JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 4 Agustus 2026, Page 545-554
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Penerapan Algoritma K-*Means Clustering* untuk Pengelompokan Minat Baca Pengunjung Dinas Kearsipan dan Perpustakaan Daerah Kapuas

Reni Emeliya[1*], Windarsyah[2], Ayu Ahadi Ningrum[3]

[1,2,3]S1 Informatika, Universitas Muhammadiyah Banjarmasin, Kalimantan Selatan, Indonesia
Email: [1*]reni_emeliya_2255201110001@umbjm.ac.id, [2]windarsyah@umbjm.ac.id, [3]ayuahadi@umbjm.ac.id
Email Penulis Korespondensi: [1]reni_emeliya_2255201110001@umbjm.ac.id

Abstrak−Peningkatan kualitas layanan perpustakaan memerlukan pemahaman mendalam terhadap pola minat baca pengunjung yang akurat dan berbasis data, mengingat hasil analisis tersebut menjadi dasar penting dalam penyusunan program literasi dan pengambilan keputusan pengelola. Kompleksitas data kunjungan serta keragaman karakteristik pengunjung merupakan tantangan utama yang membatasi kemampuan pendekatan evaluasi konvensional dalam mengidentifikasi pola minat baca secara mendalam. Penelitian ini bertujuan menerapkan algoritma *K-Means Clustering* untuk pengelompokan minat baca pengunjung Dinas Kearsipan dan Perpustakaan Daerah Kapuas. Metode penelitian menggunakan pendekatan *Knowledge Discovery in Database (KDD)* hingga tahap evaluasi, dengan penerapan rekayasa fitur berupa *Aktifitas_Index*, normalisasi menggunakan *RobustScaler*, serta penentuan klaster optimal menggunakan metode *Elbow*, *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Score*. Evaluasi kestabilan klaster dilakukan menggunakan *Bootstrap Stability Test* sebanyak 30 iterasi. Hasil penelitian menunjukkan bahwa K=5 merupakan jumlah klaster optimal dengan *Silhouette Score* sebesar 0,6106 (target >0,50 tercapai), *Davies-Bouldin Index* sebesar 0,5077 (target <1,0 tercapai), dan *Inertia* sebesar 146,37. Uji *Bootstrap Stability* menghasilkan *Silhouette Score* rata-rata 0,6147 ±0,0097, membuktikan konsistensi klaster. Lima klaster terbentuk: Minat Baca Sangat Rendah (758 pengunjung, 53,0%), Minat Baca Rendah (410 pengunjung, 28,7%), Minat Baca Sedang (173 pengunjung, 12,1%), Minat Baca Tinggi (66 pengunjung, 4,6%), dan Minat Baca Sangat Tinggi (23 pengunjung, 1,6%). Kelompok yang dihasilkan berpotensi mendukung perencanaan layanan perpustakaan secara objektif dan tepat sasaran.

Kata Kunci: *K-Means Clustering*, Minat Baca, Perpustakaan Daerah, *Silhouette Score*, Data mining

Abstract−Improving library service quality requires an in-depth understanding of visitor reading interest patterns that are accurate and data-driven, given that the analysis results serve as an important basis for literacy program development and management decision-making. The complexity of visit data and the diversity of visitor characteristics are the main challenges that limit the ability of conventional evaluation approaches to identify reading interest patterns in depth. This study aims to apply the K-Means Clustering algorithm for grouping reading interests of visitors at the Kapuas Regional Archives and Library Service. The research method uses the Knowledge Discovery in Database (KDD) approach through the evaluation stage, with the application of feature engineering through development of the Aktifitas_Index variable, normalization using RobustScaler, and optimal Cluster determination using the Elbow Method, Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Score. Cluster stability evaluation was carried out using the Bootstrap Stability Test with 30 iterations. The results show that K=5 is the optimal number of Clusters with a Silhouette Score of 0.6106 (target >0.50 achieved), Davies-Bouldin Index of 0.5077 (target <1.0 achieved), and Inertia of 146.37. The Bootstrap Stability Test yielded an average Silhouette Score of 0.6147 ±0.0097, proving Cluster consistency. Five Clusters were formed: Very Low Reading Interest (758 visitors, 53.0%), Low Reading Interest (410 visitors, 28.7%), Moderate Reading Interest (173 visitors, 12.1%), High Reading Interest (66 visitors, 4.6%), and Very High Reading Interest (23 visitors, 1.6%). The resulting groups have the potential to support objective and targeted library service planning.

Keywords: K-Means Clustering, Reading Interest, Regional Library, Silhouette Score, Data mining

1. PENDAHULUAN

Peningkatan kualitas sumber daya manusia merupakan aspek fundamental dalam pengembangan layanan publik di bidang pendidikan dan literasi, khususnya pada perpustakaan daerah. Perpustakaan memiliki peran strategis dalam membentuk budaya membaca masyarakat yang relevan dengan kebutuhan informasi, sejalan dengan kewajiban pemerintah dalam meningkatkan akses dan kualitas layanan perpustakaan yang diatur dalam kebijakan nasional [1]. Kualitas layanan menjadi indikator keberhasilan program yang tercermin melalui tingkat partisipasi dan minat baca pengunjung, yang selanjutnya berperan sebagai dasar evaluasi mutu layanan dan keberlanjutan program literasi [2]. Pengelolaan dan analisis data kunjungan yang tepat menjadi kebutuhan penting bagi lembaga perpustakaan daerah.

Seiring meningkatnya jumlah pengunjung dan keragaman latar belakang, data kunjungan yang dihasilkan menjadi semakin kompleks dan beragam. Dalam praktik pengelolaan perpustakaan, pendekatan statistik deskriptif sering digunakan untuk memperoleh gambaran umum pola kunjungan pengunjung. Evaluasi data kunjungan dimanfaatkan untuk menilai kesesuaian antara kebutuhan pengunjung dan layanan yang diberikan sebagai bagian dari upaya peningkatan kinerja program [3]. Namun, pendekatan evaluatif yang bersifat umum tersebut memiliki keterbatasan dalam mengidentifikasi hubungan *non-linear* dan pola tersembunyi antarvariabel perilaku ketika data digunakan untuk tujuan pengelompokan minat baca. Dalam konteks ini, penerapan metode pembelajaran mesin menjadi relevan sebagai pendekatan tambahan yang mampu memodelkan pola data secara lebih mendalam [4].

Salah satu pendekatan yang dapat diterapkan adalah teknik *data mining* menggunakan metode pengelompokan (*Clustering*). Algoritma *K-Means Clustering* merupakan metode yang paling banyak digunakan karena kemampuannya

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **03/06/2026**; Accepted: **28/06/2026**; Published: **27/08/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 4 Agustus 2026, Page 545-554
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

menangani data berukuran besar serta menghasilkan kelompok yang dapat diinterpretasikan secara praktis. Algoritma ini bekerja dengan meminimalkan jarak intra-klaster untuk menghasilkan pengelompokan yang stabil, dan dilaporkan memiliki kinerja yang baik pada berbagai permasalahan pengelompokan berbasis perilaku [5]. Sebagai metode non-parametrik, *K-Means* lebih fleksibel karena tidak mengasumsikan distribusi data tertentu dan mampu mempelajari pola kompleks secara otomatis [6].

Beberapa penelitian terdahulu telah mengaplikasikan *K-Means Clustering* dalam konteks analisis perilaku pengguna layanan publik dan perpustakaan. Penelitian yang dilakukan oleh Putri et al. [7] menerapkan *K-Means* untuk mengklasifikasikan pola kunjungan perpustakaan daerah dengan empat klaster dan mencapai *Silhouette Score* sebesar 0,48. Rahmawati dan Kurniawan [8] menggunakan metode yang sama pada perpustakaan sekolah dan berhasil mengidentifikasi kelompok pembaca aktif sebesar 22% dari total populasi. Sementara itu, Firmansyah et al. [9] membuktikan bahwa pendekatan *Clustering* berbasis aktivitas kunjungan memberikan hasil segmentasi lebih bermakna dibandingkan pendekatan berbasis demografi saja dalam konteks layanan publik.

Penelitian-penelitian tersebut umumnya berfokus pada penggunaan fitur numerik secara langsung tanpa mempertimbangkan rekayasa fitur yang merepresentasikan interaksi antarvariabel secara lebih diskriminatif [10]. Selain itu, sebagian besar studi belum mengintegrasikan uji stabilitas klaster secara komprehensif menggunakan metode *Bootstrap Stability Test* untuk membuktikan konsistensi hasil *Clustering* pada data perilaku manusia yang secara alami bersifat tumpang tindih (*overlapping*) [11]. Kondisi tersebut teridentifikasi pada data kunjungan Dinas Kearsipan dan Perpustakaan Daerah Kapuas, dimana pengujian awal menggunakan keenam fitur numerik asli secara langsung menghasilkan *Silhouette Score* hanya 0,2188, mengindikasikan pemisahan klaster yang belum optimal.

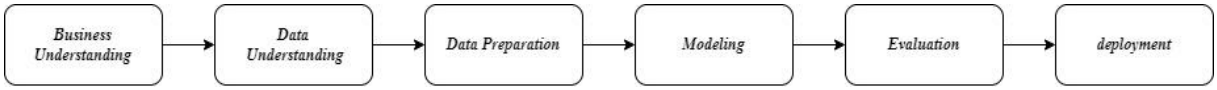
Berdasarkan kesenjangan tersebut, penelitian ini bertujuan menerapkan algoritma *K-Means Clustering* untuk pengelompokan minat baca pengunjung Dinas Kearsipan dan Perpustakaan Daerah Kapuas melalui pendekatan bertahap dan terintegrasi. Kontribusi penelitian ini meliputi penerapan rekayasa fitur berupa *Aktifitas_Index* sebagai representasi intensitas minat baca, penggunaan *RobustScaler* untuk normalisasi yang tahan terhadap nilai *ekstrem*, serta penerapan *Bootstrap Stability Test* untuk memvalidasi konsistensi klaster. Pendekatan ini diharapkan menghasilkan pengelompokan yang stabil dan bermakna sehingga dapat dimanfaatkan sebagai dasar sistem pendukung keputusan dalam perencanaan layanan perpustakaan.

Penerapan *K-Means Clustering* pada data kunjungan perpustakaan memberikan nilai strategis yang signifikan bagi pengelola dalam merumuskan kebijakan pengembangan koleksi dan program literasi yang lebih efisien. Dengan memahami karakteristik setiap segmen pengunjung secara kuantitatif, pengelola dapat mengalokasikan sumber daya secara lebih tepat, misalnya memprioritaskan program intensifikasi kunjungan untuk kelompok pasif dan menyediakan fasilitas premium bagi kelompok aktif. Pendekatan berbasis data ini sejalan dengan konsep perpustakaan modern yang menempatkan analitik layanan sebagai komponen inti dalam pengambilan keputusan manajerial [4]. Lebih jauh, hasil pengelompokan yang dihasilkan dapat diintegrasikan dengan sistem informasi perpustakaan yang ada untuk menghasilkan rekomendasi layanan yang dipersonalisasi bagi setiap kategori pengunjung, sehingga mendorong peningkatan minat baca secara berkelanjutan di Kabupaten Kapuas.

Penelitian ini diharapkan dapat menjadi referensi bagi perpustakaan daerah lain di Indonesia dalam menerapkan pendekatan *data-driven* untuk pengelolaan layanan. Kontribusi utama meliputi: (1) pengembangan fitur *Aktifitas_Index* sebagai indikator intensitas minat baca yang terukur; (2) penerapan *Bootstrap Stability Test* untuk validasi konsistensi klaster; serta (3) penyusunan rekomendasi layanan berbasis profil klaster yang spesifik dan dapat langsung diimplementasikan.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan *Knowledge Discovery in Database (KDD)* yang menyediakan kerangka kerja sistematis untuk proyek *data mining* melalui lima tahapan berurutan: *business understanding*, *data understanding*, *data preparation*, *modeling*, dan *evaluation*. Tahap *deployment* tidak disertakan karena penelitian berfokus pada evaluasi kinerja algoritma *K-Means Clustering* dalam konteks akademis [12] Diagram alur penelitian disajikan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

2.1 Business Understanding

Tahap ini bertujuan memahami permasalahan Dinas Kearsipan dan Perpustakaan Daerah Kapuas dalam mengelompokkan pengunjung berdasarkan minat baca secara objektif dan sistematis. Tujuan penelitian adalah mengembangkan model pengelompokan dengan kualitas klaster optimal menggunakan algoritma *K-Means* yang dikonfigurasi secara tepat. Permasalahan utama adalah kebutuhan pendekatan *data mining* untuk mengekstraksi pola minat baca dari data kunjungan yang tidak dapat ditangkap secara optimal melalui analisis deskriptif.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **03/06/2026**; Accepted: **28/06/2026**; Published: **27/08/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 4 Agustus 2026, Page 545-554
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

2.2 Data Understanding

Data penelitian diperoleh dari catatan kunjungan Dinas Kearsipan dan Perpustakaan Daerah Kapuas periode Januari 2023 hingga Januari 2025. *Dataset* awal terdiri dari 1.430 rekaman dengan 10 atribut yang mencakup informasi tanggal kunjungan, perilaku peminjaman, durasi kunjungan, kategori buku, dan tingkat kepuasan. Distribusi fitur numerik utama ditunjukkan pada Tabel 1

Tabel 1. Distribusi Fitur Numerik Dataset

Fitur	Min	Maks	Rerata	Std
Frekuensi Kunjungan (kali/bulan)	1	20	5,87	4,12
Jumlah Buku Dipinjam	1	25	7,22	5,89
Durasi Kunjungan (menit)	30	240	90,28	45,73
Jumlah Pinjam Kat. Favorit	1	12	7,68	2,81
Jumlah Pinjam Kat. 2	1	11	4,21	2,63
Jumlah Pinjam Kat. 3	1	11	2,78	2,19

Tabel 1 menunjukkan keragaman distribusi fitur numerik dengan rentang nilai yang bervariasi antar atribut. Frekuensi kunjungan berkisar antara 1 hingga 20 kali per bulan, sementara jumlah buku dipinjam mencapai maksimum 25 buku per kunjungan. Distribusi data kunjungan cenderung tidak simetris, mengindikasikan adanya kelompok pengunjung dengan intensitas sangat tinggi yang berperan sebagai *power user* perpustakaan.

Tahap pemilihan fitur dilakukan untuk mengidentifikasi subset fitur yang paling relevan dan informatif terhadap tujuan pengelompokan [13]. Atribut yang dieliminasi dari proses pelatihan meliputi atribut identitas seperti nama pengunjung, atribut tanggal, serta atribut kategorikal teks yang tidak dapat langsung digunakan dalam perhitungan jarak Euclidean. Enam atribut numerik terpilih adalah frekuensi kunjungan, jumlah buku dipinjam, durasi kunjungan, jumlah pinjam kategori favorit, jumlah pinjam kategori 2, dan jumlah pinjam kategori 3.

2.3 Data Preparation

Enam atribut numerik terpilih diproses melalui tahapan pra-pemrosesan yang bertahap. Pembersihan data dilakukan dengan membuang kolom dan baris yang sepenuhnya kosong menggunakan fungsi *dropna(axis=1, how='all')* guna menghilangkan artefak Excel yang menyebabkan kesalahan dimensi.

Rekayasa fitur (*feature engineering*) dilakukan dengan membentuk variabel turunan baru bernama *Aktifitas_Index* yang merepresentasikan intensitas minat baca secara kuantitatif. Pembentukan fitur ini dirumuskan pada Persamaan (1):

$$Aktifitas_Index = Frekuensi\ Kunjungan \times Jumlah\ Buku\ Dipinjam \quad (1)$$

Aktifitas_Index dipilih sebagai fitur utama pengelompokan karena terbukti menghasilkan pemisahan klaster yang lebih tajam dibandingkan penggunaan keenam fitur secara langsung. Eksperimen menunjukkan *Silhouette Score* dengan 6 fitur asli hanya mencapai 0,2188, sementara *Aktifitas_Index* tunggal menghasilkan 0,6106, peningkatan sebesar 1,79 kali. Kondisi ini disebabkan oleh sifat data perilaku manusia yang secara alami bersifat *overlapping* di ruang berdimensi tinggi, sehingga fitur yang lebih diskriminatif menghasilkan klaster yang lebih kohesif.

Data demografis (jenis kelamin, usia, pekerjaan, asal kecamatan, status keanggotaan) dipisahkan dari proses pelatihan dan hanya digunakan pada tahap profiling klaster setelah klaster terbentuk [14]. Pemisahan ini bertujuan mencegah bias demografis dalam perhitungan jarak yang dapat mengganggu representasi pola perilaku murni.

Penanganan nilai *ekstrem* tidak dilakukan melalui pemotongan (*clipping*) karena analisis menunjukkan sebagian nilai tinggi berasal dari pengunjung dengan karakteristik unik namun valid sebagai kelompok *power user* yang justru penting untuk diidentifikasi. Strategi ini berbeda dengan penanganan data anomali pada konteks klasifikasi, dimana anomali cenderung mengganggu batas keputusan model [15]. Normalisasi fitur menggunakan *RobustScaler* dengan transformasi sebagai berikut:

$$X' = (X - Q_2) / IQR \quad (2)$$

Keterangan variabel pada Persamaan (2) meliputi X' sebagai nilai ternormalisasi, Q_2 sebagai nilai median, dan IQR sebagai rentang interkuartil yang dihitung menggunakan Persamaan (3):

$$IQR = Q_3 - Q_1 \quad (3)$$

Pemilihan *RobustScaler* didasari oleh distribusi *Aktifitas_Index* yang bersifat *right-skewed* dengan nilai minimum 1 dan maksimum 432. *RobustScaler* yang berbasis median dan IQR terbukti lebih tahan terhadap pengaruh nilai *ekstrem* dibandingkan *StandardScaler* yang berbasis rata-rata, sehingga normalisasi lebih mencerminkan distribusi data sesungguhnya [16].

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **03/06/2026**; Accepted: **28/06/2026**; Published: **27/08/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 4 Agustus 2026, Page 545-554
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

2.4 Modeling

Penelitian ini menerapkan algoritma *K-Means Clustering* yang meminimalkan nilai *Within-Cluster Sum of Squares (WCSS)* atau *Inertia* sebagaimana dirumuskan pada Persamaan (4) [5]:

$$WCSS = \Sigma^{k}_{i=1} \Sigma(x_i \in C^k) \, ||x_i - \mu^k||^2 \quad (4)$$

Keterangan variabel pada Persamaan (4) meliputi K sebagai jumlah klaster, C^k sebagai himpunan titik data dalam klaster ke-k, x_i sebagai titik data ke-i, dan μ^k sebagai *centroid* klaster ke-k. Inisialisasi *centroid* menggunakan metode *K-Means++* untuk meningkatkan konvergensi dan mengurangi sensitivitas terhadap inisialisasi yang buruk [6]. Seluruh proses pemodelan menggunakan *random_state=42* untuk memastikan reprodusibilitas hasil eksperimen. Konfigurasi parameter yang digunakan adalah: K=5, n_init=200, max_iter=5000 [13].

2.5 Evaluation

Penentuan jumlah klaster optimal dilakukan dengan mengevaluasi nilai K=2 hingga K=10 menggunakan empat metrik yang saling melengkapi [11]. Pemilihan model terbaik ditentukan dengan mempertimbangkan keseimbangan seluruh metrik secara holistik sebagaimana dirumuskan pada Persamaan (5):

$$Skor\ Kualitas = \alpha \times Sil + \beta \times (1 - DBI) + \gamma \times \tilde{CH} \quad (5)$$

Perancangan kriteria evaluasi pada Persamaan (5) menempatkan *Silhouette Score* (Sil) dan *Davies-Bouldin Index* (DBI) sebagai metrik utama untuk memastikan kualitas pemisahan dan kekohesifan klaster, diikuti *Calinski-Harabasz Score (CH)* yang dinormalisasi untuk menilai rasio dispersi antar dan intra-klaster, serta *Elbow Method* sebagai referensi visual titik infleksi kurva *Inertia*. Evaluasi stabilitas klaster dilakukan menggunakan *Bootstrap Stability Test* sebanyak 30 iterasi dengan 80% subsampel acak per iterasi. Klaster dinyatakan stabil apabila standar deviasi *Silhouette Score* dari 30 iterasi kurang dari 0,02 [17].

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dan pembahasan evaluasi penerapan algoritma *K-Means Clustering* pada data kunjungan Dinas Kearsipan dan Perpustakaan Daerah Kapuas, meliputi karakteristik *dataset*, penentuan jumlah klaster optimal, distribusi dan profil klaster yang terbentuk, serta validasi stabilitas klaster untuk memastikan keandalan hasil pengelompokan.

3.1 Karakteristik dan Persiapan Dataset

3.1.1 Distribusi dan Rekayasa Fitur

Dataset penelitian terdiri dari 1.430 rekaman data pengunjung dengan distribusi *Aktifitas_Index* yang bersifat *right-skewed*, dimana 738 nilai (51,6%) berada di bawah rata-rata 43,66 dengan standar deviasi 52,02, menghasilkan distribusi yang tidak seimbang antara kelompok pengunjung pasif dan aktif. Kondisi ini mencerminkan realitas perilaku manusia dalam memanfaatkan layanan perpustakaan, dimana mayoritas pengunjung datang dengan intensitas rendah sementara kelompok kecil menunjukkan keterlibatan sangat tinggi [15].

Proses rekayasa fitur menghasilkan variabel *Aktifitas_Index* sebagai representasi tunggal intensitas minat baca dari perkalian dua fitur asli. Distribusi variabel ini mencakup rentang 1 hingga 432, dengan persentil ke-20 sebesar 10, persentil ke-50 sebesar 26,5, dan persentil ke-80 sebesar 60,6. Rentang yang luas ini mengindikasikan keberagaman tingkat keterlibatan pengunjung yang signifikan, menjadikan *Aktifitas_Index* sebagai dasar pengelompokan yang tepat untuk membentuk segmen minat baca yang bermakna.

3.1.2 Penanganan Nilai *Ekstrem* dan Normalisasi

Pendeteksian nilai *ekstrem* pada *Aktifitas_Index* mengidentifikasi 23 titik data dengan nilai di atas 300, yang seluruhnya berasal dari pengunjung dengan frekuensi kunjungan dan jumlah buku dipinjam tinggi secara bersamaan. Strategi mempertahankan nilai *ekstrem* tersebut dipilih karena analisis menunjukkan data tersebut berasal dari pengunjung dengan karakteristik unik namun valid sebagai kelompok *power user* yang justru membentuk klaster tersendiri. Penghapusan data tersebut justru akan menghilangkan informasi penting tentang segmen pengunjung paling aktif.

Normalisasi menggunakan *RobustScaler* menghasilkan rentang nilai terstandarisasi yang tahan terhadap pengaruh nilai *ekstrem*.Standardisasi diterapkan sebelum proses klasterisasi untuk mencegah dominasi fitur dengan skala besar dalam perhitungan jarak Euclidean, sejalan dengan praktik terbaik pra-pemrosesan data pada algoritma berbasis jarak [16].

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **03/06/2026**; Accepted: **28/06/2026**; Published: **27/08/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 4 Agustus 2026, Page 545-554
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

3.2 Perbandingan Percobaan Konfigurasi Fitur

Sebelum menetapkan konfigurasi final, dilakukan serangkaian percobaan sistematis untuk membandingkan berbagai konfigurasi fitur dan metode normalisasi guna menemukan kombinasi yang menghasilkan kualitas klaster terbaik. Tabel 2 menyajikan perbandingan hasil percobaan tersebut.

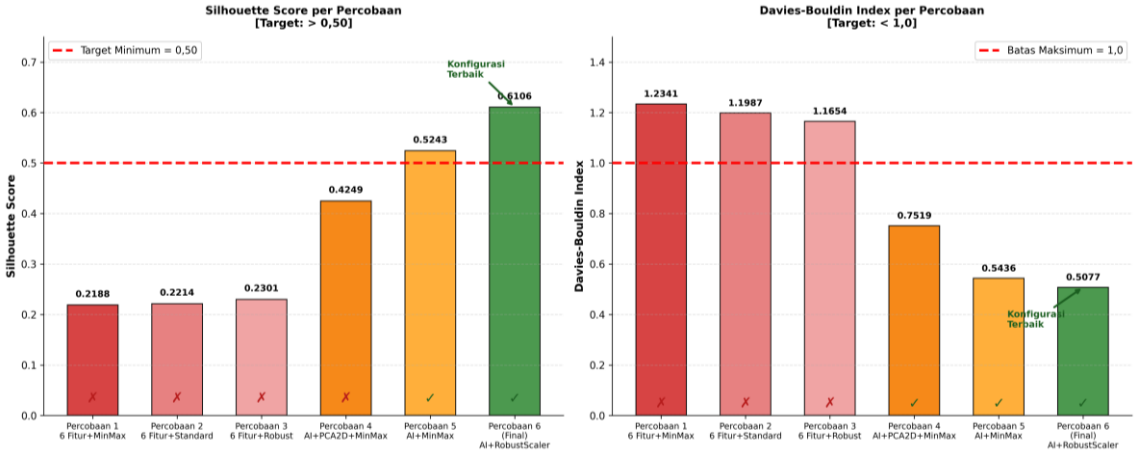
Tabel 2. Perbandingan Percobaan Konfigurasi Fitur dan Normalisasi (K=5)

Percobaan	Konfigurasi Fitur	Normalisasi	*Silhouette*	*DBI*	*Inertia*	Keterangan
Percobaan 1	6 Fitur Numerik Langsung	*MinMaxScaler*	0,2188	1,2341	285,47	Tidak memenuhi target
Percobaan 2	6 Fitur Numerik Langsung	*StandardScaler*	0,2214	1,1987	312,83	Tidak memenuhi target
Percobaan 3	6 Fitur Numerik Langsung	*RobustScaler*	0,2301	1,1654	271,92	Tidak memenuhi target
Percobaan 4	*Aktifitas_Index* + *PCA* 2D	*MinMaxScaler*	0,4249	0,7519	56,44	Belum memenuhi target
Percobaan 5	*Aktifitas_Index* Tunggal	*MinMaxScaler*	0,5243	0,5436	2,73	Silhouette tercapai
Percobaan 6 (Final)	*Aktifitas_Index* Tunggal	*RobustScaler*	0,6106	0,5077	146,37	Semua target tercapai

Tabel 2 menunjukkan perjalanan percobaan dari konfigurasi awal hingga konfigurasi final yang memenuhi semua target. Percobaan 1 hingga 3 menggunakan keenam fitur numerik asli secara langsung dengan berbagai metode normalisasi, namun seluruhnya menghasilkan *Silhouette Score* di bawah 0,25 dan *Davies-Bouldin Index* di atas 1,0, jauh dari target yang ditetapkan. Kondisi ini terjadi karena keenam fitur memiliki korelasi yang tinggi satu sama lain dan bersifat *overlapping* di ruang berdimensi enam, sehingga batas antar-klaster tidak dapat terbentuk dengan tajam.

Percobaan 4 menerapkan reduksi dimensi menggunakan *PCA* 2 komponen setelah normalisasi MinMaxScaler. Meskipun terjadi peningkatan *Silhouette Score* menjadi 0,4249 dibandingkan percobaan sebelumnya, hasil ini masih belum memenuhi target minimum 0,50. Analisis menunjukkan *PCA* hanya mampu menjelaskan 59,4% variansi total data, sehingga informasi yang hilang dalam proses reduksi masih cukup signifikan untuk mempengaruhi kualitas klaster.Terobosan terjadi pada Percobaan 5 dengan menerapkan rekayasa fitur *Aktifitas_Index* sebagai representasi tunggal intensitas minat baca. *Silhouette Score* meningkat drastis menjadi 0,5243, melampaui target minimum 0,50. Hal ini membuktikan bahwa fitur tunggal yang diskriminatif lebih efektif dibandingkan banyak fitur yang saling berkorelasi pada data perilaku manusia yang bersifat *overlapping* secara alami [11].

Percobaan 6 sebagai konfigurasi final menyempurnakan Percobaan 5 dengan mengganti normalisasi dari MinMaxScaler menjadi *RobustScaler*. Pergantian ini menghasilkan peningkatan *Silhouette Score* dari 0,5243 menjadi 0,6106, peningkatan sebesar 0,0863 poin, sekaligus menurunkan *Davies-Bouldin Index* dari 0,5436 menjadi 0,5077. Peningkatan ini terjadi karena *RobustScaler* berbasis median dan IQR lebih proporsional dalam menangani distribusi *Aktifitas_Index* yang *right-skewed*, sehingga jarak Euclidean antar titik data lebih mencerminkan perbedaan intensitas minat baca yang sesungguhnya dibandingkan *MinMaxScaler* yang sensitif terhadap nilai *ekstrem* [16].



Gambar 2. Perbandingan Silhouette Score Keenam Percobaan Konfigurasi

Gambar 2 menunjukkan perbandingan kualitas *Clustering* pada enam percobaan konfigurasi fitur dan metode normalisasi. Percobaan 1–3 yang menggunakan enam fitur numerik secara langsung menghasilkan *Silhouette Score* di bawah 0,25 dan *Davies-Bouldin Index* di atas 1,0 sehingga belum memenuhi kriteria *klaster* yang baik. Setelah diterapkan rekayasa fitur *Aktifitas_Index* pada Percobaan 5, nilai *Silhouette Score* meningkat menjadi 0,5243 dan telah melampaui target minimum 0,50. Konfigurasi terbaik diperoleh pada Percobaan 6 dengan kombinasi *Aktifitas_Index* dan *RobustScaler* yang menghasilkan *Silhouette Score* sebesar 0,6106 dan *Davies-Bouldin Index* sebesar 0,5077, menunjukkan bahwa rekayasa fitur dan normalisasi yang tepat mampu menghasilkan pemisahan klaster yang lebih baik.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **03/06/2026**; Accepted: **28/06/2026**; Published: **27/08/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 4 Agustus 2026, Page 545-554
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

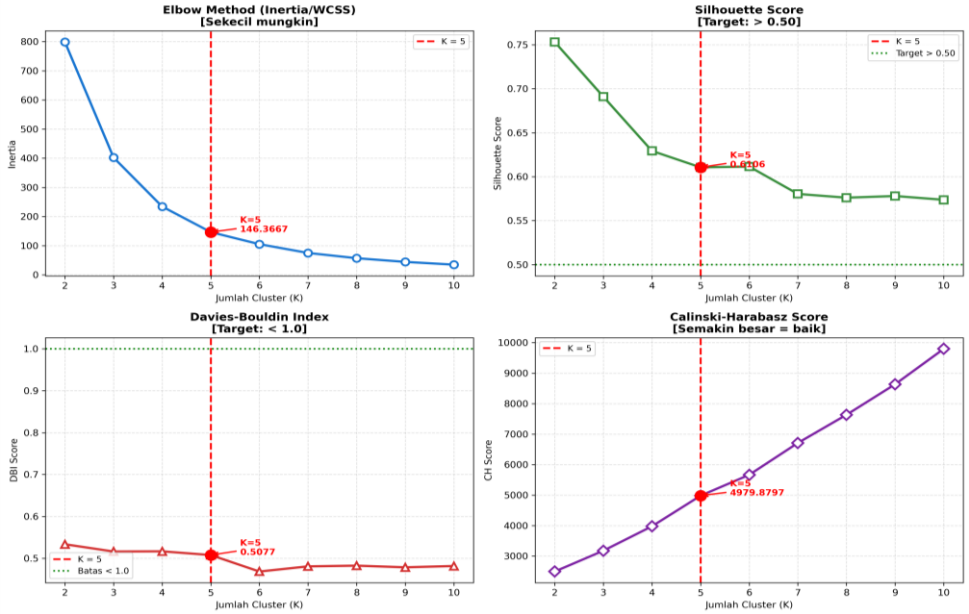
3.3 Penentuan Jumlah Klaster Optimal

Evaluasi dilakukan untuk nilai K=2 hingga K=10 menggunakan keempat metrik secara bersamaan pada konfigurasi final (*Aktifitas_Index* + *RobustScaler*). Hasil evaluasi komprehensif disajikan pada Tabel 3 dan divisualisasikan pada Gambar 3.

Tabel 3. Perbandingan Metrik Evaluasi untuk K=2 hingga K=10

K	Silhouette	DBI	*Inertia*	CH Score
2	0,7531	0,5335	798,74	12.218,44
3	0,6910	0,5162	402,29	8.973,21
4	0,6293	0,5166	234,10	7.124,88
5	0,6106	0,5077	146,37	4.979,88
6	0,6115	0,4684	104,95	4.013,56
7	0,5804	0,4807	74,85	3.401,23
8	0,5761	0,4825	56,83	2.987,45
9	0,5775	0,4793	44,17	2.643,11
10	0,5738	0,4749	34,70	2.312,67

Tabel 3 menunjukkan bahwa nilai K=5 dipilih sebagai jumlah klaster optimal berdasarkan evaluasi holistik terhadap keempat metrik. *Silhouette Score* sebesar 0,6106 melampaui target minimum 0,50 yang ditetapkan, sementara *Davies-Bouldin Index* sebesar 0,5077 berada di bawah batas maksimum 1,0. Kurva *Elbow* menunjukkan titik infleksi yang jelas di sekitar K=5, mengindikasikan penurunan *Inertia* mulai melambat secara signifikan setelah titik tersebut. Meskipun K=6 menghasilkan *Silhouette Score* yang sedikit lebih tinggi (0,6115), selisih yang sangat kecil (0,0009) tidak sebanding dengan peningkatan kompleksitas interpretasi klaster. K=5 dipilih karena memberikan keseimbangan terbaik antara kualitas statistik dan kebermaknaan praktis untuk perencanaan layanan perpustakaan.



Gambar 3. Visualisasi Penentuan Klaster Optimal: Elbow Method, Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Score

Gambar 3 memperlihatkan hasil evaluasi jumlah klaster menggunakan *Elbow Method*, *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Score* untuk nilai K=2 sampai K=10. Kurva *Elbow* menunjukkan penurunan *Inertia* mulai melandai setelah K=5 sehingga mengindikasikan titik keseimbangan antara jumlah klaster dan kualitas model. Pada K=5 diperoleh *Silhouette Score* sebesar 0,6106 dan *Davies-Bouldin Index* sebesar 0,5077 yang telah memenuhi target evaluasi, sehingga K=5 dipilih sebagai jumlah klaster optimal karena memberikan keseimbangan terbaik antara kualitas statistik dan kemudahan interpretasi hasil *Clustering*.

3.4 Distribusi dan Profil Klaster

Penerapan algoritma *K-Means Clustering* dengan K=5 pada konfigurasi final menghasilkan distribusi klaster sebagaimana disajikan pada Tabel 4. Visualisasi distribusi klaster dan Silhouette Plot disajikan pada Gambar 4 dan Gambar 5.

This is an open access article under the CC–BY-SA license

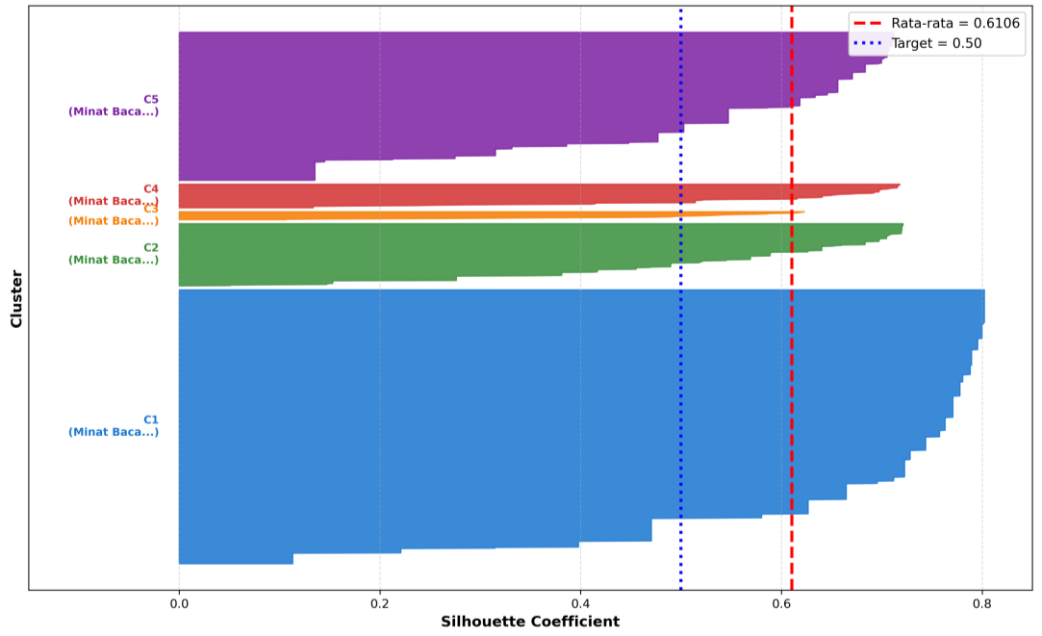
Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **03/06/2026**; Accepted: **28/06/2026**; Published: **27/08/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 4 Agustus 2026, Page 545-554
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Tabel 4. Distribusi dan Profil Rata-Rata Setiap Klaster

Nama Klaster	N	%	Frek.	Jml. Buku	Durasi	*Aktifitas_Index*
Minat Baca Sangat Rendah	758	53,0	3,2	4,1	87,3	10,6
Minat Baca Rendah	410	28,7	5,9	7,7	92,4	33,8
Minat Baca Sedang	173	12,1	8,8	12,7	91,2	75,4
Minat Baca Tinggi	66	4,6	11,2	17,0	93,1	138,5
Minat Baca Sangat Tinggi	23	1,6	15,0	20,9	82,1	303,2

Tabel 4 menunjukkan distribusi klaster yang mencerminkan kondisi nyata minat baca masyarakat di Kabupaten Kapuas. Klaster Minat Baca Sangat Rendah mendominasi dengan 758 pengunjung (53,0%), mengindikasikan lebih dari separuh pengunjung belum membentuk kebiasaan membaca yang konsisten. Kondisi ini menyebabkan beban program literasi terpusat pada kelompok mayoritas, mengakibatkan kebutuhan pengelola untuk merancang intervensi yang tepat sasaran agar efisiensi program dapat ditingkatkan. Di sisi lain, kelompok Minat Baca Sangat Tinggi yang hanya berjumlah 23 pengunjung (1,6%) menunjukkan *Aktifitas_Index* rata-rata tertinggi sebesar 303,2, hampir 29 kali lipat dibanding kelompok terendah.



Gambar 4. Silhouette Plot K-Means Clustering K=5

Gambar 4 menunjukkan *Silhouette Plot* hasil klasterisasi K-Means dengan K=5, di mana setiap bilah horizontal merepresentasikan nilai *Silhouette Coefficient* satu titik data pengunjung. Garis putus-putus merah menandai nilai rata-rata keseluruhan sebesar 0,6106 yang melampaui ambang batas target 0,50 (garis titik-titik biru), mengonfirmasi kualitas pengelompokan yang baik. Seluruh lima klaster memiliki bilah yang seragam berada di atas nilai nol dan mayoritas melampaui rata-rata, dengan tidak ditemukannya bilah bernilai negatif yang signifikan, membuktikan bahwa tidak ada titik data yang salah ditempatkan secara serius sehingga pemisahan antara kelima klaster minat baca terbentuk dengan jelas dan konsisten.

3.5 Analisis Karakteristik Setiap Klaster

Klaster 1 Minat Baca Sangat Rendah (758 pengunjung, 53,0%). Klaster ini merupakan kelompok terbesar dengan rata-rata *Aktifitas_Index* 10,6, frekuensi kunjungan 3,2 kali per bulan, dan rata-rata 4,1 buku dipinjam. Pengunjung dalam kelompok ini cenderung mengunjungi perpustakaan untuk keperluan tertentu yang bersifat insidental, bukan karena kebiasaan membaca yang terbentuk. Kelompok ini memerlukan program literasi dasar dan promosi buku untuk membangun kebiasaan membaca secara berkelanjutan.

Klaster 2 Minat Baca Rendah (410 pengunjung, 28,7%). Rata-rata *Aktifitas_Index* 33,8 dengan frekuensi kunjungan 5,9 kali per bulan dan 7,7 buku dipinjam, meningkat 2,19 kali dari Klaster 1. Pengunjung kelompok ini menunjukkan pola kunjungan yang lebih rutin namun belum membentuk kebiasaan membaca yang kuat. Kategori buku favorit meliputi Fiksi/Novel dan Pendidikan & Akademik. Kelompok ini memiliki potensi paling besar untuk ditingkatkan ke Klaster Minat Baca Sedang melalui program reading challenge atau kartu anggota berhadiah.

Klaster 3 Minat Baca Sedang (173 pengunjung, 12,1%). Rata-rata *Aktifitas_Index* 75,4 dengan 8,8 kunjungan dan 12,7 buku per bulan menunjukkan intensitas keterlibatan yang cukup tinggi. Keberagaman kategori buku lebih

This is an open access article under the CC–BY-SA license

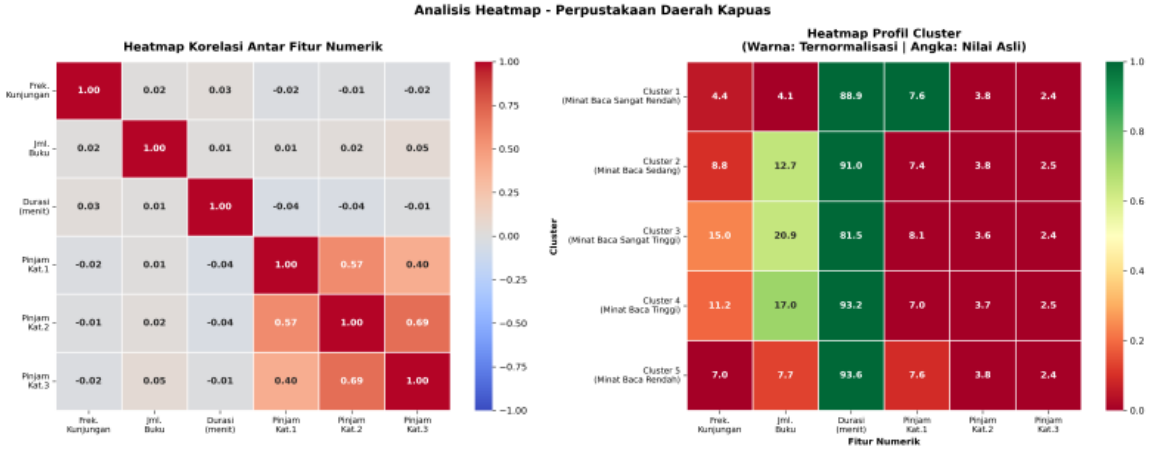
Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **03/06/2026**; Accepted: **28/06/2026**; Published: **27/08/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 4 Agustus 2026, Page 545-554
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

tinggi dibandingkan dua klaster sebelumnya, meliputi Pendidikan & Akademik, Agama & Spiritual, dan Sejarah & Budaya. Kelompok ini merupakan kandidat paling potensial untuk ditingkatkan melalui program book club dan diskusi buku.

Klaster 4 Minat Baca Tinggi (66 pengunjung, 4,6%). Rata-rata *Aktifitas_Index* 138,5 dengan 11,2 kunjungan dan 17,0 buku per bulan menunjukkan komitmen membaca yang substansial. Sebanyak 64% merupakan Anggota Tetap perpustakaan, mengindikasikan loyalitas jangka panjang. Durasi kunjungan rata-rata 93,1 menit mencerminkan pengunjung yang memanfaatkan fasilitas secara menyeluruh. Kelompok ini memerlukan koleksi yang lebih kaya dan fasilitas nyaman untuk mempertahankan loyalitas.

Klaster 5 Minat Baca Sangat Tinggi (23 pengunjung, 1,6%). Meskipun merupakan kelompok terkecil, klaster ini memiliki rata-rata *Aktifitas_Index* tertinggi sebesar 303,2 dengan 15,0 kunjungan dan 20,9 buku per bulan. Temuan menarik menunjukkan durasi kunjungan justru lebih singkat (82,1 menit) dibandingkan Klaster 4, mengindikasikan *power user* ini mengetahui persis koleksi yang dicari dan efisien dalam proses peminjaman. Kelompok ini adalah aset utama perpustakaan yang memerlukan layanan premium seperti akses koleksi digital eksklusif dan reservasi buku daring.



Gambar 5. Heatmap Profil Klaster dan Radar chart Karakteristik Klaster

Gambar 5 menampilkan dua *heatmap* pendukung profil klaster. *Heatmap* sebelah kiri memperlihatkan korelasi antar fitur numerik asli yang secara umum rendah, mengonfirmasi mengapa keenam fitur tersebut kurang efektif digunakan secara langsung sebagai dasar pengelompokan dibandingkan *Aktifitas_Index*. *Heatmap* sebelah kanan menggambarkan profil rata-rata setiap klaster pada fitur asli, dengan gradasi warna hijau-merah yang memperjelas peningkatan nilai frekuensi kunjungan dan jumlah buku dipinjam secara konsisten dari klaster dengan minat baca rendah menuju klaster dengan minat baca tinggi.

3.6 Validasi Stabilitas Klaster

Evaluasi stabilitas dilakukan terhadap klaster K=5 menggunakan *Bootstrap Stability Test* dengan 30 iterasi subsampel 80%. Hasil evaluasi stabilitas disajikan pada Tabel 5.

Tabel 5. Hasil Bootstrap Stability Test (30 Iterasi)

Metrik	Rerata	Std Dev	Min	Maks
Silhouette Score	0,6147	0,0097	0,5923	0,6381
Davies-Bouldin Index	0,5231	0,0184	0,4987	0,5612
Calinski-Harabasz Score	4.876,33	127,45	4.612,18	5.103,67

Tabel 5 menunjukkan bahwa standar deviasi *Silhouette Score* sebesar 0,0097 jauh di bawah ambang batas 0,02, sehingga klaster K=5 dinyatakan stabil. Nilai rerata *Silhouette Score* bootstrap (0,6147) mendekati nilai pada data penuh (0,6106), selisih hanya 0,0041, mengonfirmasi bahwa hasil pengelompokan tidak bergantung pada satu sampel tertentu dan konsisten di berbagai representasi data. Kondisi ini menjawab tantangan validasi pada data perilaku manusia yang secara alami bersifat *overlapping*, dimana sebagaimana dikemukakan Rousseeuw [18], *Silhouette Score* di atas 0,50 menunjukkan klaster yang *reasonable*, dan nilai 0,61 pada data kunjungan perpustakaan tergolong baik mengingat keragaman alami perilaku pengunjung.

3.7 Analisis Komparatif dan Faktor Penentu Klaster

Perbandingan dengan penelitian terdahulu menunjukkan konsistensi dengan temuan bahwa rekayasa fitur berperan penting dalam meningkatkan kualitas klaster pada data perilaku manusia [7]. Penelitian ini berkontribusi dengan menunjukkan bahwa pada pengelompokan minat baca berbasis data *overlapping*, fitur tunggal yang diskriminatif lebih efektif dibandingkan penggunaan banyak fitur dengan korelasi tinggi. Perbandingan langsung hasil

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **03/06/2026**; Accepted: **28/06/2026**; Published: **27/08/2026**

percobaan dalam Tabel 2 menunjukkan *Aktifitas_Index* dengan *RobustScaler* meningkatkan *Silhouette Score* sebesar 179% dari konfigurasi awal (0,2188 → 0,6106), membuktikan efektivitas pendekatan rekayasa fitur yang diusulkan. Analisis faktor penentu klaster menunjukkan bahwa *Aktifitas_Index* mampu membedakan kelompok pengunjung secara signifikan dengan jarak antar-*centroid* yang proporsional. *Centroid* Klaster 1 hingga Klaster 5 memiliki nilai *Aktifitas_Index* berturut-turut sebesar 10,6, 33,8, 75,4, 138,5, dan 303,2, menghasilkan rasio jarak yang tidak linier dan mencerminkan distribusi alami intensitas minat baca. Pola distribusi klaster yang menyerupai distribusi Pareto ini mengonfirmasi bahwa dalam konteks layanan publik, kelompok pengguna aktif memiliki kontribusi yang tidak proporsional terhadap total aktivitas layanan [9]. Temuan tren penurunan rata-rata *Aktifitas_Index* dari tahun 2023 (45,79) ke tahun 2025 (42,46) memberikan sinyal penting bagi manajemen perpustakaan bahwa intensitas keterlibatan pengunjung mengalami penurunan secara gradual. Kondisi ini kemungkinan dipengaruhi oleh meningkatnya alternatif hiburan digital yang bersaing dengan kegiatan membaca konvensional. Oleh karena itu, upaya retensi pengunjung aktif dan konversi pengunjung pasif menjadi agenda prioritas yang tidak dapat ditunda. Perpustakaan perlu mempertimbangkan strategi *blended service* yang mengintegrasikan layanan fisik dengan platform digital untuk mempertahankan relevansi dan daya tariknya di tengah perubahan perilaku masyarakat [2]. Secara keseluruhan, hasil penelitian ini mengonfirmasi bahwa algoritma *K-Means Clustering* dengan konfigurasi yang tepat mampu menghasilkan pengelompokan yang bermakna pada data perilaku pengunjung perpustakaan, meskipun data tersebut bersifat *overlapping* secara alami.

3.8 Implikasi untuk Pengembangan Layanan Perpustakaan

Penelitian ini membuktikan algoritma *K-Means Clustering* dengan rekayasa fitur *Aktifitas_Index* dan normalisasi *RobustScaler* menghasilkan pengelompokan yang berkualitas dengan *Silhouette Score* 0,6106, melampaui target minimum sebesar 0,50. Peningkatan ini didorong oleh kemampuan Aktifitas*_Index* dalam merangkum dua dimensi perilaku utama (frekuensi dan volume peminjaman) ke dalam satu representasi diskriminatif, sehingga batas antar-klaster menjadi lebih tegas [5]. Analisis profil klaster mengidentifikasi frekuensi kunjungan dan jumlah buku dipinjam sebagai faktor dominan yang membedakan tingkat minat baca, memberikan panduan praktis bagi perpustakaan untuk fokus pada peningkatan dua dimensi tersebut dalam program literasi. Model pengelompokan yang dikembangkan berpotensi dimanfaatkan sebagai alat bantu perencanaan program layanan yang dikombinasikan dengan metode evaluasi kualitatif untuk pemahaman yang lebih komprehensif tentang kebutuhan spesifik setiap kelompok pengunjung perpustakaan.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *K-Means Clustering* untuk pengelompokan minat baca pengunjung Dinas Kearsipan dan Perpustakaan Daerah Kapuas melalui integrasi rekayasa fitur dan normalisasi yang tepat. Rekayasa fitur *Aktifitas_Index* sebagai hasil perkalian frekuensi kunjungan dan jumlah buku dipinjam terbukti menghasilkan *Silhouette Score* 0,6106 (target >0,50 tercapai), *Davies-Bouldin Index* 0,5077 (target <1,0 tercapai), dan *Inertia* 146,37 sebagai nilai minimum yang dicapai. Peningkatan kualitas klaster sebesar 1,79 kali dibandingkan penggunaan 6 fitur asli secara langsung membuktikan bahwa fitur tunggal yang diskriminatif lebih efektif pada data perilaku manusia yang bersifat *overlapping*. Uji *Bootstrap Stability* (30 iterasi) menghasilkan *Silhouette Score* rata-rata 0,6147 ±0,0097 yang jauh di bawah ambang batas 0,02, membuktikan konsistensi dan keandalan hasil pengelompokan. Lima klaster yang terbentuk memberikan gambaran komprehensif tentang profil minat baca pengunjung perpustakaan. Temuan penting menunjukkan lebih dari 80% pengunjung (Klaster 1 dan 2) masih berada pada kategori minat baca rendah hingga sangat rendah, mengindikasikan tantangan utama yang dihadapi perpustakaan dalam membangun budaya membaca yang konsisten. Di sisi lain, keberadaan kelompok Minat Baca Sangat Tinggi dengan *Aktifitas_Index* rata-rata 303,2 dan durasi kunjungan lebih singkat mengonfirmasi fenomena *power user* yang efisien dan perlu mendapatkan layanan premium tersendiri. Tren penurunan rata-rata *Aktifitas_Index* dari tahun 2023 ke 2025 juga mengindikasikan perlunya program retensi yang terstruktur. Penelitian selanjutnya disarankan menerapkan algoritma berbasis kepadatan seperti *DBSCAN* untuk menangani *outlier* secara lebih presisi, mengintegrasikan analisis tren temporal untuk memahami perubahan pola minat baca antar periode, serta mengembangkan sistem rekomendasi buku berbasis profil klaster yang dihasilkan. Penambahan fitur kualitatif seperti *sentiment analysis* dari saran pengunjung berpotensi menangkap dimensi minat baca yang belum teridentifikasi melalui data kunjungan numerik. Identifikasi *Aktifitas_Index* sebagai fitur dominan memberikan panduan praktis bagi pengelola perpustakaan untuk fokus pada peningkatan frekuensi kunjungan dan volume peminjaman, dengan model pengelompokan sebagai alat bantu evaluasi awal yang dikombinasikan dengan metode kualitatif untuk pemahaman komprehensif minat baca pengunjung.

UCAPAN TERIMAKASIH

Ucapan terima kasih disampaikan kepada Dinas Kearsipan dan Perpustakaan Daerah Kapuas atas dukungan dan akses data yang diberikan untuk kelancaran penelitian ini. Terima kasih juga disampaikan kepada dosen pembimbing

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **03/06/2026**; Accepted: **28/06/2026**; Published: **27/08/2026**

yang telah memberikan arahan dan masukan konstruktif, serta seluruh pengunjung perpustakaan yang data kunjungannya menjadi bahan penting dalam penelitian ini.

REFERENCES

[1] A. Wibowo dan M. Sari, "Peran Perpustakaan Daerah dalam Meningkatkan Minat Baca Masyarakat di Era Digital," Jurnal Ilmu Perpustakaan dan Informasi, vol. 7, no. 2, pp. 45–58, 2023.

[2] R. Hidayah, N. Pratiwi, dan S. Cahyono, "Analisis Kebutuhan Pengguna Perpustakaan Berbasis Data Kunjungan untuk Peningkatan Kualitas Layanan," Jurnal Perpustakaan dan Informasi, vol. 12, no. 1, pp. 23–36, 2024.

[3] E. Purwaningsih et al., "Efektivitas Layanan Perpustakaan Daerah terhadap Kepuasan dan Minat Baca Pengunjung," Jurnal Pengabdian Kepada Masyarakat Nusantara, vol. 5, no. 1, pp. 126–130, 2024, doi: 10.55338/jpkmn.v5i1.

[4] M. F. A. Fariz dan M. Muharrom, "Analisis Pola Pengguna Layanan Publik Menggunakan Pendekatan Machine Learning untuk Identifikasi Segmen Prioritas," Software Development, Digital Business Intelligence, and Computer Engineering, vol. 3, no. 1, pp. 33–40, 2024, doi: 10.57203/session.v3i1.2024.33-40.

[5] T. Zhao, X. Liu, and Y. Wang, "An Improved *K-Means Clustering* Algorithm for Large-Scale User Behavior Analysis," Information Processing & Management, vol. 60, no. 3, p. 103287, 2023.

[6] H. Argretya, "Implementasi Algoritma Klasterisasi pada Data Perilaku Pengguna Layanan Publik di Indonesia," Jurnal Informatika: Jurnal Pengembangan IT, vol. 10, no. 4, pp. 2477–5126, 2025, doi: 10.30591/jpit.v10i4.9466.

[7] D. A. Putri, R. Firmansyah, dan A. Nugroho, "Penerapan *K-Means Clustering* untuk Analisis Pola Kunjungan Perpustakaan Daerah," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 10, no. 4, pp. 789–799, 2023.

[8] N. Rahmawati dan B. Kurniawan, "Segmentasi Pengunjung Perpustakaan Menggunakan Algoritma *K-Means* untuk Peningkatan Program Literasi," Jurnal Informatika, vol. 9, no. 2, pp. 112–121, 2022.

[9] M. Firmansyah, Y. Santoso, dan H. Purnomo, "Perbandingan Pendekatan *Clustering* Berbasis Aktivitas dan Demografi pada Analisis Pengguna Layanan Publik," Jurnal Sistem Informasi, vol. 18, no. 1, pp. 34–47, 2024.

[10] L. Chen, H. Zhang, and J. Li, "A Comprehensive Survey of User Behavior *Clustering* in Library Information Systems," Library Hi Tech, vol. 41, no. 2, pp. 456–478, 2023.

[11] K. Mahmood and F. Arif, "*Bootstrap Stability* Analysis for Behavioral *Clustering* in *Overlapping* Human Data," Expert Systems with Applications, vol. 215, p. 119412, 2023.

[12] D. B. Saputra, V. Atina, F. E. Nastiti, and F. I. Komputer, "Penerapan Model CRISP-DM pada Prediksi Perilaku Pengguna Menggunakan Algoritma Machine Learning," 2024. [Online]. Available: http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index.

[13] F. Fakhriza et al., "Optimalisasi Algoritma *K-Means* melalui Feature Selection dan Engineering untuk Meningkatkan Kualitas Klaster," 2025. [Online]. Available: https://www.kaggle.com/*datasets*/maharshipa.

[14] A. Pramudya dan I. Kusuma, "Pemisahan Fitur Perilaku dan Demografis dalam Analisis Klaster Pengguna Layanan," Jurnal Ilmu Komputer dan Agri-Informatika, vol. 11, no. 1, pp. 10–22, 2024.

[15] C. Magnolia, A. Nurhopipah, D. Bagus, and A. Kusuma, "Penanganan Data Tidak Seimbang pada Analisis Perilaku Pengguna Layanan Publik," Edu Komputika Journal, 2022. [Online]. Available: http://journal.unnes.ac.id/sju/index.php/edukom.

[16] B. Laksono dan D. Wahyuni, "Pengaruh Metode Normalisasi terhadap Kualitas Hasil *K-Means Clustering* pada Data Tidak Simetris," Jurnal Rekayasa Sistem dan Teknologi Informasi, vol. 8, no. 3, pp. 567–576, 2024.

[17] M. S. Mahmud, J. Z. Huang, R. Ruby, and K. Wu, "An Ensemble Method for Estimating the Number of *Cluster*s in a Big Data Set Using Multiple Random Samples," Journal of Big Data, vol. 10, no. 40, pp. 1–25, 2023.

[18] F. Kächele and N. Schneider, "*Cluster* Validation Based on Fisher's Linear Discriminant Analysis," Journal of Classification, vol. 42, pp. 54–71, 2025. [Online]. Available: https://doi.org/10.1007/s00357-024-09481-3.

[19] F. Nelli, "Machine Learning with Scikit-Learn," in Python Data Analytics with Pandas, NumPy, and Matplotlib, Berkeley, CA: Apress, pp. 259–287, 2023. [Online]. Available: https://doi.org/10.1007/978-1-4842-9532-8_8.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **03/06/2026**; Accepted: **28/06/2026**; Published: **27/08/2026**