

Clustering Mahasiswa Baru Menggunakan K-Means Untuk Perencanaan Teknologi Pembelajaran di Unilak

Elvira Asril^{*1}, Hardiano Siburian²

¹Program Studi Pendidikan Vokasi & Keteknikan, Pascasarjana, Universitas Lancang Kuning

²Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Lancang Kuning

*E-mail: elvira@unilak.ac.id

Abstract

Mapping the characteristics of new students is an important first step in designing a technology-based education system, particularly at private universities, where competition to attract new students is intense. This study uses the K-means algorithm to cluster new students at Lancang Kuning University. The goal is to identify patterns that can inform the design of adaptive learning systems. The study employs a quantitative approach with data mining techniques using new student data with variables of academic and digital literacy scores. The clustering results revealed three groups: 54 students with high scores and high digital literacy, 52 students with moderate scores and moderate digital literacy, and 44 students with low scores and low digital literacy. Three main clusters emerged from the data clustering: digitally independent students, moderate students, and students requiring assistance and guidance. These results indicate differences in the learning needs of new students and support strategic recommendations for implementing educational technology tailored to their needs. These results contribute to the use of data mining technology for data-driven learning planning at Lancang Kuning University.

Keywords: Clustering, k-means, data mining, new students, education technology

Abstrak

Pemetaan karakteristik mahasiswa baru merupakan langkah awal yang penting dalam merancang sistem pendidikan berbasis teknologi, terutama di perguruan tinggi swasta saat ini yang begitu ketat persaingannya untuk mendapatkan mahasiswa baru. Penelitian ini menggunakan algoritma K-means untuk mengelompokkan mahasiswa baru di Universitas Lancang Kuning. Tujuannya adalah untuk mengidentifikasi pola yang dapat menginformasikan desain sistem pembelajaran yang adaptif. Penelitian ini menggunakan pendekatan kuantitatif dengan teknik data mining pada data mahasiswa baru, dengan variabel nilai akademik (nilai rata-rata rapor) dan skor literasi digital. Hasil pengelompokan atau clustering ini terdapat 54 mahasiswa dengan nilai tinggi dan literasi digital tinggi, 52 mahasiswa dengan nilai sedang dan literasi digital sedang serta 44 mahasiswa dengan nilai rendah dan literasi digital rendah. Pengelompokan data menghasilkan tiga klaster utama yaitu mahasiswa yang mandiri secara digital, mahasiswa yang moderat, dan mahasiswa yang membutuhkan bantuan dan bimbingan. Artinya terdapat perbedaan kebutuhan pembelajaran mahasiswa baru, mendukung rekomendasi strategis yang diperlukan untuk menerapkan teknologi pendidikan yang cocok terhadap kebutuhan mahasiswa tersebut. Hasil ini berkontribusi pada penggunaan teknologi data mining untuk perencanaan pembelajaran berbasis data fakta di Universitas Lancang Kuning.

Kata kunci: Clustering, k-means, data mining, mahasiswa baru, teknologi pendidikan

1. PENDAHULUAN

Perguruan tinggi swasta (studi kasus Universitas Lancang Kuning disingkat Unilak), setiap tahunnya menerima ratusan bahkan ribuan mahasiswa baru dengan latar belakang yang beragam, seperti asal daerah, prestasi akademik, minat, hingga kemampuan literasi. Keberagaman ini menjadi tantangan bagi Unilak untuk memberikan layanan yang optimal, secara akademik maupun non-akademik. Selain itu persaingan dengan perguruan tinggi negeri dan sesama perguruan tinggi swasta lainnya menjadi tantangan agar Unilak semakin meningkatkan kualitas dan pelayanan kepada calon mahasiswa baru. Tantangan dalam menghadirkan sistem pembelajaran yang inklusif dan responsif terhadap kebutuhan mahasiswa juga semakin beragam. Identifikasi siswa membutuhkan intervensi atau pembelajaran tambahan serta adanya rekomendasi strategi pembelajaran berbasis data akademik ([Okgistiyan et al, 2025](#)) ([Melizah et al, 2024](#)). Teknologi pendidikan sebagai sebuah pendekatan sistematis yang dapat memberikan solusi adaptif berbasis data faktual, dalam hal ini adalah dengan cara analisis pola mahasiswa menggunakan teknik data mining.

Metode pembelajaran dengan gaya belajar kaku dan tradisional yang tidak memperhatikan kondisi mahasiswa, dapat mempengaruhi terhadap jalannya proses pembelajaran. Perbedaan kondisi, cara belajar, kemampuan menerima dan mengelola informasi adalah faktor yang menentukan keberhasilan sebuah proses

pembelajaran. Untuk itu dirasa perlu melakukan pengelompokan data mahasiswa (*clustering*) berdasarkan gaya belajar, agar memudahkan dosen untuk menentukan metode pembelajaran yang cocok dengan kondisi mahasiswanya sehingga dapat meningkatkan hasil belajar mereka (Handayani, 2022).

Penerimaan mahasiswa baru menghasilkan banyak data tentang profil mahasiswa. Data yang besar sangat penting untuk memahami kebutuhan dan perilaku mahasiswa. Namun, jumlah data yang sangat besar dan beragam ini tidak mungkin diolah secara manual. Oleh sebab itu dibutuhkan sebuah teknik pengelompokan data atau *clustering*. *Clustering* data mahasiswa adalah sangat penting dalam pengerjaan data mining, sebab dapat mengungkap struktur tersembunyi dalam data mahasiswa tersebut (Asril et al, 2015). Dari beberapa algoritma *clustering*, K-Means merupakan salah satu algoritma paling efektif dalam segmentasi data numerik yang kompleks (Ikotun et al, 2023). K-Means merupakan suatu algoritma *clustering* yang bersifat iteratif, yang bekerja dengan melakukan pengelompokan secara bertahap. Prosesnya dimulai dengan menentukan secara acak atau teracak nilai K, yang mewakili jumlah cluster yang akan terbentuk. Selanjutnya, algoritma ini secara acak menetapkan nilai awal untuk setiap pusat cluster, yang sering disebut sebagai centroid, mean, atau "mean". Metode ini bertujuan untuk membagi data menjadi beberapa kelompok atau cluster (Afidah et al, 2023).

Penelitian ini bertujuan untuk menerapkan algoritma K-Means guna mengidentifikasi kelompok karakteristik mahasiswa baru di Unilak. Hasil *clustering* ini kemudian digunakan sebagai dasar perencanaan sistem pembelajaran berbasis teknologi pendidikan yang adaptif dan personal. Perencanaan sistem pembelajaran digital yang mampu beradaptasi otomatis dengan kebutuhan mahasiswa dan memberikan pengalaman belajar yang personal, agar setiap mahasiswa mendapat dukungan belajar sesuai dengan kondisi dan potensinya masing-masing.

2. METODE

Jenis penelitian ini adalah kuantitatif deskriptif dengan pendekatan data mining. Dimana fokus penelitian adalah penerapan algoritma *K-Means* untuk klasifikasi mahasiswa baru berdasarkan dua variabel utama yaitu data akademik (nilai rata-rata rapor) dan kemampuan literasi digital. Populasi dalam penelitian ini adalah mahasiswa baru tahun akademik 2024/2025 di Unilak. Subjek penelitian yang diambil adalah 150 mahasiswa baru yang dipilih secara purposive berdasarkan kelengkapan administratif dan angket.

Data dikumpulkan melalui sistem informasi penerimaan mahasiswa baru (SMART Unilak) dan kuisioner kemampuan literasi digital. Variabel awal yang dianalisis adalah nilai rata-rata rapor, asal sekolah (SMK/SMA, asal daerah), minat belajar terhadap teknologi, serta melakukan sebaran angket sebagai pretest yaitu mengetahui tingkat literasi digital mahasiswa.

Adapun langkah analisis dimulai dengan *preprocessing* (normalisasi data) yaitu tahap pembersihan data, penentuan jumlah klaster (metode elbow), implementasi *K-Means* menggunakan python serta melakukan visualisasi klaster dan validasi internal.

Penerapan algoritma *K-Means* menggunakan bahasa pemrograman *Python*, dengan *library* seperti:

- Pandas*, untuk pengelolaan dan manipulasi dataset,
- Numpy*, untuk operasi numerik dan perhitungan array,
- Sklearn*, untuk fungsi *K-Means*,
- StandardScaler* untuk fungsi normalisasi, *matplotlib* dan *seaborn* untuk visualisasi hasil *clustering*.

3. HASIL DAN PEMBAHASAN

A. Preprocessing

Preprocessing adalah tahap awal yang sangat penting untuk mempersiapkan data mentah agar layak digunakan dalam proses pengelompokan data (*clustering*). Tujuannya adalah agar data bersih, terstruktur, konsisten dan siap untuk dihitung oleh algoritma *K-Means*. Diawali dengan pembersihan data (*cleaning*) yakni menghapus atau mengisi data yang hilang (*missing value*), menghapus duplikasi, menangani *outlier* (nilai ekstrem yang bisa mengganggu perhitungan range). Selain itu dilakukan transformasi data, dengan mengubah kategori data (misal SMK dan SMA menjadi angka). Selanjutnya adalah memilih variabel atau fitur yang relevan untuk *clustering* (hanya 2 variabel yaitu nilai rata-rata rapor dan data literasi digital), lalu melakukan normalisasi atau standarisasi data yang memiliki skala yang sama, menggunakan *StandardScaler* dari *sklearn.preprocessing* agar variabel mempunyai skala yang seimbang. Data tersebut nantinya akan mengarahkan ke penerapan teknologi pendidikan berupa pemanfaatan analisis data untuk pembelajaran, rekomendasi pengembangan sistem pembelajaran,

integrasi aplikasi *AI (Artificial Intelligent)* dalam desain instruksional pembelajaran serta meningkatkan kualitas penerapan teknologi pembelajaran melalui pemetaan kebutuhan mahasiswa.

B. Penentuan Jumlah Klaster

Berdasarkan metode elbow, didapati output = jumlah klaster optimal yaitu 3. Dimulai dengan menjalankan algoritma *K-Means* untuk berbagai nilai *k* (misal dari 1 sampai 10). Untuk setiap *k* tersebut, dihitung nilai *SSE (Sum of Squared Errors)*.

$$SSE = \sum_{i=1}^K \sum_{x \in C_k} ||x - \mu_i||^2$$

Dimana :
 C_i = klaster ke *i*
 μ_i = pusat (centroid) klaster ke-*i*
 x = data dalam klaster tersebut

Hasil clustering menunjukkan pembagian mahasiswa menjadi tiga klaster dominan yaitu klaster 1; nilai tinggi dengan literasi digital tinggi, klaster 2; nilai sedang dengan literasi digital sedang, klaster 3; nilai rendah dengan literasi digital rendah.

C. Implementasi dan Visualisasi Cluster

Setelah mendapatkan jumlah klaster = 3, model *K-Means* di inialisasi dan dilatih menggunakan *coding python* sebagai berikut:

```
from sklearn.cluster import KMeans
model = KMeans(n_clusters=3, random_state=42)
model.fit(data_scaled)
```

Selanjutnya hasil *cluster* disimpan dan divisualisasikan dalam bentuk *scatter plot* dua dimensi, dengan *coding python* sebagai berikut:

```
import matplotlib.pyplot as plt
plt.scatter(data_scaled[:,0], data_scaled[:,1], c=model.labels_, cmap='viridis')
plt.title("Hasil Clustering Mahasiswa Baru")
plt.xlabel("Nilai Rapor")
plt.ylabel("Literasi Digital")
plt.show()
```

Validasi internal dilakukan dengan menguji kualitas klaster menggunakan *Silhouette Score*:

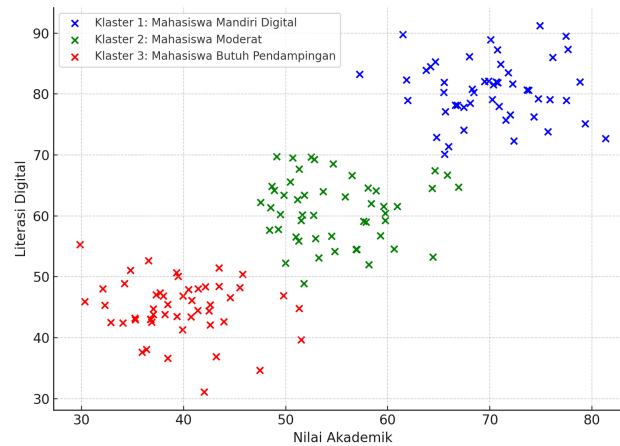
```
from sklearn.metrics import silhouette_score
score = silhouette_score(data_scaled, model.labels_)
print("Silhouette Score:", score)
```

Dengan hasil klaster yang terbentuk, mahasiswa baru dikelompokkan menjadi tiga tipe, yaitu nilai akademik dan literasi digital tinggi, nilai akademik dan literasi digital sedang, serta nilai akademik dan literasi digital rendah, Berikut hasil klaster dalam bentuk tabel:

Tabel 1. Hasil clustering

Klaster	Karakteristik Utama	Jumlah Mahasiswa
1	Nilai tinggi, literasi digital tinggi	54
2	Nilai sedang, literasi digital sedang	52
3	Nilai rendah, literasi digital rendah	44

Berikut adalah grafik hasil *clustering* data mahasiswa baru Unilak dengan tiga kluster :



Gambar 1. *Clustering* data mahasiswa baru Unilak dengan tiga kluster

Berdasarkan gambar di atas, dapat disimpulkan bahwa distribusi mahasiswa baru Unilak berdasarkan dua variable yaitu nilai rata-rata rapor (akademik) dan literasi digital adalah sebagai berikut:

1. Klaster 1 (warna biru) : Merupakan mahasiswa mandiri digital. Artinya mahasiswa yang mampu mengikuti pembelajaran daring secara efisien.
2. Klaster 2 (warna hijau) : Merupakan mahasiswa moderat. Artinya mahasiswa yang memerlukan penguatan strategi blended learning.
3. Klaster 3 (warna merah) : Merupakan mahasiswa yang membutuhkan pembimbingan. Artinya mahasiswa yang harus diberikan pelatihan TIK dasar serta bimbingan tambahan.

Dengan adanya klasterisasi diatas maka dilakukan integrasi ke sistem pembelajaran berbasis teknologi pendidikan, dengan penjelasan :

1. Klaster tinggi (Nilai akademik dan literasi digital tinggi)
Diberikan rekomendasi model pembelajaran berbasis proyek maupun *problem based learning* dengan pemberian *e-modul* (modul digital). Pada klaster ini, mahasiswa siap menerima model pembelajaran berbasis digital, penerapan teknologi dan interaktif. Memaksimalkan aplikasi *Edlink* dan SMART Unilak yang sudah berjalan selama beberapa tahun ini.
2. Klaster sedang (Nilai akademik dan literasi digital sedang)
Diberikan rekomendasi model pembelajaran adaptif berbasis *platform hybrid* seperti *edlink* dan pembelajaran offline di kelas. Ini merupakan intervensi atau dukungan berbasis digital ringan karena ada tutorialnya. Selain itu perlu dilakukan evaluasi mingguan guna melihat perkembangan mahasiswa agar kompetensi dapat ditingkatkan.
3. Klaster rendah (Nilai akademik dan literasi digital rendah)

Diberikan rekomendasi penguatan literasi digital dengan model pembelajaran berbasis modul atau panduan lengkap, dimana modul tersebut dimonitor langsung oleh dosen maupun asisten dosen. Mahasiswa klaster ini membutuhkan pembelajaran menggunakan video interaktif, pengulangan materi dan pengawasan.

Berdasarkan profil mahasiswa baru Unilak dengan basis *clustering K-Means*, maka perencanaan sistem pembelajaran mencakup penetapan tujuan belajar, desain kurikulum dan materi, rekomendasi metode belajar serta evaluasi pembelajaran. Alat dan platform digital yang dapat dimaksimalkan seperti *LMS (Learning Management System)*, *edlink* Unilak, *moodle*, *google classroom*. Dan dalam kelas-kelas pembelajaran menggunakan media interaktif seperti video pembelajaran, gaming, simulasi maupun *AR/VR* yang dibuat oleh dosen.

Hasil penelitian ini sejalan dengan penelitian [Handayani \(2022\)](#), bahwa pengelompokan mahasiswa berdasarkan gaya belajar lebih efisien dan efektif. Hal ini dapat dilihat melalui kecepatan dan ketepatan waktu dalam *clustering* mahasiswa dan memudahkan para dosen untuk menentukan

metode pembelajaran yang sesuai dengan gaya belajar mahasiswa tersebut, sehingga akhirnya dapat meningkatkan hasil belajar mahasiswa.

4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma *K-Means* efektif digunakan untuk mengidentifikasi pola karakteristik mahasiswa baru Unilak berdasarkan data nilai akademik rapor dan literasi digital. Hasil *clustering* dengan tiga kelompok utama yaitu mahasiswa mandiri digital yang mampu mengikuti pembelajaran daring secara efisien, mahasiswa moderat yang memerlukan penguatan strategi *blended learning*, serta mahasiswa yang membutuhkan pembimbingan dengan pelatihan TIK dasar dan bimbingan tambahan lainnya. Hal ini telah memberikan gambaran yang sangat jelas tentang kebutuhan pembelajaran mahasiswa baru serta rekomendasi terhadap strategi pembelajaran berbasis teknologi pendidikan di Unilak, yang dapat digunakan oleh civitas akademika sebagai landasan dalam merancang sistem pembelajaran berbasis teknologi yang adaptif dan responsif terhadap keberagaman mahasiswa barunya, yang di desain berdasarkan data profil mahasiswa.

DAFTAR PUSTAKA

- Afidah., Nayla Nur, Masrukan. (2023). Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang. *Prosiding Seminar Nasional Matematika*, Prisma 6, 729-738. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Asril., Elvira, Wiza., Fana, YuneFri., Yogi. (2015). Analisis Data Lulusan dengan Data Mining untuk Mendukung Strategi Promosi Universitas Lancang Kuning. *Jurnal Teknologi Informasi & Komunikasi Digital Zone*, 6(2),24-32. <https://journal.unilak.ac.id/index.php/dz/article/view/94/30>
- AyuPratiwi., Sekar, Ginting., Subhan Hafiz Nanda, Imam., Chairul. (2024). Penerapan Algoritma k-means dalam Data Mining untuk Mengidentifikasi Strategi Promosi di Politeknik Ganesha Medan. *Jurnal Minfo Polgan*, 13(1), 189–196. <https://doi.org/10.33395/jmp.v13i1.13509>
- Handayani. (2022). Aplikasi Data Mining Menggunakan Algoritma K-Means untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar. *Jurnal Teknologi dan Informasi*, 12(1), 46–63. <https://ojs.unikom.ac.id/index.php/jati/article/view/6733>
- Ikotun., Abiodun M, Ezugwu., Absalom E, Abualigah., Laith, Abuhaija., Belal, Heming., Jia. (2023): K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178-210. <https://doi.org/10.1016/j.ins.2022.11.139>
- Jasin., Harteti. (2021). Implementasi Guru Terhadap Model Pembelajaran Daring dimasa Pandemi Covid-19 di SDN 4 Ponelo Kepulauan. *Prosiding Seminar Nasional Pendidikan Dasar*, 63-71. <https://ejurnal.pps.ung.ac.id/index.php/PSNPD/article/view/1052>
- Okgistiyan., Muchammad, Irmayansyah. (2025). Penerapan Metode K-Means Clustering Untuk Pemetaan Kompetensi Peserta Didik Baru Berdasarkan Hasil Tes Diagnostik. *Jurnal Sains dan Teknologi Komputasi*, 1(1), 80–89. <https://doi.org/10.36350/jskom.v1i1.12>
- Maukar., Anastasia L, Marisa., Fitri, Widodo., Anang Aris. (2022). Analisis Data Penerimaan Mahasiswa Baru Berbasis K-Means. *Jurnal Informatika Dan Komputer*, 6(2), 142-147. <https://doi.org/10.26798/jiko.v6i2.558>
- Melizah, Susilo., Andri Anto Tri. (2024). Implementasi Algoritma K-Means Clustering untuk Analisis Data Nilai Akademik Mahasiswa. *Jurnal Teknologi Informasi Mura*, 16(2), 25–43. <https://doi.org/10.32767/jti.v16i2.2427>
- Norshahlan., Muhammad, Jaya., Hendra, Kustini., Rini. (2023). Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Data Calon Siswa Baru. *Jurnal Sistem Informasi Triguna Dharma TGD*, 2(6), 1042-1053. <https://doi.org/10.53513/jursi.v2i6.9148>
- Tensao., Oki Oktaviarna, Wijaya., I Nyoman Yudi Anggara, Fredlina., Ketut Queen. (2022). Analisa Data Mining dengan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Pada STMIK Primakara. *Jurnal Informatika Dan Sistem Informasi*, 14(1), 1–17. <https://doi.org/10.37424/informasi.v14i1.135>