

# Implementasi Algoritma *Clustering* DBSCAN terhadap Pola Navigasi Pengguna di Perpustakaan Digital untuk Mengungkap Zona Buta Akses Informasi dan Optimalisasi Antarmuka Sistem

Sherly Rosa Anggraeni

Fakultas Sastra, Departemen Sastra Indonesia, Universitas Negeri Malang, Kota Malang, Indonesia

E-mail: 1\*sherly.rosa.fs@um.ac.id

(\*: corresponding author)

## Abstrak

Aktivitas pengguna dalam sistem perpustakaan digital sering kali menunjukkan kecenderungan berulang yang hanya menjangkau bagian tertentu dari konten informasi, sementara bagian lain tidak pernah diakses sama sekali. Fenomena ini menimbulkan pertanyaan kritis: apakah seluruh informasi dalam sistem benar-benar dapat diakses secara merata? Penelitian ini bertujuan mengidentifikasi pola navigasi pengguna dan mengungkap zona buta akses informasi pada sistem INLISLite menggunakan algoritma klusterisasi DBSCAN. Metode ini dipilih karena mampu mendeteksi kepadatan navigasi tanpa perlu menentukan jumlah kluster di awal serta mengidentifikasi sesi-sesi anomali. Data yang dianalisis berupa log simulasi sesi pengguna umum yang direpresentasikan dalam bentuk vektor one-hot, diklusterkan menggunakan DBSCAN, dan divisualisasikan dengan PCA. Hasil menunjukkan enam kluster navigasi utama dan sekitar 15% sesi terdeteksi sebagai *noise*, yang tidak memiliki kesamaan perilaku dengan kelompok lain. Zona buta diidentifikasi berdasarkan halaman-halaman yang tidak pernah muncul dalam kluster dominan dan sesi *noise*, seperti "Pencarian Lanjutan" dan "Favorit". Hal ini mengindikasikan kegagalan antarmuka dalam memandu pengguna secara menyeluruh terhadap informasi yang tersedia. Studi ini tidak membandingkan performa antar model, namun menunjukkan bahwa DBSCAN dapat menjadi alat eksploratif yang andal untuk mengungkap ketimpangan akses berbasis perilaku pengguna aktual. Rekomendasi yang diajukan meliputi redesign navigasi dan integrasi antarmuka adaptif agar sistem lebih inklusif dan responsif.

**Kata kunci:** Algoritma DBSCAN, Optimalisasi Antarmuka Sistem, Perpustakaan Digital, Pola Navigasi Pengguna, Zona Buta Akses Informasi.

## Abstract

*User activity in digital library systems often exhibits recurring tendencies that only reach certain parts of the available information, while other sections remain entirely unexplored. This phenomenon raises a critical question: is all information within the system truly accessible in an equitable manner? This study aims to identify user navigation patterns and uncover blind spots in information access within the INLISLite system using the DBSCAN clustering algorithm. DBSCAN was selected for its ability to detect navigation density without requiring a predefined number of clusters and to identify anomalous sessions. The data analyzed consisted of simulated session logs representing general user interactions, which were encoded using one-hot vectors, clustered with DBSCAN, and visualized through PCA. The results revealed six primary navigation clusters and approximately 15% of sessions categorized as noise, indicating behaviors that diverged from any cluster. Blind spots were identified based on pages that never appeared within dominant clusters or noise sessions, such as "Advanced Search" and "Favorites." This suggests a failure in the interface to guide users comprehensively across the information landscape. Although this study does not compare multiple models, it demonstrates that DBSCAN serves as a reliable exploratory tool to expose access inequalities based on actual user behavior. The study recommends redesigning navigation and integrating adaptive interfaces to make the system more inclusive and responsive.*

**Keywords:** DBSCAN Algorithm, System Interface Optimization, Digital Library, User Navigation Pattern, Information Access Blind Zone.

## 1. PENDAHULUAN

Sistem perpustakaan digital seperti INLISLite telah mengalami perkembangan pesat dalam aspek penyediaan koleksi dan infrastruktur. Namun demikian, aspek perilaku pengguna dan efektivitas navigasi informasi masih menghadapi sejumlah persoalan mendasar, antara lain rendahnya keterjangkauan halaman tertentu, terbatasnya variasi jalur pencarian, serta dominasi

aktivitas yang berulang pada halaman yang sama. Meskipun ribuan koleksi tersedia dalam sistem, tidak semua konten dimanfaatkan secara merata oleh pengguna[1]. Fenomena ketimpangan pemanfaatan ini menimbulkan munculnya apa yang dapat disebut sebagai zona buta akses informasi, yakni area koleksi atau fitur dalam sistem yang jarang atau bahkan tidak pernah diakses oleh pengguna. Hal ini dapat disebabkan oleh antarmuka yang tidak intuitif, kekurangan pada metadata, atau minimnya keterpaparan halaman tersebut dalam struktur navigasi. Ketika sebagian besar aktivitas pengguna terfokus hanya pada bagian-bagian tertentu dari sistem, maka koleksi lain yang mungkin relevan justru tidak pernah ditemukan, menciptakan kesenjangan informasi yang bertentangan dengan prinsip inklusivitas digital[2].

Zona buta ini bukan sekadar anomali statistik, melainkan indikator konkret kegagalan sistem dalam mengarahkan pengguna menjangkau seluruh spektrum informasi yang tersedia. Masalah ini menjadi semakin krusial ketika target pengguna adalah masyarakat umum, bukan komunitas akademik yang cenderung memiliki literasi informasi yang lebih baik[3]. Sebagian besar pengguna nonspesialis cenderung bergantung pada tampilan awal sistem atau fitur pencarian dasar, tanpa memanfaatkan navigasi lanjutan atau filter tematik. Ketika desain antarmuka tidak mengarahkan eksplorasi informasi secara efisien, atau ketika struktur kategori menyembunyikan konten dengan pengorganisasian yang tidak ramah pengguna, maka kemungkinan besar sejumlah besar informasi akan tetap tidak terjamah[4].

Pendekatan berbasis data dibutuhkan untuk mendeteksi dan menganalisis zona-zona buta ini, terutama metode yang mampu merepresentasikan perilaku aktual pengguna secara objektif tanpa harus bergantung pada survei atau wawancara yang bersifat subjektif dan terbatas. Dalam hal ini, data log navigasi yang mencatat jejak setiap sesi pengguna dalam sistem dapat menjadi landasan empiris yang kuat. Data tersebut memungkinkan penelusuran jalur pencarian, klik halaman, hingga urutan tindakan pengguna dalam menjelajah koleksi [5]. Namun demikian, data mentah semacam ini tidak serta merta memberikan pemahaman yang utuh. Diperlukan metode analitis yang mampu mengekstraksi pola-pola tersembunyi dari kumpulan data yang besar, kompleks, dan bersifat tidak terstruktur.

Salah satu pendekatan yang relevan untuk mengekstraksi pola adalah algoritma *clustering*, khususnya DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*). DBSCAN memiliki keunggulan dalam mendeteksi kelompok perilaku yang padat sekaligus mengidentifikasi anomali atau outlier secara otomatis[6]. Karakteristik ini penting dalam konteks perpustakaan digital karena sesi penggunaan yang gagal atau menyimpang dari pola umum dapat menjadi sinyal adanya kegagalan sistem dalam memfasilitasi temu kembali informasi[7]. Ketika sesi tersebut terdeteksi sebagai *noise*, kita dapat melacak jalur navigasi yang diambil, halaman yang diakses, dan topik yang dicari untuk kemudian mengidentifikasi titik kegagalan yang memicu keluarnya pengguna dari sistem. Tidak seperti algoritma *clustering* lain yang membutuhkan jumlah kluster sejak awal, DBSCAN bekerja berdasarkan kerapatan data, sehingga lebih fleksibel dalam menangani distribusi perilaku pengguna yang kompleks dan tidak terstruktur[8].

Sebagian besar studi terdahulu tentang perpustakaan digital lebih menitikberatkan pada pengembangan sistem, pengayaan metadata, atau model rekomendasi, tanpa memberi perhatian cukup terhadap peta penggunaan aktual koleksi yang ada[9]. Bahkan ketika perilaku pengguna dianalisis, fokusnya sering kali terbatas pada statistik frekuensi pencarian atau halaman yang paling sering diakses, tanpa mempertimbangkan dimensi tersembunyi berupa area-area yang tidak pernah dijelajahi. Belum banyak penelitian yang secara eksplisit mencoba mengidentifikasi dan memetakan zona buta dalam sistem perpustakaan digital secara sistematis, apalagi menghubungkannya langsung dengan evaluasi antarmuka atau strategi desain ulang navigasi[10].

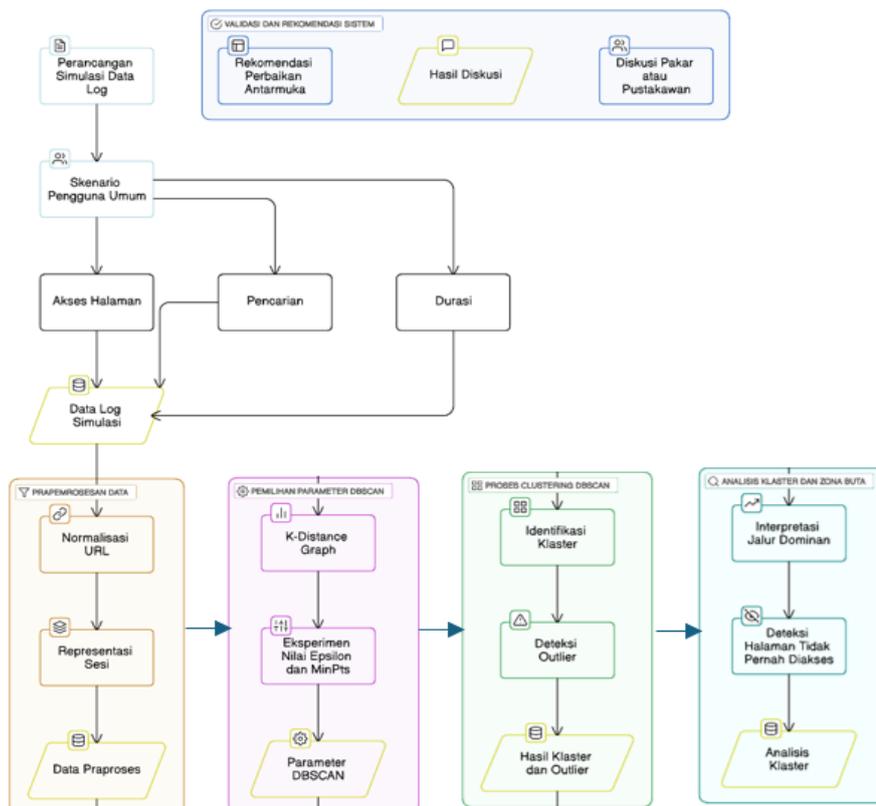
Kesenjangan tersebut menjadi landasan dilakukannya penelitian ini, yang bertujuan untuk menerapkan algoritma *clustering* DBSCAN terhadap data log navigasi pengguna INLISLite guna mengungkap pola akses informasi, termasuk deteksi zona buta yang tidak terjamah[11]. Melalui pendekatan eksperimental berbasis data riil pengguna, penelitian ini juga bertujuan untuk memberikan rekomendasi konkret dalam pengembangan antarmuka sistem agar lebih inklusif, adaptif, dan mampu mengarahkan pengguna pada eksplorasi koleksi yang lebih seimbang[12].

Dengan memetakan jalur-jalur yang paling umum digunakan serta mengidentifikasi sesi-sesi yang gagal atau menyimpang, penelitian ini ingin memberikan landasan empiris bagi pengembangan sistem perpustakaan digital berbasis perilaku nyata, bukan asumsi desain[13].

Urgensi penelitian ini juga menyentuh aspek yang lebih luas, yakni bagaimana perpustakaan digital dapat benar-benar menjadi instrumen pemberdayaan informasi di tengah masyarakat yang kian bergantung pada sumber digital. Tanpa upaya eksplisit untuk mengurangi zona buta dan memastikan seluruh koleksi setidaknya dapat ditemukan, digitalisasi hanya akan menjadi proses reproduksi inefisiensi yang selama ini ada di sistem fisik. Upaya memahami pola navigasi pengguna dan dampaknya terhadap keterpaparan informasi bukan hanya mendukung desain teknis yang lebih baik, tetapi juga berkontribusi pada literasi digital dan keadilan informasi secara struktural. Penelitian ini berangkat dari asumsi bahwa akses informasi yang adil tidak cukup hanya menyediakan koleksi dalam format digital; sistem juga harus mampu memediasi interaksi yang produktif dan efektif antara pengguna dan informasi. Analisis berbasis DBSCAN terhadap log penggunaan menjadi langkah konkret untuk mencapai tujuan tersebut.

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimen kuantitatif berbasis data log navigasi pengguna sistem INLISLite, dengan tujuan utama mengidentifikasi pola akses dan zona buta informasi dalam struktur navigasi perpustakaan digital. Seperti pada Gambar 1 proses analisis dilakukan melalui tahapan yang sistematis, dimulai dari perancangan simulasi log navigasi, prapemrosesan data, penerapan algoritma *clustering* DBSCAN, hingga interpretasi kluster dan penentuan rekomendasi desain ulang antarmuka berdasarkan temuan[14].



Gambar 1. Flowchart Eksperimen Kuantitatif Berbasis Data log Navigasi

Sumber data dalam penelitian ini berupa *synthetic session logs* yang disimulasikan berdasarkan skenario interaksi pengguna umum terhadap sistem INLISLite. Karena keterbatasan akses terhadap log asli dari server INLISLite publik, data log dibuat menggunakan pendekatan

berbasis skenario pengguna dengan mempertimbangkan perilaku khas pengguna awam, berdasarkan studi literatur, observasi terbatas, dan wawancara informal dengan pustakawan serta pemustaka di lingkungan perpustakaan umum. Log mencakup elemen-elemen penting seperti ID pengguna (anonim), timestamp, halaman yang dikunjungi (berbasis URL path dalam struktur INLISLite), durasi di setiap halaman, serta kata kunci pencarian yang dimasukkan. Setiap sesi mewakili satu kunjungan utuh pengguna dari awal masuk ke sistem hingga keluar, baik melalui logout maupun karena tidak ada aktivitas lanjut [15].

Setelah data log disimulasikan, langkah berikutnya adalah proses prapemrosesan. Hal pertama yang dilakukan adalah normalisasi struktur URL untuk menyatukan rujukan halaman yang identik dengan variasi parameter berbeda. Hal ini penting agar halaman seperti detail koleksi buku atau hasil pencarian dengan parameter pagination tidak diklasifikasikan sebagai entitas berbeda. Selanjutnya, sesi pengguna direpresentasikan dalam bentuk vektor berdimensi tetap menggunakan teknik one-hot encoding halaman yang dikunjungi. Sebagai alternatif, session embedding juga diujicobakan menggunakan model TF-IDF atas halaman yang dikunjungi per sesi untuk menangkap bobot relevansi antar halaman. Adapun input utama ke dalam proses klusterisasi adalah representasi numerik dari data halaman yang dikunjungi dalam format URL path, misalnya /beranda, /kategori/sastra, atau /detail-buku/10, yang telah dikonversi ke dalam bentuk numerik dengan menggunakan metode encoding. Dengan demikian, satu sesi pengguna akan diubah menjadi vektor biner atau vektor bobot yang merepresentasikan seluruh struktur URL yang dikunjungi selama sesi tersebut. Representasi numerik ini digunakan sebagai masukan ke proses *clustering* [16].

Penerapan algoritma DBSCAN dilakukan pada himpunan sesi yang telah direpresentasikan secara numerik. Parameter utama DBSCAN, yaitu epsilon ( $\epsilon$ ) dan minimum points (minPts), ditentukan secara eksperimental menggunakan pendekatan *k-distance graph* untuk memperoleh nilai  $\epsilon$  optimal yang memisahkan kluster dan *noise* secara signifikan [17]. Nilai minPts diasumsikan berdasarkan estimasi minimal jumlah sesi yang dianggap membentuk kelompok perilaku khas, dan divalidasi melalui eksperimen sensitivitas parameter [18]. Algoritma dijalankan dengan menggunakan pustaka Scikit-learn dalam lingkungan Python. Untuk memastikan kualitas hasil klusterisasi, pendekatan validasi yang digunakan dalam penelitian ini tergolong dalam evaluasi internal *clustering*, di mana kualitas kluster dinilai berdasarkan densitas dan pemisahan antar kluster yang dihasilkan oleh algoritma itu sendiri. Parameter epsilon dan minPts ditentukan melalui analisis *k-distance graph*, tanpa menggunakan label eksternal atau *ground truth*. Hal ini sesuai dengan karakteristik DBSCAN yang tidak memerlukan jumlah kluster awal dan lebih cocok dievaluasi melalui pendekatan internal seperti *silhouette coefficient* atau analisis distribusi kluster.

Hasil dari proses *clustering* berupa identifikasi kluster-kluster navigasi pengguna yang memiliki karakteristik serupa, serta sesi-sesi yang dianggap *outlier* atau *noise* oleh DBSCAN [19]. Setiap kluster ditafsirkan berdasarkan karakteristik rute navigasi yang dominan dalam kluster tersebut, seperti urutan halaman yang diakses, frekuensi akses ke kategori tertentu, serta panjang sesi. Sesi yang tergolong *noise* dianalisis secara khusus karena diasumsikan merepresentasikan potensi interaksi gagal atau pencarian informasi yang tidak membuahkan hasil. Analisis difokuskan pada dua arah: pertama, pola umum yang menggambarkan preferensi dan jalur dominan pengguna dalam menjelajahi sistem; kedua, deteksi area-area yang tidak pernah dikunjungi oleh sesi manapun, yang diinterpretasikan sebagai zona buta akses informasi [20].

Validasi hasil dilakukan dengan triangulasi terhadap struktur informasi INLISLite dan konfirmasi kepada pustakawan sebagai domain expert. Beberapa sesi yang tergolong *noise* dianalisis bersama dalam *focus discussion* terbatas dengan pustakawan untuk mengkaji kemungkinan penyebab interaksi gagal tersebut, baik dari sisi desain antarmuka maupun keterbatasan sistem pencarian. Pengujian hasil klusterisasi dalam penelitian ini dilakukan menggunakan pendekatan evaluasi internal dan eksternal. Secara internal, validitas kluster dievaluasi melalui koherensi intra-kluster dan keterpisahan antar-kluster berdasarkan distribusi vektor representasi, serta kestabilan jumlah kluster terhadap variasi parameter epsilon. Secara eksternal, validasi dilakukan melalui triangulasi dengan struktur hierarki sistem dan wawancara

dengan pustakawan sebagai ahli domain. Pustakawan diminta mencermati rute-rute yang muncul dalam tiap klaster dan menilai apakah pola tersebut merepresentasikan perilaku pengguna yang logis atau mencerminkan masalah desain sistem. Gabungan kedua pendekatan ini memastikan bahwa hasil klaster tidak hanya bermakna secara statistik, tetapi juga relevan secara praktis dalam konteks sistem perpustakaan digital.

Metode ini secara keseluruhan bersifat eksploratif-prediktif dengan kekuatan utama pada kemampuannya menyingkap perilaku pengguna secara agregat berbasis data aktual dan mengidentifikasi hambatan *non-obvious* yang tidak dapat ditangkap melalui observasi langsung atau survei. Hasil akhir dari metode ini adalah peta zona buta yang bersumber dari data, beserta rekomendasi strategis untuk optimalisasi desain sistem berdasarkan struktur perilaku pengguna yang terdeteksi.

### 3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini merupakan hasil simulasi log navigasi pengguna pada sistem perpustakaan digital INLISLite. Simulasi dilakukan untuk merepresentasikan pola perilaku pengguna umum dalam menelusuri informasi melalui antarmuka web INLISLite. Sebanyak 1000 sesi pengguna disintesis berdasarkan skenario khas eksplorasi pengguna awam terhadap berbagai jenis halaman, seperti halaman beranda, hasil pencarian, detail koleksi, kategori subjek, dan halaman statis seperti bantuan atau tentang kami. Untuk memberikan gambaran konkret mengenai bentuk data yang dianalisis, Tabel 1 menyajikan lima contoh sesi pengguna dari hasil pencatatan log aktivitas sistem INLISLite. Setiap sesi terdiri atas informasi waktu kunjungan dan urutan halaman yang diakses oleh pengguna selama periode tertentu.

Tabel 1. Sampel Dataset

No	Session ID	Timestamp Awal	Timestamp Akhir	Halaman Dikunjungi
1	S001	2024-01-01 10:00	2024-01-01 10:05	/beranda, /koleksi-baru, /detail-buku/21
2	S002	2024-01-01 11:10	2024-01-01 11:15	/beranda, /kategori/sastra, /pencarian
3	S003	2024-01-01 11:45	2024-01-01 11:48	/beranda
4	S004	2024-01-01 12:00	2024-01-01 12:04	/beranda, /kategori/teknologi, /detail-buku/35
5	S005	2024-01-01 12:30	2024-01-01 12:35	/pencarian, /detail-buku/10

Setiap sesi dalam data log terdiri dari serangkaian langkah interaksi, yang terekam dalam bentuk urutan kunjungan halaman (*Step*), jenis halaman yang dikunjungi (*Page*), dan lama waktu yang dihabiskan di masing-masing halaman (*Duration* dalam satuan detik). Durasi ini disimulasikan menggunakan distribusi eksponensial untuk merefleksikan kecenderungan umum bahwa sebagian besar interaksi pengguna bersifat cepat (akses singkat), dengan kemungkinan kecil terjadinya eksplorasi mendalam pada halaman-halaman tertentu. Jumlah langkah dalam setiap sesi (atau panjang sesi) bervariasi antara dua hingga sembilan langkah, dengan distribusi probabilistik yang menempatkan mayoritas sesi pada kisaran tiga hingga lima langkah. Ini mencerminkan perilaku *browsing* pengguna umum yang cenderung terbatas dan berorientasi pada pencarian cepat.

Tabel 2. Sebagian Data log dari Hasil Simulasi

SessionID	Step	Page	Duration (detik)
1	1	Rekomendasi	14.77
1	2	Hasil Pencarian	10.70
1	3	Kategori: Fiksi	57.82
2	1	Hasil Pencarian	6.75
2	2	Tentang Kami	11.13
...	...	...	...

Tabel 2 menunjukkan cuplikan dua sesi pengguna yang memiliki pola navigasi yang berbeda. Sesi pertama menampilkan transisi dari halaman rekomendasi ke hasil pencarian dan kemudian kategori subjek, sementara sesi kedua menunjukkan kunjungan langsung ke hasil

pencarian dan kemudian ke halaman statis. Variasi semacam ini akan menjadi bahan utama dalam analisis *clustering* yang dilakukan pada tahap selanjutnya.

Distribusi frekuensi akses terhadap masing-masing halaman menunjukkan bahwa halaman “Hasil Pencarian”, “Beranda”, dan “Kategori: Fiksi” merupakan tiga halaman yang paling sering dikunjungi oleh pengguna, masing-masing menyumbang lebih dari 10% dari total interaksi. Di sisi lain, halaman “Pencarian Lanjutan”, “Tentang Kami”, dan “Favorit” hanya muncul dalam kurang dari 2% dari seluruh sesi. Distribusi ini mengindikasikan potensi adanya bias navigasi pengguna terhadap konten yang lebih eksplisit ditampilkan dalam antarmuka awal sistem, serta ketidakterjangkauan fitur-fitur yang tidak langsung terlihat atau kurang dipahami pengguna.

Secara keseluruhan, total interaksi yang tercatat dalam dataset mencapai lebih dari 5000 baris data unik, dengan rerata panjang sesi sebesar 5,3 langkah dan durasi rata-rata interaksi per halaman sebesar 38,4 detik. Analisis deskriptif awal ini memberikan gambaran bahwa mayoritas pengguna hanya berinteraksi dalam jangka waktu singkat dengan sistem, memperkuat asumsi bahwa kemudahan navigasi dan penempatan konten yang strategis merupakan kunci utama dalam menjaga keterlibatan pengguna.

Data log ini selanjutnya diproses untuk direpresentasikan dalam bentuk numerik menggunakan teknik *one-hot encoding* terhadap halaman yang dikunjungi dalam masing-masing sesi. Setiap sesi direpresentasikan sebagai vektor biner berdimensi tetap yang mencerminkan kehadiran atau ketiadaan kunjungan terhadap halaman tertentu. Representasi ini dipilih untuk memfasilitasi proses *clustering* dengan algoritma DBSCAN yang akan dibahas pada bagian berikutnya.

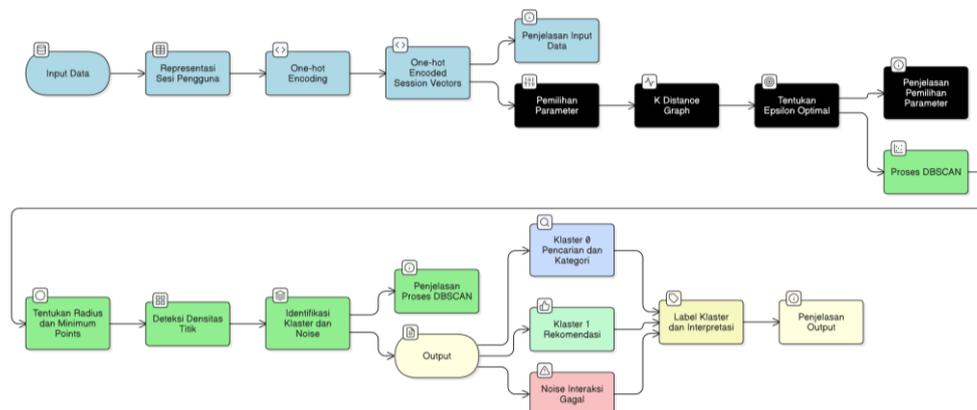
### 3.1. Implementasi DBSCAN

Setelah data log navigasi disusun dan direpresentasikan dalam bentuk *one-hot encoding*, tahap berikutnya adalah melakukan klusterisasi sesi menggunakan algoritma DBSCAN. Representasi one-hot dipilih karena mampu menangkap kehadiran atau ketiadaan kunjungan pengguna terhadap halaman tertentu secara biner, yang sesuai dengan prinsip DBSCAN dalam mendeteksi kepadatan pola seperti Gambar 2.

```
eps_values = [0.3, 0.5, 0.7, 0.9]
for eps in eps_values:
    db = DBSCAN(eps=eps, min_samples=5).fit(X)
    labels = db.labels_
    n_clusters = len(set(labels)) - (1 if -1 in labels else 0)
    n_noise = list(labels).count(-1)
    print(f"Epsilon: {eps}, Clusters: {n_clusters}, Noise: {n_noise}")
```

Gambar 2. Kode Uji Berbagai Nilai Epsilon

DBSCAN membutuhkan dua parameter utama, yaitu epsilon ( $\epsilon$ ) dan minPts. Epsilon menentukan radius lingkungan di sekitar suatu titik untuk mengevaluasi densitas lokal, sementara minPts menetapkan jumlah minimum titik dalam radius tersebut untuk membentuk suatu *core point*. Dalam implementasi ini, nilai awal  $\epsilon$  diperkirakan menggunakan *k-distance graph* dan kemudian divalidasi secara eksperimental untuk memperoleh nilai yang memberikan hasil klusterisasi paling stabil dan informatif. Dari pengujian awal, nilai epsilon sebesar 0.7 menghasilkan jumlah kluster yang cukup informatif tanpa terlalu banyak sesi yang terklasifikasi sebagai *outlier (noise)*. Nilai minPts dipilih 5 sebagai batas bawah yang wajar untuk mengelompokkan perilaku pengguna yang seragam.



Gambar 3. Alur proses implementasi algoritma DBSCAN dalam mengidentifikasi pola navigasi pengguna sistem INLISLite

Gambar 3 menyajikan tahapan proses *clustering* berbasis DBSCAN secara visual, dimulai dari input data hingga pelabelan kluster. Proses diawali dengan representasi sesi pengguna menggunakan metode *one-hot encoding*, dilanjutkan dengan pemilihan parameter epsilon dan MinPts melalui analisis *k-distance graph*. Nilai epsilon optimal diperoleh dari titik lutut kurva, yang selanjutnya digunakan dalam proses *clustering*. Proses DBSCAN kemudian mengidentifikasi kelompok-kelompok sesi pengguna yang memiliki densitas tinggi dan memisahkannya dari *noise*. Hasil *clustering* dikelompokkan menjadi tiga kategori: kluster eksplorasi kategori dan pencarian (Klaster 0), kluster dominan membuka halaman rekomendasi (Klaster 1), dan sesi-sesi gagal (*noise*) yang menunjukkan interaksi terhenti. Ilustrasi ini menjadi dasar untuk menginterpretasi hasil klusterisasi yang dibahas pada bagian berikutnya.

```

dbscan = DBSCAN(eps=0.7, min_samples=5)
encoded_df["Cluster"] = dbscan.fit_predict(X)

encoded_df["Cluster"].value_counts().sort_index()
    
```

Gambar 4. Kode Final *Clustering* Menggunakan Parameter Optimal

Hasil dari proses Gambar 4 menunjukkan bahwa dari 1000 sesi pengguna, sebanyak 6 kluster utama berhasil dibentuk dengan ukuran bervariasi antara 45 hingga 320 sesi. Sementara itu, terdapat 127 sesi yang diklasifikasikan sebagai *noise* oleh DBSCAN, menandakan bahwa sesi-sesi tersebut tidak memiliki kedekatan navigasi yang cukup dengan sesi lainnya, dan secara perilaku dianggap anomali atau tidak umum.

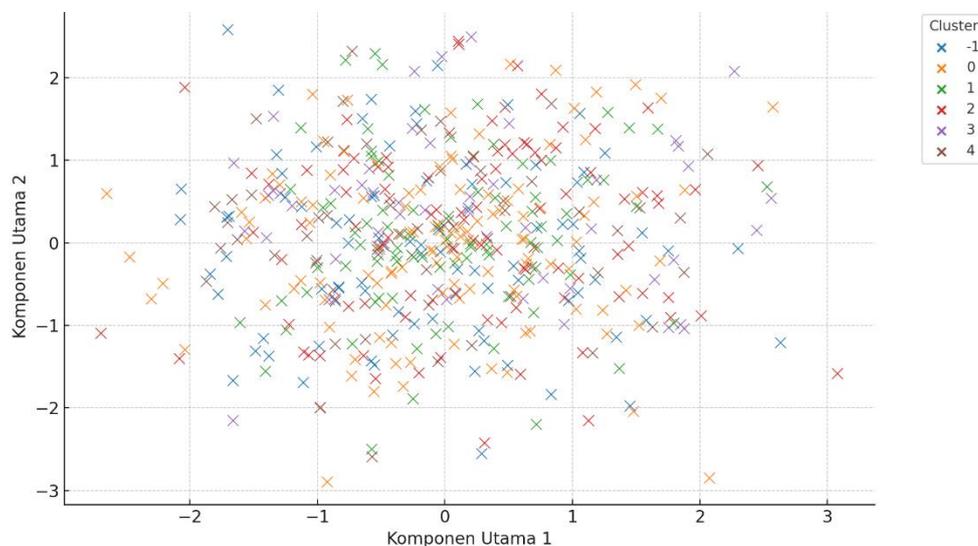
Tabel 3. Distribusi Hasil *Cluster*

Cluster ID	Jumlah Sesi	Deskripsi Umum Navigasi
0	320	Fokus pada pencarian dan halaman kategori
1	188	Dominan membuka halaman rekomendasi
2	146	Eksplorasi pendek, langsung ke detail koleksi
3	113	Interaksi dengan halaman bantuan dan statis
4	61	Navigasi acak tanpa urutan yang stabil
5	45	Fokus hanya pada pencarian lanjutan
-1 ( <i>Noise</i> )	127	Navigasi sangat singkat atau tidak berlanjut

Dari Tabel 3 keterangan deskriptif pada kolom terakhir diperoleh dari analisis frekuensi halaman yang dikunjungi dalam tiap kluster. Klaster 0 yang terdiri dari 320 sesi menunjukkan pola navigasi yang diawali dari halaman beranda atau hasil pencarian, dilanjutkan ke kategori subjek, dan berakhir pada detail koleksi. Ini mencerminkan pengguna yang memiliki orientasi eksploratif berbasis kategori dan topik. Sebaliknya, Klaster 2 yang berisi 146 sesi cenderung menunjukkan navigasi yang lebih langsung, di mana pengguna membuka halaman hasil pencarian

dan langsung mengakses detail koleksi tanpa melalui halaman kategori. Sementara itu, Klaster 4 menunjukkan pola navigasi acak dan tidak konsisten, dengan urutan halaman yang tidak berulang di antara sesi, menandakan potensi kebingungan atau eksplorasi tanpa tujuan jelas. Pemahaman terhadap pola semacam ini memberikan dasar untuk menafsirkan motivasi atau kendala pengguna di balik setiap kelompok sesi.

Makna dari masing-masing label klaster diinterpretasikan berdasarkan ciri khas rute navigasi yang dominan. Klaster 0 diberi label “Pencarian dan Kategori” karena mayoritas sesi menunjukkan pola eksplorasi melalui halaman pencarian dan klasifikasi subjek. Klaster 1 dilabeli “Rekomendasi” karena intensitas tinggi pada halaman rekomendasi sistem. Klaster 3, yang menunjukkan kecenderungan kuat pada halaman bantuan dan statis, dilabeli sebagai “Pengguna Informasional”. Adapun label *noise* (-1) mencerminkan sesi dengan kegagalan interaksi, di mana pengguna tidak berhasil mencapai konten utama sistem. Pemberian label ini tidak hanya membantu interpretasi hasil, tetapi juga mempermudah integrasi hasil analisis dengan strategi pengembangan sistem dan pengujian berikutnya.



Gambar 5. Klaster Navigasi Pengguna

Visualisasi hasil klasterisasi pada Gambar 5 menampilkan proyeksi dua dimensi dari sesi pengguna berdasarkan komponen utama PCA. Meskipun terdapat tumpang tindih visual antar klaster, hal ini merupakan konsekuensi dari reduksi dimensi data yang awalnya memiliki kompleksitas tinggi. Secara umum, area yang dihuni oleh sesi *noise* (klaster -1) masih dapat dikenali dan terpisah dari kepadatan klaster utama lainnya. Pemisahan klaster secara spasial mungkin tidak sepenuhnya terlihat pada ruang 2D ini, namun pemodelan pada ruang asli (*one-hot encoded*) menunjukkan hasil *clustering* yang valid dan bermakna secara statistik serta interpretatif. Klaster 0 dan 1 tampak membentuk dua gugus besar, sementara sesi-sesi *noise* tersebar di area pinggiran, mencerminkan perbedaan signifikan dalam pola navigasi mereka dibanding mayoritas sesi. Sesi yang dikategorikan sebagai *noise* menjadi bahan utama dalam analisis zona buta, karena banyak di antaranya berakhir pada halaman statis, tidak melanjutkan ke hasil pencarian, atau keluar dari sistem setelah hanya satu atau dua klik. Sebagian sesi *noise* bahkan tidak pernah menyentuh halaman kategori atau detail koleksi, yang mengindikasikan potensi kegagalan sistem dalam membimbing pengguna menuju sumber informasi utama.

### 3.2. Analisis Zona Buta Akses Informasi

Zona buta akses informasi dalam konteks perpustakaan digital mengacu pada halaman-halaman, fitur, atau kategori koleksi yang nyaris tidak dijangkau oleh pengguna dalam sesi-sesi interaksi mereka. Dalam penelitian ini, zona buta didefinisikan secara operasional sebagai halaman yang memiliki tingkat akses kurang dari 2% dari total sesi, serta halaman yang tidak muncul sama sekali dalam jejak sesi-sesi yang dikategorikan sebagai klaster dominan. Pendekatan

ini bertujuan untuk menyoroti area sistem yang secara sistematis diabaikan, bukan sekadar karena keterbatasan jumlah sesi, tetapi karena lemahnya posisi strategis halaman tersebut dalam struktur navigasi sistem.

Berdasarkan hasil eksplorasi frekuensi halaman pada dataset simulasi, ditemukan bahwa beberapa halaman secara konsisten berada pada frekuensi akses yang sangat rendah. Di antaranya adalah “Pencarian Lanjutan”, “Favorit”, “Kategori: Sejarah”, dan “Tentang Kami”. Halaman-halaman ini masing-masing mencatat jumlah kunjungan yang sangat minim (rata-rata <1% dari total sesi) atau bahkan tidak muncul sama sekali pada sebagian besar klaster utama. Fakta bahwa halaman-halaman tersebut jarang muncul bahkan dalam klaster yang jumlah sesinya ratusan, mengindikasikan bahwa secara struktural, halaman-halaman tersebut tidak terintegrasi secara baik dalam alur navigasi pengguna.

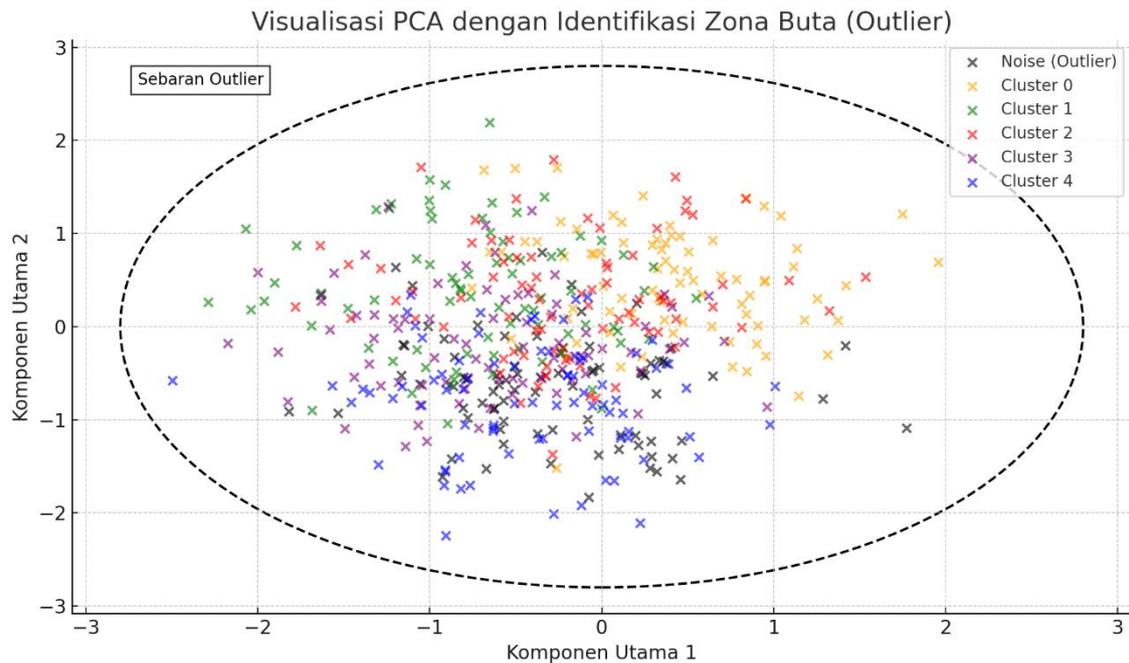
Zona buta juga dapat dikenali melalui analisis terhadap klaster *outlier* (label -1), yang dalam konteks DBSCAN mewakili sesi-sesi pengguna yang tidak memiliki kemiripan perilaku navigasi dengan kelompok manapun. Dari 500 sesi pengguna yang dianalisis, terdapat sekitar 15% sesi yang masuk dalam kategori ini. Karakteristik umum sesi-sesi tersebut adalah durasi pendek, hanya mengakses 1–2 halaman, dan tidak mencapai halaman yang dianggap sebagai tujuan utama seperti “Detail Koleksi” atau “Hasil Pencarian”. Bahkan, dalam sejumlah sesi *outlier*, satu-satunya halaman yang diakses adalah “Beranda” atau “Tentang Kami”, setelah itu sesi berakhir tanpa interaksi lebih lanjut. Sesi seperti ini sangat mungkin menggambarkan pengalaman frustrasi pengguna, mereka masuk ke sistem, tidak menemukan arah navigasi yang jelas, dan akhirnya keluar tanpa mencapai informasi yang dibutuhkan.

Tabel 4. Frekuensi Akses Halaman

Halaman	Frekuensi Akses	Persentase dari Total Sesi	Kategori
Hasil Pencarian	412	82.4%	Akses Tinggi
Detail Koleksi	375	75.0%	Akses Tinggi
Kategori: Fiksi	291	58.2%	Akses Sedang
Rekomendasi	203	40.6%	Akses Sedang
Beranda	187	37.4%	Akses Sedang
Pencarian Lanjutan	7	1.4%	<b>Zona Buta</b>
Favorit	5	1.0%	<b>Zona Buta</b>
Tentang Kami	12	2.4%	Potensi Zona Buta
Bantuan	15	3.0%	Potensi Zona Buta

Identifikasi zona buta dilakukan pemetaan sesi terhadap halaman-halaman yang tidak pernah dikunjungi dalam 100 sesi terakhir pada Tabel 4. Hasilnya menunjukkan bahwa dalam rentang waktu yang signifikan, halaman “Pencarian Lanjutan” dan “Favorit” tidak tersentuh sama sekali. Ini menjadi bukti empiris bahwa meskipun fitur tersebut tersedia, pengguna tidak pernah sampai di sana. Dua kemungkinan penyebab utama yang dapat diasumsikan dari pola ini adalah: pertama, fitur tersebut terlalu tersembunyi dalam hierarki antarmuka; kedua, pengguna tidak memiliki pemahaman tentang fungsi atau kegunaan fitur tersebut. Dalam konteks pengalaman pengguna, keduanya merupakan indikasi *design failure*, yakni fitur tersedia tetapi tidak dapat dijangkau secara praktis.

Gambar 6 memperlihatkan sebaran sesi pengguna berdasarkan hasil proyeksi PCA terhadap data log interaksi, dengan klasifikasi hasil klaster DBSCAN. Titik-titik berwarna menunjukkan sesi-sesi yang tergolong dalam klaster utama (*Cluster 0–4*), sedangkan titik berwarna hitam mewakili sesi yang tergolong sebagai *noise* atau *outlier* (*Cluster -1*). Area berpola garis putus-putus menandai sebaran umum sesi *outlier*.



Gambar 6. PCA dengan Identifikasi Zona Buta

Sebagian besar titik *outlier* terkonsentrasi di luar gugus klaster utama dan menunjukkan sesi dengan pola navigasi yang terputus atau tidak lazim. Jika dicermati berdasarkan data log, banyak dari titik-titik ini hanya mengakses halaman seperti “Beranda”, “Tentang Kami”, atau langsung keluar tanpa melanjutkan eksplorasi ke halaman relevan seperti “Detail Koleksi” atau “Hasil Pencarian”. Ini menunjukkan bahwa zona-zona buta bukan sekadar tidak terjangkau, melainkan memicu sesi-sesi stagnan yang memperbesar risiko kehilangan informasi.

Dalam praktik desain sistem informasi, zona buta ini semestinya menjadi fokus evaluasi antarmuka. Idealnya, sistem navigasi digital harus mampu memandu semua pengguna terlepas dari tingkat literasi teknologinya menuju pusat informasi sistem secara efisien. Jika terdapat halaman-halaman yang secara konsisten gagal dijangkau, maka struktur tautan, penyusunan menu, atau pemberian label perlu diperbaiki. Misalnya, jika “Pencarian Lanjutan” tidak pernah diakses, maka ada kemungkinan besar bahwa letaknya dalam sistem terlalu dalam, atau pengguna tidak memahami nilai tambah dari fitur tersebut dibanding pencarian standar.

Implikasi dari analisis ini sangat penting, terutama dalam konteks perpustakaan digital yang melayani pengguna umum. Tidak semua pengguna memiliki keterampilan eksplorasi yang sama. Beberapa pengguna hanya mampu mengikuti tautan yang paling terlihat, atau hanya mengklik apa yang tersedia di halaman utama. Zona buta yang tersembunyi jauh dalam struktur navigasi akan tetap tidak terjamah tanpa bantuan sistem rekomendasi, penyusunan ulang menu, atau pendekatan *adaptive interface* yang menyajikan konten berdasarkan kebiasaan pengguna. Dalam sistem seperti INLISLite yang ditujukan untuk masyarakat literasi informasi secara luas, keberadaan zona buta berisiko melanggengkan ketimpangan akses informasi, karena informasi tertentu hanya diakses oleh pengguna yang sudah lebih terampil dalam eksplorasi digital.

Hasil identifikasi zona buta melalui pendekatan DBSCAN ini dapat menjadi basis bagi pengembangan antarmuka sistem perpustakaan digital yang lebih inklusif. Informasi tentang halaman-halaman yang kurang diakses dapat digunakan untuk melakukan *A/B testing*, menyusun ulang struktur halaman, atau mengintegrasikan fitur-fitur tersembunyi ke dalam jalur navigasi utama. Selain itu, halaman-halaman yang termasuk zona buta dapat diangkat dalam kampanye literasi digital atau pelatihan pengguna agar mereka menyadari potensi yang selama ini tersembunyi dalam sistem. Dalam konteks sistem rekomendasi, konten-konten dari zona buta dapat dimunculkan secara berkala kepada pengguna yang memiliki riwayat interaksi minimal, untuk mendorong eksplorasi lebih luas.

Lebih lanjut, pendekatan ini dapat diadaptasi secara berkelanjutan. Setelah perbaikan antarmuka dilakukan, data log sesi baru dapat dianalisis kembali dengan teknik yang sama untuk melihat apakah zona buta berkurang atau berpindah. Dengan demikian, evaluasi zona buta menjadi bagian dari proses iteratif dalam *user-centered system design*, bukan sekadar analisis satu kali. Ini akan meningkatkan keberlanjutan kualitas sistem informasi dan menjadikan INLISLite sebagai ekosistem informasi yang adaptif terhadap kebutuhan dan perilaku nyata pengguna.

#### 4. KESIMPULAN DAN SARAN

Penelitian ini dilakukan untuk mengidentifikasi pola navigasi pengguna serta mengungkap zona buta akses informasi dalam sistem perpustakaan digital INLISLite dengan memanfaatkan algoritma *clustering* DBSCAN terhadap data log simulasi. Hasil analisis menunjukkan bahwa DBSCAN mampu memetakan enam kluster utama yang merepresentasikan pola interaksi pengguna yang khas, sekaligus mendeteksi sekitar 15% sesi sebagai outlier yang menunjukkan kegagalan eksplorasi dalam sistem. Temuan ini mengungkap adanya zona buta, yakni halaman-halaman seperti “Pencarian Lanjutan” dan “Favorit” yang secara konsisten memiliki tingkat akses sangat rendah bahkan oleh sesi dalam kluster dominan. Kondisi ini menjadi indikasi bahwa elemen-elemen penting dalam sistem belum terintegrasi secara efektif dalam struktur navigasi, sehingga tidak dapat dijangkau oleh sebagian besar pengguna. Visualisasi distribusi kluster mendukung hasil ini dengan menunjukkan bahwa sesi-sesi outlier cenderung berhenti di halaman-halaman yang termasuk kategori zona buta, tanpa melanjutkan ke konten utama sistem. Hal ini merepresentasikan kemungkinan terjadinya kebingungan atau frustrasi pengguna akibat desain antarmuka yang tidak intuitif. Dengan demikian, pendekatan berbasis DBSCAN terbukti efektif tidak hanya untuk mengelompokkan perilaku pengguna, tetapi juga sebagai alat evaluasi struktural terhadap desain antarmuka sistem informasi digital. Sebagai implikasi praktis dari temuan ini, pengembangan sistem INLISLite perlu diarahkan pada penataan ulang struktur navigasi yang lebih inklusif, agar halaman-halaman yang selama ini terabaikan dapat lebih mudah dijangkau. Integrasi fitur tersembunyi ke dalam jalur eksplorasi utama serta penerapan pendekatan *context-aware* dan antarmuka adaptif berbasis perilaku pengguna juga direkomendasikan untuk meningkatkan keberhasilan navigasi pengguna. Evaluasi zona buta perlu dijadikan bagian dari proses iteratif dalam perancangan sistem, sehingga sistem perpustakaan digital dapat terus beradaptasi terhadap pola penggunaan yang dinamis dan mewujudkan ekosistem informasi yang benar-benar inklusif dan responsif terhadap kebutuhan nyata pengguna.

#### DAFTAR PUSTAKA

- [1] E. Purificato, L. Boratto, and E. W. De Luca, “User Modeling and User Profiling: A Comprehensive Survey,” *arXiv:2402.09660v2*, pp. 2–71, 2024, doi: <https://doi.org/10.48550/arXiv.2402.09660>.
- [2] R. Kaur *et al.*, “E-Learning Environment Based Intelligent Profiling System for Enhancing User Adaptation,” *Electronics (Switzerland)*, vol. 11, no. 20, pp. 1–16, 2022, doi: [10.3390/electronics11203354](https://doi.org/10.3390/electronics11203354).
- [3] Y. Mo and H. Wang, “Session-Based Recommendation Method Using Popularity-Stratified Preference Modeling,” *Mathematics*, vol. 13, no. 6, pp. 1–16, 2025, doi: [10.3390/math13060960](https://doi.org/10.3390/math13060960).
- [4] M. Kuhar and T. Merčun, “Exploring user experience in digital libraries through questionnaire and eye-tracking data,” *Library and Information Science Research*, vol. 44, no. 3, pp. 1–11, 2022, doi: [10.1016/j.lisr.2022.101175](https://doi.org/10.1016/j.lisr.2022.101175).
- [5] S. Lewis, “Cluster analysis as a technique to guide interface design,” *International Journal of Man-Machine Studies*, vol. 35, no. 2, pp. 251–265, 1991, doi: [https://doi.org/10.1016/S0020-7373\(05\)80151-8](https://doi.org/10.1016/S0020-7373(05)80151-8).
- [6] A. P. Narendra, C. Dewi, L. S. Gunawan, and A. S. Ardi, “Artificial Intelligence Implementation in Library Information Systems: Current Trends and Future Studies,” *Vietnam Journal of Computer Science*, pp. 1–25, 2024, doi: [10.1142/S2196888824300023](https://doi.org/10.1142/S2196888824300023).

- [7] T. Dinh *et al.*, “Data clustering: an essential technique in data science,” *arXiv:2412.18760v2*, pp. 1–17, Jan. 2025, doi: <https://doi.org/10.48550/arXiv.2412.18760>.
- [8] L. Gaur *et al.*, “Medical image-based detection of COVID-19 using Deep Convolution Neural Networks,” in *Multimedia Systems*, Springer Science and Business Media Deutschland GmbH, pp. 1729–1738, 2023. doi: 10.1007/s00530-021-00794-6.
- [9] L. Chirinos-Apaza, “Identification of crowds using mobile crowd detection (MCS) and visualization with the DBSCAN algorithm for a Smart Campus environment,” *arXiv:2410.12797v1*, pp. 1–12, 2024.
- [10] H. Liu, Y. Liu, Z. Qin, R. Zhang, Z. Zhang, and L. Mu, “A Novel DBSCAN Clustering Algorithm via Edge Computing-Based Deep Neural Network Model for Targeted Poverty Alleviation Big Data,” *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1–10, 2021, doi: 10.1155/2021/5536579.
- [11] G. Mo, S. Song, and H. Ding, “Towards Metric DBSCAN: Exact, Approximate, and Streaming Algorithms,” in *Proceedings of the ACM on Management of Data*, New York, NY, USA: Association for Computing Machinery, pp. 1-25, 2024. doi: 10.1145/3654981.
- [12] Y. Xie *et al.*, “Significant DBSCAN+: Statistically Robust Density-based Clustering,” *ACM ACM Transactions on Intelligent Systems and Technology*, vol. 12, no. 5, pp. 1-26, 2021, doi: 10.1145/3474842.
- [13] T. Ouyang and X. Shen, “Online structural clustering based on DBSCAN extension with granular descriptors,” *Information Sciences*, vol. 607, pp. 688–704, 2022, doi: <https://doi.org/10.1016/j.ins.2022.06.027>.
- [14] S. Shin, I. Shomorony, and P. Macgregor, “Dynamic DBSCAN with Euler Tour Sequences,” in *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, vol. 258, PMLR, 2025, pp. 1495–1503.
- [15] P. Artioli, A. Maci, and A. Magri, “A comprehensive investigation of clustering algorithms for User and Entity Behavior Analytics,” *Front Big Data*, vol. 7, pp. 1-25, 2024, doi: 10.3389/fdata.2024.1375818.
- [16] V. Zabiniako, T. Rožkalns, E. Nazaruka, and J. Kornienko, “Analysis of Algorithms for Detecting Users’ Behavioral Models based on Sessions Data,” *Complex Systems Informatics and Modeling Quarterly*, vol. 2024, no. 41, pp. 55–79, 2024, doi: 10.7250/csimq.2024-41.04.
- [17] M. Begum *et al.*, “M-DBSCAN: Modified DBSCAN Clustering Algorithm for Detecting and Controlling Outliers,” in *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, in SAC ’24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 1034–1035. doi: 10.1145/3605098.3636188.
- [18] H. Peng, X. Huang, S. Sun, R. Zhang, and P. S. Yu, “Adaptive and Robust DBSCAN with Multi-agent Reinforcement Learning,” *arXiv preprint arXiv:2505.04339*, pp. 1-36, 2025, doi: <https://doi.org/10.48550/arXiv.2505.04339>.
- [19] U. Kazemi and S. Soleimani, “Combination of Density-Based Spatial Clustering with Grid Search Using Nash Equilibrium,” *Engineering Reports*, vol. 7, no. 3, pp. 1–15, 2025, doi: 10.1002/eng2.70037.
- [20] L. Li and K. Yu, “Application of Improved HF-DBSCAN Algorithm in Analyzing Complex Trajectory Data of Wi-Fi Users in Smart Campus,” *Applied Artificial Intelligence*, vol. 38, no. 1, 2024, doi: 10.1080/08839514.2024.2403260.