

## KLASTERISASI PELANGGAN BERDASARKAN PERILAKU KONSUMEN MENGGUNAKAN K-MEANS

Jefina Tri Kumalasari<sup>1\*</sup>,  
Agustiena Merdekawati<sup>2</sup>

Program Studi Sistem Informasi<sup>1,2</sup>  
Fakultas Teknologi Informasi<sup>1,2</sup>

Corresponden Email : [jefina.jtk@bsi.ac.id](mailto:jefina.jtk@bsi.ac.id)

Author Email : [jefina.jtk@bsi.ac.id](mailto:jefina.jtk@bsi.ac.id)<sup>1</sup>, [agustiena.atd@bsi.ac.id](mailto:agustiena.atd@bsi.ac.id)<sup>2</sup>

**Received:** October 5, 2025. **Revised:** October 28, 2025. **Accepted:** October 30, 2025. **Issue Period:** Vol.9 No.4 (2025), Pp: 1556-1572

**Abstrak:** Mengelola loyalitas pelanggan dan menjangkau pembeli merupakan tantangan terbesar yang dihadapi industri retail. Segmentasi atau pengelompokan pelanggan merupakan strategi yang dilakukan dengan memisahkan pelanggan ke dalam beberapa kelompok berdasarkan perbedaan karakteristik, perilaku, maupun kebutuhan mereka. Pembagian ini bertujuan membantu pebisnis guna memenuhi kebutuhan mereka dengan mengoptimalkan layanan dan produk. Metode Clustering digunakan untuk mengidentifikasi beberapa segmen pelanggan. Berdasarkan hasil pengujian maka terlihat pada jumlah cluster optimal adalah  $K = 3$  karena memiliki nilai tertinggi yaitu 0.33861316, hal ini didukung setelah dilakukan pengujian elbow method. Penerapan Algoritma K-Means dalam mengkluster perilaku pelanggan memberikan wawasan mengenai variasi pola pembelian di antara pelanggan. Hasil ini dapat digunakan untuk mengembangkan strategi bisnis yang lebih terarah dan efektif. Namun, kondisi ekonomi dan faktor gaya hidup turut memberi pengaruh besar sehingga pola pengeluaran dalam populasi sangat bervariasi.

Kata Kunci: Konsumen, Clustering, K-Means

**Abstract:** Managing customer loyalty and reaching buyers are the biggest challenges faced by the retail industry. Customer segmentation or grouping is a strategy that involves separating customers into several groups based on differences in their characteristics, behavior, and needs. This division aims to help businesses meet their needs by optimizing services and products. The Clustering method is used to identify several customer segments. Based on the test results, it can be seen that the optimal number of clusters is  $K = 3$  because it has the



DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

*highest value of 0.33861316, which is supported after conducting an elbow method test. The application of the K-Means Algorithm in clustering customer behavior provides insight into the variation in purchasing patterns among customers. These results can be used to develop more targeted and effective business strategies. However, economic conditions and lifestyle factors also have a significant influence, resulting in highly varied spending patterns within the population.*

**Keywords:** Consumers, Clustering, K-Means

## I. PENDAHULUAN

Pelanggan merupakan kunci sukses dalam bisnis retail. Pembagian atau pengelompokan pelanggan menjadi dasar dari analisa kebutuhan dari berbagai kebutuhan yang berbeda [1]. Pelanggan mempunyai kebutuhan yang berbeda sehingga memerlukan pelayanan yang berbeda pula. Pemahaman perilaku pelanggan dan strategi pemasaran akan secara langsung berhubungan dengan kesuksesan dan profitabilitas bisnis. Pengelompokan pelanggan berdasarkan karakteristik yang sama menjadi salah satu komponen penting dalam pemasaran.

Kemajuan teknologi memberikan dampak meningkatkan keberhasilan dalam pemasaran. Kini data transaksi harian dapat diolah menjadi pola pemecahan masalah bisnis, berbeda dengan dahulu ketika hanya dianggap sekadar catatan. Dengan memanfaatkan komputasi, data mining lalu mengekstrak data transaksi pelanggan dapat menjadi data berharga, perusahaan dapat memprediksi perilaku pelanggan, membuat suatu keputusan berbasis ilmu pengetahuan [2]. Setiap toko menjual produk atau layanan perlu untuk buat memahami sifat pelanggan yang nantinya akan mempengaruhi pada strategi pemasaran dan meningkatkan kepuasan pelanggan yang nantinya akan meningkatkan pendapatan suatu perusahaan, dengan analisis ini bertujuan untuk mengevaluasi karakteristik setiap pelanggan yang nantinya akan menyesuaikan produk dan layanan pada perusahaan. Selain itu, strategi pemasaran yang ada saat ini dianggap kurang efektif karena penawaran yang diberikan perusahaan kepada pelanggan masih bersifat umum. Perusahaan belum mampu menyajikan produk atau promosi yang sesuai dengan kebutuhan spesifik pelanggan.

Sebelum merancang strategi pemasaran, platform e-commerce perlu terlebih dahulu melakukan segmentasi pelanggan. Segmentasi ini menjadi fondasi utama bagi pelaksanaan pemasaran yang tepat sasaran. Di era big data, penggunaan machine learning sangat berperan dalam membantu platform memahami serta menganalisis perilaku konsumsi [1]. Segmentasi pelanggan merupakan proses membagi pelanggan ke dalam beberapa kelompok yang memiliki karakteristik serupa. Tujuan dari segmentasi ini adalah untuk menentukan cara terbaik dalam menangani setiap kategori pelanggan sehingga dapat meningkatkan keuntungan yang diperoleh bisnis dari masing-masing pelanggan [3]. Metode yang digunakan pada pengolahan data ini menggunakan metode K-Means dimana Pendekatan *Interpretable Machine Learning* (IML) menunjukkan efektivitas yang lebih tinggi dalam proses segmentasi pelanggan dibandingkan dengan metode konvensional. Selain itu, penerapan teknik ini memungkinkan identifikasi yang lebih akurat [4], *Interpretable machine learning* merupakan dasar yang digunakan untuk pengelompokan produk baru. Hal tersebut memberikan implikasi praktis bagi manajer dalam merancang strategi pengembangan produk yang lebih tepat sasaran, dengan tingkat efisiensi dan otomatisasi yang lebih baik dibandingkan pendekatan tradisional.

Pada penelitian sebelumnya [5] Algoritma K-Means Untuk Pengelompokan Perilaku Customer. Mengidentifikasi pola perilaku pembelian menggunakan atribut seperti frekuensi pembelian, jumlah pembelian, dan



jenis produk. Dengan mengelompokkan pelanggan ke dalam segmen homogen, perusahaan dapat merancang strategi pemasaran yang lebih efektif dan personalisasi yang lebih baik.

Analisis Clustering Topik Survey menggunakan Algoritme K-Means (Studi Kasus: Kudata)[6] Metode pada penelitian ini menggunakan scraping untuk mengumpulkan data instrumen survey, meliputi deskripsi, pertanyaan dan kombinasi keduanya pada setiap formulir di Google Forms. penelitian ini menggunakan text representation dengan metode TF IDF (Term Frequency-Inverse Document Frequency) dan reduksi dimensi menggunakan PCA (Principal Component Analysis), serta menggunakan Silhouette Score untuk menghasilkan cluster secara optimal dalam algoritme K-means. Sehingga, penelitian ini menghasilkan rekomendasi dan tren topik survey.

Penerapan K-Means Clustering untuk Segmentasi Konsumen E-Commerce Berdasarkan Pola Pembelian [7] Penelitian ini bertujuan untuk mengelompokkan konsumen e-commerce berdasarkan pola pembelian menggunakan metode K-Means Clustering. Segmentasi dilakukan dengan mempertimbangkan tiga variabel utama, yaitu frekuensi pembelian, nilai transaksi, dan waktu pembelian. Hasil penelitian menunjukkan bahwa konsumen dapat dikelompokkan ke dalam lima segmen yang memiliki karakteristik perilaku belanja yang berbeda-beda, mulai dari konsumen bernilai tinggi dan aktif, hingga konsumen pasif dengan nilai transaksi rendah.

## II. METODE DAN MATERI

Penelitian ini menggunakan data yang berasal dari kaggle dengan 2.240 baris dan 29 atribut. Analisis kepribadian pelanggan ini diharapkan memberikan pemahaman yang lebih mengenai kebutuhan konsumen, perilaku dan kekhawatiran pelanggan. Selain itu memberikan pandangan mengenai pemasaran yang baik. Daripada mempromosikan produk baru ke semua pelanggan lebih baik mengidentifikasi segmentasi pelanggan mana yang paling mungkin membeli. Sehingga menghasilkan strategi pemasaran lebih lanjut memungkinkan pengembangan produk baru berdasarkan tren dan preferensi pelanggan.

Dari data tersebut akan dibuat proses lebih lanjut dengan menggunakan platform Knime, dimana proses tersebut meliputi data preparation, modelling dengan menggunakan algoritma k-means clustering, dan visualisasi. Secara umum data mining memiliki 2 jenis task yaitu prediktif dan deskriptif. Prediktif itu data mining yang dilakukan untuk membentuk sebuah model pengetahuan yang digunakan untuk melakukan prediksi [8]. Algoritma clustering yang diterapkan, yaitu algoritma clustering, K-Means dengan melakukan pengelompokan pelanggan yang berbelanja, dimana data yang diperoleh berdasarkan karakteristik tertentu. Data yang memiliki ciri atau karakteristik sama dimasukkan ke dalam satu kelompok dan data yang memiliki karakteristik berbeda dimasukkan ke dalam kelompok lain. Metode ini memanfaatkan digunakan untuk menemukan pusat cluster (centroid) yang mengelompokkan data pelanggan sehingga setiap pelanggan berada di cluster dengan jarak terdekat ke centroid tersebut. Proses ini dilakukan berulang hingga tidak ada perubahan signifikan dalam posisi centroid [9]. Pada tahap pengujian menggunakan metode *silhouette score* bertujuan untuk mengukur seberapa baik data dalam setiap cluster dikelompokkan [8] Pada penelitian ini akan digunakan metode Elbow untuk menentukan jumlah cluster optimal.

## III. PEMBAHASAN DAN HASIL

Guna menghasilkan hasil maksimal, maka diperlukan beberapa tahap pemrosesan pengolahan terbagi menjadi beberapa tahap yaitu :

### 1. Data Understanding

Data yang digunakan meliputi data customer analisis, dimana data ini berisi informasi pelanggan yang akan digunakan untuk analisis dan segmentasi.



DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

## 2 Data Preparation

Data preparation atau persiapan data merupakan proses analisis data dengan membangun kumpulan data akhir (data yang akan diproses pada tahap pemodelan) dari data mentah. Pada tahap ini mencakup pemilihan tabel, values, dan atribut atribut data, termasuk proses pembersihan dan transformasi data untuk dijadikan masukan dalam tahap modelling (pemodelan).

a. Pengumpulan Data (*Data Collection*)

Pada tahap ini melakukan proses pengumpulan data untuk dianalisis. Jika diimplementasikan dengan platform knime maka menggunakan node CSV reader, contohnya sebagai berikut:

File Edit View Window Help										
Sheet1:2:40 - Worksheet:40										
	A	B	C	D	E	F	G	H	I	J
		Account ID	Open Date	Expiration Date	Account ID	Account Name	Expiration Date	Expiration Date	Expiration Date	Account ID
1	1	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
2	2	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
3	3	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
4	4	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
5	5	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
6	6	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
7	7	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
8	8	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
9	9	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
10	10	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
11	11	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
12	12	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
13	13	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
14	14	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
15	15	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
16	16	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
17	17	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
18	18	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
19	19	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
20	20	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
21	21	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
22	22	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
23	23	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000
24	24	1000000000	10/1/2000	10/1/2000	1000000000	1000000000	10/1/2000	10/1/2000	10/1/2000	1000000000

### Gambar 1. CSV Reader

Dapat dilihat pada gambar 1.1 merupakan contoh penggunaan dari node CSV Reader yang berguna untuk membaca data dari file csv. Dalam data tersebut terdapat 2.240 baris dan 29 kolom.

### 2.3 Pembersihan Data (*Data Cleaning*)

Pada tahap ini dilakukan proses pembersihan data sebagai berikut:

a. Mengatasi Nilai yang Hilang (*Missing Values*)

Pada dataset kami sebelumnya telah terlampir kolom atau tabel mana yang memiliki nilai yang hilang (*missing values*).

Name	Type	# Missing values
id	Number (int)	0
year_birth	Number (int)	0
education	String	0
MARRI_STATUS	String	0
income	Number (float)	24
software	Number (int)	0

### Gambar 2. Missing Values

Dapat dilihat pada gambar diatas terdapat 24 missing values pada table income. Selanjutnya akan diproses menggunakan node missing values dengan menghilangkan baris yang memiliki nilai yang hilang.

### b. Mengatasi Outliers

Konsumen yang memiliki kebiasaan yang berbeda akan menimbulkan outlier (Siti Monalisa). Pada dataset kami terdapat beberapa table atau kolom yang nilainya jauh berbeda dari sebagian besar data lainnya, hal ini akan



**DOI:** 10.52362/jisamar.v9i4.2124

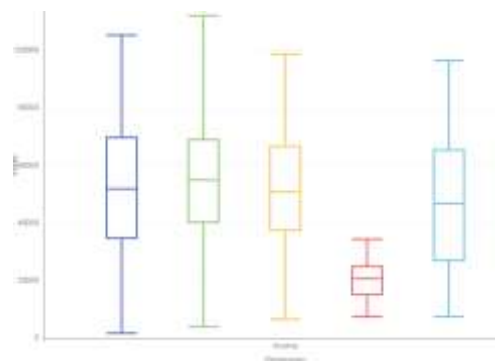
Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](#).

mempengaruhi model, terutama dalam algoritma yang sensitif terhadap outlier seperti k-means. Hal ini dapat divisualisasikan dengan node box plot sebagai berikut:



Gambar 3. Box Plot Income

Dapat dilihat pada gambar terdapat bulatan-bulatan kecil diatas box plot yang menunjukkan nilai tersebut tidak sesuai dengan nilai yang lainnya. Oleh sebab itu, kami akan memperbaikinya dengan node Numeric Outliers. Sehingga menghasilkan sebagai berikut:

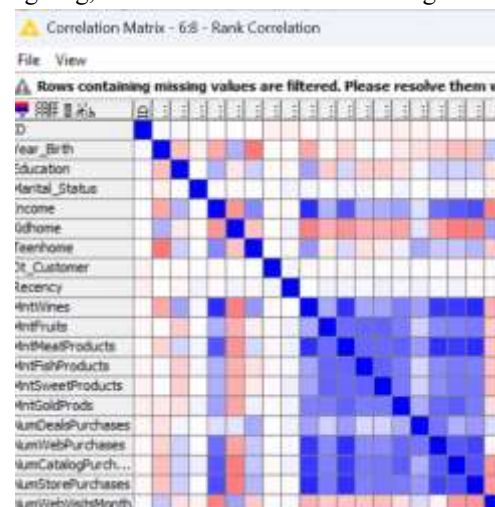


Gambar 4 Hasil Numeric Outliers

Dapat dilihat pada gambar terdapat dari node numeric outliers dengan memasukkan table agar tidak terdapat nilai diluar outliers. Dari hasil visualisasi menggunakan box plot tidak terdapat nilai diluar outliers. Untuk meningkatkan kualitas segmentasi, baris-baris yang ditandai sebagai outlier dapat dihapus dari dataset.[10]

### c. Analisa Korelasi

Analisa korelasi digunakan untuk memeriksa kekuatan hubungan linier antar variabel dalam data set. Koefisien korelasi positif mengindikasikan hubungan langsung, di mana kedua variabel meningkat atau menurun bersamaan [10]



DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

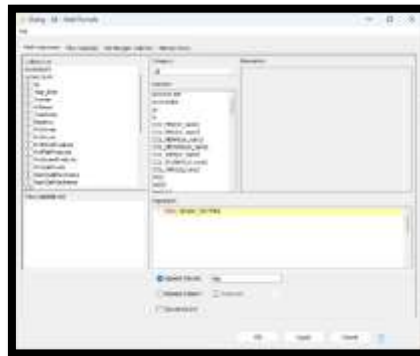
Hasil dari Correlation Matriks ini dapat dijadikan acuan variabel atau atribut mana yang tidak perlu digunakan pada proses selanjutnya.

#### d. Transformasi Data (*Data Transformation*)

Pada tahap transformasi data dilakukan proses mengubah, memodifikasi, atau mengonversi data dari satu format ke format lainnya dengan tujuan untuk modelling. Dalam tahap ini dilakukan proses sebagai berikut:

##### 1. Feature Engineering

Pada tahap ini data diproses dengan membuat kolom baru dari kombinasi kolom yang ada agar lebih jelas untuk digunakan dalam tahap modelling. Menggunakan node Math Formula dan node Rule Engine. Berikut salah satu yang data yang diolah dengan menggunakan metode Feature Engineering.

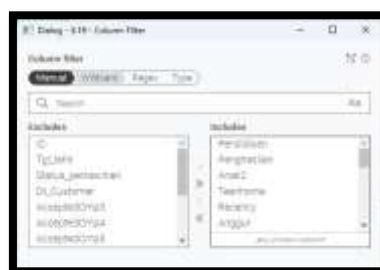


Gambar 6 Dialog Box Node Math

Dapat dilihat pada gambar diatas yang mana menggunakan node math formula untuk membuat kolom “Age” yang berasal dari rumus “2021-\$year\_birth\$” rumus tersebut akan menghasilkan usia dari customer.

##### 2. Memfilter Kolom

Pada tahap ini dilakukan proses membuang beberapa kolom yang tidak diperlukan dalam tahap modelling data menggunakan node Column Filter.



Gambar 7 Column Filter

Dapat dilihat pada gambar merupakan config dari penggunaan node Column Filter. Terdapat fitur “Excludes” dan “Includes”. Dalam fitur excludes memilih data yang akan dibuang dan pada fitur include berarti memilih kolom yang akan digunakan.





#### 4. Normalisasi (Normalizer)

Normalizer yang bertujuan untuk mengubah format numerik dalam dataset dalam rentang skala yang sama, misalnya antara 0 dan 1.



Gambar 8 Normalizer

Pada gambar diatas node Normalizer yang digunakan skala 0 dan 1 agar lebih mudah memodelling data.

#### e. Reduksi Data (Data Reduction)

Pada tahap reduksi data dilakukan pengurangan data yang digunakan dalam modelling data, sambil mempertahankan informasi yang penting. Proses ini dilakukan untuk mengurangi kompleksitas data, mempercepat pelatihan model, dan meningkatkan kinerja model, terutama ketika dataset memiliki dimensi yang sangat besar (banyak fitur atau kolom) atau ukuran yang sangat besar (banyak baris atau sampel). Pada reduksi data, menggunakan **Principal Component Analysis (PCA)**. PCA mengubah data ke dalam dimensi yang lebih rendah yang disebut principal components yang mengandung sebagian besar variansi data, membantu mengurangi jumlah fitur dengan mengurangi data ke dimensi yang lebih sedikit tanpa kehilangan banyak informasi penting.

PCA dimension 0 Number (double)	PCA dimension 1 Number (double)	PCA dimension 2 Number (double)
0.435	-0.198	0.22
-1.368	0.475	0.155
0.429	0.23	0.051
-0.295	-0.209	-0.738
-0.694	-0.459	0.302
-0.813	-0.384	-0.091
0.258	0.148	-0.057

Gambar 9. Hasil PCA

Pada gambar diatas merupakan config dan hasil dari penggunaan node PCA, Dimana kami menggunakan 3 Dimension PCA agar menjadi 3 fitur utama yang mencakup Sebagian besar data asli. Tujuannya agar mempermudah membaca pola atau klaster data lebih mudah.

#### 3. Modelling



DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

Modelling merupakan tahap pemilihan teknik pengembangan dengan menentukan algoritma yang akan digunakan. Pada tahap ini kami menggunakan algoritma K-Means dimana kami akan membuat beberapa cluster berdasarkan kesamaan. Pengujian dilakukan sebanyak enam kali dengan jumlah cluster uji mulai dari  $k=2$ ,  $k=3$ ,  $k=4$ ,  $k=5$ ,  $k=6$ ,  $k=7$ .

Yang mana  $k$  disini berarti sebagai:

1.  $k=2$  berarti terdapat 2 cluster, yaitu C0 dan C1.
2.  $k=3$  berarti terdapat 3 cluster, yaitu C0, C1, dan C2.
3.  $k=4$  berarti terdapat 4 cluster, yaitu C0, C1, C2, dan C3
4.  $k=5$  berarti terdapat 5 cluster, yaitu C0, C1, C2, C3, dan C4
5.  $k=6$  berarti terdapat 6 cluster, yaitu C0, C1, C2, C3, C4, dan C5
6.  $k=7$  berarti terdapat 7 cluster, yaitu C0, C1, C2, C3, C4, C5, dan C6

Setelah mendapat nilai  $k$  maka akan di implementasikan dalam platform KNIME menggunakan node k-means dan node *Silhouette Coefficient* untuk melihat nilai overall dari setiap pengujian cluster tersebut. Maka akan mendapat hasilnya sebagai berikut:



DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



Mean Silhouette Coefficient (Table)

Rows: 3 | Columns: 1

#	RowID	Mean Silhouette Coefficient Number (double)
1	cluster_1	0.38
2	cluster_0	0.164
3	Overall	0.317

Nilai Overall K=2

Mean Silhouette Coefficient (Table)

Rows: 4 | Columns: 1

#	RowID	Mean Silhouette Coefficient Number (double)
1	cluster_1	0.395
2	cluster_2	0.318
3	cluster_0	0.218
4	Overall	0.339

Nilai Overall k=3

Mean Silhouette Coefficient (Table)

Rows: 5 | Columns: 1

#	RowID	Mean Silhouette Coefficient Number (double)
1	cluster_0	0.179
2	cluster_3	0.274
3	cluster_2	0.366
4	cluster_1	0.206
5	Overall	0.255

Nilai Overall k=4

Mean Silhouette Coefficient (Table)

Rows: 6 | Columns: 1

#	RowID	Mean Silhouette Coefficient Number (double)
1	cluster_3	0.262
2	cluster_4	0.266
3	cluster_2	0.36
4	cluster_1	0.141
5	cluster_0	0.21
6	Overall	0.265

Nilai Overall k=5

Mean Silhouette Coefficient (Table)

Rows: 7 | Columns: 1

#	RowID	Mean Silhouette Coefficient Number (double)
1	cluster_5	0.262
2	cluster_4	0.253
3	cluster_2	0.368
4	cluster_0	0.05
5	cluster_1	0.129
6	cluster_3	0.236
7	Overall	0.251

Nilai Overall K = 6

Mean Silhouette Coefficient (Table)

Rows: 8 | Columns: 1

#	RowID	Mean Silhouette Coefficient Number (double)
1	cluster_0	0.237
2	cluster_5	0.242
3	cluster_6	0.295
4	cluster_1	0.05
5	cluster_2	0.124
6	cluster_3	0.179
7	cluster_4	0.223
8	Overall	0.203

Nilai Overall k=7

Gambar 10. Perbandingan Nilai overall Clustering

Dalam pengujian k=2, membuat 2 cluster, dari pengujian overallnya menggunakan node *silhouette coefficient* dan mendapat nilai *overall* 0.317. Dalam pengujian k=3 menghasilkan 3 cluster sebagai menghasilkan *silhouette coefficient* yaitu 0.339. dari pengujian k=4 menghasilkan overallnya menggunakan node *silhouette coefficient* dan mendapat nilai *overall* 0.255. Dalam pengujian k=5 menghasilkan *overall silhouette coefficient* bernilai 0.265. Dari pengujian k=6 nilai



DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

overallnya menggunakan node silhouette coefficient adalah 0.251. dan pengujian k=7 menghasilkan silhouette coefficient bernilai overall 0.203.

#### **4. Evaluation (Evaluasi)**

Pada tahap ini metode elbow digunakan untuk mengevaluasi hasil pengujian clustering, dimana proses ini akan menghasilkan berapa cluster yang tepat digunakan.

##### **a. Elbow Method**

Elbow method merupakan teknik yang digunakan untuk menentukan jumlah cluster yang optimal dalam klusterisasi, terutama pada algoritma K-Means. Tujuannya adalah untuk memilih jumlah cluster yang paling tepat dengan cara meminimalkan jumlah kesalahan dalam klusterisasi.

Cara membuat elbow method adalah dengan mengumpulkan hasil nilai overall dari 7 pengujian cluster dan dibuat visualisasinya dengan line chart.

Tabel 1. Hasil Pengujian Cluster

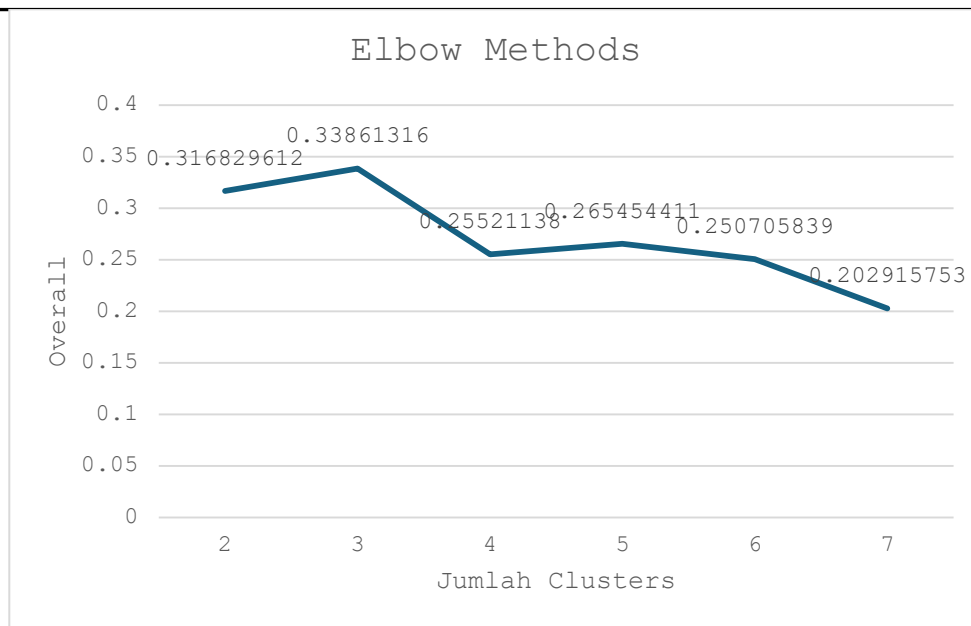
<b>Jumlah Clusters</b>	<b>Overall</b>
<b>2</b>	0.316829612
<b>3</b>	0.33861316
<b>4</b>	0.25521138
<b>5</b>	0.265454411
<b>6</b>	0.250705839
<b>7</b>	0.202915753

Berdasarkan hasil pengujian diatas maka terlihat pada jumlah cluster optimal adalah K = 3 karena memiliki nilai tertinggi yaitu 0.33861316.



DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



Gambar 11 Grafik Elbow Method

Pada grafik diatas merupakan hasil dari nilai overall pengujian 7 cluster, nilai yang membentuk siku. Jika dilihat dari grafik Berdasarkan grafik Elbow Method diatas Nilai *overall* meningkat dari 0.3168 (2 cluster) ke 0.3386 (3 cluster), setelah itu nilai *overall* menurun tajam mulai dari 4 cluster (0.2552) dan seterusnya terus turun hingga 7 cluster (0.2029). Titik “siku” (*elbow point*) terlihat jelas pada **jumlah cluster = 3, Setelah titik 3 nilai justru menurun**, menandakan penambahan cluster tidak meningkatkan kualitas pengelompokan. Jumlah cluster terbaik berdasarkan **Elbow Method** pada grafik tersebut adalah cluster **3**.

## 5. Deployment (Penerapan)

Deployment (penerapan) merupakan tahap akhir dari proses data mining. Pada proses ini menerapkan pengujian dari 4 cluster dengan visualisasi menggunakan scatter plot.



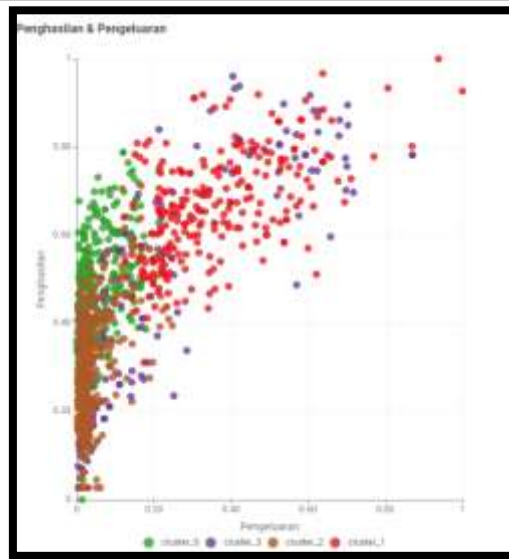
Gambar 12 Color & Scatter Plot

### 1. Visualisasi Pengeluaran & Penghasilan



DOI: 10.52362/jisamar.v9i4.2124

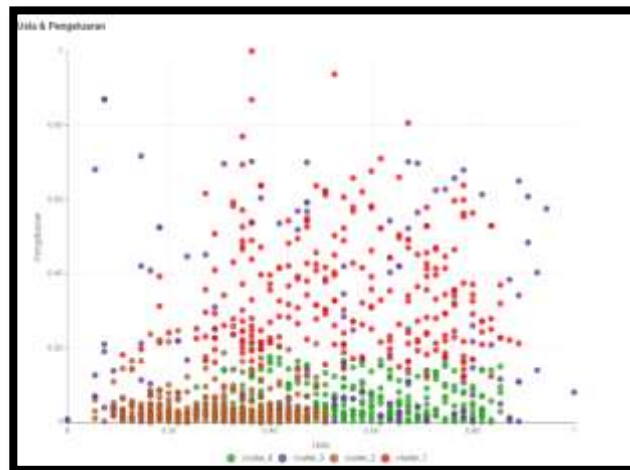
Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



Gambar 13. Penghasilan & Pengeluaran

Berdasarkan hasil pengelompokan (clustering) tersebut, dapat disimpulkan bahwa terdapat perbedaan yang jelas antara individu dalam hal tingkat penghasilan dan pengeluaran. Cluster 1 menunjukkan kelompok dengan kondisi ekonomi relatif tinggi, sedangkan Cluster 2 menggambarkan kelompok dengan kondisi ekonomi paling rendah. Cluster 0 berada sedikit di atas Cluster 2 dalam hal penghasilan dan pengeluaran, sementara Cluster 3 menunjukkan perilaku ekonomi yang adaptif, di mana pengeluaran mengikuti besarnya penghasilan. Secara keseluruhan, pola ini mencerminkan adanya variasi tingkat kesejahteraan dan perilaku finansial di antara individu dalam populasi yang dianalisis.

## 2. Visualisasi Pengeluaran & Usia



Gambar 14. Pengeluaran & Usia

Berdasarkan hasil pengelompokan tersebut, dapat disimpulkan bahwa usia berpengaruh terhadap pola pengeluaran individu. Individu berusia tua cenderung memiliki pengeluaran rendah hingga tinggi tergantung kondisi ekonominya (Cluster 0 dan 1), sedangkan individu berusia muda umumnya memiliki pengeluaran yang lebih rendah

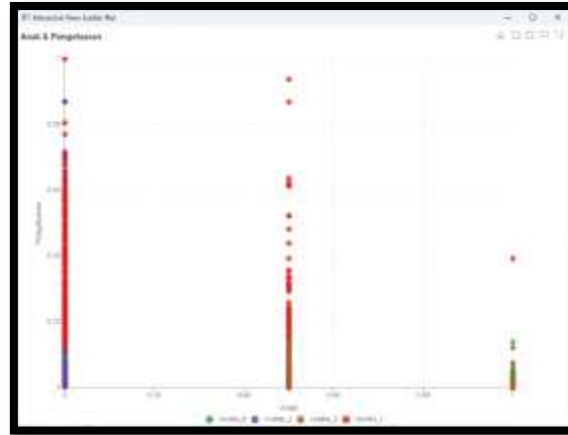


DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

(Cluster 2). Adapun Cluster 3 menunjukkan bahwa variasi usia, baik muda maupun tua, dapat disertai dengan tingkat pengeluaran yang beragam. Hal ini mengindikasikan bahwa selain faktor usia, terdapat faktor lain seperti pendapatan atau gaya hidup yang turut memengaruhi tingkat pengeluaran seseorang.

### 3. Visualisasi Pengeluaran & Anak



Gambar 15. Pengeluaran & Anak

Terdapat empat kelompok individu berdasarkan keberadaan anak di rumah dan tingkat pengeluaran. Cluster 0 (hijau) terdiri dari individu yang memiliki anak di rumah dengan pengeluaran rendah. Cluster 1 (merah) mencakup individu yang umumnya tidak memiliki anak kecil di rumah namun memiliki pengeluaran relatif tinggi. Cluster 2 (coklat) berisi individu yang pasti memiliki anak di rumah dengan pengeluaran rendah, sedangkan Cluster 3 (ungu) terdiri dari individu yang tidak memiliki anak di rumah dengan tingkat pengeluaran yang bervariasi.

### 4. Visualisasi Pengeluaran & Remaja



Gambar 16. Pengeluaran & Remaja

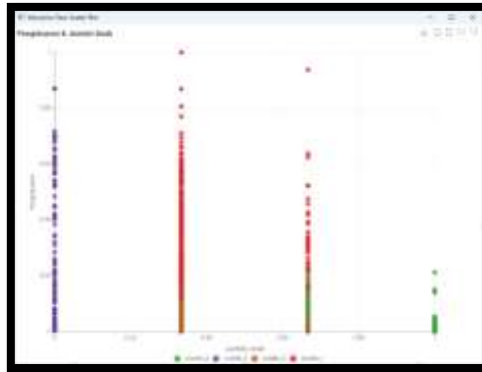
Berdasarkan hasil pengelompokan tersebut, dapat disimpulkan bahwa keberadaan remaja di rumah berpengaruh terhadap tingkat pengeluaran individu. Individu yang memiliki remaja di rumah umumnya menunjukkan pola pengeluaran yang cenderung rendah hingga tinggi tergantung kondisi ekonominya, seperti terlihat pada Cluster 0 dan Cluster 1. Sementara itu, individu yang tidak memiliki remaja di rumah memiliki pola pengeluaran yang lebih beragam (Cluster 3) atau cenderung rendah (Cluster 2). Hal ini menunjukkan bahwa keberadaan remaja dalam rumah tangga dapat memengaruhi besarnya pengeluaran keluarga.



DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

## 5. Visualisasi Pengeluaran & Jumlah Anak



Gambar 17. Pengeluaran & Jumlah Anak

Berdasarkan hasil pengelompokan tersebut, dapat disimpulkan bahwa jumlah anak dalam rumah tangga berpengaruh terhadap tingkat pengeluaran individu. Individu dengan jumlah anak lebih banyak cenderung memiliki pengeluaran yang rendah (Cluster 0), sedangkan mereka yang memiliki sedikit anak atau tidak memiliki anak menunjukkan variasi pengeluaran yang lebih luas, dari rendah hingga tinggi (Cluster 1, 2, dan 3). Hal ini menunjukkan bahwa tanggungan keluarga, khususnya jumlah anak, menjadi salah satu faktor yang memengaruhi besar kecilnya pengeluaran rumah tangga.

## 6. Visualisasi Pengeluaran & Orang Tua



Gambar 18. Pengeluaran & Orang Tua

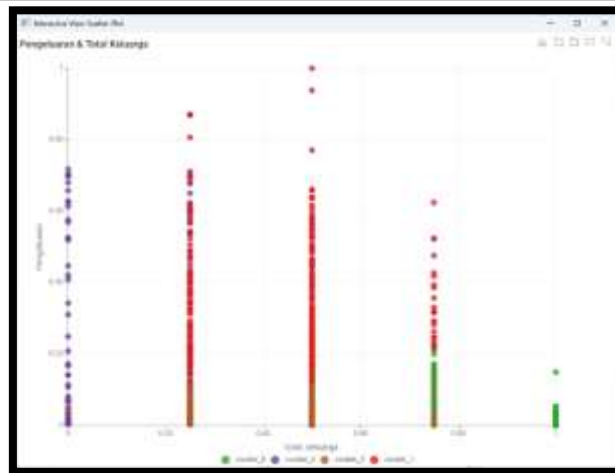
Berdasarkan hasil pengelompokan tersebut, dapat disimpulkan bahwa status sebagai orang tua berpengaruh terhadap pola pengeluaran individu. Kelompok yang merupakan orang tua cenderung memiliki pengeluaran yang lebih rendah, seperti terlihat pada Cluster 0 dan Cluster 2, sedangkan sebagian orang tua pada Cluster 1 memiliki pengeluaran relatif tinggi. Sementara itu, individu yang bukan orang tua (Cluster 3) menunjukkan pola pengeluaran yang lebih beragam, dari rendah hingga tinggi, yang mengindikasikan bahwa faktor tanggungan keluarga dapat memengaruhi besarnya pengeluaran seseorang.

## 7. Visualisasi Pengeluaran & Total Keluarga



DOI: 10.52362/jisamar.v9i4.2124

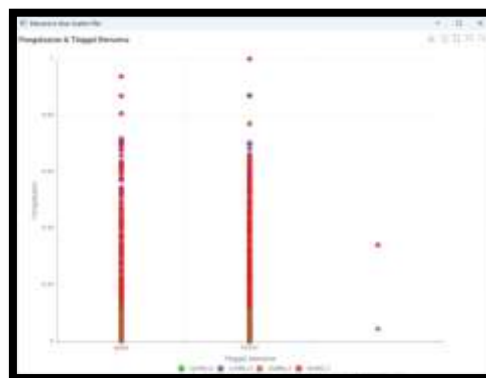
Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



Gambar 19. Pengeluaran & Total Keluarga

Berdasarkan hasil pengelompokan tersebut, dapat disimpulkan bahwa jumlah anggota keluarga berhubungan dengan tingkat pengeluaran individu. Individu dengan jumlah anggota keluarga lebih banyak cenderung memiliki pengeluaran yang rendah (Cluster 0 dan 2), sedangkan individu dengan anggota keluarga lebih sedikit menunjukkan pola pengeluaran yang lebih beragam, termasuk yang relatif tinggi (Cluster 1 dan 3). Hal ini mengindikasikan bahwa ukuran keluarga dapat memengaruhi besarnya pengeluaran, di mana keluarga kecil memiliki fleksibilitas pengeluaran yang lebih tinggi dibanding keluarga besar.

#### Visualisasi Pengeluaran & Tinggal Bersama



Gambar 20. Pengeluaran & Tinggal Bersama

Berdasarkan hasil pengelompokan tersebut, dapat disimpulkan bahwa status hubungan berpengaruh terhadap tingkat pengeluaran individu. Sebagian besar individu dalam semua kelompok merupakan mereka yang berpasangan, namun dengan tingkat pengeluaran yang berbeda-beda. Individu berpasangan umumnya memiliki pengeluaran rendah (Cluster 0 dan 2), sementara mereka yang berpasangan atau orang tua tunggal cenderung memiliki pengeluaran tinggi (Cluster 1). Adapun pada Cluster 3, individu berpasangan menunjukkan variasi pengeluaran dari rendah hingga tinggi. Hal ini menunjukkan bahwa meskipun status hubungan relatif sama, faktor lain seperti kondisi ekonomi dan tanggungan keluarga turut memengaruhi tingkat pengeluaran.

#### IV. KESIMPULAN



DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



Pada tahap ini merupakan kesimpulan dari hasil proses data mining, dimana proses ini menarik kesimpulan dari hasil visualisasi menggunakan scatter plot dan didapatkan hasilnya yaitu algoritma K-Means dalam mengkluster perilaku pelanggan memberikan wawasan mengenai variasi pola pembelian di antara pelanggan. Hasil ini dapat digunakan untuk mengembangkan strategi bisnis yang lebih terarah dan efektif, serta meningkatkan pengalaman pelanggan melalui personalisasi layanan variabel seperti usia, jumlah anak, keberadaan remaja, status orang tua, ukuran keluarga, dan status hubungan sangat memengaruhi pola pengeluaran individu. Namun, kondisi ekonomi dan faktor gaya hidup turut memberi pengaruh besar sehingga pola pengeluaran dalam populasi sangat bervariasi.

Setiap segmen pelanggan yang diidentifikasi menunjukkan perilaku menginformasikan strategi unik yang pemasaran dapat yang ditargetkan. Misalnya, pelanggan bernilai tinggi dapat diberi penghargaan dengan program loyalitas, sementara pelanggan yang tidak aktif dapat dilibatkan kembali melalui kampanye yang dipersonalisasi. Penelitian ini berkontribusi pada bidang analisis data dengan menunjukkan efektivitas algoritma pengelompokan dalam segmentasi pelanggan ritel online. Penggunaan Metode Elbow dan skor siluet memastikan keandalan hasil segmentasi, memungkinkan retailer untuk membuat keputusan berbasis data.

## REFERENASI

- [1] Z. Wu, L. Jin, J. Zhao, L. Jing, and L. Chen, "Research on Segmenting E-Commerce Customer through an Improved K-Medoids Clustering Algorithm," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/9930613.
- [2] Prof. Y. Jadhav and Dr. D. Parasar, "Customer Segmentation and Buyer Targeting Approach," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 9, no. 4, pp. 295–303, Nov. 2020, doi: 10.35940/ijrte.D5013.119420.
- [3] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability (Switzerland)*, vol. 14, no. 12, Jun. 2022, doi: 10.3390/su14127243.
- [4] J. Joung and H. Kim, "Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews," *Int J Inf Manage*, vol. 70, Jun. 2023, doi: 10.1016/j.ijinfomgt.2023.102641.
- [5] H. Mukhtar *et al.*, "ALGORITMA K-MEANS UNTUK PENGELOMPOKAN PERILAKU CUSTOMER," 2024.
- [6] M. A. Haq, W. Purnomo, and N. Y. Setiawan, "Analisis Clustering Topik Survey menggunakan Algoritme K-Means (Studi Kasus: Kudata)," 2023. [Online]. Available: <http://j->
- [7] B. Pola Pembelian Oleh, H. Safitri, S. Putri Lenggo Geni, F. Merry, and M. Wati, "Penerapan K-Means Clustering untuk Segmentasi Konsumen E-Commerce Penerapan K-Means Clustering untuk Segmentasi Konsumen E-Commerce Berdasarkan Pola Pembelian," *JUKI : Jurnal Komputer dan Informatika*, vol. 7, 2025.
- [8] A. Satriawan, R. Andreswari, and O. N. Pratiwi, "SEGMENTASI PELANGGAN TELKOMSEL MENGGUNAKAN METODE CLUSTERING DENGAN RFM MODEL DAN ALGORITMA K-MEANS



DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

---

TELKOMSEL CUSTOMER SEGMENTATION USING CLUSTERING METHOD WITH RFM MODEL AND K-MEANS ALGORITHM."

- [9] H. Mukhtar *et al.*, "ALGORITMA K-MEANS UNTUK PENGELOMPOKAN PERILAKU CUSTOMER," 2024.
- [10] A. Yusak, N. Rumapea, D. Pratiwi, and S. Sari, "Analisis Segmentasi Pelanggan Ritel Online Menggunakan K-Means Clustering Berdasarkan Model Recency, Frequency, Monetary (RFM)," *Jurnal Sains dan Teknologi*, vol. 6, no. 3, pp. 292–299, 2024, doi: 10.55338/saintek.v6i3.4607.



DOI: 10.52362/jisamar.v9i4.2124

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).