

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kementerian Pendidikan dan Kebudayaan (Kemendikbud) membagi sistem pendidikan di Indonesia menjadi tiga jenjang utama, yaitu pendidikan dasar, menengah, dan tinggi. Pendidikan menengah terdiri dari Sekolah Menengah Pertama (SMP) dan Sekolah Menengah Atas (SMA/SMK), yang masing-masing membutuhkan waktu tiga tahun untuk diselesaikan. SMA memiliki peran penting dalam membentuk kepribadian siswa dan mempersiapkan mereka untuk jenjang pendidikan yang lebih tinggi (Dharmawan & Adi Wijoyo, 2023).

Peningkatan kualitas pendidikan menjadi prioritas utama pemerintah, terutama di daerah dengan banyak sekolah yang memiliki standar dan kualitas bervariasi, seperti Kabupaten Serdang Bedagai. Jumlah sekolah yang terus bertambah dengan perbedaan dalam prestasi, fasilitas, tenaga pengajar, dan kurikulum menimbulkan tantangan bagi masyarakat dan pemerintah. Masyarakat, orang tua, dan siswa membutuhkan informasi terstruktur untuk memilih sekolah yang sesuai kebutuhan, sementara pemerintah daerah perlu data yang sistematis untuk mengidentifikasi sekolah yang memerlukan peningkatan atau layak dijadikan contoh sebagai sekolah berkualitas. Tanpa data yang jelas, akan sulit bagi pemerintah untuk merancang strategi efektif dalam meningkatkan mutu pendidikan secara menyeluruh.

Penerapan ilmu data mining adalah proses yang bertujuan untuk menggali informasi penting dari sejumlah besar data yang tersimpan dalam basis data. Berdasarkan berbagai jurnal penelitian, data mining diartikan sebagai pendekatan analitis untuk mengekstraksi pola, hubungan, atau pengetahuan tersembunyi dalam data yang besar dan kompleks (Zai & Komputer, 2022). Data mining merupakan proses menelusuri kumpulan data yang ada untuk menemukan pola-pola penting, yang kemudian digunakan untuk membangun sebuah model. Model ini dikembangkan agar dapat mengenali pola atau tren pada data baru yang belum

ada dalam basis data yang disimpan (Mahasiswa & Samudra, 2023). Dengan kata lain, data mining melibatkan pemrosesan data secara mendalam untuk mengekstrak informasi yang bernilai, yang kemudian bisa diaplikasikan untuk menganalisis data serupa di masa depan atau data yang tidak dikenal. Tujuan utama dari data mining adalah menggali wawasan yang bermanfaat dan menggunakannya untuk prediksi atau pengambilan keputusan berbasis data yang lebih akurat.

Klasterisasi adalah metode pengelompokan data ke dalam sejumlah kelompok yang disebut kluster. Beragam algoritma klasterisasi tersedia, yang umumnya terbagi menjadi dua kategori: klasterisasi hirarki dan non-hirarki. Salah satu algoritma hirarki yang banyak digunakan adalah *Average Linkage*, yang unggul karena tidak membutuhkan jumlah kluster yang ditentukan di awal. Sebaliknya, klasterisasi secara non-hierarki salah satunya yang sering digunakan adalah *k-means* dan *fuzzy c-means*, dipilih karena efisien dan mudah diterapkan, meskipun membutuhkan penentuan jumlah kluster sejak awal (Simanjorang & Yanti, 2023).

K-means adalah metode *clustering* non-hierarki yang bertujuan mengelompokkan data ke dalam satu atau beberapa *cluster*. Data dengan karakteristik serupa ditempatkan dalam *cluster* yang sama, sementara data dengan karakteristik berbeda ditempatkan dalam *cluster* lainnya. Dalam algoritma ini, jumlah *cluster* ditentukan terlebih dahulu, tetapi pemilihan titik tengah atau centroid tidak harus ditetapkan secara pasti, sehingga titik tengah dapat dipilih dari objek mana saja dalam data. Untuk penentuan centroid, data pertama dipilih sebagai centroid pertama, data kedua sebagai centroid kedua, dan seterusnya hingga mencapai jumlah centroid yang dibutuhkan. Langkah selanjutnya adalah menghitung jarak antara setiap titik data dan centroid yang telah dipilih, lalu mengelompokkan titik-titik tersebut berdasarkan jarak terdekatnya ke centroid masing-masing. Setelah semua titik dikelompokkan, langkah pertama selesai. Kemudian, dilakukan penghitungan ulang posisi centroid baru sebagai pusat atau rata-rata dari setiap kelompok yang terbentuk. Setelah diperoleh centroid baru, proses pengelompokan diulangi untuk menyesuaikan posisi titik dengan centroid

baru tersebut. Penerapan algoritma *k-means* dalam data mining ini dapat mendukung pengambilan keputusan di masa depan, karena metode *k-means* berfungsi membagi data ke dalam satu atau beberapa cluster atau mengelompokkan data ke dalam sejumlah kelompok yang berbeda (Latipa Sari & Yati Beti, 2023).

Fuzzy C-Means (FCM) adalah metode pengelompokan yang berkembang dari konsep dasar *Hard C-Means*, dengan perbedaan utama pada cara menentukan keanggotaan data dalam suatu *cluster*. Berbeda dengan *Hard C-Means* yang mengharuskan setiap data hanya menjadi anggota satu *cluster* secara eksklusif, FCM memungkinkan setiap data untuk memiliki derajat keanggotaan yang bervariasi di beberapa *cluster* sekaligus. Nilai derajat keanggotaan ini berkisar antara 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan tingkat keterkaitan yang lebih kuat antara data dan *cluster* tersebut. Proses pengelompokan ini didasarkan pada prinsip *fuzzy*, yang berarti bahwa data tidak hanya memiliki satu identitas *cluster* yang pasti, melainkan dapat menjadi bagian dari beberapa *cluster* dengan tingkat keanggotaan yang berbeda-beda, mencerminkan kompleksitas hubungan antar data dan *cluster*. Keunggulan utama dari FCM terletak pada kemampuannya untuk menangani data yang memiliki ambiguitas atau tumpang tindih antara kelompok, yang sering kali ditemukan dalam kasus dunia nyata. Dengan cara ini, FCM memberikan gambaran yang lebih fleksibel dan realistis tentang bagaimana data tersebar dan berinteraksi dalam berbagai kategori (R.S et al., 2024).

Penelitian sebelumnya dilakukan oleh Nurahman dengan judul “Klasterisasi Sekolah Menggunakan Algoritma *K-Means* berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik” (Nurahman et al., 2022). Penelitian ini bertujuan untuk membantu pemerintah Kabupaten Seruyan dalam mengelompokkan sekolah-sekolah tertentu menjadi beberapa klaster berdasarkan fasilitas, pendidik, dan tenaga pendidik. Dengan pengelompokan ini, diharapkan pemerintah dapat lebih mudah memberikan pendampingan dan pengadaan kebutuhan sekolah, terutama untuk sekolah-sekolah yang berada dalam klaster dengan kondisi rendah. Penggunaan algoritma *k-means* dalam penelitian ini bertujuan untuk

mempermudah proses klasterisasi dan menghasilkan data yang dapat diandalkan untuk pengambilan keputusan.

Pada penelitian ini dilakukan untuk mengatasi tantangan dalam pemerataan dan peningkatan mutu pendidikan di Kabupaten Serdang Bedagai, yang memiliki banyak sekolah dengan standar dan kualitas yang beragam. Penggunaan metode *K-Means* dan *Fuzzy C-Means* dalam penelitian ini menjadi sangat penting karena mampu menghasilkan klasterisasi sekolah yang lebih akurat, baik untuk mengidentifikasi pola kualitas maupun untuk menentukan prioritas pengembangan. Dengan mengaplikasikan metode *K-Means* dan *Fuzzy C-Means*, penelitian ini mampu menghasilkan informasi yang terstruktur mengenai klasterisasi Sekolah Menengah Atas (SMA) berdasarkan berbagai indikator kualitas, seperti fasilitas, tenaga pendidik, dan prestasi. Informasi ini sangat penting bagi masyarakat untuk memilih sekolah yang sesuai dengan kebutuhan mereka, sekaligus bagi pemerintah daerah untuk mengidentifikasi sekolah yang membutuhkan perhatian lebih atau yang dapat dijadikan model sekolah berkualitas.

Berdasarkan permasalahan dan uraian di atas, penulis tertarik untuk mengajukan penelitian dengan judul "Analisis Perbandingan Metode *K-Means* dan *Fuzzy C-Means* untuk Klasterisasi Sekolah Menengah Atas (SMA) Terbaik di Kabupaten Serdang Bedagai". Penelitian ini dilatarbelakangi oleh kebutuhan akan data yang terstruktur dan akurat untuk membantu masyarakat dan pemerintah daerah dalam mengidentifikasi kualitas sekolah di Kabupaten Serdang Bedagai, dengan memanfaatkan metode *K-Means* dan *Fuzzy C-Means*.

1.2 Rumusan Masalah

Dari latar belakang, rumusan masalah dalam penelitian ini yaitu :

1. Bagaimana perbandingan efektivitas algoritma *k-means* dan *fuzzy c-means* dalam mengelompokkan Sekolah Menengah Atas (SMA) terbaik berdasarkan kriteria tertentu di Kabupaten Serdang Bedagai.

2. Apa saja faktor-faktor yang mempengaruhi hasil klasterisasi Sekolah Menengah Atas (SMA) menggunakan metode *K-Means* dan *Fuzzy C-Means* di Kabupaten Serdang Bedagai.

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini yaitu:

1. Untuk membandingkan efektivitas algoritma *K-Means* dan *Fuzzy C-Means* dalam mengelompokkan Sekolah Menengah Atas (SMA) di Kabupaten Serdang Bedagai berdasarkan indikator kualitas sekolah.
2. Untuk mengidentifikasi dan menganalisis kontribusi faktor-faktor seperti akreditasi, status sekolah, jumlah guru, prestasi, dan fasilitas terhadap hasil klasterisasi yang dihasilkan oleh metode *K-Means* dan *Fuzzy C-Means*.

1.4 Manfaat Penelitian

Manfaat hasil penelitian ini adalah sebagai berikut:

1. Menjadi referensi dan sumber informasi tambahan dalam menganalisis dan menentukan sekolah Menengah Atas (SMA) terbaik di Kabupaten Serdang Bedagai berdasarkan hasil klasterisasi menggunakan metode *K-Means* dan *Fuzzy C-Means*.
2. Memberikan panduan kepada pemerintah dan masyarakat dalam mengidentifikasi Sekolah Menengah Atas (SMA) berkualitas tinggi yang dapat menjadi acuan untuk meningkatkan mutu pendidikan di Kabupaten Serdang Bedagai.
3. Membantu orang tua memahami kualitas sekolah secara objektif, tanpa terpengaruh oleh istilah seperti sekolah unggulan, favorit, atau label prestisius lainnya, melalui hasil klasterisasi yang berbasis data.
4. Mempermudah orang tua dan pihak terkait dalam membuat keputusan yang tepat dalam memilih Sekolah Menengah Atas (SMA) terbaik bagi anak-anak mereka berdasarkan analisis yang terukur dan terpercaya.

1.5 Ruang Lingkup dan Batasan Penelitian

Penelitian ini berfokus pada klasterisasi Sekolah Menengah Atas (SMA) di Kabupaten Serdang Bedagai dengan menggunakan *K-Means* dan *Fuzzy C-Means*. Data yang dianalisis mencakup berbagai aspek sekolah, seperti akreditasi, status sekolah, jumlah tenaga pengajar, fasilitas yang tersedia, dan prestasi yang diraih. Hasil penelitian ini diharapkan dapat memberikan wawasan dalam pemetaan kualitas pendidikan di wilayah tersebut.

Adapun batasan masalah dari penelitian ini sebagai berikut :

1. Data yang digunakan dalam penelitian ini bersumber dari semua Sekolah Menengah Atas (SMA) yang ada di wilayah Kabupaten Serdang Bedagai, yang diperoleh dari Dinas Pendidikan Kabupaten Serdang Bedagai, dan mencakup data tahun 2024.
2. Penelitian ini menggunakan dua metode klasterisasi, yaitu *K-Means* dan *Fuzzy C-Means*, untuk membandingkan kinerja kedua metode dalam mengelompokkan Sekolah Menengah Atas (SMA).
3. Kriteria yang digunakan dalam penelitian ini meliputi akreditasi, status sekolah, jumlah guru, fasilitas yang tersedia, dan prestasi yang diraih oleh sekolah tersebut.
4. Hasil akhir dari penelitian ini adalah mengklasifikasikan Sekolah Menengah Atas (SMA) terbaik di Kabupaten Serdang Bedagai berdasarkan kriteria yang telah ditentukan, untuk memberikan gambaran mengenai kualitas sekolah di wilayah tersebut.