

BAB I

PENDAHULUAN

1.1 Latar Belakang

Transformasi digital dalam bidang pendidikan telah mendorong peningkatan ketersediaan data akademik dalam jumlah besar, mulai dari nilai harian, hasil ujian, hingga laporan capaian pembelajaran tiap semester. Pemanfaatan data tersebut menjadi bagian penting dalam pendekatan data-driven decision making yang bertujuan meningkatkan efektivitas evaluasi dan intervensi pembelajaran (Hernández-Blanco et al., 2019). Namun demikian, pada banyak institusi pendidikan dasar, data nilai siswa masih diperlakukan sebagai dokumen administratif yang berfungsi untuk pelaporan dan penentuan kenaikan kelas, tanpa dianalisis lebih lanjut untuk mengidentifikasi pola prestasi yang tersembunyi.

Secara konseptual, nilai akademik bukan sekadar representasi angka, melainkan refleksi dari struktur kemampuan siswa yang multidimensional. Siswa dengan rata-rata nilai yang sama belum tentu memiliki pola performa yang identik pada setiap mata pelajaran. Ketergantungan pada agregasi nilai sederhana berpotensi mengaburkan variasi internal dalam data. Oleh karena itu, diperlukan pendekatan analitis yang mampu mengeksplorasi struktur kemiripan antar siswa secara lebih objektif dan sistematis (Hanaris, 2023).

Dalam kerangka ilmu data, teknik clustering pada pendekatan unsupervised learning menawarkan mekanisme untuk mengelompokkan data berdasarkan kemiripan karakteristik tanpa memerlukan label awal (Alzubi et al., 2018). Penerapan clustering dalam pendidikan telah digunakan untuk mengidentifikasi kelompok performa siswa, segmentasi tingkat risiko akademik, hingga personalisasi strategi pembelajaran (Lu et al., 2025). Di antara berbagai algoritma clustering, K-Means menjadi metode yang paling populer karena efisiensi komputasi dan kemudahan implementasinya.

Meskipun demikian, K-Means memiliki keterbatasan struktural, seperti sensitivitas terhadap inialisasi centroid dan kecenderungan menghasilkan cluster

berbentuk konveks. Alternatif seperti Hierarchical Clustering menawarkan pendekatan berbasis struktur bertingkat yang memungkinkan visualisasi hubungan antar data melalui dendrogram, sehingga memberikan perspektif segmentasi yang berbeda. Perbedaan prinsip dasar antara pendekatan partisional dan hierarkis menunjukkan bahwa hasil clustering tidak bersifat universal, melainkan bergantung pada karakteristik algoritma yang digunakan (Siswadi et al., 2025). Hal ini mengindikasikan pentingnya analisis komparatif untuk menilai konsistensi dan kualitas pengelompokan.

Selain pemilihan algoritma, validitas hasil clustering juga ditentukan oleh evaluasi kualitas cluster. Penggunaan metrik internal seperti Silhouette Score dan Davies-Bouldin Index memungkinkan penilaian objektif terhadap tingkat separasi dan kohesi cluster (Hasan, 2024). Namun, tinjauan terhadap penelitian terdahulu menunjukkan bahwa sebagian besar studi pendidikan hanya menggunakan satu algoritma clustering tanpa melakukan komparasi langsung dengan metode alternatif. Di sisi lain, penelitian yang membandingkan algoritma clustering sering kali diterapkan pada domain non-pendidikan, seperti ekonomi atau geospasial, sehingga relevansi kontekstual terhadap data akademik sekolah dasar masih terbatas.

Lebih lanjut, penelitian clustering dalam konteks pendidikan dasar umumnya bersifat statis, yaitu hanya menganalisis satu periode data tanpa mempertimbangkan dinamika perubahan performa antar semester (Saqr et al., 2023). Padahal, nilai rapor bersifat longitudinal dan merepresentasikan perkembangan akademik siswa secara temporal (Li et al., 2022). Minimnya eksplorasi terhadap dimensi waktu ini menunjukkan adanya celah penelitian dalam pemetaan pola prestasi yang lebih komprehensif.

Berdasarkan kondisi tersebut, dapat diidentifikasi adanya kebutuhan untuk melakukan analisis komparatif algoritma clustering dalam konteks pendidikan dasar, dengan evaluasi kualitas cluster yang terukur serta mempertimbangkan perubahan performa antar semester. Penelitian ini berfokus pada siswa kelas IV MI Al Chaeriyah sebagai studi kasus, dengan membandingkan algoritma K-Means dan

Hierarchical Clustering menggunakan metrik Silhouette Score dan Davies-Bouldin Index, serta menganalisis dinamika nilai antar semester.

Dengan demikian, penelitian ini tidak hanya mengimplementasikan teknik clustering, tetapi juga berupaya memberikan kontribusi metodologis melalui pendekatan komparatif yang terstruktur serta kontribusi kontekstual dalam pemanfaatan data rapor siswa sekolah dasar sebagai dasar analisis pola prestasi akademik.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka perumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana struktur pengelompokan pola prestasi akademik siswa kelas IV MI Al Chaeriyah jika dianalisis menggunakan algoritma K-Means dan Hierarchical Clustering?
2. Apakah terdapat perbedaan kualitas hasil clustering antara algoritma K-Means dan Hierarchical Clustering berdasarkan metrik evaluasi Silhouette Score dan Davies-Bouldin Index?
3. Bagaimana dinamika perubahan prestasi akademik siswa antar semester berdasarkan struktur cluster yang terbentuk?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menganalisis struktur pengelompokan pola prestasi akademik siswa menggunakan algoritma K-Means dan Hierarchical Clustering.
2. Membandingkan kualitas hasil clustering kedua algoritma berdasarkan nilai Silhouette Score dan Davies-Bouldin Index.
3. Mengidentifikasi perubahan pola prestasi akademik siswa antar semester berdasarkan hasil segmentasi cluster.

1.4 Manfaat Penelitian

Jika tujuan penelitian ini tercapai, hasil temuan diharapkan dapat memberikan kontribusi yang signifikan, baik secara teoritis, metodologis, dan praktis.

1.4.1 Manfaat Teoritis

Memberikan kontribusi terhadap pengembangan kajian clustering komparatif dalam konteks pendidikan dasar, khususnya dalam analisis pola prestasi akademik siswa.

1.4.2 Manfaat Metodologis

Menyediakan evaluasi objektif terhadap dua pendekatan clustering (partisional dan hierarkis) menggunakan lebih dari satu metrik kualitas cluster.

1.4.3 Manfaat Praktis

Memberikan gambaran segmentasi prestasi siswa kepada pihak sekolah sebagai dasar pertimbangan dalam evaluasi pembelajaran dan strategi akademik.

1.5 Batasan Masalah

Untuk menjaga agar penelitian ini tetap fokus, tidak melebar, dan dapat diselesaikan tepat waktu sesuai dengan tujuan yang telah ditetapkan, maka peneliti menetapkan batasan-batasan masalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini terbatas pada data nilai rapor siswa kelas IV MI Al Chaeriyah dalam satu tahun ajaran yang terdiri dari dua semester.
2. Algoritma clustering yang dianalisis hanya K-Means dan Hierarchical Clustering.
3. Evaluasi kualitas cluster dilakukan menggunakan metrik Silhouette Score dan Davies-Bouldin Index.
4. Penelitian tidak membahas model prediksi atau klasifikasi performa akademik (supervised learning).