

Klasifikasi Kualitas Redaman Kabel Fiber Optik Berdasarkan Terminasi Menggunakan Algoritma *K-Means*

Naufal Galfan Syah^{1*}
Arif Senja Fitriani²
Irwan Alnarus Kautsar³
Cindy Taurusta⁴

^{1,2,3,4}Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sidoarjo, Indonesia
¹galfansyahn@gmail.com, ²asfjim@umsida.ac.id, ³irwan@umsida.ac.id,
⁴cindytaurusta@umsida.ac.id

***Penulis Korespondensi:**
Naufal Galfan Syah
galfansyahn@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengklasifikasikan kualitas redaman pada jaringan fiber optik menggunakan algoritma *K-Means* dengan menggunakan 400 data pengukuran lapangan. Model klasifikasi dibangun untuk mengelompokkan data berdasarkan karakteristik parameter jaringan fiber optik yang diperoleh dari hasil pengukuran lapangan. Penelitian ini juga membandingkan performa model berdasarkan dua pendekatan, yaitu dengan menggunakan seluruh fitur dan dengan menggunakan fitur yang telah diseleksi. Evaluasi dilakukan menggunakan dua metrik utama, yaitu Sum of Squared Errors (SSE) dan Silhouette Coefficient, guna menilai kualitas hasil klusterisasi. Pengujian menunjukkan bahwa penggunaan fitur terpilih memberikan hasil klaster yang lebih baik, dengan nilai SSE sebesar 1935.5 dan Silhouette Coefficient sebesar 0.3529, dibandingkan dengan penggunaan seluruh fitur yang menghasilkan SSE sebesar 3130.93 dan Silhouette Coefficient sebesar 0.2616. Hasil menunjukkan bahwa model dengan seleksi fitur menghasilkan klaster yang lebih baik, dengan distribusi data yang seimbang (klaster 0: 197 data, klaster 1: 203 data). Hasil juga menyediakan metode klasifikasi yang dapat mengidentifikasi kondisi redaman tinggi dan rendah. Hasil ini menunjukkan bahwa pemilihan fitur yang tepat dapat memberikan peningkatan pada kualitas klaster. Secara keseluruhan, *K-Means* terbukti mampu mengelompokkan kualitas redaman secara efektif, sehingga dapat mendukung pemeliharaan dan pemantauan jaringan fiber optik secara lebih efisien.

Kata Kunci: Data Mining; Fiber Optik; Klasifikasi; *K-Means*; Kualitas Redaman.

Abstract

This study aims to classify the attenuation quality in fiber optic networks using the *K-Means* algorithm based on 400 field measurement data points. The classification model was built to group the data based on the characteristics of fiber optic network parameters obtained from field measurements. The study also compares the performance of two model approaches: one using all features and the other using selected features. Evaluation was conducted using two main metrics—Sum of Squared Errors (SSE) and Silhouette Coefficient—to assess the quality of the clustering results. The testing showed that using selected features yielded better clustering results, with an SSE of 1935.5 and a Silhouette Coefficient of 0.3529, compared to using all features, which resulted in an SSE of 3130.93 and a Silhouette Coefficient of 0.2616. The results indicate that the feature-selected model produced better clusters, with balanced data distribution (Cluster 0: 197 data points, Cluster 1: 203 data points). The outcomes also provide a classification method capable of identifying high and low attenuation conditions. These findings demonstrate that proper feature selection can enhance cluster quality. Overall, *K-Means* has proven effective in clustering attenuation quality, supporting more efficient maintenance and monitoring of fiber optic networks.

Keywords: Classification; Data Mining; Fiber Optic; *K-Means*; Signal Attenuation Quality.

1. Pendahuluan

Di era modern saat ini, permintaan akan layanan komunikasi data yang cepat, stabil, dan dapat diandalkan semakin meningkat, seiring dengan perkembangan teknologi yang pesat [1]. Salah satu solusi utama untuk memenuhi kebutuhan tersebut adalah penggunaan fiber optik, yang mentransmisikan informasi melalui cahaya, menggantikan sinyal elektrik pada kabel tembaga [2]. Teknologi ini menawarkan kecepatan tinggi dan kapasitas besar dengan tingkat redaman yang rendah, sehingga menjadi pilihan utama dibandingkan kabel tembaga tradisional [3].

Namun, meskipun memiliki banyak keunggulan, redaman yang tinggi pada fiber optik tetap menjadi tantangan serius. Redaman ini berkaitan dengan berbagai faktor, termasuk jenis dan panjang kabel, serta konfigurasi terminasi jaringan [4][5]. Kualitas redaman yang buruk dapat berdampak pada penurunan kecepatan dan stabilitas transmisi data, yang berujung pada kualitas jaringan yang kurang optimal [6]. Oleh karena itu, penting memastikan bahwa nilai redaman tidak melebihi ambang batas standar, yaitu 28 dB sesuai ITU-T G.984 [7].

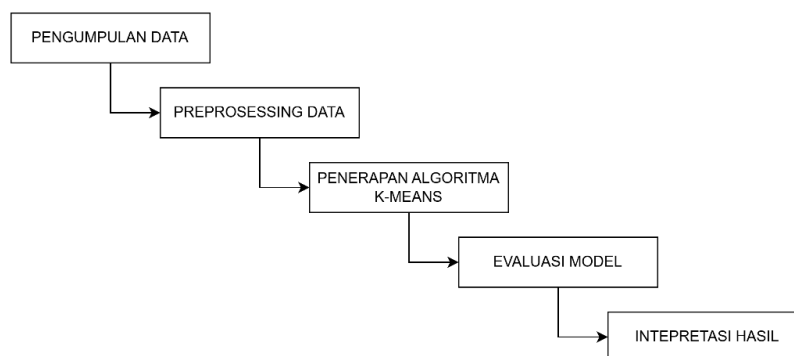
Penelitian terdahulu menganalisis kualitas jaringan fiber optik dengan menggunakan *metode link power budget*, di mana faktor teknis pada konfigurasi fiber optik terbukti memengaruhi nilai redaman secara signifikan [5]. Sementara itu penelitian lainnya menggunakan algoritma *Naive Bayes* untuk mengklasifikasikan data redaman fiber optik berdasarkan parameter-parameter teknis pada jaringan fiber optik [8]. Namun belum ada yang menerapkan algoritma *unsupervised* seperti *K-Means* untuk mengklasifikasikan kualitas redaman fiber optik.

Dalam konteks ini, klasifikasi kualitas redaman akan membantu dalam pengambilan keputusan terkait pemeliharaan jaringan serta pencegahan gangguan. *K-means* merupakan algoritma pengelompokan yang sangat dasar dan sering digunakan dalam berbagai aplikasi, baik di bidang akademis maupun industri [9]. *K-Means* dapat mengelompokkan data secara cepat, bahkan pada data dengan ukuran besar sehingga data menjadi lebih efisien [10]. Sehingga *K-Means* cocok untuk diterapkan pada jumlah data besar seperti data pengukuran lapangan dalam studi ini. *K-means* juga dapat mengklasifikasikan *dataset* menjadi kelompok-kelompok yang berbeda berdasarkan kesamaan titik data tanpa perlu menjalani proses pelatihan data terlebih dahulu [11][12]. Oleh karena itu, algoritma *K-Means* dipilih didasarkan pada kebutuhan untuk mengelompokkan data redaman fiber optik secara otomatis berdasarkan kemiripan karakteristik teknis tanpa memerlukan label awal. Selain itu, algoritma ini memiliki keunggulan dalam kecepatan komputasi dan ketahanannya terhadap outlier [13].

Penelitian ini bertujuan untuk mengklasifikasikan kualitas redaman kabel fiber optik menggunakan algoritma *K-Means*. Dengan pendekatan ini, diharapkan dapat diperoleh pengelompokan kualitas redaman yang lebih efisien dan akurat, sehingga membantu dalam pengambilan keputusan yang lebih tepat terkait pemeliharaan jaringan fiber optik dan pemilihan terminasi kabel yang optimal.

2. Metode Penelitian

Penelitian ini menggunakan algoritma *K-Means* untuk mengklasifikasikan kualitas redaman pada fiber optik. *K-Means* merupakan algoritma *machine learning unsupervised* yang digunakan untuk mengelompokkan data ke dalam beberapa klaster berdasarkan kesamaan atribut atau karakteristik [14]. *Machine learning unsupervised learning* adalah proses menemukan pola alami dalam data untuk menentukan label kelas tanpa menggunakan data pelatihan yang telah diberi label [15]. *K-Means* juga algoritma klasifikasi yang paling banyak digunakan dalam *machine learning* dan analisis data, karena kesederhanaannya, kemudahan implementasi, dan efisiensi komputasinya [16]. Tahapan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pengumpulan Data

Pengumpulan data dalam penelitian ini difokuskan pada pengukuran redaman kabel fiber optik. Data diperoleh dari hasil pengukuran lapangan, yang mencakup informasi seperti panjang kabel, konfigurasi *splitter* di *Optical Line Terminal* (OLT), *Optical Distribution Cabinet* (ODC), dan *Optical Distribution Point* (ODP), serta daya *transmit* dan *receive* dari OLT dan *Optical Network Terminal* (ONT). Data ini dapat berasal dari hasil pengukuran menggunakan *Optical Power Meter* (OPM) maupun dokumentasi sistem *monitoring* jaringan [17]. Proses pengumpulan data dilakukan dengan tujuan untuk memperoleh informasi yang akurat dan representatif mengenai kondisi nyata di lapangan. Dengan demikian, data yang diperoleh menjadi sangat relevan untuk proses klasifikasi kualitas redaman yang akan dilakukan. Data yang terkumpul kemudian diorganisir dalam format yang sesuai untuk mempermudah dan mempercepat proses analisis lebih lanjut. Hal ini penting untuk memastikan keakuratan hasil penelitian dan meminimalisir potensi kesalahan dalam pengolahan data.

Prapemrosesan Data

Setelah data terkumpul, langkah selanjutnya adalah melakukan proses pembersihan dan seleksi fitur. Prapemrosesan ini juga bertujuan untuk mengubah data yang masih bersifat kategorikal menjadi data numerik, karena komputasi *clustering* hanya dapat mengolah data numerik saja [18]. Langkah ini meliputi pemilihan atribut yang relevan, penanganan data kosong, dan normalisasi atribut numerik untuk mempersiapkan data ke tahap pemodelan. Pembersihan data bertujuan untuk meningkatkan kualitas data dengan imputasi nilai yang hilang dan penghapusan *outlier* [19]. Fitur-fitur yang tidak memberikan informasi yang berguna disebut fitur yang tidak relevan, sementara fitur-fitur yang tidak memberikan informasi lebih dari fitur yang saat ini telah dipilih disebut fitur yang redundan [20]. Seleksi fitur dalam penelitian ini dilakukan secara langsung dengan pendekatan *filter-based*, khususnya dengan mengeliminasi atribut yang memiliki variasi nilai sangat rendah. Karena menunjukkan nilai yang hampir seragam di seluruh dataset dan tidak memilih atribut-atribut yang secara teknis dianggap tidak memiliki pengaruh signifikan terhadap kualitas redaman jaringan fiber optik. Setelah itu data distandarisasi dengan menggunakan metode *StandardScaler*, yaitu dengan mengubah skala setiap fitur agar memiliki rata-rata (mean) sebesar 0 dan standar deviasi sebesar 1. Pendekatan ini bertujuan untuk menyamakan skala antar atribut numerik.

Penerapan Algoritma K-Means

Dilanjutkan dengan membuat model klasifikasi dengan algoritma *K-Means*. *K-means* adalah formulasi *clustering* yang paling populer di mana tujuannya adalah untuk memaksimalkan kesamaan yang diharapkan antara item data dan *centroid* klaster yang terkait [21]. Kesamaan suatu klaster terhadap anggotanya diukur berdasarkan kedekatan objek dengan nilai rata-rata

dalam kluster, yang dapat disebut sebagai *centroid* kluster [22]. *K-Means* mengidentifikasi pusat kluster (*centroid*) dan mengelompokkan data berdasarkan kedekatannya dengan *centroid* tersebut. Proses ini dilakukan secara iteratif hingga posisi *centroid* stabil, dan kluster yang terbentuk tidak lagi mengalami perubahan signifikan [23]. Pada proses klasifikasi ini menggunakan 2 model yaitu model pertama menggunakan *dataset* dengan atribut yang dipilih saja, kemudian untuk model kedua menggunakan *dataset* dengan semua atribut. Hal ini untuk membandingkan model manakah yang lebih baik dalam klasifikasi kualitas redaman fiber optik.

Evaluasi Model

Hasil klusterisasi dievaluasi menggunakan *Sum of Squared Errors* (SSE) dan *Silhouette Coefficient*. *Sum of Squared Errors* (SSE) yaitu mengukur total kesalahan jarak data ke *centroid* klusternya. Semakin kecil nilai SSE, semakin baik kluster yang terbentuk [24].

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} ||x - \mu_i||^2 \quad (1)$$

Keterangan (1) :

k : Jumlah kluster.

C_i : Kluster ke- i .

μ_i : *Centroid* kluster ke- i .

$||x - \mu_i||^2$: kuadrat jarak euclidean antara data x dan *centroid* kluster μ_i .

SSE digunakan untuk menghitung setiap k (kluster) dan ditampilkan menggunakan grafik SSE untuk mengidentifikasi titik *Elbow*. Jumlah kluster yang optimal dapat ditentukan berdasarkan titik *Elbow* pada grafik SSE, di mana laju penurunan SSE menjadi tajam, dan titik-titik setelahnya mengalami penurunan yang lambat dan stabil [25].

Silhouette Score digunakan untuk mengevaluasi kualitas klasifikasi dengan mengukur seberapa baik setiap data dalam suatu kelompok dibandingkan dengan kluster lain [26].

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

Keterangan (2) :

$a(i)$: Rata-rata jarak data i ke semua titik lain dalam klusternya sendiri (*intra-cluster distance*).

$b(i)$: Rata-rata jarak data i ke semua titik dalam kluster terdekat berikutnya (*inter-cluster distance*).

$S(i)$: Nilai *Silhouette Coefficient* untuk data i , berkisar antara -1 sampai dengan 1.

Silhouette Score berkisar dari -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan kluster yang lebih baik [27]. Jumlah kluster yang paling optimal ditentukan oleh nilai *Silhouette Coefficient* tertinggi [28]. Evaluasi ini penting untuk menentukan validitas dan keakuratan model klasifikasi.

Intepretasi Hasil

Tahap akhir dalam metode penelitian ini adalah melakukan interpretasi terhadap hasil klasifikasi yang diperoleh dari algoritma *K-Means*. Interpretasi dilakukan dengan menganalisis distribusi data pada masing-masing kluster yang terbentuk berdasarkan jumlah kluster optimal yang ditentukan melalui metode evaluasi seperti *Sum of Squared Errors* dan *Silhouette Coefficient*. Perbandingan hasil klasifikasi dilakukan antara *dataset* yang menggunakan semua atribut dengan *dataset* yang telah melalui proses seleksi fitur, untuk menilai model manakah yang terbaik dalam proses klasifikasi.

Hasil klasifikasi menunjukkan apakah model dengan atribut terpilih menghasilkan kluster yang lebih baik atau tidak, dengan ditunjukkan oleh nilai SSE yang lebih rendah dan *Silhouette Coefficient* yang lebih tinggi dibandingkan dengan model yang menggunakan seluruh atribut. Dengan interpretasi ini, penelitian dapat memberikan pemahaman yang lebih mendalam mengenai karakteristik data dalam setiap kluster serta mendukung proses pengambilan keputusan dan memberikan rekomendasi teknis dalam pengelolaan jaringan fiber optik, terutama dalam pemilihan konfigurasi terminasi yang dapat meminimalkan redaman dan menjaga performa jaringan secara optimal.

3. Hasil dan Pembahasan

Penelitian ini dilakukan untuk mengklasifikasikan kualitas redaman fiber optik menggunakan algoritma *K-Means* yang diimplementasikan menggunakan bahasa pemrograman *Python* pada platform *Google Colaboratory*. Data dikumpulkan ke dalam bentuk *dataset* dengan bantuan *Microsoft Excel* sebagai alat untuk pengumpulan dan pengolahan awal data. Atribut serta deskripsi pada *dataset* disajikan dalam Tabel 1, sedangkan contoh sampel data ditampilkan pada Tabel 2.

Tabel 1. Deskripsi Atribut

No. Atribut	Atribut	Deskripsi
<i>f1</i>	<i>fiber_length</i>	Panjang kabel fiber optik dalam jaringan, diukur dalam meter.
<i>f2</i>	<i>fiber_type</i>	Jenis kabel fiber optik yang digunakan
<i>f3</i>	<i>connector_type</i>	Jenis connector pada kabel fiber optik
<i>f4</i>	<i>olt_txpower</i>	Daya transmisi dari <i>Optical Line Terminal</i>
<i>f5</i>	<i>olt_rxpower</i>	Daya yang diterima di sisi OLT, diukur dalam dBm
<i>f6</i>	<i>olt_temperature</i>	Suhu perangkat OLT, diukur dalam derajat Celsius (°C)
<i>f7</i>	<i>olt_splitter</i>	Jenis <i>splitter</i> pasif di sisi <i>Optical Line Terminal</i> (OLT)
<i>f8</i>	<i>odc_splitter</i>	Jenis <i>splitter</i> pasif di sisi <i>Optical Distribution Cabinet</i> (ODC)
<i>f9</i>	<i>odp_splitter</i>	Jenis <i>splitter</i> pasif di sisi <i>Optical Distribution Point</i> (ODP)
<i>f10</i>	<i>ont_txpower</i>	Daya transmisi dari <i>Optical Network Terminal</i> (ONT)
<i>f11</i>	<i>ont_rxpower</i>	Daya yang diterima di sisi ONT
<i>f12</i>	<i>ont_temperature</i>	Suhu perangkat ONT
<i>f13</i>	<i>wavelength</i>	Panjang gelombang fiber optik
<i>f14</i>	<i>onu_link_status</i>	Status koneksi antara <i>Optical Network Unit</i> (ONU)/ONT dengan OLT
<i>f15</i>	<i>spec_status</i>	Standar kualitas redaman yang dipenuhi oleh kabel fiber optik
<i>f16</i>	<i>loss</i>	Selisih redaman dari daya transmisi OLT hingga daya terima ONT

Tabel 2. Sample Dataset

f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15	f16
1025	SM	SC	4.39	-31.26	37	1:2	1:4	1:8	1.96	-29.68	49	1490	ONLINE	UNSPEC	34.07
1064	SM	SC	4.11	-19.49	32	1:2	1:2	1:16	2.13	-21.38	46	1490	ONLINE	SPEC	25.49
1066	SM	SC	4.04	-23.20	38	1:2	1:2	1:8	2.22	-29.63	55	1490	ONLINE	UNSPEC	33.67
1089	SM	SC	4.27	-32.57	33	1:2	1:4	1:8	2.06	-30.64	39	1490	ONLINE	UNSPEC	34.91
1097	SM	SC	3.57	-32.38	34	1:2	1:2	1:8	2.33	-26.31	50	1490	ONLINE	UNSPEC	29.88

Dataset kemudian melalui tahap prapemrosesan data dan seleksi fitur untuk membentuk model pertama. Pada tahap ini, nilai-nilai atribut dikonversi ke dalam bentuk numerik, baik dalam format *ordinal numeric* maupun *categorical numeric*, serta dilakukan standarisasi data dengan menggunakan *StandardScaler*. Proses seleksi fitur pada model pertama dilakukan dengan memilih beberapa atribut yang dianggap paling relevan untuk proses klasifikasi, yaitu:

olt_txpower, *olt_rxpower*, *olt_splitter*, *odc_splitter*, *odp_splitter*, *ont_txpower*, *ont_rxpower*, *spec_status* dan *loss*. Sementara itu, pada model kedua, digunakan seluruh atribut yang tersedia dalam *dataset* tanpa seleksi fitur. Data hasil prapemrosesan untuk model pertama disajikan pada Tabel 3, sedangkan untuk model kedua ditampilkan pada Tabel 4.

Tabel 3. Hasil Prapemrosesan Dataset Model Pertama

f4	f5	f7	f8	f9	f10	f11	f14	f16
4.39	-31.26	1.96	-29.68	0	1	1	1.0	34.07
4.11	-19.49	2.13	-21.38	0	0	0	0.0	25.49
4.04	-23.20	2.22	-29.63	0	0	1	1.0	33.67
4.27	-32.57	2.06	-30.64	0	1	1	1.0	34.91
3.57	-32.38	2.33	-26.31	0	0	1	1.0	29.88

Tabel 4. Hasil Prapemrosesan Model Kedua

f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15	f16
1025	0	0	4.39	-31.26	37	0	1	1	1.96	-29.68	49	1490	0	1.0	34.07
1064	0	0	4.11	-19.49	32	0	0	0	2.13	-21.38	46	1490	0	0.0	25.49
1066	0	0	4.04	-23.20	38	0	0	1	2.22	-29.63	55	1490	0	1.0	33.67
1089	0	0	4.27	-32.57	33	0	1	1	2.06	-30.64	39	1490	0	1.0	34.91
1097	0	0	3.57	-32.38	34	0	0	1	2.33	-26.31	50	1490	0	1.0	29.88

Proses Klasifikasi *K-Means* menggunakan bahasa pemrograman *python* dengan menggunakan library yang tersedia yaitu *scikit-learn* [29]. *Scikit-learn* adalah pustaka (*library*) dalam bahasa pemrograman *Python* yang digunakan untuk penerapan algoritma *machine learning* serta berbagai tugas analisis data [30]. Klasifikasi ini menggunakan model pertama yang telah dilakukan seleksi fitur dan model kedua yang menggunakan semua fitur. Pada setiap model nantinya akan dicari pusat *centroid* dan dilakukan pengelompokan berdasarkan kedekatan dengan pusat *centroid*. Hasil pusat *centroid* dari model pertama disajikan pada Tabel 5, pada kolom f17 (*cluster_model1*) sebagai cluster dari model pertama, sedangkan hasil pusat *centroid* dari model kedua disajikan pada Tabel 6 dengan kolom f18 (*cluster_model2*) sebagai cluster dari model kedua.

Tabel 5. Centroid dari Model Pertama

f4	f5	f7	f8	f9	f10	f11	f14	f16	f17
4.031269	-29.209239	204.802	-28.706497	0.0	0.395939	0.548223	0.989848	32.737766	0
3.961281	-21.093498	199.133	-20.316946	0.0	0.477833	0.536946	0.221675	24.278227	1

Tabel 6. Centroid dari Model Kedua

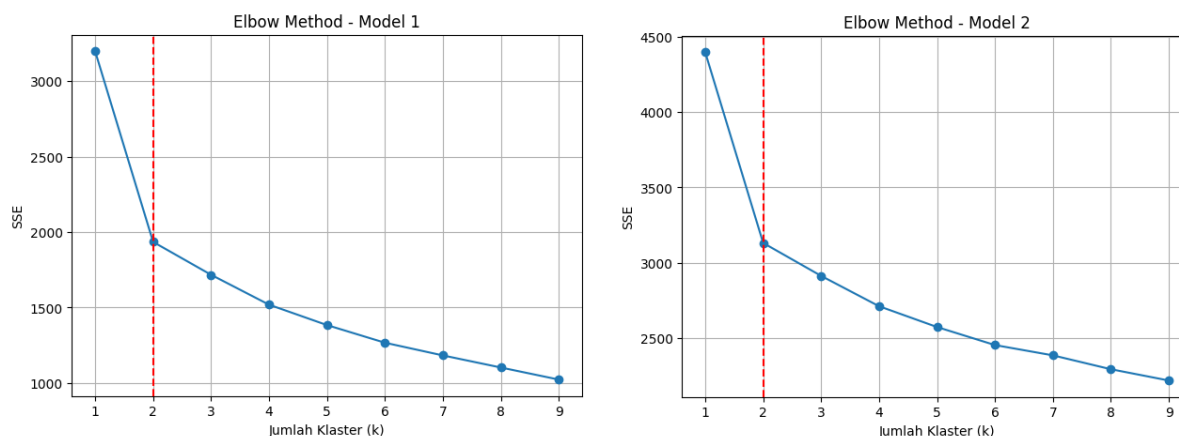
f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15	f16	f18
3.96673	0.0	0.0	4.029	-29.189	29.303	0.0	0.393	0.545	2.048	-28.675	45.318	149	0.0	0.989	32.705	0
7.374	0.0	0.0	899	798	030	0	939	455	737	707	182	0.0	0	899	606	0
3.95371	0.0	0.0	3.962	-21.072	29.044	0.0	0.480	0.539	1.990	-20.305	43.524	149	0.0	0.217	24.267	1
7.822	0.0	0.0	277	376	554	0	198	604	347	594	752	0.0	0	822	871	1

Hasil klasifikasi redaman fiber optik menggunakan algoritma *K-Means* pada model pertama dan model kedua, masing-masing dengan 400 data *instance*, menghasilkan 2 klaster pada setiap model. Pada model pertama, distribusi data menunjukkan bahwa klaster 0 terdiri atas 197 data, sedangkan klaster 1 terdiri atas 203 data. Sementara itu, pada model kedua, klaster 0 memuat 198 data, dan klaster 1 memuat 202 data. Hasil klasifikasi dari kedua model disajikan pada Tabel 7.

Tabel 7. Hasil Klasifikasi K-Means Model Pertama dan Model Kedua

f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15	f16	f17	F18
1.0 25	0	0	4.3 9	- 31. 26	37	0	1	1	1.9 6	- 29. 68	49	149 0	0	1.0	34. 07	0	0
1.0 64	0	0	4.1 1	- 19. 49	32	0	0	0	2.1 3	- 21. 38	46	149 0	0	0.0	25. 49	1	1
106 6	0	0	4.0 4	- 23. 20	38	0	0	1	2.2 2	- 29. 63	55	149 0	0	1.0	33. 67	0	0
108 9	0	0	4.2 7	- 32. 57	33	0	1	1	2.0 6	- 30. 64	39	149 0	0	1.0	34. 91	0	0
109 7	0	0	3.5 7	- 32. 38	34	0	0	1	2.3 3	- 26. 31	50	149 0	0	1.0	29. 88	0	0

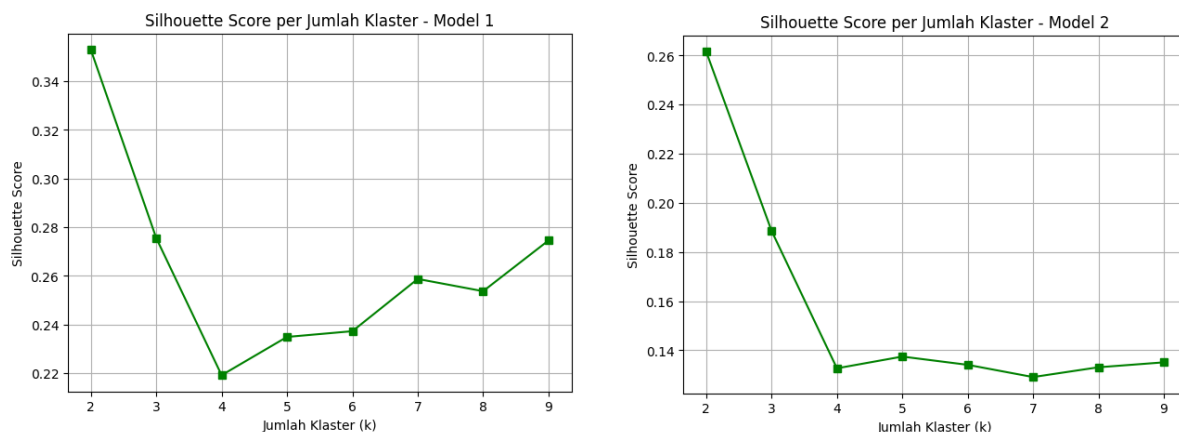
Hasil evaluasi menggunakan metode *Elbow Method* dengan perhitungan Sum of Squared Error (SSE) menghasilkan beberapa nilai SSE untuk setiap model. Pada model pertama, jumlah klaster optimal ditentukan sebanyak 2 klaster ($k=2$) dengan nilai SSE sebesar 1935.5. Sementara itu, pada model kedua, jumlah klaster optimal juga diperoleh pada $k=2$ dengan nilai SSE sebesar 3130.93. Grafik *Elbow Method* yang ditampilkan pada Gambar 2 menunjukkan adanya penurunan nilai SSE yang tajam hingga $k=2$, kemudian diikuti oleh penurunan yang lebih lambat dan landai. Pola ini mengindikasikan bahwa $k=2$ merupakan jumlah klaster yang paling optimal untuk kedua model.



Gambar 2. Grafik *Elbow Method* dengan nilai SSE Model Pertama dan Model Kedua

Evaluasi lebih lanjut dilakukan menggunakan metode *Silhouette Coefficient* dengan melakukan iterasi terhadap berbagai nilai klaster. Hasil evaluasi menunjukkan bahwa jumlah klaster paling optimal terdapat pada $k=2$ untuk kedua model. Pada model pertama, nilai *Silhouette Coefficient* terbaik tercapai pada $k=2$ dengan skor sebesar 0.3529, sedangkan pada model kedua nilai terbaiknya sebesar 0.2616. Meskipun kedua model memiliki jumlah klaster optimal yang sama, model pertama menunjukkan kualitas pemisahan klaster yang lebih baik berdasarkan nilai

Silhouette Coefficient yang lebih tinggi. Grafik perbandingan nilai *Silhouette Coefficient* antara model pertama dan model kedua ditampilkan pada Gambar 3.

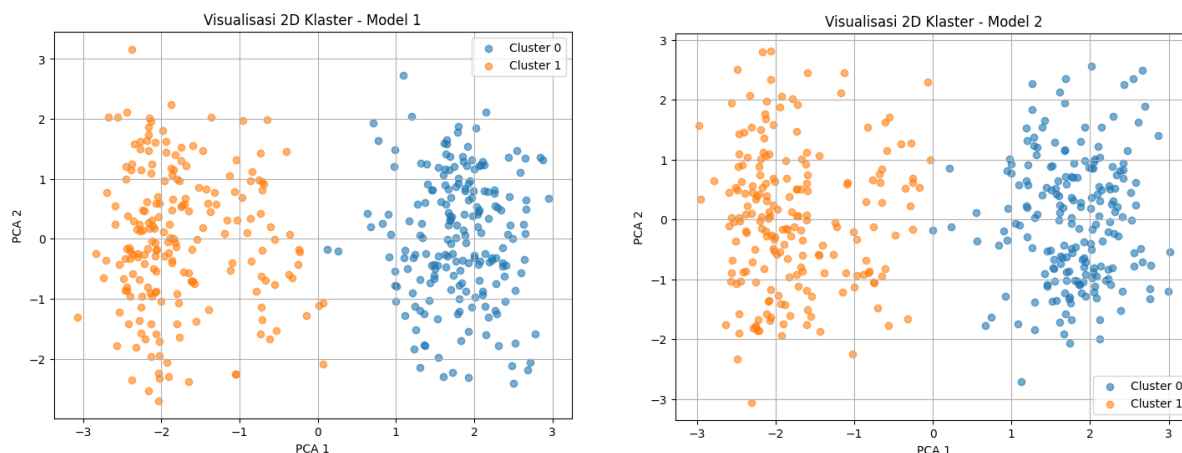


Gambar 3. Grafik *Silhouette Coefficient* Model Pertama dan Model Kedua

Berdasarkan hasil evaluasi klasifikasi, baik model pertama maupun model kedua menghasilkan jumlah kluster yang sama, yaitu dua kluster ($k=2$), sebagaimana ditentukan melalui metode *Elbow* dan *Silhouette Coefficient*. Evaluasi menggunakan metode *Elbow* menunjukkan bahwa model pertama memiliki nilai SSE lebih rendah (1935.5) dibandingkan nilai SSE pada model kedua (3130.93), yang mengindikasikan bahwa model pertama memiliki pembagian kluster yang lebih baik berdasarkan kedekatan antar data. Selain itu, model pertama juga menunjukkan performa yang lebih baik dalam hal pemisahan kluster berdasarkan nilai *Silhouette Coefficient* sebesar 0.3529, dibandingkan model kedua yang hanya mencapai 0.2616. Dengan demikian, model pertama dapat disimpulkan memiliki kualitas klasifikasi yang lebih baik dibandingkan model kedua baik dari sisi kompaksi kluster maupun pemisahan antar kluster.

Nilai *Silhouette Coefficient* bisa tetap rendah meskipun SSE menurun karena SSE mengukur kedekatan data dengan *centroid* klusternya, tanpa memperhatikan jarak antar kluster. Sementara itu, *Silhouette Coefficient* mempertimbangkan baik kedekatan dalam kluster maupun pemisahan antar kluster. Nilai *Silhouette Coefficient* yang rendah mengindikasikan adanya tumpang tindih antar kluster atau jarak antar kluster yang tidak terlalu signifikan, sehingga meskipun pembentukan kluster cukup padat, perbedaan karakteristik antar kelompok belum cukup tajam [31]. Pada penelitian Hartama dan Anjelita yang menunjukkan bahwa nilai SSE dapat menurun tetapi *Silhouette Coefficient* bisa bernilai rendah, menandakan bahwa data berada di antara dua kluster yang saling tumpang tindih [32]. Dalam penelitian lain oleh Mulyani dkk tentang pengelompokan jumlah sekolah negeri di Indonesia menggunakan metode *K-Means*, ditemukan bahwa meskipun SSE menurun, nilai *Silhouette Coefficient* tetap rendah. Hal ini menunjukkan bahwa kluster yang terbentuk tumpang tindih antar kluster [33].

Hasil distribusi kluster menunjukkan adanya keseimbangan yang cukup baik antara jumlah data pada masing-masing kluster, di mana model pertama menghasilkan kluster 0 sebanyak 197 data dan kluster 1 sebanyak 203 data, sedangkan model kedua menghasilkan kluster 0 sebanyak 198 data dan kluster 1 sebanyak 202 data. Meskipun kedua model membagi data ke dalam dua kluster dengan jumlah yang hampir seimbang, model pertama menunjukkan performa yang lebih unggul berdasarkan evaluasi kuantitatif serta pemilihan fitur yang lebih tepat guna. Visualisasi sebaran kluster pada masing-masing model disajikan pada Gambar 4, yang menggunakan metode *Principal Component Analysis* (PCA) untuk mereduksi dimensi data ke dalam dua komponen, sehingga distribusi dan pemisahan antar kluster dapat terlihat dengan lebih jelas.



Gambar 4. Grafik Visualisasi 2D Model Pertama dan Model Kedua

Interpretasi hasil klaster menunjukkan bahwa klaster 0 merepresentasikan kondisi dengan redaman tinggi yang berpotensi mengganggu kualitas layanan jaringan, sementara klaster 1 mencerminkan kondisi jaringan dengan redaman rendah atau normal. Dengan ini dapat dimanfaatkan untuk mendukung dalam mengambil keputusan dan memberikan rekomendasi dalam konfigurasi teknis mengenai fiber optik dalam pemeliharaan serta menjaga performa kualitas jaringan fiber optik.

4. Penutup

Penelitian ini berhasil menerapkan algoritma *K-Means* untuk mengklasifikasikan kualitas redaman pada jaringan fiber optik menggunakan dua pendekatan model, yakni model dengan seleksi fitur dan model tanpa seleksi fitur. Proses klasifikasi dilakukan terhadap 400 data *instance* dan menghasilkan dua klaster pada masing-masing model. Hasil evaluasi menunjukkan bahwa model dengan seleksi fitur memiliki performa yang lebih baik berdasarkan nilai SSE 1935.5 dan *Silhouette Coefficient* 0.3529, dibandingkan model tanpa seleksi fitur yang menghasilkan SSE sebesar 3130.93 dan *Silhouette Coefficient* sebesar 0.2616.

Model pertama menunjukkan pemisahan klaster yang lebih optimal dan kompak, mengindikasikan pentingnya pemilihan fitur yang tepat dalam meningkatkan kualitas klasifikasi. Klaster yang terbentuk dapat mengindikasikan kondisi redaman rendah (normal) dan redaman tinggi (berpotensi mengganggu layanan), sehingga hasil klasifikasi ini dapat dimanfaatkan dalam pengambilan keputusan teknis untuk menjaga performa jaringan fiber optik. Model ini juga berpotensi untuk diimplementasikan oleh penyedia layanan jaringan fiber optik sebagai sistem monitoring otomatis redaman, guna mendeteksi dini gangguan dan mengoptimalkan pemeliharaan jaringan secara proaktif.

Pada penelitian selanjutnya dapat dikembangkan dan mengeksplorasi dengan algoritma *clustering* lain seperti *DBSCAN* atau *Hierarchical Clustering* guna membandingkan efektivitas metode yang digunakan. Penambahan jumlah dan variasi data juga dapat meningkatkan kemampuan generalisasi model. Selain itu, penerapan teknik reduksi dimensi seperti *Principal Component Analysis* (PCA) dapat membantu dalam mengoptimalkan seleksi fitur.

5. Referensi

- [1] P. Yusman, D. U. Dari, D. Al Farabi, and R. Ardana, "ANALISIS MANAGEMENT DAN REDAMAN FIBER OPTIC PADA PT . JALUR NET INFOTEK," *BHAKTI NAGORI: Jurnal Pengabdian kepada Masyarakat*, vol. 4, pp. 31–37, 2024, doi: 10.36378/bhakti_nagori.v4i1.3886.
- [2] Muhammad Ridhwan and Lela Nurpulaela, "Analisis Penggunaan Jaringan Fiber Optik Di Area Kawasan Bijb Kertajati," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 14, pp. 467–479, 2023.
- [3] M. A. Rahmatulloh, D. Hanto, M. Yantidewi, Agitta Rianaris, and R.A. Firdaus, "Analisis Redaman Fiber Optik dengan Menggunakan Pemodelan Software Optisystem," *Jurnal Kolaboratif Sains*, vol. 6, no. 7, pp. 630–639, 2023, doi: 10.56338/jks.v6i7.3795.
- [4] M. Nurwahidah, U. A. Ahmad, R. E. Saputra, and P. Y. Pangestu, "Analisis Jarak Jangkauan Jaringan Fiber To The Home (FttH) dengan Teknologi Gigabit Passive Optical Network (Gpon) Berdasarkan Link Power Budget," *Prosiding Seminar Nasional Teknik Elektro dan Informatika*, vol. 8, no. September, pp. 203–207, 2021.
- [5] D. Setiawan, O. Yuliani, P. Studi Teknik Elektro, and F. Teknologi dan Industri ITNY, "Triple Play (3P) Di Kirana Garden Residence Yogyakarta," *Jmte*, vol. 04, no. 02, pp. 20–29, 2023.
- [6] I. M. Zukri, "Analisis Pengaruh Penggunaan Passive Splitter Pada Optical Distribution Point (Odp) Terhadap Kinerja Jaringan Di Rumah Pelanggan," *Jurnal Ilmiah Poli Rekayasa*, vol. 18, no. 1, p. 32, 2022, doi: 10.30630/jipr.18.1.249.
- [7] M. N. Ikhsanto and A. Setiawan, "Jaringan Akses Fiber To The Home (FTTH) Dengan Teknologi Gigabyte Passive Optical Network (GPON) PT. Telkom Kota Metro," *Journal Computer Science and Information Systems: J-Cosys*, vol. 4, no. 1, pp. 57–63, 2024, doi: 10.53514/jco.v4i1.497.
- [8] Suwandi, Y. Akbar, and D. I. Mulyana, "Optimasi Gigabit Passive Optical Network dan Algoritma Naïve Bayes dalam Analisis Jaringan FTTH," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 9, no. 1, pp. 186–192, Nov. 2024, doi: 10.35870/jtik.v9i1.3058.
- [9] H. Liu, J. Chen, J. Dy, and Y. Fu, "Transforming Complex Problems Into K-Means Solutions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 9149–9168, 2023, doi: 10.1109/TPAMI.2023.3237667.
- [10] S. M. Miraftabzadeh, C. G. Colombo, M. Longo, and F. Foiadelli, "K-Means and Alternative Clustering Methods in Modern Power Systems," *IEEE Access*, vol. 11, no. September, pp. 119596–119633, 2023, doi: 10.1109/ACCESS.2023.3327640.
- [11] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 20–28, 2023, doi: 10.58602/jaiti.v1i1.22.
- [12] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *Eurasip Journal on Wireless Communications and Networking*, vol. 2021, no. 1, 2021, doi: 10.1186/s13638-021-01910-w.
- [13] D. M. Putri, A. S. Ilmananda, and N. Prisanta, "Penggunaan Algoritma K-Means dan K-Medoids untuk Pengembangan Strategi Promosi Penerimaan Mahasiswa Baru," *SMATIKA JURNAL*, vol. 14, no. 02, pp. 388–398, Dec. 2024, doi: 10.32664/smatika.v14i02.1474.
- [14] R. Sharma and S. Arora, "Analysis of K-Means Clustering Algorithm," *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)*, vol. 11, no. 06, pp. 150–153, 2022, doi: DOI : 10.17577/IJERTV11IS060047.
- [15] R. Cohn and E. Holm, "Unsupervised Machine Learning Via Transfer Learning and k-Means Clustering to Classify Materials Image Data," *Integrating Materials and Manufacturing Innovation*, vol. 10, no. 2, pp. 231–244, 2021, doi: 10.1007/s40192-021-00205-8.

- [16] A. Pegado-Bardayo, A. Lorenzo-Espejo, J. Muñuzuri, and A. Escudero-Santana, "A review of unsupervised k-value selection techniques in clustering algorithms," *Journal of Industrial Engineering and Management*, vol. 17, no. 3, p. 641, 2024, doi: 10.3926/jiem.6791.
- [17] B. Adi Nugroho, A. Endang Jayati, and P. Muliandi, "Analisa Kualitas Jaringan Akses Indihome Berdasarkan Teknologi Msan Dan Gpon Di Sto Majapahit," *Teknik Elektro*, pp. 1–6, 2021.
- [18] R. D. Dana, D. Soilihudin, R. H. Silalahi, D. A. Kurnia, and U. Hayati, "Competency test clustering through the application of Principal Component Analysis (PCA) and the K-Means algorithm," *IOP Conference Series: Materials Science and Engineering*, vol. 1088, no. 1, p. 012038, 2021, doi: 10.1088/1757-899x/1088/1/012038.
- [19] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data," *Frontiers in Energy Research*, vol. 9, no. March, pp. 1–17, 2021, doi: 10.3389/fenrg.2021.652801.
- [20] S. M. Javidan, A. Banakar, K. A. Vakilian, and Y. Ampatzidis, "Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning," *Smart Agricultural Technology*, vol. 3, no. March 2022, 2023, doi: 10.1016/j.atech.2022.100081.
- [21] E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe, "Comprehensive Review of K-Means Clustering Algorithms," *International Journal of Advances in Scientific Research and Engineering*, vol. 07, no. 08, pp. 64–69, 2021, doi: 10.31695/ijasre.2021.34050.
- [22] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency," *Journal of Physics: Conference Series*, vol. 1751, no. 1, 2021, doi: 10.1088/1742-6596/1751/1/012038.
- [23] Z. Ning, J. Chen, J. Huang, U. J. Sabo, Z. Yuan, and Z. Dai, "WeDIV – An improved k-means clustering algorithm with a weighted distance and a novel internal validation index," *Egyptian Informatics Journal*, vol. 23, no. 4, pp. 133–144, 2022, doi: 10.1016/j.eij.2022.09.002.
- [24] L. P. Refialy, H. Maitimu, and M. S. Pesulima, "Perbaikan Kinerja Clustering K-Means pada Data Ekonomi Nelayan dengan Perhitungan Sum of Square Error (SSE) dan Optimasi nilai K cluster," *Techno.Com*, vol. 20, no. 2, pp. 321–329, 2021, doi: 10.33633/tc.v20i2.4572.
- [25] J. Wala, H. Herman, R. Umar, and S. Suwanti, "Heart Disease Clustering Modeling Using a Combination of the K-Means Clustering Algorithm and the Elbow Method," *Scientific Journal of Informatics*, vol. 11, no. 4, pp. 903–914, Dec. 2024, doi: 10.15294/sji.v11i4.14096.
- [26] W. B. Syamhuri, M. T. Furqon, and C. Dewi, "Pengelompokan Film Berdasarkan Alur Cerita menggunakan Metode Self Organizing Maps dan Silhouette Coefficient," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 12, pp. 5898–5904, 2022.
- [27] S. Suraya, M. Sholeh, and U. Lestari, "Evaluation of Data Clustering Accuracy using K-Means Algorithm," *International Journal of Multidisciplinary Approach Research and Science*, vol. 2, no. 01, pp. 385–396, 2023, doi: 10.59653/ijmars.v2i01.504.
- [28] A. S. Ritonga and I. Muhandhis, "Clustering Data Tweet E-Commerce Menggunakan Metode K-Means (Studi Kasus Akun Twitter Blibli Indonesia)," *Smatika Jurnal*, vol. 12, no. 01, pp. 75–84, 2022, doi: 10.32664/smatika.v12i01.665.
- [29] T. Seyam *et al.*, "Next-Generation K-Means Clustering: Mojo-Driven Performance for Big Data," *International Journal of Intelligent Information Systems*, vol. 14, no. 1, pp. 7–19, Mar. 2025, doi: 10.11648/j.ijis.20251401.12.
- [30] V. No, D. Prasetyawan, A. Mulyanto, and R. Gatra, "Edumatic : Jurnal Pendidikan Informatika Pemetaan Lintasan Karir Alumni Berdasarkan Analisis Cluster : Kombinasi K-Means dan Reduksi Dimensi Autoencoder," *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 1, pp. 198–207, 2025, doi: 10.29408/edumatic.v9i1.29713.

- [31] Muh. A. Amri, A. A. N. Risal, Muh. F. Bakri, and D. F. Surianto, "Mini-Batch K-Means Clustering Untuk Pengelompokan Kemandirian Daerah Di Sulawesi Selatan," *Jurnal Informatika Polinema*, vol. 11, no. 2, pp. 235–244, Feb. 2025, doi: 10.33795/jip.v11i2.6871.
- [32] D. Hartama and M. Anjelita, "Analysis of Silhouette Coefficient Evaluation with Euclidean Distance in the Clustering Method (Case Study: Number of Public Schools in Indonesia)," *Jurnal Mantik*, vol. 6, no. 3, pp. 3667–3677, 2022.
- [33] H. Mulyani, R. A. Setiawan, and H. Fathi, "Optimization of K Value in Clustering Using Silhouette Score (Case Study: Mall Customers Data)," *Journal of Information Technology and Its Utilization*, vol. 6, no. 2, pp. 45–50, 2023, doi: 10.56873/jitu.6.2.5243.