BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2625-2638
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1412

Perbandingan K-Means dan K-Medoids untuk Analisis Perilaku Operasional Kunjungan Perpustakaan Sekolah Menengah Atas

Comparing K-Means and K-Medoids for Operational Behavior Analysis of Senior High School Library Visits

Rian Septian Anwar[1,*], Entin Sutinah[2], Nani Agustina[2]

[1] Fakultas Teknik dan Informatika, Program Studi Teknik Elektro, Universitas Bina Sarana Informatika, Jakarta, Indonesia
[2] Fakultas Teknik dan Informatika, Program Studi Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia
Email: [1,*]rian.ptn@bsi.ac.id, [2]entin.esh@bsi.ac.id, [3]nani.nna@bsi.ac.id
Email Penulis Korespondensi: rian.ptn@bsi.ac.id
Received: **09/10/2026**, Revised: **17/10/2026**, Accepted: **30/10/2026**, Published: **31/10/2026**

Abstrak-Pencatatan kunjungan perpustakaan menyediakan informasi perilaku pemustaka, tetapi jumlah kunjungan agregat belum menjelaskan perbedaan intensitas dan kebaruan aktivitas setiap pengunjung. Kondisi ini menyulitkan pengelola perpustakaan sekolah menengah atas yang umumnya bekerja dengan petugas terbatas dalam menentukan prioritas layanan. Penelitian ini bertujuan membandingkan K-Means dan K-Medoids untuk menganalisis perilaku operasional kunjungan perpustakaan SMAN 48 Jakarta. Data mencakup 10.050 transaksi pada 18 Agustus 2025–7 Oktober 2026 yang direpresentasikan menjadi 1.033 profil pengunjung berkode anonim. Penelitian menggunakan kerangka CRISP-DM, sembilan fitur *recency–frequency*–temporal, transformasi logaritmik, standardisasi *Z-score*, serta evaluasi 14 konfigurasi dengan jumlah klaster dua sampai delapan. K-Means dengan dua klaster memperoleh *Silhouette* tertinggi sebesar 0,4114 dan Davies–Bouldin Index sebesar 1,1651, sedangkan K-Medoids terbaik menggunakan lima klaster dengan *Silhouette* 0,2825 dan Davies–Bouldin Index 1,2348. Segmentasi K-Means menghasilkan 165 pengunjung berintensitas rendah (15,97%) dan 868 pengunjung relatif lebih aktif (84,03%). Sepuluh pengulangan inisialisasi mereproduksi keanggotaan yang sama, sedangkan penghapusan fitur hari aktif yang identik dengan frekuensi menurunkan *Silhouette* menjadi 0,3896 dengan *Adjusted Rand Index* 0,9705, sehingga struktur dua segmen tetap bertahan tetapi sebagian kekuatan pemisahannya berasal dari pembobotan ganda. Kebaruan penelitian terletak pada protokol evaluasi segmentasi log kunjungan perpustakaan sekolah yang menggabungkan pembandingan dua algoritma, verifikasi ulang metrik, uji kestabilan inisialisasi, dan analisis sensitivitas fitur. Kontribusi praktisnya berupa dasar penjadwalan petugas pada jam sibuk, daftar tindak lanjut pengunjung jarang hadir, dan siklus pemutakhiran segmen per semester. Hasil bersifat deskriptif dan tidak membuktikan minat baca, kepuasan, ataupun dampak akademik pemustaka.

Kata Kunci: K-Means; K-Medoids; Segmentasi Pemustaka; Kunjungan Perpustakaan; Validasi Klaster

Abstract-Library visit records provide information about patron behavior, yet aggregate counts do not explain differences in individual activity intensity and *recency*. This gap hampers senior high school library managers with limited staff in setting service priorities. This study compares K-Means and K-Medoids to analyze the operational visit behavior of the SMAN 48 Jakarta library. The dataset contains 10,050 transactions recorded between August 18, 2025 and October 7, 2026, represented as 1,033 anonymized visitor profiles. The study follows the CRISP-DM framework, constructs nine recency–frequency–temporal features, applies logarithmic transformation and Z-score standardization, and evaluates 14 configurations with two to eight clusters. K-Means with two clusters achieves the highest Silhouette score of 0.4114 and a Davies–Bouldin Index of 1.1651, whereas the best K-Medoids configuration uses five clusters with scores of 0.2825 and 1.2348. The selected partition contains 165 low-intensity visitors (15.97%) and 868 relatively more active visitors (84.03%). Ten initialization runs reproduce the membership exactly, while removing active days, which duplicates visit frequency, lowers the Silhouette to 0.3896 with an Adjusted Rand Index of 0.9705; the two-segment structure persists although part of its separation stems from double weighting. The novelty lies in an evaluation protocol for school library visit logs that combines a two-algorithm comparison, metric re-verification, initialization stability testing, and feature sensitivity analysis. Its practical contribution comprises a basis for peak-hour staff scheduling, a follow-up list of infrequent visitors, and a semester-based segment update cycle. The results describe recorded utilization and do not establish reading interest, satisfaction, or academic impact.

Keywords: K-Means; K-Medoids; Patron Segmentation; Library Visits; Cluster Validation

1. PENDAHULUAN

Perpustakaan sekolah membutuhkan informasi yang lebih terperinci daripada total kunjungan untuk merencanakan layanan yang sesuai dengan perilaku pemustaka. Dua orang dapat memiliki jumlah kunjungan yang sama, tetapi berbeda dalam waktu terakhir hadir, sebaran hari, maupun keteraturan kunjungannya. Perbedaan tersebut memengaruhi cara pengelola menafsirkan kebutuhan layanan. Analisis transaksi perpustakaan telah dimanfaatkan untuk menemukan pola sirkulasi dan mendukung rekomendasi koleksi, sebagaimana ditunjukkan Khademizadeh, Nematollahi, dan Danesh [1]. Namun, transaksi peminjaman dan catatan kehadiran mewakili aktivitas yang berbeda. Kehadiran di perpustakaan tidak dengan sendirinya menunjukkan buku yang dibaca, tujuan kedatangan, atau keberhasilan pembelajaran. Oleh sebab itu, pemanfaatan catatan kunjungan harus diawali dengan definisi yang jelas mengenai unit analisis dan batas interpretasinya.

Pada SMAN 48 Jakarta, dataset penelitian memuat 10.050 catatan kunjungan selama 18 Agustus 2025 sampai 7 Oktober 2026. Data tersebut mencakup pengunjung siswa, guru atau staf, serta sebagian kecil kategori yang tidak diketahui. Agregasi menghasilkan 1.033 profil pengunjung, dengan rata-rata 9,73 transaksi per pengunjung. Ringkasan ini menunjukkan tersedianya bahan analisis longitudinal, tetapi belum menjelaskan apakah seluruh pengunjung

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2625-2638
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1412

menggunakan layanan dengan pola yang serupa. Informasi mengenai kelompok yang jarang hadir dan kelompok yang relatif sering hadir diperlukan untuk membedakan kebutuhan tindak lanjut. Tantangan penelitian terletak pada pembentukan representasi perilaku, pemilihan jumlah klaster, dan penerjemahan hasil komputasi menjadi rekomendasi yang dapat ditinjau oleh pengelola.

Penggunaan analitik juga perlu ditempatkan dalam konteks pengembangan layanan. Okunlaya, Syed Abdullah, dan Alias [2] mengembangkan kerangka konseptual layanan perpustakaan berbantuan kecerdasan buatan melalui analisis isi literatur. Kerangka tersebut menjelaskan hubungan teknologi dengan inovasi layanan, tetapi tidak memberikan pengujian segmentasi terhadap *log* kunjungan sekolah yang digunakan dalam penelitian ini. Dari sisi tata kelola, Jones dan Hinchliffe [3] membahas kesiapan praktisi perpustakaan menghadapi persoalan etika analitik pembelajaran. Kajian tersebut menegaskan pentingnya pertimbangan profesional terhadap penggunaan data mahasiswa. Pada konteks sekolah, kebutuhan menjaga kerahasiaan semakin relevan karena sebagian pemustaka merupakan peserta didik. Analisis ini karena itu menggunakan kode pengunjung dan melaporkan hasil agregat.

Pendekatan *recency* dan *frequency* memberi cara yang mudah dijelaskan untuk merangkum kebaruan serta intensitas aktivitas. Abbasimehr dan Bahrini [4] mengembangkan segmentasi dinamis berbasis RFM dan pengelompokan deret waktu dalam konteks pelanggan. Penelitian tersebut menunjukkan pentingnya representasi perubahan perilaku, tetapi dimensi *monetary* tidak memiliki padanan langsung dalam *log* kehadiran perpustakaan sekolah. Penelitian ini mengadaptasi unsur kebaruan dan frekuensi, kemudian menambahkan waktu kunjungan dan jeda antarkunjungan. Adaptasi tersebut tidak dimaksudkan untuk mengubah pemustaka menjadi pelanggan komersial. Tujuannya ialah memperoleh representasi operasional yang sesuai dengan data yang benar-benar tersedia, tanpa menambahkan nilai transaksi, kepuasan, atau indikator akademik yang tidak tercatat.

K-Means dipilih sebagai metode pembanding karena menghasilkan partisi berbasis kedekatan terhadap pusat kelompok yang relatif mudah ditafsirkan. Tinjauan Ikotun dan kolega [5] menunjukkan luasnya penggunaan K-Means sekaligus keterbatasannya terhadap inisialisasi, jumlah klaster, bentuk kelompok, dan pemilihan ukuran kesamaan. Ezugwu dan kolega [6] memperlihatkan bahwa karakteristik data dan kebutuhan aplikasi memengaruhi kesesuaian algoritma pengelompokan. Kedua kajian tersebut mendukung pembandingan yang terikat pada dataset tertentu, bukan klaim bahwa satu algoritma selalu unggul. K-Medoids memberikan pembanding dengan wakil kelompok berupa pengamatan aktual. Schubert dan Rousseeuw [7] mengembangkan percepatan optimasi medoid, sedangkan Lenssen dan Schubert [8] mengkaji optimasi *Medoid Silhouette* serta pemilihan jumlah klaster. Perbedaan fungsi objektif membuat perbandingan kedua algoritma perlu menggunakan metrik evaluasi yang konsisten.

Kualitas representasi fitur merupakan persoalan yang sama pentingnya dengan pemilihan algoritma. Wongoutong [9] membandingkan beberapa metode penskalaan dan menunjukkan bahwa fitur dengan satuan berbeda dapat menghasilkan perilaku pengelompokan yang berbeda apabila penskalaan diabaikan. Pada dataset ini, *recency* dinyatakan dalam hari, frekuensi dalam jumlah transaksi, dan proporsi waktu dalam rentang nol sampai satu. Standardisasi dibutuhkan untuk mengurangi dominasi satuan numerik tertentu. Meskipun demikian, standardisasi tidak menghilangkan pengaruh fitur yang berulang. Jika dua kolom identik tetap dimasukkan, aspek perilaku yang sama memperoleh kontribusi ganda dalam penghitungan jarak. Kondisi tersebut ditemukan pada frekuensi dan jumlah hari aktif sehingga perlu diuji secara terpisah.

Penilaian hasil tidak cukup dilakukan dengan satu grafik atau satu nilai maksimum. Ullmann, Hennig, dan Boulesteix [10] membedakan validasi internal, eksternal, visual, serta validasi pada data lain. Todeschini dan kolega [11], melalui perbandingan 68 indeks, menunjukkan bahwa pemilihan ukuran validitas dapat menghasilkan preferensi yang berbeda. Schubert [12] secara khusus mengkritik penggunaan kriteria siku sebagai dasar tunggal penentuan jumlah klaster. Dalam penelitian ini, *Silhouette* digunakan sebagai kriteria utama pemilihan partisi, sedangkan *Davies–Bouldin Index* digunakan sebagai ukuran pendamping. Kesepakatan atau ketidaksepakatan kedua ukuran akan dilaporkan, sehingga pembaca dapat menilai alasan pemilihan model tanpa menganggap seluruh metrik mendukung pilihan yang sama.

Kajian terdahulu yang paling dekat dengan penelitian ini dapat dibandingkan dari fokus, metode, temuan, dan keterbatasannya. Khademizadeh dan kolega [1] menerapkan teknik penambangan data pada data sirkulasi buku perpustakaan akademik dan menghasilkan sistem rekomendasi buku; kajian tersebut bertumpu pada transaksi koleksi, sedangkan penelitian ini menganalisis log kehadiran yang tidak memuat judul bacaan. Okunlaya dan kolega [2] menyusun kerangka konseptual inovasi layanan perpustakaan berbantuan kecerdasan buatan melalui analisis literatur, tetapi tidak menguji segmentasi empiris pada satu institusi. Abbasimehr dan Bahrini [4] menggabungkan model RFM dengan pengelompokan deret waktu untuk segmentasi dinamis pelanggan; penelitian ini mengadaptasi unsur *recency* dan *frequency* tanpa dimensi *monetary* yang tidak tersedia pada log perpustakaan. Wongoutong [9] menunjukkan bahwa pengabaian penskalaan fitur mengubah hasil K-Means, sehingga penelitian ini menstandarkan fitur sekaligus menguji pengaruh fitur yang identik. Ullmann dan kolega [10] menyusun kerangka validasi hasil pengelompokan pada data validasi; kerangka tersebut menjadi dasar pemisahan antara verifikasi metrik, kestabilan inisialisasi, dan validasi pada data lain yang belum dilakukan dalam penelitian ini.

Pemilihan K-Means dan K-Medoids sebagai pasangan pembanding didasarkan pada karakteristik data yang dihadapi, bukan pada anggapan bahwa dataset kunjungan memerlukan metode yang rumit. Kedua algoritma merupakan metode partisi klasik yang mudah dijalankan pada perangkat lunak umum, tetapi berbeda dalam cara menentukan wakil kelompok. K-Means menggunakan *centroid* berupa rerata dan meminimalkan jarak kuadrat sehingga peka terhadap nilai ekstrem, sedangkan K-Medoids menggunakan pengamatan aktual sebagai *medoid* dan meminimalkan jumlah jarak

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2626
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2625-2638
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1412

sehingga secara teoretis lebih tahan terhadap pencilan [5], [7]. Perbedaan ini relevan karena log kunjungan sekolah memiliki sebaran yang menceng: sebagian pengunjung hanya tercatat sekali, sebagian kecil tercatat hingga puluhan kali, dan rentang *recency* mencapai lebih dari satu tahun. Pada data seperti ini, belum dapat dipastikan apakah wakil kelompok berbasis rerata atau berbasis pengamatan aktual yang lebih sesuai, sehingga pembandingan empiris diperlukan. Segmentasi juga tidak cukup dilakukan dengan aturan ambang sederhana, misalnya membagi pengunjung berdasarkan frekuensi saja, karena kebaruan, keragaman hari, dan jeda kunjungan bekerja bersamaan dalam membentuk pola perilaku. Dengan demikian, kesederhanaan dataset justru menuntut pembandingan yang transparan agar pengelola memperoleh segmen yang dapat dipertanggungjawabkan dan dapat diulang dengan sumber daya terbatas.

Urgensi penelitian ini berkaitan dengan kondisi operasional perpustakaan sekolah menengah atas. Pengelola harus melayani seluruh warga sekolah, yang pada dataset ini mencapai lebih dari seribu pengunjung, dengan jumlah petugas dan jam layanan yang terbatas. Keputusan mengenai penjadwalan petugas, sasaran program literasi, dan tindak lanjut terhadap siswa yang jarang hadir karena itu perlu didasarkan pada bukti yang dapat diperoleh dari sistem pencatatan yang sudah ada. Kajian sebelumnya lebih banyak membahas perpustakaan akademik, kerangka konseptual, atau segmentasi pelanggan komersial, sedangkan evaluasi terukur atas segmentasi log kunjungan perpustakaan sekolah yang menggabungkan kebaruan, intensitas, dan waktu kunjungan masih terbatas. Kesenjangan tersebut bersifat kontekstual dan metodologis karena hasil segmentasi juga perlu diperiksa konsistensinya serta ketergantungannya terhadap fitur berulang. Rujukan metodologis yang bersifat teknis dalam penelitian ini diposisikan sebagai alat untuk menjamin mutu segmen, bukan sebagai tujuan, sehingga setiap pilihan metode diterjemahkan ke pertanyaan operasional pengelola: siapa yang perlu ditindaklanjuti, kapan petugas perlu disiagakan, dan seberapa sering segmen perlu diperbarui.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk (1) membandingkan K-Means dan K-Medoids pada K=2–8 dalam membentuk segmen perilaku operasional kunjungan perpustakaan SMAN 48 Jakarta, (2) menentukan partisi yang paling sesuai berdasarkan kriteria validitas yang dinyatakan sejak awal, serta (3) menjelaskan karakteristik segmen dan menerjemahkannya menjadi rekomendasi layanan. Kontribusi penelitian ini terdiri atas tiga aspek. Secara metodologis, penelitian menyajikan protokol evaluasi yang menggabungkan pembandingan 14 konfigurasi, verifikasi ulang metrik dari label tersimpan, uji kestabilan inisialisasi K-Means, dan analisis sensitivitas fitur berulang sehingga dapat direplikasi oleh perpustakaan sekolah lain. Secara empiris, penelitian memberikan profil segmen pengunjung berbasis 1.033 profil dan 10.050 transaksi yang masih jarang tersedia dalam literatur perpustakaan sekolah. Secara praktis, penelitian menghasilkan rekomendasi yang dapat ditindaklanjuti pengelola, seperti penjadwalan petugas pada interval kedatangan tinggi, verifikasi dan pendampingan pengunjung berintensitas rendah, serta siklus pembaruan segmen per semester. Penelitian ini tidak mengklaim penemuan algoritma baru atau menjadi studi pertama mengenai pengelompokan pemustaka.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian menggunakan pendekatan kuantitatif deskriptif dan komparatif terhadap data sekunder. Kerangka CRISP-DM digunakan untuk mengurutkan pemahaman kebutuhan layanan, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyajian hasil. Implementasi kerangka pada penelitian ini diperlihatkan dalam Gambar 1. Tahap pemahaman kebutuhan berfokus pada pengenalan variasi aktivitas pemustaka. Tahap pemahaman data memeriksa jumlah transaksi, rentang waktu, kelengkapan atribut, dan konsistensi identitas. Tahap persiapan data membentuk satu baris per pengunjung serta representasi numerik perilaku. Tahap pemodelan menelaah konfigurasi K-Means dan K-Medoids pada K=2–8. Tahap evaluasi mencakup pemeriksaan kembali metrik, kestabilan hasil, dan sensitivitas fitur. Penyajian hasil berbentuk tabel, grafik, serta rekomendasi operasional; penelitian ini tidak melakukan pengujian penerapan rekomendasi di lapangan.



Gambar 1. Tahapan penelitian dan pemeriksaan hasil

Pemrosesan membedakan hasil model yang sudah tersimpan dalam dataset dengan penghitungan tambahan untuk memeriksa keterulangannya. Pemisahan ini menjaga keterlacakan bukti, sebagaimana ditekankan dalam kerangka pembandingan algoritma Gagolewski [13] dan pedoman praktik pembandingan Van Mechelen dan kolega [14]. Nilai seluruh konfigurasi K=2–8 berasal dari lembar Evaluasi Model. Label K-Means K=2 dan K-Medoids K=5 berasal dari

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2627
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2625-2638
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1412

lembar Hasil Segmentasi. Statistik deskriptif dihitung kembali dari transaksi dan fitur. Verifikasi metrik menggunakan label tersebut, sementara pengujian sepuluh inisialisasi dan penghapusan fitur hari aktif merupakan analisis tambahan. Dengan demikian, hasil tambahan tidak disajikan seolah-olah merupakan catatan prosedur komputasi awal yang seluruh parameternya terdokumentasi.

2.2 Sumber Data dan Unit Analisis

Sumber penelitian adalah catatan kunjungan perpustakaan SMAN 48 Jakarta pada 18 Agustus 2025–7 Oktober 2026. Sebanyak 10.050 transaksi digunakan tanpa penarikan sampel acak. Unit pencatatan adalah kunjungan, sedangkan unit pengelompokan adalah pengunjung. Dataset menyediakan 1.033 kode pengunjung yang berkorespondensi dengan profil agregat. Identifikasi awal pada rancangan sumber mengutamakan NIS atau NIP. Penelitian ini mempertahankan pemetaan anonim yang tersedia, memeriksa keunikan kode, dan merekonsiliasi jumlah frekuensi profil terhadap total transaksi. Nama, nomor identitas, kontak, serta alamat tidak dimasukkan ke matriks model maupun tabel publikasi. Tidak ada pengunjung dengan frekuensi nol dalam dataset; karena itu, hasil tidak mencakup warga sekolah yang tidak pernah tercatat berkunjung.

Pemeriksaan kualitas meliputi keunikan ID kunjungan, kelengkapan jenis pengunjung dan jenis kelamin, serta distribusi jam pencatatan. Kategori tidak diketahui dipertahankan agar jumlah agregat tetap utuh. Transaksi pada jam yang jarang muncul tidak otomatis dihapus karena jadwal layanan resmi dan *log* sistem pembanding tidak tersedia. Perlakuan tersebut menghindari penghapusan pengamatan yang belum terbukti salah, tetapi konsekuensinya dibahas sebagai keterbatasan. Penelitian tidak menganggap seluruh hari tanpa transaksi sebagai hari tutup. Tanpa kalender operasional, pembagi berupa jumlah hari layanan efektif tidak dapat ditetapkan secara pasti. Agustus 2025 dan Oktober 2026 juga merupakan bulan pengamatan parsial sehingga totalnya tidak langsung dibandingkan sebagai satu bulan penuh.

2.3 Rekayasa Fitur dan Standardisasi

Sembilan fitur numerik digunakan untuk merepresentasikan pengunjung, sebagaimana dirangkum dalam Tabel 1. *Recency* dihitung sebagai selisih waktu penuh dalam satuan hari antara waktu acuan dan kunjungan terakhir. Waktu acuan ditetapkan pada transaksi terakhir, yaitu 7 Oktober 2026 pukul 07.26.51. Penghitungan tersebut cocok dengan seluruh nilai *recency* pada 1.033 profil. Nilai nol berarti selisih kurang dari 24 jam, sehingga tidak selalu berarti transaksi terjadi pada tanggal kalender yang sama. *Frequency* adalah jumlah transaksi, sedangkan hari aktif merupakan jumlah tanggal kunjungan yang berbeda pada profil. Keragaman hari menyatakan banyaknya jenis hari dalam sepekan yang tercatat. Rata-rata jam dan proporsi pagi serta siang merangkum preferensi waktu yang tersimpan dalam dataset.

Tabel 1. Definisi sembilan fitur dan transformasi model

Fitur	Makna pada profil pengunjung	Satuan	ln(1+x)
Recency	Selisih hari penuh dari kunjungan terakhir ke waktu acuan	Hari	Ya
Frequency	Jumlah transaksi per pengunjung	Transaksi	Ya
Hari aktif	Jumlah tanggal kunjungan berbeda pada profil	Hari	Ya
Rata-rata jam	Rerata waktu kunjungan dalam jam desimal	Jam	Tidak
Proporsi pagi	Proporsi kunjungan pagi tersimpan	0–1	Tidak
Proporsi siang	Proporsi kunjungan siang tersimpan	0–1	Tidak
Keragaman hari	Banyak jenis hari kunjungan dalam sepekan	1–5	Tidak
Rata-rata jeda	Rerata jeda yang tersedia pada profil	Hari	Ya
Simpangan baku jeda	Simpangan baku jeda yang tersedia pada profil	Hari	Ya

Sumber: Data Fitur dan rekonstruksi matriks model. Seluruh fitur distandardisasi dengan simpangan baku sampel.

Lembar sumber menyediakan pula proporsi sore, tanggal kunjungan pertama, dan tanggal kunjungan terakhir. Proporsi sore tidak dimasukkan ke sembilan fitur model karena proporsi pagi, siang, dan sore berjumlah satu pada seluruh profil. Dengan mempertahankan dua proporsi, komponen ketiga tetap dapat diturunkan. Batas jam pembentuk kategori pagi dan siang tidak dijelaskan secara eksplisit dalam metadata sumber; analisis mempertahankan nilai tersimpan dan tidak mengarang ambang pembentuknya. Demikian pula, rata-rata dan simpangan baku jeda digunakan sesuai profil yang tersedia. Pada 97 pengunjung yang hanya memiliki satu kunjungan, kedua statistik jeda bernilai nol, meskipun jeda antarkunjungan sebenarnya belum teramati. Konsekuensi pengodean ini dianalisis pada pembahasan.

Transformasi ln(1+x) diterapkan pada *recency*, *frequency*, hari aktif, rata-rata jeda, dan simpangan baku jeda. Empat fitur lainnya dipertahankan pada skala asal sebelum standardisasi. Seluruh fitur kemudian dipusatkan menggunakan rerata dan dibagi simpangan baku sampel dengan penyebut n−1, mengikuti representasi yang mereproduksi metrik dan SSE model tersimpan. Persamaan (1) menyatakan transformasi dan standardisasi. Parameter pemusatan serta skala dihitung dari 1.033 profil yang sama. Prinsip representasi, standardisasi, dan pembelajaran tanpa label mengikuti pembahasan James dan kolega [15] serta Murphy [16]. Standardisasi ini digunakan untuk deskripsi seluruh dataset, bukan untuk mengklaim performa prediksi terhadap pengunjung baru.

$$u_{ij} = ln(1 + x_{ij}) \; atau \; x_{ij}; \quad z_{ij} = \frac{u_{ij} - \mu_j}{s_j} \tag{1}$$

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2628
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2625-2638
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1412

Pada Persamaan (1), u_{ij} merupakan nilai fitur sesudah transformasi yang sesuai, μ_j adalah rerata fitur, dan s_j adalah simpangan baku sampel. Penerapan logaritma ditujukan untuk mengurangi pengaruh rentang nilai yang sangat lebar pada fitur hitungan dan jeda. Transformasi tidak mengubah status nilai nol sebagai nilai yang sudah dikodekan dalam sumber. Fitur demografis tidak digunakan dalam matriks jarak dan hanya dimanfaatkan untuk mendeskripsikan komposisi kelompok. Pendekatan ini mencegah kategori identitas berfungsi sebagai penentu partisi sekaligus mempertahankan kemungkinan pemeriksaan konteks layanan setelah kelompok terbentuk.

Pemilihan transformasi dan standardisasi didasarkan pada tiga pertimbangan statistik. Pertama, fitur hitungan dan jeda memiliki distribusi menceng ke kanan; frekuensi berkisar satu sampai 80 transaksi dengan median delapan dan rerata 9,73, sedangkan *recency* memiliki median 20 hari tetapi rerata 98,41 hari. Pada distribusi seperti ini, jarak Euclidean pada skala asal akan didominasi segelintir pengunjung bernilai ekstrem. Kedua, transformasi ln(1+x) dipilih alih-alih ln(x) karena tetap terdefinisi pada nilai nol: nilai nol dipetakan menjadi nol, urutan nilai dipertahankan, dan selisih pada nilai besar dimampatkan. Dengan demikian, fitur yang banyak bernilai nol tidak memerlukan penghapusan pengamatan atau penambahan konstanta arbitrer. Ketiga, standardisasi *Z-score* diterapkan sesudah transformasi agar setiap fitur memberi kontribusi berskala sebanding pada jarak, sejalan dengan temuan Wongoutong [9] mengenai pentingnya penskalaan. Karena *Z-score* merupakan transformasi linear, standardisasi tidak mengubah bentuk distribusi maupun urutan pengamatan, melainkan hanya menyetarakan skala. Kombinasi ini juga tidak menghilangkan sifat nilai nol sebagai hasil pengodean; nilai nol pada statistik jeda tetap membentuk titik massa setelah transformasi. Oleh karena itu, pengaruh pengodean nol tidak diasumsikan netral, melainkan diperiksa secara eksplisit melalui profil klaster dan analisis sensitivitas pada Subbab 3.5 dan 3.6.

2.4 Pembandingan Model dan Kriteria Evaluasi

K-Means meminimalkan jumlah kuadrat jarak pengunjung terhadap *centroid*, sedangkan K-Medoids meminimalkan jumlah jarak terhadap wakil kelompok yang merupakan pengamatan aktual. Persamaan (2) dan (3) menyatakan kedua fungsi objektif. Evaluasi mencakup tujuh nilai K untuk masing-masing algoritma. Nilai *cost* K-Medoids dan SSE K-Means tidak dibandingkan secara langsung karena keduanya mengoptimalkan besaran berbeda. Kriteria utama adalah rata-rata *Silhouette*, yang menilai hubungan jarak dalam kelompok dengan jarak ke kelompok terdekat. *Davies–Bouldin Index* menilai perbandingan sebaran dalam kelompok dengan jarak antarpusat. *Silhouette* yang lebih tinggi dan Davies–Bouldin yang lebih rendah menunjukkan preferensi masing-masing indeks pada representasi yang sama.

$$J_{KM} = \sum_{k=1}^{K} \sum_{i \in C_k} \|z_i - \mu_k\|^2 \tag{2}$$

$$J_{KD} = \sum_{k=1}^{K} \sum_{i \in C_k} d(z_i, m_k), \quad m_k \in C_k \tag{3}$$

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}; \quad S = \frac{1}{n} \sum_{i=1}^{n} s(i) \tag{4}$$

$$DBI = \frac{1}{K} \sum_{k=1}^{K} max_{l \neq k} \frac{S_k + S_l}{M_{kl}} \tag{5}$$

Dalam Persamaan (2)–(5), C_k menyatakan klaster, μ_k menyatakan *centroid*, m_k menyatakan medoid, dan d menyatakan jarak Euclidean. Pada *Silhouette*, a(i) adalah rata-rata jarak pengamatan i terhadap anggota klasternya, sedangkan b(i) adalah rata-rata jarak minimum terhadap klaster lain. Pada Davies–Bouldin, S_k adalah rata-rata jarak anggota terhadap *centroid* dan M_{kl} adalah jarak antarpusat. Nilai *Silhouette* 0,4114 tidak ditafsirkan sebagai akurasi 41,14% karena tidak tersedia label kebenaran. Pemilihan model bersifat relatif terhadap kandidat dan representasi yang diuji. Yerbury dan kolega [17] menunjukkan keterbatasan indeks validitas relatif untuk membandingkan paradigma kesamaan yang berbeda; karena itu, perubahan ruang fitur dibahas sebagai sensitivitas, bukan kompetisi skor yang sepenuhnya setara.

2.5 Verifikasi Hasil dan Analisis Sensitivitas

Verifikasi tambahan dilakukan menggunakan Python 3.12, NumPy 2.5.3, dan scikit-learn 1.9.1. Label tersimpan digunakan untuk menghitung kembali *Silhouette* dan Davies–Bouldin pada matriks terstandardisasi. K-Means K=2 kemudian dipasang ulang menggunakan inisialisasi k-means++, *n_init*=20, maksimum 300 iterasi, toleransi 0,0001, algoritma Lloyd, serta sepuluh nilai *random_state* dari 0 sampai 9. Parameter ini merujuk pada pengujian tambahan, bukan klaim mengenai konfigurasi awal yang tidak tersedia. Implementasi dan pengertian metrik mengikuti dokumentasi resmi scikit-learn [18]. Kesamaan keanggotaan terhadap label sumber diukur dengan *Adjusted Rand Index* atau ARI, yang tidak terpengaruh pertukaran nama klaster.

Analisis sensitivitas menghapus kolom hari aktif karena identik dengan *frequency* pada seluruh pengunjung, kemudian mengulangi K-Means K=2 pada delapan fitur dengan *random_state*=42 dan *n_init*=20. Pengujian ini memeriksa ketergantungan partisi terhadap pembobotan ganda intensitas kunjungan. Visualisasi PCA dua dimensi dihitung dari matriks sembilan fitur dan hanya digunakan untuk menampilkan partisi, bukan sebagai masukan pembentukan klaster. Pemeriksaan ini tidak mencakup *bootstrap*, validasi lintas sekolah, atau pengujian periode

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2625-2638
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1412

mendatang. K-Medoids tidak dipasang ulang karena pilihan implementasi, medoid awal, dan parameter penghentian model sumber tidak tersedia. Verifikasi terhadap metode tersebut terbatas pada metrik label K=5 yang disimpan.

Keandalan klaster dalam penelitian ini tidak disandarkan pada *Silhouette* dan *Davies–Bouldin Index* semata. Kedua indeks internal tersebut digunakan bersama tiga lapis bukti lain, yaitu (1) penghitungan ulang metrik dari label tersimpan untuk memastikan konsistensi antara fitur, transformasi, dan label, (2) kestabilan keanggotaan terhadap sepuluh pengulangan inisialisasi yang diukur dengan ARI, serta (3) kestabilan keanggotaan terhadap perubahan ruang fitur melalui penghapusan fitur hari aktif. Pendekatan berlapis ini mengikuti pembedaan validasi internal, validasi kestabilan, dan validasi pada data lain dalam kerangka Ullmann dan kolega [10], serta temuan Todeschini dan kolega [11] bahwa satu indeks tidak cukup mewakili kualitas partisi. Validasi berbasis pengambilan sampel ulang dan validasi pada periode atau sekolah lain belum dilakukan sehingga diperlakukan sebagai keterbatasan penelitian.

3. HASIL DAN PEMBAHASAN

3.1 Karakteristik dan Kualitas Data Kunjungan

Tabel 2 menyajikan karakteristik dataset yang dihitung kembali dari catatan sumber. Sebanyak 9.727 transaksi berasal dari siswa, 320 dari guru atau staf, dan tiga dari jenis pengunjung yang tidak diketahui. Pemeriksaan tidak menemukan pengulangan ID kunjungan. Hal ini menunjukkan keunikan identitas transaksi, tetapi tidak dengan sendirinya membuktikan tidak adanya pencatatan aktivitas yang serupa. Pada tingkat profil, seluruh 1.033 kode unik dan jumlah *frequency* tepat sama dengan 10.050 transaksi. Rekonsiliasi tersebut menghubungkan unit transaksi dan unit pengunjung sehingga kedua total tidak tertukar ketika proporsi layanan dibahas. Pengunjung yang tercatat lebih dari satu kali berjumlah 936 atau 90,61%, sedangkan 97 pengunjung hanya tercatat sekali.

Tabel 2. Karakteristik dan hasil pemeriksaan data

Indikator	Jumlah atau nilai	Keterangan
Transaksi kunjungan	10.050	18 Agustus 2025–7 Oktober 2026
Profil pengunjung	1.033	Kode unik; satu baris per pengunjung
Transaksi siswa	9.727 (96,79%)	Penyebut: seluruh transaksi
Transaksi guru/staf	320 (3,18%)	Penyebut: seluruh transaksi
Jenis tidak diketahui	3 (0,03%)	Dipertahankan dalam total
Jenis kelamin kosong	3.454 (34,37%)	Tidak menjadi fitur model
Pengunjung berulang	936 (90,61%)	Penyebut: seluruh pengunjung
ID kunjungan berulang	0	Bukan bukti tidak ada duplikasi aktivitas
Tanggal dengan transaksi	206	Bukan jumlah hari layanan resmi
Jam <06.00 atau ≥18.00	19 (0,19%)	Memerlukan verifikasi pencatatan

Sumber: penghitungan ulang Kunjungan 1 dan Data Fitur. Persentase dapat berbeda kecil akibat pembulatan.

Jenis kelamin kosong pada 3.454 transaksi atau 34,37% catatan. Karena kelengkapan tersebut tidak merata dan bukan fokus pembentukan klaster, jenis kelamin tidak digunakan sebagai fitur maupun dasar rekomendasi spesifik. Dataset memuat 206 tanggal dengan setidaknya satu transaksi. Angka tersebut menggambarkan tanggal yang memiliki aktivitas tercatat, bukan jumlah hari operasional resmi. Terdapat pula 19 transaksi sebelum pukul 06.00 atau sejak pukul 18.00, terdiri atas 14 transaksi pada jam 05, tiga pada jam 22, dan dua pada jam 23. Kejadian tersebut perlu dicocokkan dengan prosedur pencatatan sekolah sebelum dianggap mewakili jam kunjungan aktual. Pengamatan dipertahankan dalam analisis utama agar tidak mengubah populasi berdasarkan dugaan.

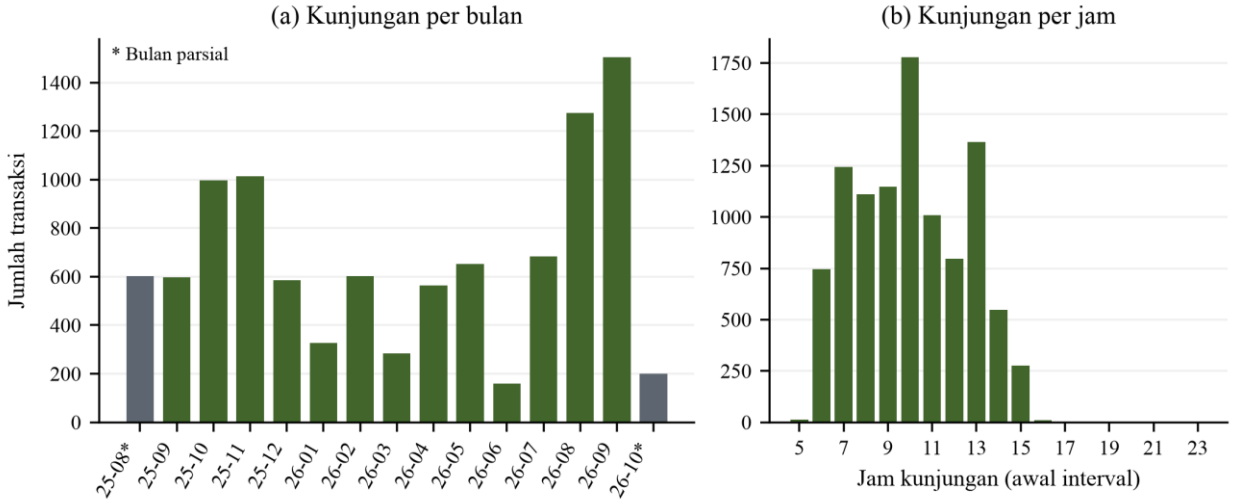
Frekuensi per pengunjung memiliki median delapan, rerata 9,73, dan rentang satu sampai 80 transaksi. *Recency* memiliki median 20 hari dan rerata 98,41 hari. Perbedaan antara rerata dan median memperlihatkan adanya pengunjung dengan jeda terakhir yang panjang. Fakta ini mendukung perlunya menyajikan lebih dari satu statistik pemusatan. Pada saat yang sama, pola tersebut tidak membuktikan alasan seseorang tidak hadir. Perubahan angkatan, masa pengamatan individu, perpindahan kelas, dan kebutuhan layanan dapat berhubungan dengan pola yang tercatat, tetapi variabel tersebut belum tersedia untuk diuji. Penelitian membatasi interpretasi pada aktivitas teramati dan tidak memberi label motivasi kepada individu.

3.2 Pola Kunjungan Menurut Waktu

Gambar 2 menggabungkan distribusi bulanan dan jam kunjungan. September 2026 merupakan bulan dengan total tertinggi, yaitu 1.504 transaksi, diikuti Agustus 2026 dengan 1.275 transaksi. Juni 2026 memiliki 160 transaksi. Perbedaan tersebut menggambarkan variasi volume pencatatan, tetapi tidak dapat langsung diatribusikan kepada libur sekolah, ujian, atau program tertentu karena kalender kegiatan tidak disertakan. Agustus 2025 hanya mencakup tanggal 18 sampai akhir bulan, sedangkan Oktober 2026 berakhir pada tanggal 7. Kedua bulan ditandai sebagai periode parsial pada grafik. Penyajian ini mencegah pembaca menafsirkan total Oktober yang rendah sebagai penurunan layanan selama satu bulan penuh.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2630
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2625-2638
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1412



Gambar 2. Distribusi transaksi menurut bulan dan jam kunjungan

Pada tingkat jam, interval 10.00–10.59 menyumbang 1.778 transaksi atau 17,69% dari keseluruhan kunjungan. Interval 13.00–13.59 menyumbang 1.365 transaksi atau 13,58%, sedangkan 07.00–07.59 menyumbang 1.245 transaksi atau 12,39%. Ketiga interval tersebut bersama-sama memuat 43,66% transaksi. Rabu menjadi hari dengan jumlah tertinggi, yaitu 2.602 transaksi atau 25,89%, diikuti Senin dengan 2.119 dan Selasa dengan 2.097 transaksi. Seluruh catatan berada pada Senin sampai Jumat. Distribusi ini dapat digunakan untuk mengidentifikasi waktu yang perlu diperiksa lebih lanjut oleh petugas. Namun, jumlah kedatangan bukan jumlah penghuni ruang secara bersamaan karena dataset tidak menyimpan waktu keluar atau lama tinggal.

Implikasi praktis pola waktu perlu dibedakan dari interpretasi kapasitas. Penambahan kesiapan petugas pada jam dengan kedatangan tinggi merupakan opsi perencanaan yang masuk akal, tetapi kebutuhan jumlah petugas atau kursi tidak dapat dihitung langsung dari *log* kedatangan. Penentuan kapasitas memerlukan durasi kunjungan, jumlah ruang, dan standar pelayanan. Demikian pula, grafik bulanan sebaiknya diperbarui bersama kalender akademik agar perubahan volume dapat dibandingkan pada periode yang sepadan. Pendekatan segmentasi dinamis Abbasimehr dan Bahrini [4] memberikan arah konseptual untuk pembaruan profil, sementara penelitian ini masih menggunakan satu jendela agregasi tetap sepanjang periode pengamatan.

3.3 Perbandingan K-Means dan K-Medoids

Hasil seluruh konfigurasi diperlihatkan dalam Tabel 3. *Silhouette* tertinggi di antara 14 kandidat adalah 0,4114 pada K-Means K=2. K-Means K=3 memperoleh 0,3583, sedangkan K=4 menghasilkan 0,2665. Pada K=5–8, nilainya berada pada kisaran 0,2922–0,3172. Pola ini tidak menunjukkan peningkatan pemisahan yang konsisten ketika jumlah klaster ditambah. Dengan kriteria utama *Silhouette*, partisi dua kelompok merupakan pilihan yang paling didukung dalam ruang sembilan fitur. Pemilihan tersebut juga menghasilkan ringkasan operasional yang mudah ditinjau, tetapi kemudahan penjelasan merupakan pertimbangan pendamping dan bukan pengganti ukuran kualitas yang telah ditetapkan.

Tabel 3. Evaluasi seluruh konfigurasi pengelompokan

Algoritma	K	*Silhouette* ↑	Davies–Bouldin ↓	Klaster terkecil	Klaster terbesar
K-Means	2	0,4114	1,1651	165	868
K-Means	3	0,3583	1,1547	120	761
K-Means	4	0,2665	1,3141	114	439
K-Means	5	0,2922	1,2255	116	271
K-Means	6	0,3064	1,1715	92	265
K-Means	7	0,3172	1,1430	50	231
K-Means	8	0,3167	1,1571	48	262
K-Medoids	2	0,2041	1,7426	483	550
K-Medoids	3	0,2086	1,5948	111	482
K-Medoids	4	0,2392	1,3916	110	369
K-Medoids	5	0,2825	1,2348	80	326
K-Medoids	6	0,2339	1,2469	97	271
K-Medoids	7	0,2567	1,2626	47	235
K-Medoids	8	0,2626	1,1362	26	252

Sumber: Evaluasi Model. Ukuran klaster dalam pengunjung; ↑ lebih tinggi lebih baik, ↓ lebih rendah lebih baik.

Untuk K-Medoids, *Silhouette* tertinggi adalah 0,2825 pada K=5. Nilai pada K=2 sebesar 0,2041 meningkat hingga K=5, kemudian berubah menjadi 0,2339, 0,2567, dan 0,2626 pada K=6, 7, dan 8. Selisih absolut antara kandidat terbaik

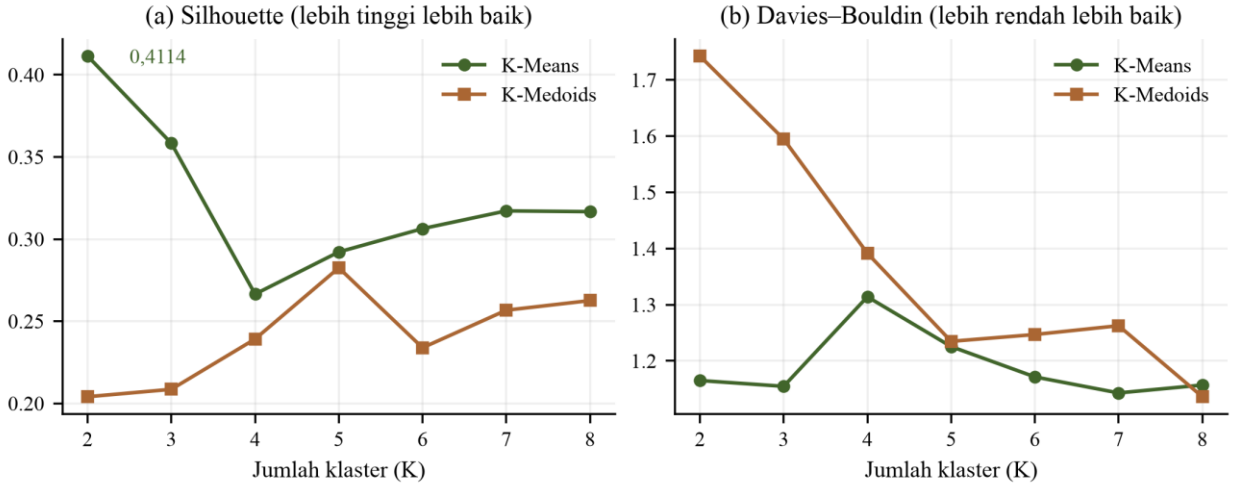
Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2631
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2625-2638
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1412

kedua algoritma adalah 0,1289. Pada K yang sama, K-Means memperoleh *Silhouette* lebih tinggi daripada K-Medoids untuk seluruh tujuh konfigurasi dalam dataset. Temuan ini mendukung pemilihan K-Means untuk representasi yang digunakan, tetapi belum menunjukkan keunggulan universal atau ketahanan yang lebih baik terhadap pencilan. Pengujian tersebut tidak memvariasikan metrik jarak, strategi inisialisasi K-Medoids, maupun perlakuan pencilan secara sistematis.

Gambar 3 menunjukkan bahwa preferensi Davies–Bouldin berbeda dari preferensi *Silhouette*. Davies–Bouldin terendah pada K-Means adalah 1,1430 untuk K=7, bukan 1,1651 pada K=2. Pada K-Medoids, nilai terendah adalah 1,1362 untuk K=8. Dengan demikian, konfigurasi K-Medoids K=8 justru memperoleh Davies–Bouldin terbaik di antara kandidat yang dicatat. Artikel ini memilih K-Means K=2 karena *Silhouette* dinyatakan sebagai kriteria utama, bukan karena konfigurasi tersebut mengungguli semua kandidat pada seluruh ukuran. Ketidakselarasan ini merupakan informasi penting: satu indeks lebih menonjolkan hubungan rata-rata jarak antaranggota, sedangkan indeks lain merangkum kemiripan antarklaster melalui pusat dan sebarannya.

Perbedaan fungsi objektif juga memengaruhi cara membaca keluaran optimasi. SSE K-Means K=2 pada sumber sebesar 6.637,0496 berhasil direproduksi dalam pengujian tambahan. *Cost* K-Medoids memiliki satuan objektif yang berbeda, sehingga angka yang lebih kecil tidak diperlakukan sebagai bukti langsung bahwa model tersebut lebih baik. Kurva SSE yang menurun saat K bertambah juga tidak cukup untuk menetapkan jumlah kelompok. Sikap ini konsisten dengan kritik Schubert [12] terhadap ketergantungan pada kriteria siku dan dengan hasil perbandingan indeks oleh Todeschini dan kolega [11]. Penggunaan beberapa ukuran dalam artikel berfungsi membuka konsekuensi pilihan, bukan mencari pembenaran bagi satu nilai K setelah hasil diketahui.



Gambar 3. Perbandingan indeks validitas untuk K=2–8

3.4 Karakteristik Segmen Pengunjung

Profil dua klaster disajikan pada Tabel 4. Klaster C1 memuat 165 pengunjung atau 15,97% populasi analitik, sedangkan C2 memuat 868 atau 84,03%. Frekuensi rata-rata C1 sebesar 1,53 transaksi dengan median satu; C2 memiliki rata-rata 11,29 dengan median sembilan. *Recency* rata-rata C1 sebesar 285,68 hari dengan median 335 hari, sedangkan C2 sebesar 62,81 hari dengan median 14,50 hari. Perbedaan ini menunjukkan bahwa intensitas dan kebaruan aktivitas merupakan ciri utama pemisahan. C1 menyumbang 253 transaksi atau 2,52% seluruh catatan, sedangkan C2 menyumbang 9.797 atau 97,48%. Proporsi transaksi berbeda dari proporsi pengunjung karena setiap orang dapat memiliki jumlah kunjungan yang berbeda.

Tabel 4. Profil klaster K-Means K=2 pada skala asal

Indikator	C1 intensitas rendah	C2 relatif lebih aktif
Pengunjung, n (%)	165 (15,97%)	868 (84,03%)
Transaksi, n (%)	253 (2,52%)	9.797 (97,48%)
Frequency, rerata; median	1,53; 1,00	11,29; 9,00
Frequency, minimum–maksimum	1–5	3–80
Recency, rerata; median (hari)	285,68; 335,00	62,81; 14,50
Recency, minimum–maksimum (hari)	0–413	0–376
Jam kunjungan, rerata desimal	10,20	10,66
Keragaman hari, rerata	1,28	3,23
Simpangan baku jeda, rerata (hari)	0,16	29,13
Pengunjung satu kali, n	97	0
Siswa; guru/staf; tidak diketahui	140; 25; 0	855; 12; 1

Sumber: Data Fitur dan Hasil Segmentasi. Persentase transaksi dan pengunjung menggunakan penyebut berbeda.

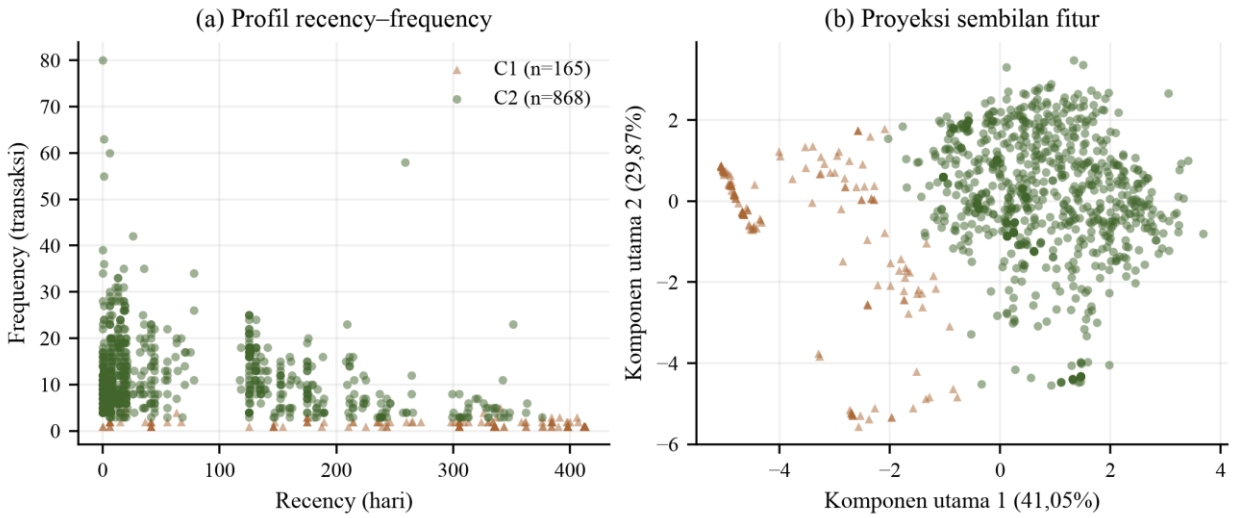
Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2632
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2625-2638
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1412

Nama segmen pada *workbook* adalah "Pasif / Lama Tidak Berkunjung" untuk C1 dan "Sangat Aktif" untuk C2. Untuk menghindari kesimpulan yang terlalu mutlak, artikel menggunakan deskripsi C1 berintensitas rendah dan C2 relatif lebih aktif. Pada C1 terdapat dua pengunjung dengan *recency* nol, sehingga tidak semua anggota kelompok ini lama tidak hadir. Sebaliknya, C2 memiliki *recency* hingga 376 hari dan frekuensi minimum tiga, sehingga tidak semua anggotanya dapat disebut sangat aktif pada akhir periode. Rentang frekuensi C1 adalah satu sampai lima, sedangkan C2 adalah tiga sampai 80. Adanya tumpang tindih pada frekuensi tiga sampai lima menegaskan bahwa keanggotaan dihasilkan dari beberapa fitur sekaligus, bukan ambang frekuensi tunggal.

Komposisi C1 terdiri atas 140 siswa dan 25 guru atau staf. C2 terdiri atas 855 siswa, 12 guru atau staf, serta satu pengunjung dengan jenis tidak diketahui. Pengamatan terakhir ini penting karena penjumlahan hanya dua kategori yang dikenal menghasilkan 867, sedangkan ukuran C2 adalah 868. Kategori tidak diketahui tetap dimasukkan ke total sehingga pelaporan konsisten. Perbedaan komposisi tidak digunakan untuk menyatakan bahwa kelompok jabatan tertentu memiliki motivasi yang lebih rendah. Durasi keterdaftaran, kebutuhan kerja, jadwal mengajar, dan akses terhadap layanan lain dapat berbeda. Variabel demografis tidak dimasukkan ke proses pengelompokan, sehingga hubungan yang terlihat bersifat deskriptif setelah partisi terbentuk.

Gambar 4 memperlihatkan sebaran *recency–frequency* dan proyeksi PCA. Dua komponen utama menjelaskan 70,92% variasi matriks terstandardisasi, terdiri atas 41,05% pada komponen pertama dan 29,87% pada komponen kedua. Proporsi ini membuat visualisasi cukup informatif untuk menggambarkan struktur umum, tetapi 29,08% variasi tetap berada di dimensi lain. Kedekatan titik pada proyeksi tidak dapat dianggap mewakili seluruh hubungan jarak sembilan dimensi. Titik di sekitar perbatasan kelompok perlu ditafsirkan bersama profil numerik, bukan dari warna grafik saja. PCA digunakan sesudah label ditetapkan dan tidak dipakai untuk memperbaiki skor pemisahan atau mengubah partisi sumber.



Gambar 4. Profil pengunjung dan proyeksi PCA menurut label K-Means K=2

3.5 Verifikasi Metrik dan Kestabilan Partisi

Penghitungan ulang terhadap label K-Means K=2 menghasilkan *Silhouette* 0,4113952718 dan Davies–Bouldin 1,1650648793, sesuai dengan angka sumber pada tingkat pembulatan yang dilaporkan. Pada K-Medoids K=5, hasilnya adalah 0,2825199899 dan 1,2348147542. Selisih numerik yang sangat kecil pada *Silhouette* berada sekitar orde 10^{-10} dan tidak memengaruhi empat angka desimal. Kecocokan ini menguatkan konsistensi antara fitur, transformasi, label, dan metrik. Namun, verifikasi metrik K-Medoids tidak setara dengan reproduksi seluruh proses optimasinya. Tanpa rincian implementasi awal, tidak dapat dipastikan apakah partisi tersebut merupakan hasil terbaik yang dapat dicapai semua variasi algoritma medoid.

Tabel 5. Verifikasi dan sensitivitas K-Means K=2

Pengujian	Ukuran klaster	*Silhouette*	DBI	ARI[1]
Sumber; sembilan fitur	165 / 868	0,4114	1,1651	1,0000
Pemodelan ulang; 10 seed[2]	165 / 868	0,4114	1,1651	1,0000
Tanpa hari aktif; delapan fitur	163 / 870	0,3896	1,2566	0,9705

[1] Dibandingkan dengan label sumber. [2] *random_state* 0–9, *n_init*=20 per pengulangan; seluruh hasil sama. Skor antar-ruang fitur tidak setara secara langsung.

Tabel 5 menunjukkan bahwa sepuluh pengulangan K-Means dengan *random_state* 0–9 dan 20 inisialisasi per pengulangan menghasilkan ARI 1,0000 terhadap partisi sumber. SSE dan *Silhouette* juga sama pada tingkat pelaporan. Hasil ini menunjukkan kestabilan terhadap inisialisasi yang diuji pada matriks tetap. Temuan tersebut tidak berarti partisi kebal terhadap pengambilan sampel ulang, perubahan jumlah pengunjung, atau pergeseran periode. Kestabilan inisialisasi

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

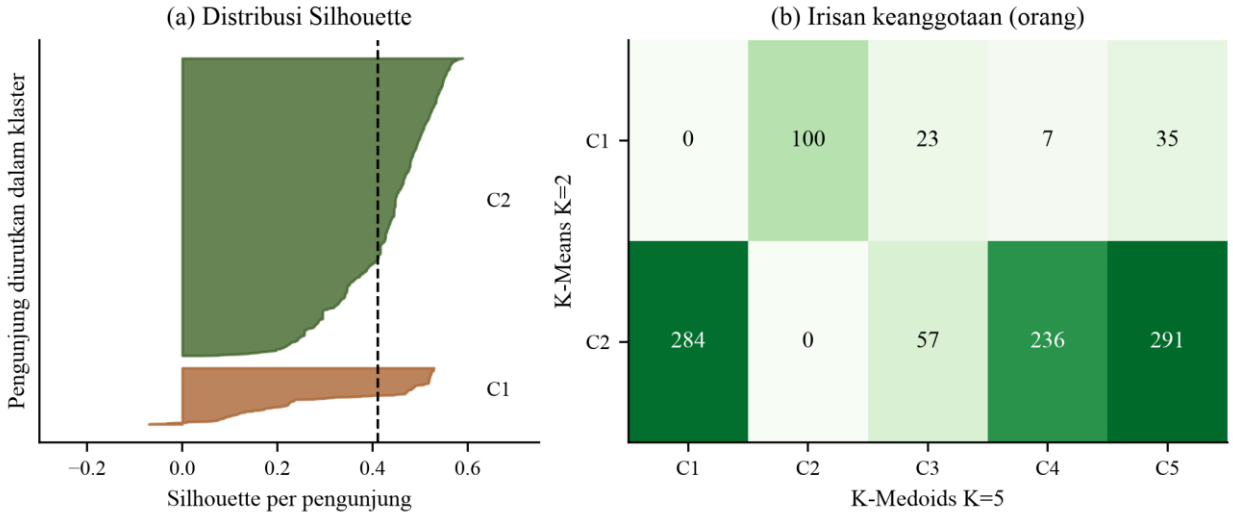
BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2625-2638
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1412

memeriksa satu sumber variasi komputasi, sedangkan kestabilan populasi dan waktu memerlukan rancangan berbeda. Pembedaan ini mengikuti kerangka validasi Ullmann dan kolega [10], yang menempatkan validasi pada data lain sebagai aspek tersendiri dari evaluasi internal.

Ketika hari aktif dihapus, K-Means K=2 menghasilkan ukuran kelompok 163 dan 870, *Silhouette* 0,3896, Davies–Bouldin 1,2566, dan ARI 0,9705 terhadap partisi awal. Hasil tersebut menunjukkan bahwa sebagian besar struktur keanggotaan bertahan meskipun pembobotan ganda frekuensi dihilangkan. Akan tetapi, penurunan *Silhouette* tidak boleh langsung ditafsirkan sebagai bukti bahwa mempertahankan fitur identik lebih benar. Ruang kesamaan telah berubah dari sembilan menjadi delapan dimensi. Yerbury dan kolega [17] mengingatkan bahwa indeks relatif bukan alat yang netral untuk memilih seluruh unsur paradigma kesamaan. Pada penerapan lanjutan, delapan fitur dapat dipertimbangkan demi menghindari penggandaan kontribusi, dengan keputusan akhir ditinjau bersama tujuan layanan dan validasi tambahan.

3.6 Perbedaan Resolusi Segmentasi dan Pengaruh Pengodean Fitur

Gambar 5 memperlihatkan hubungan label K-Means K=2 dengan K-Medoids K=5. Seratus anggota C1 K-Means masuk ke C2 K-Medoids, 35 ke C5, 23 ke C3, dan tujuh ke C4. Sebanyak 868 anggota C2 K-Means tersebar menjadi 284 pada C1 K-Medoids, 57 pada C3, 236 pada C4, serta 291 pada C5. Nama C1 sampai C5 hanya pengenal lokal masing-masing model dan tidak menunjukkan urutan tingkat aktivitas yang setara. Oleh karena itu, C1 pada K-Means tidak boleh langsung disamakan dengan C1 pada K-Medoids. Matriks tersebut memperlihatkan perbedaan resolusi pembagian, bukan matriks kesalahan klasifikasi terhadap label kebenaran.



Gambar 5. Distribusi *Silhouette* dan irisan label kedua algoritma

Distribusi *Silhouette* pada Gambar 5 juga memperlihatkan lima pengunjung dengan nilai negatif. Nilai tersebut berarti bahwa rata-rata jarak pengunjung itu terhadap kelompok lain lebih kecil daripada terhadap kelompoknya sendiri. Jumlah yang kecil tidak menghapus kebutuhan meninjau anggota di sekitar perbatasan klaster. Garis putus-putus menunjukkan rata-rata *Silhouette* seluruh pengunjung, sehingga variasi kualitas keanggotaan tetap terlihat di balik satu angka ringkasan.

Salah satu sumber pemisahan yang perlu mendapat perhatian adalah statistik jeda. Seluruh 97 pengunjung dengan satu kunjungan berada pada C1, sehingga mencapai 58,79% anggota klaster tersebut. Rata-rata dan simpangan baku jeda mereka dikodekan nol. Padahal, nol pada kasus ini berarti tidak tersedia pasangan kunjungan untuk menghitung selisih, bukan perilaku berkunjung tanpa jeda. Secara keseluruhan, simpangan baku jeda bernilai nol pada 153 profil, termasuk profil dengan sangat sedikit interval. Rerata simpangan baku jeda C1 hanya 0,16 hari, sedangkan C2 sebesar 29,13 hari. Perbedaan besar tersebut dapat merefleksikan gabungan perilaku dan aturan pengodean. Karena itu, “keteraturan” tidak disimpulkan hanya dari simpangan baku yang kecil pada kelompok jarang hadir.

Analisis lanjutan sebaiknya membedakan jeda yang tidak teramati dari jeda yang teramati bernilai kecil. Alternatifnya ialah menambahkan indikator kecukupan jumlah kunjungan, memperlakukan pengunjung satu kali sebagai kelompok deskriptif awal, atau menghitung statistik jeda hanya pada profil yang memenuhi jumlah interval minimum. Alternatif tersebut belum dibandingkan pada penelitian ini, sehingga tidak dilaporkan sebagai perbaikan yang sudah terbukti. Demikian pula, frekuensi kumulatif dapat dipengaruhi oleh panjang kesempatan pengamatan. Pengunjung yang baru pertama kali tercatat pada akhir periode memiliki waktu lebih sedikit untuk mengumpulkan transaksi. Fitur frekuensi per lama keterdaftaran dapat dipertimbangkan apabila tanggal keanggotaan yang sebenarnya tersedia, karena tanggal kunjungan pertama belum tentu tanggal mulai menjadi warga sekolah.

3.7 Pembahasan

Pola kunjungan menurut waktu memberikan informasi operasional yang lebih dalam apabila dibaca bersama karakteristik segmen. Konsentrasi 43,66% transaksi pada interval 07.00, 10.00, dan 13.00 menunjukkan bahwa kedatangan pemustaka

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2634
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2625-2638
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1412

tidak tersebar merata, melainkan berkumpul pada beberapa jendela waktu yang kemungkinan berkaitan dengan jeda dalam jadwal pembelajaran. Dugaan tersebut masih perlu dicocokkan dengan jadwal pelajaran dan istirahat sekolah, tetapi implikasinya sudah jelas bagi pengelola: beban layanan bersifat memuncak sehingga penempatan petugas dengan intensitas yang sama sepanjang hari kurang efisien. Rabu yang menyumbang 25,89% transaksi juga menunjukkan bahwa perencanaan dapat dilakukan per hari, bukan hanya per jam. Pada tingkat segmen, rerata jam kunjungan C1 dan C2 relatif berdekatan, yaitu 10,20 dan 10,66, sedangkan keragaman hari berbeda tajam, yaitu 1,28 dan 3,23. Artinya, pembeda utama kedua segmen bukan jam kedatangan, melainkan keberulangan dan keluasan hari kunjungan. Pengunjung berintensitas rendah datang pada jam yang serupa dengan pengunjung aktif, tetapi belum menjadikan perpustakaan bagian dari rutinitas mingguan. Temuan ini menggeser fokus intervensi dari perubahan jam layanan menuju upaya membangun kebiasaan berkunjung berulang.

Keunggulan K-Means atas K-Medoids pada dataset ini dapat dijelaskan dari kesesuaian antara struktur data, fungsi objektif, dan ukuran evaluasi. Pertama, transformasi ln(1+x) dan standardisasi telah memampatkan nilai ekstrem sehingga keunggulan teoretis K-Medoids terhadap pencilan menjadi kurang menentukan. Kedua, struktur dominan dalam data adalah pemisahan berdasarkan intensitas dan kebaruan yang membentuk dua kelompok tidak seimbang berukuran 165 dan 868. *Centroid* K-Means dapat bergeser mengikuti massa data yang padat, sedangkan *medoid* terikat pada pengamatan aktual. Pada K=2, K-Medoids justru menghasilkan kelompok yang hampir seimbang, yaitu 483 dan 550, dengan *Silhouette* 0,2041 dan Davies–Bouldin 1,7426. Pola ini mengindikasikan bahwa K-Medoids membagi data menurut arah variasi lain yang tidak memisahkan kelompok jarang hadir secara tegas. Ketiga, *Silhouette* dan Davies–Bouldin dihitung dengan jarak Euclidean yang selaras dengan fungsi objektif jarak kuadrat K-Means, sedangkan K-Medoids meminimalkan jumlah jarak tanpa kuadrat. Keselarasan ini dapat memberi keuntungan sistematis bagi K-Means sehingga keunggulan yang diamati perlu dibaca sebagai keunggulan pada representasi dan ukuran yang digunakan, bukan keunggulan universal. Meskipun demikian, konsistensi keunggulan *Silhouette* K-Means pada seluruh tujuh nilai K menunjukkan bahwa perbedaan tersebut tidak disebabkan oleh satu konfigurasi kebetulan.

Analisis sensitivitas memberikan kesimpulan yang tegas mengenai kualitas model. Penghapusan hari aktif menurunkan *Silhouette* dari 0,4114 menjadi 0,3896 atau sekitar 5,3% dan menaikkan Davies–Bouldin dari 1,1651 menjadi 1,2566 atau sekitar 7,9%, sementara ARI tetap sebesar 0,9705. Kombinasi ketiga angka tersebut berarti bahwa keanggotaan segmen hampir tidak berubah, tetapi sebagian kekuatan pemisahan pada model sembilan fitur berasal dari pembobotan ganda intensitas kunjungan. Dengan kata lain, struktur dua segmen merupakan temuan yang substantif dan bukan artefak fitur berulang, sedangkan nilai *Silhouette* 0,4114 lebih tepat dibaca sebagai batas atas kualitas pemisahan. Meskipun skor antar-ruang fitur tidak setara secara langsung, nilai 0,3896 pada model delapan fitur masih lebih tinggi daripada seluruh konfigurasi lain yang dicatat dalam Tabel 3, sehingga keputusan memilih dua segmen tetap bertahan. Atas dasar itu, model delapan fitur direkomendasikan sebagai model operasional karena menghasilkan partisi yang hampir identik dengan representasi yang lebih bersih.

Dibandingkan penelitian Khademizadeh dan kolega [1], penelitian ini menggunakan kehadiran sebagai peristiwa utama, sehingga luaran yang paling tepat adalah segmentasi pemanfaatan ruang atau layanan, bukan rekomendasi judul buku. Hubungannya dengan kerangka Okunlaya dan kolega [2] terletak pada penggunaan analitik untuk membantu keputusan layanan. Akan tetapi, penelitian ini memberikan pengukuran empiris pada satu sekolah dan tidak mengevaluasi keseluruhan transformasi organisasi. Perbandingan tersebut memperjelas kontribusi yang spesifik: *log* sederhana dapat dipakai untuk mengenali variasi aktivitas, tetapi ketersediaan data membatasi jenis keputusan yang didukung. Klaim mengenai efektivitas koleksi atau keberhasilan akademik memerlukan hubungan dengan data lain yang belum dianalisis.

Dari sisi metode, hasil mendukung perhatian Ikotun dan kolega [5] pada karakteristik K-Means serta perhatian Wongoutong [9] pada penskalaan fitur, tanpa menganggap kedua kajian menjamin skor tertentu pada dataset ini. Pemilihan dua klaster berbeda dari optimasi jumlah kelompok berbasis *Medoid Silhouette* oleh Lenssen dan Schubert [8], karena penelitian ini mengevaluasi *Silhouette* biasa pada label dari dua algoritma dan tidak menjalankan algoritma optimasi *Medoid Silhouette*. Pengujian fitur berulang dan pelaporan perbedaan metrik juga sejalan dengan prinsip pembandingan yang dapat ditelusuri [13], [14]. Keunggulan *Silhouette* dalam satu kasus tetap membutuhkan pengujian pada sekolah dan periode lain sebelum digeneralisasikan.

Agregasi selama lebih dari satu tahun dapat menyembunyikan perpindahan pola penggunaan. Iglesias Vázquez dan Zseby [19] mengembangkan *Temporal Silhouette* untuk validasi pengelompokan aliran data ketika terjadi perubahan distribusi. Penelitian ini tidak menerapkan indeks tersebut karena menggunakan profil agregat statis, tetapi kajian itu menjelaskan mengapa evaluasi berkala diperlukan. Pemisahan yang terlihat stabil terhadap inisialisasi pada hari analisis belum tentu stabil setelah pergantian semester. Pembaruan triwulanan atau per semester dapat dijadikan rancangan evaluasi berikutnya, dengan jendela waktu yang dinyatakan secara eksplisit, dokumentasi perubahan fitur, serta pemeriksaan perpindahan anggota kelompok. Rancangan tersebut akan menguji keberlanjutan manfaat segmentasi, bukan sekadar mengulang grafik yang sama.

3.8 Implikasi Layanan dan Keterbatasan Penelitian

Bagi pengelola, C1 dapat menjadi titik awal peninjauan kebutuhan layanan setelah status keanggotaan dan kesempatan hadir diperiksa. Program pengenalan koleksi atau pengingat kegiatan dapat dirancang untuk kelompok yang masih aktif sebagai warga sekolah dan memang membutuhkan dukungan. Identitas individu tetap dikelola secara terbatas oleh pihak berwenang. C2 dapat menjadi dasar untuk meninjau kebutuhan layanan berulang, terutama pada interval kedatangan tinggi. Kedua rekomendasi tersebut merupakan usulan berbasis pola, bukan intervensi yang telah terbukti

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2635
This work is licensed under a Creative Commons Attribution 4.0 International License.

meningkatkan kunjungan. Evaluasinya memerlukan indikator terpisah, misalnya perubahan frekuensi pada periode sepadan, umpan balik pemustaka, atau penggunaan koleksi setelah program berlangsung.

Agar rekomendasi tersebut dapat dijalankan dengan sumber daya perpustakaan SMA yang umumnya terbatas, hasil segmentasi diterjemahkan menjadi empat langkah operasional. Pertama, jadwal petugas dan pustakawan siswa disusun mengikuti puncak kedatangan pada interval 07.00, 10.00, dan 13.00 serta hari Rabu, misalnya dengan memusatkan layanan sirkulasi dan bantuan penelusuran pada jendela tersebut dan menempatkan pekerjaan teknis seperti pengolahan koleksi di luar jam puncak. Kedua, daftar kode pengunjung C1 direkap setiap akhir bulan, dicocokkan dengan data siswa aktif oleh petugas berwenang, kemudian ditindaklanjuti secara kolektif melalui wali kelas atau guru mata pelajaran, misalnya dengan kunjungan kelas terjadwal atau penugasan berbasis koleksi perpustakaan, tanpa mempublikasikan identitas individu. Ketiga, segmentasi diperbarui setiap semester menggunakan ekspor log yang sama dan model delapan fitur sehingga perpindahan pengunjung dari C1 ke C2 dapat dipakai sebagai indikator sederhana keberhasilan program. Keempat, prosedur pencatatan diperbaiki, termasuk verifikasi transaksi di luar pukul 06.00–18.00 dan pengisian jenis kelamin yang kosong pada 34,37% catatan, agar kualitas data periode berikutnya lebih baik. Langkah-langkah ini tidak memerlukan perangkat lunak berbayar dan dapat dijalankan dengan lembar kerja serta skrip analisis yang telah digunakan dalam penelitian ini.

Tata kelola data merupakan bagian dari penerapan, bukan sekadar penghapusan nama pada gambar. Jones dan Hinchliffe [3] menunjukkan bahwa analitik perpustakaan berhubungan dengan kesiapan etis praktisi. Pedoman American Library Association [20] mendukung pembatasan penggunaan dan akses terhadap informasi sistem perpustakaan. Pada penelitian ini, publikasi hanya menyajikan kode dan statistik agregat. Namun, pengodean anonim dalam keluaran tidak berarti sumber asli sepenuhnya anonim karena masih mengandung identitas langsung. Akses, retensi, serta dasar penggunaan data harus ditetapkan institusi sebelum distribusi dataset. Data individual tidak boleh dipublikasikan hanya karena nama kolom telah diganti dengan kode.

Keterbatasan utama penelitian adalah cakupan satu sekolah, ketiadaan label eksternal, dan penggunaan satu periode agregasi. Detail awal optimasi K-Medoids dan batas kategori waktu belum tersedia, sehingga keterulangan penuh belum dapat dipastikan. Tidak tersedia kalender layanan, durasi kunjungan, data peminjaman, status keanggotaan, ataupun variabel hasil belajar. Pengodean nol pada jeda yang tidak teramati dan identitas frekuensi dengan hari aktif dapat memengaruhi struktur jarak. Pengujian tambahan hanya mencakup inisialisasi K-Means dan satu bentuk ablasi fitur. Oleh sebab itu, hasil paling kuat adalah konsistensi partisi deskriptif pada data yang dianalisis. Pengujian sebab-akibat, manfaat program, atau ketepatan prediksi di masa depan berada di luar bukti penelitian ini.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa segmentasi berbasis kebaruan, frekuensi, dan waktu kunjungan dapat merangkum perbedaan aktivitas pemustaka perpustakaan SMAN 48 Jakarta. Sebanyak 10.050 transaksi selama 18 Agustus 2025–7 Oktober 2026 direpresentasikan menjadi 1.033 profil pengunjung dan digunakan untuk membandingkan K-Means serta K-Medoids pada K=2–8. Berdasarkan *Silhouette* sebagai kriteria utama, K-Means K=2 menjadi kandidat terpilih dengan nilai 0,4114 dan *Davies–Bouldin Index* sebesar 1,1651. K-Medoids terbaik berdasarkan kriteria yang sama menggunakan K=5 dengan *Silhouette* 0,2825. Pilihan tersebut tidak merupakan kemenangan pada seluruh metrik karena Davies–Bouldin terendah terdapat pada konfigurasi lain. Keunggulan *Silhouette* K-Means pada seluruh nilai K berkaitan dengan struktur dua kelompok yang tidak seimbang serta keselarasan fungsi objektifnya dengan ukuran berbasis jarak Euclidean. Dua segmen K-Means terdiri atas 165 pengunjung berintensitas rendah dan 868 pengunjung relatif lebih aktif, masing-masing mencakup 15,97% dan 84,03% populasi analitik. Perbedaan rerata frekuensi, sebesar 1,53 dan 11,29 transaksi, serta rerata *recency*, sebesar 285,68 dan 62,81 hari, memberikan dasar kuantitatif untuk membedakan prioritas peninjauan layanan. Verifikasi tambahan mereproduksi metrik utama dan SSE model, sedangkan sepuluh pengulangan inisialisasi menghasilkan kesamaan keanggotaan sempurna terhadap partisi sumber. Penghapusan fitur hari aktif yang identik dengan frekuensi tetap mempertahankan struktur yang sangat mirip dengan ARI 0,9705, meskipun komposisi dan nilai indeks berubah. Penurunan *Silhouette* menjadi 0,3896 menegaskan bahwa struktur dua segmen bersifat substantif, tetapi sebagian kekuatan pemisahannya berasal dari pembobotan ganda sehingga model delapan fitur lebih layak digunakan secara operasional. Temuan penting lainnya adalah bahwa pengunjung satu kali memiliki statistik jeda nol yang merepresentasikan keterbatasan pengamatan, sehingga simpangan baku kecil tidak selalu menunjukkan keteraturan perilaku. Implikasi praktis penelitian bersifat operasional, yaitu penyusunan jadwal petugas dan pustakawan siswa mengikuti puncak kedatangan pada interval 07.00, 10.00, dan 13.00 serta hari Rabu, rekap bulanan pengunjung berintensitas rendah yang ditindaklanjuti secara kolektif melalui wali kelas setelah status keanggotaannya dikonfirmasi, pemutakhiran segmen setiap semester dengan model delapan fitur sebagai indikator perpindahan pengunjung, serta perbaikan prosedur pencatatan kunjungan. Rekomendasi tersebut perlu diuji melalui pelaksanaan program dan indikator hasil yang terpisah. Kontribusi penelitian terletak pada pembandingan yang transparan, rekonsiliasi data, dan interpretasi segmen yang membatasi klaim sesuai bukti. Penelitian berikutnya perlu melengkapi metadata pemrosesan, menguji perlakuan jeda yang tidak tersedia, menggunakan jendela waktu yang sepadan, dan melakukan validasi pada sekolah lain. Data durasi, peminjaman, kalender akademik, serta umpan balik pemustaka juga diperlukan agar segmentasi dapat dihubungkan secara lebih kuat dengan kebutuhan layanan. Hasil penelitian ini menjelaskan pola pemanfaatan yang

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2636
This work is licensed under a Creative Commons Attribution 4.0 International License.

tercatat dan tidak mengukur secara langsung minat baca, kepuasan, kemampuan akademik, ataupun penyebab rendahnya kunjungan.

5. PERNYATAAN PENULIS

5.1 Kontribusi Penulis

Konseptualisasi: R. S. A., E. S., dan N. A.;
Metodologi: R. S. A., dan E. S.;
Perangkat Lunak: R. A. S.;
Validasi: R. S. A., E. S., dan N. A.;
Analisis Formal: R. S. A., dan E. S.;
Investigasi: R. S. A.;
Sumber Daya: R. S. A., dan N. A.;
Kurasi Data: R. S. A.;
Penulisan Draf Awal: R. S. A.;
Penulisan, Peninjauan, dan Penyuntingan: R. S. A., E. S., dan N. A.;
Visualisasi: R. S. A., dan N. A.;
Seluruh penulis telah membaca dan menyetujui versi manuskrip yang telah diterbitkan.

5.2 Ketersediaan Data

Data yang disajikan dalam penelitian ini tersedia berdasarkan permintaan kepada penulis pertama.

5.3 Pendanaan

Penelitian ini tidak menerima pendanaan eksternal.

5.4 Pernyataan Dewan Peninjau Institusional

Tidak Berlaku (*Not applicable*)

5.5 Pernyataan Persetujuan Berdasarkan Informasi

Tidak Berlaku (*Not applicable*)

5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan pribadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI dalam Proses Penulisan

Para penulis menggunakan AI Generatif untuk meningkatkan kejelasan penulisan dalam artikel ini. Para penulis telah meninjau dan menyunting konten yang dihasilkan dengan bantuan AI serta bertanggung jawab penuh atas versi akhir publikasi.

REFERENCES

[1] S. Khademizadeh, Z. Nematollahi, and F. Danesh, “Analysis of book circulation data and a book recommendation system in academic libraries using data mining techniques,” Library & Information Science Research, vol. 44, no. 4, Art. no. 101191, 2022. doi: 10.1016/j.lisr.2022.101191.
[2] R. O. Okunlaya, N. Syed Abdullah, and R. A. Alias, “Artificial intelligence (AI) library services innovative conceptual framework for the digital transformation of university education,” Library Hi Tech, vol. 40, no. 6, pp. 1869-1892, 2022. doi: 10.1108/lht-07-2021-0242.
[3] K. M. Jones and L. J. Hinchliffe, “Ethical issues and learning analytics: Are academic library practitioners prepared?,” The Journal of Academic Librarianship, vol. 49, no. 1, Art. no. 102621, 2023. doi: 10.1016/j.acalib.2022.102621.
[4] H. Abbasimehr and A. Bahrini, “An analytical framework based on the *recency*, *frequency*, and *monetary* model and time series clustering techniques for dynamic segmentation,” Expert Systems with Applications, vol. 192, Art. no. 116373, 2022. doi: 10.1016/j.eswa.2021.116373.
[5] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, “K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data,” Information Sciences, vol. 622, pp. 178-210, 2023. doi: 10.1016/j.ins.2022.11.139.
[6] A. E. Ezugwu et al., “A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects,” Engineering Applications of Artificial Intelligence, vol. 110, Art. no. 104743, 2022. doi: 10.1016/j.engappai.2022.104743.
[7] E. Schubert and P. J. Rousseeuw, “Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms,” Information Systems, vol. 101, Art. no. 101804, 2021. doi: 10.1016/j.is.2021.101804.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

[8] L. Lenssen and E. Schubert, "*Medoid Silhouette* clustering with automatic cluster number selection," Information Systems, vol. 120, Art. no. 102290, 2024. doi: 10.1016/j.is.2023.102290.
[9] C. Wongoutong, "The impact of neglecting feature scaling in k-means clustering," PLOS ONE, vol. 19, no. 12, Art. no. e0310839, 2024. doi: 10.1371/journal.pone.0310839.
[10] T. Ullmann, C. Hennig, and A. L. Boulesteix, "Validation of cluster analysis results on validation data: A systematic framework," WIREs Data Mining and Knowledge Discovery, vol. 12, no. 3, Art. no. e1444, 2022. doi: 10.1002/widm.1444.
[11] R. Todeschini, D. Ballabio, V. Termopoli, and V. Consonni, "Extended multivariate comparison of 68 cluster validity indices. A review," Chemometrics and Intelligent Laboratory Systems, vol. 251, Art. no. 105117, 2024. doi: 10.1016/j.chemolab.2024.105117.
[12] E. Schubert, "Stop using the elbow criterion for k-means and how to choose the number of clusters instead," ACM SIGKDD Explorations Newsletter, vol. 25, no. 1, pp. 36-42, 2023. doi: 10.1145/3606274.3606278.
[13] M. Gagolewski, "A framework for benchmarking clustering algorithms," SoftwareX, vol. 20, Art. no. 101270, 2022. doi: 10.1016/j.softx.2022.101270.
[14] I. Van Mechelen et al., "A white paper on good research practices in benchmarking: The case of cluster analysis," WIREs Data Mining and Knowledge Discovery, vol. 13, no. 6, Art. no. e1511, 2023. doi: 10.1002/widm.1511.
[15] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning: With Applications in R, 2nd ed. Springer US, 2021. doi: 10.1007/978-1-0716-1418-1.
[16] K. P. Murphy, Probabilistic Machine Learning: An Introduction. MIT Press, 2022. [Online]. Available: https://mitpress.mit.edu/9780262046824/probabilistic-machine-learning/
[17] L. W. Yerbury, R. J. G. B. Campello, G. C. Livingston, Jr., M. Goldsworthy, and L. O'Neil, "On the Use of Relative Validity Indices for Comparing Clustering Approaches," ACM Transactions on Knowledge Discovery from Data, vol. 19, no. 8, pp. 1-53, 2025. doi: 10.1145/3748726.
[18] scikit-learn developers, "Clustering," 2026. Accessed: Oct. 9, 2026. [Online]. Available: https://scikit-learn.org/stable/modules/clustering.html
[19] F. Iglesias Vázquez and T. Zseby, "*Temporal silhouette*: validation of stream clustering robust to concept drift," Machine Learning, vol. 113, no. 4, pp. 2067-2091, 2024. doi: 10.1007/s10994-023-06462-2.
[20] American Library Association, "Library Privacy Guidelines for Library Management Systems," 2022. Accessed: Oct. 9, 2026. [Online]. Available: https://www.ala.org/advocacy/privacy/guidelines/library-management-systems

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.