



## Evaluasi Performa Deteksi Penyakit Diabetes Dengan Fuzzy C-Means Dan K-Means Clustering

**Roy Efendi Subarja**

Universitas Graha Nusantara Padangsidempuan

Email: [subardjaroy@gmail.com](mailto:subardjaroy@gmail.com)

**Billy Hendrik**

Universitas Putra Indonesia "YPTK" Padang

Email: [billy\\_hendrik@upiypk.ac.id](mailto:billy_hendrik@upiypk.ac.id)

Korespondensi penulis: [subardjaroy@gmail.com](mailto:subardjaroy@gmail.com)

**Abstract.** *The increasing prevalence of diabetes has led to a growing need for accurate and efficient disease detection methods. This research focuses on evaluating the performance of diabetes detection using Fuzzy C-Means and K-Means clustering algorithms. The study aims to compare the effectiveness of these two clustering techniques in identifying diabetes cases based on relevant medical data. A dataset comprising various health parameters and diagnostic indicators was utilized for experimentation. The Fuzzy C-Means and K-Means algorithms were implemented to cluster the dataset, and their detection performance was assessed using metrics such as sensitivity, specificity, accuracy, and F1-score. The results indicate that both clustering methods exhibit promising potential for diabetes detection, with variations in their performance based on different evaluation criteria. This research contributes to a deeper understanding of the applicability of clustering algorithms in diabetes detection and provides insights into their strengths and limitations. Further optimization and validation of these algorithms could lead to enhanced diagnostic accuracy and early intervention in diabetes management.*

**Keywords:** *Diabetes, Fuzzy C-Means, K-Means, Clustering*

**Abstrak.** Peningkatan prevalensi diabetes telah mendorong kebutuhan akan metode deteksi penyakit yang akurat dan efisien. Penelitian ini berfokus pada evaluasi performa deteksi diabetes menggunakan algoritma Fuzzy C-Means dan K-Means clustering. Tujuan dari penelitian ini adalah membandingkan efektivitas kedua teknik pengelompokan ini dalam mengidentifikasi kasus diabetes berdasarkan data medis yang relevan. Dataset yang terdiri dari berbagai parameter kesehatan dan indikator diagnostik digunakan untuk eksperimen. Algoritma Fuzzy C-Means dan K-Means diimplementasikan untuk mengelompokkan dataset, dan performa deteksinya dievaluasi menggunakan metrik seperti sensitivitas, spesifisitas, akurasi, dan skor F1. Hasil penelitian menunjukkan bahwa kedua metode pengelompokan menunjukkan potensi yang menjanjikan untuk deteksi diabetes, dengan variasi performa berdasarkan kriteria evaluasi yang berbeda. Penelitian ini memberikan pemahaman lebih mendalam tentang aplikabilitas algoritma pengelompokan dalam deteksi diabetes dan memberikan wawasan tentang kelebihan dan keterbatasan masing-masing. Optimisasi dan validasi lebih lanjut dari algoritma-algoritma ini dapat menghasilkan akurasi diagnostik yang lebih baik dan intervensi dini dalam pengelolaan diabetes.

**Kata kunci:** *Diabetes, Fuzzy C-Means, K-Means, Clustering*

### LATAR BELAKANG

Diabetes merupakan salah satu masalah kesehatan global yang signifikan. Prevalensinya terus meningkat dan menjadi beban besar bagi sistem kesehatan serta kualitas hidup individu. Deteksi dini dan penanganan yang tepat sangat penting dalam mengurangi komplikasi dan dampak negatif yang disebabkan oleh penyakit ini. Dalam upaya untuk meningkatkan deteksi dini diabetes, banyak penelitian telah dilakukan untuk mengembangkan metode-metode baru yang lebih akurat dan efisien. Salah satu pendekatan yang telah digunakan

adalah penggunaan teknik pengelompokan data, seperti Fuzzy C-Means dan K-Means clustering. Teknik-teknik ini dapat membantu mengidentifikasi pola-pola yang tersembunyi dalam data klinis dan memisahkan individu dengan risiko tinggi diabetes. Namun, meskipun telah ada usaha dalam menerapkan pengelompokan data untuk deteksi diabetes, masih ada kebutuhan untuk mengevaluasi secara lebih mendalam performa dan efektivitas dari kedua teknik pengelompokan ini dalam konteks spesifik deteksi penyakit ini. Evaluasi ini akan membantu mengidentifikasi keunggulan dan batasan masing-masing metode, serta memberikan panduan bagi praktisi medis dalam memilih pendekatan yang paling sesuai. Oleh karena itu, penelitian ini bertujuan untuk melakukan evaluasi performa deteksi penyakit diabetes menggunakan algoritma Fuzzy C-Means dan K-Means clustering. Dengan menggali lebih dalam tentang kemampuan kedua metode ini dalam mengenali pola-pola yang berkaitan dengan diabetes pada data klinis, diharapkan penelitian ini dapat memberikan kontribusi dalam pengembangan strategi deteksi dini yang lebih efektif dan akurat dalam penanganan diabetes.

Jadi metode K-Means akan mengelompokkan suatu data yang menjadi anggota salah satu cluster yang memenuhi syarat, dan data tersebut tidak menjadi anggota cluster yang lainnya. Namun, pada metode Fuzzy C-Means Clustering data ditentukan dengan derajat keanggotaan yang terbesar, dengan kata lain data dapat menjadi anggota lebih dari satu cluster. Berasal dari latar belakang tes gula darah yang harus dilakukan beberapa kali untuk mendapatkan hasil yang akurat dan keterbatasan tenaga medis untuk menentukan deteksi pada penyakit diabetes pada beberapa indikator dengan data yang banyak, maka penelitian ini akan merancang aplikasi deteksi penyakit diabetes dengan metode Fuzzy C-Means Clustering dan K-Means Clustering.

## **KAJIAN TEORITIS**

Diabetes mellitus, atau diabetes, adalah suatu penyakit kronis yang ditandai oleh kadar glukosa darah yang tinggi akibat gangguan dalam produksi insulin atau respons tubuh terhadap insulin. Diabetes memiliki beberapa tipe, yang paling umum adalah tipe 1 dan tipe 2. Tipe 1 melibatkan kerusakan sel-sel pankreas yang menghasilkan insulin, sedangkan tipe 2 melibatkan resistensi insulin dan produksi insulin yang tidak memadai. Deteksi dini diabetes sangat penting karena penyakit ini dapat menyebabkan komplikasi serius, seperti penyakit jantung, gagal ginjal, dan gangguan mata.

Deteksi dini melibatkan pengukuran glukosa darah, tes toleransi glukosa oral, dan pengukuran hemoglobin A1c (HbA1c) sebagai indikator pengendalian gula darah dalam jangka waktu tertentu. Selain itu, analisis data medis dapat menjadi alat yang kuat dalam deteksi dini

diabetes. Penggunaan teknik pengolahan data seperti Fuzzy C-Means dan K-Means clustering dapat membantu mengidentifikasi pola-pola yang mungkin tidak terlihat secara langsung oleh dokter. Dalam hal ini, Fuzzy C-Means memungkinkan adanya overlap antara kelompok, sedangkan K-Means menghasilkan kelompok yang saling eksklusif. Evaluasi performa dari metode pengelompokan ini melibatkan pengukuran sensitivitas (kemampuan mendeteksi kasus positif), spesifisitas (kemampuan mendeteksi kasus negatif), akurasi, dan skor F1. Evaluasi ini penting untuk memahami seberapa baik kedua metode ini dalam mengklasifikasikan individu sebagai risiko diabetes.

- a. Penderita diabetes mellitus, baik tipe 1 maupun tipe 2, dapat memiliki sejumlah ciri-ciri khas yang berkaitan dengan perubahan metabolisme dan gangguan dalam regulasi glukosa darah. Namun, penting untuk diingat bahwa tidak semua penderita diabetes akan mengalami semua ciri-ciri ini, dan gejala dapat bervariasi antara individu.

Berikut adalah beberapa ciri-ciri umum penderita diabetes:

- a) Polifagia (Nafsu Makan Berlebihan)
- b) Polidipsia (Dahaga Berlebihan)
- c) Poluri (Produksi Urine Berlebihan)
- d) Penurunan berat badan tanpa sebab yang jelas
- e) Kelelahan
- f) Penglihatan kabur
- g) Tekanan darah tinggi

Selain beberapa ciri-ciri fisik, penyakit diabetes juga dapat diindikasikan dengan menggunakan tes toleransi kandungan glukosa dalam darah di laboratorium atau Glucose Tolerance Test, indikatornya dapat dilihat pada tabel dibawah:

**Tabel 1.** Hasil Glucose Tolerance Test pada indikasi penyakit diabetes

<b>Kondisi Darah yang Diambil</b>	<b>Prediabetes</b>	<b>Diabetes</b>
Puasa	100 – 125 mg/dL	➤ 126 mg/dL
Setelah 2 Jam	140 – 199 mg/dL	➤ 200 mg/dL

- b. Tujuan Rancangan

Tujuan utama dari rancangan penelitian ini adalah untuk menguji dan membandingkan performa deteksi penyakit diabetes menggunakan algoritma Fuzzy C-Means dan K-Means clustering. Rancangan ini memiliki beberapa tujuan dan kegunaan yang meliputi:

- **Evaluasi Algoritma Pengelompokan:** Rancangan ini bertujuan untuk mengevaluasi efektivitas dan kinerja dari dua metode pengelompokan, yaitu Fuzzy C-Means dan K-Means, dalam mengidentifikasi kasus diabetes. Dengan mengadopsi pendekatan eksperimen, penelitian ini akan memberikan pemahaman mendalam tentang bagaimana kedua metode ini dapat digunakan dalam konteks deteksi diabetes.
- **Pengembangan Metode Deteksi Dini:** Melalui pengujian performa dan perbandingan kedua metode, penelitian ini bertujuan untuk memberikan panduan bagi para profesional medis dan peneliti dalam memilih metode yang paling sesuai untuk mendeteksi dini kasus diabetes. Hasil penelitian ini dapat membantu pengembangan strategi deteksi dini yang lebih efisien dan akurat.
- **Kontribusi dalam Manajemen Diabetes:** Dengan memahami performa kedua metode pengelompokan, penelitian ini dapat memberikan kontribusi dalam pengembangan pendekatan baru dalam manajemen diabetes. Deteksi dini yang lebih baik dapat membantu dalam pencegahan komplikasi dan perencanaan perawatan yang lebih tepat bagi pasien diabetes.
- **Validasi Teknik Analisis Data:** Rancangan ini dapat memberikan validasi empiris terhadap efektivitas Fuzzy C-Means dan K-Means clustering dalam deteksi diabetes. Ini dapat membantu menguatkan landasan teoritis dan praktis penggunaan kedua metode dalam analisis data klinis.
- **Kontribusi pada Literatur Ilmiah:** Hasil penelitian ini dapat berkontribusi pada literatur ilmiah dalam bidang deteksi penyakit menggunakan teknik pengelompokan data. Informasi yang dihasilkan dapat digunakan oleh peneliti lain dalam mengembangkan penelitian lebih lanjut tentang deteksi penyakit lainnya menggunakan pendekatan serupa.

Kegunaan dari sistem ini adalah mengelompokkan data penderita diabetes dan non diabetes dari dataset. Sehingga pada hasil prosesnya didapat dua nilai centroid. Nilai centroid tersebut adalah nilai titik tengah data diabetes dan non diabetes yang menjadi acuan pada data

pengelompokan. Sepuluh baris pertama dataset penderita diabetes dan non-diabetes yang digunakan adalah:

**Tabel 2.** Barisan Dataset

A	B	C	D	E	F	G	H	Result	Detail
8	133	72	0	0	32.9	0.27	39	1	Diabetes
3	158	76	36	245	31.6	0.851	28	1	Diabetes
11	103	68	40	0	46.2	0.126	42	0	Non Diabetes
5	99	54	28	83	34	0.499	30	0	Non Diabetes
5	139	80	35	160	31.6	0.361	25	1	Diabetes
1	89	66	23	94	28.1	0.167	21	0	Non Diabetes
2	155	52	27	540	38.7	0.24	25	1	Diabetes
0	125	96	0	0	22.5	0.262	21	0	Non Diabetes
6	104	74	18	156	29.9	0.722	41	1	Diabetes
6	96	0	0	0	23.7	0.19	28	0	Non Diabetes

## METODE PENELITIAN

Metode penelitian yang akan digunakan dalam evaluasi performa deteksi penyakit diabetes dengan menggunakan Fuzzy C-Means dan K-Means clustering merupakan pendekatan empiris yang melibatkan beberapa langkah penting untuk mengumpulkan, menganalisis, dan mengevaluasi data. Berikut adalah deskripsi rinci tentang metode penelitian yang akan diimplementasikan:

- a. **Pengumpulan Data:** Data medis yang relevan akan dikumpulkan dari sumber yang sah, seperti catatan medis pasien, basis data klinik, atau sumber lain yang berkaitan. Data ini akan mencakup berbagai parameter kesehatan seperti kadar glukosa darah, berat badan, tekanan darah, riwayat keluarga, dan faktor risiko lainnya yang dapat berperan dalam deteksi diabetes.
- b. **Preprocessing Data:** Data yang terkumpul akan diproses untuk memastikan kualitas dan konsistensi. Langkah-langkah preprocessing meliputi penanganan missing values dengan pengisian data yang hilang, normalisasi data untuk mengubah skala variabel menjadi seragam, dan penghapusan outlier yang dapat mempengaruhi analisis.
- c. **Pemilihan Jumlah Cluster (k):** Sebelum menerapkan algoritma pengelompokan, jumlah kluster (k) perlu ditentukan. Ini dapat dilakukan melalui pendekatan seperti Elbow Method, Silhouette Analysis, atau berdasarkan pengetahuan domain.

- d. **Implementasi Algoritma Fuzzy C-Means dan K-Means:** Kedua algoritma akan diimplementasikan menggunakan bahasa pemrograman atau perangkat lunak analisis data. Data yang telah dipreprocessing akan dianalisis menggunakan Fuzzy C-Means untuk menghasilkan keanggotaan setiap data pada setiap kluster, serta menggunakan K-Means untuk mengelompokkan data ke dalam kluster eksklusif.
- e. **Evaluasi Performa:** Hasil dari kedua algoritma akan dievaluasi menggunakan metrik evaluasi performa seperti sensitivitas, spesifisitas, akurasi, dan skor F1. Metrik-metrik ini akan memberikan gambaran tentang kemampuan algoritma dalam mengklasifikasikan kasus diabetes berdasarkan data medis yang ada.
- f. **Validasi Silang (Cross-Validation):** Untuk memastikan generalisasi hasil, metode validasi silang akan diterapkan. Data akan dibagi menjadi subset untuk pelatihan dan pengujian, sehingga algoritma akan diuji pada data yang tidak digunakan dalam pelatihan.
- g. **Analisis Hasil dan Interpretasi:** Data hasil evaluasi performa dan validasi silang akan dianalisis secara mendalam. Hasil perbandingan antara Fuzzy C-Means dan K-Means serta interpretasi signifikansi hasil akan disajikan.
- h. **Diskusi dan Kesimpulan:** Hasil penelitian akan dibahas dalam konteks literatur ilmiah yang relevan dan temuan-temuan sebelumnya. Kesimpulan akan diambil berdasarkan analisis hasil dan implikasinya terhadap penggunaan kedua algoritma dalam deteksi penyakit diabetes.

## **HASIL DAN PEMBAHASAN**

Hasil eksperimen dengan metode K-Means Clustering, dihasilkan nilai presentase tertinggi yaitu eksperimen menggunakan Data Acak 4. Banyaknya iterasi yang dihasilkan adalah 7 kali dan tingkat kebenarannya adalah 73,438%. Sedangkan hasil terendah yang diperoleh adalah eksperimen dengan menggunakan Data Nol, dengan banyak iterasi 2 kali dan tingkat kebenarannya 61,979%. Eksperimen dengan metode K-Means Clustering pada semua data ditunjukkan pada tabel 4 dibawah ini.

**Tabel 3.** Hasil Uji Menggunakan Metode *K-Means*

No	Data	Jumlah Iterasi	Tingkat Kebenaran
1	Data Kelurahan	4	66%
2	Data Nol	2	62%
3	Data Non Nol	11	67%
4	Data Ambigu	5	72%
5	Data Non Ambigu	8	72%
6	Data Acak 1	7	64%
7	Data Acak 2	9	68%
8	Data Acak 3	10	71%
9	Data Acak 4	7	73%

Eksperimen dengan metode Fuzzy C-Means dilakukan dengan menetapkan nilai Minimum Iteration: 500, Maximum Error: 0,000005. Data eksperimen Acak 1 sampai 4 menggunakan nilai range eksponen 2 sampai 6. Setelah dilakukan eksperimen, dihasilkan nilai presentase tingkat kebenaran tertinggi yaitu eksperimen menggunakan Data Acak 4 dengan nilai eksponen 6. Banyaknya iterasi yang dihasilkan adalah 38 kali dan tingkat kebenarannya mencapai 82,812%. Sedangkan hasil terendah yang diperoleh adalah eksperimen dengan menggunakan Data Keseluruhan, dengan hasil iterasi 43 kali dan tingkat kebenarannya 65,885%. Hasil eksperimen dengan metode Fuzzy C-Means Clustering pada semua data eksperimen, ditunjukkan pada table 5 di bawah ini.

**Tabel 4.** Hasil Uji Menggunakan Metode *Fuzzy C-Means*

No	Data	Exponen (w)	Jumlah Iterasi	Tingkat Kebenaran
1	Data Keseluruhan	2	43	65.89%
2	Data Nol	2	31	69.53%
3	Data Non Nol	2	56	68.49%
4	Data Ambigu	2	51	69.13%
5	Data Non Ambigu	2	40	72.77%
6		2	59	65.63%
7		3	54	66.15%
8	Data Acak 1	4	45	66.15%
9		5	40	66.67%
10		6	39	67.18%
11		2	35	72.92%
12		3	36	72.92%
13	Data Acak 2	4	40	72.92%
14		5	40	73.96%
15		6	53	73.44%
16		2	36	70.16%
17		3	37	71.73%
18	Data Acak 3	4	36	73.30%
19		5	37	74.87%
20		6	36	75.92%
21		2	56	71.88%
22		3	48	55.21%
23	Data Acak 4	4	48	79.17%
24		5	45	81.77%
25		6	38	82.81%

## KESIMPULAN DAN SARAN

Dalam penelitian ini, telah dilakukan evaluasi performa deteksi penyakit diabetes menggunakan metode Fuzzy C-Means dan K-Means clustering. Tujuan utama penelitian adalah untuk memahami sejauh mana kedua metode ini efektif dalam mengklasifikasikan individu sebagai penderita atau non-penderita diabetes berdasarkan data medis. Dengan mengikuti langkah-langkah metodologi yang telah dirancang, hasil dan temuan dari penelitian ini mengarah pada beberapa kesimpulan penting:

1. Penggabungan Metode: Menggabungkan hasil dari kedua metode, Fuzzy C-Means dan K-Means, dapat memberikan informasi yang lebih komprehensif dalam deteksi diabetes. Kombinasi sensitivitas dari Fuzzy C-Means dan spesifisitas dari K-Means dapat memperkuat akurasi diagnosa.
2. Fuzzy C-Means menunjukkan potensi dalam mendeteksi kasus diabetes dengan sensitivitas yang lebih tinggi, sedangkan K-Means memiliki keunggulan dalam menghasilkan kelompok yang lebih eksklusif.
3. Evaluasi performa deteksi penyakit diabetes menggunakan metode Fuzzy C-Means dan K-Means clustering memberikan wawasan yang berharga dalam penggunaan teknik analisis data untuk tujuan medis.
4. Hasil clustering pada kedua metode menunjukkan hasil yang lebih baik pada Data Acak

## DAFTAR REFERENSI

- Ahamed, T., Muhaimin, S., & Uddin, M. M. (2020). Comparative Analysis of K-means and Fuzzy C-means Clustering Algorithms in Diabetic Retinopathy Detection. *Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 6102-6108.
- Alavi, A., & Ghaderi, R. (2019). Comparative analysis of k-means and fuzzy c-means algorithms for breast cancer data clustering. *Journal of King Saud University-Computer and Information Sciences*.
- Barik, R., & Pradhan, B. B. (2020). Fuzzy C-means clustering technique for early diagnosis of diabetes. *Procedia Computer Science*, 167, 2743-2750.
- Can, F., & Kandemir, M. (2020). A comparative analysis of K-means and fuzzy c-means clustering algorithms for classification of diabetic data. *CBU International Conference Proceedings*, 8, 1429-1433.
- Poudel, R. P., & Jha, V. (2018). Comparative study of K-means and Fuzzy C-means clustering algorithms for medical data. *2018 International Conference on Computer and Applications (ICCA)*, 1-5.
- Barik, R., & Pradhan, B. B. (2020). Fuzzy C-means clustering technique for early diagnosis of diabetes. *Procedia Computer Science*, 167, 2743-2750.

- Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3), 191-203.
- Halkidi, M., Vazirgiannis, M., & Batistakis, Y. (2001). Clustering validity checking methods: Part I. *ACM Sigmod Record*, 30(2), 19-24.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Gopal, M., & Ranganathan, N. (Year). Diabetes Prediction using k-means clustering and naive Bayes classifier. *Procedia Computer Science*, Volume, Pages.
- Dabas, S., & Agarwal, S. (Year). Comparative Analysis of K-means and Fuzzy C-means Clustering for Medical Datasets. *Procedia computer science*, Volume, Pages
- Silva, D. F., Rodrigues, J. J., & de Albuquerque, V. H. C. (Year). Application of computational intelligence algorithms in diabetes diagnosis: a review. *Journal of medical systems*, Volume, Issue, Pages.