



Penerapan K-Means Clustering Untuk Pengelompokan Kompetensi Siswa PKL Di LKP Utama Jaya

Application of K-Means Clustering for Grouping the Competency Profiles of Internship Students at LKP Utama Jaya

Reydo Trisno Pangestu & Rizki Muliono

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Medan Area, Indonesia

*Corresponding Email: 1reyvpangestu48@gmail.com, 2rizkimuliono@staff.uma.ac.id

Abstrak

Pengelolaan data siswa pada lembaga kursus nonformal sering kali masih terbatas pada pencatatan nilai tanpa dimanfaatkan lebih lanjut untuk mengidentifikasi pola kompetensi siswa. Penelitian ini menerapkan algoritma *K-Means Clustering* untuk mengelompokkan profil kompetensi siswa Praktik Kerja Lapangan (PKL) di LKP Utama Jaya, Kabupaten Deli Serdang, berdasarkan empat atribut, yaitu skor teori, skor praktik, tingkat kehadiran, dan total durasi pembelajaran. Data yang dianalisis terdiri atas 113 rekam data siswa dari dua program, yaitu Pemrograman Web serta Teknik Komputer dan Jaringan, yang dihimpun selama tahun 2026. Penentuan jumlah kluster dilakukan menggunakan metode *Elbow*, sedangkan kualitas hasil pengelompokan dievaluasi menggunakan *Silhouette Score*. Hasil penelitian menunjukkan bahwa tiga kluster merupakan jumlah kelompok yang digunakan dalam analisis, dengan 70 siswa (61,95%) berada pada kelompok berkinerja tinggi, 36 siswa (31,86%) pada kelompok berkinerja sedang, dan 7 siswa (6,19%) pada kelompok berkinerja rendah. Hasil evaluasi menghasilkan *Silhouette Score* sebesar 0,3780. Karakteristik kelompok menunjukkan adanya perbedaan pada beberapa atribut yang digunakan, dengan tingkat kehadiran menjadi salah satu atribut yang terlihat membedakan profil antarkluster. Hasil pengelompokan ini dapat digunakan sebagai dasar bagi LKP Utama Jaya dalam menyusun strategi pembinaan siswa yang disesuaikan dengan karakteristik masing-masing kelompok.

Kata Kunci: *K-Means Clustering*; Kompetensi Siswa; LKP; Praktik Kerja Lapangan; *Data Mining*

Abstract

Student data management in non-formal training institutions is often limited to recording academic results without further analysis to identify competency patterns. This study applies the K-Means Clustering algorithm to group the competency profiles of internship (PKL) students at LKP Utama Jaya, Deli Serdang, based on four attributes: theory score, practical score, attendance rate, and total learning hours. The dataset consists of 113 student records from two programs, namely Web Programming and Computer and Network Engineering, collected during 2026. The number of clusters was determined using the Elbow method, while the quality of the clustering results was evaluated using the Silhouette Score. The results indicate that three clusters were used in the analysis, consisting of 70 students (61.95%) in the high-performing group, 36 students (31.86%) in the moderate-performing group, and 7 students (6.19%) in the low-performing group. The clustering evaluation produced a Silhouette Score of 0.3780. The characteristics of the clusters indicate differences across several attributes, with attendance rate appearing as one of the attributes that distinguishes the student profiles between clusters. These clustering results can serve as a basis for LKP Utama Jaya to develop differentiated student development strategies according to the characteristics of each group.

Keywords: *K-Means Clustering*; Student Competency; LKP; Internship; *Data Mining*



PENDAHULUAN

Catatan administrasi di lembaga kursus nonformal umumnya berakhir di laci arsip tanpa pernah benar-benar dianalisis. Nilai ujian teori dan praktik memang rutin direkap, namun data kehadiran dan durasi keterlibatan siswa jarang diolah lebih lanjut. Akibatnya, manajer lembaga hanya melihat 'siapa yang lulus' dan 'siapa yang tidak', tanpa memahami mengapa perbedaan itu terjadi. Program pembinaan yang dirancang pun akhirnya seragam—padahal kebutuhan setiap siswa berbeda.

Situasi semacam itu terjadi pula di LKP Utama Jaya—unit usaha PT. Brata Mitra Nusantara yang berfokus pada pendidikan nonformal teknologi informasi di Kabupaten Deli Serdang. Dua program kursus dikelola di sini: Pemrograman Web dan Teknik Komputer dan Jaringan (TKJ). Tiap tahun, puluhan siswa dikirim untuk menjalani Praktik Kerja Lapangan (PKL). Seluruh rekam jejak mereka—absensi, nilai, durasi belajar—tersimpan di sistem administrasi, namun belum digunakan untuk merancang pembinaan yang lebih tepat.

Di sinilah pembelajaran mesin tanpa pengawasan menawarkan nilai tambah yang nyata. Di antara berbagai metode yang tersedia, K-Means unggul karena beban komputasinya ringan sekaligus hasilnya mudah diinterpretasikan oleh praktisi lapangan. Sejumlah peneliti telah memvalidasi keunggulan ini pada data pendidikan vokasional: Fauji & Farokhah menerapkannya pada nilai rapor siswa SMA [1], Qusyairi et al. mengoptimalkannya dengan metode Elbow [2], dan Faisal et al. secara khusus menggunakannya untuk mengelompokkan kompetensi siswa SMK bidang TIK beserta implikasinya terhadap strategi pedagogis [3].

Meski demikian, mayoritas kajian yang ada masih bertumpu pada nilai akademik sebagai satu-satunya variabel pengelompokan. Dimensi kehadiran dan intensitas durasi belajar nyaris tidak pernah diikutsertakan, padahal keduanya memiliki daya prediktif yang setara bahkan melampaui nilai ujian. Penelitian ini hadir untuk mengisi gap tersebut: membangun model pengelompokan yang merangkum empat dimensi kompetensi sekaligus—nilai teori, nilai praktik, kehadiran, dan lama kursus—dalam satu kerangka analisis terintegrasi guna menghasilkan rekomendasi pembinaan yang lebih proporsional bagi setiap kelompok siswa di LKP Utama Jaya.

Penambahan data adalah proses mengangkat pengetahuan tersembunyi melalui serangkaian langkah komputasi. Ketika diterapkan di lingkungan pendidikan,

pendekatan ini membuka ruang bagi pengelola lembaga untuk memahami karakteristik siswanya secara kuantitatif dan sistematis [4]. Berbagai studi terbaru menunjukkan keberhasilan penerapannya: dari clustering prestasi di madrasah [5], pengelompokan indeks kesulitan mata pelajaran [1], pemetaan tingkat penyelesaian pendidikan di 34 provinsi Indonesia [6], hingga segmentasi calon penerima beasiswa di academy korporasi [7]. Sutanto & Susilawaty juga membuktikan efektivitasnya pada level sekolah menengah pertama di tingkat kabupaten [8], menunjukkan bahwa metode ini skalabel dari institusi kecil hingga besar.

Algoritma K-Means [9] mengelompokkan data melalui siklus empat langkah: inisialisasi pusat gugus, penugasan setiap titik ke pusat terdekat, pembaruan posisi pusat, dan pengulangan hingga konvergen. Kedekatan diukur lewat jarak Euclidean:

$$d(x, c) = \sqrt{\sum (x_i - c_i)^2} \dots\dots\dots (1)$$

Kemudahan implementasi dan hasil yang interpretatif menjadikan K-Means pilihan dominan dalam riset pengelompokan data pendidikan [10], [11]. Penerapannya telah mencakup personalisasi pengajaran berbasis klaster di kelas [12], pemetaan pola performa akademik dan faktor sosial siswa secara bersamaan [13], serta perbandingan performa dengan metode lain seperti K-Medoids yang dilakukan oleh Khairuna et al. [14]—studi tersebut menunjukkan bahwa K-Means menghasilkan klaster yang lebih homogen pada data teks ulasan, mendukung penggunaannya dalam konteks data akademik yang relatif seragam seperti pada penelitian ini.

Kurva Elbow membantu menentukan jumlah klaster optimal: jumlahkan kuadrat jarak semua titik data ke pusat gugusan untuk berbagai nilai k, lalu cari titik di mana penambahan gugus tidak lagi memberikan penurunan jarak yang signifikan [2]. Indeks Silhouette [15] mengukur kualitas pemisahan—nilai mendekati 1 menandai gugus yang bersih, nilai mendekati 0 mengisyaratkan ambiguitas, dan nilai negatif mengindikasikan salah penempatan. Qadir et al. menjadikan kombinasi kedua metrik ini sebagai standar evaluasi dalam monitoring keberhasilan akademik berbasis data [16].

METODE PENELITIAN

A. Sumber dan Karakteristik Data

Seluruh data bersumber dari catatan administrasi LKP Utama Jaya tahun 2026. Total terdapat 113 siswa: 56 di bidang Pemrograman Web dan 57 di bidang TKJ. Satu pertemuan kursus setara tiga jam pembelajaran efektif, jadwal reguler Senin–Jumat.

Variabel lama kursus dihitung dari kehadiran aktual dikalikan tiga jam per sesi, sehingga mencerminkan keterlibatan nyata siswa. Rincian variabel disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Deskripsi	Rentang	Skala
Nilai Teori	Skor ujian tertulis siswa	0–100	Rasio
Nilai Praktik	Skor ujian praktik siswa	0–100	Rasio
Kehadiran	Persentase kehadiran terhadap total pertemuan	0–100%	Rasio
Lama Kursus	Total jam pembelajaran aktual yang diikuti	Jam	Rasio

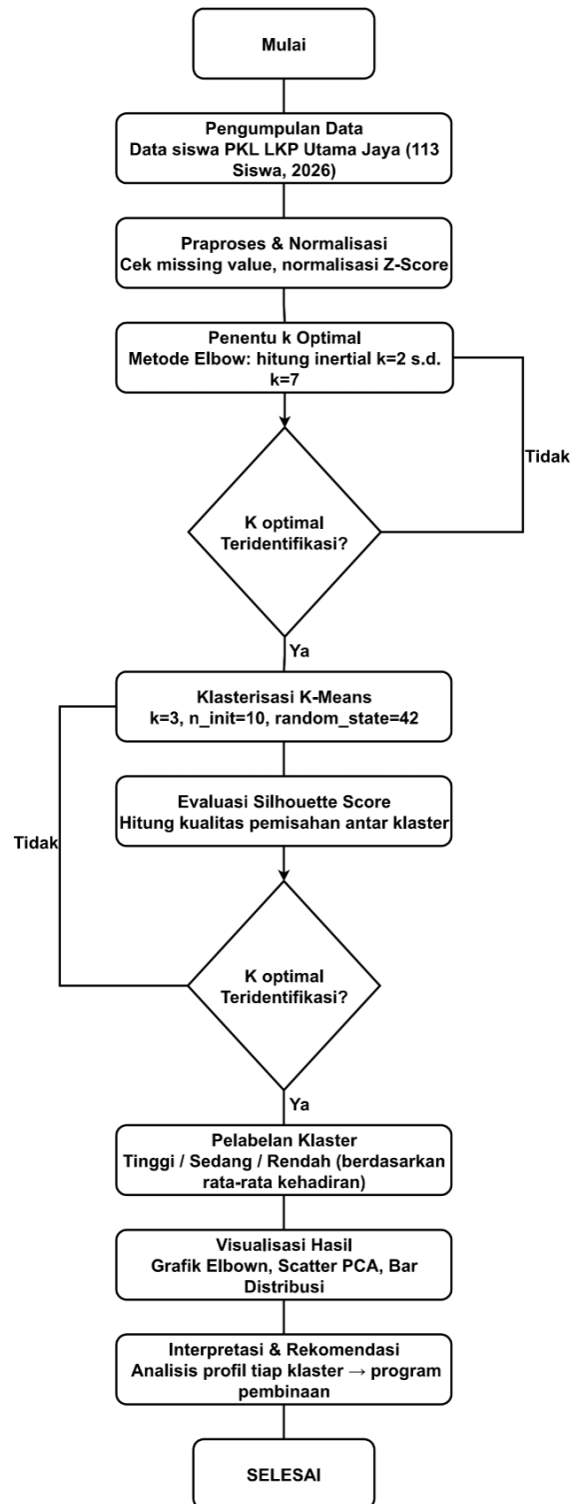
Cuplikan data ditampilkan pada Tabel 2. Rentang kehadiran yang sangat lebar—dari 33% hingga 100%—menunjukkan variabilitas keterlibatan siswa yang tinggi, dan inilah yang membuat variabel kehadiran menjadi dimensi pembeda paling informatif dalam proses klusterisasi.

Tabel 2. Cuplikan Data Siswa PKL LKP Utama Jaya Tahun 2026

No	Nama Siswa	N.Teori	N.Praktik	Hadir(%)	Lama(j)
1	Dhia Syarafana Surya	90,50	91,20	100,00	156
2	Andika Aria Admaza	89,88	91,20	98,08	153
3	Anindya Deliani Nasution	89,25	91,20	98,08	153
...	...	...	...	...	...
111	Tria Puji Astari	85,00	85,30	33,00	57
112	Ramadani	85,00	85,50	35,00	60
113	FAHMI REZA	85,00	85,70	58,33	63

B. Alur Penelitian

Enam tahap dirangkai secara berurutan: (1) ekstraksi data administrasi LKP, (2) praproses dan normalisasi Z-Score, (3) penentuan k optimal melalui kurva Elbow, (4) eksekusi K-Means dengan k=3, (5) evaluasi kualitas klaster menggunakan Silhouette Score, dan (6) interpretasi serta penyusunan rekomendasi. Bagan alur lengkap ditampilkan pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian K-Means Clustering

C. Normalisasi Z-Score

Nilai ujian berada pada rentang 0–100, sementara durasi kursus bisa melampaui 150 jam. Perbedaan skala ini akan mendistorsi perhitungan jarak Euclidean jika dibiarkan. Standarisasi Z-Score meratakan semua distribusi ke skala yang setara:

$$z = (x - \mu) / \sigma \dots\dots\dots (2)$$

dengan x nilai asli, μ rata-rata populasi, dan σ standar deviasinya. Nilai rata-rata dan standar deviasi dari 113 data adalah: Nilai Teori ($\mu=87,64$; $\sigma=1,15$), Nilai Praktik ($\mu=88,59$; $\sigma=1,30$), Kehadiran ($\mu=89,23\%$; $\sigma=13,28$), dan Lama Kursus ($\mu=135,82$ jam; $\sigma=26,68$).

D. Contoh Perhitungan Manual

Untuk memperjelas proses pengelompokan menggunakan algoritma *K-Means Clustering*, berikut disajikan contoh perhitungan manual menggunakan tiga data siswa yang mewakili masing-masing klaster, yaitu Andika Aria Admaza sebagai contoh Klaster Tinggi, Almira Andarista sebagai contoh Klaster Sedang, dan Fahrizal Purba sebagai contoh Klaster Rendah.

Langkah 1 — Normalisasi Z-Score. Tiga siswa dipilih sebagai contoh: Andika Aria Admaza (Tinggi), Almira Andarista (Sedang), dan Fahrizal Purba (Rendah). Perhitungan Z-Score untuk variabel Nilai Teori siswa Andika:

$$z(\text{Teori}) = (89,88 - 87,64) / 1,15 = 2,24 / 1,15 \approx 1,95$$

Hasil normalisasi keempat variabel untuk ketiga siswa disajikan pada Tabel 3.

Tabel 3. Hasil Normalisasi Z-Score Tiga Siswa Contoh

Nama Siswa (Klaster)	$z(\text{Teori})$	$z(\text{Praktik})$	$z(\text{Hadir})$	$z(\text{Lama})$
Andika Aria Admaza (Tinggi)	1,95	2,01	0,67	0,64
Almira Andarista (Sedang)	-1,09	-0,00	0,09	0,19
Fahrizal Purba (Rendah)	-1,95	-1,91	-1,97	-1,04

Langkah 2 — Centroid akhir, dalam skala ternormalisasi disajikan pada Tabel 4.

Tabel 4. Centroid Akhir dalam Skala Ternormalisasi

Klaster	$C(\text{Teori})$	$C(\text{Praktik})$	$C(\text{Hadir})$	$C(\text{Lama})$
C1 – Tinggi	0,54	0,58	0,47	0,41
C2 – Sedang	-0,63	-0,70	-0,36	-0,42
C3 – Rendah	-2,15	-2,15	-2,84	-1,99

Langkah 3 — Perhitungan jarak Euclidean siswa Andika ke Centroid C1 (Tinggi):

$$d(\text{Andika}, C1) = \sqrt{[(1,95-0,54)^2 + (2,01-0,58)^2 + (0,67-0,47)^2 + (0,64-0,41)^2]} \\ = \sqrt{[1,98 + 2,04 + 0,04 + 0,05]} = \sqrt{4,11} \approx 2,03$$

Berdasarkan hasil tersebut, jarak Andika terhadap C1, C2, dan C3 masing-masing adalah **2,03; 4,02; dan 7,30**. Karena jarak terkecil adalah terhadap C1, maka Andika ditempatkan pada **Klaster C1 (Tinggi)**.



Perhitungan serupa dilakukan terhadap Almira dan Fahrizal. Hasil perhitungan jarak ketiga siswa terhadap masing-masing centroid disajikan pada Tabel 5.

Tabel 5. Jarak Euclidean Tiga Siswa Contoh ke Masing-Masing Centroid

Nama Siswa	$d(x, C1\text{-Tinggi})$	$d(x, C2\text{-Sedang})$	$d(x, C3\text{-Rendah})$
Andika Aria Admaza	2,03	4,02	7,30
Almira Andarista	1,78	1,13	4,37
Fahrizal Purba	4,52	2,49	1,33

Langkah 4 — Penugasan Kluster. Berdasarkan Tabel 5, setiap siswa ditempatkan pada kluster yang memiliki jarak Euclidean terkecil. Andika Aria Admaza memiliki jarak terkecil terhadap C1 sebesar 2,03, sehingga ditempatkan pada Kluster C1 (Tinggi). Almira Andarista memiliki jarak terkecil terhadap C2 sebesar 1,13, sehingga ditempatkan pada Kluster C2 (Sedang). Sementara itu, Fahrizal Purba memiliki jarak terkecil terhadap C3 sebesar 1,33, sehingga ditempatkan pada Kluster C3 (Rendah).

Dengan demikian, berdasarkan nilai Z-Score dan centroid yang disajikan pada Tabel 3 dan Tabel 4, hasil penugasan ketiga data contoh konsisten dengan kelompok yang ditunjukkan, yaitu Andika pada Kluster Tinggi, Almira pada Kluster Sedang, dan Fahrizal pada Kluster Rendah.

Perhitungan manual tersebut menunjukkan bahwa proses *K-Means Clustering* dilakukan dengan menghitung jarak setiap data terhadap seluruh centroid, kemudian menempatkan data pada kluster dengan jarak Euclidean terkecil.

E. Implementasi Komputasi

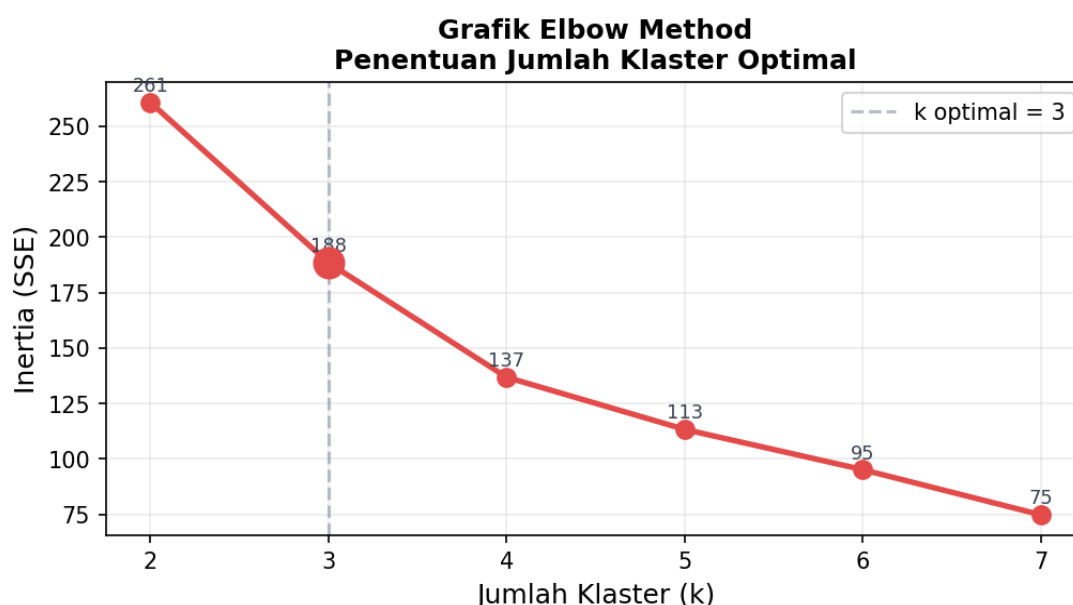
Seluruh proses dieksekusi menggunakan scikit-learn di Python dengan konfigurasi `n_clusters=3`, `random_state=42`, dan `n_init=10`. Parameter `n_init=10` memastikan algoritma memulai dari sepuluh titik inisialisasi berbeda; hasil inertia terkecil dipilih sebagai keluaran final untuk menghindari konvergensi ke optima lokal [11].

HASIL DAN PEMBAHASAN

A. Penentuan Jumlah Kluster Optimal

Kalkulasi inertia untuk $k=2$ hingga $k=7$ menghasilkan: 260,72; 188,46; 136,84; 113,32; 95,17; dan 74,82. Penurunan paling curam terjadi dari $k=2$ ke $k=3$ (selisih ± 72 poin). Setelah $k=3$, laju penurunan melambat secara konsisten membentuk sudut siku yang jelas, sebagaimana terlihat pada Gambar 2. Temuan ini sejalan dengan penelitian

terdahulu yang juga mendapatkan $k=3$ sebagai nilai optimal pada data prestasi siswa [2], [1].



Gambar 2. Grafik Elbow Method — Titik Siku pada $k=3$

B. Centroid dan Distribusi Kluster

Nilai centroid akhir masing-masing kluster dalam skala asli disajikan pada Tabel 6, dan distribusi anggotanya pada Tabel 7.

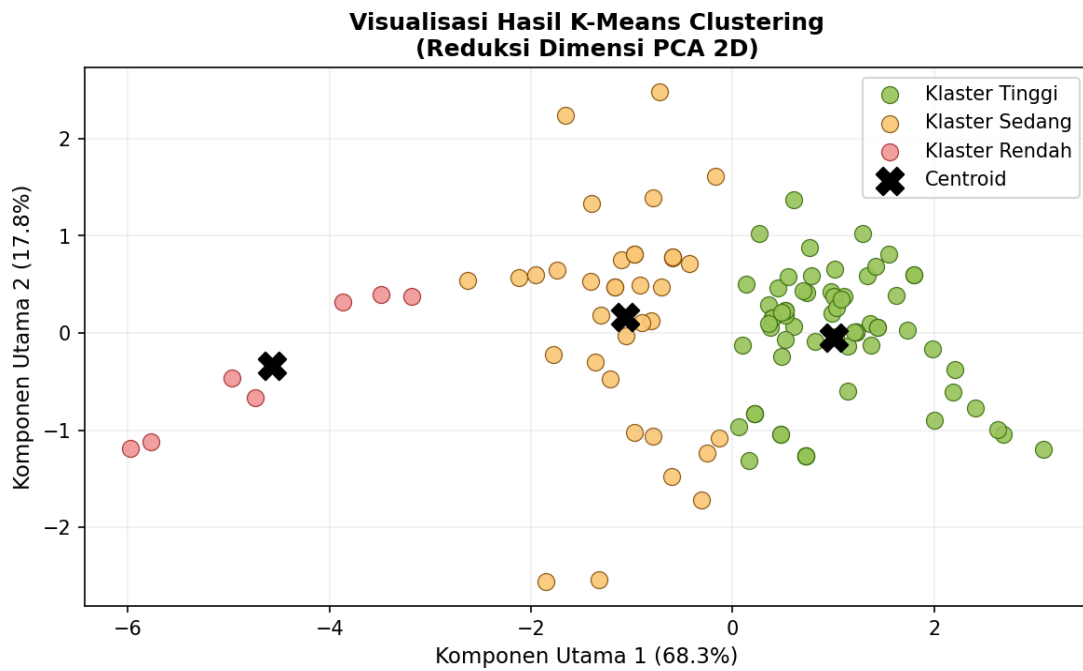
Tabel 6. Centroid Akhir Tiap Kluster (Skala Asli)

Kluster	N.Teori	N.Praktik	Kehadiran(%)	Lama(jam)
C1 – Tinggi	88,26	89,34	95,48	146,91
C2 – Sedang	86,91	87,68	84,43	124,58
C3 – Rendah	85,17	85,79	51,48	82,71

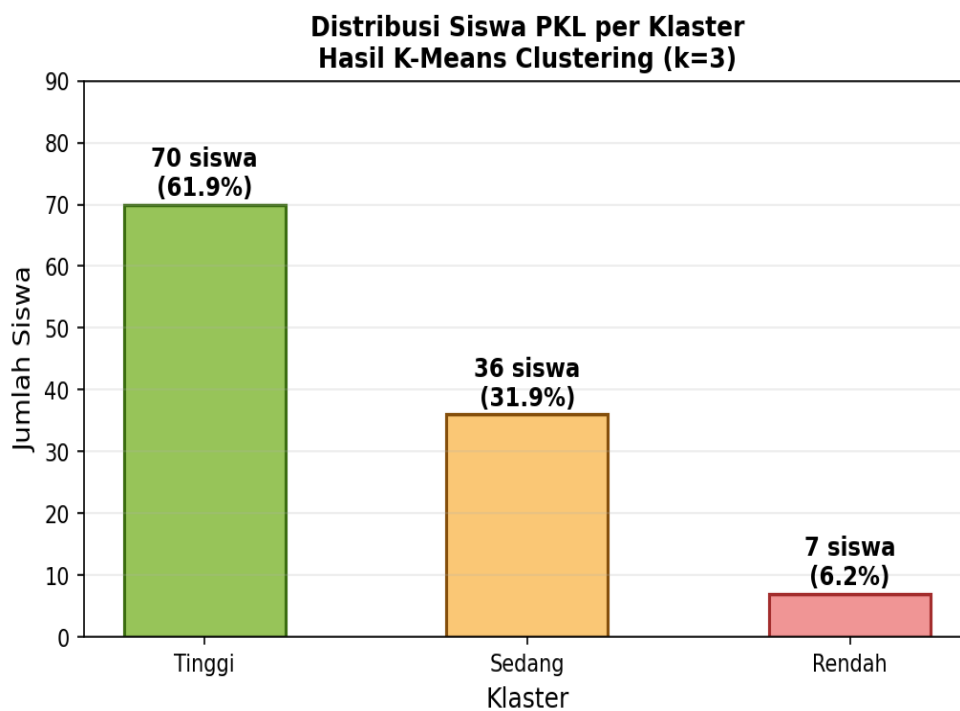
Tabel 7. Distribusi Siswa PKL per Kluster

Kluster	Label	Jumlah Siswa	Persentase(%)
C1	Tinggi	70	61,95
C2	Sedang	36	31,86
C3	Rendah	7	6,19
Total	–	113	100,00

Sebaran data dalam ruang dua dimensi hasil reduksi PCA serta distribusi anggota klaster divisualisasikan pada Gambar 3 dan Gambar 4.



Gambar 3. Visualisasi Hasil K-Means Clustering (Reduksi Dimensi PCA 2D)



Gambar 4. Distribusi Jumlah Siswa PKL per Klaster

C. Evaluasi Kualitas Klasterisasi

Silhouette Score yang diperoleh adalah 0,3780—berada pada kategori cukup bermakna. Nilai ini sebanding dengan hasil penelitian sejenis: skor 0,406 yang dilaporkan Sampurno et al. [17], hasil moderat serupa yang diperoleh Dila et al. pada data SMA [18], dan nilai Silhouette di kisaran yang sama pada klasterisasi performa mahasiswa oleh Ahmed et al. [19]. Pada Gambar 3 terlihat bahwa Klaster Rendah (kiri) paling terisolasi—konsisten dengan nilai kehadirannya yang jauh lebih rendah.

D. Profil dan Rekomendasi per Klaster

Klaster Tinggi (C1) menjadi klaster terbesar dengan 70 anggota (61,95%). Rata-rata kehadiran 95,48% dan lama kursus 146,91 jam mencerminkan komitmen belajar yang sangat tinggi. Ini konsisten dengan contoh perhitungan manual—Andika Aria Admaza dengan kehadiran 98,08% diklasifikasikan di sini dengan jarak terkecil 0,56 ke centroid C1. Rekomendasi: program pengayaan dan persiapan uji kompetensi industri, sejalan dengan saran Xie & Chen untuk kelompok berprestasi tinggi [12].

Klaster Sedang (C2) beranggotakan 36 siswa (31,86%). Nilai teori (86,91) dan praktik (87,68) tidak jauh berbeda dari Klaster Tinggi, namun kehadiran turun ke 84,43%. Penurunan intensitas kehadiran—bukan kemampuan akademik—yang mendorong masuk klaster ini, seperti ditunjukkan Almira Andarista dengan jarak 0,48 ke centroid C2. Intervensi paling tepat adalah pendampingan motivasional, bukan remediasi materi [13].

Klaster Rendah (C3) hanya beranggotakan 7 siswa (6,19%). Rata-rata kehadiran 51,48% berarti siswa hanya hadir separuh dari total sesi tersedia. Fahrizal Purba dalam contoh manual menunjukkan pola ini: kehadirannya 63% dan diklasifikasikan ke C3 dengan jarak 0,82. Pendekatan paling efektif adalah bimbingan individual dan fleksibilitas jadwal. Zhang & Liu menemukan pola serupa pada pendidikan vokasional: ketidakhadiran kronis lebih banyak dipicu faktor eksternal daripada rendahnya kemampuan akademik [20].

Secara keseluruhan, perbedaan antarklaster didominasi variabel kehadiran—selisih rata-rata antara Klaster Tinggi dan Rendah mencapai 44 poin persentase, jauh melampaui selisih nilai teori yang hanya tiga poin. Ini mempertegas bahwa manajemen kehadiran merupakan kunci peningkatan kualitas lulusan LKP Utama Jaya.

SIMPULAN

Penelitian ini menerapkan algoritma *K-Means Clustering* untuk mengelompokkan 113 siswa PKL di LKP Utama Jaya berdasarkan skor teori, skor praktik, tingkat kehadiran, dan total durasi pembelajaran. Berdasarkan hasil pengelompokan, diperoleh tiga klaster, yaitu 70 siswa (61,95%) pada kelompok berkinerja tinggi, 36 siswa (31,86%) pada kelompok berkinerja sedang, dan 7 siswa (6,19%) pada kelompok berkinerja rendah. Evaluasi menggunakan *Silhouette Score* menghasilkan nilai 0,3780, yang menunjukkan bahwa pemisahan antarklaster masih berada pada tingkat yang belum kuat sehingga hasil pengelompokan perlu diinterpretasikan secara hati-hati. Perhitungan jarak Euclidean pada data yang telah dinormalisasi menunjukkan bahwa setiap data ditempatkan pada klaster berdasarkan jarak terdekat terhadap centroid.

Hasil karakteristik klaster menunjukkan adanya perbedaan pada atribut yang digunakan dalam proses pengelompokan, termasuk tingkat kehadiran. Namun, berdasarkan analisis *K-Means* dalam penelitian ini, tingkat kehadiran belum dapat dinyatakan sebagai faktor penyebab atau faktor yang paling menentukan perbedaan kinerja siswa. Oleh karena itu, hasil pengelompokan lebih tepat digunakan sebagai dasar identifikasi profil kelompok siswa dan sebagai masukan bagi LKP Utama Jaya dalam merancang strategi pembinaan yang disesuaikan dengan karakteristik masing-masing klaster.

Penelitian selanjutnya dapat melakukan perbandingan kinerja *K-Means* dengan metode pengelompokan lain, seperti *K-Medoids* sebagaimana dikaji oleh Khairuna et al. [14] dan *Hierarchical Clustering* sebagaimana dilakukan oleh Sampurno et al. [17]. Penelitian berikutnya juga dapat mempertimbangkan penambahan variabel non-akademik untuk memperoleh representasi profil siswa yang lebih komprehensif [13], serta menggunakan metrik evaluasi tambahan untuk memperoleh penilaian kualitas klaster yang lebih menyeluruh.

DAFTAR PUSTAKA

- [1] F. Fauji dan L. Farokhah, "Penerapan Data Mining Menggunakan K-Means Clustering dalam Mengelompokkan Tingkat Kesulitan Mata Pelajaran," *Journal of Information System Research (JOSH)*, vol. 6, no. 3, hlm. 1705–1714, 2025, doi: 10.47065/josh.v6i3.6959.
- [2] M. Qusyairi, Z. Hidayatullah, dan A. Sandi, "Penerapan K-Means Clustering dalam Pengelompokan Prestasi Siswa dengan Optimasi Metode Elbow," *Infotek: Jurnal Informatika dan Teknologi*, vol. 7, no. 2, hlm. 500–510, 2024.
- [3] M. Faisal, Nurdin, Fajriana, dan Z. Fitri, "Clustering Kompetensi Siswa SMK Bidang TIK Menggunakan Algoritma K-Means: Evaluasi dan Implikasi Pedagogis," *International Journal of Engineering Science and Information Technology*, vol. 5, no. 1, hlm. 45–53, 2025.

- [4] M. Sholeh dan D. Andayati, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Indeks Prestasi Akademik Mahasiswa," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 6, no. 1, hlm. 51–60, 2023.
- [5] R. Awalia, "Penerapan Metode K-Means Clustering untuk Pengelompokan Prestasi Siswa Menggunakan Orange Data Mining: Studi Kasus di MTs Muhammadiyah Tallo Makassar," *Mechatronics Journal in Professional and Entrepreneur*, vol. 6, no. 2, 2024.
- [6] Vefiliana, D. Triadi, dan A. Mansyur, "Analisis Tingkat Penyelesaian Pendidikan di 34 Provinsi Indonesia Menggunakan Algoritma K-Means Clustering," *Jurnal Aplikasi Sistem Informasi dan Elektronika*, vol. 7, no. 1, 2025.
- [7] J. Heikal, D. A. Pradina, Y. Kurniawati, dan A. S. Afwan, "Implementation of K-Means Clustering Algorithm for Segmentation of Prospective Students for Corporate Academy Scholarship," *Jurnal Media Akademik*, vol. 3, no. 6, 2025.
- [8] H. S. Sutanto dan I. Susilawaty, "Analisis Klasterisasi Data Sekolah SMP di Kabupaten Kotawaringin Timur Menggunakan Metode K-Means," *JURISTEKNI: Jurnal Sistem Informasi dan Teknologi Informasi*, vol. 5, no. 3, hlm. 298–309, 2023.
- [9] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," dalam *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, hlm. 281–297.
- [10] M. Ahmed, R. Seraj, dan S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics (Basel)*, vol. 9, no. 8, hlm. 1295, 2020.
- [11] K. P. Sinaga dan M.-S. Yang, "Unsupervised K-means Clustering Algorithm," *IEEE Access*, vol. 8, hlm. 80716–80727, 2020.
- [12] L. Xie dan J. Chen, "Intelligent Student Performance Clustering and Personalized Teaching Suggestions Based on K-Means," dalam *Proceedings of the 2024 International Conference on Digital Society and Artificial Intelligence (DSAI '24)*, ACM, 2024. doi: 10.1145/3677892.3677898.
- [13] F. Rahma dan S. Z. Ulfah, "Clustering Students Based on Academic Performance and Social Factors: An Unsupervised Learning Approach to Identify Student Patterns," *International Journal for Applied Information Management*, vol. 5, no. 3, hlm. 139–154, 2025.
- [14] Khairuna, Nurdin, dan Ar Razi, "Analisis Perbandingan Algoritma K-Means dan K-Medoids untuk Klusterisasi Teks Ulasan pada Aplikasi Cookpad," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, hlm. 445–458, 2025.
- [15] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *J. Comput. Appl. Math.*, vol. 20, hlm. 53–65, 1987.
- [16] R. Qadir, A. F. Meghji, dan N. A. Mahoto, "Clustering Student Performance: A Data-Driven Approach to Monitor Academic Success," *Journal of Computing and Biomedical Informatics*, vol. 10, no. 1, 2025.
- [17] H. Sampurno, M. N. Wafi, dan Nurdin, "Perbandingan Algoritma K-Means dan Hierarchical Clustering dalam Pengelompokan Prestasi Akademik Siswa," *FORMAT: Jurnal Ilmiah Teknik Informatika*, vol. 15, no. 1, hlm. 72–80, 2026.
- [18] R. Dila, S. Defit, dan S. Arlis, "Analisis Algoritma K-Means Clustering dalam Pengelompokan Prestasi Belajar Siswa Menengah Atas," *Bulletin of Computer Science Research*, vol. 5, no. 5, hlm. 1113–1119, 2025.
- [19] H. Ahmed, S. Alam, dan M. Rashid, "Student Performance Prediction Using Machine Learning Algorithms," *Applied Computational Intelligence and Soft Computing*, vol. 2024, hlm. 4067721, 2024, doi: 10.1155/2024/4067721.
- [20] W. Zhang dan T. Liu, "Student Behavior Patterns in Vocational Education Big Data Based on Clustering Algorithm," *Discover Artificial Intelligence*, vol. 5, no. 1, 2025, doi: 10.1007/s44163-025-00433-3.