Penerapan Algoritma K-Means untuk Mengelompokkan Siswa Berdasarkan Nilai Rapor dan Ekstrakurikuler di SMA Hang Tuah 1 Surabaya

1st Najla Aqilha Sya'bania
Departemen Sistem Informasi
Telkom University
Surabaya, Indonesia
najlaaqs@student.telkomuniversity.ac.id

2nd Agus Sulistya
Departemen Sistem Informasi
Telkom University
Surabaya, Indonesia
Sulistya@telkomuniversity.ac.id

3rd Sri Hidayati
Departemen Sistem Informasi
Telkom University
Surabaya, Indonesia
srihidayati@telkomuniversity.ac.id

Abstrak* — Pemanfaatan data akademik dan non-akademik di lingkungan sekolah masih belum dilakukan secara optimal, khususnya dalam mendukung perencanaan pembelajaran yang sesuai dengan karakteristik siswa. SMA Hang Tuah 1 Surabaya memiliki data nilai rapor dan keaktifan ekstrakurikuler yang cukup lengkap, namun belum dimanfaatkan sebagai dasar pengambilan keputusan pembelajaran berbasis data. Penelitian ini bertujuan mengelompokkan siswa berdasarkan nilai rapor dan aktivitas ekstrakurikuler menggunakan algoritma K-Means Clustering sebagai dasar penerapan pembelajaran diferensiasi. Metode penelitian mengacu pada kerangka kerja CRISP-DM yang meliputi tahap business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Data siswa diolah melalui proses pembersihan dan normalisasi, kemudian jumlah cluster optimal ditentukan menggunakan metode Elbow dan Silhouette Score. Hasil penelitian menunjukkan bahwa tiga cluster merupakan jumlah paling optimal dengan nilai Silhouette Score sebesar 0,4152. Setiap cluster memiliki karakteristik akademik dan non-akademik yang berbeda, sehingga dapat digunakan sebagai dasar penyesuaian strategi pembelajaran. Hasil pengelompokan divisualisasikan melalui dashboard interaktif berbasis Streamlit untuk memudahkan guru dan pihak sekolah dalam melakukan pemantauan dan analisis secara real-time. Penelitian ini menunjukkan bahwa pendekatan berbasis data dapat membantu sekolah memahami kondisi siswa secara lebih objektif dan mendukung penerapan pembelajaran diferensiasi.

***Kata kunci*— K-Means Clustering, Nilai Akademik, Ekstrakurikuler, Streamlit, Pembelajaran Diferensiasi.**

I. PENDAHULUAN

Pendidikan memiliki peran penting dalam membentuk kualitas sumber daya manusia. Proses pembelajaran yang efektif tidak hanya bergantung pada kurikulum, tetapi juga pada kemampuan sekolah dalam memahami karakteristik dan kebutuhan siswa. Setiap siswa memiliki perbedaan dalam kemampuan akademik, minat, serta tingkat keaktifan non-akademik, sehingga pendekatan pembelajaran yang seragam sering kali kurang optimal [1].

SMA Hang Tuah 1 Surabaya merupakan salah satu sekolah menengah atas dengan jumlah siswa yang besar dan latar belakang kemampuan yang beragam. Sekolah ini menyimpan data nilai rapor siswa dari beberapa semester serta data keaktifan ekstrakurikuler. Namun, pemanfaatan data tersebut masih terbatas pada penilaian administratif dan belum digunakan secara maksimal untuk pemetaan karakteristik siswa secara menyeluruh.

Pembelajaran diferensiasi menjadi salah satu pendekatan yang dianjurkan dalam Kurikulum Merdeka karena memberikan keleluasaan kepada guru untuk menyesuaikan pembelajaran dengan kebutuhan siswa . Untuk mendukung penerapan pembelajaran diferensiasi secara objektif, diperlukan pemetaan siswa berdasarkan data yang relevan. Salah satu pendekatan yang dapat digunakan adalah data mining, khususnya teknik clustering, untuk mengelompokkan siswa berdasarkan kemiripan karakteristik tertentu.

Penelitian ini menerapkan algoritma K-Means untuk mengelompokkan siswa berdasarkan nilai rapor dan keaktifan ekstrakurikuler. Hasil pengelompokan diharapkan dapat memberikan gambaran objektif mengenai kondisi siswa dan menjadi dasar bagi guru dalam merancang strategi pembelajaran yang lebih tepat sasaran.

II. KAJIAN TEORI

Kajian teori disusun untuk memberikan landasan konseptual yang mendukung pelaksanaan penelitian ini. Pembahasan teori difokuskan pada konsep data mining, algoritma K-Means Clustering, serta pembelajaran diferensiasi yang berkaitan dengan pengelompokan siswa berdasarkan karakteristik akademik dan non-akademik.

A. Data Mining

Data mining merupakan proses penggalian pengetahuan dari kumpulan data dalam jumlah besar untuk menemukan pola, hubungan, atau informasi baru yang bernilai [2]. Proses ini melibatkan tahapan pemahaman masalah, pengolahan data, pemilihan metode analisis, hingga interpretasi hasil. Dalam bidang pendidikan, data mining dimanfaatkan untuk menganalisis data akademik dan non-akademik siswa guna mendukung pengambilan keputusan berbasis data [3]. Pemanfaatan teknik data mining memungkinkan sekolah memperoleh gambaran objektif mengenai karakteristik siswa sehingga dapat digunakan sebagai dasar perencanaan strategi pembelajaran yang lebih tepat sasaran[4]..

B. Algoritma K-Means Clustering

K-Means merupakan algoritma clustering yang bertujuan mengelompokkan data ke dalam sejumlah cluster berdasarkan tingkat kemiripan tertentu [5]. Algoritma ini bekerja dengan menentukan pusat cluster (*centroid*) dan mengelompokkan

data ke dalam cluster dengan jarak terdekat secara iteratif hingga posisi centroid stabil. K-Means banyak digunakan karena memiliki konsep yang sederhana dan efisien dalam mengolah data berukuran besar. Untuk menentukan jumlah cluster yang optimal, metode Elbow dan Silhouette Coefficient sering digunakan sebagai alat evaluasi kualitas pengelompokan. Dalam konteks pendidikan, K-Means dapat membantu mengelompokkan siswa berdasarkan karakteristik akademik dan non-akademik sehingga memudahkan analisis perbedaan kemampuan dan potensi siswa[6].

C. Pembelajaran Diferensiasi

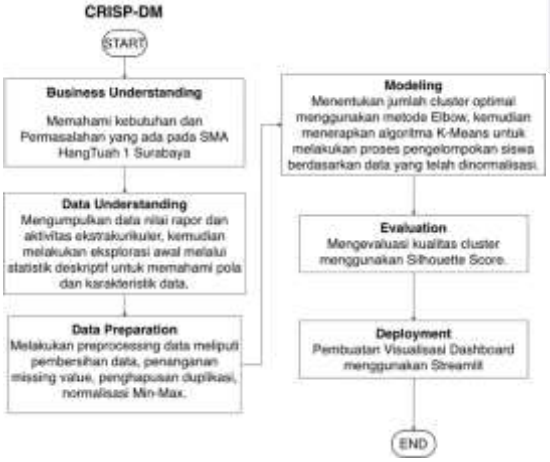
Pembelajaran diferensiasi merupakan pendekatan pembelajaran yang menyesuaikan proses, konten, dan hasil belajar dengan kebutuhan, minat, serta kemampuan siswa. Pendekatan ini bertujuan memberikan kesempatan belajar yang optimal bagi setiap siswa, mengingat adanya perbedaan karakteristik dalam satu kelas. Dalam Kurikulum Merdeka, pembelajaran diferensiasi dianjurkan sebagai strategi untuk meningkatkan keterlibatan dan capaian belajar siswa. Penerapan pembelajaran diferensiasi memerlukan pemetaan karakteristik siswa secara akurat [7]. Oleh karena itu, hasil pengelompokan siswa berbasis data melalui teknik clustering dapat digunakan sebagai dasar dalam merancang strategi pembelajaran diferensiasi yang lebih objektif dan terarah[8].

D. Crisp-DM

Penelitian ini menggunakan metodologi *Cross Industry Standard Process* for Data Mining (CRISP-DM) sebagai kerangka kerja dalam proses analisis data [9]. Penerapan CRISP-DM membantu mengarahkan penelitian mulai dari perumusan tujuan analisis hingga interpretasi dan penyajian hasil clustering. Melalui kerangka ini, proses pengelompokan siswa menggunakan algoritma K-Means dapat dilakukan secara sistematis dan terstruktur, sehingga hasil yang diperoleh relevan dengan tujuan analisis dan mudah diinterpretasikan[10]

III. METODE

Metodologi penelitian ini mengacu pada kerangka kerja CRISP-DM yang terdiri dari enam tahapan utama. Pendekatan ini dipilih karena memberikan alur kerja yang sistematis dalam proses data mining, mulai dari pemahaman masalah hingga penerapan hasil analisis. Alur metodologi penelitian ditunjukkan pada Gambar dibawah ini.



GAMBAR 1
(ALUR METODOLOGI PENELITIAN)

A. Business Understanding

Tahap business understanding dilakukan dengan mengidentifikasi permasalahan di SMA Hang Tuah 1 Surabaya, yaitu belum optimalnya pemanfaatan data akademik dan ekstrakurikuler dalam mendukung proses pembelajaran. Tujuan utama penelitian ini adalah mengelompokkan siswa berdasarkan karakteristik akademik dan non-akademik sebagai dasar penerapan pembelajaran diferensiasi.

B. Data Understanding

Tahap data understanding meliputi proses pengenalan struktur dan karakteristik data yang digunakan dalam penelitian. Dataset yang digunakan terdiri dari data nilai rapor siswa dari semester satu hingga semester lima serta data keaktifan ekstrakurikuler dari empat angkatan siswa. Pada tahap ini dilakukan pemeriksaan kelengkapan data, tipe data, serta distribusi nilai untuk memastikan data layak digunakan dalam proses analisis.

C. Data Preparation

Tahap data preparation mencakup proses pembersihan data, penanganan data kosong, serta normalisasi data. Normalisasi dilakukan untuk memastikan seluruh variabel berada pada skala yang sebanding sehingga tidak ada variabel tertentu yang mendominasi proses clustering.

D. Modeling

Tahap modeling dilakukan untuk membangun model pengelompokan siswa menggunakan algoritma K-Means Clustering. Sebelum proses pengelompokan dilakukan, penentuan jumlah cluster optimal ditentukan terlebih dahulu menggunakan metode Elbow dengan menganalisis pola penurunan nilai Within Cluster Sum of Squares (WCSS). Selanjutnya, algoritma K-Means diterapkan untuk mengelompokkan siswa berdasarkan karakteristik nilai akademik dan tingkat keaktifan ekstrakurikuler.

E. Evaluation

Tahap evaluation dilakukan untuk menilai kualitas hasil pengelompokan yang dihasilkan oleh algoritma K-Means. Evaluasi dilakukan menggunakan Silhouette Score sebagai metrik penilaian kualitas cluster. Nilai Silhouette Score yang diperoleh digunakan untuk memastikan bahwa jumlah cluster yang dipilih telah menghasilkan struktur pengelompokan yang layak sebelum dilakukan analisis lebih lanjut.

F. Deployment

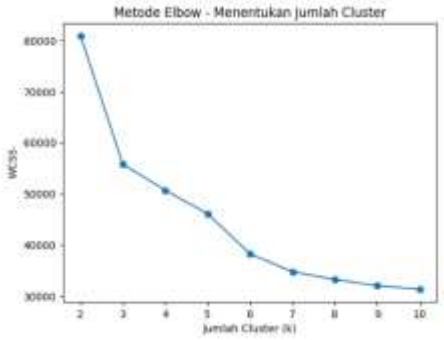
Tahap deployment dilakukan dengan mengimplementasikan hasil clustering ke dalam dashboard interaktif berbasis Streamlit. Dashboard ini dirancang untuk memudahkan guru dan pihak sekolah dalam melihat hasil pengelompokan siswa secara visual dan melakukan pemantauan secara real-time.

IV. HASIL DAN PEMBAHASAN

Penelitian ini menerapkan algoritma K-Means untuk mengelompokkan siswa SMA Hang Tuah 1 Surabaya berdasarkan dua variabel utama, yaitu nilai akademik yang direpresentasikan oleh nilai rapor dan keaktifan siswa dalam kegiatan ekstrakurikuler. Sebelum proses clustering dilakukan, data terlebih dahulu melalui tahap prapemrosesan, meliputi pemeriksaan kelengkapan data dan normalisasi, guna memastikan setiap variabel memiliki kontribusi yang seimbang dalam proses pengelompokan.

A. Modeling

Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow dengan menganalisis pola penurunan nilai *Within Cluster Sum of Squares* (WCSS). Metode ini bertujuan untuk mengidentifikasi titik di mana penurunan WCSS mulai melandai, yang mengindikasikan jumlah cluster paling efisien.

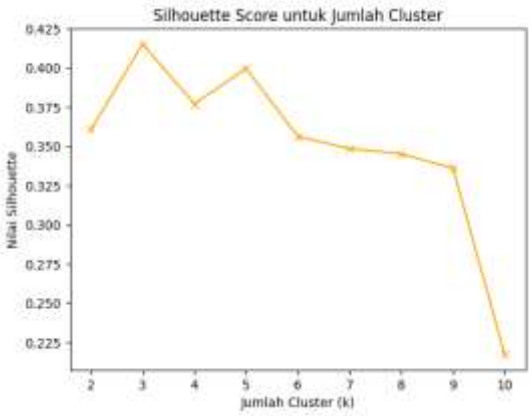


GAMBAR 2
(METODE ELBOW)

Berdasarkan Gambar 2, terlihat bahwa penurunan WCSS yang signifikan terjadi hingga jumlah cluster ke-3, kemudian cenderung melandai pada cluster berikutnya. Oleh karena itu, jumlah cluster optimal yang digunakan dalam penelitian ini adalah tiga cluster. Pemilihan jumlah cluster ini dianggap mampu merepresentasikan perbedaan karakteristik siswa secara lebih jelas tanpa menyebabkan pengelompokan yang terlalu rinci.

B. Evaluation

Selain metode Elbow, evaluasi kualitas hasil clustering juga dilakukan menggunakan Silhouette Score. Metode ini mengukur sejauh mana suatu objek memiliki kemiripan dengan cluster-nya sendiri dibandingkan dengan cluster lain. Nilai Silhouette Score berada pada rentang -1 hingga 1, di mana nilai mendekati 1 menunjukkan hasil pengelompokan yang semakin baik.



GAMBAR 3
(SILHOUETTE SCORE)

Hasil pengujian menunjukkan bahwa nilai Silhouette Score yang diperoleh sebesar 0,4152. Nilai ini mengindikasikan bahwa struktur cluster yang terbentuk berada pada kategori cukup baik dan memiliki tingkat pemisahan antar cluster yang memadai. Dengan demikian, hasil clustering dapat digunakan untuk analisis lanjutan dan penerapan rekomendasi pembelajaran.

C. Karakteristik Tiap Cluster

Berdasarkan hasil pengelompokan, masing-masing cluster memiliki karakteristik akademik dan non-akademik yang berbeda. Ringkasan karakteristik setiap cluster disajikan pada Tabel 1.

TABEL 1
(KARAKTERISTIK TIAP CLUSTER)

Cluster	Karakteristik Nilai Akademik	Karakteristik Ekstrakurikuler	Karakteristik Umum
Cluster 0	Tinggi	Sedang	Siswa dengan capaian akademik stabil
Cluster 1	Sedang	Rendah	Siswa yang memerlukan pendampingan motivasi
Cluster 2	Rendah	Tinggi	Siswa dengan potensi non-akademik menonjol

Berdasarkan Tabel 1, cluster pertama didominasi oleh siswa dengan nilai akademik tinggi serta tingkat keaktifan ekstrakurikuler yang sedang. Siswa pada cluster ini umumnya mampu mengikuti proses pembelajaran dengan baik dan memiliki potensi untuk diberikan program pengayaan atau pembelajaran dengan tingkat kesulitan yang lebih tinggi.

Cluster kedua terdiri dari siswa dengan capaian akademik sedang namun memiliki tingkat keaktifan ekstrakurikuler yang relatif rendah. Kondisi ini menunjukkan bahwa siswa dalam cluster ini memerlukan perhatian khusus, terutama dalam aspek motivasi belajar dan peningkatan keterlibatan dalam aktivitas sekolah.

Sementara itu, cluster ketiga menunjukkan karakteristik siswa dengan nilai akademik yang lebih rendah tetapi memiliki tingkat keaktifan ekstrakurikuler yang tinggi. Hal ini mengindikasikan adanya potensi kuat pada aspek non-akademik, sehingga pendekatan pembelajaran yang kontekstual dan berbasis minat dapat menjadi strategi yang tepat untuk mengembangkan potensi siswa pada cluster ini.

D. Visualisasi Dashboard Hasil Clustering

Untuk memberikan gambaran yang lebih jelas mengenai hasil pengelompokan siswa, dilakukan visualisasi distribusi cluster menggunakan grafik dua dimensi. Visualisasi ini memperlihatkan sebaran siswa berdasarkan nilai akademik dan keaktifan ekstrakurikuler, serta posisi masing-masing cluster.



GAMBAR 4
(VISUALISASI DASHBOARD)

Gambar diatas menyajikan distribusi jumlah dan persentase siswa pada masing-masing cluster hasil penerapan algoritma K-Means. Perbedaan proporsi siswa antar cluster mengindikasikan adanya variasi karakteristik akademik dan non-akademik yang cukup jelas, sehingga data siswa tidak bersifat homogen dan membentuk pola pengelompokan yang terstruktur.

E. Interpretasi Karakteristik Setiap Cluster

Untuk memahami lebih lanjut makna dari pola pengelompokan tersebut, dilakukan interpretasi terhadap karakteristik akademik dan non-akademik pada masing-masing cluster.



GAMBAR 5
(INTERPRETASI KARAKTERISTIK)

Gambar diatas menyajikan ringkasan karakteristik akademik dan non-akademik dari setiap cluster yang terbentuk. Interpretasi ini menunjukkan bahwa setiap cluster memiliki profil capaian akademik dan tingkat keterlibatan ekstrakurikuler yang berbeda, sehingga memperkuat hasil pengelompokan yang diperoleh melalui algoritma K-Means.

F. Implikasi Pembelajaran Diferensiasi

Variasi karakteristik yang teridentifikasi pada masing-masing cluster menunjukkan bahwa pendekatan pembelajaran yang seragam kurang efektif untuk diterapkan kepada seluruh siswa. Oleh karena itu, hasil clustering dapat dimanfaatkan sebagai dasar penerapan pembelajaran diferensiasi, di mana strategi pembelajaran disesuaikan dengan karakteristik akademik dan non-akademik setiap kelompok siswa. Siswa dengan capaian akademik tinggi dapat diarahkan pada program pengayaan, sementara siswa dengan capaian akademik sedang dan rendah memerlukan strategi pendampingan serta penguatan motivasi belajar. Di sisi lain, siswa dengan potensi non-akademik yang menonjol dapat difasilitasi melalui pendekatan pembelajaran kontekstual dan berbasis aktivitas. Temuan ini menegaskan bahwa hasil clustering tidak hanya bersifat analitis, tetapi juga memiliki implikasi praktis dalam perencanaan pembelajaran yang lebih adaptif dan berbasis data.

V. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means untuk mengelompokkan siswa berdasarkan nilai rapor dan keaktifan ekstrakurikuler di SMA Hang Tuah 1 Surabaya. Hasil penelitian menunjukkan bahwa tiga cluster merupakan jumlah yang paling optimal dan mampu menggambarkan perbedaan karakteristik siswa secara jelas. Pengelompokan ini memberikan dasar objektif bagi sekolah dalam menerapkan pembelajaran diferensiasi sesuai dengan kebutuhan masing-masing kelompok siswa. Implementasi hasil clustering dalam bentuk dashboard interaktif berbasis Streamlit juga mempermudah pemanfaatan hasil analisis oleh guru dan pihak sekolah. Penelitian selanjutnya dapat mengembangkan variabel yang digunakan, membandingkan metode clustering lain, serta menambahkan analisis historis untuk memantau perubahan karakteristik siswa dari waktu ke waktu.

REFERENSI

[1] S. Assyifa.dkk, “Pemenuhan Hak dan Kewajiban Warga Negara Indonesia di Bidang Pendidikan,” *Madani J. Multidisiplin*, vol. 1, no. 4, pp. 248–252, 2023.

[2] M. A. Wardana, “Penerapan Data Mining Pengelompokan Data Penjualan Motor di PT . Nusantara Surya Sakti Soppeng Menggunakan Metode K-Means Clustering,” vol. 8, no. April, 2025.

[3] Y. H. Chrisnanto and G. Abdullah, “The uses of educational data mining in academic performance analysis at higher education institutions (case study at UNJANI),” *Matrix J. Manaj. Teknol. dan Inform.*, vol. 11, no. 1, pp. 26–35, 2021, doi: 10.31940/matrix.v11i1.2330.

[4] D. Darlinda and J. N. Utamajaya, “Sistem Pendukung Keputusan Penerima Beasiswa Program Indonesia Pintar Menggunakan Metode Algoritma K-Means Clustering,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 167, 2022, doi: 10.30865/jurikom.v9i2.3971.

[5] S. Suraya, M. Sholeh, and D. Andayati, “Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa,” *Skanika*, vol. 6, no. 1, pp. 51–60, 2023, doi: 10.36080/skanika.v6i1.2982.

[6] B. Hartono, S. Eniyati, and K. Hadiono, “Perbandingan Metode Perhitungan Jarak pada Nilai Centroid dan Pengelompokan Data Menggunakan K-Means Clustering,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, p. 503, 2023, doi: 10.30865/json.v4i3.6021.

[7] K. Merdeka and D. I. Sd, “PEMBELAJARAN DIFERENSIASI DALAM,” vol. 7, pp. 10714–10721, 2024.

[8] B. Ii, K. Manggala, and K. Makassar, “Analisis penerapan pembelajaran berdiferensiasi dalam kurikulum merdeka di kelas iv upt spf sd inpres tello baru ii kecamatan manggala kota makassar,” vol. 2, no. 3, 2024.

[9] A. Mulyana, Y. Hermawan, J. Saputri, S. Informasi, F. Informatika, and I. Bisnis, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Rekomendasi Pilihan Program Studi Pada Mahasiswa Baru Penerimaan mahasiswa baru merupakan agenda tahunan yang dilaksanakan oleh perguruan tinggi untuk menjaring calon mahasiswa potensial,” vol. 5, no. 1, pp. 60–72, 2024, doi: 10.31284/j.kernel.2024.v5i1.7624.

[10] Mubarok, “Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi) Implementasi Data Mining untuk Clustering Lowongan,” vol. 9, no. June, pp. 703–712, 2025.