

Analisis Komparatif Algoritma K-Means dan K-Medoids dalam Klasterisasi Minat Belajar Siswa MDTA

William Ramdhan^{1*}

¹ Fakultas Ilmu Komputer, Sistem Informasi, Universitas Royal, Kisaran, Indonesia

Email: ^{1*}william.ramdhan052@gmail.com

Email Penulis Korespondensi: william.ramdhan052@gmail.com

Abstrak– Minat belajar merupakan salah satu faktor penting yang memengaruhi keberhasilan siswa dalam mengikuti proses pembelajaran. Namun, identifikasi tingkat minat belajar siswa sering kali dilakukan secara subjektif sehingga kurang mampu menggambarkan kondisi siswa secara menyeluruh. Penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Means dan K-Medoids dalam mengelompokkan minat belajar siswa MDTA Qur'an Kisaran berdasarkan data akademik dan nonakademik. Dataset yang digunakan terdiri dari 119 siswa dengan enam variabel, yaitu nilai akademik, persentase kehadiran, penilaian hafalan, kemampuan bacaan Al-Qur'an, penilaian akhlak, dan partisipasi siswa. Proses klasterisasi dilakukan dengan jumlah cluster sebanyak tiga kelompok yang merepresentasikan kategori minat belajar tinggi, sedang, dan rendah. Evaluasi kualitas cluster dilakukan menggunakan Davies-Bouldin Index (DBI) dan Silhouette Coefficient. Hasil penelitian menunjukkan bahwa algoritma K-Means menghasilkan nilai DBI sebesar 1,5839 dan Silhouette sebesar 0,2186, sedangkan algoritma K-Medoids menghasilkan nilai DBI sebesar 1,5950 dan Silhouette sebesar 0,2046. Berdasarkan hasil evaluasi tersebut, algoritma K-Means menunjukkan kinerja yang lebih baik dibandingkan K-Medoids dalam mengelompokkan minat belajar siswa. Hasil klasterisasi yang diperoleh dapat digunakan sebagai pendukung pengambilan keputusan dalam pembinaan siswa, penyusunan strategi pembelajaran, serta proses seleksi peserta Pekan Olahraga dan Seni Antar Diniyah (PORSADIN) secara lebih objektif dan berbasis data.

Kata Kunci: Data Mining; K-Means; K-Medoids; Klasterisasi; Minat Belajar; MDTA

Abstract- Learning interest is one of the important factors that influence student success in following the learning process. However, identification of students' learning interest levels is often done subjectively so that it is less able to describe the student's condition as a whole. This study aims to compare the performance of the K-Means and K-Medoids algorithms in grouping the learning interests of MDTA Qur'an Kisaran students based on academic and non-academic data. The dataset used consists of 119 students with six variables, namely academic grades, attendance percentage, memorization assessment, Qur'an reading ability, moral assessment, and student participation. The clustering process was carried out with three clusters representing the categories of high, medium, and low learning interest. Cluster quality evaluation was carried out using the Davies-Bouldin Index (DBI) and Silhouette Coefficient. The results showed that the K-Means algorithm produced a DBI value of 1.5839 and a Silhouette of 0.2186, while the K-Medoids algorithm produced a DBI value of 1.5950 and a Silhouette of 0.2046. Based on the evaluation results, the K-Means algorithm performed better than K-Medoids in clustering student learning interests. The clustering results can be used to support decision-making in student development, developing learning strategies, and selecting participants for the Inter-Islamic Sports and Arts Week (PORSADIN) in a more objective and data-driven manner.

Keywords: Data Mining; K-Means; K-Medoids; Clustering; Learning Interest; MDTA

1. PENDAHULUAN

Madrasah Diniyah Takmiliyah Awaliyah (MDTA) merupakan lembaga pendidikan keagamaan nonformal yang berperan dalam meningkatkan kemampuan membaca Al-Qur'an, pemahaman ajaran Islam, serta pembentukan karakter dan akhlak peserta didik. Keberhasilan proses pembelajaran di lingkungan MDTA tidak hanya ditentukan oleh kualitas tenaga pendidik dan kurikulum yang diterapkan, tetapi juga dipengaruhi oleh tingkat minat belajar siswa. Minat belajar merupakan salah satu faktor internal yang berpengaruh terhadap keterlibatan siswa dalam proses pembelajaran. Siswa yang memiliki minat belajar tinggi cenderung lebih aktif, disiplin, antusias, dan memiliki motivasi yang kuat dalam mengikuti kegiatan belajar, sedangkan siswa dengan minat belajar rendah cenderung pasif, kurang fokus, dan menunjukkan tingkat partisipasi yang lebih rendah selama proses pembelajaran berlangsung.

MDTA Qur'an Kisaran membina puluhan siswa dengan karakteristik yang beragam. Perbedaan latar belakang, kemampuan akademik, kemampuan membaca Al-Qur'an, hafalan, tingkat kehadiran, serta perilaku siswa menyebabkan tingkat minat belajar yang dimiliki setiap siswa juga berbeda. Namun demikian, proses identifikasi minat belajar siswa selama ini masih dilakukan berdasarkan pengamatan langsung guru terhadap aktivitas siswa di kelas. Pendekatan tersebut memiliki keterbatasan karena bersifat subjektif dan belum didukung oleh analisis data yang terstruktur. Akibatnya, pihak madrasah mengalami kesulitan dalam mengidentifikasi kelompok siswa yang memiliki minat belajar tinggi, sedang, maupun rendah secara objektif. Kondisi ini juga berdampak pada proses pengambilan keputusan, termasuk dalam menentukan siswa yang layak dipersiapkan sebagai perwakilan madrasah pada kegiatan Pekan Olahraga dan Seni Antar Diniyah (PORSADIN).

Perkembangan teknologi informasi telah mendorong pemanfaatan pendekatan berbasis data (*data-driven decision making*) dalam berbagai bidang, termasuk pendidikan. Salah satu pendekatan yang berkembang pesat adalah *Educational Data Mining* (EDM), yaitu penerapan teknik data mining untuk menggali pola dan informasi tersembunyi dari data pendidikan guna mendukung pengambilan keputusan yang lebih efektif [1]. Melalui pendekatan ini, data akademik dan non-akademik siswa dapat diolah menjadi informasi yang berguna untuk memahami karakteristik peserta didik, mengevaluasi proses pembelajaran, serta menyusun strategi pendidikan yang lebih tepat sasaran. Penelitian Liu

menunjukkan bahwa teknik data mining mampu membantu proses evaluasi pendidikan melalui analisis pola data yang sebelumnya sulit diidentifikasi secara manual[2].

Salah satu teknik yang banyak digunakan dalam data mining adalah *clustering*, yaitu proses pengelompokan data berdasarkan tingkat kemiripan karakteristik yang dimiliki. Teknik ini bertujuan untuk menempatkan objek yang memiliki karakteristik serupa ke dalam kelompok yang sama sehingga pola tertentu dapat lebih mudah dikenali. Algoritma K-Means merupakan salah satu metode clustering yang paling populer karena memiliki proses komputasi yang sederhana, mudah diterapkan, dan mampu menangani data dalam jumlah besar secara efisien[2]. Dalam konteks pendidikan, K-Means telah banyak dimanfaatkan untuk mengelompokkan siswa berdasarkan performa akademik, karakteristik belajar, maupun tingkat keterlibatan dalam proses pembelajaran[3][4]. Meskipun demikian, algoritma K-Means memiliki kelemahan karena pusat cluster ditentukan berdasarkan nilai rata-rata (*centroid*) sehingga cukup sensitif terhadap keberadaan data pencilan (*outlier*)[5].

Sebagai alternatif, algoritma K-Medoids menggunakan objek data aktual sebagai pusat cluster (*medoid*), sehingga lebih tahan terhadap pengaruh data ekstrem dan perubahan nilai yang signifikan. Karakteristik tersebut menjadikan K-Medoids sering digunakan sebagai metode pembandingan dalam berbagai penelitian clustering. Penelitian Badhera dkk. menunjukkan bahwa K-Means dan K-Medoids sama-sama mampu melakukan segmentasi siswa berdasarkan performa akademik, namun kualitas cluster yang dihasilkan dapat berbeda tergantung karakteristik data yang digunakan[5]. Hasil penelitian tersebut menunjukkan pentingnya melakukan evaluasi terhadap kedua algoritma sebelum menentukan metode yang paling sesuai untuk suatu kasus tertentu.

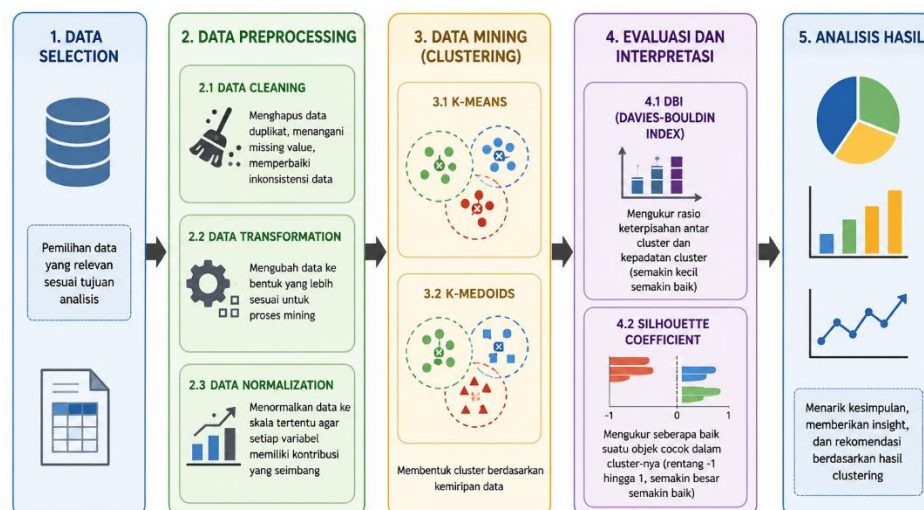
Berbagai penelitian terdahulu telah membuktikan efektivitas penerapan metode clustering dalam bidang pendidikan. Kurniawan dan Hermawan menerapkan algoritma K-Means dan K-Medoids untuk membangun mekanisme pembelajaran yang dipersonalisasi berdasarkan preferensi belajar siswa dan menunjukkan bahwa kedua algoritma mampu mengelompokkan siswa berdasarkan karakteristik yang dimiliki[3]. Qomariyah dan Siregar melakukan studi komparatif antara K-Means dan K-Medoids pada data akademik mahasiswa menggunakan indeks validitas cluster dan menemukan adanya perbedaan performa kedua algoritma dalam menghasilkan kualitas cluster[6]. Penelitian Nagaswetha dkk. memanfaatkan K-Means untuk menganalisis performa siswa dan menghasilkan kelompok-kelompok yang dapat digunakan sebagai dasar evaluasi pembelajaran[4]. Selain itu, Siagian dan Maulana membandingkan kinerja K-Means dan K-Medoids dalam klasifikasi siswa berprestasi dan menemukan bahwa masing-masing algoritma memiliki keunggulan yang berbeda dalam menghasilkan pengelompokan data[7]. Kaligis dan Yulianto juga menunjukkan bahwa perbandingan beberapa algoritma clustering dapat memberikan informasi mengenai metode yang paling sesuai berdasarkan kualitas cluster yang dihasilkan[8].

Meskipun berbagai penelitian tersebut telah berhasil menerapkan metode clustering dalam bidang pendidikan, masih terdapat beberapa celah penelitian (*research gap*) yang perlu dikaji lebih lanjut. Pertama, sebagian besar penelitian terdahulu berfokus pada data prestasi akademik siswa sekolah formal maupun mahasiswa perguruan tinggi. Kedua, penelitian yang membahas pengelompokan minat belajar siswa pada lingkungan Madrasah Diniyah Takmiliyah Awaliyah (MDTA) masih sangat terbatas. Ketiga, indikator yang digunakan dalam penelitian sebelumnya umumnya berfokus pada nilai akademik, sedangkan karakteristik pendidikan keagamaan seperti kemampuan membaca Al-Qur'an, hafalan Al-Qur'an, nilai akhlak, keaktifan pembelajaran, dan tingkat kehadiran belum banyak dimanfaatkan sebagai variabel clustering. Keempat, masih sedikit penelitian yang membandingkan kinerja algoritma K-Means dan K-Medoids untuk mendukung proses seleksi siswa dalam kegiatan keagamaan dan kompetisi tingkat madrasah seperti PORSADIN.

Berdasarkan permasalahan dan gap penelitian tersebut, penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Means dan K-Medoids dalam mengelompokkan minat belajar siswa MDTA berdasarkan variabel nilai akademik keagamaan, tingkat kehadiran, keaktifan siswa, hafalan Al-Qur'an, kemampuan membaca Al-Qur'an, dan nilai akhlak. Evaluasi kualitas cluster dilakukan menggunakan indeks validitas cluster untuk menentukan metode yang menghasilkan pengelompokan terbaik. Hasil penelitian diharapkan dapat memberikan informasi yang lebih objektif mengenai karakteristik minat belajar siswa, membantu guru dalam merancang strategi pembelajaran yang sesuai dengan kondisi siswa, serta mendukung proses seleksi peserta PORSADIN secara lebih efektif, transparan, dan berbasis data.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *data mining* untuk mengelompokkan minat belajar siswa Madrasah Diniyah Takmiliyah Awaliyah (MDTA). Proses penelitian mengadopsi tahapan *Knowledge Discovery in Databases* (KDD) yang terdiri atas seleksi data, *preprocessing*, transformasi data, proses *data mining*, evaluasi, dan interpretasi hasil[9][10]. Model KDD dipilih karena mampu memberikan alur kerja yang sistematis dalam menggali pola dan pengetahuan dari kumpulan data yang tersedia. Tahapan penelitian diawali dengan pengumpulan data siswa MDTA Qur'an Kisaran, kemudian dilanjutkan dengan proses pembersihan dan transformasi data. Selanjutnya dilakukan proses klasterisasi menggunakan algoritma K-Means dan K-Medoids. Hasil klasterisasi yang diperoleh dievaluasi menggunakan metode validasi cluster untuk menentukan algoritma yang menghasilkan kualitas cluster terbaik. Tahapan penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian Data Mining

2.1. Data Selection

Data yang digunakan dalam penelitian ini merupakan data siswa MDTA Qur'an Kisaran Tahun Ajaran 2025/2026. Objek penelitian difokuskan pada siswa kelas II dan III karena pada jenjang tersebut siswa telah memiliki kemampuan dasar yang relatif stabil serta berpotensi mengikuti kegiatan Pekan Olahraga dan Seni Antar Diniyah (PORSADIN). Variabel yang digunakan dalam proses clustering terdiri atas enam atribut utama, yaitu nilai akademik keagamaan, tingkat kehadiran (absensi), keaktifan dalam pembelajaran, hafalan Al-Qur'an, kemampuan membaca Al-Qur'an, dan nilai akhlak. Data diperoleh melalui dokumentasi madrasah, hasil penilaian guru, dan kuesioner yang diberikan kepada siswa.

2.2. Data Preprocessing

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data sebelum proses clustering dilaksanakan. Menurut Han et al.[10], kualitas data sangat memengaruhi hasil analisis sehingga diperlukan proses persiapan data yang memadai sebelum diterapkan pada algoritma data mining[11].

2.3. Data Cleaning

Tahap *data cleaning* bertujuan untuk mengidentifikasi dan memperbaiki data yang tidak lengkap, duplikat, maupun inkonsisten sehingga dapat mengurangi potensi kesalahan pada proses analisis[12].

2.4. Data Transformation

Data yang masih berbentuk kategori atau deskripsi kualitatif diubah ke dalam bentuk numerik agar dapat diproses oleh algoritma clustering. Transformasi data diperlukan untuk memastikan seluruh atribut memiliki format yang seragam dan dapat dihitung secara matematis[10].

2.5 Data Normalization

Atribut memiliki rentang nilai yang berbeda, dilakukan normalisasi menggunakan metode Min-Max Normalization. Normalisasi bertujuan untuk menghindari dominasi atribut tertentu dalam proses perhitungan jarak sehingga setiap variabel memiliki kontribusi yang seimbang terhadap pembentukan cluster [12][13].

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Keterangan:

x' = nilai hasil normalisasi

x = nilai data awal

$x_{\min}$ = nilai minimum dari atribut

$x_{\max}$ = nilai maksimum dari atribut

2.6 Algoritma K Means

K-Means merupakan algoritma clustering non-hierarki yang mengelompokkan data ke dalam sejumlah cluster berdasarkan kedekatan terhadap pusat cluster (*centroid*) [14],[15]. Algoritma ini termasuk salah satu metode clustering yang paling banyak digunakan karena sederhana, mudah diimplementasikan, dan memiliki efisiensi komputasi yang baik pada data berukuran besar[10].

Langkah-langkah algoritma K-Means adalah sebagai berikut:

1. Menentukan jumlah cluster ((k)).
2. Memilih centroid awal secara acak.
3. Menghitung jarak setiap data terhadap centroid.
4. Menempatkan data ke cluster dengan jarak terdekat.
5. Menghitung centroid baru berdasarkan rata-rata anggota cluster.
6. Mengulangi proses hingga tidak terjadi perubahan anggota cluster.

Perhitungan jarak dilakukan menggunakan Euclidean Distance yang merupakan ukuran jarak paling umum dalam algoritma clustering[12].

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Keterangan:

$d(x,y)$ = jarak Euclidean antara objek x dan objek y ;

x_i = nilai atribut ke- i pada objek x ;

y_i = nilai atribut ke- i pada objek y ;

n = jumlah atribut yang digunakan dalam proses clustering.

2.7 Algoritma K-Medoids

K-Medoids merupakan metode clustering yang menggunakan objek data aktual sebagai pusat cluster (*medoid*). Berbeda dengan K-Means yang menggunakan nilai rata-rata sebagai pusat cluster, K-Medoids memilih salah satu objek dalam cluster sebagai representasi kelompok sehingga lebih tahan terhadap pengaruh *outlier*[16].

Tahapan algoritma K-Medoids meliputi:

1. Menentukan jumlah cluster.
2. Memilih medoid awal.
3. Menghitung jarak setiap objek terhadap medoid.
4. Menghitung total biaya (*cost*) cluster.
5. Menukar medoid dengan objek lain apabila menghasilkan nilai biaya yang lebih kecil.
6. Mengulangi proses hingga tidak terdapat perubahan medoid.

Menurut Kaufman dan Rousseeuw[16], penggunaan medoid sebagai pusat cluster mampu menghasilkan pengelompokan yang lebih stabil pada data yang mengandung pencilan.

Evaluasi dan Interpretasi

Evaluasi dilakukan untuk menentukan algoritma yang menghasilkan kualitas cluster terbaik. Penelitian ini menggunakan dua metode validasi internal yaitu Davies-Bouldin Index (DBI) dan Silhouette Coefficient.

Davies-Bouldin Index (DBI)

Davies-Bouldin Index merupakan ukuran validitas cluster yang mengukur rasio antara variasi dalam cluster (*intra-cluster*) dan jarak antar cluster (*inter-cluster*)[17]. Semakin kecil nilai DBI yang diperoleh, maka kualitas cluster yang dihasilkan semakin baik karena menunjukkan tingkat homogenitas yang tinggi dan pemisahan yang jelas antar cluster.

Silhouette Coefficient

Silhouette Coefficient digunakan untuk mengukur tingkat kesesuaian suatu objek terhadap cluster tempat objek tersebut berada dibandingkan dengan cluster lainnya[18]. Nilai Silhouette berada pada rentang -1 hingga 1. Semakin mendekati nilai 1 menunjukkan bahwa objek berada pada cluster yang tepat dan memiliki tingkat pemisahan yang baik dengan cluster lain.

Penggunaan DBI dan Silhouette Coefficient secara bersamaan memberikan evaluasi yang lebih komprehensif terhadap kualitas cluster yang dihasilkan oleh algoritma K-Means maupun K-Medoids [12],[18].

2.8 Analisis Hasil

Tahap akhir penelitian dilakukan dengan membandingkan hasil clustering yang diperoleh dari algoritma K-Means dan K-Medoids berdasarkan nilai Davies-Bouldin Index dan Silhouette Coefficient. Algoritma yang menghasilkan nilai DBI paling rendah dan nilai Silhouette paling tinggi ditetapkan sebagai metode terbaik[13][19]. Selanjutnya, setiap cluster dianalisis berdasarkan karakteristik nilai rata-rata atribut untuk mengidentifikasi kelompok siswa dengan kategori minat belajar tinggi, sedang, dan rendah. Hasil analisis digunakan sebagai dasar penyusunan rekomendasi strategi pembelajaran dan pendukung proses seleksi peserta PORSADIN secara lebih objektif dan berbasis data.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Dataset

Pada penelitian ini digunakan data siswa MDTA Qur'an Kisaran Tahun Ajaran 2025/2026 sebanyak 119 siswa. Data yang digunakan terdiri dari enam atribut yang merepresentasikan indikator minat belajar siswa, yaitu nilai akademik, persentase kehadiran, penilaian hafalan, kemampuan bacaan Al-Qur'an, penilaian akhlak, dan partisipasi siswa dalam kegiatan pembelajaran yang tersedia pada Tabel 1.

Tabel 1. Dataset Data Siswa dan Nilai siswa MDTA Qur'an

No	Nama Siswa	Nilai Akademik	Presentasi Kehadiran	Penilaian Hafalan	Bacaan Qur'an	Penilaian Akhlak	Partisipasi Siswa
1	Azzam Abdillah Pratama	90	78	95	77	89	75
2	Fakhri Alwi Yusnah	76	74	65	42	39	55
3	Hizra Maliqa Az Zahra	6	60	78	64	56	62
4	Jaffi Khalis Batubara	71	77	71	61	54	43
...							
118	Salsabila Putri	71	61	70	74	60	63
119	Daffa Aqilla	94	84	94	68	67	89

3.2 Variabel Operasional

Variabel nilai akademik, persentase kehadiran, penilaian hafalan, kemampuan bacaan Al-Qur'an, penilaian akhlak, dan partisipasi siswa dipilih karena mampu merepresentasikan dimensi keterlibatan belajar siswa yang mencakup aspek akademik, perilaku, dan partisipasi aktif selama proses pembelajaran. Kehadiran dan partisipasi merupakan indikator utama *behavioral engagement*, sedangkan capaian akademik dan kemampuan keagamaan mencerminkan hasil keterlibatan siswa dalam kegiatan belajar [20]. Definisi operasional masing-masing variabel yang digunakan dalam penelitian ini disajikan pada Tabel 2.

Tabel 2. Variabel Operasional

Variabel	Deskripsi	Skala Likert
Nilai Akademik	Tingkat pencapaian hasil belajar siswa pada mata pelajaran keagamaan	1 = Sangat Rendah, 2 = Rendah, 3 = Sedang, 4 = Tinggi, 5 = Sangat Tinggi
Presentase Kehadiran	Tingkat kehadiran siswa dalam mengikuti kegiatan pembelajaran	1 = Sangat Rendah, 2 = Rendah, 3 = Sedang, 4 = Tinggi, 5 = Sangat Tinggi
Penilaian Hafalan	Kemampuan siswa dalam menghafal Al-Qur'an dan materi hafalan	1 = Sangat Kurang, 2 = Kurang, 3 = Cukup, 4 = Baik, 5 = Sangat Baik
Bacaan Al-Qur'an	Kemampuan siswa membaca Al-Qur'an sesuai kaidah tajwid	1 = Sangat Kurang, 2 = Kurang, 3 = Cukup, 4 = Baik, 5 = Sangat Baik
Penilaian Akhlak	Perilaku dan sikap siswa selama mengikuti kegiatan pembelajaran	1 = Sangat Kurang, 2 = Kurang, 3 = Cukup, 4 = Baik, 5 = Sangat Baik
Partisipasi Siswa	Tingkat keaktifan siswa dalam proses pembelajaran	1 = Sangat Tidak Aktif, 2 = Tidak Aktif, 3 = Cukup Aktif, 4 = Aktif, 5 = Sangat Aktif

3.3 Transformasi Data

Tahap transformasi data dilakukan sebagai bagian dari proses *data preprocessing* untuk mengubah data mentah menjadi format yang lebih sesuai bagi analisis clustering. Seluruh atribut penelitian dikonversi berdasarkan kriteria variabel operasional sehingga menghasilkan skala penilaian yang seragam. Transformasi ini tidak hanya mempermudah

interpretasi karakteristik siswa, tetapi juga membantu meningkatkan kualitas proses pengelompokan karena setiap atribut dapat direpresentasikan dalam bentuk yang setara. Hasil transformasi data yang akan digunakan pada tahap clustering ditunjukkan pada Tabel 3.

Tabel 3. Hasil Transformasi Data Berdasarkan Variabel Operasional

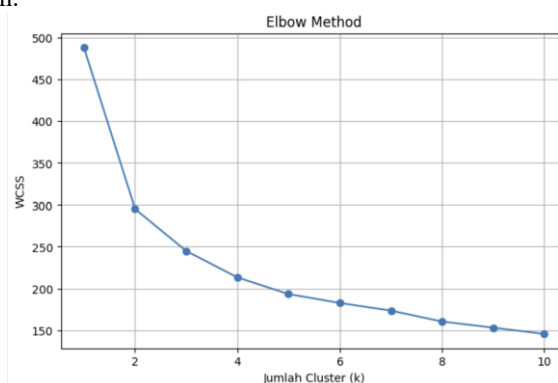
No	Nama Siswa	Nilai Akademik	Presentasi Kehadiran	Penilaian Hafalan	Bacaan Qur'an	Penilaian Ahlak	Partisipasi Siswa
1	Azzam Abdillah Pratama	5	4	5	4	5	4
2	Fakhri Alwi Yusnah	4	4	4	2	2	3
3	Hizra Maliqa Az Zahra	3	3	4	3	3	3
4	Jaffi Khalis Batubara	4	4	4	3	3	2
...							
118	Salsabila Putri	4	3	4	4	3	3
119	Daffa Aqilla	5	5	5	4	4	5

Transformasi data dilakukan berdasarkan variabel operasional penelitian, dimana proses ini menghasilkan data dengan rentang nilai 1–5 pada seluruh atribut sehingga setiap variabel memiliki kontribusi yang seimbang dalam pembentukan cluster. Oleh karena itu, tahap normalisasi tidak dilakukan karena keseragaman skala telah dicapai melalui proses transformasi data.

3.4 K-Means

Sebelum proses clustering dilakukan, terlebih dahulu ditentukan jumlah cluster optimal menggunakan metode *Elbow*. Metode ini digunakan dengan menghitung nilai *Within Cluster Sum of Squares* (WCSS) pada beberapa jumlah cluster yang berbeda. Nilai WCSS menunjukkan tingkat variasi data di dalam cluster, dimana semakin kecil nilai WCSS menunjukkan semakin baik tingkat homogenitas cluster yang terbentuk.

Hasil pengujian *Elbow Method* ditunjukkan pada Gambar 2. Terlihat bahwa penurunan nilai WCSS terjadi secara signifikan pada rentang $k = 1$ hingga $k = 3$. Setelah $k = 3$, penurunan nilai WCSS cenderung melandai sehingga penambahan jumlah cluster tidak memberikan peningkatan kualitas cluster yang signifikan. Oleh karena itu, jumlah cluster optimal yang digunakan dalam penelitian ini adalah sebanyak tiga cluster yang merepresentasikan kategori minat belajar tinggi, sedang, dan rendah.



Gambar 2. Pengujian Elbow Method

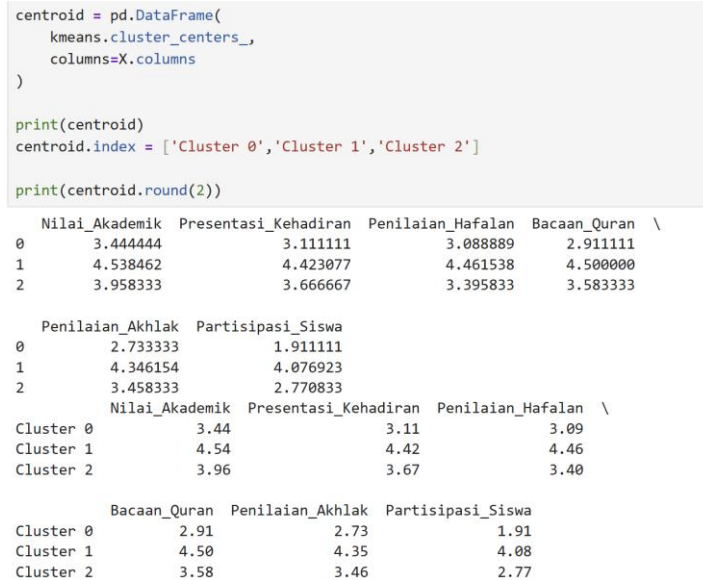
Hasil proses klasterisasi menggunakan algoritma K-Means menghasilkan tiga kelompok siswa dengan jumlah anggota yang berbeda pada setiap cluster. Distribusi anggota cluster menunjukkan bahwa sebagian besar siswa berada pada kelompok minat belajar sedang dan tinggi, sedangkan sebagian lainnya berada pada kelompok minat belajar rendah.

Tabel 4. Hasil Distribusi Anggota Cluster

No	Cluster	Jumlah Siswa
1	C0	43
2	C1	39
3	C2	35

Untuk memahami karakteristik masing-masing cluster yang terbentuk, dilakukan analisis terhadap nilai centroid hasil klasterisasi K-Means. Nilai centroid merepresentasikan rata-rata nilai setiap variabel pada masing-masing cluster sehingga dapat digunakan untuk menggambarkan profil atau karakteristik kelompok siswa yang terbentuk. Dalam

penelitian ini, analisis centroid dilakukan berdasarkan enam variabel, yaitu nilai akademik, persentase kehadiran, penilaian hafalan, kemampuan bacaan Al-Qur'an, penilaian akhlak, dan partisipasi siswa dalam pembelajaran. Semakin tinggi nilai centroid pada suatu cluster menunjukkan semakin baik tingkat keterlibatan dan minat belajar siswa pada kelompok tersebut. Hasil perhitungan centroid untuk masing-masing cluster disajikan pada Gambar 3 dan Tabel 5.



Gambar 3. Nilai Centroid Hasil Klasterisasi K-Means

Tabel 5. Nilai Centroid Hasil Klasterisasi Menggunakan Algoritma K-Means

Cluster	Nilai Akademik	Kehadiran	Hafalan	Bacaan Quran	Akhlak	Partisipasi
Cluster 0	3.44	3.11	3.09	2.91	2.73	1.91
Cluster 1	4.54	4.42	4.46	4.50	4.35	4.08
Cluster 2	3.96	3.67	3.40	3.58	3.46	2.77

Berdasarkan Tabel 5, Cluster 1 memiliki nilai centroid tertinggi pada seluruh variabel sehingga dikategorikan sebagai kelompok siswa dengan minat belajar tinggi. Cluster 2 memiliki nilai centroid pada tingkat menengah dan dikategorikan sebagai kelompok siswa dengan minat belajar sedang. Sementara itu, Cluster 0 memiliki nilai centroid terendah pada sebagian besar variabel, terutama pada aspek partisipasi siswa, sehingga dikategorikan sebagai kelompok siswa dengan minat belajar rendah. Hasil ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan siswa berdasarkan karakteristik minat belajar secara objektif menggunakan enam indikator penelitian.

Namun, interpretasi centroid saja belum cukup untuk menunjukkan kualitas pengelompokan yang dihasilkan. Oleh karena itu, diperlukan proses evaluasi cluster untuk mengukur tingkat homogenitas dalam cluster serta tingkat pemisahan antar cluster. Pada penelitian ini, evaluasi dilakukan menggunakan Davies-Bouldin Index (DBI) dan Silhouette Coefficient.

Tabel 6. Nilai Centroid Hasil Klasterisasi Menggunakan Algoritma K-Means

Metode	DBI	Silhouette
K-Means	1.5839	0.2186

Untuk mengukur kualitas cluster yang terbentuk, dilakukan evaluasi menggunakan Davies-Bouldin Index (DBI) dan Silhouette Coefficient. Nilai DBI digunakan untuk mengukur tingkat homogenitas cluster dan pemisahan antar cluster, dimana nilai yang lebih kecil menunjukkan kualitas cluster yang lebih baik. Sementara itu, Silhouette Coefficient digunakan untuk mengukur tingkat kesesuaian objek terhadap cluster tempatnya berada, dimana nilai yang semakin mendekati 1 menunjukkan kualitas cluster yang semakin baik. Berdasarkan Tabel 6, algoritma K-Means menghasilkan nilai DBI sebesar 1.5839 dan nilai Silhouette sebesar 0.2186 yang menunjukkan bahwa proses klasterisasi mampu membentuk kelompok siswa dengan karakteristik yang relatif homogen. Selanjutnya, untuk mengetahui metode yang menghasilkan kualitas pengelompokan yang lebih baik, dilakukan proses klasterisasi menggunakan algoritma K-Medoids dengan jumlah cluster yang sama, yaitu tiga cluster.

3.5 K-Medoids

Proses klasterisasi menggunakan algoritma K-Medoids dengan jumlah cluster sebanyak tiga kelompok selesai dilakukan, tahap berikutnya adalah mengamati distribusi anggota pada masing-masing cluster. Distribusi cluster memberikan gambaran mengenai sebaran siswa pada setiap kelompok yang terbentuk, sehingga dapat diketahui proporsi

anggota pada masing-masing cluster sebelum dilakukan analisis lebih lanjut terhadap karakteristik cluster berdasarkan nilai medoid yang dihasilkan. Hasil distribusi anggota cluster ditunjukkan pada Tabel 7.

Tabel 7. Hasil Distribusi Anggota Cluster

No	Cluster	Jumlah Siswa
1	C0	26
2	C1	43
3	C2	50

Berdasarkan Tabel 7, algoritma K-Medoids berhasil mengelompokkan 119 siswa ke dalam tiga cluster, yaitu Cluster 0 sebanyak 26 siswa, Cluster 1 sebanyak 43 siswa, dan Cluster 2 sebanyak 50 siswa. Hasil ini menunjukkan bahwa sebagian besar siswa tergabung dalam Cluster 2, sedangkan Cluster 0 memiliki jumlah anggota paling sedikit. Perbedaan sebaran anggota cluster tersebut mencerminkan adanya variasi karakteristik minat belajar siswa yang terbentuk dari proses klusterisasi.

Meskipun distribusi anggota cluster telah menunjukkan sebaran siswa pada setiap kelompok, karakteristik masing-masing cluster perlu dianalisis lebih lanjut melalui nilai medoid sebagai pusat representatif cluster. Nilai medoid digunakan untuk menggambarkan karakteristik yang paling mewakili anggota pada setiap kelompok sehingga pola minat belajar siswa dapat diidentifikasi dengan lebih jelas. Hasil nilai medoid untuk masing-masing cluster disajikan pada Gambar 4 dan Tabel 8.

```
# Nilai Medoid Hasil Klusterisasi K-Medoids
medoid_data = df.iloc[medoid_index]

print(medoid_data)
```

	Nilai_Akademik	Presentasi_Kehadiran	Penilaian_Hafalan	Bacaan_Quran	\
82	5	4	4	4	4
66	4	4	4	4	4
37	4	3	3	3	3

	Penilaian_Akhlak	Partisipasi_Siswa	Cluster_KMeans	Cluster_KMedoids
82	4	4	1	0
66	3	3	2	1
37	3	2	0	2

Gambar 4. Nilai Centroid Hasil Klusterisasi K-Medoids
Tabel 8. Nilai Centroid Hasil Klusterisasi Menggunakan Algoritma K-Medoids

Cluster	Nilai Akademik	Kehadiran	Hafalan	Bacaan Quran	Akhlak	Partisipasi
Cluster 0	5	4	4	4	4	4
Cluster 1	4	4	4	4	3	3
Cluster 2	4	3	3	3	3	2

Nilai medoid menunjukkan karakteristik yang mewakili setiap cluster hasil klusterisasi menggunakan algoritma K-Medoids. Cluster 0 memiliki nilai tertinggi pada hampir seluruh variabel, meliputi nilai akademik, kehadiran, hafalan, bacaan Al-Qur'an, akhlak, dan partisipasi siswa. Karakteristik tersebut menggambarkan siswa yang aktif mengikuti pembelajaran, memiliki capaian akademik yang baik, serta menunjukkan keterlibatan yang tinggi dalam kegiatan belajar. Oleh karena itu, Cluster 0 dapat dikategorikan sebagai kelompok siswa dengan minat belajar tinggi.

Cluster 1 memperlihatkan capaian yang cukup baik dan relatif seimbang pada sebagian besar variabel. Meskipun tidak setinggi Cluster 0, kelompok ini menunjukkan tingkat keterlibatan belajar yang cukup konsisten, baik dari aspek akademik maupun perilaku selama proses pembelajaran. Dengan karakteristik tersebut, Cluster 1 dikategorikan sebagai kelompok siswa dengan minat belajar sedang. Di sisi lain, Cluster 2 memiliki nilai yang cenderung lebih rendah dibandingkan dua cluster lainnya, terutama pada aspek partisipasi siswa, hafalan, dan kemampuan bacaan Al-Qur'an. Kondisi ini mengindikasikan bahwa keterlibatan siswa dalam proses pembelajaran masih perlu ditingkatkan. Oleh karena itu, Cluster 2 dikategorikan sebagai kelompok siswa dengan minat belajar rendah.

Setelah karakteristik masing-masing cluster berhasil diidentifikasi melalui nilai medoid, tahap berikutnya adalah melakukan evaluasi kualitas cluster. Evaluasi dilakukan menggunakan Davies-Bouldin Index (DBI) dan Silhouette Coefficient untuk mengetahui tingkat homogenitas dan pemisahan cluster yang dihasilkan oleh algoritma K-Medoids. Hasil evaluasi cluster disajikan pada Tabel 9.

Tabel 9. Nilai Centroid Hasil Klusterisasi Menggunakan Algoritma K-Medoids

Metode	DBI	Silhouette
K-Medoids	1.5950	0.2046

Nilai Davies-Bouldin Index (DBI) dan Silhouette Coefficient digunakan untuk mengevaluasi kualitas cluster yang

dihasilkan oleh algoritma K-Medoids. Nilai DBI yang semakin kecil menunjukkan bahwa cluster yang terbentuk memiliki tingkat homogenitas yang lebih baik dan pemisahan antar cluster yang lebih jelas. Sebaliknya, nilai Silhouette yang semakin mendekati 1 menunjukkan bahwa anggota cluster memiliki tingkat kesesuaian yang lebih tinggi terhadap kelompoknya masing-masing. Hasil evaluasi menunjukkan bahwa algoritma K-Medoids menghasilkan nilai DBI sebesar 1,5950 dan nilai Silhouette sebesar 0,2046. Nilai tersebut menunjukkan bahwa proses klasterisasi mampu membentuk kelompok siswa berdasarkan karakteristik minat belajar yang dimiliki, meskipun tingkat pemisahan antar cluster masih berada pada kategori sedang. Hasil ini menunjukkan bahwa algoritma K-Medoids dapat digunakan untuk mengidentifikasi pola minat belajar siswa, namun kualitas cluster yang dihasilkan perlu dibandingkan dengan algoritma lain untuk menentukan metode yang paling sesuai.

3.6 Evaluasi dan Perbandingan Kinerja Algoritma K-Means dan K-Medoids

Untuk menentukan algoritma yang menghasilkan kualitas cluster terbaik, dilakukan perbandingan antara algoritma K-Means dan K-Medoids berdasarkan nilai Davies-Bouldin Index (DBI) dan Silhouette Coefficient yang diperoleh dari masing-masing metode. Hasil perbandingan kedua algoritma disajikan pada Tabel 10.

Tabel 10. Perbandingan Kinerja Algoritma K-Means dan K-Medoids

Metode	DBI	Silhouette
K-Means	1.5839	0.2186
K-Medoids	1.5950	0.2046

Hasil perbandingan menunjukkan bahwa algoritma K-Means menghasilkan nilai Davies-Bouldin Index (DBI) yang lebih rendah, yaitu 1,5839, dibandingkan K-Medoids sebesar 1,5950. Selain itu, nilai Silhouette Coefficient yang diperoleh K-Means sebesar 0,2186 juga lebih tinggi dibandingkan K-Medoids yang sebesar 0,2046. Meskipun selisih nilai yang diperoleh relatif kecil, hasil tersebut menunjukkan bahwa K-Means mampu membentuk cluster dengan tingkat homogenitas yang sedikit lebih baik serta pemisahan antar cluster yang lebih jelas dibandingkan K-Medoids.

Pengelompokan yang dihasilkan berhasil membagi siswa ke dalam tiga kategori minat belajar, yaitu tinggi, sedang, dan rendah, berdasarkan enam indikator yang digunakan dalam penelitian, meliputi nilai akademik, kehadiran, hafalan, bacaan Al-Qur'an, akhlak, dan partisipasi siswa. Hasil ini menunjukkan bahwa pendekatan clustering dapat dimanfaatkan untuk membantu madrasah dalam memahami karakteristik siswa secara lebih objektif dan berbasis data.

Secara keseluruhan, algoritma K-Means dapat direkomendasikan sebagai metode yang lebih sesuai untuk mengelompokkan minat belajar siswa pada data MDTA Qur'an Kisaran Tahun Ajaran 2025/2026. Informasi yang diperoleh dari hasil klasterisasi diharapkan dapat menjadi bahan pertimbangan dalam penyusunan strategi pembelajaran, pembinaan siswa, serta mendukung proses seleksi peserta Pekan Olahraga dan Seni Antar Diniyah (PORSADIN) secara lebih terarah dan objektif.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means dan K-Medoids untuk mengelompokkan minat belajar siswa MDTA Qur'an Kisaran Tahun Ajaran 2025/2026 berdasarkan enam indikator, yaitu nilai akademik, kehadiran, hafalan Al-Qur'an, bacaan Al-Qur'an, akhlak, dan partisipasi siswa. Hasil klasterisasi menunjukkan bahwa siswa dapat dikelompokkan ke dalam tiga kategori minat belajar, yaitu tinggi, sedang, dan rendah, sehingga mampu memberikan gambaran karakteristik siswa secara lebih objektif dan berbasis data. Berdasarkan hasil evaluasi menggunakan Davies-Bouldin Index (DBI) dan Silhouette Coefficient, algoritma K-Means menghasilkan nilai DBI sebesar 1,5839 dan Silhouette sebesar 0,2186, sedangkan algoritma K-Medoids menghasilkan nilai DBI sebesar 1,5950 dan Silhouette sebesar 0,2046. Hasil tersebut menunjukkan bahwa algoritma K-Means memiliki performa yang lebih baik dibandingkan K-Medoids dalam mengelompokkan data minat belajar siswa pada penelitian ini.

Dengan demikian, algoritma K-Means dapat direkomendasikan sebagai metode yang lebih sesuai untuk pengelompokan minat belajar siswa MDTA Qur'an Kisaran. Hasil penelitian ini diharapkan dapat menjadi pendukung dalam penyusunan strategi pembelajaran, pembinaan siswa, serta proses seleksi peserta PORSADIN secara lebih terarah, objektif, dan berbasis data.

REFERENCES

- [1] A. Muis *et al.*, *DATA MINING : Konsep , Metode , dan Aplikasi*, 1st ed. Medan: FAASLIB SERAMBI MEDIA, 2025.
- [2] R. Liu, "Data Analysis of Educational Evaluation Using K-Means," *Comput. Intell. Neurosci.*, vol. 2022, p. 10, 2022, doi: <https://doi.org/10.1155/2022/3762431>.
- [3] R. Wenda and D. Kurniawan, "Personalized Student Learning Mechanisms using K- Means and K-Medoids Clustering Algorithms based on Individual Preferences," *Int. J. Comput. Appl.*, vol. 185, no. 46, pp. 13–19, 2023.
- [4] I. N. M. Alawi, S. J. S., Shaharane and J. M. Jamil, "CLUSTERING STUDENT PERFORMANCE DATA USING k-MEANS ALGORITHMS," *J. Comput. Innov. Anal.*, vol. 2, no. 1, pp. 41–55, 2023.
- [5] U. Badhera, A. Verma, and P. Nahar, "Applicability of K-medoids and K-means algorithms for segmenting students based on their scholastic performance," *J. Stat. Manag. Syst. ISSN*, vol. 22, no. 7, p. 11, 2022, doi: 10.1080/09720510.2022.2130566.
- [5] Q. Qomariyah and M. U. Siregar, "Comparative Study of K-Means Clustering Algorithm and K-Medoids Clustering in Student Data Clustering," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 2, pp. 91–99, 2022.

- [6] F. N. Siagian and H. Maulana, “Perbandingan Kinerja K-Means dan K-Medoids dalam Klasifikasi Siswa Berprestasi di SMP Muhammadiyah 60 Medan,” *Inven. J. Sci. , Technol. , Innov.*, vol. 1, no. 3, 2026.
- [7] S. Kaligis, Gideon Bartolomeus Yulianto, “ANALISA PERBANDINGAN ALGORITMA K-MEANS, K- MEDOIDS, DAN X-MEANS UNTUK PENGELOMPOKKAN KINERJA PEGAWAI,” *J. Penerapan Teknol. Inf. dan Komun.*, vol. 01, no. 3, pp. 179–193, 2022.
- [8] C. Schröer, F. Kruse, J. Marx, F. Kruse, and J. Marx, “ScienceDirect ScienceDirect A Systematic Literature Review A Systematic Literature Review on Applying Process Model on Applying CRISP-DM Process Model,” *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [9] J. Han, J. Pei, and H. Tong, *DATA MINING CONCEPTS AND TECHNIQUES*, FOURTH. Cambrige: Katey Birtcher, 2023.
- [10] W. Ramdhan, N. Nurwati, and E. Rahayu, “ANALISIS DATA EKSPLORATORI DAN CLUSTERING K-MODES UNTUK PEMETAAN STATUS GIZI BALITA PADA KASUS STUNTING DI KABUPATEN ASAHAN,” *J. Sci. Soc. Res.*, vol. 4307, no. 4, pp. 3635–3644, 2025, [Online]. Available: <https://jurnal.goretanpena.com/index.php/JSSR/article/view/4192/2204>
- [11] F. Y. Arini, L. A. Djuanda, A. Hisma, and P. Kristianto, “Performance Evaluation of Gradient Boosting Techniques for Predicting Customer Purchase Decisions,” vol. 7, no. 2, pp. 1441–1454, 2026.
- [12] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, “A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data,” *Front. Energy Res.*, vol. 9, no. March, pp. 1–17, 2021, doi: 10.3389/fenrg.2021.652801.
- [13] T. ALASALI and Y. ORTAKCI, “Clustering Techniques in Data Mining: A Survey of Methods, Challenges, and Applications,” *J. Comput. Sci.*, vol. 9, no. 1, pp. 32–50, 2024.
- [14] O. Chorna, P. Didyk, S. Titov, and O. Titova, “USAGE OF CLUSTERING ALGORITHMS FOR AUTOMATING ROUTE PLANNING Summary of the main material,” *Системы обработки информации*, vol. 1, no. 176, pp. 115–123, 2024, doi: 10.30748/soi.2024.176.14.
- [15] E. Schubert and P. J. Rousseeuw, “Fast and eager k-medoids clustering : O (k) runtime improvement of the PAM, CLARA, and CLARANS algorithms,” *Inf. Syst.*, vol. 101, p. 101804, 2021, doi: 10.1016/j.is.2021.101804.
- [16] S. Pitafi, T. Anwar, and Z. Sharif, “A Taxonomy of Machine Learning Clustering Algorithms , Challenges , and Future Realms,” 2023, doi: <https://doi.org/10.3390/app13063529> Academic.
- [17] M. Shutaywi and N. N. Kachouie, “Silhouette Analysis for Performance Evaluation in Machine with Applications to Clustering,” *Entropy*, pp. 1–17, 2021, doi: <https://doi.org/10.3390/e23060759> 1.
- [18] A. A. A. Fernandes, M. Koehler, N. Konstantinou, P. Pankin, and N. W. Paton, “Data Preparation : A Technological Perspective and Review,” *SN Comput. Sci.*, vol. 4, no. 4, pp. 1–20, 2025, doi: 10.1007/s42979-023-01828-8.
- [19] S. E. Kassab, M. Al Eraky, W. El Sayed, H. Hamdy, and H. Schmidt, “Measurement of student engagement in health professions education : a review of literature,” *BMC Med. Educ.*, pp. 1–14, 2023, doi: 10.1186/s12909-023-04344-8.