Arus Jurnal Sains dan Teknologi (AJST)
Website: http://jurnal.ardenjaya.com/index.php/ajst
Email: jurnal.ajst@ardenjaya.com

Segmentasi Tipologi Pembelajar Berbasis Learning Analytics pada Jejak Digital LMS Menggunakan K-Means Clustering

INFO PENPULIS

Nurul Awaliyah J.
Universitas Muhammadiyah Makassar
nurul.awaliyah@student.unismuh.ac.id

Titin Wahyuni
Universitas Muhammadiyah Makassar
titinwahyuni@unismuh.ac.id

Muhyiddin AM Hayat
Universitas Muhammadiyah Makassar
muhyiddin@unismuh.ac.id

INFO ARTIKEL

ISSN: 3026-3603
Vol. 4, No. 2 Oktober 2026
http://jurnal.ardenjaya.com/index.php/ajst

© 2026 Arden Jaya Publisher All rights reserved

Saran Penulisan Referensi:

Awaliyah J., N., Wahyuni, T., Hayat, M. AM. (2026). Segmentasi Tipologi Pembelajar Berbasis Learning Analytics pada Jejak Digital LMS Menggunakan K-Means Clustering. *Arus Jurnal Sains dan Teknologi , 4* (2), 246-256.

Abstrak

Pemanfaatan learning management system (LMS) menghasilkan jejak digital berukuran besar, namun jejak tersebut masih jarang dianalisis secara sistematis untuk memahami variasi pola belajar mahasiswa. Penelitian ini bertujuan menerapkan K-Means Clustering guna membentuk tipologi pembelajar dari log audit LMS Akademik Program Studi Informatika Universitas Muhammadiyah Makassar sekaligus mendeskripsikan karakteristik setiap segmen yang terbentuk. Data penelitian berupa ekspor log audit 24–28 Juli 2026 sebanyak 50.000 baris permintaan HTTP; setelah penyaringan menurut peran pengguna tersisa 45.476 baris milik 275 mahasiswa. Path endpoint dinormalisasi dari 2.400 nilai unik menjadi 74 pola, lalu diagregasi menjadi 13 fitur deskriptif per-mahasiswa yang mencakup volume, tempo, ritme temporal, dan profil belajar. Seluruh fitur ditransformasi dengan QuantileTransformer untuk meredam kemencengan, kemudian dimodelkan dengan K-Means (n_init = 20) dan Agglomerative Clustering pada k = 2 sampai 8. Jumlah segmen ditetapkan memakai prinsip parsimoni, yaitu k terkecil yang silhouette-nya masih dalam toleransi 0,015 dari nilai terbaik dengan setiap segmen beranggota minimal 20 mahasiswa. Aturan tersebut memilih k = 3 dengan silhouette 0,3018, stabil terhadap inisialisasi (rerata Adjusted Rand Index 0,9949) dan sepadan dengan Agglomerative Clustering pada 85,8% keanggotaan. Ketiga tipologi yang terbentuk adalah Pengejar Asesmen (159 mahasiswa; 57,8%) dengan rasio akses asesmen terhadap baca materi 12,38 banding satu, Partisipasi Rendah (79 mahasiswa; 28,7%), dan Pembelajar Intensif Seimbang (37 mahasiswa; 13,5%) dengan rasio 2,11 banding satu. Uji Kruskal-Wallis menunjukkan perbedaan signifikan pada seluruh 13 fitur ($p < 0,001$). Uji silang dengan penanda deteksi anomali pada data yang sama memperlihatkan sebaran yang sangat tidak merata ($\chi^2 = 33,81$; $p < 0,001$): segmen mayoritas hanya menyumbang 5 dari 14 penanda. Temuan ini menunjukkan bahwa segmentasi mengungkap persoalan sistemik yang luput dari deteksi anomali dan dapat menjadi dasar perlakuan pembelajaran yang dibedakan menurut segmen.

Kata Kunci: Learning Analytics, K-Means Clustering, Tipologi Pembelajar, Log Audit LMS, Educational Data Mining

Abstract

The use of learning management systems (LMS) produces large volumes of digital traces, yet these traces are rarely analysed systematically to understand the variety of student learning patterns. This study applies K-Means Clustering to construct a learner typology from the audit log of the Academic LMS of the Informatics Study Program, Universitas Muhammadiyah Makassar, and to describe the characteristics of each resulting segment. The data are an audit log export covering 24–28 July 2026 containing 50,000 HTTP request records; after filtering by user role, 45,476 records belonging to 275 students remain. Endpoint paths were normalised from 2,400 unique values into 74 patterns and then aggregated into 13 descriptive per-student features covering volume, tempo, temporal rhythm, and learning profile. All features were transformed using a QuantileTransformer to dampen skewness and modelled with K-Means (n_init = 20) and Agglomerative Clustering for k = 2 to 8. The number of segments was determined by a parsimony rule, namely the smallest k whose silhouette remains within a 0.015 tolerance of the best value while every segment holds at least 20 students. The rule selected k = 3 with a silhouette of 0.3018, stable across initialisations (mean Adjusted Rand Index 0.9949) and agreeing with Agglomerative Clustering on 85.8% of memberships. The three typologies are Assessment Chasers (159 students; 57.8%) with an assessment-to-reading access ratio of 12.38 to one, Low Participation (79 students; 28.7%), and Balanced Intensive Learners (37 students; 13.5%) with a ratio of 2.11 to one. A Kruskal-Wallis test shows significant differences across all 13 features ($p < 0.001$). Cross-tabulation with anomaly-detection flags on the same data reveals a highly uneven distribution ($\chi^2 = 33.81$; $p < 0.001$): the majority segment accounts for only 5 of 14 flags. Segmentation therefore exposes a systemic issue that anomaly detection cannot capture, and provides a basis for differentiated instructional treatment.

Keywords: Learning Analytics, K-Means Clustering, Learner Typology, LMS Audit Log, Educational Data Mining

A. Pendahuluan

Program studi memerlukan pemahaman tentang cara mahasiswanya belajar sebelum dapat merancang materi, asesmen, dan intervensi yang tepat sasaran. Selama ini pemahaman tersebut lebih banyak bersandar pada kesan dosen di kelas atau pada nilai akhir yang baru tersedia setelah perkuliahan berakhir. Keduanya datang terlambat dan tidak memerinci bagaimana mahasiswa memperoleh nilai itu: apakah melalui pembacaan bahan ajar yang tekun, atau semata-mata melalui pengerjaan asesmen.

Digitalisasi perkuliahan sesungguhnya telah menyediakan bahan pemetaan yang jauh lebih objektif. Learning management system (LMS) merekam seluruh lalu lintas permintaan HTTP penggunanya sebagai log audit: satu baris untuk satu permintaan, lengkap dengan waktu, pengguna, jenis aksi, endpoint, dan sesi. Log semacam ini merekam perilaku belajar apa adanya tanpa membebani mahasiswa dengan instrumen tambahan. Sayangnya, sebagaimana lazim di perguruan tinggi Indonesia, data aktivitas semacam ini masih jarang dianalisis secara sistematis untuk memahami variasi pola belajar (Nurfaizah, 2026), padahal pemanfaatan learning analytics terbukti dapat menjadi dasar pengambilan keputusan manajerial berbasis bukti pada institusi pendidikan (Hakim, 2025).

Mengubah tumpukan log menjadi peta tipologi merupakan tugas deskriptif dalam kerangka Learning Analytics dan Educational Data Mining. Tugas ini berbeda secara paradigmatik dari deteksi anomali yang hanya menyoroti segelintir individu menyimpang (Mutmainah dan Yustanti, 2024): deteksi anomali memperlakukan pencilan sebagai sinyal yang dicari, sedangkan segmentasi memperlakukannya sebagai derau. Konsekuensinya, perilaku yang dilakukan mayoritas populasi tidak akan pernah tertangkap deteksi anomali dan hanya dapat diungkap melalui pendekatan deskriptif. Karena label tipe pembelajar tidak tersedia sebelumnya, pengelompokan dilakukan secara unsupervised dengan K-Means Clustering yang sederhana, efisien, dan mudah ditafsirkan sebab profil tiap segmen terbaca langsung dari nilai median fiturnya.

Kelayakan pendekatan ini didukung banyak preseden mutakhir di dalam negeri. Ikhsan (2021) mengklaster log Moodle dan memperoleh tiga kelompok perilaku peserta sebagai dasar intervensi; Pratama, Ikhsan, dan Solichin (2025) membandingkan K-Means dengan K-Medoids pada log aktivitas LMS mahasiswa; Saputra, Agustin, Zaelani, dan Siddiq (2026) memanfaatkan

log e-learning untuk mendeteksi perubahan pola belajar antar-periode; sedangkan Nurfaizah (2026) mengungkap dua kelompok pola aktivitas belajar dari data LMS. Pada sumber data yang berbeda, Pramdhana, Nur, Sumadikarta, dan Aprillia (2026) memakai kerangka learning analytics pada hasil kuis daring untuk mengidentifikasi pola keterlibatan mahasiswa, Ritonga dan Lubis (2025) menyegmentasi pengguna LMS dari data kuesioner, dan Agusriandi, Elihami, Syarif, dan Samad (2022) mengklaster log LMS pada objek tutor. K-Means juga diadopsi luas pada data akademik agregat, mulai dari ketepatan waktu kelulusan (Azhara, Nugraha, dan Piqri, 2021), capaian pembelajaran lulusan (Agustia dan Finandhita, 2025), nilai dan kehadiran (Simanjuntak, Gulo, dan Sipayung, 2026), gaya belajar (Setiawan dan Mailoa, 2025), keaktifan berkegiatan (Hasibuan, Fahlome, Nabila, Arrahman, dan Furqan, 2026), hingga hasil belajar yang dikaitkan dengan self regulated learning (Syaifuddin, Wijayanti, dan Sudibyo, 2022). Di lingkungan Universitas Muhammadiyah Makassar sendiri, Suriani et al. (2026) telah memadukan K-Means dengan jaringan saraf tiruan untuk mendukung pembelajaran adaptif.

Meski demikian, studi-studi tersebut umumnya bekerja pada fitur agregat keaktifan, pada data persepsi, atau pada satu platform asesmen, dan kerap menetapkan jumlah klaster secara ad-hoc tanpa melaporkan kepekaannya terhadap ambang yang dipilih. Akibatnya tipologi yang dihasilkan berujung pada tingkatan aktif hingga pasif, sementara perbedaan orientasi belajar antar-mahasiswa dengan volume aktivitas setara tidak terungkap. Tipologi yang dibentuk dari clickstream tingkat permintaan HTTP sebuah LMS institusional, dengan prosedur pemilihan jumlah segmen yang terdokumentasi dan diuji sensitivitasnya, belum tersedia.

Penelitian ini karena itu mengajukan dua pertanyaan. Pertama, bagaimana penerapan K-Means Clustering untuk mengelompokkan mahasiswa berdasarkan pola aktivitas belajar pada log audit LMS Akademik? Kedua, bagaimana karakteristik tiap segmen pembelajar yang terbentuk? Kontribusi yang ditawarkan ada tiga: 13 fitur deskriptif yang menambahkan dimensi orientasi belajar berupa rasio akses asesmen terhadap baca materi di samping dimensi volume dan ritme temporal; prosedur pemilihan jumlah segmen berbasis prinsip parsimoni yang eksplisit beserta uji sensitivitasnya; serta uji silang dengan hasil deteksi anomali pada data yang sama untuk menunjukkan bahwa kedua pendekatan mengungkap titik buta satu sama lain.

B. Metodologi

Jenis Penelitian dan Sumber Data

Penelitian ini bersifat kuantitatif deskriptif dengan pendekatan unsupervised learning. Data yang digunakan adalah data sekunder berupa ekspor log audit LMS Akademik Program Studi Informatika Universitas Muhammadiyah Makassar pada rentang 2026/07/24 sampai 2026/07/28 sebanyak 50.000 baris permintaan HTTP. Data diperoleh secara read-only dengan izin pengelola sistem. Perlu dicatat bahwa rentang lima hari tersebut bukan pilihan desain: permintaan ekspor diajukan untuk rentang yang lebih panjang, tetapi sistem membatasi keluaran pada 50.000 baris terbaru sehingga yang terambil hanya lima hari terakhir yang kebetulan bertepatan dengan periode asesmen. Seluruh identitas mahasiswa dianonimkan menggunakan hash SHA-256 sebelum keluaran disimpan, dan setiap berkas keluaran diperiksa bebas data pribadi.

Karena unit analisis penelitian adalah mahasiswa sebagai subjek pembelajar, baris log disaring menurut peran pengguna. Baris milik dosen, Admin Prodi, dan Super Admin merekam aktivitas mengelola pembelajaran, bukan aktivitas belajar; baris System merupakan proses terjadwal tanpa pelaku manusia; sedangkan baris tanpa peran adalah permintaan pra-autentikasi yang tidak memuat identitas pengguna. Penyaringan menyisakan 45.476 baris (91,0%) milik 275 mahasiswa unik yang terentang pada 1.207 sesi dan 22 jenis aksi. Langkah ini penting karena pengguna non-mahasiswa berjumlah sedikit namun berintensitas akses tinggi dan berpola khas, sehingga apabila ikut diklasterkan akan menjadi pencilan ekstrem yang menarik pusat klaster.

Ekstraksi Fitur

Path endpoint dinormalisasi dengan menyamarkan UUID dan ID numerik sehingga 2.400 path unik memampat menjadi 74 pola bermakna. Tanpa normalisasi, setiap materi dan setiap soal tampak sebagai endpoint yang berbeda sehingga dua mahasiswa yang membuka bahan ajar berlainan tidak dikenali sebagai perilaku sejenis. Dari pola itulah sinyal baca materi, akses asesmen, jawab kuis, banding, dan perbaikan diidentifikasi.

Seluruh baris kemudian diagregasi menjadi matriks 275 × 13 fitur deskriptif per-mahasiswa dengan imputasi nol bagi mahasiswa yang tidak memiliki aktivitas pada suatu sinyal. Fitur tersebut dikelompokkan menjadi empat dimensi: volume dan tempo (jumlah permintaan, jumlah sesi, jumlah hari aktif, permintaan per sesi, burstiness), ritme temporal (rasio aktivitas malam, entropi jam akses), profil belajar (jumlah baca materi, jumlah akses asesmen, jumlah jawaban kuis, rasio akses asesmen terhadap baca materi), serta interaksi administratif (akses banding dan akses perbaikan). Fitur pelanggaran akses dan telemetri proctoring sengaja tidak digunakan karena tujuan penelitian adalah memahami pola belajar, bukan menilai pelanggaran.

Transformasi dan Pemodelan

Distribusi fitur cacah perilaku sangat menceng; nilai kemencengan tertinggi mencapai 7,95 pada fitur jumlah jawaban kuis dan 11 dari 13 fitur menceng ke kanan. Sebagian kecil mahasiswa hiperaktif karena itu dapat mendominasi jarak Euklides dan mendistorsi struktur klaster. QuantileTransformer dipakai untuk memetakan tiap fitur ke distribusi normal berbasis peringkat sehingga pengaruh nilai ekstrem teredam tanpa membuang data; transformasi berbasis peringkat dipilih alih-alih standardisasi z-score karena z-score tetap mempertahankan bentuk kemencengan distribusi.

Pemodelan dilakukan dengan K-Means yang mempartisi objek ke dalam k klaster dengan meminimumkan jumlah kuadrat jarak setiap objek ke centroid klasternya. Sebagaimana diuraikan Agustia dan Finandhita (2025) serta Simanjuntak, Gulo, dan Sipayung (2026), algoritme bekerja iteratif: inisialisasi k centroid, penugasan objek ke centroid terdekat, penghitungan ulang centroid, dan pengulangan hingga konvergen. Karena hasilnya bergantung pada inisialisasi, algoritme dijalankan dengan n_init = 20 lalu diambil hasil terbaik. Agglomerative Clustering dengan kriteria peleburan Ward disertakan sebagai pembanding struktur, mengikuti praktik Pratama, Ikhsan, dan Solichin (2025) yang mengadu dua algoritme pada data log yang sama. Kedua algoritme dijalankan pada rentang k = 2 sampai 8.

Penentuan Jumlah Segmen

Tanpa label kebenaran, kualitas segmentasi dinilai dengan metrik validasi internal: silhouette score, indeks Calinski-Harabasz, dan indeks Davies-Bouldin. Persoalan yang jarang dibahas adalah perilaku metrik terhadap kenaikan k: silhouette kerap terus naik karena segmen besar terfragmentasi menjadi pecahan kecil yang rapat, bukan karena struktur data makin terungkap. Penelitian ini karena itu menerapkan prinsip parsimoni, yaitu memilih k terkecil yang silhouette-nya masih berada dalam toleransi 0,015 dari nilai terbaik dan setiap segmennya beranggota minimal 20 mahasiswa. Ambang tersebut kemudian divariasikan (12 kombinasi toleransi 0,010–0,020 dan ukuran segmen minimum 15–30) untuk menguji kepekaan keputusan.

Pengujian dan Analisis

1. **Kualitas internal.** Ketiga metrik validasi internal dihitung untuk seluruh kombinasi algoritme dan k pada rentang 2 sampai 8, lalu diuraikan pula per segmen untuk memeriksa kemerataan mutu penempatan.
2. **Kestabilan.** K-Means dijalankan ulang dengan lima nilai random seed (0, 1, 7, 42, 2026); kesamaan komposisi segmen antar-percobaan diukur dengan Adjusted Rand Index (ARI) pada sepuluh pasangan hasil dan dinyatakan stabil apabila reratanya minimal 0,80.
3. **Kesepakatan antar-algoritme.** Komposisi segmen K-Means dibandingkan dengan Agglomerative Clustering pada k terpilih menggunakan ARI dan pencocokan profil median.
4. **Perbedaan antar-segmen.** Karena distribusi fitur tidak normal, perbedaan antar-segmen diuji dengan Kruskal-Wallis ($\varepsilon = 0{,}05$) beserta ukuran efek ε^2.
5. **Komplementaritas.** Sebaran mahasiswa ter-flag anomali pada tiap segmen diuji dengan khi-kuadrat untuk memeriksa apakah kedua pendekatan menyoroti kelompok yang berbeda.

Interpretasi dilakukan dengan memprofilkan tiap segmen dari nilai median fiturnya, menamainya berdasarkan peringkat relatif antar-segmen, dan memvisualisasikan sebarannya melalui proyeksi Principal Component Analysis (PCA) dua komponen sebagaimana diterapkan Hasibuan, Fahlome, Nabila, Arrahman, dan Furqan (2026). Seluruh tahapan diringkas pada

Gambar 1.



Gambar 1. Alur penelitian dari ekspor log audit sampai tipologi pembelajar

C. Hasil dan Pembahasan

Hasil

Deskripsi Data dan Fitur

Dari 50.000 baris ekspor, 45.476 baris (90,95%) berasal dari 275 mahasiswa dan dipakai sebagai bahan analisis, sedangkan 4.524 baris sisanya milik dosen, Admin Prodi, Super Admin, System, dan permintaan pra-autentikasi disaring keluar. Tercatat 284 pengguna unik dengan 6 akun di antaranya tercatat pada lebih dari satu peran, sehingga penyaringan dilakukan pada tingkat baris dan bukan tingkat akun.

Aktivitas terpusat pada 2026/07/25 dengan 16.835 permintaan dan memuncak pada pukul 20.00 yang menampung 17,0% permintaan; hanya 6,9% permintaan terjadi pada dini hari (00.00–05.00). Sebanyak 113 mahasiswa tidak tercatat membuka bahan ajar sama sekali, sedangkan hanya 1 mahasiswa yang tidak mengakses asesmen. Ketimpangan itu sudah tampak sebelum pengelompokan dilakukan. Statistik ringkas seluruh 13 fitur disajikan pada Tabel 1.

Tabel 1. Statistik deskriptif 13 fitur perilaku belajar

Fitur	Keterangan	Min	Median	Rerata	Maks	Simp. baku	Kemencengan
n_request	Jumlah permintaan HTTP	2,00	93,00	165,37	840,00	172,91	1,56
n_sesi	Jumlah sesi	1,00	3,00	4,39	23,00	4,07	1,82
n_hari_aktif	Jumlah hari aktif	1,00	2,00	1,71	5,00	0,85	1,13
req_per_sesi	Permintaan per sesi	2,00	31,67	36,66	156,50	25,76	1,97
burstiness_req	Burstiness permintaan	0,00	7,37	7,73	24,55	5,03	0,81
night_ratio	Rasio aktivitas malam (00–05)	0,00	0,00	0,08	1,00	0,21	3,16
entropi_jam	Entropi jam akses	0,00	1,00	1,11	3,13	0,85	0,22
n_baca_materi	Jumlah baca materi	0,00	3,00	13,79	399,00	32,67	7,18
n_akses_asesmen	Jumlah akses asesmen	0,00	40,00	82,79	665,00	107,46	2,49
n_jawab_kuis	Jumlah jawaban kuis	0,00	0,00	1,72	99,00	8,19	7,95
rasio_ujian_materi	Rasio asesmen : materi	0,00	6,67	20,06	303,00	37,40	4,58
n_banding	Jumlah akses banding	0,00	1,00	4,47	70,00	10,66	3,66
n_perbaikan	Jumlah akses perbaikan	0,00	3,00	6,34	48,00	8,00	2,05

Tabel 1 memperlihatkan dua hal yang menentukan langkah selanjutnya. Pertama, rerata konsisten jauh di atas median pada hampir seluruh fitur cacah, misalnya jumlah baca materi bermedian 3 namun berrerata 13,79 dengan maksimum 399. Kedua, kemencengan yang tinggi tersebut menjustifikasi pemakaian transformasi kuantil sebelum pengelompokan.

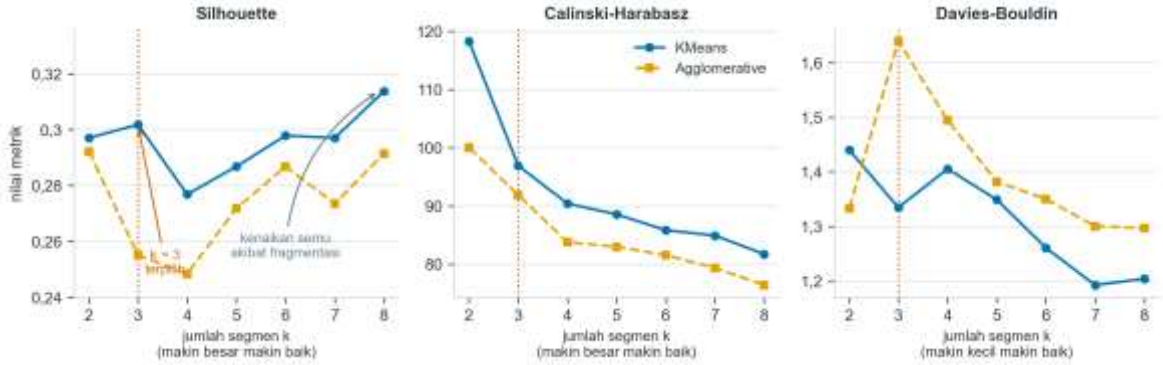
Penentuan Jumlah Segmen

Kedua algoritme dijalankan pada rentang k = 2 sampai 8, menghasilkan 14 kombinasi yang metrik validasi internalnya disajikan pada Tabel 2.

Tabel 2. Metrik validasi internal seluruh kombinasi algoritme dan k (baris tebal menandai kombinasi terpilih)

Algoritme	k	Silhouette	Calinski-Harabasz	Davies-Bouldin	Segmen terkecil
K-Means	2	0,2970	118,30	1,4401	103
Agglomerative	2	0,2922	100,04	1,3329	68
K-Means	**3**	**0,3018**	**96,92**	**1,3347**	**37**
Agglomerative	3	0,2551	91,87	1,6394	68
K-Means	4	0,2768	90,42	1,4053	38
Agglomerative	4	0,2484	83,79	1,4944	48
K-Means	5	0,2868	88,57	1,3490	28
Agglomerative	5	0,2718	82,98	1,3819	22
K-Means	6	0,2978	85,84	1,2606	19
Agglomerative	6	0,2868	81,55	1,3502	22
K-Means	7	0,2970	84,92	1,1924	19
Agglomerative	7	0,2735	79,37	1,3002	22
K-Means	8	0,3137	81,72	1,2040	20
Agglomerative	8	0,2914	76,41	1,2969	22

Silhouette tertinggi di antara 14 kombinasi justru diperoleh K-Means pada k = 8 (0,3137), tetapi kenaikan itu semu. Tabulasi silang keanggotaan menunjukkan 91,6% anggota pada k = 8 hanyalah pecahan dari segmen pada k = 3, sehingga tambahan segmen tidak memunculkan tipe perilaku baru melainkan memotong tipe yang sudah ada. Aturan parsimoni karena itu memilih K-Means pada k = 3 dengan silhouette 0,3018, yaitu peringkat 2 dan hanya 0,0119 di bawah nilai terbaik. Nilai Calinski-Harabasz dan Davies-Bouldin pada k terpilih juga tergolong baik (96,92 dan 1,3347). Perbandingan ketiga metrik sepanjang rentang k disajikan pada Gambar 2.



Gambar 2. Perbandingan tiga metrik validasi internal pada rentang k = 2 sampai 8

Uji sensitivitas atas 12 kombinasi ambang menunjukkan bahwa k = 3 bertahan pada 2 kombinasi, sedangkan pengetatan toleransi menjadi 0,010 menghasilkan k = 8 dan pelonggarannya menjadi 0,020 menghasilkan k = 2. Keputusan jumlah segmen dengan demikian memang peka terhadap pilihan ambang, sehingga k = 3 sebaiknya dipahami sebagai tingkat kedalaman deskripsi yang paling berguna, bukan jumlah segmen yang benar secara mutlak. Kepekaan ini dilaporkan apa adanya karena justru menjadi bagian dari kontribusi metodologis penelitian.

Profil dan Penamaan Segmen

Setiap segmen diprofilkan dari nilai median fiturnya lalu dinamai berdasarkan peringkat relatif antar-segmen. Hasilnya disajikan pada Tabel 3.

Tabel 3. Profil median tiap segmen tipologi pembelajar

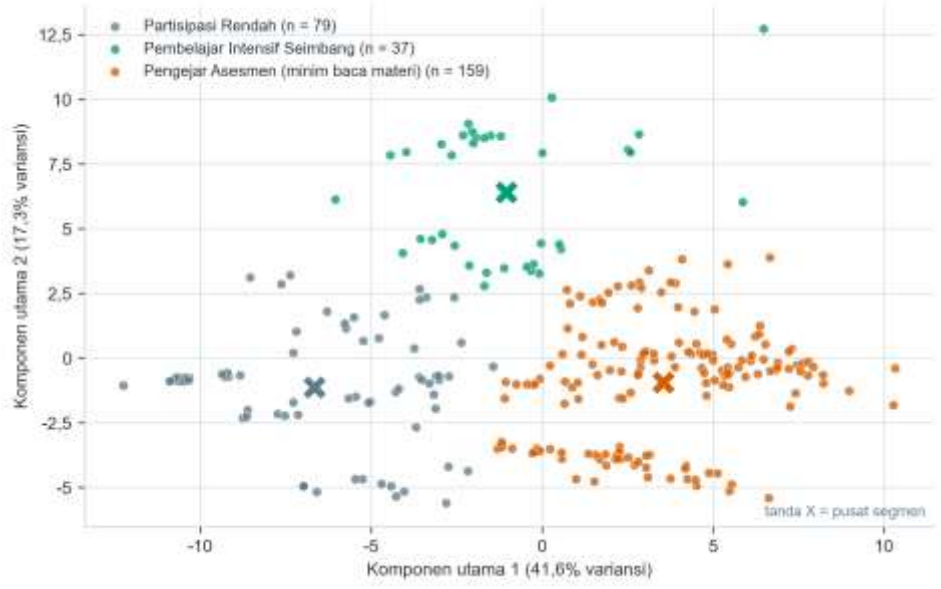
Fitur (median)	Pengejar Asesmen	Partisipasi Rendah	Pembelajar Intensif Seimbang
Jumlah mahasiswa	159	79	37

Fitur (median)	Pengejar Asesmen	Partisipasi Rendah	Pembelajar Intensif Seimbang
Proporsi populasi	57,8%	28,7%	13,5%
Jumlah permintaan HTTP	150	25	181
Jumlah sesi	4	1	4
Jumlah hari aktif	2	1	1
Permintaan per sesi	32	19	39
Burstiness permintaan	8,56	3,37	6,04
Rasio aktivitas malam (00–05)	0,00	0,00	0,00
Entropi jam akses	1,57	0,00	1,00
Jumlah baca materi	4	0	31
Jumlah akses asesmen	63	8	89
Jumlah jawaban kuis	0	0	5
Rasio asesmen : materi	12,38	4,00	2,11
Jumlah akses banding	2	0	0
Jumlah akses perbaikan	8	1	0

Karakteristik ketiga tipologi diuraikan sebagai berikut.

1. **Pengejar Asesmen** 159 mahasiswa (57,8%). Segmen terbesar. Volume aktivitasnya menengah (150 permintaan pada 4 sesi selama 2 hari), tetapi komposisinya timpang: median baca materi 4 kali berbanding 63 kali akses asesmen, atau 12,38 banding satu. Median akses perbaikan dan banding juga tertinggi (8 dan 2), menandakan orientasi pada perolehan nilai alih-alih penguasaan bahan. Mereka aktif dan tampak wajar, tetapi nyaris melewati seluruh bahan ajar.
2. **Partisipasi Rendah** 79 mahasiswa (28,7%). Keterlibatan paling rendah pada semua dimensi: median 25 permintaan dalam 1 sesi pada 1 hari, tanpa akses bahan ajar dan tanpa jawaban kuis. Entropi jam aksesnya nol, artinya seluruh aktivitas terjadi pada satu jam saja. Pola ini khas mahasiswa yang masuk sekadar memeriksa kewajiban lalu keluar, dan merupakan kelompok berisiko putus keterlibatan yang paling jelas.
3. **Pembelajar Intensif Seimbang** 37 mahasiswa (13,5%). Segmen terkecil sekaligus paling produktif: median 181 permintaan, 31 kali baca materi, 5 jawaban kuis, dan rasio asesmen terhadap materi hanya 2,11 banding satu atau sekitar seperenam rasio pengejar asesmen. Mereka menempuh siklus belajar yang utuh: membaca, mengerjakan, lalu mengirim jawaban.

Proyeksi PCA pada Gambar 3 menampung 58,9% variansi total (41,6% dan 17,3%). Komponen pertama dibentuk oleh volume dan keteraturan aktivitas, komponen kedua oleh intensitas membaca materi. Ketiga segmen menempati wilayah yang berbeda dengan tumpang tindih di perbatasan, sejalan dengan nilai silhouette yang moderat.



Gambar 3. Proyeksi PCA dua komponen utama dengan pewarnaan menurut segmen

Perbedaan Antar-Segmen

Karena distribusi fitur tidak normal, perbedaan antar-segmen diuji dengan Kruskal-Wallis. Hasil selengkapnya disajikan pada Tabel 4.

Tabel 4. Hasil uji Kruskal-Wallis perbedaan antar-segmen pada setiap fitur

Fitur	Keterangan	H	p	ε^2
entropi_jam	Entropi jam akses	156,32	$1{,}14 \times 10^{-34}$	0,567
n_sesi	Jumlah sesi	154,78	$2{,}45 \times 10^{-34}$	0,562
n_perbaikan	Jumlah akses perbaikan	152,10	$9{,}39 \times 10^{-34}$	0,552
n_akses_asesmen	Jumlah akses asesmen	127,55	$2{,}01 \times 10^{-28}$	0,462
n_request	Jumlah permintaan HTTP	127,21	$2{,}38 \times 10^{-28}$	0,460
n_jawab_kuis	Jumlah jawaban kuis	122,71	$2{,}26 \times 10^{-27}$	0,444
n_hari_aktif	Jumlah hari aktif	122,40	$2{,}64 \times 10^{-27}$	0,443
n_banding	Jumlah akses banding	102,38	$5{,}87 \times 10^{-23}$	0,369
n_baca_materi	Jumlah baca materi	81,16	$2{,}38 \times 10^{-18}$	0,291
rasio_ujian_materi	Rasio asesmen : materi	67,51	$2{,}19 \times 10^{-15}$	0,241
burstiness_req	Burstiness permintaan	38,50	$4{,}36 \times 10^{-9}$	0,134
req_per_sesi	Permintaan per sesi	26,37	$1{,}88 \times 10^{-6}$	0,090
night_ratio	Rasio aktivitas malam (00–05)	18,49	$9{,}64 \times 10^{-5}$	0,061

Seluruh 13 fitur berbeda signifikan antar-segmen dengan nilai p terbesar $9{,}64 \times 10^{-5}$, jauh di bawah ε = 0,05. Pembeda terkuat berdasarkan ukuran efek adalah entropi jam akses (ε^2 = 0,567), jumlah sesi (ε^2 = 0,562), jumlah akses perbaikan (ε^2 = 0,552). Urutan ini bermakna: yang paling membedakan tipologi bukanlah besar kecilnya volume aktivitas, melainkan keteraturan ritme akses dan intensitas keterlibatan. Ketiga segmen karena itu bukan pembagian sembarang atas populasi yang homogen.

Kestabilan dan Kesepakatan Antar-Algoritme

K-Means diulang dengan lima random seed lalu 10 pasangan hasilnya dibandingkan memakai Adjusted Rand Index. Rerata ARI 0,9949 dengan nilai terendah 0,9872, jauh di atas ambang 0,80 yang ditetapkan. Perbedaan antar-percobaan hanya menyangkut beberapa mahasiswa di perbatasan segmen, sehingga komposisi segmen dapat dinyatakan stabil terhadap inisialisasi.

Pada k = 3, ARI antara K-Means dan Agglomerative Clustering adalah 0,6595. Bila tiap segmen K-Means dipasangkan dengan klaster Agglomerative yang paling banyak memuat anggotanya, 236 dari 275 mahasiswa (85,8%) ditempatkan secara sepadan. Kesepakatannya kuat tetapi tidak sempurna: Agglomerative memecah segmen pengejar asesmen menjadi dua kelompok menurut tingkat aktivitas dan menggabungkan sebagian anggotanya dengan pembelajar intensif. Hal itu wajar karena kriteria Ward meminimumkan pertambahan variansi secara bertahap dan tidak dapat meninjau ulang peleburan yang telah terjadi. Yang penting, kedua algoritme sama-sama memisahkan kelompok partisipasi rendah dan sama-sama menempatkan orientasi asesmen sebagai pembeda utama, sehingga struktur segmen dapat dinyatakan melekat pada data dan bukan artefak satu algoritme.

Penguraian silhouette per segmen menunjukkan mutu penempatan yang merata, dengan hanya 0,7% anggota populasi bernilai negatif, yaitu mahasiswa yang lebih dekat ke segmen tetangga. Proporsi sekecil itu menegaskan bahwa silhouette keseluruhan yang moderat disebabkan batas antar-segmen yang landai, bukan salah penempatan yang meluas.

Uji Silang dengan Deteksi Anomali

Untuk menguji komplementaritas pendekatan deskriptif dan pendekatan deteksi, hasil segmentasi ditabulasi silang dengan penanda anomali dari penelitian pendamping pada data yang sama. Hasilnya disajikan pada Tabel 5 dan Gambar 4.

Tabel 5. Sebaran mahasiswa ter-flag anomali pada setiap segmen

Segmen	Jumlah mahasiswa	Ter-flag anomali	Ter-flag ganda	Proporsi ter-flag
Pengejar Asesmen	159	5	2	3,1%
Partisipasi Rendah	79	0	0	0,0%
Pembelajar Intensif Seimbang	37	9	6	24,3%
Seluruh populasi	**275**	**14**	**8**	**5,1%**



Gambar 4. Proporsi populasi tiap segmen dan sebaran flag deteksi anomali di dalamnya

Penanda anomali tersebar sangat tidak merata antar-segmen (χ^2 = 33,81; p < 0,001), dan ketidakmerataan itulah temuan pokoknya. Pembelajar Intensif Seimbang hanya 13,5% populasi tetapi menampung 9 dari 14 penanda (24,3% anggotanya). Hal itu wajar karena model anomali memberi skor tinggi pada volume aktivitas ekstrem. Sebaliknya Partisipasi Rendah yang mencakup 28,7% populasi tidak memiliki satu pun penanda, karena perilakunya yang seragam justru tampak paling wajar bagi model pencilan.

Pembahasan

Penerapan K-Means pada Log Audit LMS

Pertanyaan pertama penelitian terjawab melalui rangkaian yang terbukti layak: normalisasi path (2.400 nilai unik menjadi 74 pola), agregasi menjadi 13 fitur per-mahasiswa, transformasi kuantil, dan pemodelan pada k = 2 sampai 8. Aturan parsimoni menetapkan k = 3 dengan silhouette 0,3018, stabil terhadap inisialisasi (ARI 0,9949), dan sepadan dengan algoritme pembanding pada 85,8% keanggotaan. Nilai silhouette tersebut tergolong moderat, namun besaran seperti itu lazim pada data perilaku pendidikan karena batas antar-kelompok memang landai. Sebagai pembanding, Agustia dan Finandhita (2025) melaporkan silhouette 0,2728 pada pengelompokan capaian pembelajaran lulusan dan Setiawan dan Mailoa (2025) melaporkan 0,340 pada pengelompokan gaya belajar. Kontribusi metodologis penelitian ini terletak pada prosedur pemilihan k yang eksplisit beserta uji sensitivitas yang dilaporkan apa adanya, serta pada pemakaian log tingkat permintaan HTTP yang selama ini jarang dipakai untuk membentuk tipologi.

Karakteristik Segmen dan Perbandingan dengan Penelitian Terdahulu

Jumlah tipologi yang diperoleh sejalan dengan penelitian sejenis: Ikhsan (2021) memperoleh tiga kelompok perilaku dari log Moodle dan Nurfaizah (2026) dua kelompok pola aktivitas dari data LMS. Kesamaan ini memperkuat keyakinan bahwa pembelajar daring memang terpolakan pada segelintir tipe. Perbedaannya terletak pada dimensi yang terungkap. Kedua penelitian tersebut memakai fitur agregat keaktifan sehingga tipologinya berujung pada tingkatan aktif hingga pasif, sebagaimana juga terjadi pada segmentasi berbasis persepsi pengguna (Ritonga dan Lubis, 2025). Penambahan fitur rasio akses asesmen terhadap baca materi memunculkan dimensi baru, yaitu orientasi belajar: Pengejar Asesmen memiliki volume aktivitas yang setara dengan Pembelajar Intensif Seimbang (150 berbanding 181 permintaan), tetapi komposisi perilakunya berbeda sama sekali (12,38 berbanding 2,11 banding satu). Pada kerangka fitur konvensional keduanya akan melebur menjadi satu kelompok mahasiswa aktif dan persoalan pokoknya luput.

Temuan paling penting justru muncul dari uji silang dengan deteksi anomali. Pengejar Asesmen mencakup 57,8% populasi tetapi hanya menyumbang 5 penanda (3,1% anggotanya). Mengakses asesmen nyaris tanpa membaca materi mustahil dikenali sebagai anomali justru karena dilakukan mayoritas populasi; bagi model pencilan, perilaku mayoritas adalah definisi kenormalan. Inilah bukti bahwa kedua pendekatan mengungkap titik buta satu sama lain: deteksi anomali menemukan individu yang perlu ditindaklanjuti (Mutmainah dan Yustanti, 2024), sedangkan segmentasi menemukan persoalan sistemik yang tidak akan pernah muncul pada daftar peringkat anomali. Implikasinya, institusi yang hanya memasang sistem deteksi anomali pada LMS-nya berisiko memperoleh rasa aman yang keliru.

Implikasi Praktis

Tiap segmen memiliki penyebab keterlambatan belajar yang berbeda, sehingga perlakuan seragam tidak akan efektif. Rekomendasi perlakuan per segmen dirangkum pada Tabel 6.

Tabel 6. Rekomendasi perlakuan pembelajaran menurut segmen

Segmen	Persoalan pokok	Rekomendasi perlakuan
Pengejar Asesmen	Volume aktivitas memadai, tetapi bahan ajar nyaris dilewati; orientasi pada perolehan nilai	Menautkan asesmen dengan bahan ajar melalui prasyarat keterbacaan sebelum percobaan, memecah asesmen besar menjadi kuis formatif berbasis materi, serta menampilkan umpan balik yang merujuk bagian materi tertentu
Partisipasi Rendah	Nyaris tidak menyentuh sistem; berisiko putus keterlibatan	Peringatan dini otomatis pada minggu pertama tanpa aktivitas, penjangkauan oleh penasihat akademik, serta penelusuran hambatan non-akademik seperti akses perangkat, jaringan, dan beban kerja
Pembelajar Intensif Seimbang	Sudah menempuh siklus belajar utuh; berisiko kurang tertantang	Pengayaan berupa bahan lanjutan dan tugas terbuka, serta pelibatan sebagai tutor sebaya bagi dua segmen lainnya

Perlu ditegaskan bahwa keanggotaan segmen adalah deskripsi pola kelompok pada periode pengamatan, bukan penilaian permanen atas individu. Karena itu perlakuan sebaiknya diarahkan pada tingkat kelas atau kelompok, sedangkan keputusan yang menyangkut individu tetap memerlukan penelusuran tersendiri.

Keterbatasan Penelitian

Empat keterbatasan perlu dicatat. Pertama, rentang data hanya lima hari dan bertepatan dengan periode asesmen akibat pembatasan ekspor sistem, sehingga tipologi ini adalah pola perilaku periode asesmen dan belum tentu berlaku pada perkuliahan reguler. Kedua, keputusan jumlah segmen peka terhadap pilihan ambang sebagaimana ditunjukkan uji sensitivitas. Ketiga, seluruh fitur berasal dari jejak klik sehingga tidak dapat membedakan membaca dengan sungguh-sungguh dari sekadar membuka halaman. Keempat, data capaian akademik pada periode yang sama tidak tersedia dan validasi pemaknaan oleh dosen serta pengelola prodi belum dilaksanakan, sehingga hubungan tipologi dengan hasil belajar dan kesesuaiannya dengan pengalaman lapangan belum dapat diuji. Keempat keterbatasan tersebut sekaligus menjadi arah penelitian lanjutan, bersama pengujian dengan algoritme yang tidak mengandaikan klaster berbentuk bola seperti Gaussian Mixture Model.

D. Kesimpulan

Penelitian ini menunjukkan bahwa K-Means Clustering dapat diterapkan pada log audit LMS Akademik melalui empat tahap, yaitu normalisasi path (2.400 nilai unik menjadi 74 pola), agregasi menjadi 13 fitur per-mahasiswa, transformasi kuantil untuk meredam kemencengan, dan pemodelan pada k = 2 sampai 8. Aturan parsimoni memilih k = 3 dengan silhouette 0,3018. Hasil tersebut stabil terhadap inisialisasi (rerata ARI 0,9949 dari 10 pasangan seed) dan sepadan dengan Agglomerative Clustering pada 85,8% keanggotaan, sehingga segmentasi yang terbentuk dapat dinyatakan andal.

Terbentuk tiga tipologi pembelajar yang berbeda signifikan pada seluruh 13 fitur menurut uji Kruskal-Wallis ($p < 0{,}001$), yaitu Pengejar Asesmen sebanyak 159 mahasiswa (57,8%) dengan rasio akses asesmen terhadap baca materi 12,38 banding satu; Partisipasi Rendah sebanyak 79 mahasiswa (28,7%) dengan median aktivitas terendah pada seluruh dimensi; dan Pembelajar Intensif Seimbang sebanyak 37 mahasiswa (13,5%) dengan rasio 2,11 banding satu. Pembeda terkuat antar-segmen adalah keteraturan ritme akses dan intensitas keterlibatan, bukan besar kecilnya volume aktivitas. Karakteristik tersebut dikuatkan oleh sebaran penanda anomali yang tidak merata antar-segmen ($\chi^2 = 33{,}81$; $p < 0{,}001$): mengejar asesmen tanpa membaca materi luput dari deteksi anomali justru karena dilakukan mayoritas populasi, sehingga segmentasi mengungkap persoalan sistemik yang tidak muncul pada daftar peringkat anomali.

Bagi program studi, hasil ini dapat langsung dipakai untuk membedakan perlakuan pembelajaran menurut segmen. Bagi pengelola LMS, instrumentasi log perlu dilengkapi penanda durasi baca dan identitas materi yang stabil agar kedalaman keterlibatan dapat diukur, bukan sekadar jumlah klik. Bagi penelitian selanjutnya, perluasan rentang data menjadi satu semester penuh dan penautan dengan data capaian akademik merupakan arah yang paling menjanjikan.

E. Referensi

Agusriandi, Elihami, Syarif, I., & Samad, I. S. (2022). Model analisis aktivitas tutor dalam learning management system berdasarkan data log menggunakan K-Means dan deteksi outlier. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 9(4), 709–716. https://doi.org/10.25126/jtiik.202294764

Agustia, R. D., & Finandhita, A. (2025). Pengelompokan mahasiswa berdasarkan capaian pembelajaran lulusan menggunakan K-Means Clustering di Program Studi Teknik Informatika UNIKOM. *KOMPUTA: Jurnal Ilmiah Komputer dan Informatika*, 14(2), 63–74. https://doi.org/10.34010/komputa.v14i2.17932

Azhara, A. R., Nugraha, R. A., & Piqri, A. (2021). Penerapan K-Means Clustering tingkat ketepatan waktu kelulusan mahasiswa Program Studi Teknik Informatika STMIK Nurdin Hamzah Jambi. *Jurnal SIMETRIS*, 12(1), 1–15. https://jurnal.umk.ac.id/index.php/simet

Hakim, A. F. (2025). Pemanfaatan learning analytics dalam pengambilan keputusan manajerial di sekolah: Peluang dan tantangan di era transformasi digital. *AL-MARSUS: Jurnal Manajemen Pendidikan Islam*, 3(2), 152–169. https://doi.org/10.30983/al-marsus.v3i2.10707

Hasibuan, N. A. M., Fahlome, D., Nabila, P. S., Arrahman, S., & Furqan, M. (2026). Segmentasi keaktifan mahasiswa Universitas Islam Negeri Sumatera Utara dalam kegiatan kampus menggunakan K-Means Clustering. *Syntax: Journal of Software Engineering, Computer Science and Information Technology*, 7(1), 46–52. https://jurnal.itscience.org/index.php/syntax

Ikhsan, E. (2021). Penerapan K-Means Clustering dari log data Moodle untuk menentukan perilaku peserta pada pembelajaran daring. *SISTEMASI: Jurnal Sistem Informasi*, 10(2), 414–422. http://sistemasi.ftik.unisi.ac.id

Mutmainah, M., & Yustanti, W. (2024). Studi komparasi Local Outlier Factor (LOF) dan Isolation Forest (IF) pada analisis anomali kinerja dosen. *JINACS (Journal of Informatics and Computer Science)*, 6(2), 532–540. https://doi.org/10.26740/jinacs.v6n02.p532-540

Nurfaizah. (2026). Analisis pola aktivitas belajar mahasiswa pada learning management system menggunakan teknik clustering. *Saturnus: Jurnal Teknologi dan Sistem Informasi*, 4(1), 176–186. https://doi.org/10.61132/saturnus.v4i1.1365

Pramdhana, D. S., Nur, E., Sumadikarta, I., & Aprillia, Y. T. (2026). Learning analytics berbasis Kahoot menggunakan algoritma K-Means Clustering untuk mengidentifikasi pola keterlibatan mahasiswa dalam pembelajaran. *Jurnal Teknologi Informasi*, 12(1), 174–187. https://ejournal.urindo.ac.id/index.php/TI

Pratama, A., Ikhsan, R. D., & Solichin, A. (2025). Analisis klaster pembelajaran daring mahasiswa berdasarkan aktifitas log data pengguna learning management system (LMS) menggunakan metode K-Means dan K-Medoids. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 10(4), 3714–3730. https://doi.org/10.29100/jipi.v10i4.9143

Ritonga, A. R., & Lubis, H. (2025). Analisis penggunaan learning management system pada mahasiswa Sistem Informasi Universitas Harapan Medan dengan menggunakan metode K-Means. *Jurnal Manajemen Sistem Informasi (JMSI)*, 7(1), 119–128. https://jurnal.unhar.ac.id/index.php/jmsi

Saputra, D. S., Agustin, S. D., Zaelani, F., & Siddiq, G. (2026). Deteksi perubahan pola belajar mahasiswa berdasarkan log aktivitas e-learning menggunakan metode K-Means Clustering. *Kohesi: Jurnal Multidisiplin Saintek*, 10(8), 1–7. https://ejournal.cahayailmubangsa.institute/index.php/kohesi

Setiawan, K., & Mailoa, E. (2025). Pengelompokan gaya belajar siswa untuk mendukung efektifitas pembelajaran dengan K-Means. *JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 16(2), 26–35. http://ejurnal.provisi.ac.id/index.php/JTIKP

Simanjuntak, T., Gulo, J. A., & Sipayung, S. P. (2026). Pengelompokan mahasiswa berdasarkan nilai dan kehadiran menggunakan K-Means. *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, 5(1), 1–8. https://doi.org/10.55123/jomlai.v5i1.7283

Suriani, Faisal, M., Darniati, Agusalim, E., Kuba, M. S. S., Lee, S. L., & Nurdiansyah. (2026). Algoritma hybrid K-Means dan jaringan saraf tiruan untuk meningkatkan prestasi akademik siswa. *Buletin Sistem Informasi dan Teknologi Islam (BUSITI)*, 7(2), 128–138. https://doi.org/10.33096/busiti.v7i2.3467

Syaifuddin, M. W., Wijayanti, S., & Sudibyo, N. A. (2022). Analisis hasil belajar, gender dan self regulated learning (SRL) mahasiswa menggunakan algoritma K-Means Clustering. *Khatulistiwa: Jurnal Pendidikan dan Sosial Humaniora*, 2(2), 141–147. https://doi.org/10.55606/khatulistiwa.v2i2.540