Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533

Segmentasi Perilaku Pemustaka Menggunakan DBSCAN untuk Optimalisasi Layanan Perpustakaan Digital

Candra Naya*, Ermanto, Unggul Prima Dhani

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi
Ruko Bekasi Mas, Jl. Ahmad Yani, Marga Jaya, Kec. Bekasi Sel., Kota Bks, Jawa Barat, Indonesia
Email: [1,*]candranaya@pelitabangsa.ac.id, [2]ermanto@pelitabangsa.ac.id, [3]unggul.pd@gmail.com
Email Penulis Korespondensi: candranaya@pelitabangsa.ac.id
Submitted: **30/06/2026;** Accepted: **24/07/2026;** Published: **24/07/2026**

Abstrak−Perkembangan perpustakaan digital menghasilkan data transaksi peminjaman yang semakin besar sehingga diperlukan teknik analisis untuk memahami karakteristik dan pola perilaku pengguna sebagai dasar pengambilan keputusan. Penelitian ini bertujuan mengelompokkan perilaku pemustaka menggunakan algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) sehingga diperoleh segmentasi pengguna berdasarkan tingkat aktivitas peminjaman buku. Dataset yang digunakan adalah Book-Crossing Dataset yang terdiri atas 278.858 transaksi rating, 271.379 data pengguna, dan 271.360 data buku. Tahapan penelitian meliputi Exploratory Data Analysis (EDA), preprocessing data, feature engineering, standardisasi menggunakan StandardScaler, penentuan parameter ε melalui K-Distance Graph, proses clustering menggunakan DBSCAN, evaluasi cluster menggunakan Silhouette Score, serta visualisasi hasil menggunakan Principal Component Analysis (PCA). Hasil penelitian menunjukkan bahwa parameter ε = 0,5 dan MinPts = 5 menghasilkan tiga cluster dengan 244 pengguna teridentifikasi sebagai noise. Nilai Silhouette Score sebesar 0,5996 menunjukkan bahwa cluster yang terbentuk memiliki kualitas pemisahan yang cukup baik. Analisis karakteristik cluster berhasil mengidentifikasi kelompok pengguna dengan tingkat aktivitas rendah, sedang, dan sangat tinggi, sehingga memberikan informasi yang bermanfaat bagi pengelola perpustakaan dalam menyusun strategi personalisasi layanan, pengembangan koleksi, peningkatan sistem rekomendasi buku, serta pengambilan keputusan berbasis data untuk meningkatkan kualitas layanan perpustakaan digital.

Kata kunci: DBSCAN; Clustering; Perpustakaan Digital; Analisis Perilaku Pengguna; Book-Crossing Dataset.

Abstract−The rapid growth of digital libraries has generated increasingly large borrowing transaction data, creating the need for analytical techniques to understand user behavior patterns and support data-driven library management. This study aims to cluster library users based on their borrowing behavior using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to identify user segments according to their borrowing activity. The study employed the Book-Crossing Dataset, consisting of 278,858 rating transactions, 271,379 user records, and 271,360 book records. The research methodology included Exploratory Data Analysis (EDA), data preprocessing, feature engineering, feature standardization using StandardScaler, ε parameter selection through the K-Distance Graph, DBSCAN clustering, cluster evaluation using the Silhouette Score, and visualization using Principal Component Analysis (PCA). The experimental results indicate that ε = 0.5 and MinPts = 5 produced three clusters, with 244 users identified as noise. The obtained Silhouette Score of 0.5996 demonstrates a reasonably good clustering quality. Furthermore, the resulting clusters successfully represent users with low, moderate, and very high borrowing activities, providing valuable insights for developing personalized library services, improving book recommendation systems, supporting collection development, and facilitating data-driven decision-making to enhance the overall quality of digital library services.

Keywords: DBSCAN; Clustering; Digital Library; User Behavior Analysis; Book-Crossing Dataset.

1. PENDAHULUAN

Transformasi digital telah mengubah berbagai aspek layanan perpustakaan dari sistem konvensional menuju perpustakaan digital yang mampu menyediakan akses informasi secara cepat, fleksibel, dan tanpa batasan ruang maupun waktu [1], [2], [3]. Selain menyediakan koleksi digital, perpustakaan modern juga dituntut mampu memahami karakteristik dan perilaku pemustaka sehingga layanan yang diberikan menjadi lebih personal, efisien, dan sesuai dengan kebutuhan pengguna [4], [5]. Aktivitas pemustaka pada perpustakaan digital menghasilkan data dalam jumlah besar, seperti riwayat pemberian rating buku, frekuensi interaksi dengan koleksi, serta intensitas penggunaan layanan [6], [7]. Data tersebut memiliki potensi besar untuk dimanfaatkan sebagai dasar pengambilan keputusan dalam pengembangan layanan perpustakaan apabila dianalisis menggunakan teknik data mining [8], [9].

Salah satu pendekatan yang banyak digunakan dalam analisis perilaku pengguna adalah clustering [10], [11]. Teknik ini bertujuan mengelompokkan objek berdasarkan tingkat kemiripan karakteristik sehingga objek dalam kelompok yang sama memiliki karakteristik yang relatif homogen, sedangkan objek antar kelompok memiliki karakteristik yang berbeda [12]. Dalam konteks perpustakaan digital, hasil segmentasi dapat dimanfaatkan untuk mengidentifikasi kelompok pemustaka aktif, pemustaka dengan aktivitas sedang, pemustaka pasif, maupun pengguna yang memiliki pola interaksi tidak umum (outlier) [13]. Informasi tersebut dapat digunakan sebagai dasar penyusunan strategi rekomendasi koleksi, peningkatan layanan digital, evaluasi koleksi, hingga penyusunan program literasi informasi yang lebih tepat sasaran [14].

Berbagai algoritma clustering telah digunakan dalam penelitian sebelumnya, di antaranya K-Means, Hierarchical Clustering, Gaussian Mixture Model, dan Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [15], [16]. Algoritma K-Means merupakan metode yang paling banyak digunakan karena

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533

memiliki proses komputasi yang sederhana [17]. Namun, metode ini memiliki beberapa keterbatasan, seperti harus menentukan jumlah cluster di awal proses serta sensitif terhadap keberadaan outlier. Pada data perilaku pengguna perpustakaan, karakteristik aktivitas pemustaka umumnya tidak membentuk kelompok berbentuk bulat (spherical cluster) dan sering mengandung data dengan kepadatan yang berbeda, sehingga pendekatan berbasis centroid kurang optimal [18].

Sebagai alternatif, DBSCAN menawarkan pendekatan berbasis kepadatan (density-based clustering) yang mampu mengidentifikasi kelompok data dengan bentuk yang tidak beraturan tanpa harus menentukan jumlah cluster sejak awal [19]. Selain itu, DBSCAN memiliki kemampuan mendeteksi data noise atau outlier secara otomatis sehingga lebih sesuai diterapkan pada data perilaku pengguna yang memiliki variasi aktivitas cukup tinggi. Karakteristik tersebut menjadikan DBSCAN banyak diterapkan pada berbagai bidang, seperti segmentasi pelanggan, analisis media sosial, deteksi anomali jaringan, analisis mobilitas, serta pengelompokan data kesehatan [20]. Meskipun demikian, penerapan DBSCAN pada analisis perilaku pemustaka di lingkungan perpustakaan digital masih relatif terbatas.

Beberapa penelitian terdahulu telah membahas segmentasi perilaku pengguna menggunakan teknik clustering. Penelitian pertama menggunakan algoritma K-Means untuk mengelompokkan pengguna perpustakaan berdasarkan frekuensi peminjaman buku sehingga diperoleh beberapa kelompok pengguna dengan tingkat aktivitas yang berbeda [21]. Penelitian kedua memanfaatkan analisis perilaku pembaca melalui data transaksi perpustakaan untuk meningkatkan sistem rekomendasi koleksi digital [22]. Penelitian ketiga menerapkan DBSCAN pada data pelanggan layanan daring dan menunjukkan bahwa metode tersebut mampu mendeteksi kelompok pengguna aktif maupun data outlier dengan lebih baik dibandingkan metode berbasis centroid [23]. Penelitian lainnya mengombinasikan analisis perilaku pengguna dengan teknik recommender system untuk meningkatkan kualitas layanan perpustakaan digital [24]. Selain itu, terdapat penelitian yang memanfaatkan data rating buku dalam membangun sistem rekomendasi, namun penelitian tersebut lebih berfokus pada prediksi preferensi pengguna dibandingkan segmentasi perilaku pengguna secara menyeluruh [25].

Berdasarkan penelitian-penelitian tersebut, masih terdapat beberapa kesenjangan penelitian (research gap). Pertama, sebagian besar penelitian perpustakaan masih menggunakan algoritma K-Means sehingga hasil segmentasi sangat dipengaruhi oleh jumlah cluster yang telah ditentukan sebelumnya. Kedua, sebagian besar penelitian lebih menitikberatkan pada pembangunan sistem rekomendasi buku dibandingkan analisis karakteristik perilaku pemustaka. Ketiga, penelitian yang memanfaatkan dataset Book-Crossing umumnya digunakan untuk membangun model collaborative filtering atau prediksi rating, sedangkan pemanfaatannya sebagai dasar segmentasi perilaku pengguna menggunakan DBSCAN masih sangat terbatas. Keempat, keberadaan pengguna dengan pola aktivitas yang tidak umum sering kali diabaikan, padahal informasi tersebut penting dalam pengembangan strategi layanan perpustakaan digital.

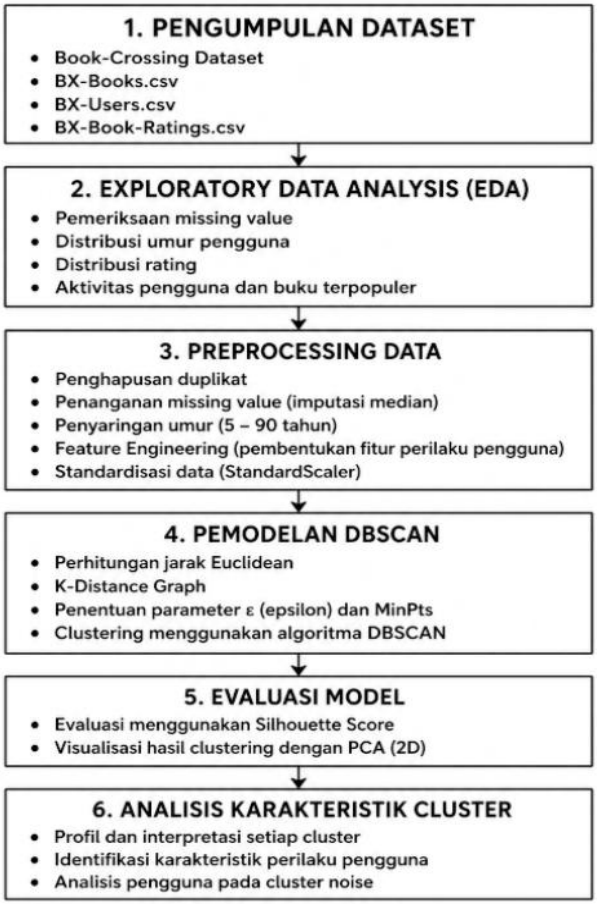
Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan penerapan algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) untuk melakukan segmentasi perilaku pemustaka menggunakan dataset Book-Crossing yang terdiri atas data pengguna, data buku, dan data rating. Karakteristik perilaku pengguna dibentuk melalui proses feature engineering berdasarkan aktivitas pemberian rating, jumlah buku yang dinilai, rata-rata rating, jumlah buku unik yang diakses, serta karakteristik demografi pengguna. Selanjutnya dilakukan normalisasi data dan proses clustering menggunakan DBSCAN untuk memperoleh kelompok pengguna yang memiliki pola aktivitas serupa. Kualitas hasil segmentasi dievaluasi menggunakan Silhouette Score, sedangkan visualisasi cluster dilakukan menggunakan Principal Component Analysis (PCA) agar distribusi masing-masing kelompok dapat diamati secara lebih jelas.

Kontribusi utama penelitian ini terletak pada pemanfaatan algoritma DBSCAN untuk melakukan segmentasi perilaku pemustaka berdasarkan aktivitas interaksi terhadap koleksi digital, bukan hanya berdasarkan jumlah peminjaman atau rekomendasi buku. Hasil segmentasi diharapkan mampu memberikan informasi yang lebih komprehensif mengenai karakteristik pengguna sehingga dapat dimanfaatkan sebagai dasar pengembangan layanan perpustakaan digital yang lebih adaptif, seperti personalisasi rekomendasi koleksi, peningkatan layanan kepada pemustaka kurang aktif, serta identifikasi pengguna dengan pola perilaku khusus. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi praktis bagi pengelola perpustakaan digital sekaligus memperkaya penerapan teknik data mining dalam analisis perilaku pemustaka.

2. METODOLOGI PENELITIAN

Penelitian ini bertujuan melakukan segmentasi perilaku pemustaka menggunakan algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) pada Book-Crossing Dataset. Tahapan penelitian meliputi pengumpulan dataset, eksplorasi data (Exploratory Data Analysis/EDA), preprocessing data yang mencakup feature engineering dan standardisasi menggunakan StandardScaler, pemodelan menggunakan DBSCAN, evaluasi hasil clustering menggunakan Silhouette Score, analisis karakteristik cluster, serta penyusunan implikasi hasil segmentasi untuk mendukung optimalisasi layanan perpustakaan digital. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Dataset

Tahap pertama penelitian adalah mengumpulkan Book-Crossing Dataset yang diperoleh dari Kaggle. Dataset terdiri atas tiga file utama, yaitu BX-Books.csv, BX-Users.csv, dan BX-Book-Ratings.csv, yang memuat informasi buku, pengguna, dan aktivitas pemberian rating. Melalui proses feature engineering, data pengguna dan data rating diintegrasikan untuk membentuk enam atribut perilaku, yaitu Total_Ratings, Average_Rating, Unique_Books, Zero_Ratings, Explicit_Ratings, dan Age. Seluruh proses pengolahan data dilakukan menggunakan bahasa pemrograman Python pada lingkungan Google Colaboratory, sedangkan ringkasan dataset yang digunakan disajikan pada Tabel 1.

Tabel 1. Ringkasan Dataset Penelitian

Dataset	Deskripsi	Jumlah Atribut	Atribut Utama
BX-Books.csv	Informasi koleksi buku	8	ISBN, Book-Title, Book-Author, Year-Of-Publication, Publisher
BX-Users.csv	Informasi pengguna	3	User-ID, Location, Age
BX-Book-Ratings.csv	Data interaksi pengguna terhadap buku	3	User-ID, ISBN, Book-Rating

Keterangan Tabel 1. Book-Crossing Dataset terdiri atas tiga file utama yang saling terhubung melalui atribut User-ID dan ISBN. File BX-Books.csv menyediakan informasi bibliografi buku, BX-Users.csv memuat data demografi pengguna, sedangkan BX-Book-Ratings.csv berisi aktivitas pemberian rating terhadap buku. Pada penelitian ini, ketiga file digabungkan untuk membentuk dataset perilaku pemustaka melalui proses feature engineering sehingga menghasilkan atribut-atribut yang digunakan sebagai masukan algoritma DBSCAN dalam proses segmentasi perilaku pengguna perpustakaan digital.

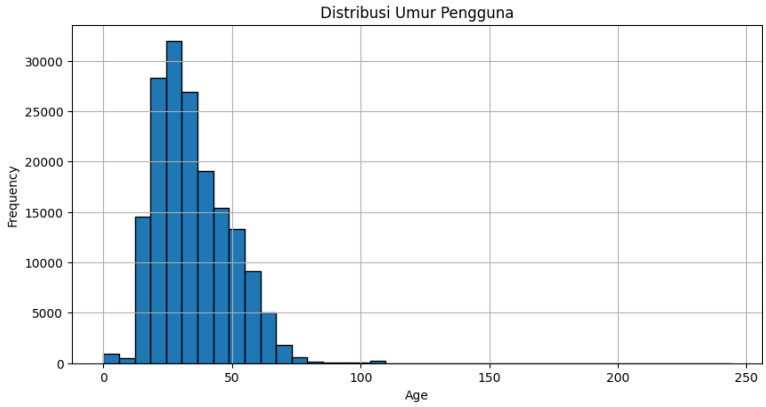
2.2 Eksplorasi Data (Exploratory Data Analysis)

Tahap Exploratory Data Analysis (EDA) dilakukan untuk memahami karakteristik Book-Crossing Dataset sebelum proses preprocessing dan clustering. Analisis meliputi identifikasi missing value, distribusi atribut, karakteristik pengguna, serta pola interaksi antara pengguna dan koleksi buku. Hasil EDA menjadi dasar dalam menentukan proses pembersihan data dan pembentukan fitur yang digunakan pada algoritma DBSCAN.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533

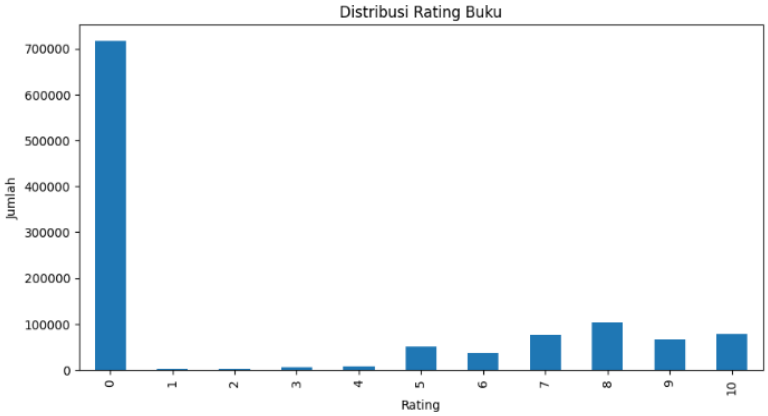
Gambar 2. Visualisasi Missing Value pada Dataset Book-Crossing

Gambar 2 memperlihatkan distribusi nilai kosong pada setiap atribut dalam Book-Crossing Dataset. Terlihat bahwa sebagian besar atribut telah lengkap, sedangkan atribut **Age** masih memiliki sejumlah missing value. Oleh karena itu, dilakukan imputasi menggunakan nilai median agar kualitas data tetap terjaga sebelum proses clustering.



Gambar 3. Distribusi Umur Pemustaka

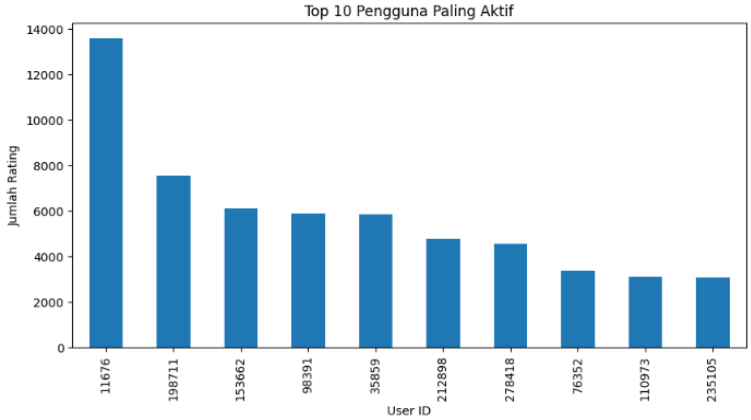
Gambar 3 menunjukkan bahwa mayoritas pengguna berada pada rentang usia dewasa produktif. Selain itu, terdapat sejumlah nilai ekstrem pada atribut umur sehingga dilakukan penyaringan dengan mempertahankan pengguna yang berusia antara 5 hingga 90 tahun agar data yang digunakan lebih representatif.



Gambar 4. Distribusi Rating Buku

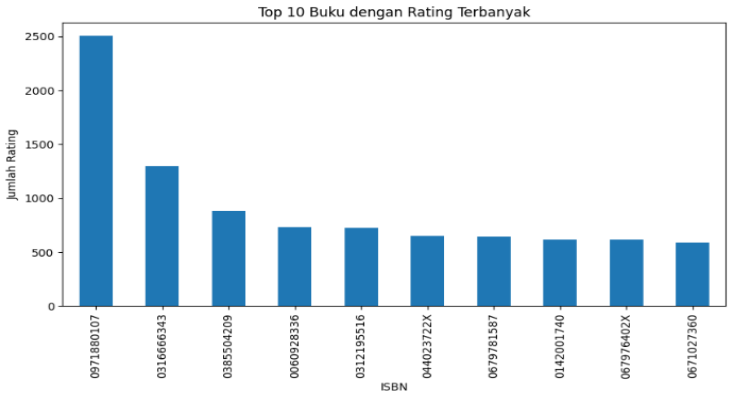
Gambar 4 memperlihatkan distribusi nilai rating yang diberikan pengguna terhadap koleksi buku. Sebagian besar interaksi berupa implicit rating (nilai 0), sedangkan explicit rating memiliki frekuensi yang lebih rendah. Hal ini menunjukkan bahwa sebagian besar pengguna melakukan interaksi terhadap koleksi tanpa memberikan penilaian secara eksplisit.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533



Gambar 5. Sepuluh Pengguna dengan Aktivitas Rating Tertinggi

Gambar 5 menunjukkan sepuluh pengguna dengan jumlah pemberian rating terbanyak pada dataset. Terlihat bahwa hanya sebagian kecil pengguna memiliki tingkat aktivitas yang sangat tinggi, sedangkan mayoritas pengguna memberikan rating dalam jumlah yang jauh lebih sedikit. Kondisi ini mengindikasikan adanya variasi intensitas penggunaan perpustakaan digital.



Gambar 6. Sepuluh Buku dengan Jumlah Rating Terbanyak

Gambar 6 memperlihatkan sepuluh buku yang memperoleh jumlah rating terbanyak dari pengguna. Hasil ini menunjukkan bahwa tingkat popularitas koleksi tidak merata, di mana beberapa buku memiliki frekuensi interaksi yang jauh lebih tinggi dibandingkan koleksi lainnya. Informasi tersebut menjadi indikasi adanya preferensi pengguna terhadap koleksi tertentu.

2.3 Preprocessing Data

Tahap preprocessing bertujuan meningkatkan kualitas data sebelum proses clustering menggunakan algoritma DBSCAN. Proses ini meliputi penghapusan data duplikat, penanganan missing value, penyaringan data pengguna, penggabungan dataset, dan standardisasi data. Nilai missing value pada atribut Age diimputasi menggunakan median, sedangkan data pengguna dengan umur di luar rentang 5–90 tahun dihapus untuk meningkatkan kualitas data. Selanjutnya, dataset BX-Book-Ratings.csv digabungkan dengan BX-Users.csv berdasarkan atribut User-ID sehingga diperoleh dataset perilaku pemustaka. Seluruh fitur kemudian distandardisasi menggunakan StandardScaler agar memiliki rata-rata (mean) 0 dan simpangan baku (standard deviation) 1, sehingga setiap atribut memiliki skala yang seimbang pada proses perhitungan jarak dalam algoritma DBSCAN.

Tabel 2. Tahapan Preprocessing Data

Tahapan	Proses	Tujuan
Penghapusan Data Duplikat	Menghapus data yang memiliki informasi sama pada ketiga dataset	Menghindari perhitungan aktivitas pengguna secara berulang
Penanganan Missing Value	Mengisi nilai kosong pada atribut Age menggunakan median	Mempertahankan kelengkapan data tanpa mengubah distribusi secara signifikan
Penyaringan Data	Mempertahankan pengguna dengan umur 5–90 tahun	Menghilangkan data yang tidak realistis
Penggabungan Dataset	Menggabungkan data pengguna dan data rating berdasarkan User-ID	Membentuk dataset perilaku pengguna

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533

Tahapan	Proses	Tujuan
Standardisasi Data	Menggunakan StandardScaler	Menyamakan skala seluruh variabel sebelum proses DBSCAN

Tahapan preprocessing dilakukan secara berurutan untuk menghasilkan dataset yang bersih, konsisten, dan siap digunakan pada proses feature engineering serta clustering. Proses ini berperan penting dalam meningkatkan kualitas hasil segmentasi karena algoritma DBSCAN sangat dipengaruhi oleh kualitas data masukan dan perhitungan jarak antar objek.

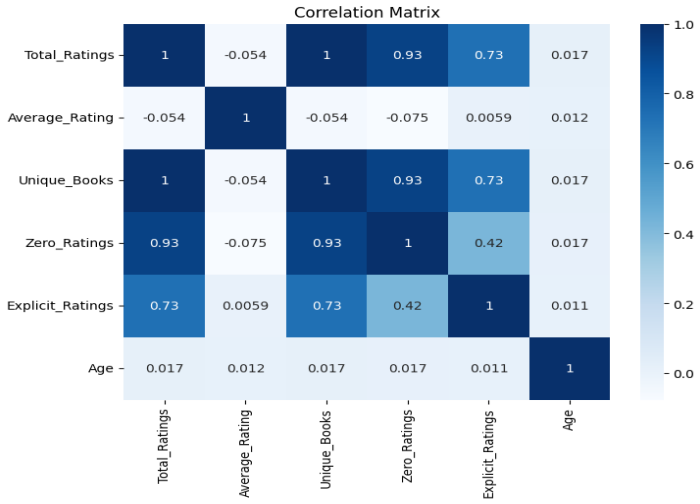
2.3.1 Feature Engineering

Setelah preprocessing, dilakukan feature engineering untuk membentuk atribut yang merepresentasikan perilaku pemustaka berdasarkan aktivitas pemberian rating. Proses ini dilakukan dengan mengelompokkan data berdasarkan User-ID dan mengagregasi aktivitas pengguna sehingga setiap pengguna direpresentasikan oleh satu baris data. Hasil feature engineering menghasilkan enam atribut, yaitu Total_Ratings, Average_Rating, Unique_Books, Zero_Ratings, Explicit_Ratings, dan Age, yang digunakan sebagai representasi tingkat aktivitas, preferensi pengguna, keberagaman koleksi yang diakses, jenis interaksi, dan karakteristik demografi. Dataset hasil transformasi ini selanjutnya digunakan sebagai masukan pada proses clustering menggunakan algoritma DBSCAN.

Tabel 3. Fitur Perilaku Pemustaka Hasil Feature Engineering

Fitur	Deskripsi	Tipe
Total_Ratings	Jumlah seluruh rating yang diberikan oleh pengguna	Integer
Average_Rating	Nilai rata-rata rating pengguna	Float
Unique_Books	Jumlah buku berbeda yang pernah diberi rating	Integer
Zero_Ratings	Jumlah implicit rating (rating = 0)	Integer
Explicit_Ratings	Jumlah explicit rating (rating > 0)	Integer
Age	Umur pengguna	Integer

Keterangan Tabel 3. Keenam fitur tersebut digunakan sebagai representasi perilaku pemustaka pada proses clustering. Fitur Total_Ratings, Average_Rating, Unique_Books, Zero_Ratings, dan Explicit_Ratings diperoleh melalui proses agregasi berdasarkan User-ID, sedangkan atribut Age diambil dari dataset pengguna sebagai informasi demografi. Seluruh fitur kemudian distandardisasi menggunakan StandardScaler sebelum diproses menggunakan algoritma DBSCAN agar setiap variabel memiliki skala yang sebanding dan tidak mendominasi proses penghitungan jarak antar data.



Gambar 7. Heatmap Korelasi Antar Fitur Perilaku Pemustaka

Berdasarkan Gambar 7 terlihat bahwa Total_Ratings, Unique_Books, Zero_Ratings, dan Explicit_Ratings memiliki korelasi positif yang cukup kuat. Hal ini menunjukkan bahwa pengguna yang aktif memberikan rating cenderung mengakses lebih banyak koleksi buku. Sementara itu, atribut Age memiliki korelasi yang relatif rendah terhadap atribut perilaku lainnya sehingga berperan sebagai informasi demografi pendukung.

2.3.2 Standardisasi Data

Sebelum proses clustering, seluruh fitur hasil feature engineering (Total_Ratings, Average_Rating, Unique_Books, Zero_Ratings, Explicit_Ratings, dan Age) distandardisasi menggunakan metode StandardScaler. Standardisasi bertujuan menyamakan skala setiap variabel agar tidak terjadi dominasi atribut tertentu pada

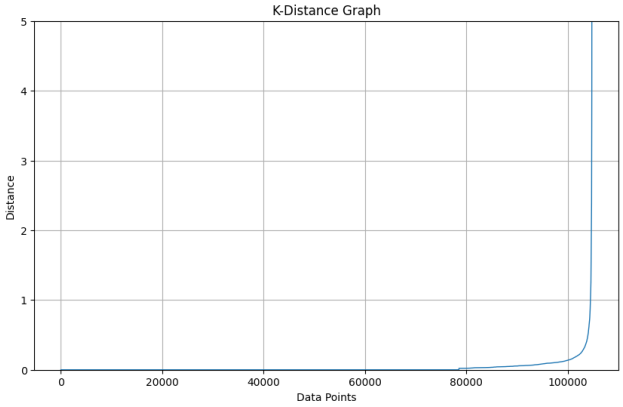
Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533

perhitungan jarak dalam algoritma DBSCAN. Melalui proses ini, setiap fitur ditransformasikan sehingga memiliki nilai rata-rata (mean) 0 dan simpangan baku (standard deviation) 1, sehingga seluruh variabel memberikan kontribusi yang seimbang dalam pembentukan cluster.

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

Dalam proses standardisasi data, z merupakan nilai yang diperoleh setelah data melalui proses standardisasi, sedangkan x adalah nilai asli dari data sebelum ditransformasi. Selanjutnya, μ menunjukkan nilai rata-rata (mean) data, sementara σ merepresentasikan simpangan baku (standard deviation) yang digunakan untuk mengukur penyebaran data terhadap nilai rata-ratanya.

2.4 Pemodelan DBSCAN



Gambar 8. K-Distance Graph untuk Penentuan Nilai ε pada DBSCAN

2.4.1 Parameter DBSCAN

Algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) menggunakan dua parameter utama, yaitu epsilon (ε) dan minimum samples (MinPts). Parameter ε menentukan radius pencarian tetangga (neighborhood), sedangkan MinPts menunjukkan jumlah minimum tetangga agar suatu data dikategorikan sebagai core point. Pada penelitian ini, nilai MinPts ditetapkan sebesar 5, sedangkan nilai ε ditentukan menggunakan K-Distance Graph berdasarkan jarak terhadap tetangga ke-5 (k-nearest neighbors). Nilai ε dipilih pada titik perubahan kurva (elbow point), yang menunjukkan batas antara area berkepadatan tinggi dan data yang mulai teridentifikasi sebagai noise. Berdasarkan hasil visualisasi pada Gambar 8, penelitian ini menggunakan ε = 0,5, karena menghasilkan batas kepadatan yang optimal dan segmentasi perilaku pemustaka yang stabil.

Euclidean Distance

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{m}(x_{ik} - x_{jk})^2} \quad (2)$$

Dalam perhitungan jarak antar data, $d(x_i, x_j)$ menunjukkan jarak antara dua data, yaitu data ke-i dan data ke-j. m merepresentasikan jumlah fitur yang digunakan dalam proses pengukuran. Sementara itu, x_{ik} merupakan nilai fitur ke-k yang terdapat pada data ke-i dan menjadi dasar perhitungan jarak.

ε-Neighborhood

$$N_\varepsilon(p) = \{q \in D \mid dist(p, q) \leq \varepsilon\} \quad (3)$$

Core Point

$$| N_\varepsilon(p) | \geq MinPts \quad (4)$$

Silhouette Score

$$S(i) = \frac{b(i)-a(i)}{max(a(i),b(i))} \quad (5)$$

2.4.2 Clustering Menggunakan DBSCAN

Setelah proses preprocessing, feature engineering, standardisasi, dan penentuan parameter selesai, dilakukan proses clustering menggunakan algoritma DBSCAN. Proses ini memanfaatkan enam atribut perilaku pengguna yang telah distandardisasi, yaitu Total_Ratings, Average_Rating, Unique_Books, Zero_Ratings, Explicit_Ratings, dan Age. Dengan parameter ε = 0,5 dan MinPts = 5, DBSCAN mengelompokkan data berdasarkan tingkat kepadatan menggunakan Euclidean Distance, kemudian mengklasifikasikan setiap data sebagai core point, border

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533

point, atau noise. Hasil clustering berupa label cluster untuk setiap pengguna selanjutnya digunakan pada proses evaluasi menggunakan Silhouette Score serta divisualisasikan dengan Principal Component Analysis (PCA).

2.5 Evaluasi Model

2.5.1 Evaluasi Cluster

Kualitas hasil clustering dievaluasi menggunakan Silhouette Score, yaitu metrik evaluasi internal yang mengukur tingkat kekompakan data dalam suatu cluster (cohesion) dan tingkat pemisahan antar cluster (separation). Metode ini dipilih karena tidak memerlukan data berlabel (ground truth) sehingga sesuai untuk mengevaluasi hasil segmentasi menggunakan algoritma DBSCAN. Nilai Silhouette Score berada pada rentang −1 hingga 1, di mana nilai yang mendekati 1 menunjukkan kualitas cluster yang semakin baik, sedangkan nilai yang mendekati 0 atau negatif menunjukkan adanya tumpang tindih antar cluster atau pengelompokan yang kurang optimal.

Silhouette Score dihitung berdasarkan rata-rata jarak setiap data terhadap anggota pada cluster yang sama dan terhadap cluster terdekat lainnya. Pada penelitian ini, perhitungan hanya dilakukan terhadap data yang berhasil dikelompokkan ke dalam cluster, sedangkan data noise (label −1) tidak disertakan dalam proses evaluasi. Secara matematis, Silhouette Score dihitung menggunakan Persamaan (3).

$$S(i) = \frac{b(i)-a(i)}{\max(a(i), b(i))} \quad (6)$$

Dalam perhitungan Silhouette Score, S(i) menunjukkan nilai Silhouette Score untuk data ke-i. Selanjutnya, a(i) merupakan rata-rata jarak antara data ke-i dan seluruh anggota dalam klaster yang sama. Sementara itu, b(i) menunjukkan rata-rata jarak minimum data ke-i terhadap anggota klaster terdekat lainnya sebagai dasar evaluasi kualitas pengelompokan.

Semakin tinggi nilai rata-rata Silhouette Score yang diperoleh, semakin baik kualitas segmentasi yang dihasilkan. Selain evaluasi kuantitatif, hasil clustering juga divisualisasikan menggunakan Principal Component Analysis (PCA) untuk memproyeksikan data berdimensi tinggi ke dalam ruang dua dimensi sehingga distribusi cluster dan keberadaan data noise dapat diamati dengan lebih jelas.

2.5.2 Visualisasi Cluster

Setelah proses clustering, hasil segmentasi divisualisasikan menggunakan Principal Component Analysis (PCA) untuk mereduksi enam atribut hasil feature engineering menjadi dua komponen utama sehingga distribusi cluster dapat diamati secara lebih jelas. Setiap titik merepresentasikan seorang pemustaka dengan warna yang menunjukkan keanggotaan cluster, sedangkan data noise tetap ditampilkan sebagai hasil identifikasi algoritma DBSCAN. Visualisasi ini bertujuan untuk mengamati tingkat pemisahan antar cluster, kepadatan data, serta keberadaan outlier, sehingga memudahkan interpretasi hasil segmentasi. Hasil visualisasi ditampilkan pada Gambar 9.

$$PC_i = Xw_i \quad (7)$$

Dalam proses reduksi dimensi menggunakan metode Principal Component Analysis (PCA), X merepresentasikan matriks data yang berisi sejumlah observasi dan variabel. W merupakan eigenvector yang menunjukkan arah komponen utama dalam ruang data. Sementara itu, Z adalah data hasil reduksi dimensi yang diperoleh melalui transformasi matriks data ke dalam ruang komponen utama.

2.6 Analisis Karakteristik Cluster

Tahap analisis karakteristik cluster bertujuan menginterpretasikan hasil segmentasi yang diperoleh menggunakan algoritma **Density-Based Spatial Clustering of Applications with Noise (DBSCAN)**. Analisis dilakukan dengan menghitung nilai rata-rata setiap atribut perilaku, yaitu **Total_Ratings, Average_Rating, Unique_Books, Zero_Ratings, Explicit_Ratings**, dan **Age** pada masing-masing cluster. Nilai rata-rata tersebut digunakan untuk mengidentifikasi perbedaan tingkat aktivitas pengguna, kecenderungan pemberian rating, keberagaman koleksi yang diakses, serta karakteristik demografi setiap kelompok. Selain itu, data **noise** turut dianalisis sebagai representasi pengguna yang memiliki pola perilaku berbeda dari mayoritas sehingga tidak tergabung dalam cluster tertentu. Hasil analisis karakteristik ini memberikan gambaran mengenai profil setiap kelompok pemustaka dan menjadi dasar dalam penyusunan rekomendasi pengembangan layanan perpustakaan digital, seperti personalisasi layanan, peningkatan sistem rekomendasi buku, dan pengelolaan koleksi yang lebih sesuai dengan kebutuhan pengguna.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Clustering

Setelah melalui tahapan preprocessing, feature engineering, standardisasi data, dan penentuan parameter DBSCAN, proses segmentasi perilaku pemustaka dilakukan menggunakan algoritma Density-Based Spatial

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533

Clustering of Applications with Noise (DBSCAN) dengan parameter ε = 0,5 dan MinPts = 5. Algoritma mengelompokkan pengguna berdasarkan tingkat kepadatan data menggunakan enam atribut perilaku, yaitu Total_Ratings, Average_Rating, Unique_Books, Zero_Ratings, Explicit_Ratings, dan Age.

Setelah parameter DBSCAN ditentukan menggunakan K-Distance Graph (ε = 0,5 dan MinPts = 5), proses clustering diterapkan pada 20.000 data pengguna yang telah melalui preprocessing dan standardisasi. Hasil clustering menghasilkan tiga cluster utama serta satu kelompok noise. Distribusi jumlah anggota pada masing-masing cluster disajikan pada Tabel 4.

Tabel 4. Hasil Clustering Menggunakan DBSCAN

Cluster	Jumlah Pengguna	Persentase (%)
Noise (-1)	244	1,22
Cluster 0	19.747	98,74
Cluster 1	4	0,02
Cluster 2	5	0,02
Total	20.000	100,00

Dominasi Cluster 0 menunjukkan bahwa sebagian besar pengguna memiliki karakteristik perilaku yang relatif homogen dengan tingkat aktivitas yang rendah hingga sedang. Sebaliknya, Cluster 1 dan Cluster 2 hanya berisi sedikit pengguna sehingga dapat dikategorikan sebagai kelompok dengan karakteristik khusus atau aktivitas ekstrem. Keberadaan 244 data noise menunjukkan bahwa DBSCAN mampu mengidentifikasi pengguna yang memiliki pola perilaku berbeda dari mayoritas tanpa memaksakan seluruh data masuk ke dalam cluster tertentu. Karakteristik ini merupakan salah satu keunggulan DBSCAN dibandingkan metode berbasis centroid.

Berdasarkan Tabel 4 terlihat bahwa 98,74% pengguna berada pada Cluster 0, sedangkan hanya sebagian kecil pengguna membentuk Cluster 1 dan Cluster 2. Selain itu, sebanyak 1,22% pengguna dikategorikan sebagai noise. Hasil tersebut menunjukkan bahwa sebagian besar pemustaka memiliki karakteristik perilaku yang relatif serupa, sedangkan sejumlah kecil pengguna memiliki pola aktivitas yang sangat berbeda sehingga dipisahkan oleh algoritma DBSCAN.

Kualitas hasil segmentasi dievaluasi menggunakan **Silhouette Score**, yang mengukur tingkat kekompakan (cohesion) dalam suatu cluster dan tingkat pemisahan (separation) antar cluster. Semakin tinggi nilai Silhouette Score, semakin baik kualitas cluster yang dihasilkan.

Tabel 5. Hasil Evaluasi Clustering

Metrik	Nilai
Silhouette Score	0,5996

Nilai Silhouette Score sebesar 0,5996 menunjukkan bahwa hasil segmentasi memiliki kualitas yang cukup baik. Nilai tersebut mengindikasikan bahwa pengguna yang berada pada cluster yang sama memiliki tingkat kemiripan yang tinggi, sedangkan pemisahan antar cluster juga cukup jelas.

Nilai 0,5996 menunjukkan bahwa algoritma DBSCAN mampu membentuk cluster yang memiliki struktur kepadatan yang baik. Walaupun distribusi anggota cluster tidak seimbang, hasil evaluasi menunjukkan bahwa objek pada masing-masing cluster memiliki tingkat kedekatan yang relatif tinggi dibandingkan dengan objek pada cluster lainnya. Dengan demikian, hasil segmentasi dapat digunakan sebagai dasar untuk menganalisis karakteristik perilaku pemustaka.

3.2 Visualisasi Hasil Clustering

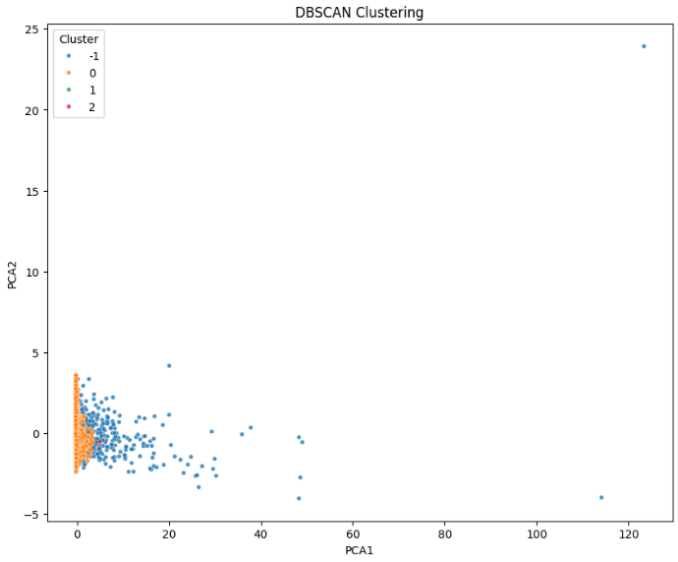
Untuk mempermudah interpretasi hasil segmentasi, dilakukan reduksi dimensi menggunakan Principal Component Analysis (PCA). Metode ini mentransformasikan enam atribut perilaku pemustaka menjadi dua komponen utama sehingga distribusi cluster dapat divisualisasikan dalam bentuk scatter plot. Visualisasi tersebut menampilkan penyebaran setiap cluster beserta data noise yang dihasilkan oleh algoritma DBSCAN. Hasil visualisasi ditunjukkan pada Gambar 9.

Berdasarkan Gambar 9, sebagian besar pemustaka terkonsentrasi pada Cluster 0 yang membentuk kelompok dengan kepadatan tinggi, sedangkan Cluster 1 dan Cluster 2 memiliki jumlah anggota yang lebih sedikit dan tersebar pada area yang berbeda. Selain itu, terlihat sejumlah titik yang berada di luar kelompok utama dan teridentifikasi sebagai noise, yang menunjukkan adanya pengguna dengan pola perilaku berbeda dari mayoritas. Visualisasi ini memperlihatkan bahwa algoritma DBSCAN mampu membentuk segmentasi yang jelas sekaligus mengidentifikasi outlier, sehingga memudahkan analisis karakteristik setiap cluster pada tahap pembahasan berikutnya.

Proses clustering dilakukan menggunakan library Scikit-learn pada Python dengan parameter eps=0.5 dan min_samples=5. Selanjutnya hasil cluster direduksi menggunakan Principal Component Analysis (PCA) menjadi dua dimensi sehingga distribusi antar cluster dapat divisualisasikan.

.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533



Gambar 9. Visualisasi Hasil Clustering Menggunakan PCA

Berdasarkan Gambar 9, mayoritas pemustaka tergabung dalam Cluster 0 yang memiliki kepadatan tinggi, sedangkan Cluster 1, Cluster 2, dan kelompok noise berada di luar kelompok utama. Hasil ini menunjukkan bahwa algoritma DBSCAN mampu mengelompokkan pengguna berdasarkan kepadatan data serta mengidentifikasi outlier secara efektif tanpa memerlukan penentuan jumlah cluster di awal proses.

Visualisasi PCA memperlihatkan bahwa Cluster 0 membentuk area dengan kepadatan tinggi, sedangkan Cluster 1 dan Cluster 2 terpisah pada area yang berbeda. Penyebaran tersebut mengindikasikan bahwa fitur hasil feature engineering mampu membedakan karakteristik masing-masing kelompok pengguna. Selain itu, titik-titik noise berada di luar area kepadatan utama sehingga memperlihatkan kemampuan DBSCAN dalam mendeteksi outlier secara efektif.

3.3 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) mampu mengelompokkan perilaku pemustaka berdasarkan tingkat kepadatan data tanpa menentukan jumlah cluster di awal. Dengan parameter ε = 0,5 dan MinPts = 5, DBSCAN menghasilkan tiga cluster serta mengidentifikasi 244 pengguna sebagai noise. Nilai Silhouette Score sebesar 0,5996 menunjukkan bahwa kualitas pemisahan cluster berada pada kategori cukup baik, sehingga struktur kelompok yang terbentuk memiliki tingkat homogenitas internal dan pemisahan antarcluster yang memadai.

Analisis karakteristik cluster menunjukkan adanya perbedaan tingkat aktivitas pemustaka. Cluster 0 merupakan kelompok terbesar yang didominasi pengguna dengan aktivitas rendah hingga sedang, sehingga mencerminkan karakteristik mayoritas pengguna perpustakaan digital. Sebaliknya, Cluster 2 terdiri atas pengguna dengan intensitas interaksi yang sangat tinggi (power users), sedangkan Cluster 1 merepresentasikan kelompok pengguna dengan karakteristik yang lebih spesifik. Selain itu, keberadaan kelompok noise menunjukkan bahwa DBSCAN mampu mengidentifikasi pengguna dengan pola perilaku yang tidak mengikuti kepadatan mayoritas. Kemampuan mendeteksi outlier ini menjadi salah satu keunggulan DBSCAN dibandingkan metode berbasis centroid, seperti K-Means, yang cenderung memaksa seluruh data masuk ke dalam cluster tertentu.

Temuan penelitian ini sejalan dengan berbagai penelitian sebelumnya yang melaporkan bahwa DBSCAN efektif untuk mengelompokkan data dengan distribusi yang tidak seragam dan mampu mendeteksi noise secara otomatis. Pada konteks perpustakaan digital, hasil segmentasi memberikan informasi yang bermanfaat bagi pengelola untuk memahami karakteristik pemustaka berdasarkan tingkat aktivitas dan pola interaksi terhadap koleksi. Informasi tersebut dapat dimanfaatkan sebagai dasar dalam pengembangan layanan yang lebih adaptif, seperti personalisasi rekomendasi buku, pengelolaan koleksi berdasarkan preferensi pengguna, serta penyusunan strategi promosi yang lebih tepat sasaran.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Kualitas hasil clustering sangat dipengaruhi oleh pemilihan parameter ε dan MinPts, sehingga diperlukan proses penentuan parameter yang tepat menggunakan K-Distance Graph. Selain itu, penelitian ini hanya menggunakan satu dataset publik sehingga karakteristik cluster yang dihasilkan belum tentu merepresentasikan perilaku pengguna pada perpustakaan digital lainnya. Oleh karena itu, penelitian selanjutnya dapat memanfaatkan dataset dari institusi perpustakaan yang berbeda, membandingkan DBSCAN dengan algoritma clustering lainnya seperti HDBSCAN, OPTICS, atau K-Means, serta mengintegrasikan hasil segmentasi ke dalam sistem rekomendasi buku untuk mengevaluasi manfaatnya secara langsung terhadap kualitas layanan perpustakaan digital.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) mampu melakukan segmentasi perilaku pemustaka berdasarkan karakteristik aktivitas pengguna pada Book-Crossing Dataset. Dengan parameter ε = 0,5 dan MinPts = 5, DBSCAN berhasil membentuk tiga cluster utama serta mengidentifikasi 244 pengguna (1,22%) sebagai noise. Hasil evaluasi menggunakan Silhouette Score sebesar 0,5996 menunjukkan bahwa kualitas cluster yang dihasilkan tergolong cukup baik, sehingga setiap kelompok memiliki tingkat kemiripan internal yang tinggi dan pemisahan yang memadai antarcluster. Segmentasi yang terbentuk mampu membedakan kelompok pengguna dengan tingkat aktivitas rendah hingga sedang, kelompok dengan karakteristik aktivitas khusus, serta kelompok pengguna yang sangat aktif. Selain itu, kemampuan DBSCAN dalam mendeteksi data noise memberikan nilai tambah karena pengguna dengan pola perilaku yang berbeda dari mayoritas tetap dapat diidentifikasi tanpa dipaksakan masuk ke dalam cluster tertentu. Informasi yang dihasilkan dari proses segmentasi ini memberikan gambaran mengenai karakteristik perilaku pemustaka yang dapat dimanfaatkan oleh pengelola perpustakaan digital untuk mendukung pengembangan layanan yang lebih adaptif, seperti personalisasi rekomendasi buku, pengelolaan koleksi berdasarkan pola penggunaan, penyusunan program peningkatan keterlibatan pengguna, serta pengambilan keputusan berbasis data. Dengan demikian, hasil penelitian membuktikan bahwa pendekatan clustering berbasis kepadatan efektif digunakan untuk memahami perilaku pengguna pada lingkungan perpustakaan digital. Meskipun demikian, hasil segmentasi masih dipengaruhi oleh pemilihan parameter DBSCAN dan penggunaan satu dataset publik. Oleh karena itu, penelitian selanjutnya disarankan membandingkan DBSCAN dengan algoritma clustering lain, seperti HDBSCAN, OPTICS, atau K-Means, serta menguji model pada dataset perpustakaan yang berbeda agar diperoleh hasil segmentasi yang lebih robust, konsisten, dan memiliki tingkat generalisasi yang lebih baik.

UCAPAN TERIMAKASIH

Penulis menyampaikan apresiasi kepada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa atas dukungan akademik dan fasilitas penelitian yang diberikan selama pelaksanaan penelitian ini. Penulis juga mengucapkan terima kasih kepada Direktorat Penelitian dan Pengabdian kepada Masyarakat (DPPM) Universitas Pelita Bangsa atas dukungan terhadap kegiatan penelitian dan publikasi ilmiah. Ucapan terima kasih turut disampaikan kepada pengelola Book-Crossing Dataset yang tersedia melalui platform Kaggle, sehingga penelitian ini dapat dilakukan menggunakan dataset terbuka yang berkualitas. Selain itu, penulis mengucapkan terima kasih kepada seluruh pihak yang telah memberikan bantuan, masukan, dan dukungan selama penyusunan penelitian ini.

REFERENCES

[1] N. Vitriana, “Transformasi perpustakaan di era digital native,” Librarium: Library and Information Science Journal, vol. 1, no. 1, pp. 59–69, Mar. 2024, doi: 10.53088/librarium.v1i1.693.
[2] A. F. Fiqori, I. Irawati, U. Faruq, and A. Fahriza, “Transformasi Digital dalam Pengelolaan Perpustakaan di SMA Negeri 5 Pekanbaru,” Social Science Academic, vol. 4, no. 1, pp. 183–192, May 2026, doi: 10.37680/ssa.9667.
[3] B. S. Nahar, “Digital Transformation in Library Science: Enhancing Access, Preservation, and User Engagement,” Journal of Advanced Research in Library and Information Science, vol. 12, no. 4, pp. 14–18, Dec. 2025, doi: 10.24321/2395.2288.2025012.
[4] K. Azzahra and E. Rahmah, “Strategi Pustakawan dalam Menghadapi Pemustaka Generasi Z di Perpustakaan Universitas Negeri Padang,” Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence), vol. 5, no. 1, pp. 94–100, Apr. 2025, doi: 10.55382/jurnalpustakaai.v5i1.957.
[5] E. Utama, J. Junaidi, and Z. Zamzami, “Pengaruh Kompetensi Pustakawan Terhadap Kepuasan Pemustaka dengan Fasilitas Perpustakaan Sebagai Variabel Intervening,” Jurnal Pustaka Budaya, vol. 13, no. 1, pp. 85–105, Jan. 2026, doi: 10.31849/avs0a353.
[6] M. Nahotko, M. Zych, A. Januszko-Szakiel, and M. Jaskowska, “Big data-driven investigation into the maturity of library research data services (RDS),” The Journal of Academic Librarianship, vol. 49, no. 1, p. 102646, Jan. 2023, doi: 10.1016/j.acalib.2022.102646.
[7] S. R. Shimray, A. Subaveerapandiyan, and N. Ahmad, “Digital transformation in academic libraries: e-resources, OPACs and AI in information discovery,” Reference Services Review, vol. 53, no. 2, pp. 238–255, Oct. 2025, doi: 10.1108/RSR-12-2024-0078.
[8] T. Landivar, R. Rendon, and L. Siguenza-Guzman, “Decision-Making of the University Libraries’ Digital Collection Through the Publication and Citation Patterns Analysis. A Literature Review,” 2022, pp. 80–94. doi: 10.1007/978-3-031-03884-6_6.
[9] P. Roy, “Transforming higher education libraries with data analytics, business intelligence, and business analytics: A review,” Journal of Librarianship and Information Science, vol. 58, no. 1, pp. 23–40, Mar. 2026, doi: 10.1177/09610006241307028.
[10] A. F. Zabidi, “Penerapan Algoritma K-Means untuk Pengelompokan Koleksi Perpustakaan dengan Data Mining,” Media Jurnal Informatika, vol. 16, no. 2, p. 233, Dec. 2024, doi: 10.35194/mji.v16i2.4814.
[11] X. Zhang and J. Zhang, “Analysis and research on library user behavior based on apriori algorithm,” Measurement: Sensors, vol. 27, p. 100802, Jun. 2023, doi: 10.1016/j.measen.2023.100802.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1225−1236
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10533

[12] M. Pirniakan, A. Adibifar, F. S. Litooee, M. Tangestanizadeh, and M. Etebar Zadeh, "Enhancing topic modeling in digital libraries through keyword-centric self-supervised learning," Journal of Librarianship and Information Science, vol. 58, no. 2, pp. 733–749, Jun. 2026, doi: 10.1177/09610006251339838.
[13] M. Marzuki, S. F. Z. Azero, N. A. A. Mohd Zamzuri, and M. R. Abdul Kadir, "A Systematic Literature Review of User Behavior and Personalization in Digital Libraries," International Journal of Research and Innovation in Social Science, vol. IX, no. I, pp. 4830–4842, 2025, doi: 10.47772/IJRISS.2025.9010372.
[14] H. Salmi Addin, H. Anggraini, H. Nur Riya Putri Yenti, F. Wandan Sari, and I. Hidayat, "Strategi Pengembangan Koleksi Perpustakaan Digital," Media Informasi, vol. 33, no. 1, pp. 88–95, Jun. 2024, doi: 10.22146/mi.v33i1.11481.
[15] H. Yin, A. Aryani, S. Petrie, A. Nambissan, A. Astudillo, and S. Cao, "A rapid review of clustering algorithms," Array, vol. 30, p. 100904, Jul. 2026, doi: 10.1016/j.array.2026.100904.
[16] Sherly Rosa Anggraeni, "Implementasi Algoritma Clustering DBSCAN terhadap Pola Navigasi Pengguna di Perpustakaan Digital untuk Mengungkap Zona Buta Akses Informasi dan Optimalisasi Antarmuka Sistem," SKANIKA: Sistem Komputer dan Teknik Informatika, vol. 8, no. 2, pp. 232–243, Jul. 2025, doi: 10.36080/skanika.v8i2.3524.
[17] R. Sigit, "Penerapan Algoritma K-Means Clustering dalam Menganalisis Pola Peminjaman Buku di Perpustakaan," The Indonesian Journal of Computer Science, vol. 13, no. 5, Oct. 2024, doi: 10.33022/ijcs.v13i5.4317.
[18] Andre, N. Suciati, H. Fabroyir, and E. Pardede, "Educational Data Mining Clustering Approach: Case Study of Undergraduate Student Thesis Topic," IEEE Access, vol. 11, pp. 130072–130088, 2023, doi: 10.1109/ACCESS.2023.3332818.
[19] A. Sharma, R. K. Gupta, and A. Tiwari, "Improved Density Based Spatial Clustering of Applications of Noise Clustering Algorithm for Knowledge Discovery in Spatial Data," Math. Probl. Eng., vol. 2016, pp. 1–9, 2016, doi: 10.1155/2016/1564516.
[20] N. P. Sutramiani, I. M. T. Arthana, P. F. Lampung, S. Aurelia, M. Fauzi, and I. W. A. S. Darma, "The Performance Comparison of DBSCAN and K-Means Clustering for MSMEs Grouping based on Asset Value and Turnover," Journal of Information Systems Engineering and Business Intelligence, vol. 10, no. 1, pp. 13–24, Feb. 2024, doi: 10.20473/jisebi.10.1.13-24.
[21] S. P. Tamba, M. D. Batubara, W. Purba, M. Sihombing, V. M. Mulia Siregar, and J. Banjarnahor, "Book data grouping in libraries using the k-means clustering method," J. Phys. Conf. Ser., vol. 1230, no. 1, p. 012074, Jul. 2019, doi: 10.1088/1742-6596/1230/1/012074.
[22] T. Xu, "Optimization of Big Data and Intelligent Algorithms in Library Resource Management and Utilization," in Proceedings of the 2025 4th International Conference on Artificial Intelligence and Education, New York, NY, USA: ACM, Nov. 2025, pp. 200–205. doi: 10.1145/3797552.3797586.
[23] Sherly Rosa Anggraeni, "Implementasi Algoritma Clustering DBSCAN terhadap Pola Navigasi Pengguna di Perpustakaan Digital untuk Mengungkap Zona Buta Akses Informasi dan Optimalisasi Antarmuka Sistem," SKANIKA: Sistem Komputer dan Teknik Informatika, vol. 8, no. 2, pp. 232–243, Jul. 2025, doi: 10.36080/skanika.v8i2.3524.
[24] S. A. Pratama, "Pengembangan Sistem Rekomendasi Buku Menggunakan Collaborative Filtering," Jurnal Komputer, vol. 2, no. 2, pp. 81–86, Jun. 2024, doi: 10.70963/jk.v2i2.112.
[25] D. Udariansyah, "Recommender System for Book Review based on Clustering Algorithms," Journal of Applied Data Sciences, vol. 6, no. 1, pp. 225–235, Jan. 2024, doi: 10.47738/jads.v6i1.492.