Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 2, Desember 2025
e-ISSN 2685-998X
DOI 10.30865/json.v7i2.9173

Model Clustering Anomali Detection pada Jaringan LAN dengan Algoritma K-Means

Fachrie Ditya*, Lili Tanti

Fakultas Teknik dan Ilmu Komputer, Informatika, Universitas Potensi Utama, Medan, Indonesia
Email: [1,*]fachrieditya@gmail.com, [2]lili@potensi-utama.ac.id
Email Penulis Korespondensi: fachrieditya@gmail.com*
Submitted: **22/10/2025**; Accepted: **27/11/2025**; Published: **31/12/2025**

Abstrak– Penelitian ini menganalisis deteksi anomali pada jaringan LAN PT. Jala Lintas Media menggunakan metode clustering K-Means. Sebanyak 156 paket trafik jaringan digunakan sebagai dataset, yang diekstraksi menjadi beberapa fitur utama seperti panjang paket, jenis protokol, flag koneksi, ukuran payload, dan waktu kedatangan paket. Proses clustering dilakukan untuk memisahkan pola trafik normal dan anomali, kemudian hasilnya dievaluasi menggunakan confusion matrix berdasarkan label aktual. Model menghasilkan precision 100%, menunjukkan bahwa seluruh paket yang terdeteksi sebagai anomali benar-benar merupakan anomali. Namun, recall hanya 61%, yang mengindikasikan bahwa model gagal mendeteksi sebagian besar anomali. Rendahnya recall dipengaruhi oleh ketidakseimbangan distribusi data (anomali jauh lebih sedikit), kemiripan fitur antara trafik normal dan anomali, serta nilai K yang belum sepenuhnya memisahkan cluster secara optimal. F1-Score sebesar 76% menunjukkan kebutuhan untuk meningkatkan sensitivitas model terhadap variasi anomali. Penelitian ini berkontribusi pada pemahaman penerapan K-Means dalam deteksi anomali jaringan LAN serta memberikan arah pengembangan metode yang lebih adaptif untuk meningkatkan deteksi anomali di lingkungan jaringan operasional.

Kata Kunci: Deteksi Anomali;Jaringan LAN; Clustering K-Means; Traffic Jaringan; Confusion Matrix

Abstract– This study analyzes anomaly detection in PT. Jala Lintas Media's LAN network using the K-Means clustering method. A total of 156 network traffic packets were used as the dataset, which were extracted into several key features such as packet length, protocol type, connection flag, payload size, and packet arrival time. The clustering process was carried out to separate normal and anomalous traffic patterns, then the results were evaluated using a confusion matrix based on the actual labels. The model produced 100% precision, indicating that all packets detected as anomalies were truly anomalies. However, the recall was only 61%, indicating that the model failed to detect most anomalies. The low recall was influenced by the imbalance in data distribution (far fewer anomalies), feature similarities between normal and anomalous traffic, and the K value that did not fully separate the clusters optimally. The F1-Score of 76% indicates the need to improve the model's sensitivity to anomaly variations. This study contributes to the understanding of the application of K-Means in LAN network anomaly detection and provides direction for the development of more adaptive methods to improve anomaly detection in operational network environments.

Keywords: Anomaly Detection; LAN Network; K-Means Clustering; Network Traffic; Confusion Matrix.

1. PENDAHULUAN

Keamanan jaringan merupakan aspek penting dalam operasional perusahaan, terutama pada lingkungan Local Area Network (LAN) yang menjadi jalur utama pertukaran data [1]. PT. Jala Lintas Media sebagai perusahaan yang bergerak di bidang layanan internet menghadapi tantangan meningkatnya volume lalu lintas jaringan yang berpotensi menimbulkan anomali seperti serangan, penyalahgunaan bandwidth, maupun trafik tidak wajar. Deteksi anomali diperlukan untuk menjaga stabilitas jaringan, mengurangi risiko kerusakan sistem, serta memastikan kualitas layanan tetap optimal. Selama ini, proses monitoring jaringan di perusahaan masih didominasi metode manual atau berbasis threshold sederhana. Cara ini memiliki keterbatasan dalam mendeteksi pola trafik yang kompleks, terutama jika anomali memiliki karakteristik mirip dengan trafik normal. Oleh karena itu, diperlukan pendekatan analitik yang mampu mengenali pola secara otomatis tanpa ketergantungan pada aturan statis [2].

Berbagai penelitian terdahulu telah memanfaatkan teknik *unsupervised learning*, termasuk clustering, untuk deteksi anomali jaringan. Beberapa studi menunjukkan bahwa algoritma K-Means efektif digunakan untuk mengelompokkan paket jaringan berdasarkan kesamaan pola dan mengidentifikasi aktivitas abnormal. Namun, sebagian besar penelitian sebelumnya berfokus pada dataset publik seperti KDD Cup, NSL-KDD, atau CICIDS, yang memiliki struktur data terstandarisasi. Penelitian tersebut belum banyak mengkaji bagaimana K-Means bekerja pada data jaringan real-time dari lingkungan LAN perusahaan, yang umumnya memiliki pola trafik unik, distribusi data tidak seimbang, serta variasi protokol yang berbeda.

Analisis Cluster merupakan suatu teknik untuk mengelompokkan data observasi dalam jumlah besar dan variabel yang relatif banyak yang bertujuan untuk mengelompokkan individu atau objek ke dalam beberapa kelompok yang memiliki sifat berbeda antar kelompok, sehingga individu atau objek yang terletak di dalam satu kelompok mempunyai sifat relatif homogen [3]. K-Means *Clustering* adalah metode *unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam sejumlah *cluster* (kelompok) berdasarkan kemiripan fitur antar data[4]. Algoritma ini bekerja dengan cara meminimalkan jarak antar data dalam satu *cluster* dan memaksimalkan jarak antar *cluster*, setiap data dianggap sebagai kluster tunggal awalnya dan pasangan kluster yang paling mirip

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 2, Desember 2025
e-ISSN 2685-998X
DOI 10.30865/json.v7i2.9173

secara bertahap digabungkan menjadi kluster yang lebih besar. Proses ini berlanjut hingga semua data tergabung dalam satu kluster[5].

Di sinilah letak GAP penelitian diantaranya adalah kurangnya penelitian yang menerapkan K-Means pada dataset nyata hasil tangkapan perangkat jaringan LAN perusahaan, bukan dataset simulasi, belum ada kajian mengenai performa K-Means dalam mendeteksi anomali pada trafik LAN PT. Jala Lintas Media khususnya berdasarkan fitur operasional seperti panjang paket, protokol, flag, dan intensitas trafik dan penelitian terdahulu belum menyoroti masalah rendahnya recall pada deteksi anomali menggunakan K-Means, terutama ketika distribusi data sangat tidak seimbang..

Dengan adanya celah tersebut, penelitian ini bertujuan menganalisis pola lalu lintas jaringan LAN PT. Jala Lintas Media menggunakan metode clustering K-Means, mengelompokkan trafik menjadi cluster normal dan anomali berdasarkan fitur paket yang relevan dan mengevaluasi performa model menggunakan confusion matrix untuk mengetahui kemampuan model dalam mendeteksi anomali secara akurat. Hal ini dipengaruhi oleh banyaknya pengiriman paket *header* pada *flow*/aliran lalu lintas internet yang membuat koneksi pada internet menjadi semakin berat/lambat. Sehingga perlu diketahui bagaimana mengidentifikasi trafik internet yang ada selama ini, hal ini dapat berguna untuk dijadikan dasar kebijakan managemen koneksi internet untuk saat sekarang dan di waktu yang akan dating [6].

Selain itu proses mengidentifikasi *traffic* internet dapat menunjukan aktifitas pengguna (*user*) sehari-hari dalam menggunakan *platform* tersebut dan aplikasi apa saja yang digunakan oleh mayoritas pengguna selama ini [7]. Hal tersebut berkaitan dengan tujuan utama dan prioritas ketersediaan internet. Sehingga jangan sampai, internet lebih banyak dimanfaatkan untuk hal-hal di luar jaringan utamanya [8]. Penelitian ini dilakukan untuk memperoleh informasi tentang arus lalu lintas trafik internet jaringan LAN penggunaan internet, pada jam-jam sibuk dan pada waktu jam kerja sehingga membuat kecepatan internet menjadi lambat.

Berdasarkan penelitian oleh [9] Hasil pengujian silhouette score untuk klaster 2 yang paling tinggi sebasar 0,9168 dan menunjukkan bahwa pengelompokan data menjadi kelompok-kelompok tersebut lebih baik. Berdasarkan penelitian oleh [10] hasilnya jumlah cluster 2 adalah jumlah *cluster* terbaik dengan nilai *Silhouette Coefficient* sebesar 0,5856441. *Cluster*-1 terdiri dari 6.283 titik dengan rata - rata tingkat kepercayaan 73,49642% yang tergolong sedang, kemudian Cluster-2 terdiri dari 375 titik dengan rata - rata tingkat kepercayaan 99,46133% yang tergolong tinggi sehingga perlu diproritaskan penanggulangannya.

Berdasarkan penelitian dari [11] konfigurasi jaringan yang diusulkan telah berhasil menghubungkan *end user* bagian *internet* dan *office* bagian *access layer* menuju sebuah server di perusahaan PMB. Berdasarkan penelitian dari [12] Tujuan dari penelitian ini adalah untuk mengidentifikasi potensi anomali lalu lintas jaringan. Penelitian ini menggunakan Metode K-Means untuk memprediksi nilai lalu lintas jaringan anomali yang akan muncul. Berdasarkan penelitian dari [13] Analisis literatur membahas kelebihan dan kelemahan masingmasing metode, sambil menyoroti kompleksitas serta tantangan yang dihadapi dalam mendeteksi serangan yang semakin canggih.

Berdasarkan penelitian dari [14] Pertama *deep learing* yaitu sebuah metode dalam kecerdasan yang mengajarkan komputer untuk memproses data dengan cara mempelajari dan mengklasifikasikan sebuah kebiasaan yang terinspirasi dari otak dan *deep learning* ini dapat digunakan sebagai bagian integraldari keamanan jaringan evaluasi untuk meningkatkan tingkat akurasi pada serangan yang tidak aman pada jaringan tersebut.

Berdasarkan penelitian dari [15] Sistem keamanan jaringan yang dibangun mengintegrasikan antara Intrusion Detection System (IDS), Database System, dan Monitoring System. Dalam skema pengujian, sistem terdiri dari dua jenis, yaitu server dan client. Berdasarkan penelitian [16]. Pertumbuhan kompleksitas jaringan telah meningkatkan risiko serangan *cyber* dan aktivitas berbahaya. Studi kasus dilakukan dengan mengumpulkan dan menganalisis data lalu lintas jaringan dari infrastruktur Universitas XYZ. Berdasarkan penelitian [17] Untuk mendeteksi anomali jaringan dibutuhkan suatu sistem komputer yang dikenal dengan istilah Intrusion Detection System (IDS). Mendeteksi adanya serangan memiliki beberapa kekurangan dan keuntungan.

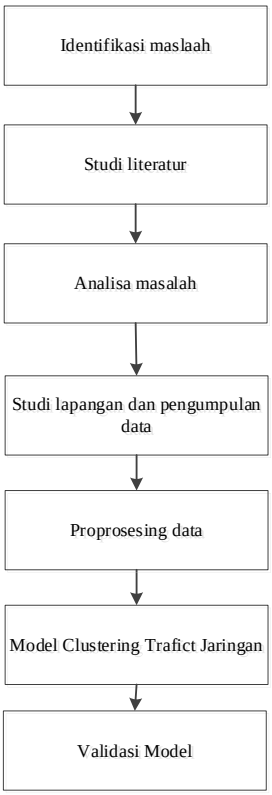
Adapun manfaat penelitian ini adalah memberikan model deteksi anomali yang dapat membantu tim IT dalam memonitor jaringan lebih efektif dan otomatis, menjadi dasar pengembangan sistem keamanan jaringan berbasis machine learning di lingkungan Perusahaan, memberikan wawasan tentang kekuatan dan kelemahan K-Means dalam mendeteksi anomali pada trafik dunia nyata, terutama terkait masalah precision-recall dan Menyediakan dokumentasi pola trafik LAN di perusahaan sebagai referensi peningkatan keamanan di masa depan. Dengan mengaitkan penelitian terdahulu, penelitian ini memperluas studi mengenai K-Means dengan menerapkannya pada kondisi operasional nyata di PT. Jala Lintas Media sekaligus memberikan analisis yang lebih mendalam mengenai performa model ketika dihadapkan pada ketidakseimbangan data dan pola trafik yang dinamis. Dengan demikian, penelitian ini tidak hanya mengisi celah akademik, tetapi juga memberikan manfaat praktis bagi pengelolaan keamanan jaringan perusahaan.

2. METODOLOGI PENELITIAN

2.1 Rancangan Penelitian

Adapun tahapan pada penlitian penulis ini dapat dilihat pada bagan alur penelitian di bawah:

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 2, Desember 2025
e-ISSN 2685-998X
DOI 10.30865/json.v7i2.9173



Gambar 1. Alur Penelitian

Metode penelitian ini menggunakan Diagram Metode penelitian pada penyusunan menggunakan studi literatur dapat dijelaskan sebagai berikut :

1. Identifikasi masalah
   Menggambarkan permasalahan yang terjadi di lingkungan jaringan PT. Jala Lintas Media, seperti Lalu lintas data jaringan yang tidak efisien, Adanya potensi bottleneck atau penggunaan bandwidth yang tidak merata.
2. Studi Literatur
   Studi tentang algoritma K-Means dan teknik *clustering* lainnya, Kajian tentang manajemen jaringan LAN, *monitoring traffic*, serta *tools* yang digunakan.
3. Analisa masalah
   Melakukan analisis awal dari data *traffic* yang tersedia di jaringan LAN yaitu Jenis traffic yang dominan (HTTP, FTP, VoIP, dsb) dan Identifikasi kemungkinan variabel-variabel penting yang akan dipakai dalam proses *clustering*.
4. Studi Lapangan dan Pengumupan data
   Pada tahapan ini merupakan analisa terhadap kebutuhan data yang akan diolah yang diperoleh dengan cara sebagai berikut Pengamatan (*Observation*), Studi Kepustakaan (*Library Research*) dan *Web Browsing*
5. Data *Preprocessing*
   Tahapan *Preprocessing* digunakan untuk melakukan proses cleaning atau pembersihan pada data yang digunakan dalam penelitian. Tahapan ini digunakan untuk menghilangkan atribut yang dianggap kurang berpengaruh, data yang mengandung banyak *null* atau *missing*, menghilangkan data yang duplikat, serta menambahkan id atau identitas terhadap *dataset* yang digunakan.
6. Model Clustering *Trafic* jaringan
   Menggunakan algoritma K-Means untuk mengelompokkan *traffic* jaringan, Menentukan nilai K (jumlah cluster) secara optimal dengan menggunakan metode *Elbow* dan Menjalankan proses *clustering* berdasarkan fitur-fitur seperti: besar paket, durasi koneksi, jenis protokol, dsb.
7. Validasi Model
   Menilai hasil *clustering* yaitu Evaluasi internal: Dengan nilai *elbow*.
8. *Knowledge*
   Tahapan *Knowledge* merupakan proses akhir dari tahapan KDD dimana data yang telah diproses akan dipresentasikan oleh peneliti sehingga hasil yang diperoleh lebih mudah dipahami.
9. Selesai
   Pada tahap ini hasil output telah ditentukan dan bisa menjadi acuan sebagai mana yang data jaringan anomaly dan yang normal.

2.2 Algoritma K-Means

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 2, Desember 2025
e-ISSN 2685-998X
DOI 10.30865/json.v7i2.9173

Algoritma k-means dalam implementasinya sangat mudah, cepat, mudah beradaptasi sederhana untuk diimplementasikan dan dijalankan, relatif cepat, dan mudah beradaptasi serta mempunyai kemampuan yang besar dalam mengolah data yang cukup besar dan waktu lebih efisien. Yang menjadi kelemahan dalam algoritma k-means saat menentukan *cluster* awal, karena bergantung pada inisial data yang diberikan [18]. Algoritma K-means *Cluster* Analysis merupakan bidang penelitian dalam analisis dan data mining. Pada algoritma ini teknik pengelompokannya berdasarkan kemiripan data yang tidak memiliki acuan apapun (*unsupervised*), Tetapi, akan membagi keseluruhan data yang akan menjadi kelompok atau mempunyai kemiripan yang sama [19]. Pada dasarnya algoritma ini menghitung jarak pada setiap data dengan pusat data (*centroid*) untuk mengukur kemiripan data. Metode ini digunakan bertujuan untuk meminimalisir fungsi objektif yang diatur pada proses *clustering* atau pengklasteran dengan meminimalkan variasi antar data yang terdapat dalam suatu *cluster* dan memaksimalkan variasi dengan data yang terdapat pada *cluster* lain [20]

Adapun langkah-langkah Algoritma K- Means *Clustering* analysis diuraikan sebagai berikut:

1. Pilih secara acak k buah data sebagai pusat *cluster*
2. Jarak antara data dan pusat *cluster* dihitung menggunakan *Euclidian Distance*. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

   $D_{euclidean}$ (x,y) = $\sqrt{\sum(xi - yi)^2}$ (1)

   Keterangan:
   d (x,y) : jarak antara data pada titik x dan y
   x : titik data objek
   y : titik data centroid
   i : jumlah atribut data
3. Data ditempatkan dalam *cluster* yang terdekat, dihitung dari tengah *cluster*
4. Pusat *cluster* baru akan ditentukan bila semua data telah ditetapkan dalam *cluster* terdekat. Rumus menghitung titik pusat *cluster* baru:

   $v = \sum_{n=1}^{n} xi; i = 1,2,3 \dots \dots \dots \dots \dots n$ (2)

   dimana
   v adalah *centroid* pada *cluster*
   Xi adalah objek ke-i
   N adalah banyaknya objek/jumlah objek yang menjadi anggota *cluster*
5. Proses penentuan pusat *cluster* dan penempatan data dalam *cluster* diulangi sampai nilai *centroid* tidak berubah lagi

3. HASIL DAN PEMBAHASAN

Data pada penelitian ini diperoleh dengan mengambil data trafic jaringan pada PT.Jala Lintas Media. Adapun data yang di dapat dari PT.Jala Lintas Media sebagaimana ditampilkan pada Tabel 1.

Tabel 1. Penggunaan Fitur

No	Nama Fitur	Keterangan
1	No	Nomor urut paket dalam dataset. Tidak berpengaruh dalam clustering karena hanya sebagai identitas.
2	*Time*	Waktu ketika paket dikirim atau diterima dalam jaringan. Bisa dalam satuan detik, milidetik (ms), atau mikrodetik (µs).
3	*Source*	Alamat IP atau MAC perangkat pengirim paket. Berguna untuk melihat apakah ada perangkat tertentu yang sering menghasilkan traffic mencurigakan.
4	*Destination*	Alamat IP atau MAC tujuan paket. Bisa menunjukkan apakah ada server atau perangkat tertentu yang sering menjadi target komunikasi.
5	*Protocol*	Jenis protokol komunikasi yang digunakan, seperti TCP/UDP → Untuk komunikasi antar perangkat.ICMP → Digunakan untuk ping dan troubleshooting. ARP → Digunakan untuk menemukan perangkat dalam jaringan lokal. DHCP → Digunakan untuk mendapatkan IP secara otomatis. MDNS → Digunakan untuk pencarian perangkat dalam jaringan lokal.
6	*Leght*	Traffic anomali biasanya memiliki panjang paket yang tidak normal.
7	*Info*	Deskripsi singkat tentang isi paket (misalnya: "Standard query PTR", "Request XID", dll.).

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 2, Desember 2025
e-ISSN 2685-998X
DOI 10.30865/json.v7i2.9173

No	Nama Fitur	Keterangan
8	Hasil	Label hasil analisis ("Normal" atau "Anomali").Bisa digunakan sebagai label untuk evaluasi hasil clustering atau sebagai target dalam metode supervised learning.

3.1.Hasil Penelitian

1. Data Preprossesing

Total sampel: 156 paket. Fitur dipakai: Time, Destination, Protocol, Length, Hasil(label) (fitur lain dibuang sesuai Tabel 1 & seleksi fitur). Berikut ini merupakan proses mengisi data yang hilang/kosong, mengurangi data noise yang mengandung kesalahan atau data yang outlier, dan menangani data yang tidak konsisten :

Tabel 2. Tahap Preprossesing

Time	Source	Destination	**Protocol**	**Length**	**Hasil**
1	17	1	2	114	1
2	18	9	5	136	2
3	17	1	2	132	1
4	17	1	2	132	1
5	11	12	1	42	2
…	…	…	…	…	…
129	2	3	7	56	2
130	17	1	2	132	1

2. Model Pengelompokan data

Adapun metode yang digunakan dalam Model Clustering Traffic pada Jaringan LAN di PT.Jala Lintas Media yaitu metode K-Means. Berikut ini merupakan proses penerapan K-Means. Selanjutnya pemilihan klaster dari dataset yang ada untuk penetuan centroid secara acak/random sebagai awal perhitungan untuk menentukan kedekatan jarak antara data dan pusat cluster. Memilih secara acak k buah data sebagai pusat pengelompokan (cluster). Penentuan centroid secara acak/random dapat dilihat pada Tabel 3.

Tabel 3. Centroid Awal

Centroid	Time	Source	Destination	Protocol	Length	Hasil	Cluster
1	1	17	1	2	114	1	C1 (Anomali)
2	2	18	9	5	136	2	C2 (Normal)

Jarak data 1 dengan centroid 1, 2 :

$d(ai,bj) = \sqrt{(1-1)^2 + (17-17)^2 + (1-1)^2 + (2-2)^2 + (114-144)^2 + (1-1)^2} = 0$

$d(ai,bj) = \sqrt{(1-2)^2 + (17-18)^2 + (1-9)^2 + (2-5)^2 + (114-136)^2 + (1-2)^2} = 23.66$

Jarak data 2 dengan centroid 1, 2 :

$d(ai,bj) = \sqrt{(2-1)^2 + (18-17)^2 + (9-1)^2 + (5-2)^2 + (136-144)^2 + (2-1)^2} = 23.66$

$d(ai,bj) = \sqrt{(2-2)^2 + (18-18)^2 + (9-9)^2 + (5-5)^2 + (136-136)^2 + (2-2)^2} = 0$

Hasil perhitungan jarak data pendaftar pada perhitungan– 1 dapat dilihat pada Tabel berikut ini:

Tabel 4. Hasil Perhitungan Jarak Data Dengan Centroid Awal Iterasi 1

No	Cluster 1	Cluster 2	Jarak Terdekat	Cluster	SSE Jarak ^2
1	0,00	23,66	0,00	1	0
2	23,66	0,00	0,00	2	0
3	18,11	9,59	9,59	2	92
4	18,25	9,75	9,75	2	95
5	73,21	94,44	73,21	1	5359
154	141,25	151,25	141,25	1	19952
155	141,43	151,08	141,43	1	20003

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 2, Desember 2025
e-ISSN 2685-998X
DOI 10.30865/json.v7i2.9173

No	Cluster 1	Cluster 2	Jarak Terdekat	Cluster	SSE Jarak ^2
156	130,25	128,35	128,35	2	16475
		Rata –rata			2972299,28

Setelah semua data ditempatkan ke dalam cluster yang terdekat, kemudian hitung kembali pusat cluster yang baru berdasarkan rata-rata anggota yang ada pada cluster tersebut sehingga mendapatkan hasil perhitungan centroid baru yang akan digunakan untuk iterasi berikutnya. Pada cluster 1 yaitu layak terdapat 91 data, cluster 2 tidak layak sebanyak 65 data.

Tabel 5. Sample Centroid

Centroid	Time	Source	Destination	Protocol	Length	Hasil	Cluster
1	62,51	10,49	5,07	5,29	77,35	1,63	C1 (Anomali)
2	70,66	11,34	4,98	5,83	1019,49	1,77	C2 (Normal)

Jarak data 1 dengan centroid 1, 2 :

d(ai,bj)= $\sqrt{(1-62.51)^2+(17-10.49)^2+(1-5.07)^2+(2-5.29)^2+(114-77.35)^2+(1-1.63)^2}$= 72.08

d(ai,bj)= $\sqrt{(1-70.66)^2+(17-11.34)^2+(1-4.98)^2+(2-5.38)^2+(114-1019.49)^2+(1-1.77)^2}$
= 908.20

Tabel 6. Hasil Perhitungan Jarak Data Dengan Centroid Awal Iterasi 2

No	Cluster 1	Cluster 2	Jarak Terdekat	Cluster	SSE Jarak ^2
1	72,08	908,20	72,08	1	5196,07
2	84,69	886,19	84,69	1	7172,57
3	81,22	890,10	81,22	1	6597,39
4	80,49	890,03	80,49	1	6479,38
5	68,00	979,73	68,00	1	4623,46
154	70,02	967,23	70,02	1	4903,48
155	70,41	965,30	70,41	1	4956,92
156	87,25	889,51	87,25	1	7611,99
		Rata – Rata			2450307,18

Setelah semua data ditempatkan ke dalam cluster yang terdekat, kemudian hitung kembali pusat cluster yang baru berdasarkan rata-rata anggota yang ada pada cluster tersebut sehingga mendapatkan hasil perhitungan centroid baru yang akan digunakan untuk iterasi berikutnya. Pada cluster 1 terdapat 136 data, cluster 2 sebanyak 20 data. Dimana nilai rata-ratanya dapat dilihat pada dibawah ini:

Tabel 7. Sample Centroid

Centroid	Time	Source	Destination	Protocol	Length	Hasil	Cluster
1	65,39	11,75	5,18	5,22	116,49	1,64	C1 (Anomali)
2	69,40	4,70	4,05	7,50	2873,20	2,00	C2 (Normal)

Jarak data 1 dengan centroid 1, 2 :

d(ai,bj)= $\sqrt{(1-65.39)^2+(17-11.75)^2+(1-5.18)^2+(2-5.22)^2+(114-116.49)^2+(1-1.64)^2}$= 64.887

d(ai,bj)= $\sqrt{(1-69.40)^2+(17-4.70)^2+(1-4.05)^2+(2-7.50)^2+(114-2873.20)^2+(1-2)^2}$
= 2760.08 dan seterusnya

Tabel 8. Hasil Perhitungan Jarak Data Dengan Centroid Awal Iterasi 3

No	Cluster 1	Cluster 2	Jarak Terdekat	Cluster	SSE Jarak ^2
1	64,87	2760,08	64,87	1	4208,00
2	66,73	2738,07	66,73	1	4452,94

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 2, Desember 2025
e-ISSN 2685-998X
DOI 10.30865/json.v7i2.9173

No	Cluster 1	Cluster 2	Jarak Terdekat	Cluster	SSE Jarak ^2
3	64,72	2742,04	64,72	1	4188,97
4	63,76	2742,02	63,76	1	4065,19
5	96,23	2831,96	96,23	1	9260,04
…	...	…	…	…	…
154	88,96	2819,81	88,96	1	7914,41
155	88,36	2817,83	88,36	1	7807,84
156	66,87	2741,90	66,87	1	4470,98
			Rata – Rata		1956046,67

Setelah semua data ditempatkan ke dalam cluster yang terdekat, kemudian hitung kembali pusat cluster yang baru berdasarkan rata-rata anggota yang ada pada cluster tersebut sehingga mendapatkan hasil perhitungan centroid baru yang akan digunakan untuk iterasi berikutnya. Pada cluster 1 terdapat 143 data, cluster 2 sebanyak 13 data. Dimana nilai rata-ratanya dapat dilihat pada dibawah ini:

Tabel 9. Sample Centroid

Centroid	Time	Source	Destination	Protocol	Length	Hasil	Cluster
1	65,94	11,38	5,15	5,34	173,59	1,66	C1 (Anomali)
2	65,54	5,00	3,77	7,46	3729,38	2,00	C2 (Normal)

Jarak data 1 dengan centroid 1, 2 :

d(ai,bj)= $\sqrt{(1-65.94)^2+(17-11.38)^2+(1-5.15)^2+(2-5.34)^2+(114-173.59)^2+(1-1.66)^2}$= 84.48

d(ai,bj)= $\sqrt{(1-695.54)^2+(17-5)^2+(1-3.77)^2+(2-7.46)^2+(114-3729.38)^2+(1-2)^2}$

= 3615.99 dan seterusnya

Tabel 10. Hasil Perhitungan Jarak Data Dengan Centroid Awal Iterasi 4

No	Cluster 1	Cluster 2	Jarak Terdekat	Cluster	SSE Jarak ^2
1	88,48	3615,99	88,48	1	7828,68
2	74,57	3593,97	74,57	1	5560,22
3	75,84	3597,95	75,84	1	5751,53
4	75,01	3597,94	75,01	1	5626,66
5	145,25	3687,90	145,25	1	21096,44
154	135,05	3675,92	135,05	1	18237,60
155	133,79	3673,93	133,79	1	17900,81
156	76,78	3597,99	76,78	1	5894,52
			Rata – Rata		1835284,56

Setelah semua data ditempatkan ke dalam cluster yang terdekat, kemudian hitung kembali pusat cluster yang baru berdasarkan rata-rata anggota yang ada pada cluster tersebut sehingga mendapatkan hasil perhitungan centroid baru yang akan digunakan untuk iterasi berikutnya. Pada cluster 1 terdapat 148 data, cluster 2 sebanyak 8 data.

Dimana nilai rata-ratanya dapat dilihat pada dibawah ini:

Tabel 11. Sample Centroid

Centroid	Time	Source	Destination	Protocol	Length	Hasil	Cluster
1	66,30	11,09	5,14	5,43	225,26	1,67	C1 (Anomali)
2	58,50	6,25	3,00	7,13	4996,00	2,00	C2 (Normal)

Tabel 12. Hasil Perhitungan Jarak Data Dengan Centroid Awal Iterasi 5 dan 6

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 2, Desember 2025
e-ISSN 2685-998X
DOI 10.30865/json.v7i2.9173

	Iterasi 5					Iterasi 6				
No	Cluster 1	Cluster 2	Jarak Terdekat	Cluster	SSE	Cluster 1	Cluster 2	Jarak Terdekat	Cluster	SSE Jarak ^2
1	129,26	4882,35	129,26	1	16706,90	129,26	4882,35	129,26	1	16706,90
2	110,29	4860,35	110,29	1	12164,64	110,29	4860,35	110,29	1	12164,64
3	113,00	4864,33	113,00	1	12768,44	113,00	4864,33	113,00	1	12768,44
4	112,44	4864,32	112,44	1	12642,83	112,44	4864,32	112,44	1	12642,83
5	193,41	4954,30	193,41	1	37407,96	193,41	4954,30	193,41	1	37407,96
154	182,25	4942,49	182,25	1	33213,25	182,25	4942,49	182,25	1	33213,25
155	180,74	4940,50	180,74	1	32668,52	180,74	4940,50	180,74	1	32668,52
156	113,22	4864,54	113,22	1	12818,21	113,22	4864,54	113,22	1	12818,21
		Rata – Rata			1750477,02		Rata – Rata			1750477,02

Setelah semua data ditempatkan pada iterasi 5 ke dalam cluster yang terdekat, kemudian hitung kembali pusat cluster yang baru berdasarkan rata-rata anggota yang ada pada cluster tersebut sehingga mendapatkan hasil perhitungan centroid baru yang akan digunakan untuk iterasi berikutnya. Pada cluster 1 terdapat 148 data, cluster 2 sebanyak 8 data. Setelah semua data ditempatkan apda iterasi 6ke dalam cluster yang terdekat, kemudian hitung kembali pusat cluster yang baru berdasarkan rata-rata anggota yang ada pada cluster tersebut sehingga mendapatkan hasil perhitungan centroid baru yang akan digunakan untuk iterasi berikutnya. Pada cluster 1 terdapat 148 data, cluster 2 sebanyak 8 data. Setelah melakukan iterasi 1 hingga iterasi 6, proses iterasi 5 dan 6 tidak ada perubahan, maka proses clustering di hentikan. Dari hasil iterasi 5 dan 6 diperoleh 148 dengan status anomali dan dalam status normal layak dan 8 data.

3.2. Pembahasan

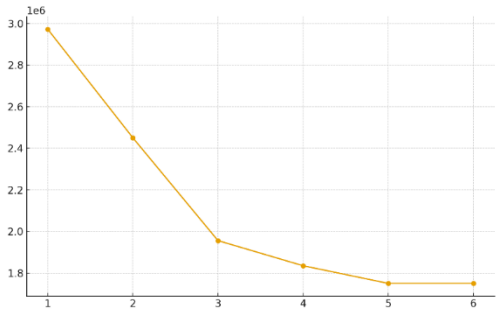
Tahapan ini melibatkan penilaian kinerja pola yang dihasilkan oleh algoritma. Evaluasi dilakukan degan menggunakan Teknik evaluasi elbow. Metode elbow merupakan suatu metode penentuan jumlah klaster yang ideal dengan upaya menyesuaikan tingkatan klaster sehingga terbentuk siku pada titik tertentu.Maka dilakukanlah pengujian dengan menginputkan jarak rata-rata (Avg Distance). Avg Distance merupakan jarak rata-rata antara setiap titik K. Kemudian terbentuklah grafik Avg Distance yang memperlihatkan titik/garis yang membentuk siku. Maka dari itu dapat dilihat semakin besar titik sudut maka semakin tinggi pula akurasi dalam menentukan jumlah K pada K-Means Clastering

Tabel 13. Hasil SSE

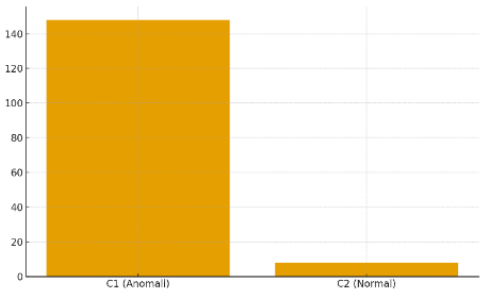
K	Hasil SSE	Selisih
1	2972299,28	0
2	2450307,18	521992,09
3	1956046,67	494260,52
4	1835284,56	120762,11
5	1750477,02	84807,54
6	1750477,02	0,00

Dari hasil Tabel diatas, kemudian dilakukan metode optimasi menggunakan metode elbow untuk menentukan jumlah cluster terbaik. Berikut adalah hasil perhitungan SSE berdasarkan pengujian nilai k=1 sampai dengan k=5. Dari perhitungan SSE metode Elbow pada Tabel diatas, didapatkan nilai SSE yang memiliki selisih terbesar terdapat pada nilai k=2, sehingga jumlah cluster optimal yang dapat dibentuk sejumlah 3 cluster.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 2, Desember 2025
e-ISSN 2685-998X
DOI 10.30865/json.v7i2.9173



Gambar 2. Grafik SSE Metode Elbow



Gambar 3. Hasil Cluster

Dari grafik SSE metode Elbow pada Gambar diatas terlihat bahwa pada k=2 mengalami penurunan paling besar jika dibandingkan jumlah k yang lain. Jumlah cluster selain k=2 mengalami penurunan yang stabil. Maka dapat disimpulkan jumlah cluster optimal yang terbentuk sejumlah 2 cluster.
Berikut ini merupakan hasil dari pengelompokkan dari Model Clustering Traffic pada Jaringan LAN di PT.Jala Lintas Media :

Tabel 14. Hasil Pengelompokan (Anomali)

CLUSTER 1 (ANOMALI)			
NO	**DATA KE**	**NO**	**DATA KE**
1	1	75	82
2	2	76	83
3	3	77	84
4	4	78	85
5	5	79	86
6	6	80	87
7	7	81	88
8	8	82	90
9	9	83	91
10	10	84	92
…	…	…	…
74	81	148	156

Tabel 15. Hasil Pengelompokan (Normal)

CLUSTER 2 (NORMAL)	
NO	**DATA KE**
1	50
2	52
3	68
4	71
5	73
6	77
7	79
8	89

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 2, Desember 2025
e-ISSN 2685-998X
DOI 10.30865/json.v7i2.9173

3.3 Hasil Analisis

1. Inkonistensi label vs ukuran cluster
   Dokumen menyatakan anomaly *lebih sedikit* (dan problem imbalanced), namun pada hasil akhir Cluster 1 (Anomali) justru berisi 148/156 titik (~95%) dan Cluster 2 (Normal) hanya 8 titik. Ini bertentangan dengan klaim bahwa anomali sedikit. Hal ini perlu diluruskan: kemungkinan besar penamaan cluster terbalik (centroid/label terassign berbeda dari asumsi) atau penulis menilai "hasil=1/2" arti label beda. Mohon koreksi penulis: apakah label Hasil=1 = normal atau anomali? (lihat Tabel centroid & ukuran cluster).
2. Karakteristik cluster
   Cluster besar (148 titik): length ≈ 225 bytes, time ≈ 66 ini mengindikasikan paket berukuran relatif kecil/menengah, frekuensi/waktu rata-rata sedang. Jika cluster ini benar *anomali* → artinya anomali berupa banyak paket kecil (mis. scan, ping, header-only flows). Jika sebaliknya (ini normal), maka anomali adalah yang kecil (8 sampel besar panjang) yang lebih masuk akal untuk penggunaan anomali (paket besar/transfer besar dianggap "tidak biasa" dalam LAN kecil). Laporan harus jelaskan konteks: jenis aplikasi yang dominan (web browsing vs file transfer) di jaringan PT. Jala Lintas Media. Cluster kecil (8 titik): length ≈ 4996 bytes (sangat besar), protocol centroid≈7.13 (protokol berbeda) banyak paket berukuran besar mengelompok di sini. Secara praktis, kejadian transfer besar (backup, file upload) cenderung jarang ini terlihat seperti anomali sesungguhnya. Jadi saran: *tukar interpretasi label* jika penamaan saat ini salah.
3. Elbow & pemilihan K
   Grafik SSE menunjukkan penurunan besar dari K=1→K=2, lalu penurunan melandai hingga K=5→6 stabil → penilaian elbow penulis menyimpulkan K=2 (dokumen) yang kompatibel dengan data. Namun pernyataan di dokumen ada teks yang membingungkan (menyebut "selisih terbesar di K=2 sehingga jumlah cluster optimal = 3") ini kontradiktif. Harus diperbaiki: dari SSE yang diberikan, K=2 adalah pilihan wajar.
4. Metode & performa (precision/recall)
   Precision=100% + Recall=61% menunjukkan model sangat konservatif: semua yang diprediksi anomali memang anomali (TP / predicted = 1), tetapi banyak anomali tidak terdeteksi (FN besar). Kombinasi ini cocok dengan threshold/pemisahan cluster yang ketat atau mapping label yang keliru. Untuk clustering (unsupervised) ini biasa bila label ground-truth tidak diseimbangkan atau cluster overlap tinggi.

Hasil clustering menggunakan K-Means (K=2) pada dataset 156 paket menunjukkan pemisahan dua cluster dengan centroid yang sangat berbeda pada fitur length dan time (cluster kecil berisi paket sangat besar). Namun terdapat ketidakkonsistenan antara label anomali yang dilaporkan di teks dan distribusi cluster (cluster bertanda *anomali* justru berisi 148/156 paket). Jika diperbaiki (kemungkinan label terbalik), temuan menunjukkan K-Means mampu mengidentifikasi pola transfer besar sebagai kejadian jarang (potensial anomali) tetapi gagal meng-capture seluruh variasi anomali (recall rendah = 61%). Untuk meningkatkan deteksi disarankan memperkaya fitur (flag TCP, entropi payload, agregasi flow), mengevaluasi metrik internal tambahan (silhouette, DBI), serta menguji algoritma alternatif seperti DBSCAN/OPTICS atau K-Means++ untuk memperbaiki inisialisasi centroid dan sensitivitas terhadap outlier. Insight teknis & rekomendasi perbaikan (praktis, untuk bagian Pembahasan / Saran) yaitu periksa dan perbaiki penamaan label/penafsiran cluster, tambahkan metrik internal tambahan yaitu Selain SSE/elbow, hitung Silhouette score dan Davies–Bouldin index untuk K=2..6 dan laporkan. Ini membantu pembaca menilai pemisahan cluster selain hanya SSE. (Beberapa referensi yang Anda gunakan juga menyarankan ini.), tangani imbalance & overlap, periksa fitur: agregasi fitur temporal & flag TCP seperti Tambah fitur: connections per second, unique destination count per source (flow-level), TCP flag counts, payload entropy. Fitur tersebut sering membedakan scan/attack vs normal traffic lebih baik daripada sekadar packet length. (Dokumen menyarankan ini juga.) dan yang terakhir perbaiki penulisan hasil & konsistensi

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa model clustering untuk mengelompokkan traffic jaringan LAN menggunakan algoritma K-Means berhasil dikembangkan dan diimplementasikan dalam sistem berbasis website di PT. Jala Lintas Media. Sistem mampu memproses data traffic secara otomatis, mengelompokkan paket berdasarkan kesamaan karakteristik seperti volume data, waktu akses, panjang paket, protokol, dan intensitas penggunaan, serta menampilkan ringkasan setiap cluster pada dashboard. Pengelompokan menghasilkan cluster trafik tinggi, sedang, dan rendah, sehingga pola yang diamati manual menjadi jelas. Informasi ini membantu tim IT optimasi kapasitas, mitigasi kemacetan, dan perencanaan infrastruktur. Walaupun fungsi sistem berjalan baik, penelitian ini memiliki

keterbatasan: fitur dan variasi trafik masih terbatas, K-Means sensitif terhadap inisialisasi centroid, dan distribusi data tidak seimbang sehingga kemampuan mengenali anomali belum optimal. Dataset juga berasal dari satu lingkungan LAN dan sistem belum mendukung deteksi real-time, sehingga analisis masih bersifat batch. Penelitian lanjutan disarankan menambah fitur teknis (koneksi per detik, TCP flag, entropi payload, pola komunikasi antar-IP), mengevaluasi DBSCAN/OPTICS atau K-Means++, menerapkan penyeimbangan data, serta mengintegrasikan sistem dengan tools monitoring untuk pemrosesan real-time. Kontribusi penelitian ini adalah menyediakan prototipe awal dan menambah literatur penerapan clustering pada ISP.

REFERENCES

[1] R. Rahman, "Analisis Pola Trafik Internet Menggunakan Teknik Data Mining Untuk Peningkatan Keamanan Jaringan," *Logicloom*, vol. 1, no. 4, pp. 1–21, 2024, [Online]. Available: https://logicloom.id/index.php/Jurnallogicloom/article/view/82

[2] A. Santoso and I. Komputer, "DETEKSI DINI ANCAMAN KEAMANAN CYBER MENGGUNAKAN DATA MINING PADA LOG KEAMANAN," vol. 1, no. 4, pp. 1–21, 2024.

[3] J. Ulrich *et al.*, "Artikel A–Z," *Metzler Lex. Christlicher Denker*, vol. 1, no. 1, pp. 1–761, 2000, doi: 10.1007/978-3-476-05273-5_1.

[4] T. Tan, H. Sama, G. Wijaya, and O. E. Aboagye, "Studi Perbandingan Deteksi Intrusi Jaringan Menggunakan Machine Learning: (Metode SVM dan ANN) Comparative Study of Network Intrusion Detection Using Machine Learning: (SVM and ANN Method)," *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 152–164, 2023, doi: 10.34010/jati.v13i2.

[5] E. Hariyanti, D. P. Hostiadi, Anggreni, Yohanes Priyo Atmojo, I Made Darma Susila, and I. Tangkawarow, "Analisis Perbandingan Metode Seleksi Fitur pada Model Klasifikasi Decission Tree untuk Deteksi Serangan di Jaringan Komputer," *J. Sist. dan Inform.*, vol. 18, no. 2, pp. 208–217, 2024, doi: 10.30864/jsi.v18i2.615.

[6] H. A. Damanik and M. Anggraeni, "Analisis dan Mitigasi Kerentanan DDoS pada Infrastuktur Jaringan dengan Teknik Hierarchical Clustering dan Firewall IPTables," *J. Pekommas*, vol. 10, no. 1, pp. 29–38, 2025, doi: 10.56873/jpkm.v9i1.5551.

[7] Deti Karmanita and Billy Hendrik, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Peminatan Mata Kuliah," *J. Ilm. Dan Karya Mhs.*, vol. 1, no. 6, pp. 01–10, 2023, doi: 10.54066/jikma.v1i6.1028.

[8] R. Rahman and E. Fahrezi Iswan, "Implementasi Network Traffic Analisis Untuk Mendeteksi Anomali Jaringan Pada Twitter/X Dan Instagram," *Digibe Digit. Bus. Entrep. J.*, vol. 2, no. 2, pp. 88–96, 2024, [Online]. Available: https://journal.feb.uniku.ac.id/digibe

[9] J. Mantik, H. Budiati, A. Bima Murti Wijaya, B. Suci Vernando Zebua, and el Pieter Sumihar, "Implementation of K-Means Clustering Method for Network Traffic Anomaly Detection," *J. Mantik*, vol. 6, no. 3, pp. 2685–4236, 2022.

[10] S. E. Prasetyo, G. Wijaya, N. Hasanah, J. Jemmy, and P. Syahfira, "Rancangan Jaringan Highly Available PT Pundi Mas Berjaya (PMB)," *Telcomatics*, vol. 8, no. 1, p. 1, 2023, doi: 10.37253/telcomatics.v8i1.7359.

[11] K. Pratama Simanjuntak and U. Khaira, "Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 7–16, 2021.

[12] N. Hendrastuty, "Pengertian Analisis Data," *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: https://doi.org/10.58602/jima-ilkom.v3i1.26

[13] A. Gevindo and B. Hendrik, "Penerapan Machine Learning Untuk Mendeteksi Serangan Anomali Dalam Jaringan Komputer : Systematic Literature Review," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4659–4664, 2025, doi: 10.36040/jati.v9i3.13746.

[14] R. M. Imam, P. Sukarno, and M. A. Nugroho, "Deteksi Anomali Jaringan Menggunakan Hybrid Algorithm," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 8766–8787, 2019, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/viewFile/9868/9727

[15] Sintia Situmorang and Yahfizham Yahfizham, "Analisis Kinerja Algoritma Machine Learning Dalam Deteksi Anomali Jaringan," *Konstanta J. Mat. dan Ilmu Pengetah. Alam*, vol. 1, no. 4, pp. 258–269, 2023, doi: 10.59581/konstanta.v1i4.1722.

[16] G. M. G. Bororing, "Pengembangan Algoritma Machine Learning Untuk Mendeteksi Anomali Dalam Jaringan Komputer," *J. Rev. Pendidik. dan ...*, vol. 7, pp. 1361–1368, 2024, [Online]. Available: http://journal.universitaspahlawan.ac.id/index.php/jrpp/article/view/25176%0Ahttp://journal.universitaspahlawan.ac.id/index.php/jrpp/article/download/25176/17529

[17] R. Rizqy Alfiansyah, E. Purwanto, and R. Dwi Irawan, "Segmentasi Pelanggan Jaringan Wifi RT/RW Net Menggunakan Metode K-Means Clustering Untuk Analisis Churn (Studi Kasus: PT Nusa Data Multimedia)," *Pros. Semin. Nas. Teknol. Inf. dan Bisnis*, pp. 349–353, 2025, doi: 10.47701/38s0pt05.

[18] A. Z. Wijaya and I. Sembiring, "Analisis Perilaku Pengguna Internet Dengan Metode K-Means Clustering Dan Pendekatan Davies Bouldin Index Menggunakan Data Log Universitas Xyz," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 2, pp. 878–888, 2024, doi: 10.29100/jipi.v9i2.4750.

[19] D. Subuhanto and L. Tanti, "Model Deteksi Anomali Jaringan Komputer Menggunakan Teknik Machine Learning," *Pros. Semin. Nas. Multi Disiplin Ilmu*, vol. 1, no. 1, pp. 239–259, 2024.

[20] A. Fauzan, N. Suarna, I. Ali, and H. Susana, "Penerapan Algoritma K-Means Clustering Untuk Meningkatkan Model Pengelompokan Dan Kinerja Jaringan Wi-Fi Secara Optimal," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, 2025, doi: 10.23960/jitet.v13i2.6272.