinggi, dan riwayat merokok. Proses analisis dilakukan menggunakan aplikasi *RapidMiner* melalui beberapa tahapan, mulai dari pencarian dan pra-pemrosesan data, penerapan algoritma *K-Means*, visualisasi hasil, evaluasi

DOI : https://doi.org/10.52158/rjti.v4i3.3518

jumlah klaster dengan metode *Elbow*, hingga validasi klasterisasi.

Hasil pengelompokan menghasilkan tiga *cluster* , yaitu *cluster* pertama terdiri dari pasien dengan risiko rendah (45%), yang memiliki kadar gula darah normal dan tidak ada yang terdiagnosis diabetes. *Cluster* kedua menunjukkan kelompok dengan risiko menengah (35%) yang mulai menunjukkan gejala pra-diabetes serta peningkatan kadar gula darah dan beberapa faktor risiko lain yang perlu diperhatikan dan dipantau lebih lanjut. Sedangkan *cluster* ketiga berisi pasien dengan risiko tinggi (20%) yang memiliki kadar gula darah sangat tinggi, di mana sebagian besar sudah berada dalam fase diabetes dan menghadapi kondisi yang lebih serius, meskipun ada juga beberapa yang belum terdiagnosis secara resmi dan membutuhkan penanganan medis yang lebih intensif.

Hasil penelitian ini menunjukkan bahwa algoritma *K-Means* berpotensi diterapkan dalam sistem pendukung keputusan medis untuk membantu menyaring pasien yang membutuhkan perhatian lebih lanjut. Dengan klasifikasi berbasis karakteristik kesehatan, tenaga medis dapat lebih mudah memprioritaskan pasien dengan risiko tinggi untuk intervensi dini. Pendekatan ini tidak hanya meningkatkan efisiensi pengambilan keputusan klinis, tetapi juga dapat dimanfaatkan untuk merancang program pencegahan yang lebih terarah. Sebagai tambahan, hasil klasterisasi juga memberikan gambaran terstruktur mengenai variasi profil kesehatan pasien, sehingga mempermudah identifikasi dini risiko diabetes berdasarkan data objektif. Interpretasi terhadap *cluster* ketiga mengindikasikan potensi pra-diabetes atau kemungkinan keterlambatan diagnosis, sehingga hasil ini dapat menjadi dasar untuk penguatan deteksi dini di fasilitas layanan kesehatan. Dengan integrasi lebih lanjut terhadap data tambahan atau sistem monitoring digital, model seperti ini memiliki potensi besar untuk dikembangkan sebagai bagian dari sistem pemantauan kesehatan yang lebih cerdas dan adaptif di masa mendatang.

Berdasarkan hasil yang diperoleh, peneliti menyadari bahwa masih banyak hal yang bisa dikembangkan dari penelitian ini. Untuk itu, peneliti menyarankan agar penelitian selanjutnya bisa menggunakan data yang lebih banyak dan beragam, misalnya dengan mengambil data dari berbagai daerah dan latar belakang pasien yang berbeda. Tujuannya agar hasil pengelompokan menjadi lebih akurat dan bisa digunakan secara lebih luas. Selain itu, penggunaan algoritma *K-Means* juga bisa dibandingkan dengan metode lainnya, seperti *DBSCAN* atau *Gaussian Mixture* Model, agar dapat diketahui metode mana yang paling efektif dalam mengelompokkan data pasien. Penelitian selanjutnya juga sebaiknya mencoba menggabungkan algoritma dengan sistem prediksi berdasarkan waktu, seperti melihat perubahan kadar gula darah dari waktu ke waktu, sehingga potensi risiko diabetes bisa terdeteksi lebih awal. Tidak kalah penting, akan lebih baik jika data medis dikombinasikan dengan informasi tentang gaya hidup pasien, seperti kebiasaan makan, olahraga, dan pola tidur. Dengan begitu, sistem yang dikembangkan di masa depan diharapkan tidak hanya bisa mengelompokkan pasien berdasarkan data kesehatan saja, tapi juga dapat memberikan saran pencegahan yang lebih tepat dan sesuai dengan kondisi masing-masing pasien. Tambahan variabel ini juga memungkinkan model menghasilkan rekomendasi yang lebih personal dan relevan, sehingga hasil penelitian dapat berkontribusi lebih besar bagi upaya pencegahan dan pengelolaan diabetes secara komprehensif.

Daftar Rujukan

[1] N. Sunanto and G. Falah, “Penerapan Algoritma C4.5 Untuk Membuat Model Prediksi Pasien Yang Mengidap Penyakit Diabetes,” Rabit J. Teknol. dan Sist. Inf. Univrab, vol. 7, no. 2, pp. 208–216, 2022, doi: 10.36341/rabit.v7i2.2435.

[2] A. E. Satriatama et al., “Analisis Klaster Data Pasien Diabetes untuk Identifikasi Pola dan Karakteristik Pasien,” J. Teknol. Dan Sist. Inf. Bisnis, vol. 5, no. 3, pp. 172–182, 2023, doi: 10.47233/jteksis.v5i3.828.

[3] A. M. Argina, “Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes,” Indones. J. Data Sci., vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.

[4] K. Ogurtsova et al., “IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021,” Diabetes Res. Clin. Pract., vol. 183, 2022, doi: 10.1016/j.diabres.2021.109118.

[5] C. Wang et al., “Unsupervised *cluster* analysis of clinical and metabolite characteristics in patients with chronic complications of T2DM: an observational study of real data,” Front. Endocrinol. (Lausanne)., vol. 14, no. October, pp. 1–12, 2023, doi: 10.3389/fendo.2023.1230921.

[6] W. Aulia, A. Putera Utama Siahaan, L. Marlina, and M. Iqbal, “Analisis Algoritma *K-Means Cluster* ing Dalam Identifikasi Tingkat Risiko Penyakit Berdasarkan Data Rekam Medis Pasien,” J. Sci. Soc. Res., vol. 4307, no. 3, pp. 3457–3465, 2025, [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR

[7] Wijoyo A, Saputra A, Ristanti S, Sya’ban S, Amalia M, and Febriansyah R, “Pembelajaran Machine Learning,” OKTAL (Jurnal Ilmu Komput. dan Sci., vol. 3, no. 2, pp. 375–380, 2024, [Online]. Available: https://journal.mediapublikasi.id/index.php/oktal/article/view/2305

[8] M. Pandia, “Kajian Literatur Multimedia Retrieval : Machine Learning Untuk Pengenalan Wajah,” J. Ilmu Komput. dan Sist. Inf., vol. 7, no. 1, pp. 161–166, 2024, doi: 10.55338/jikomsi.v7i1.2758.

[9] E. Retnoningsih and R. Pramudita, “Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python,” Bina Insa. Ict J., vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.

[10] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, “Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner,” JBASE - J. Bus. Audit Inf. Syst., vol. 4, no. 1, pp. 13–21, 2021, doi: 10.30813/jbase.v4i1.2729.

[11] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” J. KomtekInfo, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.

[12] M. Qusyairi, Z. Hidayatullah, A. Sandi, and V. No, “Infotek : Jurnal Informatika dan Teknologi Penerapan *K-Means Cluster* ing Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode *Elbow* Infotek : Jurnal Informatika dan Teknologi Perkembangan teknologi saat ini berkembang dengan sangat pesat ini terbukti,” vol. 7, no. 2, pp. 500–510, 2024.

DOI : https://doi.org/10.52158/rjti.v4i3.3518

[13] A. I. Silitonga, Z. A. Nabila, C. R. Z. Lubis, N. Safitri, and H. Haryadi, "Klasterisasi Gizi Buruk Dan Stunting Di Provinsi Sumatera Utara Menggunakan *K-Means Cluster* ing," Method. J. Tek. Inform. dan Sist. Inf., vol. 10, no. 2, pp. 13–18, 2024, doi: 10.46880/mtk.v10i2.3147.

[14] Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, "Penerapan Algoritma *K-Means* Untuk *Cluster* ing Data Kemiskinan Provinsi Banten Menggunakan *RapidMiner* ," CESS (Journal Comput. Eng. Syst. Sci., vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess.v5i2.18519.

[15] L. Hanum, "Pengelompokan Gaya Belajar Mahasiswa Menggunakan Metode *K-Means* dan Validasi Menggunakan Davies Bouldin Index," J. J-MendiKKom (Jurnal Manajemen, Pendidik. dan Ilmu Komputer), vol. 2, no. 1, 2025.

[16] Lukman, Rachmasari Pramita Wardhani, Selvia Sarungu, and Irma Andrianti, "Penggunaan Metode Seven Tool Dengan Diagram Scatter Dalam Pembelajaran Pengendalian Mutu Secara Statistik," J. Teknosains Kodepena, vol. 5, no. 1, pp. 27–33, 2024, doi: 10.54423/teknosains.v5i1.81.

DOI : https://doi.org/10.52158/rjti.v4i3.3518