

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes *mellitus* adalah kondisi yang menyebabkan gangguan pada metabolisme tubuh dalam distribusi gula. Penderita diabetes *mellitus* mengalami ketidakmampuan dalam produksi insulin yang memadai atau dalam penggunaan insulin secara efektif, yang berakibat pada peningkatan kadar gula darah (Umam et al., 2020). Kondisi ini dapat mengakibatkan berbagai komplikasi kesehatan seperti penyakit jantung, kerusakan saraf, gangguan penglihatan, dan penyakit ginjal (Prasatya et al., 2020). Kondisi diatas menjadi acuan pengelompokkan jenis diabetes *mellitus*. Klasifikasi jenis diabetes ada beberapa macam, yaitu Diabetes *Mellitus* Tipe I (DMT1), Diabetes *Mellitus* Tipe II (DMT2), dan Diabetes *Mellitus* Tipe Gestasional (DMTG) (Marzel, 2020).

DMT1 disebabkan oleh reaksi autoimun terhadap protein sel pankreas, yang menyebabkan sel-sel pankreas rusak sehingga tidak mampu menghasilkan insulin yang cukup. Sel-sel pankreas juga menghasilkan hormon insulin untuk mengontrol kadar glukosa di dalam darah dengan mengontrol penggunaan dan penyimpanan glukosa. Namun, peningkatan kadar gula darah adalah akibat dari sejumlah faktor genetik yang mempengaruhi sekresi insulin, resistensi terhadap insulin, serta faktor lingkungan seperti obesitas, pola makan yang tidak sehat, kurang berolahraga, stres, dan proses penuaan. (Lestari et al., 2021).

Indonesia memiliki sebagai salah satu jumlah penderita diabetes *mellitus* tertinggi di dunia. *International Diabetes Federation* (IDF) mencatat bahwa pada tahun 2019, sekitar 463 juta orang di seluruh dunia berusia 20 hingga 79 tahun menderita diabetes *mellitus*, atau sekitar 9,3% dari populasi dalam rentang usia tersebut. IDF juga mencatat prevalensi diabetes *mellitus* pada tahun 2019 adalah sekitar 9,65% pada laki-laki dan 9% pada perempuan. (Lubis & Kanzasabilla, 2021). Dengan sekitar 10,7 juta orang yang menderita diabetes *mellitus*, Indonesia termasuk dalam sepuluh besar negara dengan jumlah penderita diabetes tertinggi (Salisa, 2023).

Dalam penelitian Riskesdas dari tahun 2013 hingga 2018, prevalensi diabetes *mellitus* meningkat sebesar 1,8%, naik dari 6,9% pada tahun 2013 menjadi 8,5%;

studi Riskesdas tahun 2018 menunjukkan bahwa prevalensi diabetes *mellitus* di Indonesia lebih tinggi pada perempuan, sebesar 1,7%, dibandingkan dengan laki-laki, sebesar 1,4%. (Erda et al., 2021). Pada 2018, Provinsi Jawa Timur memiliki jumlah penderita diabetes *mellitus* tertinggi di Indonesia, mencapai 2,7% dari 4% populasi (Duriana Wijaya et al., 2019). Pada periode 2019-2020, jumlah penderita diabetes *mellitus* meningkat sebanyak 33.751 individu, dari 841.994 pada tahun 2019 menjadi 875.745 pada tahun 2020 (Novianti et al., 2022). Berdasarkan meningkatnya prevalensi di wilayah Indonesia yang semakin tinggi setiap tahunnya ada salah satu cara untuk mencegahnya yaitu penting untuk memahami faktor-faktor yang mempengaruhi terjadinya penyakit ini dalam masyarakat dan mengetahuinya sedini mungkin (Chrismis Novalinda Ginting, Firman Halawa, Hestin Darmaris Lase, 2021). Karena dua pertiga penderita diabetes *mellitus* di Indonesia tidak menyadari bahwa mereka menderita penyakit ini, tidak mengetahui kemungkinan komplikasi, dan tidak mengetahui bagaimana mendapatkan perawatan medis (Arania, 2023).

Pada penelitian yang dilakukan (Chrismis Novalinda Ginting, Firman Halawa, Hestin Darmaris Lase, 2021) menunjukkan bahwa pola kesehatan, perilaku dan gaya hidup memiliki pengaruh signifikan terhadap kejadian diabetes *mellitus*. Namun mirisnya, saat ini masih banyak masyarakat yang tidak peduli terhadap pola kesehatan tubuh mereka masing-masing, kepadatan produktivitas dan fasilitas transportasi yang mendukung kebutuhan sehari-hari membuat sebagian orang kurang berolahraga atau melakukan aktivitas fisik, sehingga mengganggu proses metabolisme tubuh. Kondisi ini adalah salah satu dari sejumlah faktor yang turut berperan dalam terjadinya diabetes *mellitus* (Mexsi, 2023). Penyakit diabetes dapat dicegah secara efektif dengan cara mendeteksinya secara dini, sehingga individu dapat menerapkan langkah pencegahan seperti pola hidup sehat guna mencegah perkembangan penyakit diabetes (Nugraha & Sabaruddin, 2021). Untuk melakukan hal tersebut, tentunya akan lebih memudahkan jika dibantu dengan teknologi terkini agar menciptakan efisiensi untuk mengklasterisasi penyakit diabetes berdasarkan variabel yang tersedia. Salah satunya adalah dengan menggunakan perhitungan matematika dan *machine learning* dengan pendekatan metode *Fuzzy C-Means* serta memperbandingkannya melalui teknik lainnya seperti *K-means* untuk mengetahui

efektivitas algoritma *Clustering* pada dataset yang tersedia karena pada saat ini efektivitas algoritma *Fuzzy C-Means* dapat unggul dalam berbagai dataset yang berbeda pada penelitian-penelitian terdahulu.

Machine learning merupakan sistem yang dirancang untuk dapat mempelajari secara mandiri tanpa perlu arahan langsung dari penggunanya. *Machine learning* menggunakan matematika, statistika, dan *data mining* untuk membuat sistem mampu menganalisis data tanpa memerlukan program ulang atau program khusus (Sihombing & Arsani, 2021). Terdapat metode pendekatan yang dapat dilakukan dengan *machine learning*, yaitu *supervised learning* adalah metode dimana mesin dilatih menggunakan data yang telah diberi label, *unsupervised learning* adalah mesin diizinkan bekerja sendiri untuk menemukan informasi karena data tidak berlabel, *Semi-Supervised learning* adalah pendekatan yang menggabungkan elemen dari *supervised learning* dan *unsupervised learning* lalu *reinforcement learning* adalah mesin yang dapat belajar dan beradaptasi sendiri dengan lingkungan (Kristiawan & Widjaja, 2021).

Pada penelitian ini, metode yang digunakan *unsupervised learning* dengan teknik klasterisasi karena dataset yang tersedia tidak berlabel. Klasterisasi merupakan proses pengelompokan data ke dalam beberapa kelompok dimana data dalam satu kelompok memiliki kemiripan yang relatif rendah (Suroyo, 2019). Secara umum, sebuah kluster terdiri dari sekumpulan objek dalam satu kelompok yang tidak hanya mirip satu sama lain, tetapi juga berbeda dari objek dalam kelompok lain. Algoritma *clustering* dibagi menjadi dua jenis utama, yaitu hierarkis dan partisional. Algoritma hierarkis menemukan kluster secara bertahap dengan menetapkan kluster sebelumnya, sementara algoritma partisional seperti *Fuzzy C-Means* dan *K-means*, menentukan semua kelompok pada waktu yang bersamaan. Kemudian, metode *Clustering* yang diterapkan pada penelitian ini menggunakan metode partisi seperti *Fuzzy C-Means* dan *K-means* sebagai alat perbandingan.

Fuzzy C-Means termasuk dalam klasterisasi partisi yang membagi objek ke dalam beberapa klaster. Setiap klaster memiliki titik pusat (centroid) dengan memiliki tujuan meminimalkan relasi antara data objek dan pusat klasternya masing-masing. *Fuzzy C-Means* merupakan teknik klasterisasi data yang menggunakan derajat keanggotaan untuk menentukan klaster bagi setiap data. Pada

teknik *Fuzzy C-Means*, penempatan titik data dengan *cluster* tertentu ditentukan oleh tingkat keanggotaannya (Ulinuha, 2020). Menurut konsep Fuzzy, objek atau data tidak ditentukan sebagai anggota dengan *value* yang tegas seperti 1 (anggota) atau 0 (tidak anggota), tetapi diberikan derajat keanggotaan yang bervariasi dari 0 hingga 1. Semakin tinggi derajat keanggotaan suatu data, semakin kuat hubungannya sebagai anggota, dan sebaliknya, semakin rendah nilai derajat keanggotaannya, semakin lemah hubungannya sebagai anggota (Anggara Cahya Putra & Dwi Hartomo, 2021).

Metode algoritma *K-means* pertama dicetuskan oleh Mac Queen pada 1967 dan menerapkan teknik klusterisasi yang mengolah kumpulan dataset menjadi berbagai kelompok. Data dikelompokkan ke dalam kelompok dengan karakteristik yang sebanding dan berbeda satu sama lain. Tujuan dari kelompokisasi ini untuk mengurangi perbedaan antara kelompok data dan memberikan peningkatan pada perbedaan antara kelompok tersebut. (Haris Kurniawan et al., 2020). Metode algoritma *K-means* menghasilkan titik centroid sebagai acuan utama selama proses penyelesaiannya. Metode algoritma *K-means* mengelompokkan objek data ke dalam suatu dataset dalam *cluster* berdasarkan jarak terdekat mereka. Kemudian setelah iterasi *K-means* selesai, setiap objek dalam dataset tersebut akan terpecah menjadi bagian dari sebuah cluster. *Cluster Value* dihitung dengan mendapatkan semua objek untuk dilakukan penemuan *cluster* mana yang paling dekat dengan mereka.

Fuzzy C-Means, penyempurnaan dari metode algoritma *K-means*, dapat mengubah data untuk menjadi anggota dari semua *cluster* sesuai dengan tingkat keanggotaan mereka. Inilah menjadi perbedaan dari *K-means*, di mana *Fuzzy C-Means* memberikan fleksibilitas dalam keanggotaan *cluster* dibandingkan dengan pendekatan yang lebih kaku dari *K-means*. Permasalahan yang ada pada penelitian ini berfokus pada kedua algoritma mana yang lebih unggul dalam mengklusterkan dataset yang tersedia dan urgensi masalah dari penyakit diabetes mengenai deteksi dini penyakit diabetes dengan menggunakan teknologi terkini.

Menurut penelitian yang diteliti oleh (Anggrawan & Mayadi, 2023) menjelaskan bahwa penelitian tersebut dilakukan dengan metode algoritma *Fuzzy C-Means* yang menghasilkan hasil yang akurat dibanding dengan algoritma KNN

dalam mendiagnosis penyakit diabetes. Hasil pengujian menunjukkan sebesar 96% untuk *Fuzzy C-Means* dan KNN 83,3%. Penelitian yang dilakukan oleh (Prमितasari & Nataliani, 2021) juga menunjukkan bahwa metode algoritma *Fuzzy C-Means* unggul dibandingkan metode algoritma *K-means* dalam menilai kinerja karyawan. Hasil yang ditunjukkan *Fuzzy C-Means* 76% dan *K-means* 44%.

Perubahan dibuat pada penelitian ini meliputi data variabel medis yang berjumlah 8 variabel dengan rincian variabel meliputi usia, BMI, sistolik, diastolik, hba1c, gula darah 2 jam PP, gula darah acak, dan gula darah puasa merupakan beberapa faktor resiko yang turut berperan dalam penyakit diabetes *mellitus*. Hasil riset yang dilakukan oleh (Komariah & Rahayu, 2020) menunjukkan bahwa variabel-variabel yang dapat berhubungan pada kadar gula darah puasa meliputi usia dan pola makan. Faktor usia dikaitkan dengan perubahan fisiologis pada usia lanjut, dimana selaras dengan bertambahnya usia, fungsi tubuh memasuki tahap penurunan, termasuk efektivitas hormon insulin. Peningkatan kadar gula darah terjadi karena insulin tidak berfungsi dengan baik karena penurunan ini terjadi.

Menurut penelitian (Komariah & Rahayu, 2020) menunjukkan bahwa di RSUD Kota Makassar, ada korelasi antara status gizi dan kepatuhan diet pada kadar glukosa darah acak dan pada kadar gula darah 2 jam postprandial (PP) pada penderita diabetes *mellitus*. Banyaknya makanan yang memiliki indeks glikemik tinggi, secara langsung dapat meningkatkan kadar gula darah, memainkan peran penting dalam meningkatkan kadar gula darah. Indeks glikemik merupakan ukuran yang dapat menunjukkan cepat atau lambatnya makanan yang mengandung karbohidrat dapat meningkatkan kadar gula darah setelah dikonsumsi. Salah satu faktor risiko utama penderita diabetes adalah obesitas, yang dapat diidentifikasi melalui ukuran indeks massa tubuh (BMI). Pada penderita obesitas, kadar gula dalam darah meningkat karena sel menjadi kurang sensitif terhadap insulin. Penggunaan BMI membantu dalam mengidentifikasi individu yang berisiko tinggi untuk diabetes dan komplikasi terkait, serta penting dalam penilaian dan manajemen kesehatan secara keseluruhan.

Penelitian ini menggunakan metode algoritma *Clustering Fuzzy C-Means* dengan metode pembandingan algoritma *K-means* digunakan untuk dijadikan perbandingan dalam hasil akurasi dan efektivitas dalam dataset yang sudah tersedia,

dengan jumlah *cluster* yang sama, yaitu 2. Penelitian ini berfokus pada dataset medis praduga diabetes, dengan output yang dihasilkan berupa klasterisasi pasien berdasarkan variabel yang menentukan apakah mereka menderita diabetes atau tidak. Adapun fokus dari kedua algoritma ini terletak pada penentuan *cluster* yang sudah ditentukan, dengan membandingkan dan menganalisis kedua pendekatan tersebut. Adapun tujuan utama dari penelitian ini adalah mengetahui faktor yang paling umum dan terbesar yang menyebabkan diagnosa sebagai penderita diabetes *mellitus* dengan memberikan wawasan yang lebih mendalam tentang manfaat dan potensi penggunaan algoritma *Clustering* tersebut dalam kehandalan model klasterisasi untuk mendukung proses diagnosis dan manajemen penyakit diabetes *mellitus*.

1.1. Rumusan Masalah

Berdasarkan latar belakang yang sudah dibahas, rumusan masalah penelitian ini dijelaskan sebagai dibawah ini:

- 1) Indikator apa saja yang mempengaruhi seseorang dikategorikan sebagai penyandang penyakit diabetes *mellitus* berdasarkan data medis yang tersedia.
- 2) Bagaimana penerapan algoritma *Fuzzy C-Means* dapat mengelompokkan faktor-faktor risiko Diabetes *Mellitus* berdasarkan data medis, dan sejauh mana akurasi dan efektivitasnya dibandingkan dengan algoritma *K-means*

1.2. Batasan Masalah

Batasan masalah adalah elemen penting pada penelitian yang digunakan untuk mencegah asumsi yang terlalu luas. Batasan masalah disajikan dalam bentuk poin-poin seperti berikut:

- 1) Penelitian ini menerapkan pendekatan dua teknik *Clustering*, yaitu metode algoritma *Fuzzy C-Means* dan *K-means*, dalam mengelompokkan faktor-faktor risiko Diabetes *Mellitus*.
- 2) Penggunaan dataset yang dipakai dalam penelitian ini terbatas pada catatan medis Puskesmas Sembayat yang mencakup faktor-faktor risiko Diabetes *Mellitus*.

- 3) Evaluasi kinerja algoritma *Clustering* akan didasarkan pada metrik evaluasi seperti akurasi *Clustering*, nilai Silhouette, dan interpretasi visual dari hasil *Clustering*.

1.3. Tujuan Penelitian

Menerapkan algoritma *Clustering Fuzzy C-Means* adalah tujuan utama penelitian ini dalam mengidentifikasi variabel indikator penyebab Diabetes *Mellitus* pada pasien. Berikut poin tujuan utama dari penelitian ini adalah:

- 1) Mengidentifikasi variabel indikator utama yang menyebabkan seseorang dikategorikan sebagai penderita diabetes *mellitus* berdasarkan data medis yang tersedia.
- 2) Menerapkan algoritma *Fuzzy C-Means* untuk mengelompokkan variabel indikator risiko diabetes *mellitus* berdasarkan data medis, serta mengevaluasi akurasi dan efektivitasnya dibandingkan dengan algoritma *K-means*.

1.4. Manfaat Penelitian

Diharapkan bahwa penelitian ini akan bermanfaat bagi pihak-pihak yang membutuhkannya, baik secara teoritis maupun praktis, termasuk:

- 1) Manfaat teoritis

Penelitian ini akan berkontribusi pada pengembangan pengetahuan ilmiah di bidang analisis penyebab Diabetes *Mellitus*. Dengan bantuan penerapan teknik algoritma *Clustering Fuzzy C-Means* dengan menggunakan tambahan metode algoritma *K-means* sebagai pendukung dalam menganalisis variabel indikator penyebab Diabetes *Mellitus*, penelitian ini akan memperdalam pemahaman kita tentang bagaimana metode *Clustering* dapat diterapkan dalam konteks kesehatan. Hasil penelitian ini diharapkan memberikan pemahaman yang lebih mendalam terhadap kelebihan dan kekurangan dari metode algoritma *K-means* dan *Fuzzy C-Means* dalam menganalisis variabel indikator Diabetes *Mellitus*.

2) Manfaat praktis

- a. Bagi penulis, Penelitian ini dapat memberikan pengetahuan dan keterampilan terhadap penggunaan algoritma *Clustering* untuk analisis data kesehatan. Ini akan meningkatkan kompetensi penulis dalam penelitian dan analisis data.

- b. Bagi peneliti selanjutnya, diharapkan penelitian menjadi referensi penting bagi peneliti seterusnya yang ingin menggunakan metode *Clustering* dalam menganalisis variabel indikator penyebab penyakit. Mereka dapat mengambil manfaat dari efektivitas algoritma *Fuzzy C-Means* yang diterapkan dalam penelitian ini.