



Analisis Klasterisasi Data Mahasiswa Menggunakan Algoritma K-Means untuk Mendukung Pengambilan Keputusan Akademik

Yasir Muin¹, Nukman²

¹Teknologi Informasi, Universitas Khairun, Indonesia

²Teknologi Informasi, Institut Teknologi dan Kesehatan Aspirasi, Indonesia

Email : yasirmuin@unkhair.ac.id¹ , nukman.itka@gmail.com²

ABSTRAK

Penelitian ini bertujuan menganalisis pola karakteristik akademik mahasiswa menggunakan algoritma *K-Means Clustering* sebagai pendukung pengambilan keputusan akademik. Data mahasiswa dapat mengandung berbagai informasi akademik, seperti indeks prestasi, jumlah SKS yang ditempuh, kehadiran, dan capaian nilai mata kuliah, tetapi data tersebut belum selalu dimanfaatkan secara optimal untuk mengidentifikasi kelompok mahasiswa dengan karakteristik yang serupa. Penelitian menggunakan pendekatan *Educational Data Mining* dengan tahapan pemahaman data, prapemrosesan, transformasi, penerapan algoritma K-Means, dan evaluasi kualitas kluster. Penentuan jumlah kluster dilakukan dengan mempertimbangkan metode *Elbow* dan/atau *Silhouette Score*, sedangkan hasil kluster dianalisis berdasarkan karakteristik masing-masing kelompok. Pendekatan ini diharapkan mampu menghasilkan informasi mengenai pola akademik mahasiswa sehingga program studi dapat menyusun strategi pembinaan, monitoring, dan evaluasi akademik secara lebih terarah. Hasil akhir penelitian tidak dimaksudkan sebagai pengganti keputusan akademik, melainkan sebagai informasi pendukung berbasis data.

Kata Kunci: *K-Means Clustering*, *Educational Data Mining*, data mahasiswa, performa akademik, klasterisasi, pengambilan keputusan.

ABSTRACT

This study aims to analyze patterns of student academic characteristics using the K-Means Clustering algorithm to support academic decision-making. Student data encompasses various academic details—such as grade point averages, credit hours completed, attendance, and course grades—yet this information is not always optimally utilized to identify groups of students with similar characteristics. The study employs an Educational Data Mining approach comprising stages of data understanding, preprocessing, transformation, K-Means algorithm application, and cluster quality evaluation. The number of clusters is determined using the Elbow method and/or the Silhouette Score, while the resulting clusters are analyzed based on the specific characteristics of each group. This approach is expected to yield insights into student academic patterns, enabling study programs to formulate more targeted strategies for academic guidance, monitoring, and evaluation. The study's findings are not intended to replace academic decisions but rather to serve as data-driven supporting information.

Keywords: *K-Means Clustering*, *Educational Data Mining*, student data, academic performance, clustering, decision making.

*Corresponding author: yasirmuin@unkhair.ac.id

© 2025 TAMPIASIH

PENDAHULUAN

Perkembangan teknologi informasi telah menyebabkan perguruan tinggi menghasilkan dan menyimpan data akademik dalam jumlah yang semakin besar. Data tersebut dapat berasal dari sistem informasi akademik, rekam nilai mahasiswa, aktivitas pembelajaran, kehadiran, pengambilan mata kuliah, hingga informasi mengenai perkembangan studi. Ketersediaan data dalam jumlah besar memberikan peluang bagi perguruan tinggi untuk mengembangkan proses pengambilan keputusan yang lebih berbasis bukti. Dalam konteks pendidikan tinggi, pendekatan *Educational Data Mining* (EDM) dan *Learning Analytics* berkembang sebagai metode untuk menemukan pola, hubungan, serta informasi yang dapat digunakan untuk memahami proses dan hasil pembelajaran. Kajian Aldowah, Al-Samarraie, dan Fauzy menunjukkan bahwa penerapan EDM dan *learning analytics* dalam pendidikan tinggi dapat digunakan untuk berbagai kebutuhan analitik, termasuk analisis pembelajaran, perilaku, prediksi, dan visualisasi. (Aldowah et al., 2019) Romero dan Ventura juga menjelaskan bahwa perkembangan EDM telah menghasilkan berbagai pendekatan analitik yang dapat digunakan untuk memahami data pendidikan dan mendukung peningkatan proses pembelajaran. (Romero & Ventura, 2020) Dengan demikian, pemanfaatan data akademik tidak hanya berfungsi sebagai dokumentasi administratif, tetapi dapat dikembangkan menjadi sumber informasi strategis bagi perguruan tinggi.

Salah satu permasalahan dalam pengelolaan akademik adalah keberagaman karakteristik mahasiswa. Mahasiswa dalam suatu program studi tidak selalu mempunyai tingkat capaian akademik, pola pengambilan SKS, kehadiran, dan perkembangan studi yang sama. Apabila evaluasi hanya dilakukan berdasarkan nilai rata-rata atau peringkat, informasi mengenai pola kelompok mahasiswa yang mempunyai karakteristik serupa dapat kurang terlihat. Oleh karena itu, diperlukan metode analisis yang mampu mengelompokkan data mahasiswa berdasarkan kemiripan karakteristiknya. *Clustering* merupakan salah satu pendekatan dalam data mining yang dapat digunakan untuk membentuk kelompok data berdasarkan tingkat kemiripan tertentu tanpa memerlukan label kelas sebelumnya. Dalam penelitian pendidikan, pendekatan klasterisasi telah digunakan untuk mengidentifikasi pola performa dan perilaku mahasiswa. Alalawi, Shaharane, dan Jamil menunjukkan bahwa data mahasiswa yang mencakup informasi demografis, akademik, dan aktivitas dapat dimanfaatkan untuk menghasilkan informasi yang lebih berguna bagi institusi pendidikan. (Sultan Alalawi et al., 2023) Sementara itu, Mohd Talib, Abd Majid, dan Sahran menggunakan K-Means untuk mengidentifikasi pola perilaku mahasiswa pada pendidikan tinggi dan menunjukkan bahwa pengelompokan dapat membantu mengidentifikasi karakteristik kelompok mahasiswa sebelum dilakukan analisis lanjutan. (Mohd Talib et al., 2023)

Algoritma K-Means dipilih dalam penelitian ini karena merupakan salah satu algoritma *partitioning clustering* yang relatif sederhana dan sesuai untuk data numerik. Algoritma tersebut bekerja dengan menentukan sejumlah pusat klaster (*centroid*), menghitung jarak setiap objek terhadap pusat klaster, menetapkan objek pada klaster terdekat, kemudian memperbarui posisi centroid secara iteratif sampai kondisi konvergensi tercapai. Dalam konteks data akademik, K-Means dapat digunakan untuk mengelompokkan mahasiswa berdasarkan kombinasi variabel seperti IPK, jumlah SKS, nilai mata kuliah, kehadiran, atau indikator akademik lain yang tersedia. Penelitian Alberto, Sari, Samsuryadi, dan Primanita sebelumnya menggunakan K-Means untuk mengelompokkan performa akademik mahasiswa berdasarkan indeks prestasi selama beberapa semester dan menunjukkan bahwa metode tersebut dapat menghasilkan kelompok performa akademik secara konsisten. Penelitian Amalia dkk. juga menunjukkan penerapan K-Means pada data nilai akademik yang kemudian dikombinasikan dengan evaluasi *Silhouette Width* dan pengembangan dashboard untuk membantu pemantauan performa. (Amalia et al., 2021) Temuan tersebut memperlihatkan bahwa

klasterisasi dapat menjadi pendekatan deskriptif untuk mengubah data akademik yang kompleks menjadi kelompok yang lebih mudah diinterpretasikan.

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada analisis klasterisasi data mahasiswa menggunakan algoritma K-Means untuk mendukung pengambilan keputusan akademik. Tujuan utama penelitian adalah mengidentifikasi kelompok mahasiswa berdasarkan karakteristik akademiknya, menentukan jumlah klaster yang paling sesuai, serta menginterpretasikan karakteristik setiap klaster sebagai informasi pendukung bagi pengelola program studi. Proses analisis dilakukan melalui tahapan pengumpulan dan pemahaman data, pembersihan data, pemilihan atribut, normalisasi atau standardisasi, penerapan K-Means, evaluasi kualitas klaster, dan interpretasi hasil. Penggunaan evaluasi seperti *Elbow Method*, *Silhouette Score*, atau *Davies-Bouldin Index* diperlukan agar pemilihan jumlah klaster tidak semata-mata didasarkan pada asumsi. Penelitian mengenai klasifikasi performa akademik mahasiswa juga memperlihatkan pentingnya evaluasi kualitas klaster sebelum hasilnya digunakan untuk interpretasi pendidikan. Dengan pendekatan tersebut, hasil penelitian diharapkan dapat memberikan gambaran objektif mengenai segmentasi mahasiswa dan membantu perguruan tinggi merancang kebijakan akademik yang lebih tepat sasaran, misalnya monitoring mahasiswa yang membutuhkan pendampingan, pengembangan program pengayaan bagi kelompok tertentu, dan evaluasi strategi pembelajaran. Namun demikian, hasil klaster tetap harus diposisikan sebagai informasi pendukung karena keputusan akademik tidak seharusnya ditentukan hanya berdasarkan hasil algoritma.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *Educational Data Mining* untuk menganalisis data akademik mahasiswa. Kerangka kerja penelitian dapat menggunakan konsep CRISP-DM yang terdiri atas *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Pendekatan tersebut sesuai untuk penelitian yang bertujuan mengolah data pendidikan menjadi informasi yang dapat digunakan dalam proses pengambilan keputusan. Penelitian mengenai analisis performa akademik juga telah menggunakan CRISP-DM sebagai kerangka pengembangan proses data mining karena tahapan tersebut memberikan alur yang sistematis dari pemahaman kebutuhan sampai evaluasi hasil. Pada tahap *business understanding*, penelitian diarahkan untuk mengetahui kebutuhan program studi terhadap informasi mengenai karakteristik mahasiswa. Pada tahap *data understanding*, peneliti melakukan identifikasi sumber data, jumlah record, atribut, tipe data, dan kondisi kelengkapan data. Data dapat berasal dari sistem informasi akademik atau basis data akademik resmi perguruan tinggi dengan tetap memperhatikan aspek kerahasiaan dan perlindungan data mahasiswa. Variabel yang digunakan perlu dipilih berdasarkan relevansinya terhadap tujuan penelitian, sehingga tidak semua atribut administratif harus dimasukkan ke dalam proses klasterisasi.

Tahap berikutnya adalah persiapan data (*data preparation*). Data yang diperoleh perlu diperiksa untuk menemukan nilai kosong, duplikasi, kesalahan input, dan data yang tidak sesuai dengan karakteristik variabel. Data mahasiswa yang tidak lengkap dapat ditangani menggunakan metode yang sesuai dengan karakteristik permasalahan, sedangkan data duplikat perlu dihapus agar tidak memberikan bobot berlebihan terhadap observasi tertentu. Selanjutnya dilakukan pemilihan atribut, misalnya IPK, jumlah SKS ditempuh, jumlah SKS lulus, rata-rata nilai mata kuliah, persentase kehadiran, dan jumlah semester yang telah ditempuh. Pemilihan atribut harus disesuaikan dengan dataset aktual dan tujuan penelitian. Variabel dengan skala yang berbeda perlu dinormalisasi atau

distandardisasi karena K-Means menggunakan perhitungan jarak sehingga variabel dengan rentang nilai lebih besar dapat mendominasi hasil kluster. Setelah preprocessing selesai, data disusun menjadi matriks yang terdiri atas baris sebagai mahasiswa dan kolom sebagai atribut akademik. Pendekatan pengolahan seperti ini sejalan dengan penelitian K-Means dalam bidang pendidikan yang melakukan preprocessing sebelum proses pembentukan kluster. Studi terbaru bahkan menunjukkan pentingnya normalisasi, evaluasi beberapa kandidat jumlah kluster, serta pengujian stabilitas hasil kluster.

Pada tahap *modeling*, algoritma K-Means digunakan untuk membagi mahasiswa ke dalam sejumlah kelompok berdasarkan kemiripan atribut. Secara matematis, K-Means berupaya meminimalkan jumlah kuadrat jarak antara setiap data dengan centroid kluster tempat data tersebut berada. Jarak

Euclidean dapat dirumuskan sebagai $d(x_i, c_j) = \sqrt{\sum_{m=1}^p (x_{im} - c_{jm})^2}$, dengan x_i sebagai data mahasiswa, c_j sebagai centroid kluster, dan p sebagai jumlah atribut. Proses dimulai dengan menentukan nilai K , kemudian menetapkan centroid awal, menghitung jarak setiap data terhadap centroid, menetapkan data ke centroid terdekat, menghitung kembali centroid berdasarkan rata-rata anggota kluster, dan mengulangi proses sampai perubahan keanggotaan atau centroid memenuhi kondisi berhenti. Beberapa nilai K perlu diuji untuk mendapatkan jumlah kluster yang paling sesuai. *Elbow Method* dapat digunakan dengan mengamati perubahan nilai *Within-Cluster Sum of Squares* (WCSS), sedangkan *Silhouette Score* dapat digunakan untuk mengukur kualitas pemisahan dan kekompakan kluster. *Davies-Bouldin Index* juga dapat digunakan sebagai indikator tambahan, dengan nilai yang lebih rendah secara umum menunjukkan pemisahan kluster yang lebih baik. Penelitian tentang K-Means pada data akademik mahasiswa telah menggunakan pendekatan evaluasi tersebut untuk memperoleh konfigurasi kluster yang lebih dapat dipertanggungjawabkan.

Tahap terakhir adalah evaluasi dan interpretasi. Evaluasi dilakukan dengan membandingkan hasil beberapa konfigurasi K , misalnya $K = 2$, $K = 3$, $K = 4$, dan $K = 5$, kemudian menentukan konfigurasi yang memberikan kualitas kluster dan interpretasi akademik paling baik. Hasil evaluasi selanjutnya digunakan untuk menghitung karakteristik setiap kluster berdasarkan nilai rata-rata atau median setiap atribut. Misalnya, apabila sebuah kluster mempunyai rata-rata IPK dan jumlah SKS lulus yang relatif tinggi, kelompok tersebut dapat diinterpretasikan sebagai kelompok dengan capaian akademik relatif tinggi. Sebaliknya, kelompok dengan nilai akademik dan progres studi relatif rendah dapat menjadi kelompok yang perlu mendapatkan perhatian lebih lanjut. Interpretasi tersebut tidak boleh langsung dianggap sebagai label “mahasiswa bermasalah”, karena K-Means hanya menunjukkan pola kemiripan data. Hasil kluster perlu divalidasi oleh pihak akademik dan dibandingkan dengan konteks kebijakan institusi. Penelitian mengenai EDM menegaskan bahwa data mining dapat membantu pengambilan keputusan pendidikan, tetapi hasil analitik perlu diterjemahkan secara hati-hati dalam konteks pendidikan.

HASIL DAN PEMBAHASAN

Berdasarkan rancangan penelitian, hasil awal yang perlu disajikan adalah karakteristik dataset setelah proses preprocessing. Tabel hasil penelitian sebaiknya memuat jumlah data awal sebanyak [460] mahasiswa, kemudian jumlah data setelah penghapusan duplikasi sebanyak [357], jumlah data setelah penanganan nilai kosong sebanyak [103], serta atribut yang digunakan dalam proses clustering. Apabila variabel penelitian terdiri atas IPK, jumlah SKS ditempuh, jumlah SKS lulus, rata-rata nilai, dan persentase kehadiran, maka masing-masing variabel perlu ditampilkan dalam bentuk statistik deskriptif seperti nilai minimum, maksimum, rata-rata, dan standar deviasi. Tahap ini penting karena karakteristik awal data dapat memengaruhi interpretasi hasil kluster. Misalnya, variabel yang mempunyai variasi sangat tinggi perlu mendapatkan perhatian pada tahap standardisasi. Proses

preprocessing juga perlu dilaporkan secara transparan agar penelitian dapat direplikasi. Dalam penelitian EDM, kualitas data merupakan bagian penting karena kesalahan, ketidaklengkapan, atau ketidaksesuaian atribut dapat menghasilkan pola yang kurang representatif. Kajian sistematis mengenai EDM menunjukkan bahwa clustering merupakan salah satu teknik yang banyak digunakan untuk menemukan pola dalam data pendidikan. Dengan demikian, hasil preprocessing tidak hanya menjadi tahap teknis, tetapi merupakan bagian dari validitas penelitian.

Pada tahap penentuan jumlah klaster, hasil pengujian dapat disajikan menggunakan nilai *WCSS*, *Silhouette Score*, dan/atau *Davies-Bouldin Index*. Berdasarkan hasil pengolahan dataset aktual, nilai tersebut perlu dimasukkan ke dalam $K = 2$ menghasilkan *Silhouette* [3.30], $K = 3$ menghasilkan [3.20], $K = 4$ menghasilkan [3.15], dan $K = 5$ menghasilkan [3.15]. Nilai K terbaik dipilih berdasarkan kombinasi antara kualitas statistik dan interpretasi akademik. Apabila metode *Elbow* menunjukkan titik perubahan paling jelas pada $K = [\text{nilai}]$, sedangkan nilai *Silhouette Score* juga relatif tinggi pada konfigurasi tersebut, maka jumlah tersebut dapat dipilih sebagai konfigurasi utama. Pendekatan ini lebih kuat daripada menetapkan jumlah klaster hanya berdasarkan asumsi peneliti. Penelitian Mohd Talib, Abd Majid, dan Sahran memperlihatkan bahwa K-Means dapat digunakan untuk mengidentifikasi pola mahasiswa pada pendidikan tinggi, sedangkan penelitian lain menunjukkan pentingnya membandingkan beberapa konfigurasi klaster menggunakan ukuran validitas. Penelitian terbaru tentang profil capaian akademik juga membandingkan *Elbow Method*, *Silhouette Score*, dan stabilitas klaster sebelum menentukan jumlah kelompok yang digunakan. Oleh sebab itu, hasil pemilihan K dalam penelitian ini perlu didukung oleh bukti kuantitatif, bukan hanya visualisasi.

Setelah jumlah klaster ditetapkan, analisis dilanjutkan dengan mengidentifikasi karakteristik masing-masing kelompok. Apabila hasil penelitian menghasilkan tiga klaster, misalnya, interpretasi dapat dilakukan berdasarkan nilai centroid setiap variabel. Klaster 1 memiliki centroid IPK [nilai], SKS ditempuh [nilai], rata-rata nilai [nilai], dan kehadiran [nilai]; Klaster 2 memiliki karakteristik [nilai]; sedangkan Klaster 3 mempunyai karakteristik [nilai]. Berdasarkan pola tersebut, peneliti dapat memberikan nama deskriptif seperti “kelompok capaian akademik relatif tinggi”, “kelompok capaian akademik menengah”, dan “kelompok capaian akademik relatif rendah”. Penamaan tersebut harus dilakukan setelah melihat centroid, bukan sebelum algoritma dijalankan. Temuan ini dapat dibandingkan dengan penelitian terdahulu. Alberto dkk. menggunakan K-Means untuk mengelompokkan performa akademik mahasiswa berdasarkan indeks prestasi beberapa semester. Penelitian Amalia dkk. juga menunjukkan bahwa hasil clustering dapat dikembangkan menjadi dashboard untuk membantu pemantauan performa akademik. Sementara itu, Alalawi, Shaharane, dan Jamil menunjukkan bahwa data mahasiswa dapat memberikan informasi yang berguna bagi institusi ketika dianalisis dengan teknik clustering. Oleh karena itu, hasil penelitian ini berpotensi memberikan nilai praktis apabila karakteristik klaster diterjemahkan menjadi informasi yang dapat dipahami oleh pengelola akademik.

Implikasi hasil klasterisasi terhadap pengambilan keputusan akademik perlu dibahas secara proporsional. Kelompok dengan capaian akademik relatif tinggi dapat menjadi dasar informasi untuk merancang program pengayaan atau pengembangan kompetensi, sedangkan kelompok dengan capaian akademik relatif rendah dapat menjadi sasaran monitoring dan pendampingan akademik. Namun, klasterisasi tidak boleh digunakan secara tunggal untuk menentukan status mahasiswa, memberikan sanksi, atau mengambil keputusan yang berdampak signifikan terhadap mahasiswa. Keputusan akademik tetap memerlukan pertimbangan dosen, program studi, kebijakan institusi, serta kondisi individual mahasiswa. Hasil penelitian lebih tepat digunakan sebagai *decision support* yang membantu pihak akademik menentukan kelompok mana yang perlu diperiksa lebih lanjut. Pendekatan semacam ini konsisten dengan perkembangan EDM yang menempatkan analitik

sebagai alat untuk mendukung peningkatan proses pendidikan dan pengambilan keputusan berbasis data. Dengan demikian, kontribusi utama penelitian bukan hanya menghasilkan beberapa kelompok mahasiswa, tetapi menyediakan metode yang sistematis untuk mengubah data akademik menjadi informasi yang dapat digunakan dalam monitoring dan evaluasi. Setelah angka hasil aktual dimasukkan, pembahasan dapat diperkuat dengan membandingkan distribusi anggota setiap kluster, nilai centroid, nilai validitas kluster, serta implikasi akademiknya.

KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Means dapat digunakan sebagai pendekatan *Educational Data Mining* untuk mengidentifikasi kelompok mahasiswa berdasarkan kemiripan karakteristik akademik. Data akademik yang pada awalnya tersimpan dalam bentuk record individual dapat diolah melalui proses preprocessing, standardisasi, clustering, dan evaluasi sehingga menghasilkan kelompok mahasiswa dengan karakteristik tertentu. Pendekatan ini relevan karena mahasiswa dalam suatu program studi memiliki kondisi akademik yang beragam sehingga analisis berbasis kelompok dapat memberikan informasi tambahan dibandingkan hanya menggunakan statistik deskriptif atau peringkat akademik. K-Means memiliki mekanisme yang relatif sederhana untuk mengelompokkan data numerik berdasarkan kedekatan terhadap centroid. Penelitian terdahulu telah memperlihatkan penggunaannya dalam analisis performa dan perilaku mahasiswa, termasuk pada konteks pendidikan tinggi. Dengan demikian, penggunaan K-Means dalam penelitian ini dapat menjadi salah satu alternatif untuk melakukan segmentasi data mahasiswa secara sistematis.

Pemilihan jumlah kluster merupakan bagian penting dalam penelitian karena jumlah kelompok akan menentukan bentuk interpretasi hasil. Oleh sebab itu, penelitian tidak cukup hanya menjalankan algoritma K-Means dengan satu nilai K, tetapi perlu membandingkan beberapa kemungkinan jumlah kluster menggunakan metode evaluasi seperti *Elbow Method*, *Silhouette Score*, dan *Davies-Bouldin Index*. Konfigurasi terbaik harus ditentukan berdasarkan kualitas pemisahan data sekaligus relevansi interpretasi akademiknya. Apabila hasil pengolahan aktual menunjukkan bahwa konfigurasi [K terbaik] mempunyai performa evaluasi paling baik, maka konfigurasi tersebut dapat digunakan sebagai model utama penelitian. Pendekatan evaluasi tersebut juga didukung oleh penelitian yang menggunakan K-Means dalam konteks data akademik dan membandingkan kualitas beberapa konfigurasi kluster. Penggunaan beberapa indikator validasi akan meningkatkan transparansi penelitian dan mengurangi risiko pemilihan jumlah kluster secara subjektif.

Hasil klusterisasi dapat dimanfaatkan sebagai informasi pendukung untuk memahami kondisi akademik mahasiswa secara lebih terstruktur. Berdasarkan centroid dan distribusi anggota, masing-masing kluster dapat diidentifikasi berdasarkan karakteristik akademiknya. Kelompok dengan capaian relatif tinggi dapat diarahkan pada strategi pengayaan, kelompok dengan capaian menengah dapat memperoleh pemantauan berkala, sedangkan kelompok yang menunjukkan capaian relatif rendah dapat menjadi sasaran evaluasi dan pendampingan akademik lebih lanjut. Namun, interpretasi tersebut harus mempertimbangkan bahwa K-Means merupakan metode eksploratif yang menghasilkan kelompok berdasarkan atribut yang dimasukkan ke dalam model. Oleh karena itu, hasil kluster tidak dapat dianggap sebagai ukuran mutlak kemampuan mahasiswa. Informasi hasil clustering perlu dikombinasikan dengan pengetahuan akademik, evaluasi dosen, kebijakan program studi, dan konteks individual mahasiswa. Prinsip tersebut penting agar penerapan analitik pendidikan tetap berorientasi pada peningkatan kualitas pembelajaran dan bukan sekadar klasifikasi administratif.

Secara keseluruhan, penelitian ini memberikan kerangka analisis untuk memanfaatkan data akademik sebagai dasar pendukung pengambilan keputusan yang lebih berbasis data. Implementasi K-Means dapat membantu program studi menemukan pola yang mungkin sulit terlihat melalui analisis manual, terutama ketika jumlah mahasiswa dan atribut akademik semakin besar. Pengembangan selanjutnya dapat dilakukan dengan menambahkan variabel yang relevan, membandingkan K-Means dengan algoritma clustering lain, menguji stabilitas klaster, serta mengembangkan dashboard akademik yang menyajikan hasil clustering secara mudah dipahami oleh pengelola program studi. Penelitian juga perlu memperhatikan aspek privasi, keamanan data, dan interpretasi hasil agar informasi mahasiswa tidak digunakan secara tidak tepat. Dengan demikian, hasil klasterisasi sebaiknya diposisikan sebagai sistem pendukung keputusan akademik, bukan sebagai pengganti keputusan manusia. Pendekatan tersebut sejalan dengan perkembangan EDM dan *learning analytics* yang menempatkan pemanfaatan data sebagai sarana untuk meningkatkan proses pendidikan, monitoring, dan perencanaan akademik.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada rekan-rekan sejawat yang telah memberikan masukan dan saran konstruktif demi menyempurnakan karya ilmiah ini.

DAFTAR PUSTAKA

- Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13–49. <https://doi.org/10.1016/j.tele.2019.01.007> (ScienceDirect)
- Alalawi, S. J. S., Shaharane, I. N. M., & Jamil, J. M. (2023). Clustering student performance data using k-means algorithms. *Journal of Computational Innovation and Analytics*, 2(1). <https://doi.org/10.32890/jcia2023.2.1.3> (UUM e-journals System)
- Amalia, N. L. R., Supianto, A. A., Setiawan, N. Y., Zilvan, V., Yuliani, A. R., & Ramdan, A. (2021). Student academic mark clustering analysis and usability scoring on dashboard development using K-Means algorithm and system usability scale. *Jurnal Ilmu Komputer dan Informasi*, 14(2). <https://doi.org/10.21609/jiki.v14i2.980> (Jiki)
- Andini, R. F. D., Liantoni, F., & Budianto, A. (2025). Clustering student competencies using the K-Means algorithm. *Ultimatics: Jurnal Teknik Informatika*, 17(1), 99–106. <https://doi.org/10.31937/ti.v17i1.4071> (E-Journals UMN)
- Anwar, R. A. H., & Liantoni, F. (2025). Identifying academic performance patterns among PTIK students using K-Means clustering. *Ultimatics: Jurnal Teknik Informatika*, 17(2). <https://doi.org/10.31937/ti.v17i2.3998> (E-Journals UMN)
- Hegazi, M. O., & Abugroon, M. A. (2016). The state of the art on educational data mining in higher education. *International Journal of Computer Trends and Technology*, 31(1), 46–56. <https://doi.org/10.14445/22312803/IJCTT-V31P109> (DOI)

- Khasanah, A. U., & Harwati. (2019). Educational data mining techniques approach to predict student's performance. *International Journal of Information and Education Technology*, 9(2), 115–118. <https://doi.org/10.18178/ijiet.2019.9.2.1184> ([Jurnal Teknologi Informasi dan Pendidikan](#))
- Li, Y., & Gou, J. (2019). Educational data mining for students' performance based on fuzzy C-means clustering. *The Journal of Engineering*, 2019(11), 8245–8250. <https://doi.org/10.1049/joe.2019.0938> ([IET Research](#))
- Mohd Talib, N. I., Abd Majid, N. A., & Sahran, S. (2023). Identification of student behavioral patterns in higher education using K-Means clustering and support vector machine. *Applied Sciences*, 13(5), 3267. <https://doi.org/10.3390/app13053267> ([MDPI](#))
- Panjaitan, B. (2023). Clustering data akademik mahasiswa Fakultas Teknik USNI dengan algoritma K-Means. *Jurnal Satya Informatika*, 2(2), 1–8. <https://doi.org/10.59134/jsk.v2i2.425> ([FT USNI](#))
- Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 41(4), 1432–1462. <https://doi.org/10.1016/j.eswa.2013.08.042> ([ScienceDirect](#))
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355> ([Wiley Interdisciplinary Reviews](#))
- Ismaya, N., & Zyen, K. (2022). Penentuan siswa berprestasi menggunakan metode K-Means clustering di SMP Takhasus Al Qur'an Sadamiyyah. *JTINFO: Jurnal Teknik Informatika*, 1(1), 64–48. <https://doi.org/10.02220/jtinfo.v1i1.152> ([Jurnal Unisnu](#))
- Durachman, Y., & Abdul Rahman, A. W. (2025). Clustering student behavioral patterns: A data mining approach using K-Means for analyzing study hours, attendance, and tutoring sessions in educational achievement. *Artificial Intelligence in Learning*, 1(1). <https://doi.org/10.63913/ail.v1i1.48> ([DOI](#))
- M. F. A., Abidin, D. Z., & Jasmir. (2026). K-Means clustering with Elbow method and validity indices for classifying student academic achievement based on knowledge scores at SDN 48 Kota Jambi. *Jurnal Teknik Informatika (Jutif)*, 7(1). <https://doi.org/10.52436/1.jutif.2026.7.1.5349> ([JUTIF](#))
- Limbong, J. J. A., & Likumahwa, F. M. (2025). Implementation of the K-Means algorithm for clustering students' web programming course grades using Silhouette Score. *Journal of System and Computer Engineering*, 6(3), 271–280. <https://doi.org/10.61628/jsce.v6i3.2100> ([Jurnal Universitas Pancasakti](#))
- Risal, A. A. N., Andayani, D. D., Suherman, M. I., & Kaswar, A. B. (2024). Utilizing the K-Means clustering algorithm for analyzing student achievement assessment at SMK Negeri 1 Gowa. *Journal of Embedded Systems, Security and Intelligent Systems*, 5(1). <https://doi.org/10.59562/jessi.v5i1.2178> ([Journal Unm](#))