P-ISSN : 2580-4316
https://doi.org/10.37817/ikraith-informatika.v10i2 E-ISSN : 2654-8054

Penerapan Algoritma K-Means untuk Identifikasi Jurusan Unggulan Berdasarkan Nilai Siswa (Studi Kasus: MTs Tarbiyatul Banin

Riyan Hadi Prabowo[1], Shasha Ramadhani Putri*[2], Abdullah Muttaqin[3], Atsna Rifqi Habibur Rahman[4], Muhammad Arifin[5]
[1,2,3,4,5]Magister Sistem Informasi, Universitas Muria Kudus, Kudus

E-mail: [1]re.wirohadi@gmail.com, *[2]202505010@std.umk.ac.id, [3]muttaqinab@gmail.com
[4]atsnarifqi@gmail.com, [5]arifin.m@umk.ac.id

ABSTRAK

Penelitian ini dilakukan untuk mengidentifikasi jurusan unggulan di MTs Tarbiyatul Banin Pekalongan, Pati, dengan menerapkan algoritma K-Means Clustering pada data nilai 17 mata pelajaran dari 226 siswa. Tahapan metodologi meliputi praproses data berupa pembersihan dan standarisasi fitur, penentuan jumlah klaster optimal menggunakan metode Elbow yang menghasilkan tiga klaster, proses pengelompokan dengan K-Means, serta evaluasi menggunakan Silhouette Coefficient. Hasil penelitian menunjukkan tiga klaster performa, yaitu kategori Unggul dengan rata-rata nilai tertinggi, kategori Menengah dengan performa standar, dan kategori Dasar dengan performa terendah. Nilai Silhouette Coefficient yang diperoleh sebesar 0,2208. Berdasarkan proporsi siswa yang masuk ke dalam kategori Unggul, jurusan Sains Riset Non Boarding D memiliki persentase tertinggi sebesar 53,12%, diikuti Tahfidz Boarding A sebesar 36,36%, dan Sains Riset Boarding C sebesar 31,25%. Dengan demikian, jurusan Sains Riset Non Boarding D dan Tahfidz Boarding A secara objektif teridentifikasi sebagai jurusan unggulan karena memiliki konsentrasi siswa berprestasi tinggi yang paling dominan.

Kata kunci : ***Data Mining, K-Means Clustering, jurusan unggulan, nilai mata pelajaran, MTs***

ABSTRACT

This study was conducted to identify superior majors at MTs Tarbiyatul Banin Pekalongan, Pati, by applying the K-Means Clustering algorithm to grade data from 17 subjects across 226 students. The methodology consisted of data preprocessing including data cleaning and feature standardization, determination of the optimal number of clusters using the Elbow method which yielded three clusters, clustering using K-Means, and evaluation using the Silhouette Coefficient. The results revealed three performance clusters, namely the Superior category with the highest average grades, the Intermediate category with standard performance, and the Basic category with the lowest performance. The Silhouette Coefficient value obtained was 0.2208. Based on the proportion of students classified into the Superior category, the Sains Riset D Non Boarding major recorded the highest percentage at 53.12%, followed by Tahfidz Boarding A at 36.36%, and Sains Riset C Boarding at 31.25%. Therefore, Sains Riset D Non Boarding and Tahfidz Boarding A were objectively identified as superior majors due to their dominant concentration of high-achieving students.

P-ISSN : 2580-4316
E-ISSN : 2654-805 https://doi.org/10.37817/ikraith-informatika.v10i2

Keyword : ***Data Mining, K-Means Clustering, leading academic programs, subject scores, Islamic junior high school (MTs)***

1. PENDAHULUAN

Di era digital saat ini, pemanfaatan teknologi informasi khususnya data mining telah berkembang pesat di berbagai sektor, termasuk pendidikan (Sindrawati dkk., 2024). Data mining merupakan proses penggalian pola tersembunyi dari kumpulan data yang besar. Salah satu teknik yang populer adalah clustering (pengelompokan), yang bertujuan untuk membagi data ke dalam kelompok-kelompok sehingga anggota dalam satu kelompok memiliki kemiripan yang tinggi sementara antar kelompok memiliki perbedaan yang signifikan (Sindrawati dkk., 2024). Dalam konteks pendidikan, clustering dapat digunakan untuk mengelompokkan siswa berdasarkan berbagai kriteria seperti nilai akademik, minat, bakat, dan kemampuan keagamaan (Nafuri dkk., 2022).

Permasalahan umum yang dihadapi sekolah adalah proses penentuan jurusan bagi siswa baru yang seringkali masih dilakukan secara manual dan subjektif. Hal ini berpotensi menyebabkan ketidaksesuaian antara jurusan yang dipilih dengan kemampuan dan minat siswa, yang pada akhirnya berdampak pada performa belajar dan motivasi mereka. Di MTs Tarbiyatul Banin Pekalongan, terdapat beberapa jurusan unggulan seperti Tahfidz Boarding, Tahfidz Non Boarding, Sains Riset Boarding, Sains Riset Non Boarding, serta Reguler. Namun, selama ini penetapan jurusan unggulan masih bersifat persepsi umum tanpa didasarkan pada bukti data yang sistematis, sehingga sekolah kesulitan dalam mengevaluasi secara objektif jurusan mana yang benar-benar unggul dalam capaian akademik.

Algoritma K-Means Clustering merupakan salah satu algoritma clustering yang paling sederhana dan banyak digunakan di berbagai bidang karena efisiensi dan skalabilitasnya (MacQueen, 1967). Algoritma ini bekerja dengan cara mempartisi data ke dalam k klaster, di mana setiap data akan masuk ke dalam klaster dengan jarak terdekat terhadap pusat klaster (centroid). Dibandingkan dengan algoritma clustering lainnya seperti K-Medoids, K-Means memiliki keunggulan dalam kecepatan komputasi yang lebih tinggi pada dataset yang besar dan bebas dari outlier yang signifikan (Syamfithriani dkk., 2023).

Beberapa penelitian terdahulu telah membuktikan relevansi K-Means dalam konteks pendidikan. Muasaroh & Fatah mengimplementasikan K-Means untuk optimasi pembentukan kelas unggulan dengan mempertimbangkan prestasi akademik dan kondisi ekonomi siswa (Muasaroh & Fatah, 2024). Hutagalung, dkk berhasil memetakan siswa kelas unggulan berdasarkan nilai akademik menggunakan K-Means sehingga sekolah dapat memberikan rekomendasi kelas yang lebih tepat sasaran (Hutagalung dkk., 2022). Dewi, dkk menunjukkan bahwa K-Means mampu mengidentifikasi karakteristik siswa berprestasi secara akurat berdasarkan keaktifan dalam proses pembelajaran (Dewi dkk., 2022). Qusyairi, dkk membuktikan bahwa kombinasi K-Means dengan metode Elbow mampu meningkatkan akurasi pengelompokan prestasi siswa secara otomatis (Qusyairi dkk., 2024). Berdasarkan kajian literatur tersebut, terdapat celah penelitian (*research gap*) di mana belum ada studi yang secara spesifik mengidentifikasi jurusan unggulan pada madrasah dengan karakteristik jurusan berbasis keagamaan seperti Tahfidz

https://doi.org/10.37817/ikraith-informatika.v10i2
P-ISSN : 2580-4316
E-ISSN : 2654-8054

Boarding menggunakan K-Means Clustering.

Algoritma K-Means adalah algoritma clustering partisi yang populer dan efisien (MacQueen, 1967). Algoritma ini bekerja dengan mempartisi data ke dalam k klaster berdasarkan jarak Euclidean terhadap centroid, dan iterasi berlanjut hingga centroid konvergen. Kelebihan K-Means adalah komputasi yang cepat dan mudah diimplementasikan, namun kelemahannya adalah sensitif terhadap outlier dan inisialisasi centroid awal (Wani, 2024). Untuk mengevaluasi kualitas klaster, digunakan Silhouette Coefficient yang mengukur seberapa mirip suatu titik data dengan klasternya sendiri dibandingkan klaster lain (Rousseeuw, 1987), dengan nilai mendekati 1 mengindikasikan klaster yang padat dan terpisah dengan baik.

Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma K-Means Clustering dalam mengidentifikasi jurusan unggulan siswa berdasarkan data nilai 17 mata pelajaran dari 226 siswa MTs Tarbiyatul Banin Pekalongan. Diharapkan hasil penelitian ini dapat menjadi rekomendasi bagi institusi pendidikan dalam melakukan evaluasi dan pengembangan jurusan secara lebih efektif, efisien, dan berorientasi pada pengembangan potensi siswa secara optimal.

2. METODE PENELITIAN

2.1 Instrumen dan Subjek Penelitian

Subjek penelitian adalah 226 siswa MTs Tarbiyatul Banin Pekalongan (setelah pembersihan data dari 227 menjadi 226 karena satu data tidak lengkap) yang terbagi dalam 8 kelas. Instrumen data berupa 17 nilai mata pelajaran per siswa, yaitu: QH, AA, FIK, SKI, BAR, PP, BINDO, MTK, IPA, IPS, BING, PJOK, INFO, SBP, BJAWA, KKUNG, dan KNU. Semua nilai berskala 0-100. Data sekunder diperoleh dari bagian kurikulum dan kesiswaan dalam format Excel.

Tabel 1 Distribusi Siswa per Kelas/Jurusan

Kode Kelas	*Jurusan*	*Jumlah Siswa*
A	TAHFIDZ BOARDING	22
B	TAHFIDZ NON BOARDING	31
C	SAINS RISET BOARDING	16
D	SAINS RISET NON BOARDING	32
E	REGULER	32
F	REGULER	31
G	REGULER	32
H	REGULER	30
Total		226

2.2 Prosedur Penelitian

Penelitian ini menggunakan metode eksperimen dengan model pengembangan prosedural. Tahapan penelitian dirancang secara sistematis untuk memastikan kualitas hasil klasterisasi yang optimal. Secara keseluruhan, prosedur penelitian terdiri atas lima tahapan utama yang saling berkaitan, yaitu pengumpulan data, pra-pemrosesan data, penentuan jumlah klaster optimal, implementasi algoritma K-Means, serta evaluasi hasil klasterisasi.

2.3 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data primer yang diperoleh secara langsung dari bagian kurikulum MTs Tarbiyatul Banin Pekalongan. Data tersebut berupa nilai akademik siswa dari 17 mata pelajaran, yang mencakup mata pelajaran umum maupun mata pelajaran peminatan, dan dikumpulkan dari 226 siswa aktif yang terdistribusi ke dalam beberapa jurusan

P-ISSN : 2580-4316
E-ISSN : 2654-805 https://doi.org/10.37817/ikraith-informatika.v10i2

yang tersedia di sekolah tersebut. Data disediakan dalam format Microsoft Excel dan memuat informasi berupa identitas siswa, jurusan, serta nilai akhir masing-masing mata pelajaran.

Pengumpulan data dilakukan dengan metode dokumentasi, yaitu dengan mengambil data nilai dari sistem administrasi kurikulum sekolah secara resmi melalui izin yang telah diperoleh dari pihak terkait. Pemilihan data nilai akademik sebagai atribut klasterisasi didasarkan pada pertimbangan bahwa nilai akademik merupakan representasi kuantitatif yang paling objektif dari kompetensi dan prestasi siswa, sehingga dapat digunakan sebagai dasar pengelompokan yang valid [12].

2.4 Pra-pemrosesan Data

Tahap pra-pemrosesan data merupakan langkah krusial yang sangat menentukan kualitas hasil klasterisasi. Data mentah yang diperoleh dari lapangan umumnya mengandung berbagai permasalahan seperti nilai yang hilang, data tidak konsisten, maupun perbedaan skala antar fitur, yang apabila tidak ditangani dapat menghasilkan pengelompokan yang bias dan tidak akurat. Proses pra-pemrosesan dalam penelitian ini terdiri atas dua tahapan utama, yaitu data cleaning dan normalisasi data.

2.4.1 Data Cleaning

Data cleaning dilakukan untuk memastikan bahwa data yang digunakan bersih dari inkonsistensi dan ketidaklengkapan sebelum masuk ke tahap analisis. Hasil pemeriksaan menemukan sebanyak 1 (satu) data yang tidak lengkap, yaitu terdapat nilai kosong pada satu atau lebih atribut mata pelajaran. Data tersebut kemudian dihapus menggunakan pendekatan listwise deletion sehingga jumlah data yang diproses lebih lanjut adalah 226 data. Penghapusan data yang tidak lengkap merupakan salah satu pendekatan umum yang diterapkan dalam tahap data cleaning untuk mencegah terjadinya bias pada hasil analisis klasterisasi (Setiawan & Triayudi, 2024).

2.4.2 Normalisasi Data (Standarisasi Fitur)

Setelah tahap data cleaning selesai, dilakukan normalisasi data menggunakan metode standardisasi Z-score yang diimplementasikan melalui fungsi StandardScaler dari library scikit-learn Python. Normalisasi data merupakan tahapan wajib yang perlu dilakukan sebelum menerapkan algoritma berbasis jarak seperti K-Means, karena perbedaan skala antar fitur dapat menyebabkan fitur dengan rentang nilai lebih besar mendominasi perhitungan jarak Euclidean dan menghasilkan klaster yang tidak representatif terhadap keseluruhan data (Wongoutong, 2024).

Standardisasi Z-score bekerja dengan mengubah setiap nilai fitur sehingga memiliki rata-rata nol ($\mu = 0$) dan standar deviasi satu ($\sigma = 1$) (Ultsch & Lötsch, 2022). Formula yang digunakan adalah sebagai berikut:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

di mana x adalah nilai asli, μ adalah rata-rata fitur, dan σ adalah standar deviasi fitur. Setelah proses standardisasi, seluruh 17 fitur nilai mata pelajaran berada pada skala yang seragam sehingga setiap mata pelajaran memiliki kontribusi yang setara dan proporsional dalam proses pengelompokan siswa.

2.5 Penentuan Jumlah Klaster Optimal

Penentuan jumlah klaster yang tepat merupakan salah satu tantangan utama dalam penerapan algoritma K-Means, karena nilai k harus ditentukan sebelum proses klasterisasi dilakukan. Pada penelitian ini, penentuan jumlah klaster optimal dilakukan dengan menggunakan dua metode secara bersamaan pada rentang nilai k = 2 hingga k = 8, yaitu metode Elbow dan Silhouette Coefficient.

https://doi.org/10.37817/ikraith-informatika.v10i2
P-ISSN : 2580-4316
E-ISSN : 2654-8054

2.5.1 Metode Elbow

Metode Elbow mengevaluasi nilai Within-Cluster Sum of Squares (WCSS) atau Sum of Squared Errors (SSE) untuk setiap kandidat nilai k. Nilai k optimal ditentukan pada titik di mana penurunan WCSS mulai melandai secara signifikan dan membentuk sudut seperti siku (elbow) pada grafik, yang mengindikasikan bahwa penambahan jumlah klaster berikutnya tidak lagi memberikan peningkatan kualitas pengelompokan yang berarti (Wicaksono dkk., 2024).

2.5.2 Validasi dengan Silhouette Coefficient

Sebagai validator tambahan, Silhouette Coefficient turut dihitung untuk setiap kandidat nilai k. Nilai Silhouette Coefficient rata-rata yang lebih tinggi mengindikasikan bahwa data dalam suatu klaster lebih kohesif secara internal dan lebih terpisah dari klaster lainnya. Kombinasi kedua metode ini digunakan untuk meningkatkan keandalan keputusan dalam memilih nilai k yang paling optimal (Juanita & Cahyono, 2024). Berdasarkan hasil evaluasi kedua metode tersebut, nilai k = 3 ditetapkan sebagai jumlah klaster yang optimal untuk dataset penelitian ini.

2.6 Implementasi Algoritma K-Means

Implementasi algoritma K-Means dilakukan menggunakan bahasa pemrograman Python dengan memanfaatkan library scikit-learn. Jarak Euclidean antara titik data x dan centroid c dihitung menggunakan formula (Wani, 2024):

$$d(x, c) = \sqrt{\sum(x_i - c_i)^2} \quad (2)$$

Proses iteratif K-Means berlangsung dengan dua langkah utama yang diulang secara bergantian: (1) tahap assignment, yaitu setiap titik data ditetapkan ke klaster dengan centroid terdekat, dan (2) tahap update, yaitu centroid setiap klaster diperbarui berdasarkan rata-rata seluruh anggota klaster tersebut. Iterasi berlanjut hingga posisi centroid tidak mengalami perubahan signifikan atau hingga batas iterasi maksimum tercapai. Parameter yang digunakan ditunjukkan pada Tabel 2.

Tabel 2 Parameter Implementasi K-Means

Parameter	Nilai	Keterangan
n_clusters	3	Jumlah klaster sesuai hasil penentuan optimal
init	'k-means++'	Metode inisialisasi centroid
n_init	10	Jumlah pengulangan inisialisasi berbeda
max_iter	300	Batas maksimum iterasi per pengulangan
random_state	42	Nilai seed untuk reproduktivitas hasil

2.7 Evaluasi Hasil Klasterisasi dan Analisis Jurusan Unggulan

Evaluasi kualitas klasterisasi dilakukan menggunakan Silhouette Coefficient, yang merupakan metrik internal yang tidak memerlukan label data acuan (ground truth) sehingga sesuai untuk mengevaluasi hasil klasterisasi yang bersifat unsupervised. Nilai Silhouette Coefficient untuk setiap titik data i dihitung menggunakan formula (Chicco dkk., 2025):

$$SC(i) = \frac{(b(i) - a(i))}{\max(a(i), b(i))} \quad (3)$$

di mana a(i) adalah rata-rata jarak antara titik data i dengan seluruh titik lain dalam klaster yang sama (intra-cluster distance), dan b(i) adalah rata-rata jarak minimum

P-ISSN : 2580-4316
E-ISSN : 2654-805 https://doi.org/10.37817/ikraith-informatika.v10i2

antara titik data i dengan seluruh titik dalam klaster terdekat lainnya (nearest-cluster distance) (Juanita & Cahyono, 2024). Selain evaluasi kuantitatif, dilakukan pula analisis karakteristik masing-masing klaster secara kualitatif berdasarkan nilai centroid akhir setiap klaster untuk memberikan label kategori yang bermakna (Unggul, Menengah, atau Dasar). Tahapan terakhir adalah perhitungan proporsi siswa per jurusan yang masuk ke dalam klaster Unggul, yang kemudian digunakan sebagai dasar penentuan jurusan unggulan secara objektif berbasis data.

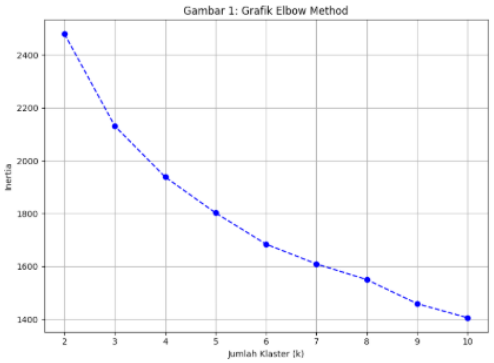
3. HASIL DAN PEMBAHASAN

3.1 Pra-pemrosesan Data

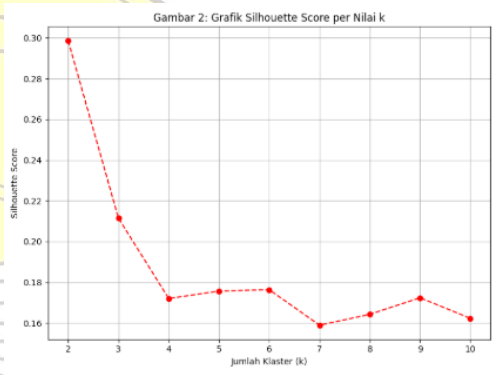
Data awal terdiri dari 227 record nilai siswa yang mencakup 17 mata pelajaran. Tahap cleaning dilakukan dengan menghapus 1 data yang tidak lengkap sehingga tersisa 226 data bersih. Selanjutnya, seluruh fitur dinormalisasi menggunakan StandardScaler (Z-score normalization) agar setiap atribut memiliki bobot yang seimbang dalam perhitungan jarak Euclidean. Normalisasi ini penting karena nilai antar mata pelajaran berpotensi memiliki distribusi yang berbeda meskipun berada pada skala 0-100 (Wongoutong, 2024).

3.2 Penentuan Jumlah Klaster Optimal

Untuk menentukan jumlah klaster optimal, digunakan kombinasi metode Elbow berdasarkan nilai inertia (Within-Cluster Sum of Squares/WCSS) dan Silhouette Coefficient pada rentang k=2 hingga k=8. Hasil analisis menunjukkan bahwa penurunan inertia mulai melandai secara signifikan pada k=3. Konfirmasi menggunakan Silhouette Coefficient menghasilkan nilai tertinggi pada k=3 sebesar 0,2208, yang memperkuat keputusan pemilihan k=3 sebagai jumlah klaster optimal (Orsoni dkk., 2023).



Gambar 1. Grafik Elbow Method untuk Penentuan Jumlah Klaster Optimal



Gambar 2. Grafik Silhouette Score per Jumlah Klaster

3.3 Hasil Klasterisasi K-Means

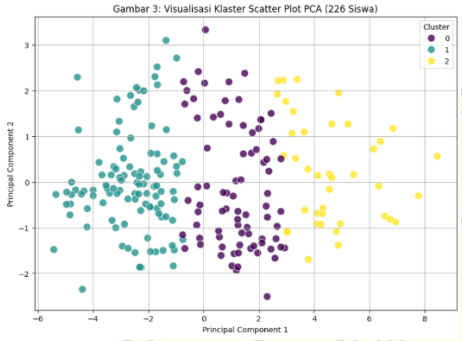
Implementasi K-Means dengan parameter n_clusters=3, init='k-means++', n_init=10, max_iter=300, dan random_state=42 menghasilkan tiga klaster dengan distribusi: Klaster 0 sebanyak 88 siswa (38,94%), Klaster 1 sebanyak 75 siswa (33,19%), dan Klaster 2 sebanyak 63 siswa (27,88%). Berdasarkan analisis nilai rata-rata centroid per mata pelajaran:

a. Klaster 2 (Kategori Unggul): kelompok siswa dengan rata-rata nilai tertinggi pada hampir seluruh mata pelajaran, menunjukkan performa akademik yang konsisten dan merata di semua bidang studi.
b. Klaster 0 (Kategori Menengah): memiliki performa rata-rata yang berada di antara klaster unggul dan

https://doi.org/10.37817/ikraith-informatika.v10i2
P-ISSN : 2580-4316
E-ISSN : 2654-8054

klaster dasar, mencerminkan kemampuan akademik yang cukup memadai namun belum optimal.

c. Klaster 1 (Kategori Dasar): kelompok siswa dengan rata-rata nilai terendah, yang memerlukan perhatian dan intervensi pedagogis lebih lanjut dari pihak sekolah.



Gambar 3. Visualisasi Hasil Klasterisasi K-Means

3.4 Evaluasi Kualitas Klaster

Evaluasi kualitas klaster menggunakan Silhouette Coefficient menghasilkan nilai sebesar 0,2208. Berdasarkan interpretasi Rousseeuw, nilai ini berada pada rentang 0,2-0,5 yang menunjukkan bahwa struktur klaster yang terbentuk tergolong cukup baik (*reasonable structure*) (Rousseeuw, 1987). Nilai Silhouette Coefficient yang tidak terlalu tinggi ini dapat disebabkan oleh karakteristik data akademik yang secara alamiah memiliki distribusi yang tidak terlalu terpisah antar kelompok, mengingat seluruh siswa berada dalam satu institusi dengan kurikulum yang relatif seragam. Hal ini sejalan dengan temuan Niwana, dkk yang juga memperoleh nilai Silhouette Coefficient pada kisaran yang serupa pada data akademik tingkat menengah (Nirwana dkk., 2022).

3.5 Distribusi Klaster per Jurusan dan Identifikasi Jurusan Unggulan

Untuk mengidentifikasi jurusan unggulan, dihitung persentase siswa dari masing-masing kelas yang masuk ke dalam Klaster 2 (Klaster Unggul). Hasil perhitungan ditampilkan pada Tabel 3 berikut.

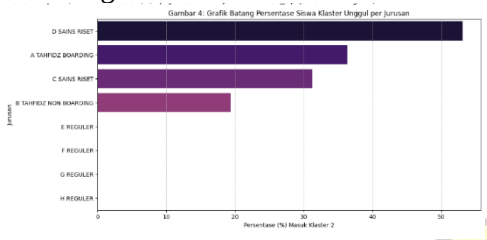
Tabel 3 Proporsi Siswa Klaster Unggul (Klaster 2) per Kelas/Jurusan

Jurusan	*Total Siswa*	*Siswa Klaster Unggul*	*Persentase (%)*
D SAINS RISET NON BOARDING	32	17	53,12
A TAHFIDZ BOARDING	22	8	36,36
C SAINS RISET BOARDING	16	5	31,25
B TAHFIDZ NON BOARDING	31	6	19,35
E REGULER	32	0	0,00
F REGULER	31	0	0,00
G REGULER	32	0	0,00
H REGULER	30	0	0,00

Berdasarkan Tabel 3, D Sains Riset Non Boarding memiliki proporsi siswa terbanyak di Klaster Unggul yakni sebesar 53,12%, yang berarti lebih dari separuh siswanya tergolong berprestasi tinggi. Posisi kedua ditempati A Tahfidz Boarding (36,36%), diikuti C Sains Riset Boarding (31,25%), dan B Tahfidz Non Boarding (19,35%). Sementara itu, seluruh kelas Reguler (E, F, G, H) tidak memiliki siswa yang masuk ke dalam Klaster Unggul, yang mengindikasikan bahwa performa akademik kelas Reguler secara keseluruhan berada di bawah kelas-kelas unggulan lainnya. Dengan demikian, secara objektif berbasis data, D Sains Riset Non Boarding dan A Tahfidz Boarding teridentifikasi sebagai dua kelas dengan performa akademik tertinggi di

P-ISSN : 2580-4316
E-ISSN : 2654-805 https://doi.org/10.37817/ikraith-informatika.v10i2

MTs Tarbiyatul Banin Pekalongan. Temuan ini sejalan dengan penelitian (Hutagalung dkk., 2022) yang juga berhasil mengidentifikasi kelas unggulan secara objektif melalui metode K-Means Clustering .



Gambar 4. Persentase Siswa Klaster Unggul per Kelas/Jurusan

4. KESIMPULAN

Berdasarkan implementasi algoritma K-Means Clustering pada data nilai 17 mata pelajaran dari 226 siswa MTs Tarbiyatul Banin Pekalongan, dapat disimpulkan sebagai berikut. Pertama, jumlah klaster optimal yang dihasilkan adalah k=3 dengan nilai Silhouette Coefficient sebesar 0,2208 yang tergolong cukup baik (*reasonable structure*), menunjukkan bahwa data nilai siswa MTs Tarbiyatul Banin Pekalongan dapat dikelompokkan secara bermakna meskipun karakteristik akademik antar siswa dalam satu institusi cenderung homogen. Kedua, tiga klaster performa yang terbentuk adalah Klaster 2 (Kategori Unggul) dengan rata-rata nilai tertinggi di hampir seluruh mata pelajaran, Klaster 0 (Kategori Menengah) dengan performa standar, dan Klaster 1 (Kategori Dasar) dengan rata-rata nilai terendah yang memerlukan intervensi lebih lanjut. Ketiga, pada tingkat kelas, D Sains Riset Non Boarding (53,12%), A Tahfidz Boarding (36,36%), dan C Sains Riset Boarding (31,25%) merupakan tiga kelas dengan konsentrasi siswa berprestasi tinggi yang paling dominan. Keempat, berdasarkan bukti data yang sistematis, D Sains Riset Non Boarding dan A Tahfidz Boarding secara objektif teridentifikasi sebagai kelas unggulan di MTs Tarbiyatul Banin Pekalongan. Hasil ini membuktikan bahwa algoritma K-Means Clustering dapat menjadi alat bantu pengambilan keputusan yang efektif dan objektif dalam konteks evaluasi akademik di institusi pendidikan berbasis keagamaan.

DAFTAR PUSTAKA

Chicco, D., Campagner, A., Spagnolo, A., Ciucci, D., & Jurman, G. (2025). The Silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index , Calinski-Harabasz index , Shannon entropy , and Gap statistic for unsupervised clustering internal evaluation of two convex clusters. *PeerJ Computer Science*. https://doi.org/10.7717/peerj-cs.3309

Dewi, F. P., Aryni, P. S., & Umaidah, Y. (2022). Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *7*(2), 111–121.

Hutagalung, J., Syahputra, Y. H., & Tanjung, Z. P. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. *Jurnal Teknik Informatika dan Sistem Informasi*, *9*(1), 606–620.

Juanita, S., & Cahyono, R. D. (2024). Klastering K-Means dengan Perbandingan Metode Elbow dan Silhouette untuk Pengelompokan Obat Berdasarkan Ulasan. *Jurnal Teknik Informatika (JUTIF)*, *5*(1), 283–289.

MacQueen, J. (1967). Some Methods For Classification and Analysis of Multivariate Observations. *In Proceedings of the 5th Berkeley Symposium on Mathematical*

Statistics and Probability, *1*, 281–297.

Muasaroh, Y. I., & Fatah, Z. (2024). Implementasi RapidMiner dalam Optimasi Pembentukan Kelas Unggulan Menggunakan K-Means Clustering. *Jurnal Advance Research Informatika*, *3*(1), 66–72.

Nafuri, A. F. M., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering Analysis for Classifying Student Academic Performance in Higher Education. *Applied Science*, *12*.

Nirwana, S. D., Jambak, M. I., & Bardadi, A. (2022). Perbandingan Algoritma K-Means dan K-Medoids dalam Clustering Rata-rata Penambahan Kasus Covid-19 Berdasarkan Kota/Kabupaten di Provinsi Sumatera Selatan. *JSiI : Jurnal Sistem Informasi*, *9*(2), 126–131. https://doi.org/10.30656/jsii.v9i2.5127

Orsoni, M., Giovagnoli, S., Garofalo, S., Magri, S., Benvenuti, M., Mazzoni, E., & Benassi, M. (2023). Preliminary evidence on machine learning approaches for clusterizing students ' cognitive profile. *Heliyon*, *9*(3). https://doi.org/10.1016/j.heliyon.2023.e14506

Qusyairi, M., Hidayatullah, Z., & Sandi, A. (2024). Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow. *Infotek : Jurnal Informatika dan Teknologi*, *7*(2), 500–510.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Setiawan, I. D., & Triayudi, A. (2024). Penerapan Data Mining Dengan Menggunakan Algoritma Clustering K-Means Untuk Pembagian Jurusan Pada Sekolah Menengah Atas. *Journal of Computer System and Informatics (JoSYC)*, *5*(2), 380–392. https://doi.org/10.47065/josyc.v5i2.4970

Sindrawati, Syaripudin, D., & Raharja, A. R. (2024). Penerapan Algoritma K-Means Clustering pada Data Nilai Siswa untuk Menentukan Kelompok Penerima Beasiswa.pdf. *SisInfo*, *6*(2), 47–55.

Syamfithriani, T. S., Mirantika, N., & Trisudarmo, R. (2023). Perbandingan Algoritma K-Means dan K-Medoids Untuk Pemetaan Daerah Penanganan Diare Pada Balita di Kabupaten Kuningan. *Jurnal Sistem Informasi Bisnis*, 132–139. https://doi.org/10.21456/vol12iss2pp132-139

Ultsch, A., & Lötsch, J. (2022). Euclidean distance-optimized data transformation for cluster analysis in biomedical data (EDOtrans). *BMC Bioinformatics*, *23*. https://doi.org/https://doi.org/10.1186/s12859-022-04769-w

Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: Exploring limitations and innovative solutions. *PeerJ Computer Science*. https://doi.org/10.7717/peerj-cs.2286

Wicaksono, A. P., Widjaja, S., Nugroho, M. F., Christina, & Putri, P. (2024). Analisa Nilai K terhadap Pengelompokkan Penjualan Tiket pada K-Means dengan Metode Elbow dan Silhouette Elbow and Silhouette Methods for K Value Analysis of Ticket Sales Grouping. *SISTEMASI: Jurnal Sistem Informasi*, *13*(1), 28–38.

Wongoutong, C. (2024). The impact of neglecting feature scaling in k-means clustering. *PLOS ONE*, 1–19. https://doi.org/10.1371/journal.pone.0310839