

Analisis Data Mining dengan Algoritma K-Means Clustering untuk Menentukan Siswa Berprestasi di MTS miftahul ulum bengkak

Zaehol Fatah¹, Siti Maghfiroh²

¹Sistem Informasi, Sains dan Teknologi, Universitas Ibrahimy, Indonesia

²Teknologi Informasi, Sains dan Teknologi, Universitas Ibrahimy, Indonesia

Email: zaeholfatah@gmail.com¹, 03siti24@gmail.com*²

Abstrak

Tujuan dari penelitian ini adalah melakukan analisis terhadap data nilai siswa di MTs Miftahul Ulum Bengkak dengan menerapkan algoritma K-Means Clustering, sehingga dapat diperoleh pengelompokan siswa berdasarkan tingkat prestasinya. secara objektif dan efisien. Latar belakang penelitian ini adalah masih digunakannya metode manual dalam penentuan siswa berprestasi yang cenderung subjektif dan memakan waktu lama. Penelitian ini menggunakan lima jenis data nilai siswa sebagai bahan analisis. Penelitian ini berfokus pada lima mata pelajaran inti, yaitu Bahasa Indonesia, Bahasa Inggris, Matematika, IPA, dan IPS, yang digunakan sebagai dasar dalam proses analisis. Proses analisis dilakukan menggunakan perangkat lunak RapidMiner Studio dengan tahapan menentukan jumlah cluster, perhitungan jarak menggunakan *Euclidean Distance*, dan pembentukan cluster berdasarkan kemiripan nilai rata-rata siswa. Berdasarkan hasil analisis, penerapan algoritma K-Means mampu mengelompokkan data siswa ke dalam tiga kelompok utama yang mewakili perbedaan tingkat prestasi., yaitu Cluster 0 yang berisi 7 siswa berprestasi tinggi, Cluster 1 berisi 3 siswa berprestasi sedang, dan Cluster 2 berisi 1 siswa berprestasi rendah. Berdasarkan hasil analisis, metode K-Means Clustering terbukti efektif dalam mengelompokkan data secara akurat sesuai dengan karakteristik yang dimiliki masing-masing siswa. dan dapat membantu pihak sekolah dalam mengidentifikasi siswa berprestasi berdasarkan data nilai akademik.

Kata Kunci: Data, K-Means Clustering, Siswa Berprestasi, MTs Miftahul Ulum Bengkak.

Abstract

The purpose of this study is to analyze student grade data at MTs Miftahul Ulum Bengkak by applying the K-Means Clustering algorithm, so that student grouping can be obtained based on their achievement level. objectively and efficiently. The background of this study is that manual methods are still used in determining high-achieving students which tend to be subjective and time-consuming. This study uses five types of student grade data as analysis material. This study focuses on five core subjects, namely Indonesian, English, Mathematics, Science, and Social Studies, which are used as the basis for the analysis process. The analysis process is carried out using RapidMiner Studio software with the stages of determining the number of clusters, calculating distances using *Euclidean Distance*, and forming clusters based on the similarity of students' average scores. Based on the results of the analysis, the application of the K-Means algorithm is able to group student data into three main groups that represent differences in achievement levels, namely Cluster 0 which contains 7 high-achieving students, Cluster 1 contains 3 medium-achieving students, and Cluster 2 contains 1 low-achieving student. Based on the analysis results, the K-Means Clustering method proved effective in accurately grouping data according to each student's characteristics and can help schools identify high-achieving students based on academic grades.

Keywords: *Data Mining, K-Means Clustering, High-Achieving Students, MTs Miftahul Ulum Bengkak.*

PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi (TIK) pada masa kini membawa pengaruh yang signifikan terhadap berbagai sektor, terutama pendidikan. TIK memungkinkan proses pembelajaran, pengolahan data, serta penyampaian informasi dilakukan dengan cepat, akurat, dan efisien (Asmawi et al., 2019). Dalam konteks pendidikan, pemanfaatan teknologi informasi tidak hanya mendukung proses belajar-mengajar, tetapi juga membantu pihak sekolah dalam melakukan evaluasi dan pengambilan keputusan berbasis data.

Pendidikan berbasis teknologi informasi membuka peluang besar dalam menciptakan sistem pendidikan yang lebih efektif dan objektif. Salah satu bentuk penerapan teknologi dalam bidang pendidikan adalah penggunaan data mining, yaitu teknik untuk menggali informasi dan pola tersembunyi dari kumpulan data besar. Melalui pendekatan ini, data siswa dapat diolah menjadi informasi yang berguna untuk mendukung proses penilaian, evaluasi, dan peningkatan mutu pendidikan (Kharis et al., 2022).

Sejalan dengan hal tersebut, penelitian di berbagai negara juga menunjukkan efektivitas *educational data mining (EDM)* dalam mendukung proses pengambilan keputusan akademik. Menurut Alturki dan Alturki (2021), EDM mampu memprediksi prestasi siswa secara dini sehingga lembaga pendidikan dapat memberikan intervensi yang tepat waktu. Selain itu, Alshareef et al. (2020) menjelaskan bahwa penerapan metode clustering dan klasifikasi dalam EDM dapat meningkatkan efisiensi analisis data akademik secara signifikan.

Namun pada kenyataannya, di berbagai lembaga pendidikan, termasuk di MTs Miftahul Ulum Bengkak, proses penentuan siswa berprestasi masih dilakukan secara manual dan subjektif, berdasarkan pengamatan guru dan rekap nilai akademik. Proses ini tidak hanya memakan waktu, tetapi juga berisiko menimbulkan ketidakakuratan dan bias dalam pengambilan keputusan. Kondisi tersebut menunjukkan adanya kesenjangan antara kondisi ideal di

mana penilaian siswa seharusnya dilakukan secara objektif dan berbasis data dengan kondisi nyata di lapangan (Qodariyah et al., 2023).

Oleh karena itu, diperlukan inovasi sistem yang dapat membantu pihak sekolah dalam mengelompokkan siswa berdasarkan tingkat prestasi menggunakan pendekatan ilmiah dan berbasis teknologi. Salah satu metode yang dapat digunakan adalah algoritma K-Means Clustering, sebuah metode analisis data yang mampu mengelompokkan data ke dalam beberapa kategori berdasarkan kemiripan atribut (Apriandi et al., 2024). Dengan menerapkan metode ini, sekolah dapat mengidentifikasi kelompok siswa berprestasi tinggi, sedang, dan rendah secara otomatis dan objektif. Menurut (Handoko dan Purnomo 2021), penerapan algoritma K-Means dalam dunia pendidikan dapat meningkatkan objektivitas dalam menilai prestasi belajar siswa karena prosesnya berbasis data. Pendapat serupa dikemukakan oleh Fadillah dan Arifin (2022) yang menjelaskan bahwa K-Means efektif dalam mengelompokkan data akademik berdasarkan kesamaan nilai rata-rata antar siswa. Selain itu, (Alpaydin 2020) menegaskan bahwa K-Means termasuk algoritma *unsupervised learning* yang sederhana namun efisien dalam menemukan pola tersembunyi dari kumpulan data besar. Dengan menerapkan metode ini, sekolah dapat mengidentifikasi kelompok siswa berprestasi tinggi, sedang, dan rendah secara otomatis, akurat, dan bebas dari unsur subjektivitas guru.

Penelitian ini bertujuan untuk mengembangkan sistem analisis data berbasis data mining dengan algoritma K-Means Clustering untuk menentukan siswa berprestasi di MTs Miftahul Ulum Bengkak. Pendekatan ini menjadi inovasi baru karena mengintegrasikan teknologi informasi dengan sistem penilaian prestasi siswa, sehingga menghasilkan proses penentuan yang lebih cepat, akurat, dan bebas dari subjektivitas.

METODE

Pada penelitian ini ada beberapa langkah –langkah yang jelas di perlukan Sebagai acuan dalam proses pengembangan penelitian, kerangka kerja ini menjelaskan tahapan-tahapan yang dilakukan selama penelitian berlangsung. Rangkaian langkah tersebut disajikan secara visual pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

1. Identifikasi Masalah

Tahap awal dalam penelitian ini dimulai dengan mengidentifikasi permasalahan utama yang menjadi fokus kajian. Langkah ini dilakukan untuk memahami situasi yang terjadi pada objek penelitian serta faktor-faktor yang memengaruhinya, sehingga diperoleh gambaran yang jelas mengenai dasar permasalahan yang akan dikaji. Melalui proses ini, peneliti dapat menentukan pendekatan yang tepat dalam mengelompokkan prestasi siswa berdasarkan nilai mata pelajaran yang diperoleh.

2. Studi Literatur

Mengumpulkan teori dan referensi yang berkaitan dengan *data mining*, *algoritma K-Means*, serta konsep penentuan siswa berprestasi sebagai dasar dalam analisis data

3. Pengumpulan Data

Pengumpulan data secara langsung di lapangan melalui observasi dan dokumentasi terhadap data nilai akademik belajar siswa

4. Data Mining

Data mining merupakan kegiatan yang dilakukan untuk menemukan pola atau informasi yang menarik dalam sekumpulan data tertentu dengan

memanfaatkan teknik

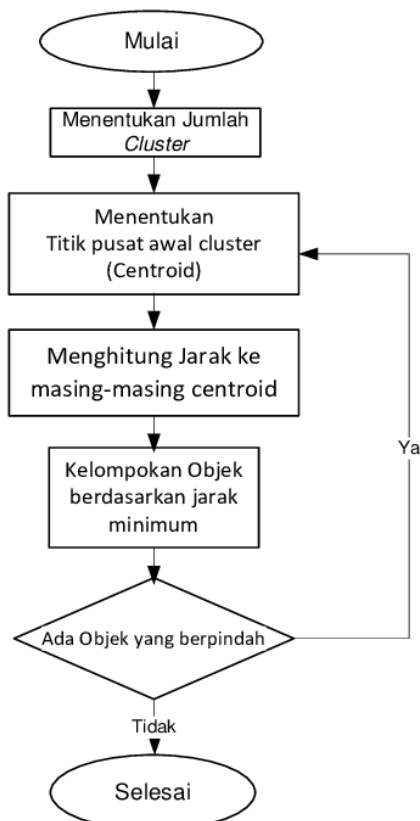
atau pendekatan tertentu. Berbagai teknik, metode, atau algoritma yang digunakan dalam data mining sangat beragam. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan keseluruhan proses Penemuan Pengetahuan dalam Database (KDD) (Mardi, 2014).

5. Clustering

Clustering merupakan salah satu metode dalam *data mining* yang digunakan untuk mengelompokkan data menjadi beberapa kelompok (*cluster*) berdasarkan tingkat kesamaan atribut atau karakteristik antar data. Objek-objek dengan kemiripan tertentu akan tergabung dalam satu *cluster*, sedangkan objek yang memiliki perbedaan ciri akan ditempatkan pada *cluster* yang lain. Pada penelitian ini, metode *clustering* diterapkan untuk mengelompokkan siswa berdasarkan rata-rata nilai, tingkat kehadiran, serta kedisiplinan, sehingga dapat diidentifikasi kelompok siswa dengan prestasi tinggi, sedang, maupun rendah (Apriandi et al., 2024).

6. K-Means

Metode K-Means adalah salah satu algoritma *clustering* yang banyak digunakan dalam bidang *data mining* karena mampu mengelompokkan data berdasarkan tingkat kesamaan karakteristik antar objek. Proses kerja algoritma ini dimulai dengan menetapkan sejumlah titik pusat kelompok (*centroid*), kemudian menghitung jarak setiap data terhadap *centroid* terdekat menggunakan ukuran jarak, misalnya *Euclidean Distance*. Tahapan pengelompokan dilakukan berulang kali hingga posisi setiap *centroid* tidak lagi berubah secara signifikan. Dalam penelitian ini, algoritma K-Means diterapkan untuk mengelompokkan siswa ke dalam tiga kategori prestasi, yaitu tinggi, sedang, dan rendah (Apriandi et al., 2024). Langkah-langkah penerapan algoritma K-Means dijelaskan sebagai berikut:



Gambar 2. Metode Yang digunakan

Berikut Uraian diagram di atas :

1. Menentukan jumlah cluster.
2. Menentukan centroid awal secara acak.
3. Menghitung jarak setiap data yang ada terhadap setiap pusat cluster.
4. Mengelompokkan data hasil perhitungan jarak data pada centroid.
5. Menentukan centroid baru dari hasil pengklasteran.
6. Menghitung kembali jarak setiap data menggunakan pusat cluster yang baru.

HASIL DAN PEMBAHASAN

Hasil

Proses pengelompokan data pada penelitian ini dilakukan menggunakan perangkat lunak RStudio dengan metode K-Means. Data yang dianalisis berasal dari MTs Miftahul Ulum Bengkak, Kabupaten Situbondo, berupa nilai akademik siswa tahun ajaran 2023/2024 pada lima mata pelajaran utama.

Tabel 1. Data Nilai Siswa

No	Nama	MTK	Bahasa Indonesia	Bahasa Inggris	IPA	IPS
1	Alfina	89	90	79	78	79
2	Atikha	89	92	75	85	89
3	Ayu Maulida	84	91	83	78	79
4	Bilqis Jumrotul	89	94	85	78	89
5	Cindy Rahayu	89	86	83	78	83
6	Dalilatul	84	78	79	80	81
7	Dara Ayu	89	83	79	78	81
8	Davina	89	94	77	78	81
9	Devi wulandari	84	79	83	78	81
10	Faridatul jannah	89	95	87	78	81

Gambar tersebut menunjukkan proses impor data nilai siswa ke dalam perangkat lunak RapidMiner. Data yang dimasukkan terdiri atas lima variabel utama, yaitu Matematika, Bahasa Indonesia, Bahasa Inggris, IPA, dan IPS. Pada tahap ini, sistem mengenali baris pertama sebagai header (judul kolom) dan menampilkan seluruh data yang akan digunakan untuk analisis menggunakan algoritma K-Means. Data tersebut sebelumnya telah disusun dalam format Excel berdasarkan nilai akademik siswa MTs Miftahul Ulum Bengkulu tahun ajaran 2023/2024 pada Gambar 3.

A	B	C	D	E	F	
1	Nama	MTK	Bahasa Indonesia	Bahasa Inggris	IPA	IPS
2						
3	Affina	89.000	90.000	79.000	78.000	79.000
4	Abikha	89.000	92.000	75.000	85.000	89.000
5	Ayu Maulida	84.000	91.000	83.000	78.000	79.000
6	Bilqis Jumrotul	89.000	94.000	85.000	78.000	89.000
7	Cindy Rahayu	89.000	85.000	83.000	78.000	83.000
8	Dalilatul	84.000	78.000	79.000	80.000	81.000
9	Dara Ayu	89.000	83.000	79.000	78.000	81.000
10	Dawna	89.000	94.000	77.000	78.000	81.000
11	Devi wulandari	84.000	79.000	83.000	78.000	81.000
12	Fardatul jannah	89.000	95.000	87.000	78.000	81.000

Gambar 3. Data Nilai Siswa

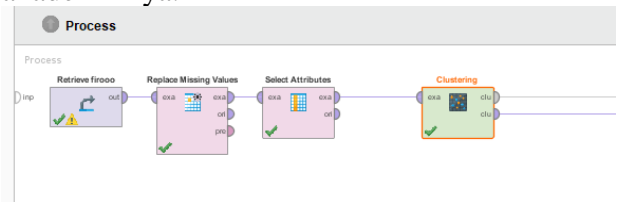
Gambar ini memperlihatkan tahap pengaturan format data di RapidMiner setelah proses impor selesai. Pada tahap ini setiap kolom diberi tipe data yang sesuai dengan karakteristiknya, seperti *integer* untuk kolom nilai mata pelajaran dan *polynomial* untuk kolom nama siswa. Pengaturan ini bertujuan agar sistem dapat mengenali jenis data dengan benar sebelum dilakukan proses analisis menggunakan algoritma K-Means. Data yang digunakan merupakan nilai akademik siswa MTs Miftahul Ulum Bengkulu tahun ajaran 2023/2024 pada Gambar 4.

Nama	MTK	Bahasa Indonesia	Bahasa Inggris	IPA	IPS
polynomial	integer	integer	integer	integer	integer
1	?	?	?	?	?
2	89	90	79	78	79
3	89	92	75	85	89
4	84	91	83	78	79
5	89	94	85	78	89
6	89	85	83	78	83
7	84	78	79	80	81
8	89	83	79	78	81
9	89	94	77	78	81
10	84	79	83	78	81
11	89	95	87	78	81

Gambar 4. Format Kolom

Gambar ini menampilkan tahapan penerapan algoritma K-Means pada aplikasi RapidMiner. Proses dimulai dengan pemanggilan dataset melalui operator *Retrieve*, dilanjutkan dengan pembersihan data kosong menggunakan *Replace Missing*

Values, serta penyaringan atribut yang relevan melalui *Select Attributes*. Tahap utama dilakukan pada operator *Clustering* dengan pengaturan jumlah cluster ($k = 3$), metode jarak *Mixed Euclidean Distance*, dan jumlah iterasi maksimum 10. Langkah-langkah tersebut digunakan untuk mengelompokkan data nilai siswa MTs Miftahul Ulum Bengkulu tahun ajaran 2023/2024 menjadi tiga kelompok berdasarkan kesamaan karakteristik akademiknya.



Gambar 5. Process

Gambar 6 Merupakan hasil clusterisasi dari data nilai siswa datanya sudah terkumpul data rata-rata nilai dari paling besar sampai paling kecil tertuang dalam tabel pada gambar berikut:

Row No.	Nama	cluster	MTK	Bahasa Indo...	Bahasa Ingg...	IPA	IPS
1	?	cluster_0	88	88	81	79	82
2	Affina	cluster_0	89	90	79	78	79
3	Abikha	cluster_2	89	92	75	85	89
4	Ayu Maulida	cluster_0	84	91	83	78	79
5	Bilqis Jumrotul	cluster_0	89	94	85	78	89
6	Cindy Rahayu	cluster_0	89	85	83	78	83
7	Dalilatul	cluster_1	84	78	79	80	81
8	Dara Ayu	cluster_1	89	83	79	78	81
9	Dawna	cluster_0	89	94	77	78	81
10	Devi wulandari	cluster_1	84	79	83	78	81
11	Fardatul jann...	cluster_0	89	95	87	78	81

Gambar 6. Hasil clusterisasi tabel

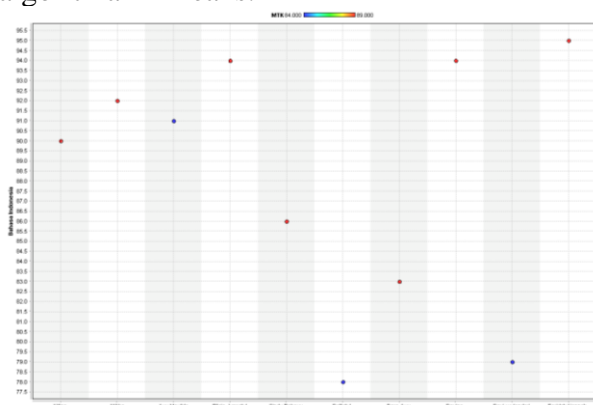
Gambar 7 merupakan hasil clusterisasi data nilai siswa yang berbentuk *cluster model*. Pada hasil tersebut terdapat Cluster 0 dengan 7 items yang menunjukkan bahwa 7 siswa termasuk dalam kelompok dengan nilai rata-rata tertinggi pertama, Cluster 1 dengan 3 items yang menunjukkan 3 siswa memiliki nilai rata-rata tertinggi ke-2, dan Cluster 2 dengan 1 item yang menggambarkan 1 siswa dengan nilai rata-rata terendah. Dengan demikian, total keseluruhan terdapat 11 data nilai rata-rata siswa yang telah dikelompokkan berdasarkan tingkat pencapaian nilainya, yaitu 7 data nilai rata-rata tertinggi pertama, 3 data nilai rata-rata tertinggi ke-2, dan 1 data nilai rata-rata siswa terendah.

Cluster Model

```
Cluster 0: 7 items
Cluster 1: 3 items
Cluster 2: 1 items
Total number of items: 11
```

Gambar 7. Cluster Model

Gambar ini menampilkan hasil visualisasi data nilai siswa dalam bentuk diagram sebar (*scatter plot*) menggunakan RapidMiner. Sumbu-X menunjukkan nama siswa, sedangkan sumbu-Y merepresentasikan nilai mata pelajaran Bahasa Indonesia. Warna titik menandakan nilai mata pelajaran Matematika, di mana variasi warna menggambarkan perbedaan skor antar siswa. Visualisasi ini membantu memperlihatkan pola perbandingan antara nilai Bahasa Indonesia dan Matematika pada setiap siswa MTs Miftahul Ulum Bengkak tahun ajaran 2023/2024 sebelum dilakukan proses pengelompokan menggunakan algoritma K-Means.



Gambar 8 Visualisasi Data

Pembahasan

Pada penelitian terdahulu yang pertama berjudul “Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means” yang dilakukan oleh (Erdi Amos Saputra dan Yessica Nataliani 2021) penelitian ini membahas tentang penerapan algoritma K-Means Clustering dalam pengelompokan data nilai siswa berdasarkan hasil akademik dan tingkat kehadiran untuk menentukan kelompok siswa berprestasi tinggi, sedang, dan rendah. Metode yang digunakan adalah

algoritma K-Means dengan tiga cluster utama, serta dilakukan pengujian menggunakan Simple Additive Weighting (SAW) untuk menentukan urutan prestasi siswa tertinggi di setiap cluster. Hasil dari penelitian ini menunjukkan bahwa penerapan metode K-Means berhasil mengelompokkan siswa secara efektif dengan tingkat kesesuaian hasil pengelompokan mencapai akurasi sebesar 85%, yang menunjukkan bahwa model clustering mampu mengelompokkan siswa secara objektif berdasarkan data nilai akademik dan kehadiran. (Saputra & Nataliani, 2021).

Pada penelitian terdahulu yang kedua berjudul “Penerapan Algoritma K-Means Clustering untuk Menentukan Siswa Berprestasi di SMAN 1 Tanjungpinang” oleh (Nia Dwi Rahayu, Fathur Rahman, dan Dinda Maharani 2023) penelitian ini membahas tentang pengelompokan data nilai siswa menggunakan algoritma K-Means untuk menentukan tingkat prestasi siswa secara objektif berdasarkan nilai akademik. Data yang digunakan meliputi nilai rata-rata dari beberapa mata pelajaran utama seperti Matematika, Bahasa Indonesia, Bahasa Inggris, dan IPA. Proses pengelompokan dilakukan menjadi tiga cluster, yaitu siswa berprestasi tinggi, sedang, dan rendah. Hasil penelitian menunjukkan bahwa metode K-Means mampu mengelompokkan siswa secara efektif, dengan hasil evaluasi Davies-Bouldin Index (DBI) sebesar 0.144, yang berarti kualitas cluster sangat baik karena mendekati nilai nol. Dengan demikian, algoritma K-Means terbukti dapat digunakan untuk membantu pihak sekolah dalam menentukan kategori siswa berprestasi secara cepat dan akurat. (Rahayu et al., 2023).

Penelitian terdahulu yang ketiga berjudul “Penerapan Algoritma K-Means Clustering untuk Menentukan Prestasi Belajar Siswa” oleh Utami dan Mulyana (2022). Penelitian ini membahas tentang penerapan algoritma K-Means dalam pengelompokan data nilai siswa guna menentukan kategori prestasi akademik di tingkat sekolah menengah. Data yang digunakan mencakup nilai dari beberapa

mata pelajaran inti, yaitu Matematika, Bahasa Indonesia, Bahasa Inggris, IPA, dan IPS, dengan total sampel sebanyak 50 siswa. Penelitian ini berfokus pada penerapan algoritma K-Means sebagai metode utama tanpa kombinasi metode lain. Hasil pengujian menunjukkan bahwa sistem berhasil mengelompokkan data ke dalam tiga cluster, yaitu kelompok siswa berprestasi tinggi, sedang, dan rendah. Evaluasi hasil cluster menggunakan *Davies-Bouldin Index (DBI)* menghasilkan nilai sebesar **0.196**, yang mengindikasikan bahwa hasil pengelompokan tergolong baik dan mampu menggambarkan perbedaan tingkat prestasi antar siswa secara akurat (Utami & Mulyana, 2022).

Penelitian lain oleh Hendrian (2018) juga menggunakan pendekatan data mining untuk mengelompokkan siswa dalam konteks pemberian bantuan pendidikan. Hasilnya menunjukkan bahwa metode klasifikasi mampu menghasilkan model prediksi yang akurat terhadap data akademik. Sementara itu, penelitian terbaru oleh Sarker et al. (2024) menggunakan berbagai teknik *educational data mining* dalam menganalisis kinerja siswa sekolah menengah di Bangladesh, dan menemukan bahwa metode clustering berbasis K-Means memberikan hasil terbaik dalam memvisualisasikan perbedaan tingkat prestasi siswa.

Pada penelitian ini yang berjudul “Analisis Data Mining dengan Algoritma K-Means Clustering untuk Menentukan Siswa Berprestasi di MTs Miftahul Ulum Bengkak”, dilakukan penerapan algoritma K-Means Clustering untuk mengelompokkan data nilai siswa berdasarkan tingkat prestasi akademik. Data yang digunakan mencakup lima mata pelajaran utama, yaitu Matematika, Bahasa Indonesia, Bahasa Inggris, IPA, dan IPS, dengan total sampel sebanyak 11 siswa. Proses analisis dilakukan menggunakan aplikasi RapidMiner Studio dengan menentukan jumlah cluster sebanyak tiga, yaitu kategori siswa berprestasi tinggi, sedang, dan rendah. Hasil pengelompokan menunjukkan bahwa Cluster 0 berisi 7 siswa (63,6%) dengan kategori berprestasi tinggi, Cluster 1 berisi 3 siswa (27,3%) dengan

kategori berprestasi sedang, dan Cluster 2 berisi 1 siswa (9,1%) dengan kategori berprestasi rendah. Proses K-Means dilakukan secara iteratif menggunakan perhitungan jarak Euclidean Distance hingga posisi centroid stabil. Hasil ini menunjukkan bahwa algoritma K-Means berhasil mengelompokkan data nilai siswa secara objektif dan akurat, serta dapat membantu pihak sekolah dalam mengidentifikasi siswa berprestasi berdasarkan data akademik yang dimiliki.

SIMPULAN (PENUTUP)

Hasil penelitian ini menunjukkan beberapa kesimpulan penting sebagai berikut: Data penelitian ini bersumber dari nilai akademik siswa MTs Miftahul Ulum Bengkak yang meliputi lima mata pelajaran utama, yaitu Matematika, Bahasa Indonesia, Bahasa Inggris, IPA, dan IPS. Jumlah data yang digunakan sebanyak 11 siswa. Proses pengelompokan dilakukan menggunakan algoritma K-Means Clustering pada aplikasi RapidMiner Studio dengan tiga cluster, yaitu berprestasi tinggi, sedang, dan rendah. Operator yang digunakan meliputi *Retrieve* untuk memanggil dataset, *K-Means* untuk membentuk model clustering, dan *Cluster Model* untuk melihat hasil pengelompokan. Hasil penelitian menunjukkan bahwa Cluster 0 berisi 7 siswa (63,6%) yang dikategorikan sebagai berprestasi tinggi, Cluster 1 berisi 3 siswa (27,3%) sebagai berprestasi sedang, dan Cluster 2 berisi 1 siswa (9,1%) sebagai berprestasi rendah. Proses iterasi K-Means menggunakan jarak Euclidean Distance hingga posisi centroid stabil dan menghasilkan pengelompokan yang akurat. Hasil ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan data nilai siswa secara objektif dan dapat membantu pihak sekolah dalam menilai serta mengidentifikasi prestasi siswa secara efisien.

UCAPAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada dosen pembimbing yang telah memberikan arahan, saran, serta motivasi selama proses penelitian berlangsung. Ucapan terima kasih juga disampaikan kepada keluarga, teman, dan

sahabat yang selalu memberikan dukungan moral maupun semangat hingga penelitian ini dapat diselesaikan. Semoga hasil penelitian ini dapat memberikan manfaat bagi pembaca, khususnya dalam penerapan teachable machine.

DAFTAR PUSTAKA

- Apriandi, D., Sari, R. M., & Sarif, M. I. (2024). Analisis clustering untuk menentukan siswa berprestasi di SMK Swasta TI Panca Dharma Stungkit menggunakan metode K-Means. *Jurnal Minfo Polgan*, 13(1), 1–8. <https://ejournal.polgan.ac.id/index.php/minfo/article/view/3100>
- Asmawi, M., Syafei, W., & Yamin, M. (2019). Pemanfaatan teknologi informasi dan komunikasi dalam dunia pendidikan. *Jurnal Teknologi dan Pembelajaran*, 5(1), 12–19.
- Kharis, S. A. A., & Zili, A. H. A. (2022). Learning analytics dan educational data mining pada data pendidikan. *Jurnal Riset Pembelajaran Matematika Sekolah (JRPMS)*, 6(2), 133–142. <https://journal.unesa.ac.id/index.php/jrpms/article/view/22987>
- Qodariyah, L., Rengganis, D., & Setiawati, E. (2023). Penerapan algoritma K-Means untuk clustering data nilai siswa dalam menentukan siswa berprestasi. *Jurnal Informatika dan Komputer (JIKO)*, 3(1), 55–63. <https://doi.org/10.33395/jiko.v3i1.234>
- Mardi, Y. (2014). Data mining: Klasifikasi menggunakan algoritma C4.5. *Jurnal Edik Informatika – Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, 1(2), 75–82. <https://doi.org/10.31537/jei.v1i2.145>
- Saputra, E. A., & Nataliani, Y. (2021). Analisis pengelompokan data nilai siswa untuk menentukan siswa berprestasi menggunakan metode clustering K-Means. *Journal of Information Systems and Informatics (JISI)*, 3(1), 78–87. <https://doi.org/10.24002/jisi.v3i1.101>
- Utami, R. D., & Mulyana, E. (2022). Penerapan algoritma K-Means clustering untuk menentukan prestasi belajar siswa. *Jurnal Teknologi Informasi dan Komunikasi (JTIK)*, 9(1), 33–42. <https://doi.org/10.33395/jtik.v9i1.2189>
- Rahayu, N. D., Rahman, F., & Maharani, D. (2023). Penerapan algoritma K-Means clustering untuk menentukan siswa berprestasi di SMAN 1 Tanjungpinang. *Jurnal Komputer dan Informatika (JUKI)*, 2(3), 101–109. <https://doi.org/10.33395/juki.v2i3.175>
- Alpaydin, E. (2020). *Introduction to Machine Learning* (4th ed.). MIT Press.
- Handoko, D., & Purnomo, W. (2021). Implementasi algoritma K-Means dalam pengelompokan data akademik siswa. *Jurnal Teknologi Informasi dan Komputer*, 4(2), 56–63.
- Fadillah, R., & Arifin, M. (2022). Penerapan K-Means untuk analisis prestasi akademik siswa. *Jurnal Informatika Edukasi*, 9(1), 23–31.
- Alturki, S., & Alturki, N. (2021). Using Educational Data Mining to Predict Students' Academic Performance for Applying Early Interventions. *Journal of Information Technology Education: Innovations in Practice*, 20, 121–137. <https://www.informingscience.org/Publications/4835>
- Alshareef, F., Alhakami, H., Alsubait, T., & Baz, A. (2020). Educational Data Mining Applications and Techniques. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 11(4). <https://doi.org/10.14569/IJACSA.2020.0110494>
- Hendrian, S. (2018). Algoritma klasifikasi data mining untuk memprediksi siswa dalam memperoleh bantuan dana pendidikan. *Faktor: Jurnal Ilmiah*, 11(3). https://journal.lppmunindra.ac.id/index.php/Faktor_Exacta/article/view/2777
- Sarker, S., Rahman, M., & Alam, M. (2024). Analyzing students' academic performance using educational data mining techniques. *Heliyon*, 10(3), e25519. <https://www.sciencedirect.com/science/article/pii/S2666920X24000663>