**Analisis Cluster Mahasiswa Berdasarkan Faktor Gaya Belajar Dan Nilai Akademik Menggunakan Algoritma K-means**

**Rico Herdianto1, Risky Fauzan Gumilang2, Muhammad Syahrevi3,**

**Nurul Isnayni4, Cindi Debby5,**

1,2,3,4,5 Universitas Bina Sarana Informatika

e-mail: 1xxxx@xxxx.xxx, 215220266@bsi.ac.id

|  |  |  |
| --- | --- | --- |
| Diterima  31-05-2025 | Direvisi  08-06-2025 | Disetujui  01-03-2020 |

**Abstrak** - Gaya belajar merupakan aspek penting dalam proses pendidikan karena memengaruhi cara mahasiswa menyerap dan memahami informasi. Dataset yang digunakan berasal dari situs Kaggle berjudul Students Performance in Exams, yang mencakup 10.000 entri data mahasiswa dengan atribut seperti gaya belajar dan nilai akhir. Metode yang digunakan adalah pendekatan kuantitatif eksploratif, dengan preprocessing data dan implementasi algoritma K-Means melalui perangkat lunak Weka. Hasil klasterisasi menunjukkan adanya dua kelompok utama mahasiswa: klaster pertama didominasi oleh mahasiswa dengan gaya belajar visual dan kinestetik serta capaian nilai akhir tinggi, sedangkan klaster kedua terdiri dari mahasiswa dengan gaya belajar reading/writing dan kinestetik serta capaian nilai lebih rendah. Temuan ini menunjukkan bahwa gaya belajar visual memiliki kecenderungan berkontribusi terhadap hasil akademik yang lebih baik. Penelitian ini diharapkan dapat menjadi dasar dalam penerapan strategi pembelajaran adaptif di lingkungan pendidikan tinggi.

Kata Kunci: K-Means, Gaya Belajar, Klasterisasi,

**PENDAHULUAN**

Gaya belajar merupakan cara atau pendekatan individu dalam memahami, menyerap, dan mengelola informasi saat proses belajar berlangsung. Dalam konteks pendidikan tinggi, memahami gaya belajar mahasiswa sangat penting agar proses pembelajaran dapat dirancang sesuai dengan karakteristik peserta didik. Gaya belajar yang umum digunakan adalah visual, auditori, dan kinestetik. Perbedaan gaya belajar dapat memengaruhi capaian hasil belajar, yang biasanya direpresentasikan dalam bentuk nilai akademik. Oleh karena itu, mengelompokkan mahasiswa berdasarkan gaya belajar dan nilai akademik dapat menjadi strategi untuk menyesuaikan metode pengajaran yang lebih efektif dan tepat sasaran.

***Data mining*** merupakan proses menganalisis data dari berbagai sudut pandang untuk menemukan pola atau informasi yang bernilai. Informasi yang diperoleh dari proses ini dapat dimanfaatkan untuk mendukung pengambilan keputusan yang lebih tepat, meningkatkan efisiensi operasional, serta mengurangi biaya yang dikeluarkan (Rizki Agam Syahputra, 2024). Penambangan data (*Data Mining)* ditentukan oleh proses mengekstraksi dan mengidentifikasi informasi penting menggunakan matematika, statistik, kecerdasan buatan dan pembelajaran mesin, dan mendapatkan wawasan dari database besar (Prasetyo & Dewayanto, 2024).

***Machine learning*** suatu teknologi yang memungkinkan mesin belajar secara mandiri tanpa instruksi langsung dari pengguna. Pembelajaran ini melibatkan berbagai disiplin ilmu seperti statistika, matematika, dan data mining, sehingga mesin dapat mengenali pola dari data dan mengambil keputusan tanpa harus diprogram ulang secara eksplisit (Hairani et al., 2022).

*Clustering* merupakan metode dalam data mining yang bertujuan untuk membagi data ke dalam beberapa kelompok *(cluster)*, di mana setiap kelompok terdiri dari data yang memiliki karakteristik atau pola yang serupa, Tujuan dari pengelompokan ini adalah sama Meminimalkan fungsi target yang ditentukan Proses clustering yang biasa dicoba Meminimalkan variasi cluster Memaksimalkan variasi antar cluster (Priyatman et al., 2019).

Algoritma K-Means *clustering* adalah salah satu metode yang efektif dan efisien dalam mengelompokkan data berdasarkan kemiripan, sehingga mampu menghasilkan pembagian kelompok yang sesuai dengan karakteristik data yang dianalisis (Nasari & Darma, 2015). Algoritma K-Means memulai proses klasterisasi dengan memilih sejumlah titik pusat cluster awal secara acak dari keseluruhan data. Titik-titik ini berfungsi sebagai representasi awal dari setiap kelompok. Selanjutnya, algoritma akan mengevaluasi setiap data berdasarkan kedekatan jaraknya terhadap pusat cluster yang telah dipilih. Setiap data kemudian akan dialokasikan ke cluster dengan jarak terdekat, sehingga proses pembentukan kelompok dilakukan berdasarkan kemiripan karakteristik (Alkhairi & Windarto, 2019).

Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam mengelompokkan mahasiswa berdasarkan gaya belajar dan pencapaian akademik, guna menghasilkan informasi yang dapat dimanfaatkan untuk merancang strategi pembelajaran yang lebih optimal.

**METODE PENELITIAN**

Penelitian ini dilaksanakan dengan mengacu pada metode Cross Industry Standard Process for Data Mining (CRISP-DM). Metode ini memberikan kerangka kerja sistematis dalam pelaksanaan data mining (Suhanda et al., 2020). Adapun alur tahapan dalam pendekatan CRISP-DM dijelaskan sebagai berikut:

1. *Business Understanding*

Merupakan tahap awal dalam proses CRISP-DM yang berfokus pada identifikasi tujuan bisnis, pemahaman terhadap kondisi dan konteks permasalahan yang sedang dihadapi, serta perumusan tujuan penelitian yang dapat diselesaikan melalui pendekatan data mining (Dhewayani et al., 2022).

Ini bertujuan untuk mengelompokkan mahasiswa berdasarkan gaya belajar dan nilai akhir guna mendukung strategi pembelajaran yang lebih tepat sasaran. Dataset yang digunakan bersumber dari Kaggle dengan judul Students Performance in Exams, yang memuat informasi seperti gaya belajar dan nilai akademik mahasiswa. Dataset dapat diakses: <https://www.kaggle.com/datasets/spscientist/students-performance-in-exams>.

1. *Data Understanding*

Langkah yang dilakukan meliputi pemilihan tabel dan atribut yang akan dikonversi ke dalam basis data baru sebagai sumber data mentah untuk keperluan proses data mining (Budiman et al., 2014).

Dengan semua atribut yang ada pada dataset Students Performance in Exams pada penelitian jurnal ini memilih atribut utama yaitu Preferred\_Learning\_Style dan Final\_Grade untuk mengelompokan gaya belajar pada nilai akhiri mahasiswa.

1. *Data Preparation*

Pada tahap ini, data awal yang telah dipilih kemudian diproses lebih lanjut agar siap digunakan dalam proses data mining (Muhammad, 2019).

Tahap ini mencakup pembersihan dan penyesuaian data, khususnya pada atribut Preferred Learning Style dan Final Grade, agar siap digunakan dalam proses klasterisasi.

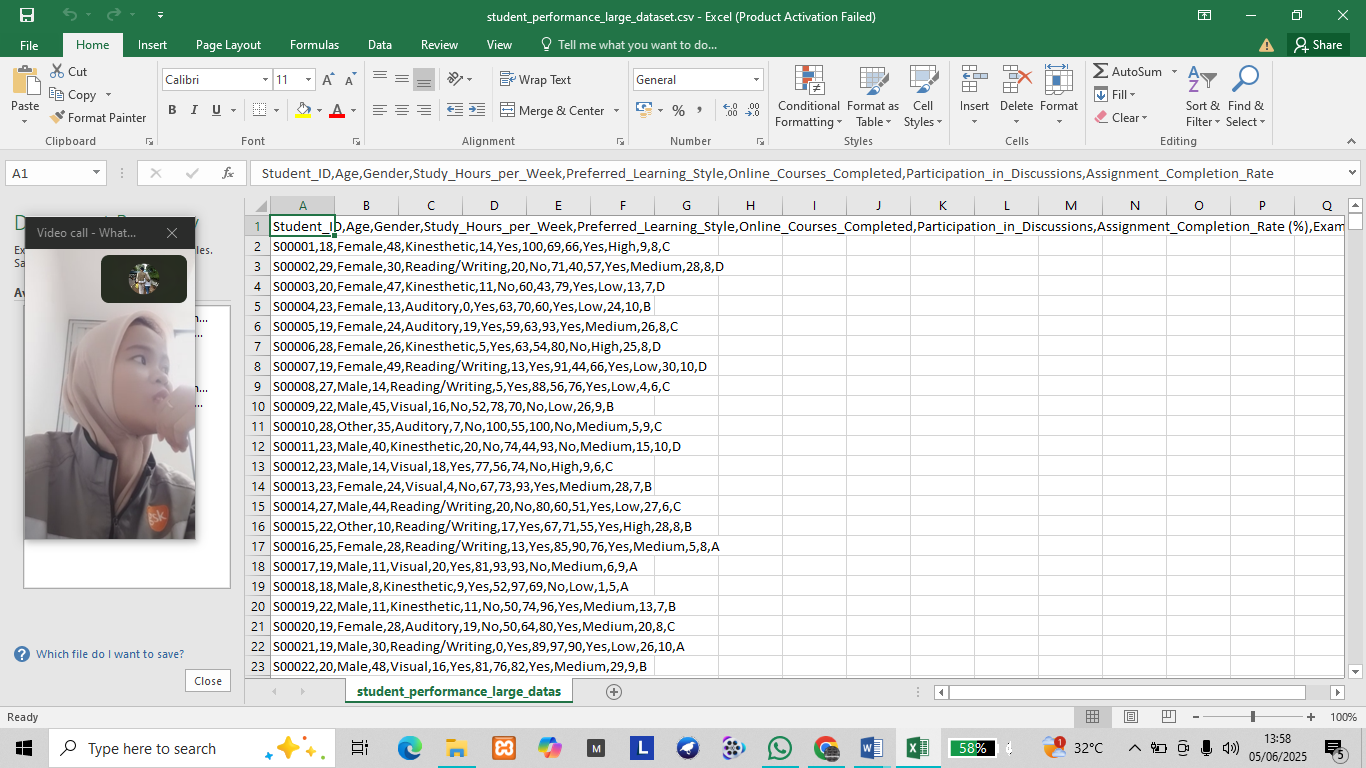
1. *Modelling*

Selanjutnya dilakukan sebuah Model pengetahuan dan pengujian dijalankan (Muhammad, 2019).

proses klasterisasi dilakukan menggunakan algoritma K-Means dengan bantuan perangkat lunak Weka. Data yang telah dipersiapkan sebelumnya diimpor ke Weka dalam format CSV, lalu dipilih algoritma SimpleKMeans.

**HASIL DAN PEMBAHASAN**

**Tabel 1.** Dataset *Student Performance In Examps*

**

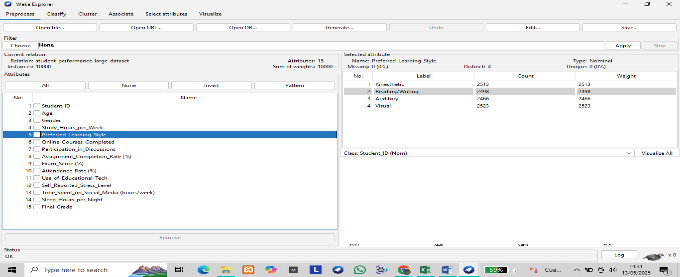
*Sumber : Kaggle.com*

Berdasarkan **Tabel 1,** Dataset yang digunakan dalam penelitian ini diperoleh dari situs Kaggle dengan judul Students Performance in Exams. Untuk keperluan analisis awal, digunakan sebanyak 23 data sampel. Setiap baris data merepresentasikan satu mahasiswa dengan sejumlah atribut penting yang relevan terhadap tujuan penelitian.

Adapun atribut yang terdapat dalam dataset ini adalah sebagai berikut:

* Student\_ID: Merupakan kode identitas unik bagi setiap mahasiswa.
* Age: Usia mahasiswa dalam satuan tahun.
* Gender: Jenis kelamin mahasiswa, terdiri dari kategori Male dan Female.
* Study\_Hours\_per\_Week: Jumlah rata-rata waktu belajar mahasiswa dalam satu minggu.
* Preferred\_Learning\_Style: Gaya belajar yang paling disukai mahasiswa, seperti Visual, Auditory, Kinesthetic, atau Reading/Writing.
* Online\_Courses\_Completed: Jumlah kursus online yang telah diselesaikan oleh mahasiswa.
* Participation\_in\_Study\_Groups: Partisipasi mahasiswa dalam kelompok belajar, dengan jawaban Yes atau No.
* Attendance\_Rate: Tingkat kehadiran mahasiswa, dikategorikan menjadi Low, Medium, dan High.
* Final\_Grade: Nilai akhir mahasiswa yang menjadi indikator utama performa akademik.

Dua atribut utama yang menjadi analisis adalah Preferred\_Learning\_Style dan Final\_Grade, karena keduanya dinilai memiliki pengaruh signifikan terhadap proses belajar dan hasil capaian akademik mahasiswa.

**

*Sumber : Di Olah Penulis Menggunakan Software Weka*

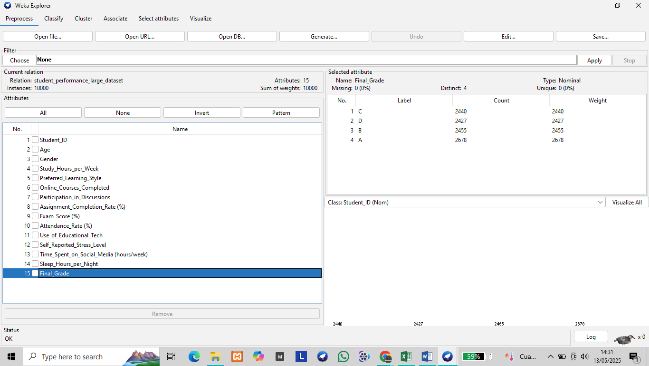
**Gambar 1.** Visualisasi atribut Preferred\_Learning\_Style pada dataset mahasiswa menggunakan Weka

Berdasarkan **Gambar 1**, yang merupakan hasil visualisasi analisis atribut *Preferred\_Learning\_Style* menggunakan perangkat lunak *Weka*, dapat diketahui bahwa mahasiswa dalam dataset ini terbagi ke dalam empat kategori utama gaya belajar, yaitu: *Kinesthetic, Reading/Writing, Auditory*, dan *Visual.* Dari total 10.000 entri data yang dianalisis, distribusi masing-masing gaya belajar adalah sebagai berikut:

* *Kinesthetic*: 2.513 mahasiswa
* *Reading/Writing:* 2.498 mahasiswa
* *Auditory:* 2.466 mahasiswa
* *Visual:* 2.523 mahasiswa

Hal ini menunjukkan bahwa **mahasiswa memiliki cara belajar yang berbeda-beda,** namun jumlahnya hampir sama banyak. Artinya, semua jenis gaya belajar cukup umum ditemui pada mahasiswa. Adapun penjelasan masing-masing gaya belajar adalah sebagai berikut:

* Mahasiswa dengan gaya belajar *Visual* cenderung menyukai informasi dalam bentuk gambar, diagram, dan visualisasi data.
* Mereka yang memiliki gaya belajar *Auditory* lebih efektif ketika belajar melalui diskusi, ceramah, atau media audio.
* Gaya belajar *Reading/Writing* menunjukkan preferensi terhadap materi dalam bentuk teks tertulis, baik berupa catatan maupun buku.
* Sedangkan gaya belajar *Kinesthetic* cenderung mengutamakan pengalaman langsung, praktik lapangan, dan pendekatan berbasis gerakan atau simulasi.

****

*Sumber : Di Olah Penulis Menggunakan Software Weka*

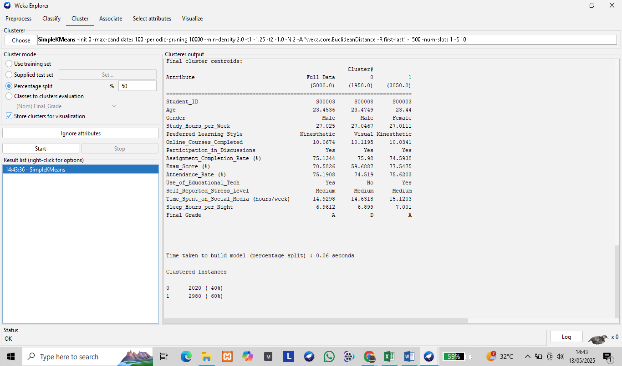
**Gambar 2.** Visualisasi atribut Final\_Grade pada dataset mahasiswa menggunakan Weka

Selanjutnya, sebagaimana ditunjukkan pada **Gambar 2**, atribut *Final\_Grade* (nilai akhir) dari seluruh 10.000 mahasiswa dalam dataset dianalisis guna mengetahui distribusi performa akademik mahasiswa berdasarkan kategori nilai. Atribut ini telah diklasifikasikan ke dalam empat kategori utama, yaitu A, B, C, dan D, yang masing-masing merepresentasikan tingkat pencapaian akademik dari yang tertinggi hingga terendah. Distribusi dari keempat kategori nilai tersebut adalah sebagai berikut:

* Nilai A: 2.678 mahasiswa
* Nilai B: 2.455 mahasiswa
* Nilai C: 2.440 mahasiswa
* Nilai D: 2.427 mahasiswa

Terdapat kecenderungan bahwa mayoritas mahasiswa mampu mencapai nilai akademik yang tergolong tinggi, ditunjukkan oleh dominasi kategori A sebagai kelompok dengan jumlah terbesar. Secara statistik, lebih dari 26% dari total populasi berhasil memperoleh nilai A, suatu pencapaian yang dapat dikatakan sangat positif dalam konteks performa akademik.

Salah satu pendekatan yang digunakan dalam penelitian ini adalah algoritma *K-Means*, yang memungkinkan pengelompokan mahasiswa berdasarkan kemiripan atribut gaya belajar (*Preferred\_Learning\_Style*) dan nilai akhir (*Final\_Grade*). Proses klasterisasi ini tidak hanya bertujuan untuk menemukan struktur atau pola tersembunyi dalam data, tetapi juga untuk membentuk dasar bagi pengembangan sistem pembelajaran adaptif yang mampu menyajikan strategi belajar berdasarkan kelompok karakteristik yang serupa.



*Sumber : Di Olah Penulis Menggunakan Software Weka*

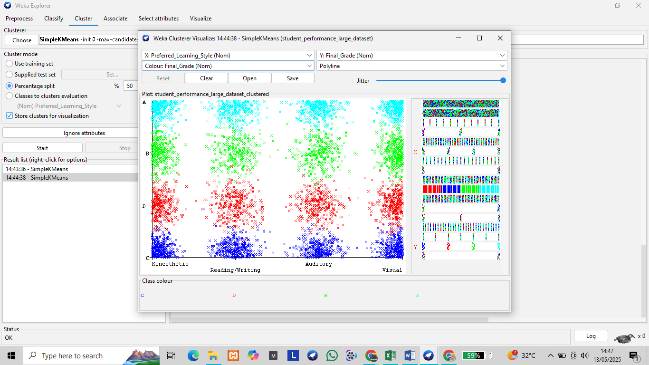
**Gambar 3.** Hasil Klasifikasi menggunakan *cluster* dengan logika *K-means*

Hasil **Gambar 3,** Proses klasterisasi dalam penelitian ini dilakukan dengan menerapkan algoritma *K-Means* yang diimplementasikan menggunakan perangkat lunak *Weka.* Dataset yang digunakan mencakup sebanyak 10.000 entri data mahasiswa, dengan berbagai atribut yang relevan terhadap performa akademik, seperti *Exam Score*, *Assignment Completion Rate*, *Final Grade*, *Study Hours per Week*, dan *Preferred Learning Style*.

Secara rinci, *Cluster 1* menunjukkan karakteristik mahasiswa dengan performa akademik tinggi, ditandai dengan nilai rata-rata *Exam Score* sebesar 77,5%, *Assignment Completion Rate* sebesar 74,5%, serta dominasi nilai akhir (*Final Grade*) pada kategori A. Sebaliknya, Cluster 0 mencerminkan kelompok mahasiswa dengan performa akademik lebih rendah, dengan nilai rata-rata *Exam Score* sebesar 59,6%, *Assignment Completion Rate* sebesar 75,1%, dan dominasi *Final Grade* pada kategori D. Perlu dicatat bahwa tingkat penyelesaian tugas pada kedua klaster relatif sebanding, namun perbedaan mencolok terlihat pada capaian nilai ujian dan nilai akhir secara keseluruhan. Hal ini mengindikasikan bahwa penyelesaian tugas semata belum cukup menjamin pencapaian akademik yang optimal, dan bahwa faktor lain turut memengaruhi perbedaan tersebut.

Terkait atribut *Preferred Learning Style,* hasil klasterisasi menunjukkan bahwa *Cluster 1* didominasi oleh mahasiswa dengan gaya belajar Visual dan Kinestetik, sedangkan *Cluster 0* lebih banyak terdiri dari mahasiswa dengan gaya belajar Kinestetik dan *Reading/Writing.* Meskipun gaya belajar *Kinestetik* muncul pada kedua *klaster,* kombinasi dengan gaya *Visual* tampaknya memberikan kontribusi yang lebih positif terhadap performa akademik. Ini mengindikasikan bahwa penggunaan media visual, pendekatan pembelajaran berbasis pengamatan, dan representasi grafis informasi dapat membantu memperkuat pemahaman dan hasil belajar mahasiswa secara signifikan.

Hasil klasterisasi ini juga dapat digunakan sebagai dasar dalam pengembangan sistem rekomendasi pembelajaran adaptif (adaptive learning systems) yang menyesuaikan materi, metode, dan evaluasi berdasarkan profil belajar mahasiswa secara individual. Secara keseluruhan, implementasi algoritma *K-Means* dalam penelitian ini memberikan gambaran awal tentang bagaimana pendekatan machine learning dapat diterapkan untuk mendukung pengambilan keputusan berbasis data dalam konteks pendidikan tinggi.

**

*Sumber : Di Olah Penulis Menggunakan Software Weka*

**Gambar 4.** Hasil Visualisi Klasifikasi menggunakan cluster dengan logika *K-means*

Visualisasi yang ditampilkan pada Gambar **4** Visualisasi hasil klasterisasi yang ditampilkan pada Gambar 4 memperlihatkan hubungan antara gaya belajar mahasiswa (Preferred Learning Style) dan pencapaian nilai akhir (Final Grade). Dari gambar tersebut terlihat adanya kecenderungan bahwa mahasiswa dengan gaya belajar **Visual dan Auditory** lebih sering memperoleh nilai A dan B. Sebaliknya, mahasiswa dengan gaya belajar **Kinestetik** serta **Reading/Writing** cenderung mendominasi kelompok dengan nilai C dan D.

Pola ini mengindikasikan bahwa gaya belajar yang lebih selaras dengan metode pengajaran digital, seperti visualisasi dan interaksi audio, berkontribusi terhadap peningkatan performa akademik. Hal ini dapat dijelaskan karena mahasiswa visual lebih mudah memahami materi melalui media grafis, animasi, dan diagram, sementara mahasiswa auditory merespon lebih baik terhadap presentasi verbal dan diskusi.

Dengan demikian, hasil visualisasi ini tidak hanya bersifat deskriptif, tetapi juga menawarkan **nilai tambah** berupa pemahaman yang lebih mendalam mengenai faktor-faktor yang memengaruhi keberhasilan akademik mahasiswa. Hal ini dapat menjadi dasar pengambilan keputusan dalam menyusun kebijakan pembelajaran yang **adaptif, inklusif, dan berbasis data**.

**KESIMPULAN**

Berdasarkan hasil penelitian yang telah dilakukan terhadap data mahasiswa menggunakan algoritma K-Means, dapat disimpulkan bahwa proses klasterisasi mampu mengelompokkan mahasiswa berdasarkan gaya belajar dan nilai akhir secara efektif. Dataset yang digunakan terdiri dari 10.000 entri mahasiswa dengan berbagai atribut, namun penelitian ini secara khusus memfokuskan pada dua atribut utama, yaitu Preferred Learning Style (gaya belajar) dan Final Grade (nilai akhir), karena keduanya dinilai memiliki peran penting dalam memengaruhi proses belajar dan performa akademik.

Hasil analisis menunjukkan bahwa gaya belajar mahasiswa terbagi secara proporsional ke dalam empat kategori utama, yaitu Visual, Auditory, Reading/Writing, dan Kinesthetic. Tidak ada dominasi signifikan di antara keempat gaya belajar tersebut, sehingga distribusinya tergolong merata. Hal ini mengindikasikan bahwa mahasiswa memiliki preferensi belajar yang beragam, dan setiap jenis gaya belajar perlu mendapatkan perhatian yang sama dalam proses pembelajaran di lingkungan pendidikan tinggi.

Distribusi nilai akhir mahasiswa juga menunjukkan pola yang relatif seimbang, dengan mayoritas mahasiswa berada pada kategori nilai A, disusul oleh B, C, dan D. Hal ini menunjukkan bahwa secara umum performa akademik mahasiswa dalam dataset tersebut tergolong baik. Namun demikian, distribusi ini tidak secara langsung menunjukkan hubungan antara gaya belajar dengan nilai akhir, sehingga diperlukan analisis lanjutan menggunakan pendekatan klasterisasi.

Melalui penerapan algoritma K-Means pada aplikasi Weka, proses klasterisasi menghasilkan dua kelompok utama. Klaster pertama (Cluster 1) didominasi oleh mahasiswa dengan capaian akademik tinggi, ditandai dengan dominasi nilai A, nilai rata-rata ujian yang tinggi, dan banyaknya mahasiswa dengan gaya belajar Visual dan Auditory. Klaster kedua (Cluster 0) mencerminkan kelompok mahasiswa dengan capaian akademik yang lebih rendah, dengan dominasi gaya belajar Reading/Writing dan Kinesthetic, serta nilai akhir yang lebih banyak berada pada kategori C dan D.

Hasil dari proses klasterisasi ini memberikan gambaran awal yang sangat berguna dalam mengembangkan sistem pembelajaran yang lebih adaptif dan berbasis data. Dengan memanfaatkan hasil analisis klaster, institusi pendidikan dapat merancang strategi pengajaran, penyusunan materi ajar, dan metode evaluasi yang disesuaikan dengan karakteristik masing-masing kelompok mahasiswa. Pendekatan ini juga dapat digunakan sebagai dasar dalam pengembangan sistem rekomendasi pembelajaran individual (personalized learning system), yang mampu memberikan pengalaman belajar yang lebih relevan dan efektif bagi setiap mahasiswa.

Secara keseluruhan, penelitian ini membuktikan bahwa penerapan algoritma K-Means dalam analisis gaya belajar dan nilai akhir mahasiswa mampu memberikan kontribusi yang signifikan dalam memahami pola belajar serta mendukung pengambilan keputusan berbasis data di lingkungan pendidikan tinggi. Hasil ini juga memperkuat urgensi integrasi antara data analytics, machine learning, dan teknologi pendidikan dalam rangka meningkatkan kualitas pembelajaran secara menyeluruh.

**REFERENSI**

Muhammad, B. (2019). Implementasi Data Mining untuk Prediksi Standar Hidup Layak Berdasarkan Tingkat Kesehatan dan Pendidikan Masyarakat. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, *2*(2), 33–37. <https://jurnal.tau.ac.id/index.php/siskom-kb/article/view/58>

Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*, *12*(1), 64–77. <https://doi.org/10.34010/jati.v12i1.6674>

Budiman, I., Prahasto, T., & Christyono, Y. (2014). Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma. *Jurnal Sistem Informasi Bisnis*, *1*(3), 129–134. <https://doi.org/10.21456/vol1iss3pp129-134>

Suhanda, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika Dan Komputer*, *6*(2). <https://doi.org/10.37012/jtik.v6i2.299>

Hairani, H., Bumigora, U., & Colab, G. (2022). *PELATIHAN IMPLEMENTASI MACHINE LEARNING PADA BIDANG*. *2*(2), 305–310. <https://doi.org/10.30821/adma.v2i2.xxxx>

Rizki Agam Syahputra, M. R. H. (2024). *Metode Analisis Kesehatan Dengan Mengguakan Mechine Learning Atau Artificial Inteligenci Atau Data Mining Literature Review*.

Alkhairi, P., & Windarto, A. P. (2019). Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara. *Seminar Nasional Teknologi Komputer & Sains*, 762–767. <https://prosiding.seminarid.com/index.php/sainteks/article/download/228/223>

Nasari, F., & Darma, S. (2015). *Seminar Nasional Teknologi Informasi dan Multimedia 2015 PENERAPAN K-MEANS CLUSTERING PADA DATA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS POTENSI UTAMA)*. 6–8.

Priyatman, H., Sajid, F., & Haldivany, D. (2019). JEPIN (Jurnal Edukasi dan Penelitian Informatika) Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika*, *5*(1), 62–66.

Prasetyo, S., & Dewayanto, T. (2024). Penerapan Machine Learning, Deep Learning, Dan Data Mining Dalam Deteksi Kecurangan Laporan Keuangan-a Systematic Literature Review. *Diponegoro Journal of Accounting*, *13*(3), 1–12. <http://ejournal-s1.undip.ac.id/index.php/accounting>