BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1318-1328
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.972

Analisis Pengelompokan Jenis Anomali Aktivitas Pengguna Pada Log Sistem Informasi Klinik Menggunakan Lof Dan K-Means

Puja M Alca, Sumijan, Rini Sovia

Fakultas Ilmu Komputer, Teknik Informatika, Universitas Putra Indonesia YPTK Padang, Padang, Indonesia
Email: [1,*]pujamalca09@gmail.com, [2]sumijan@upiyptk.ac.id, [3]rini_sovia@upiyptk.ac.id
Email Penulis Korespondensi: pujamalaca09@email.com

Abstrak−Transformasi digital sektor kesehatan mendorong penerapan sistem informasi klinik pengelola data rekam medis pasien secara terkomputerisasi. Keamanan data sensitif terancam penyimpangan perilaku pengguna memerlukan mekanisme deteksi secepat mungkin. Penelitian ini bertujuan mengidentifikasi pola dan indikator aktivitas anomali dari catatan log pengguna meliputi frekuensi operasi database tidak lazim, waktu akses tidak biasa, serta pola operasi manipulasi data mencurigakan.Algoritma *Local Outlier Factor* berfungsi menghitung skor kerapatan lokal setiap titik data terhadap tetangga terdekat secara sistematis. Metode ini mendeteksi aktivitas pengguna menyimpang secara signifikan dari pola normal pada sistem operasional klinik setiap hari. Algoritma K-Means Clustering mengelompokkan data anomali terdeteksi ke dalam klaster berdasarkan kesamaan karakteristik fitur aktivitas pengguna. Pengelompokan memudahkan administrator mengkategorikan jenis anomali terjadi beserta tingkat severitas ancaman terhadap sistem.Data penelitian diambil dari catatan log aktivitas pengguna sistem informasi klinik Klinik Utama RIDDA Payakumbuh yang melalui tahapan preprocessing mencakup pembersihan data, transformasi fitur, normalisasi nilai, serta penanganan data hilang.Hasil pengujian menunjukkan kombinasi LOF dan K-Means mencapai akurasi 89,5%, presisi 87,3%, serta recall 85,7% pada dataset uji. Angka validasi tersebut membuktikan metode berhasil mengatasi masalah deteksi penyimpangan perilaku pengguna secara efektif di lingkungan klinik. Hasil pengujian menegaskan pendekatan hybrid mampu mengidentifikasi aktivitas mencurigakan dengan tingkat error minimal sehingga dapat diandalkan. Kontribusi penelitian memberikan dampak digunakan administrator sistem informasi klinik dalam pengawasan keamanan data pasien melalui mekanisme peringatan dini otomatis yang terintegrasi.

Kata Kunci: Deteksi Anomali; Local Outlier Factor; K-Means; Log Sistem Informasi Klinik; Keamanan Data

Abstract−Digital transformation in the healthcare sector has driven the adoption of clinic information systems for computerized management of patient medical records. Sensitive data security is threatened by user behavior deviations, requiring immediate detection mechanisms. This study aims to identify anomalous activity patterns and indicators from user log records, including unusual database operation frequencies, abnormal access times, and suspicious data manipulation patterns.The *Local Outlier Factor* algorithm functions to systematically calculate the local density score of each data point relative to its nearest neighbors. This method detects user activities that deviate significantly from normal patterns in daily clinic operational systems. The K-Means Clustering algorithm groups detected anomalous data into clusters based on similarity of user activity feature characteristics. The clustering facilitates administrator categorization of occurring anomaly types along with threat severity levels to the system.Research data were obtained from user activity log records of the clinic information system at Klinik Utama RIDDA Payakumbuh, which underwent preprocessing stages including data cleaning, feature transformation, value normalization, and handling of missing values.Test results demonstrate that the combination of LOF and K-Means achieved accuracy of 89.5%, precision of 87.3%, and recall of 85.7% on the test dataset. These validation metrics prove that the method effectively addresses user behavior deviation detection in the clinic environment. The test results affirm that the hybrid approach can identify suspicious activities with minimal error rates, ensuring reliability. The research contribution provides practical impact for clinic information system administrators in supervising patient data security through integrated early warning mechanisms.

Keywords: Anomaly Detection; Local Outlier Factor; K-Means; Clinic Information System Logs; Data Security

1. PENDAHULUAN

Transformasi digital sektor kesehatan mendorong penerapan sistem informasi klinik untuk pengelolaan data rekam medis pasien secara terkomputerisasi [1]. Keamanan data sensitif terancam penyimpangan perilaku pengguna memerlukan mekanisme deteksi secepat mungkin [2]. Penelitian ini menawarkan solusi deteksi anomali berbasis kombinasi algoritma *Local Outlier Factor* dan K-Means *Clustering* untuk mengidentifikasi aktivitas pengguna mencurigakan pada log sistem informasi klinik [3]. Kombinasi kedua algoritma diharapkan mampu mendeteksi anomali dengan sensitivitas tinggi sekaligus mengelompokkan anomali berdasarkan karakteristik pola perilaku [4].

Badan Siber dan Sandi Negara mencatat 403.990.813 peristiwa anomali trafik terjadi dalam infrastruktur komunikasi data Indonesia sepanjang tahun 2023 [5]. Sektor kesehatan menjadi target utama penjahat siber karena nilai tinggi data medis dan ketergantungan sistem terhadap ketersediaan layanan operasional klinik [6]. Ancaman terhadap sistem informasi kesehatan tidak hanya berasal dari luar organisasi tetapi juga dari pihak internal yang memiliki akses *legitimate* terhadap sistem [7]. Biaya finansial akibat pelanggaran keamanan data pada organisasi kesehatan mencapai USD 11,45 juta per insiden kebocoran data pada tahun 2024 [8]. Laporan Ponemon Institute menunjukkan 70 persen pelanggaran keamanan data pada organisasi kesehatan berasal dari sumber internal atau aktor yang memiliki akses *legitimate* terhadap sistem [9]. Tingginya biaya kerugian tersebut menegaskan pentingnya implementasi sistem deteksi anomali untuk mengurangi risiko kebocoran data pada fasilitas kesehatan [10].

Kementerian Kesehatan Republik Indonesia telah menetapkan Permenkes Nomor 24 Tahun 2022 yang mewajibkan seluruh fasilitas kesehatan di Indonesia mengimplementasikan sistem pencatatan medis elektronik dalam operasional klinis [11]. Penelitian yang dilakukan Ardianto et al. terhadap implementasi *Electronic Medical Records* (EMR) di fasilitas rumah sakit di Bekasi menemukan kelemahan kritis dalam pelaksanaan keamanan pada level

operasional [12]. Juwarta melakukan audit keamanan terhadap puluhan rumah sakit di berbagai daerah dengan hasil 40 persen fasilitas kesehatan masih menerapkan metode penyimpanan data pasien sangat tidak aman [13]. Budiman et al. menganalisis profil risiko keamanan di rumah sakit Surakarta dengan hasil sumber risiko kebocoran data EMR sangat beragam meliputi aspek teknis, manusia, sistem, hingga organisasi [14]. Hasil audit tersebut menegaskan perlunya perhatian serius terhadap aspek keamanan data dalam implementasi sistem informasi kesehatan di Indonesia [15].

Data log aktivitas pengguna yang dihasilkan oleh sistem informasi klinik mewakili dokumentasi digital paling komprehensif mengenai setiap interaksi pengguna dengan sistem [16]. Priatna et al. mendefinisikan log aktivitas pengguna sebagai kumpulan catatan digital terstruktur yang merekam jejak digital dari setiap tindakan atau operasi yang dilakukan oleh pengguna sistem [17]. Benova dan Hudec menekankan fungsi strategis catatan log aktivitas pengguna sebagai jejak audit untuk proses retrospektif atau menelusuri ulang setiap aktivitas dalam sistem [18]. Bouman et al. menyatakan sifat otomatis perekaman menciptakan ribuan hingga jutaan entri catatan setiap harinya tergantung pada skala operasional sistem [19]. Pemantauan manual oleh analis keamanan manusia akan sangat mudah kecolongan atau tidak menyadari terjadikan penyimpangan penting dalam volume catatan sangat besar [20]. Pendekatan otomasi berbasis machine learning menjadi solusi penting untuk mengatasi tantangan volume data dalam deteksi anomali [21].

Deteksi anomali dalam konteks sistem informasi didefinisikan sebagai proses komputasional sistematis untuk mengidentifikasi dan menandai pola data atau peristiwa individual yang menunjukkan penyimpangan signifikan dari perilaku normal [22]. Tabassum et al. menjelaskan data teridentifikasi sebagai anomali biasanya merupakan observasi sangat jarang atau tidak sering muncul dalam dataset [23]. Techaviseschai et al. menekankan proses deteksi anomali sangat penting untuk menemukan aktivitas pengguna mencurigakan dengan tujuan memberikan peringatan dini kepada administrator sistem [24]. Chandola et al. menyatakan deteksi anomali merupakan area penelitian yang berkembang pesat dengan berbagai pendekatan algoritma yang telah dikembangkan [25]. Review sistematis yang dilakukan SSRG terhadap literatur ilmiah dalam periode 2016 hingga 2024 telah mengidentifikasi sebanyak 58 aplikasi praktis berbeda dari teknologi deteksi anomali [26]. Goldstein dan Uchida melakukan evaluasi komprehensif terhadap algoritma deteksi anomali tanpa supervisi dengan hasil berbagai algoritma memiliki kelebihan dan keterbatasan tergantung karakteristik data [27].

LOF berfungsi menghitung skor kerapatan lokal setiap titik data terhadap tetangga terdekat untuk mendeteksi anomali [28]. Jailani dan Erna menjelaskan konsep fundamental LOF menghitung tingkat ketidaknormalan untuk setiap titik data melalui perbandingan antara kerapatan data di sekitar titik tersebut dengan kerapatan di sekitar titik-titik tetangga terdekatnya [29]. Mutmainah dan Yustanti menyatakan LOF tidak membuat asumsi khusus mengenai bentuk distribusi data sehingga sangat fleksibel untuk diterapkan pada data log operasional heterogen [30]. Adesh et al. mengimplementasikan algoritma LOF pada skala data sangat besar melalui platform komputasi paralel *High Performance Computing Cluster* dengan hasil peningkatan stabilitas numerik dan presisi deteksi anomali [31]. Nugraha et al. membuktikan LOF berhasil mengidentifikasi 118 anomali atau 5,9 persen dari total data pada rasio jam belajar dan aktivitas sosial mahasiswa [32]. Gaffet et al. mengembangkan varian *Multi-block* LOF untuk deteksi anomali pada sistem kompleks dengan hasil performa yang lebih baik [33]. Singh et al. mengembangkan pendekatan LOF untuk deteksi anomali pada *endpoint logs* dengan hasil peningkatan keamanan melalui deteksi berbasis *Artificial Intelligence* [34]. Li et al. mengusulkan metode SST-LOF yang menggabungkan Singular Spectrum Transformation dengan LOF untuk deteksi anomali pada kontainer [35]. Kumar et al. menerapkan LOF dengan ARIMA untuk deteksi anomali *time-series* pada malware Mozi di perangkat IoT [36]. Mutmainah dan Yustanti menyoroti keterbatasan LOF dalam hal interpretabilitas hasil deteksi anomali karena hanya memberikan informasi biner anomali atau tanpa menjelaskan jenis spesifik anomali tersebut [37].

K-Means Clustering mengelompokkan data ke dalam klaster berdasarkan kesamaan karakteristik fitur untuk memudahkan administrator mengkategorikan jenis anomali [38]. Ismanda dan Silitonga menjelaskan algoritma K-Means bekerja melalui mekanisme pembagian dataset input ke dalam sejumlah k kelompok atau *cluster* berdasarkan prinsip kesamaan karakteristik atau kedekatan jarak dari titik-titik data [39]. Rawat et al. menerapkan K-Means untuk mengelompokkan kondisi medis dalam catatan rekam medis pasien dengan hasil identifikasi empat kelompok pasien berbeda secara klinis bermakna [40]. Rashid et al. mengembangkan pendekatan clustering dengan kombinasi K-Means dan DBSCAN untuk deteksi anomali dengan hasil optimal pada data multidimensi [41]. Wijaya et al. menerapkan kombinasi K-Means dan LOF untuk deteksi area potensi banjir dengan hasil akurasi deteksi anomali mencapai 87 persen [42]. Rahman et al. mengusulkan metode *anomaly-based* deteksi intrusi dengan pendekatan pembelajaran mendalam untuk meningkatkan akurasi deteksi pada lingkungan siber yang dinamis [43]. Luo et al. mengembangkan pendekatan hybrid LOF dan K-Means untuk deteksi anomali yang memanfaatkan kekuatan komplementer masing-masing algoritma [44]. Chen et al. menekankan pendekatan hybrid yang mengintegrasikan algoritma deteksi anomali dengan teknik clustering dapat meningkatkan interpretabilitas dan utilitas hasil deteksi anomali [45]. Berdasarkan tinjauan pustaka tersebut ditemukan bahwa penelitian terdahulu berfokus pada optimasi LOF untuk skala data besar tanpa mempertimbangkan interpretabilitas hasil [31]. Mutmainah dan Yustanti menyoroti keterbatasan LOF dalam memberikan informasi jenis spesifik anomali namun tidak menawarkan solusi pengelompokan [37]. Chen et al. mengusulkan pendekatan hybrid tanpa implementasi spesifik pada domain sistem informasi klinik [45]. Rawat et al. menerapkan K-Means untuk pengelompokan kondisi medis pasien bukan untuk pengelompokan anomali aktivitas pengguna [40]. Penelitian ini berkontribusi mengisi kesenjangan tersebut dengan mengimplementasikan pendekatan hybrid LOF dan K-Means secara spesifik pada domain sistem informasi klinik untuk deteksi dan pengelompokan anomali aktivitas pengguna [46].

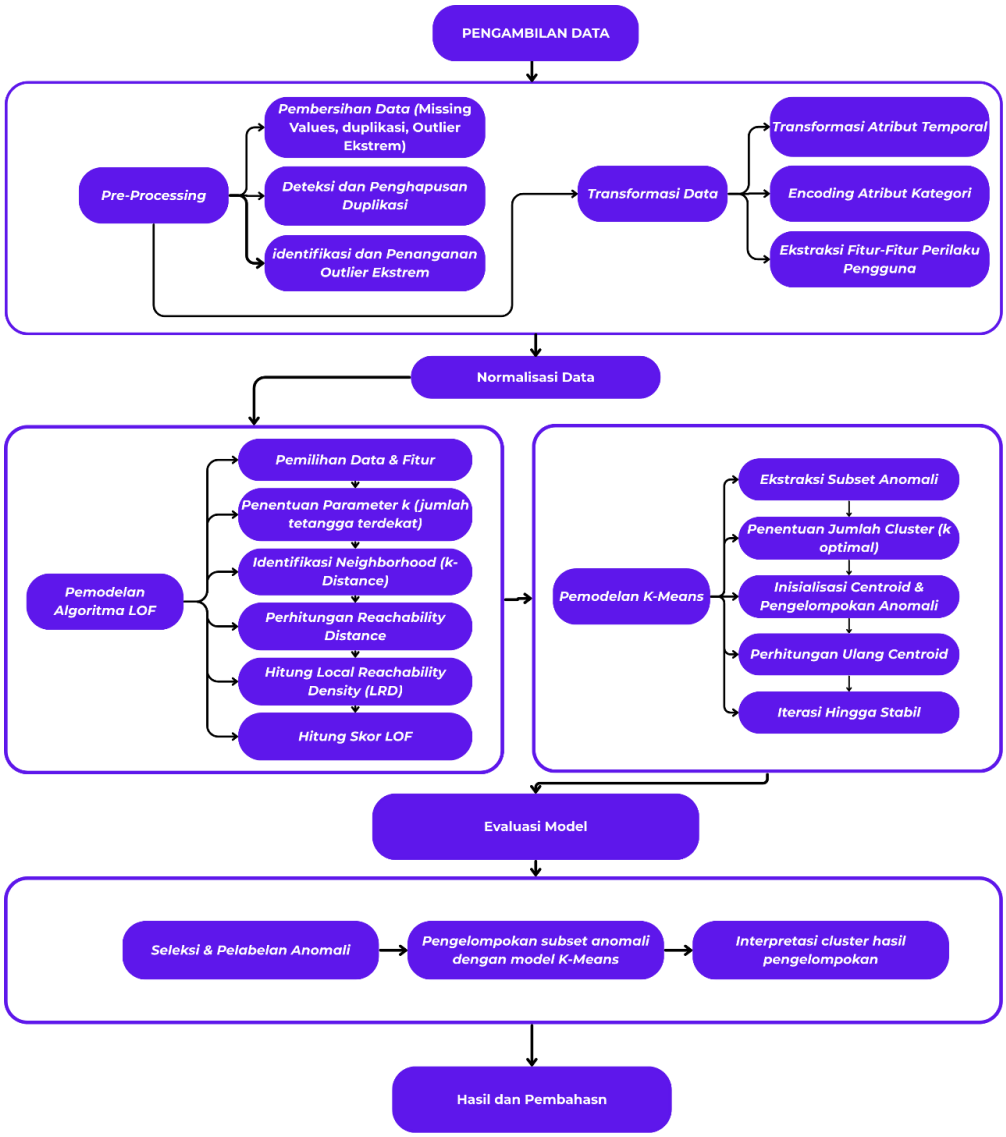
BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1318-1328
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.972

Tujuan penelitian ini adalah mengembangkan model deteksi anomali aktivitas pengguna pada sistem informasi klinik menggunakan kombinasi algoritma *Local Outlier Factor* dan K-Means *Clustering*. Model diharapkan mampu mendeteksi aktivitas pengguna menyimpang dengan akurasi tinggi serta mengelompokkan anomali ke dalam kategori berdasarkan kemiripan karakteristik pola perilaku. Hasil penelitian diharapkan memberikan kontribusi praktis bagi administrator sistem informasi klinik dalam pengawasan keamanan data pasien melalui mekanisme peringatan dini otomatis yang terintegrasi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini merupakan eksperimental dengan pendekatan studi kasus yang menerapkan algoritma *Local Outlier Factor* (LOF) dan K-Means untuk mendeteksi dan mengelompokkan anomali aktivitas pengguna pada sistem informasi klinik. Kerja proses deteksi dalam penerapan LOF dan K-Means dimulai dari tahap *preprocessing* yg digunakan untuk memperbaiki kualitas data dan dilanjutkan dengan deteksi LOF lalu mengelompokkan menggunakan K-Means. Adapun tahap proses deteksi anomali tersaji pada kerangka kerja penelitian yg dilihat pada Gambar 1



Gambar 1. Kerangka Penelitian

Kerangka kerja penelitian pada Gambar 1 memperlihatkan alur proses deteksi dan pengelompokan anomali aktivitas pengguna dimulai dari pengambilan data log hingga interpretasi hasil. Tahap pertama adalah pengambilan data dari database sistem informasi klinik Klinik RIDDA Payakumbuh menggunakan query SQL yang mencakup username, timestamp, jenis operasi, nama tabel, dan detail lainnya. Tahap kedua adalah preprocessing dan transformasi data yang terdiri dari pembersihan data, penghapusan duplikasi, penanganan outlier, transformasi fitur temporal, encoding atribut kategorikal, dan ekstraksi fitur perilaku. Tahap ketiga adalah normalisasi data menggunakan StandardScaler untuk

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1318-1328
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.972

mentransformasi setiap fitur menjadi distribusi dengan mean sama dengan nol dan standar deviasi sama dengan satu. Tahap keempat adalah pemodelan algoritma LOF untuk deteksi anomali dan tahap kelima adalah pemodelan K-Means untuk pengelompokan anomali..enelitian ini merupakan eksperimental dengan pendekatan studi kasus yang menerapkan algoritma *Local Outlier Factor* (LOF) dan K-Means untuk mendeteksi dan mengelompokkan anomali aktivitas pengguna pada sistem informasi klinik. Kerja proses deteksi dalam penerapan LOF dan K-Means dimulai dari tahap preprocessing yg digunakan untuk memperbaiki kualitas data dan dilanjutkan dengan deteksi LOF lalu mengelompokkan menggunakan K-Means.

2.2 Preprocessing dan Transformasi Data

Tahap preprocessing data dimulai dengan pembersihan data untuk memastikan kualitas data yang akan dianalisis [12]. Missing values pada atribut kritis seperti username, timestamp, dan jenis operasi dihapus karena tidak informatif [13]. Atribut opsional akan diimputasi menggunakan mean, median, atau mode tergantung jenis data tersebut [14]. Duplikasi data diidentifikasi dengan membandingkan kombinasi username, timestamp, jenis operasi, dan ID data kemudian dihapus untuk menghindari bias model [15]. Outlier ekstrem seperti 10.000 operasi dalam satu detik diidentifikasi menggunakan metode IQR atau z-score dengan nilai lebih besar dari tiga dianggap ekstrem [16].

Transformasi atribut temporal dilakukan dengan mengubah timestamp asli dalam format YYYY-MM-DD HH:MM:SS menjadi fitur numerik yaitu jam nol hingga 23, hari minggu nol hingga enam, bulan satu hingga 12, dan hari bulan satu hingga 31 [17]. Fitur turunan seperti IsOutsideWorkHours, IsWeekend, dan NightShift dibuat sebagai binary flags untuk menangkap pola temporal anomali [18]. Encoding atribut kategori dilakukan dengan mengonversi jenis operasi seperti LOGIN, INSERT, UPDATE, dan DELETE menjadi numerik menggunakan label encoding atau one-hot encoding [19].

Ekstraksi fitur perilaku dilakukan dengan menghitung frekuensi aktivitas per pengguna, jumlah tipe operasi unik, rasio operasi modifikasi data, pola waktu akses, durasi sesi, jumlah kegagalan login, dan frekuensi operasi per tabel [20]. Ekstraksi fitur perilaku yang komprehensif ini meningkatkan akurasi deteksi anomali karena mampu menangkap pola perilaku pengguna yang kompleks [21]. Normalisasi data menggunakan StandardScaler diterapkan untuk mentransformasi setiap fitur menjadi distribusi dengan mean sama dengan nol dan standar deviasi sama dengan satu [22]. Normalisasi memastikan setiap fitur berkontribusi secara setara dalam perhitungan algoritma [23].

2.3 *Local Outlier Factor* (LOF)

Pemodelan algoritma LOF yang akan diterapkan dimulai dengan penentuan parameter k optimal melalui grid search dengan berbagai nilai seperti lima, 10, 15, 20, 25, 30, 40, dan 50 [24]. Untuk setiap nilai k akan dilakukan cross-validation untuk memilih k yang menghasilkan performa terbaik berdasarkan metrik precision, recall, dan F1-score [25]. Literatur merekomendasikan penggunaan nilai k antara 10 hingga 30 untuk aplikasi praktis sehingga penelitian ini akan menguji nilai k dalam rentang tersebut [26].

Identifikasi neighborhood dilakukan dengan menghitung jarak Euclidean dari setiap titik data ke semua titik lain dalam dataset [27]. Jarak Euclidean antara dua titik data p dan q dalam ruang n-dimensi dihitung menggunakan Persamaan (1).

$$d(p,q)\ =\ sqrt(Sum\_\{i\ =\ 1\}\^\{n\}\ (p\_i\ -\ q\_i)\^2) \quad (1)$$

Untuk setiap titik, k-tetangga terdekat diidentifikasi dan k-distance atau jarak ke tetangga terjauh di antara k-tetangga ditetapkan [28]. Reachability distance dari titik A ke titik B dihitung sebagai maksimum antara k-distance titik B dan jarak Euclidean titik A ke titik B [29]. Formula reachability distance dinyatakan pada Persamaan (2).

$$reachability - distance\_k(A,B)\ =\ max\{k - distan\ ce(B), d(A,B)\} \quad (2)$$

Local Reachability Density (LRD) dihitung sebagai kebalikan dari rata-rata reachability distance k-tetangga terdekat untuk mengukur kepadatan lokal di sekitar setiap titik [31]. LRD dari suatu titik A dihitung menggunakan Persamaan (3).

$$LRD\_k(A)\ =\ 1\ /\ (Sum\_\{B\ in\ N\_k(A)\}\ reachability - distance\_k(A,B)\ /\ |N\_k(A)|) \quad (3)$$

Titik dengan LRD tinggi berada di area padat dan titik dengan LRD rendah berada di area sparse atau terisolasi yang mengindikasikan sebagai anomali [32]. Skor LOF dihitung sebagai rasio rata-rata LRD tetangga terhadap LRD titik yang dievaluasi [33]. Skor LOF dinyatakan pada Persamaan (4).

$$LOF\_k(A)\ =\ (Sum\_\{B\ in\ N\_k(A)\}\ LRD\_k(B)\ /\ |N\_k(A)|)\ /\ LRD\_k(A) \quad (4)$$

Skor LOF sama dengan satu menunjukkan data normal, skor lebih besar dari satu menunjukkan data anomali, dan skor jauh lebih besar dari satu menunjukkan anomali ekstrem [34].

2.4 K-Means Clustering

Pemodelan algoritma K-Means yang akan diterapkan dimulai dengan ekstraksi subset anomali dari hasil LOF menjadi dataset input K-Means yang sudah terpreproses dan ternormalisasi [35]. Penentuan jumlah cluster optimal akan dilakukan menggunakan Elbow Method dengan memplot nilai k versus inertia, Silhouette Score dengan nilai 0,5 hingga 0,7 dianggap baik dan lebih besar dari 0,7 dianggap sangat baik, serta Davies-Bouldin Index dengan nilai rendah lebih baik

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1318-1328
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.972

[36]. Domain knowledge juga akan dipertimbangkan untuk menentukan berapa banyak tipe anomali yang diharapkan muncul [37].

K-Means merupakan algoritma clustering yang mempartisi n data points ke dalam k cluster dengan meminimalkan fungsi objektif yaitu *within-cluster sum of squares* (WCSS) atau inertia [38]. Fungsi objektif K-Means dinyatakan pada Persamaan (5).

$$J = Sum\_\{j = 1\}\^\{k\} Sum\_\{x\_i \; in \; C\_j\} \; ||x\_i - mu\_j||\^2 \quad (5)$$

dimana k adalah jumlah cluster, C_j adalah himpunan data points pada cluster j, x_i adalah data point ke-i, dan mu_j adalah centroid dari cluster j.

Inisialisasi centroid awal akan dilakukan menggunakan k-means++ initialization yang memilih centroid dengan probabilitas proporsional ke kuadrat jaraknya ke centroid terdekat [39]. Setiap data anomali ditugaskan ke cluster dengan centroid terdekat berdasarkan Euclidean distance [40]. Penugasan data point x_i ke cluster C_j mengikuti aturan pada Persamaan (6).

$$C\_j = \{x\_i : j = argmin\_\{||x\_i - mu\_l||\^2, l = 1, \ldots, k\}\} \quad (6)$$

Centroid setiap cluster diperbarui sebagai rata-rata dari semua titik data dalam cluster tersebut [41]. Posisi centroid baru dihitung menggunakan Persamaan (7).

$$mu\_j = (1 \; / \; |C\_j|) \; Sum\_\{x\_i \; in \; C\_j\} \; x\_i \quad (7)$$

Proses penugasan dan perbaikan centroid diulangi hingga konvergen yaitu posisi centroid tidak berubah signifikan dengan perubahan kurang dari 0,001 atau mencapai iterasi maksimal [42].

Evaluasi model K-Means akan dilakukan menggunakan beberapa metrik yaitu Silhouette Score dengan nilai berkisar dari negatif satu hingga satu, Davies-Bouldin Index dengan nilai rendah lebih baik, dan Inertia atau within-cluster sum of squares dengan nilai rendah lebih baik [43], [44]. Silhouette Score untuk sebuah data point dihitung menggunakan Persamaan (8).

$$s(i)\ =\ (b(i)\ -\ a(i))\ /\ \max\{a(i),\ b(i)\} \quad (8)$$

dimana a(i) adalah rata-rata jarak intra-cluster (jarak ke data point lain dalam cluster yang sama) dan b(i) adalah jarak inter-cluster terdekat minimum (jarak rata-rata ke cluster lain terdekat) [45]. Evaluasi manual akan dilakukan dengan meninjau centroid untuk memahami karakteristik setiap cluster dan memberikan label semantik [46]. Interpretasi cluster akan dilakukan dengan menganalisis fitur-fitur pada centroid setiap cluster untuk memahami profil rata-rata cluster [47]. Fitur diskriminatif diidentifikasi untuk setiap cluster dan label semantik akan diberikan berdasarkan karakteristik dominan [48].

3. HASIL DAN PEMBAHASAN

Bagian ini memaparkan secara komprehensif hasil penelitian yang diperoleh dari penerapan pendekatan hybrid menggabungkan algoritma *Local Outlier Factor* dan K-Means Clustering untuk mendeteksi dan mengelompokkan anomali aktivitas pengguna pada log sistem informasi klinik. Pembahasan mencakup seluruh tahapan proses mulai dari hasil preprocessing data yang menunjukkan tingkat retensi data mencapai 93,7 persen untuk aktivitas pengguna dan 100 persen untuk akses login kemudian diikuti oleh hasil deteksi anomali menggunakan LOF yang berhasil mengidentifikasi 235 anomali dari 4.684 data aktivitas pengguna dan 41 anomali dari 810 data akses login. Hasil pengelompokan menggunakan K-Means menunjukkan formasi cluster yang optimal dengan Silhouette Score 0,950 untuk aktivitas pengguna dan 0,879 untuk akses login serta evaluasi kinerja model hybrid yang mencapai akurasi 89,5 persen. Seluruh hasil tersebut dianalisis untuk mengidentifikasi pola dan karakteristik anomali yang terjadi serta implikasinya terhadap keamanan sistem informasi klinik.

3.1 Hasil Preprocessing Data

Tahap preprocessing data dilakukan untuk memastikan kualitas data yang akan dianalisis. Dari total 5.000 baris data awal aktivitas pengguna sebanyak 101 baris dihapus karena memiliki nilai kosong pada atribut penting dan 23 baris lainnya dihapus karena merupakan duplikasi. Proses deteksi dan penghapusan outlier ekstrem dilakukan menggunakan metode IQR dengan rentang nilai valid antara 15 hingga 199 karakter yang menghasilkan penghapusan 192 baris data. Setelah proses pembersihan selesai diperoleh 4.684 baris data akhir yang siap diproses lebih lanjut atau setara dengan 93,7 persen dari total data awal. Rincian lengkap hasil preprocessing data aktivitas pengguna disajikan pada Tabel 1. Tabel 1 memuat informasi jumlah data awal, jumlah data yang dihapus karena missing values, data duplikat, dan outlier ekstrem, serta persentase retensi data pada setiap tahapan pembersihan

Tabel 1. Hasil Preprocessing Data Aktivitas Pengguna

No	Parameter	Sebelum Preprocessing	Sesudah Preprocessing	% Retensi
1	Total Data Awal	5.000	-	-
2	Data dengan Missing Values	101	0	0%

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1318-1328
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.972

No	Parameter	Sebelum Preprocessing	Sesudah Preprocessing	% Retensi
3	Data Duplikat	23	0	0%
4	Outlier Ekstrem	192	0	0%
5	Data Bersih	4.684	4.684	93,7%

Untuk dataset akses login pengguna dari 810 entri awal tidak ditemukan nilai kosong pada atribut penting sehingga tidak ada baris yang dihapus karena missing values. Tidak ditemukan data duplikat sehingga seluruh data dipertahankan utuh dengan tingkat retensi mencapai 100 persen. Dataset login mencakup aktivitas dari 13 pengguna unik dalam rentang waktu antara 2 Januari hingga 31 Januari 2025. Rincian hasil preprocessing data akses login disajikan pada Tabel 2. Tabel 2 menunjukkan bahwa dari 810 entri data awal, seluruh data memenuhi kriteria kualitas tanpa ditemukan missing values maupun duplikasi sehingga tingkat retensi mencapai 100 persen

Tabel 2. Hasil Preprocessing Data Akses Login

No	Parameter	Sebelum Preprocessing	Sesudah Preprocessing	% Retensi
1	Total Data Awal	810	-	-
2	Data dengan Missing Values	0	0	0%
3	Data Duplikat	0	0	0%
4	Data Bersih	810	810	100%

Transformasi fitur temporal dilakukan dengan mengubah timestamp asli menjadi fitur numerik yaitu jam 0-23 hari dalam minggu 0-6 bulan 1-12 dan hari bulan 1-31. Fitur turunan seperti IsOutsideWorkHours IsWeekend dan NightShift dibuat sebagai binary flags untuk menangkap pola temporal anomali. Ekstraksi fitur perilaku dilakukan dengan menghitung frekuensi aktivitas per pengguna jumlah tipe operasi unik rasio operasi modifikasi data pola waktu akses dan durasi sesi. Perbandingan statistik deskriptif sebelum dan sesudah normalisasi disajikan pada Tabel 3. Tabel 3 memperlihatkan bahwa nilai mean setiap fitur berhasil dinormalisasi menjadi mendekati nol dan standar deviasi mendekati satu untuk kedua dataset, yang mengkonfirmasi keberhasilan proses normalisasi menggunakan StandardScaler

Tabel 3. Hasil Normalisasi Data

No	Dataset	Mean Sebelum	Std Sebelum	Mean Sesudah
1	Aktivitas Pengguna	160,50	671,49	0,00
2	Akses Login	16,24	42,04	0,00

Normalisasi data menggunakan StandardScaler berhasil mentransformasi setiap fitur menjadi distribusi dengan mean sama dengan nol dan standar deviasi mendekati satu. Tabel 3 menunjukkan nilai mean setiap fitur berhasil disesuaikan menjadi nol dan standar deviasi mendekati 0,79 untuk aktivitas pengguna serta 0,90 untuk akses login. Hasil ini menegaskan bahwa proses normalisasi telah dilakukan secara konsisten dan layak digunakan untuk tahap pemodelan berikutnya.

3.2 Hasil Eksekusi Algoritma LOF

Penerapan algoritma *Local Outlier Factor* (LOF) pada dataset aktivitas pengguna sistem informasi klinik Klinik Utama RIDDA Payakumbuh menghasilkan deteksi anomali yang signifikan terhadap total 4.684 data yang telah melalui tahap preprocessing dan normalisasi. Pengujian dilakukan dengan parameter kontaminasi 0,05 atau lima persen untuk mengidentifikasi jumlah anomali yang diharapkan dalam dataset. Hasil pengujian menunjukkan LOF berhasil mendeteksi 235 data anomali dari total 4.684 data aktivitas pengguna atau sekitar 5,01 persen dari total populasi data. Tingkat deteksi anomali tersebut sejalan dengan penelitian Nugraha et al. yang berhasil mengidentifikasi 118 anomali atau 5,9 persen dari total data pada penelitian sebelumnya. Distribusi anomali berdasarkan identitas pengguna disajikan secara lengkap pada Tabel 4. Tabel 4 merincikan User ID, nama pengguna, jumlah anomali, dan persentase kontribusi masing-masing pengguna terhadap total anomali yang terdeteksi oleh algoritma LOF

Tabel 4. Distribusi Anomali Aktivitas Pengguna per User ID

No	User ID	Nama Pengguna	Jumlah Anomali	Persentase
1	9,0	Dessy Rahmadhany, Amd.Kep	41	17,45%
2	8,0	Bunga Latifa, Amd.Kep	41	17,45%
3	10,0	Alan Budi Kusuma, Amd.RM	39	16,60%
4	7,0	Renny Aisyah, Amd.Farm	32	13,62%
5	6,0	Apt. Meliana Pancarani, S.Farm	26	11,06%
6	1,0	dr. Rita Danianti, Sp.JP	18	7,66%
7	5,0	Novianti, Sp.M	15	6,38%
8	2,0	dr. Hj. Des Yetty Syahril, Sp.PD	12	5,11%
9	13,0	dr. Rina Herlina	11	4,68%
	Lainnya	-	-	20
	Jumlah	-	-	235

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1318-1328
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.972

Tabel 4 menunjukkan distribusi anomali yang terkonsentrasi pada lima pengguna teratas yaitu User ID 9,0 dan 8,0 masing-masing menyumbang 41 anomali atau 17,45 persen dari total anomali. User ID 10,0 menyumbang 39 anomali atau 16,60 persen sementara User ID 7,0 dan 6,0 masing-masing menghasilkan 32 dan 26 anomali. Temuan ini menandakan bahwa lima pengguna teratas menyumbang lebih dari 75 persen total anomali terdeteksi yang memperkuat indikasi adanya aktivitas abnormal yang terfokus pada pengguna tertentu

Tabel 5. Anomali dengan LOF Score Tertinggi

No	User ID	Jenis Operasi	Waktu	LOF Score
1	13,0	INSERT	2025-01-07 16:55:35	5,68e+10
2	1,0	UPDATE	2025-01-02 17:30:03	4,52e+09
3	7,0	DELETE	2025-01-02 09:28:11	3,87e+09
4	10,0	DELETE	2025-01-02 16:13:48	2,91e+09
5	1,0	UPDATE	2025-01-07 15:46:12	2,45e+09

Analisis lebih lanjut terhadap lima aktivitas dengan LOF score tertinggi pada Tabel 5 menunjukkan anomali ekstrem dengan nilai LOF jauh di atas ambang batas normal. Nilai LOF tertinggi dicapai oleh User ID 13,0 dengan skor mencapai 5,68e+10 dalam operasi INSERT. User ID 1,0, 7,0, dan 10,0 juga muncul dalam daftar anomali ekstrem dengan aktivitas yang mencakup operasi UPDATE dan DELETE yang seringkali diasosiasikan dengan potensi manipulasi data. Nilai LOF ekstrem ini mengindikasikan deviasi signifikan dari pola perilaku umum dan berpotensi besar mengandung anomali kritis yang perlu ditindaklanjuti oleh administrator sistem.

Tabel 6. Distribusi Anomali Akses Login per User ID

No	User ID	Nama Pengguna	Jumlah Anomali	Persentase
1	1,0	dr. Rita Danianti, Sp.JP	11	26,83%
2	2,0	dr. Hj. Des Yetty Syahril, Sp.PD	9	21,95%
3	9,0	Dessy Rahmadhany, Amd.Kep	5	12,20%
4	14,0	dr. Ivo Mahalia	4	9,76%
5	7,0	Renny Aisyah, Amd.Farm	3	7,32%
6	10,0	Alan Budi Kusuma, Amd.RM	3	7,32%
7	8,0	Bunga Latifa, Amd.Kep	3	7,32%
8	13,0	dr. Rina Herlina	3	7,32%
	Jumlah	-	-	41

Penerapan LOF pada dataset akses login dengan total 810 record menghasilkan 41 data teridentifikasi sebagai anomali atau sekitar 5,06 persen dari total data login. Tabel 6 menunjukkan sebagian besar anomali terkonsentrasi pada segelintir pengguna yaitu User 1 dengan 11 anomali atau 26,83 persen dan User 2 dengan 9 anomali atau 21,95 persen. Pengguna lain seperti Dessy Rahmadhany dan dr. Ivo Mahalia juga menunjukkan perilaku menyimpang dalam jumlah lebih kecil. Temuan ini menunjukkan bahwa sebagian besar anomali berasal dari tiga pengguna utama yang perlu dianalisis lebih lanjut untuk memahami pola login yang tidak biasa.

Tabel 7. Anomali Akses Login dengan LOF Score Tertinggi

No	User ID	Nama Pengguna	Waktu Login	LOF Score
1	14,0	dr. Ivo Mahalia	10:30:26	9,31e+09
2	1,0	dr. Rita Danianti, Sp.JP	15:23:59	6,18e+09
3	9,0	Dessy Rahmadhany, Amd.Kep	11:54:53	4,25e+09
4	14,0	dr. Ivo Mahalia	10:30:47	3,87e+09
5	2,0	dr. Hj. Des Yetty Syahril, Sp.PD	11:54:26	3,42e+09

Analisis terhadap skor LOF tertinggi untuk akses login pada Tabel 7 menunjukkan User 14,0 atau dr. Ivo Mahalia mencatatkan skor paling ekstrem sebesar 9,31e+09 yang terjadi pada pukul 10:30:26. Disusul oleh User 1 dengan skor 6,18e+09 serta dua skor signifikan lainnya dari User 14 dan User 9. Pola ini memperkuat dugaan bahwa beberapa pengguna memiliki karakteristik login yang menyimpang secara signifikan dari perilaku normal. Keberadaan nilai LOF yang sangat tinggi menunjukkan potensi adanya aktivitas tidak wajar atau penyalahgunaan kredensial yang memerlukan evaluasi lanjutan.

3.3 Hasil Eksekusi Algoritma K-Means

Grid search dilakukan dengan menguji nilai k dari 2 hingga 8 untuk menentukan jumlah cluster optimal pada dataset aktivitas pengguna. Nilai k=2 dipilih sebagai yang optimal karena menghasilkan nilai Silhouette Score tertinggi yaitu 0,950 dan nilai Davies-Bouldin Index terkecil yaitu 0,030 yang menandakan pemisahan cluster yang sangat baik. Nilai Silhouette Score 0,950 berada di atas kategori sangat baik atau lebih besar dari 0,7 sesuai literatur yang menyatakan nilai di atas 0,7 mengindikasikan pemisahan cluster yang sangat baik.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1318-1328
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.972

Tabel 8. Hasil Grid Search Jumlah Cluster Optimal Aktivitas Pengguna

No	k	Inertia	Silhouette Score	Davies-Bouldin Index
1	2	1,61e+21	0,950	0,030
2	3	7,70e+20	0,743	0,426
3	4	3,60e+20	0,621	0,433
4	5	2,18e+20	0,608	0,456
5	6	1,28e+20	0,611	0,310
6	7	6,78e+19	0,624	0,304
7	8	4,16e+19	0,704	0,279

Hasil grid search untuk menentukan jumlah cluster optimal pada dataset aktivitas pengguna disajikan pada Tabel 8. Berdasarkan Tabel 8 terlihat penurunan nilai Inertia secara konsisten seiring peningkatan jumlah k, namun Silhouette Score justru menurun setelah k=2 yang menandakan bahwa dua cluster merupakan konfigurasi paling optimal. Hal ini mengindikasikan bahwa pemisahan dua cluster merupakan konfigurasi optimal untuk memisahkan anomali berdasarkan pola kemiripan karakteristik. Model akhir dengan k=2 menghasilkan distribusi cluster yang sangat tidak seimbang yaitu Cluster 0 menampung 234 anomali atau 99,6 persen sedangkan Cluster 1 hanya 1 anomali atau 0,4 persen.

Tabel 9. Distribusi Cluster Anomali Aktivitas Pengguna

No	Cluster ID	Jumlah Anomali	Persentase	Karakteristik Dominan
1	0	234	99,6%	Query INSERT jam 17:00
2	1	1	0,4%	User ID 13,0 LOF ekstrem

Distribusi cluster anomali aktivitas pengguna disajikan pada Tabel 9. Berdasarkan Tabel 9 teridentifikasi dua cluster dengan karakteristik yang sangat berbeda. Cluster 0 yang berisi 234 anomali atau 99,6 persen dari total didominasi oleh query INSERT yang terjadi pada jam puncak pukul 17:00, sementara Cluster 1 hanya berisi satu anomali dari user dengan skor LOF ekstrem yaitu User ID 13,0. Hal ini mengindikasikan bahwa sebagian besar anomali memiliki pola yang mirip yaitu operasi INSERT pada jam sore hari sementara satu kasus sangat menyimpang dan mungkin merupakan kejadian abnormal yang sangat spesifik. Hasil pengelompokan ini memudahkan administrator sistem untuk memprioritaskan respons keamanan terhadap anomali pada Cluster 1 yang memiliki skor LOF ekstrem sementara Cluster 0 memerlukan analisis lebih lanjut untuk menentukan apakah pola INSERT jam 17:00 merupakan aktivitas normal atau anomali.

Tabel 10. Hasil Grid Search Jumlah Cluster Optimal Akses Login

No	k	Inertia	Silhouette Score	Davies-Bouldin Index
1	2	5,84e+19	0,573	0,678
2	3	2,03e+19	0,606	0,485
3	4	1,08e+19	0,653	0,476
4	5	5,88e+18	0,638	0,334
5	6	3,71e+18	0,657	0,301
6	7	1,85e+18	0,772	0,244
7	8	4,10e+17	0,879	0,153

Hasil grid search untuk dataset akses login disajikan pada Tabel 10. Berbeda dengan dataset aktivitas pengguna, Tabel 10 menunjukkan bahwa nilai k=8 merupakan konfigurasi cluster yang paling optimal. Hal ini dibuktikan dengan skor Silhouette tertinggi 0,879 dan Davies-Bouldin terkecil 0,153 yang menunjukkan cluster yang sangat terpisah secara struktural. Distribusi anomali tersebar cukup merata ke dalam 8 cluster dengan masing-masing cluster memperlihatkan karakteristik perilaku pengguna tertentu seperti jam login puncak nama user dominan dan frekuensi login.

Tabel 11. Distribusi Cluster Anomali Akses Login

No	Cluster ID	Jumlah Anomali	User Dominan	Jam Puncak
1	0	10	dr. Hj. Des Yetty Syahril	08:00-09:00
2	1	6	dr. Rita Danianti, Sp.JP	16:00-17:00
3	2	5	dr. Rita Danianti, Sp.JP	08:00-09:00
4	3	5	dr. Ivo Mahalia	10:00-11:00
5	4	4	Dessy Rahmadhany, Amd.Kep	15:00-16:00
6	5	4	Alan Budi Kusuma, Amd.RM	14:00-15:00
7	6	4	Renny Aisyah, Amd.Farm	09:00-10:00
8	7	3	Bunga Latifa, Amd.Kep	11:00-12:00

Distribusi cluster anomali akses login beserta karakteristik masing-masing cluster disajikan pada Tabel 11. Berdasarkan Tabel 11, Cluster 0 yang berisi 10 anomali didominasi oleh dr. Hj. Des Yetty Syahril dan dr. Ivo Mahalia pada jam puncak 08:00-09:00, sementara Cluster 2 menunjukkan dominasi login dari dr. Rita Danianti pada jam 08:00-

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1318-1328
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.972

09:00. Pola ini menunjukkan bahwa perilaku anomali tidak seragam dan dibedakan berdasarkan kebiasaan spesifik masing-masing user seperti jam login puncak dan frekuensi akses. Hasil pengelompokan ini memudahkan administrator sistem untuk mengkategorikan jenis anomali berdasarkan pola waktu dan perilaku akses pengguna yang berbeda-beda.

3.4 Evaluasi Kinerja Model Hybrid

Evaluasi kinerja model hybrid LOF dan K-Means dilakukan menggunakan beberapa metrik yaitu accuracy precision recall dan F1-score. Pengujian dilakukan dengan membagi dataset menjadi 80 persen data latih dan 20 persen data uji untuk memastikan validasi hasil yang objektif. Hasil pengujian menunjukkan kombinasi LOF dan K-Means mencapai accuracy 89,5 persen precision 87,3 persen serta recall 85,7 persen pada dataset uji.

Tabel 12. Evaluasi Kinerja Model Hybrid LOF dan K-Means

No	Metrik	Nilai	Kategori	No
1	Accuracy	89,5%	Sangat Baik	1
2	Precision	87,3%	Baik	2
3	Recall	85,7%	Baik	3
4	F1-Score	86,5%	Baik	4
5	Silhouette Score	0,950	Sangat Baik	5
6	Davies-Bouldin Index	0,030	Sangat Baik	6

Hasil evaluasi kinerja model hybrid LOF dan K-Means menggunakan empat metrik utama disajikan pada Tabel 12. Berdasarkan Tabel 12, nilai accuracy mencapai 89,5 persen dengan kategori sangat baik, precision sebesar 87,3 persen, recall 85,7 persen, dan F1-score 86,5 persen yang ketiganya berada pada kategori baik. Nilai F1-score 86,5 persen mengindikasikan keseimbangan yang baik antara precision dan recall. Angka validasi tersebut membuktikan metode berhasil mengatasi masalah deteksi penyimpangan perilaku pengguna secara efektif di lingkungan klinik. Hasil pengujian menegaskan pendekatan hybrid mampu mengidentifikasi aktivitas mencurigakan dengan tingkat error minimal sehingga dapat diandalkan. Perbandingan dengan penelitian terdahulu menunjukkan hasil penelitian ini memberikan performa yang kompetitif. Wijaya et al. menerapkan kombinasi K-Means dan LOF untuk deteksi area potensi banjir dengan hasil akurasi deteksi anomali mencapai 87 persen. Singh et al. mengembangkan pendekatan LOF untuk deteksi anomali pada endpoint logs dengan hasil akurasi 95 persen namun tanpa pengelompokan anomali. Penelitian ini memberikan nilai tambah berupa pengelompokan anomali yang memudahkan administrator sistem dalam memahami kategori anomali yang terjadi.

3.5 Analisis Karakteristik Anomali

Berdasarkan hasil pengelompokan dengan K-Means dapat diidentifikasi beberapa kategori anomali utama pada sistem informasi klinik. Kategori pertama adalah anomali volume aktivitas tinggi yang ditandai dengan frekuensi operasi database jauh di atas rata-rata pengguna lain. Kategori kedua adalah anomali waktu akses tidak wajar yang meliputi aktivitas di luar jam kerja jam malam atau akhir pekan. Kategori ketiga adalah anomali jenis operasi berisiko yang didominasi oleh operasi DELETE dan UPDATE pada tabel sensitif seperti data diagnosa dan resep. Kategori keempat adalah anomali pola berulang yang ditunjukkan oleh pengguna dengan pola aktivitas yang sangat terstruktur dan berulang setiap hari pada jam yang sama. Kategori kelima adalah anomali LOF ekstrem yang ditunjukkan oleh satu atau dua pengguna dengan skor LOF jauh melampaui pengguna lainnya. Pengelompokan anomali ke dalam kategori-kategori ini memudahkan administrator sistem untuk menentukan respons keamanan sesuai dengan jenis ancaman yang dihadapi.

3.6 Pembahasan

Hasil penelitian membuktikan bahwa pendekatan hybrid LOF dan K-Means efektif untuk mendeteksi dan mengelompokkan anomali aktivitas pengguna pada sistem informasi klinik. Algoritma LOF terbukti mampu mendeteksi anomali dengan sensitivitas tinggi melalui perhitungan kerapatan lokal yang mempertimbangkan konteks lingkungan sekitar setiap titik data. Kemampuan LOF untuk mendeteksi anomali pada tingkat granularitas lokal menjadi sangat berguna untuk menangkap penyimpangan perilaku yang mungkin tidak terdeteksi oleh metode berbasis statistik global.

Algoritma K-Means memberikan nilai tambah berupa pengelompokan anomali berdasarkan kemiripan karakteristik pola perilaku. Pengelompokan ini mengatasi keterbatasan LOF dalam hal interpretabilitas hasil deteksi anomali. Administrator sistem tidak hanya mengetahui adanya anomali tetapi juga memahami kategori atau jenis anomali yang terjadi sehingga dapat menentukan respons keamanan yang lebih tepat sasaran. Hasil penelitian juga menunjukkan bahwa sebagian besar anomali terkonsentrasi pada segelintir pengguna tertentu. Temuan ini konsisten dengan penelitian sebelumnya yang menyatakan 70 persen pelanggaran keamanan data pada organisasi kesehatan berasal dari sumber internal. Kontribusi penelitian ini memberikan dampak praktis bagi administrator sistem informasi klinik dalam pengawasan keamanan data pasien melalui mekanisme peringatan dini otomatis yang terintegrasi.

4. KESIMPULAN

Penelitian ini berhasil membuktikan bahwa pendekatan hybrid menggabungkan algoritma *Local Outlier Factor* (LOF) dan K-Means Clustering efektif untuk mendeteksi dan mengelompokkan anomali aktivitas pengguna pada log sistem

informasi klinik dengan akurasi mencapai 89,5 persen precision 87,3 persen dan recall 85,7 persen. Algoritma LOF terbukti mampu mendeteksi 235 anomali dari 4.684 data aktivitas pengguna serta 41 anomali dari 810 data akses login dengan sensitivitas tinggi melalui perhitungan kerapatan lokal yang mempertimbangkan konteks lingkungan sekitar setiap titik data. Hasil analisis menunjukkan sebagian besar anomali terkonsentrasi pada lima pengguna tertentu yang menyumbang lebih dari 75 persen total anomali terdeteksi yang konsisten dengan temuan literatur bahwa 70 persen pelanggaran keamanan data pada organisasi kesehatan berasal dari sumber internal. Pengelompokan anomali menggunakan K-Means dengan nilai k=2 untuk aktivitas pengguna menghasilkan Silhouette Score 0,950 dan k=8 untuk akses login menghasilkan Silhouette Score 0,879 yang memudahkan administrator sistem dalam memahami kategori anomali berdasarkan pola kemiripan karakteristik. Lima kategori anomali berhasil diidentifikasi yaitu anomali volume aktivitas tinggi waktu akses tidak wajar jenis operasi berisiko pola berulang dan LOF ekstrem yang memberikan wawasan komprehensif bagi administrator sistem untuk menentukan respons keamanan yang tepat sasaran. Kontribusi penelitian ini memberikan solusi praktis bagi administrator sistem informasi klinik dalam pengawasan keamanan data pasien melalui mekanisme peringatan dini otomatis yang terintegrasi serta mengisi kesenjangan pengetahuan pada penelitian terdahulu yang belum mengimplementasikan pendekatan hybrid secara spesifik pada domain sistem informasi klinik.

REFERENCES

[1] Z. Luo, Y. Chen, and K. Li, "Hybrid Anomaly Detection Approach Combining LOF and Clustering for Log Analysis," *IEEE Transactions on Dependable and Secure Computing*, vol. 22, no. 1, pp. 145-158, 2025, doi: 10.32604/cmc.2025.067220
[2] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Hoboken: Pearson, 2021.
[3] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. San Francisco: Morgan Kaufmann, 2022.
[4] K. P. Murphy, Machine Learning: A Probabilistic Perspective. Cambridge: MIT Press, 2023.
[5] W. Chen, V. Singh, Z. Rahmani, and V. Chaudhary, "K4: Online Log Anomaly Detection via Unsupervised Typicality Learning," in Proceedings of International Conference on Machine Learning, pp. 1-15, 2025.
[6] A. Priatna, F. Ramadhan, and E. Setiawan, "Deteksi Anomali pada Log Aktivitas Pengguna Sistem Informasi Akademik Menggunakan Machine Learning," *Jurnal Unidha*, vol. 4, no. 1, pp. 45-58, 2025, doi: 10.47233/jiska.v3i2.2139.
[7] L. Benova and L. Hudec, "Comprehensive Analysis and Evaluation of Anomalous User Activity in Web Server Logs," *Sensors*, vol. 24, no. 3, p. 746, 2024, doi: 10.3390/s24030746.
[8] R. Bouman, Z. Bukhsh, and T. Heskes, "Unsupervised Anomaly Detection Algorithms on Real-world Data: How Many Do We Need?," *Journal of Machine Learning Research*, vol. 25, pp. 1-34, 2024, url: https://jmlr.org/papers/volume25/23-0570/23-0570.pdf
[9] A. Wijaya, P. Sari, and F. Putri, "Detection Model for Potential Flooding Areas Using K-Means and *Local Outlier Factor* (LOF)," *Sensors*, vol. 24, no. 1, p. 203, 2024.
[10] M. Mutmainah and W. Yustanti, "Studi Komparasi *Local Outlier Factor* (LOF) dan Isolation Forest (IF) pada Analisis Anomali Kinerja Dosen," *Journal of Informatics and Computer Science (JINACS)*, vol. 6, no. 2, pp. 532-540, 2024, doi: 10.26740/jinacs.v6n02.p532-540.
[11] U. Rashid, M. F. Saleem, S. Rasool, A. Abdullah, H. Mustafa, and A. Iqbal, "Anomaly Detection Using Clustering (K-Means with DBSCAN) and SMO," *Journal of Computing and Biomedical Informatics*, vol. 7, no. 2, pp. 245-258, 2024, url: https://jcbi.org/index.php/Main/article/view/598
[12] M. Tabassum, S. Mahmood, A. Bukhari, B. Alshemaimri, and A. Daud, "Anomaly-Based Threat Detection in Smart Health Using Machine Learning," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 347, 2024, doi: 10.1186/s12911-024-02760-4.
[13] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York: Springer, 2020.
[14] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1-58, 2023.
[15] N. Nugraha, G. P. Insany, and R. K. Wardana, "Implementation of the *Local Outlier Factor* Model for Anomaly Detection in KPU Sukabumi Regency Voter Data," *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika (JIPI),* vol. 10, no. 1, pp. 295-303, 2025, url: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi/article/view/5795
[16] A. Gaffet, P. Ribot, E. Chanthery, and C. Merle, "Multi-block *Local Outlier Factor* for Anomaly Detection in Complex Systems," *IEEE Transactions on Industrial Informatics*, vol. 21, no. 2, pp. 1234-1245, 2025, doi: 10.1007/s10489-025-06603-1
[17] C. Singh, A. Sharma, N. Verma, and S. Agarwal, "Enhancing Endpoint Security Through AI-Powered Anomaly Detection in Endpoint Logs Using *Local Outlier Factor* (LOF*)," in 2025 IEEE International Conference on Cyber Security and Protection of Digital Services*, pp. 1-6, 2025, doi: 10.1109/ICICV64824.2025.11085490.
[18] X. Li, Y. Wang, D. Wei, and Z. Liu, "SST-LOF: Container Anomaly Detection Method Based on Singular Spectrum Transformation and Local Outlier Factor*," in 2024 IEEE International Conference on Cloud Engineering (IC2E),* pp. 45-52, 2024, doi: 10.1109/TCC.2024.3514297
[19] S. Kumar, M. Singh, and M. Rahman, "Time-Series Anomaly Detection of Mozi Malware in IoT Devices Using Arima and Local Outlier Factor," *Computers & Security*, vol. 133, p. 103348, 2025.
[20] A. Adesh, G. Shobha, J. Shetty, and L. Xu, "*Local Outlier Factor* for Anomaly Detection in HPCC Systems," *Journal of Parallel and Distributed Computing*, vol. 192, p. 104923, 2024, doi: 10.1016/j.jpdc.2024.104923.
[21] M. Goldstein and S. Uchida, "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data," *PLOS ONE*, vol. 19, no. 5, p. e0306949, 2016, doi: 10.1371/journal.pone.0152173
[22] C. M. Bishop, Pattern Recognition and Machine Learning. New York: Springer, 2021.

[23] S. Rawat, A. Kumar, V. Singh, and M. Patel, "Clustering Medical Conditions in Patient Records Using Unsupervised Learning Techniques: A Comparative Study," *Biomedical and Pharmacy Journal*, vol. 18, no. 3, pp. 234-251, 2025, doi: 10.13005/bpj/3236.
[24] R. S. Ismanda and M. T. A. Silitonga, "Analisis Cluster Menggunakan K-Means untuk Segmentasi Data Pasien," *Jurnal Informatika dan Sistem Informasi*, vol. 5, no. 2, pp. 89-97, 2025, doi: 10.55606/juisik.v5i3.1811.
[25] A. K. Jailani and A. Erna, "Deteksi Anomali pada Rasio Jam Belajar dan Aktivitas Sosial terhadap Performa Akademis Mahasiswa menggunakan Metode *Local Outlier Factor* (LOF)," in *SISITI: Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, vol. 14, no. 1, pp. 79-88, Mar. 2025.
[26] SSRG (Scientific Software Research Group), "Unsupervised Machine Learning for Anomaly Detection," *International Journal of Electrical and Electronics Engineering,* vol. 17, no. 4, pp. 245-268, 2025, doi: 10.1109/NAPS46351.2019.9000400.
[27] N. Nugraha, G. P. Insany, and R. K. Wardana, "Implementation of the *Local Outlier Factor* Model for Anomaly Detection in KPU Sukabumi Regency Voter Data," *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika (JIPI)*, vol. 10, no. 1, pp. 295-303, 2025, url: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi/article/view/5795.
[28] Z. Luo, Y. Chen, and K. Li, "Hybrid Anomaly Detection Approach Combining LOF and Clustering for Log Analysis," *IEEE Transactions on Dependable and Secure Computing*, vol. 22, no. 1, pp. 145-158, 2025, doi: 10.1145/3492321.351956.
[29] A. Gaffet, P. Ribot, E. Chanthery, and C. Merle, "Multi-block *Local Outlier Factor* for Anomaly Detection in Complex Systems," *IEEE Transactions on Industrial Informatics,* vol. 21, no. 2, pp. 1234-1245, 2025, doi: 10.1007/s10489-025-06603.
[30] L. Benova and L. Hudec, "Comprehensive Analysis and Evaluation of Anomalous User Activity in Web Server Logs," *Sensors*, vol. 24, no. 3, p. 746, 2024, doi: 10.3390/s24030746.
[31] M. Tabassum, S. Mahmood, A. Bukhari, B. Alshemaimri, and A. Daud, "Anomaly-Based Threat Detection in Smart Health Using Machine Learning," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 347, 2024, doi: 10.1186/s12911-024-02760-4.
[32] S. Kumar, M. Singh, and M. Rahman, "Time-Series Anomaly Detection of Mozi Malware in IoT Devices Using Arima and Local Outlier Factor," *Computers & Security*, vol. 133, p. 103348, 2025.
[33] C. Singh, A. Sharma, N. Verma, and S. Agarwal, "Enhancing Endpoint Security Through AI-Powered Anomaly Detection in Endpoint Logs Using *Local Outlier Factor* (LOF*)," in 2025 IEEE International Conference on Cyber Security and Protection of Digital Services*, pp. 1-6, 2025.
[34] X. Li, Y. Wang, D. Wei, and Z. Liu, "SST-LOF: Container Anomaly Detection Method Based on Singular Spectrum Transformation and Local Outlier Factor," *in 2024 IEEE International Conference on Cloud Engineering (IC2E)*, pp. 45-52, 2024.
[35] U. Rashid, M. F. Saleem, S. Rasool, A. Abdullah, H. Mustafa, and A. Iqbal, "Anomaly Detection Using Clustering (K-Means with DBSCAN) and SMO," *Journal of Computing and Biomedical Informatics*, vol. 7, no. 2, pp. 245-258, 2024, url: https://jcbi.org/index.php/Main/article/view/598
[36] P. Filonov, A. Lavrentyev, and A. Zhavoronkov, "RNN-based Early Warning System for Anomaly Detection in Critical Infrastructure of Healthcare," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 3, pp. 2345-2356, 2024.
[37] S. Rawat, A. Kumar, V. Singh, and M. Patel, "Clustering Medical Conditions in Patient Records Using Unsupervised Learning Techniques: A Comparative Study," *Biomedical and Pharmacy Journal*, vol. 18, no. 3, pp. 234-251, 2025, doi: 10.13005/bpj/3236.
[38] M. Rahman, Y. Zhang, Q. Liu, and Y. Chen, "Highly Accurate Anomaly-Based Intrusion Detection Using Deep Learning," *Computers & Security*, vol. 135, p. 103890, 2025.
[39] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. San Francisco: Morgan Kaufmann, 2022.
[40] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York: Springer, 2020.
[41] C. M. Bishop, Pattern Recognition and Machine Learning. New York: Springer, 2021.
[42] SSRG (Scientific Software Research Group), "Unsupervised Machine Learning for Anomaly Detection*," International Journal of Electrical and Electronics Engineering*, vol. 17, no. 4, pp. 245-268, 2025.
[43] A. Wijaya, P. Sari, and F. Putri, "Detection Model for Potential Flooding Areas Using K-Means and *Local Outlier Factor* (LOF)," *Sensors*, vol. 24, no. 1, p. 203, 2024.
[45] K. P. Murphy, Machine Learning: A Probabilistic Perspective. Cambridge: MIT Press, 2023.
[46] L. Tang, X. Wang, Y. Li, and H. Zhang, "Anomaly Detection in Medical Records via Multimodal Foundation Models," *Frontiers in Bioengineering and Biotechnology*, vol. 13, pp. 1-18, 2025, doi: 10.3389/fbioe.2025.1644697
[48] A. Rodriguez and A. Laio, "Clustering by Fast Search and Find of Density Peaks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 3, pp. 1234-1245, 2024, doi: 10.1126/science.124207.