**Jurnal Kolaborasi**
**Sains dan Ilmu Terapan**
Volume 5 Nomor 1 Juli Desember 2026 | Page -9-18
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT*
Penerbit : Utiliti Project Solution

# Analisis Matematis pada Algoritma K-Means Clustering dalam Pengelompokkan Data

**Yulia Utami[1*], Desi Vinsensia[2], Sinta Suwanda[3], Muhammad Rizal Mohaimin[4]**
[1,3,4]Teknik Informatika, STMIK Pelita Nusantara, Medan, Indonesia
[3]Bisnis Digital, STMIK Pelita Nusantara, Medan, Indonesia
Email: [1]yuliautami14071990@gmail.com, [2]desivinsensia87@gmail.com, [3]sintasuwanda21@gmail,
[4]muhammadrizalmuhaimin@email.com

Email Penulis Korespondensi: [1]yuliautami14071990@gmil.com

**Info Artikel**

**Kata Kunci:**
Segmentasi Pelanggan
K-Means Clustering
Data Mining
Metode Elbow
Analisis Konsumen

**Abstrak**

Dunia bisnis ritel selalu dihadapkan pada satu pertanyaan mendasar: bagaimana memahami pelanggan secara lebih personal? Studi ini berupaya menjawab pertanyaan tersebut dengan memanfaatkan pendekatan *data mining*, tepatnya algoritma *K-Means Clustering*, untuk memetakan konsumen sebuah toko susu berdasarkan rekam jejak transaksi mereka. Dua variabel kunci dijadikan tolok ukur, yaitu intensitas berbelanja (frekuensi) dan total uang yang dibelanjakan (pengeluaran). Rangkaian analisis dimulai dari tahap pembersihan data, lalu berlanjut pada penentuan jumlah klaster paling pas menggunakan metode *Elbow*. Hasil evaluasi menunjukkan bahwa membagi pelanggan ke dalam empat segmen merupakan pilihan yang paling optimal. Proses komputasi yang dijalankan dengan Python berhasil mengelompokkan konsumen berdasarkan kemiripan karakteristik. Setiap klaster memiliki titik pusat (*centroid*) unik yang merepresentasikan tingkat aktivitas dan besaran pengeluaran masing-masing kelompok. Cluster 3 merepresentasikan pelanggan dengan frekuensi dan pengeluaran rendah, Cluster 1 menandai kelompok menengah, Cluster 2 menghimpun konsumen beraktivitas tinggi, dan Cluster 0 menempati posisi puncak sebagai segmen premium dengan nilai transaksi dan frekuensi tertinggi. Temuan ini menegaskan efektivitas *K-Means Clustering* dalam membedah perilaku belanja konsumen, sekaligus menyediakan fondasi kokoh bagi perancangan strategi pemasaran yang lebih presisi dan tepat sasaran.

**Keywords:**
Customer Segmentation
K-Means Clustering
Data Mining
Elbow Method
Consumer Analysis

**Abstract**

This study employs a data mining approach, specifically the K-Means Clustering algorithm, to segment customers of a milk store based on their transaction patterns. Two key variables purchase frequency and total customer spending serve as the basis for the analysis. The process begins with data preprocessing, followed by determining the optimal number of clusters using the Elbow method to identify the best K value. The results indicate that four clusters represent the optimal configuration. The implementation of K-Means using Python produces customer groupings based on shared characteristics, where each cluster is defined by a distinct centroid that reflects differing levels of purchasing activity and expenditure. Cluster 3 represents customers with low frequency and low spending, Cluster 1 describes those with moderate activity, Cluster 2 comprises customers with high activity levels, while Cluster 0 captures the premium segment with the highest transaction values and frequency. These findings demonstrate that K-Means Clustering is effective in grouping customers based on purchasing behavior and can serve as a robust foundation for designing more targeted marketing strategies.




Submitted: **17/06/2026**; Accepted: **29/06/2026**; Published:**08/07/2026**

**Jurnal Kolaborasi**
**Sains dan Ilmu Terapan**
Volume 5 Nomor 1 Juli Desember 2026 | Page -9-18
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT*
Penerbit : Utiliti Project Solution

# 1. PENDAHULUAN

Kemajuan ilmu data dan *machine learning* membawa kebutuhan yang semakin tinggi terhadap teknik pengelompokan (*clustering*) untuk mengekstraksi pola dari dataset tanpa label. Di antara sekian banyak pilihan yang tersedia, *K-Means Clustering* menonjol sebagai algoritma yang paling banyak digunakan. Metode ini beroperasi dengan membagi data ke dalam sejumlah k kelompok berdasarkan kedekatan setiap titik terhadap pusat klaster (*centroid*). Prosesnya berlangsung secara iteratif dengan sasaran memperkecil fungsi objektif yang dihitung dari total jarak kuadrat antara data dan centroid. Secara matematis, tujuannya adalah menemukan susunan klaster yang sekompak mungkin, di mana anggota di dalamnya seragam namun terpisah jelas dari kelompok lain. Beberapa penelitian relevan antara lain: Perbaikan inisialisasi dan stabilitas: Salah satu fokus riset terbaru adalah mengatasi sensitivitas terhadap inisialisasi centroid. Inisialisasi yang buruk sering membuat K-Means terjebak pada *local minimum* atau menghasilkan cluster yang tidak stabil. Berbagai pendekatan baru memperbaiki tahap inisialisasi ini, baik secara heuristik maupun sistematis, untuk mempercepat konvergensi dan meningkatkan kualitas cluster. Misalnya, varian baru yang mengoptimalkan titik awal memberikan hasil yang lebih konsisten dibandingkan metode acak standar, sehingga mengurangi jumlah iterasi yang diperlukan dan meningkatkan akurasi clustering.[1]. Penelitian modern banyak menggabungkan K-Means dengan teknik lain — seperti dalam *collaborative filtering* atau pembelajaran fitur — untuk memperluas kemampuan clustering di ruang masalah yang lebih kompleks. Contohnya, algoritma berbasis K-Means yang ditingkatkan digunakan dalam rekomendasi berbasis rating pengguna dengan memasukkan bobot varians fitur sebagai bagian dari perhitungan kemiripan, yang meningkatkan akurasi prediksi secara signifikan dibandingkan metode klasik.[2]. Pada dasarnya, *clustering* adalah salah satu teknik dalam ranah *data mining* yang berfungsi menyatukan data-data berciri serupa ke dalam kelompok yang sama. Mengadopsi teknik *K-Means Clustering*, studi ini menyasar data kunjungan wisatawan mancanegara ke Indonesia sepanjang tahun 2024. Kerangka kerja *Knowledge Discovery in Database* (KDD) menjadi panduan metodologisnya, yang mencakup lima tahapan berurutan: seleksi data, pra-pemrosesan, transformasi, penggalian data, dan evaluasi. Proses klasterisasi dijalankan menggunakan platform RapidMiner. Untuk menemukan jumlah klaster yang paling pas, peneliti menguji serangkaian skenario, mulai dari k = 2 hingga k = 7. Hasil pengujian menunjukkan bahwa konfigurasi tiga klaster (k = 3) merupakan yang terbaik, ditandai dengan perolehan nilai *Davies-Bouldin Index* (DBI) paling rendah. Melalui pendekatan ini, wisatawan berhasil dipetakan ke dalam tiga segmen: kelompok dengan intensitas kunjungan rendah, sedang, dan tinggi. Temuan semacam ini menjadi masukan berharga bagi para perumus kebijakan dalam menyusun strategi pengembangan pariwisata yang lebih terarah, sekaligus mendukung perencanaan kapasitas dan pemasaran musiman demi memaksimalkan potensi sektor pariwisata Indonesia. Secara praktis, K-Means telah diterapkan dalam berbagai domain seperti pengelompokan nilai mahasiswa, keamanan siber, manajemen koleksi perpustakaan, dan distribusi tenaga kerja kesehatan, yang menunjukkan fleksibilitas dan utilitasnya dalam menyederhanakan struktur data [4][5][6][7]. Selain itu, pemilihan jarak dan metode validasi cluster menjadi faktor penting dalam menentukan optimalitas hasil pengelompokan [8]. Berbagai studi sebelumnya telah memanfaatkan algoritma K-Means secara luas, baik untuk memetakan segmen pelanggan, menelaah performa penjualan, maupun mengklasifikasikan portofolio produk. Namun, sebagian besar penelitian tersebut dilakukan pada perusahaan ritel berskala besar atau menggunakan karakteristik data yang berbeda dengan kondisi di Toko Susu Asia. Toko Susu Asia memiliki karakteristik data yang unik, yaitu transaksi penjualan yang didominasi oleh produk susu dengan berbagai merek, jenis, ukuran kemasan, dan tingkat permintaan yang berfluktuasi sesuai kebutuhan konsumen. Selain itu, data penjualan yang dimiliki relatif sederhana, terdiri atas atribut-atribut numerik seperti jumlah penjualan, frekuensi transaksi, dan nilai penjualan, sehingga sesuai untuk diterapkan metode K-Means yang bekerja efektif pada data numerik. Metode Elbow dipilih karena sesuai dengan karakteristik data penjualan Toko Susu Asia yang bersifat numerik dan memiliki jumlah atribut yang relatif terbatas. Selain mudah diimplementasikan dan diinterpretasikan, metode ini juga banyak digunakan sebagai tahap awal dalam menentukan nilai K pada algoritma K-Means. Secara visual, metode *Elbow* ditampilkan melalui grafik agar titik siku sebagai penanda jumlah klaster ideal yang dapat teramati dengan jelas. Prinsipnya adalah mencari nilai k terkecil yang masih mampu mempertahankan nilai *Within Sum of Squares* (WSS) pada level yang rendah[9]. Berlandaskan pendekatan inilah, studi ini diarahkan untuk memetakan kelompok-kelompok produk. Hasil pemetaan tersebut diharapkan dapat menjadi fondasi dalam pengambilan keputusan strategis, mencakup manajemen persediaan, perancangan taktik pemasaran, hingga perencanaan stok barang.

# 2. METODOLOGI PENELITIAN

### 2.1 Pendekatan Penelitian

Secara metodologis, penelitian ini berdiri di atas fondasi kuantitatif deskriptif. Fokus utamanya bukan sekadar menjalankan algoritma, melainkan membedah sisi matematis dari *K-Means Clustering* yang memahami bagaimana rumus dan iterasi bekerja untuk menghasilkan pengelompokan data yang bermakna.Pendekatan ini menggabungkan perhitungan


Submitted: **17/06/2026**; Accepted: **29/06/2026**; Published:**08/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli Desember 2026 | Page -9-18
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT* Penerbit : Utiliti Project Solution

manual untuk memahami mekanisme iteratif algoritma, serta *implementasi Python* untuk memproses dataset nyata dan memvisualisasikan hasil clustering. Analisis ini bertujuan untuk memahami fungsi objektif, centroid, dan jarak Euclidean dalam pembentukan cluster. Objek penelitian dilakukan pada toko Susu Asia yang beralamat dijalan besar tembung**.**

**2.2 Analisis Data**

Beberapa penelitian dalam konteks Indonesia menunjukkan penerapan K-Means pada berbagai domain pengelompokan data. Misalnya, penelitian Yulianto dkk. mengimplementasikan K-Means dengan perangkat seperti RapidMiner untuk mengelompokkan kunjungan wisatawan asing berdasarkan data kedatangan di Provinsi Jawa Timur, sehingga dapat mengidentifikasi pola kunjungan dan cluster wilayah yang memiliki karakteristik berbeda dalam aktivitas wisatawan. Proses ini mencakup tahapan pra-pemrosesan, penentuan jumlah cluster optimal, serta evaluasi hasil menggunakan metrik indeks seperti Davies-Bouldin Index [10]. Di bidang sosial ekonomi, K-Means digunakan untuk memetakan kemiskinan di tingkat kabupaten/kota di Indonesia berdasarkan indikator sosial ekonomi, seperti angka kemiskinan, pendidikan, dan infrastruktur. Analisis ini membantu mengelompokkan wilayah prioritas untuk perumusan kebijakan yang lebih efektif [11]. Secara matematis fungsi objektif K-Means dapat dituliskan sebagai berikut:

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

Keterangan
$C_i$ = sebagai cluster ke- i
$x$ = sebagai titik data
$\mu_i$ = sebagai centroid cluster ke-i

Secara umum, literatur Indonesia menekankan bahwa pengelompokan data dengan K-Means memerlukan perhatian pada tahapan pra-proses seperti normalisasi dan penentuan jumlah cluster yang optimal (*k*), serta penggunaan indeks validasi cluster untuk memastikan bahwa hasil clustering mencerminkan struktur data secara statistik. Temuan-temuan ini memperkuat pemahaman bahwa pengelompokan data bukan hanya soal aplikasi algoritma, tetapi juga tentang pemilihan parameter dan metode evaluasi yang sesuai agar cluster yang terbentuk dapat memberikan informasi yang bermakna bagi pengguna. Langkah – langkah analisis data sebagai berikut:

1. Jumlah cluster menggunakan metode Elbow. Untuk memudahkan pengamatan, metode *Elbow* umumnya ditampilkan dalam wujud grafik. Di sanalah siku jadi titik belok yang menjadi kunci penentuan jumlah klaster yang dapat terlihat secara kasatmata. Ide dasarnya adalah memilih angka k terkecil yang masih sanggup menjaga nilai *Within Sum of Squares* (WSS) tetap rendah [9]. Ketika siku sudah terlampaui, setiap tambahan klaster baru hanya memberi perbaikan yang semakin tipis dan tidak lagi berarti. Dengan demikian, nilai k yang persis berada di titik siku itulah yang dianggap paling optimal. Dalam terminologi *K-Means*, dikenal pula istilah *Within-Cluster Sum of Squares* (WCSS). Konsep ini sejatinya identik dengan SSE, yakni mengukur total jarak kuadrat antara setiap titik data terhadap centroid klaster masing-masing. Karakteristiknya khas: semakin banyak klaster K yang dibentuk, semakin kecil nilai SSE/WCSS yang dihasilkan. Namun, laju penurunannya tidak seragam. Pada mulanya, penurunan terjadi secara drastis, lalu melambat dan membentuk lekukan menyerupai siku. Di titik perubahan laju inilah kita dapat menyimpulkan bahwa penambahan klaster tidak lagi memberikan kontribusi yang signifikan. SSE untuk suatu nilai k dirumuskan melalui Persamaan 2. Metode *Elbow* bekerja dengan menghitung nilai SSE ini untuk tiap kandidat k yang diujikan[12].

$$SSE\ (K) = \sum_{K=1}^{K} \sum_{X_i \epsilon C_k} |x_i - \mu_k|^2 \qquad (2)$$

   Keterangan
   K = *cluster* ke- *c*
   $x_i$ = jarak data obyek ke – *i*
   $\mu_k$ = pusat *cluster* ke- *i*
2. Inisialisasi centroid awal.
3. Hitung Hitung jarak Euclidean dari seluruh titik data terhadap masing-masing centroid
   Ketika *K-Means* perlu menentukan seberapa mirip dua titik data, jarak Euclidean adalah penggaris yang paling sering digunakan. Rumusnya cukup elegan: akar dari jumlah kuadrat selisih antara dua titik pada setiap dimensi fitur yang diukur.Penggunaan jarak Euclidean sangat populer dalam konteks K-Means karena memberikan ukuran intuitif


Submitted: **17/06/2026**; Accepted: **29/06/2026**; Published:**08/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli Desember 2026 | Page -9-18
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT* Penerbit : Utiliti Project Solution

terhadap kedekatan spasial antara objek data dan titik pusat cluster. Dalam praktik, penelitian penerapan K-Means pada data spasial fasilitas kesehatan di Kota Semarang memanfaatkan konstruksi jarak untuk menganalisis hubungan antara lokasi fasilitas dan konsentrasi peserta sehingga cluster dengan aksesibilitas yang serupa dapat diidentifikasi [13]. Pengukuran jarak antar titik data dan centroid dilakukan menggunakan jarak Euclidean, yang merupakan metrik yang paling umum dipakai dalam K-Means. Rumus jarak Euclidean antaratitik data $x_1, x_2, \dots, x_m$ dan centroid $\mu = \mu_1, \mu_2, \dots, \mu_m$ pada ruang berdimensi $m$ adalah:

$$d(x, \mu) = \sqrt{\sum_{k=1}^{m} (x_k - \mu_k)^2} \quad (3)$$

Jarak ini digunakan untuk menentukan titik data masuk ke cluster mana pada setiap iterasi, yaitu data akan ditempatkan pada cluster dengan centroid terdekat berdasarkan nilai Euclidean terkecil. Dengan cara ini, jarak Euclidean menjadi dasar matematis untuk evaluasi kedekatan dan kompaknya cluster [14].

4. Tentukan cluster berdasarkan jarak minimum.
5. Hitung ulang centroid berdasarkan titik data baru dalam cluster
   Jika sebuah cluster diibaratkan sebagai kumpulan titik-titik yang berkerumun, maka centroid adalah jantungnya. Secara matematis, ia tak lain adalah titik pusat yang posisinya ditentukan oleh rata-rata dari seluruh titik data yang menghuni cluster tersebut. *Centroid* μ berfungsi sebagai acuan pusat cluster dan diperbarui secara iteratif berdasarkan rata-rata koordinat titik-titik anggota cluster sehingga nilai fungsi objektif jarak kuadrat dapat diminimalkan. Perhitungan centroid memberikan gambaran posisi tengah cluster dan menjadi dasar dalam menentukan pola pengelompokan data yang konsisten dari iterasi ke iterasi. Permasalahan pemilihan centroid awal sering dibahas dalam kajian aplikasi K-Means karena pilihan awal ini dapat mempengaruhi hasil akhir clustering dan konvergensi algoritma [15]. Secara matematis, centroid $\mu_i$ dari cluster ke-i dengan anggota data $x_1, x_2, \dots, x_n$ dihitung menggunakan rumus

$$\mu_i = \frac{1}{n} \sum_{j=1}^{n} x_j \quad (4)$$

   Keterangan

   $n =$ jumlah titik data dalam cluster $C_i$

   Penentuan centroid ini bersifat iteratif: setiap kali titik data berpindah cluster, posisi *centroid* diperbarui untuk meminimalkan fungsi objektif jarak kuadrat, sehingga cluster menjadi lebih kompak secara internal [10].
6. Ulangi seluruh tahapan ini hingga tidak terjadi lagi perpindahan titik data antar *cluster*.

## 3. HASIL DAN PEMBAHASAN

Data awal dan variable dalam penelitian

**Tabel 1.** Tabel Variabel

| No | Pelanggan | Frekuensi (X1) | Pengeluaran (X2) |
|---|---|---|---|
| 1 | P1 | 2 | 350 |
| 2 | P2 | 2 | 380 |
| 3 | P3 | 3 | 400 |
| 4 | P4 | 3 | 450 |
| 5 | P5 | 3 | 420 |
| 6 | P6 | 3 | 440 |
| 7 | P7 | 4 | 480 |
| 8 | P8 | 4 | 520 |
| 9 | P9 | 4 | 490 |
| 10 | P10 | 5 | 500 |
| 11 | P11 | 5 | 550 |
| 12 | P12 | 5 | 570 |
| 13 | P13 | 5 | 560 |
| 14 | P14 | 6 | 520 |


Submitted: **17/06/2026**; Accepted: **29/06/2026**; Published:**08/07/2026**

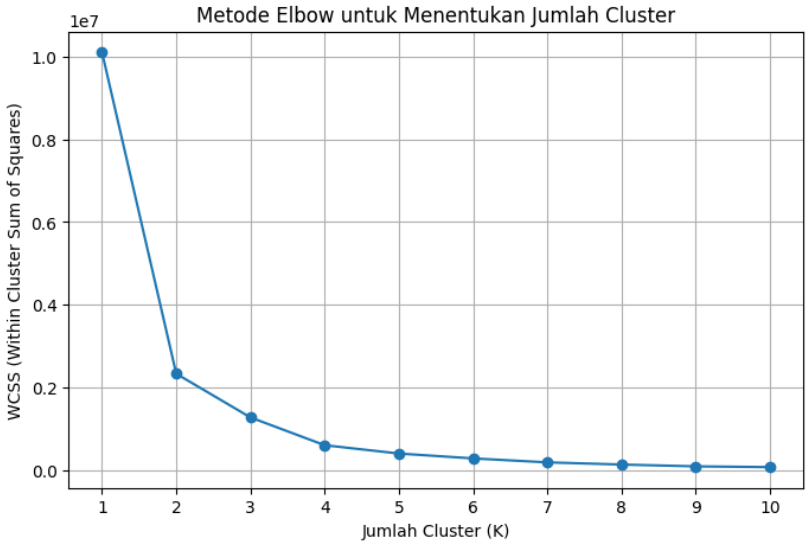
**Jurnal Kolaborasi**
**Sains dan Ilmu Terapan**
Volume 5 Nomor 1 Juli Desember 2026 | Page -9-18
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT*
Penerbit : Utiliti Project Solution

| | | | |
|---|---|---|---|
| 15 | P15 | 6 | 680 |
| 16 | P16 | 6 | 630 |
| 17 | P17 | 7 | 600 |
| 18 | P18 | 7 | 650 |
| 19 | P19 | 7 | 720 |
| 20 | P20 | 7 | 710 |
| 21 | P21 | 8 | 780 |
| 22 | P22 | 8 | 790 |
| 23 | P23 | 9 | 700 |
| 24 | P24 | 9 | 850 |
| 25 | P25 | 9 | 880 |
| 26 | P26 | 10 | 800 |
| 27 | P27 | 10 | 950 |
| 28 | P28 | 10 | 970 |
| 29 | P29 | 11 | 900 |
| 30 | P30 | 11 | 1050 |
| 31 | P31 | 11 | 1080 |
| 32 | P32 | 12 | 1200 |
| 33 | P33 | 12 | 1300 |
| 34 | P34 | 12 | 1150 |
| 35 | P35 | 12 | 1180 |
| 36 | P36 | 13 | 1250 |
| 37 | P37 | 13 | 1280 |
| 38 | P38 | 14 | 1400 |
| 39 | P39 | 14 | 1350 |
| 40 | P40 | 14 | 1380 |
| 41 | P41 | 15 | 1480 |
| 42 | P42 | 15 | 1480 |
| 43 | P43 | 16 | 1450 |
| 44 | P44 | 16 | 1380 |
| 45 | P45 | 17 | 1450 |
| 46 | P46 | 17 | 1680 |
| 47 | P47 | 18 | 1780 |
| 48 | P48 | 19 | 1680 |
| 49 | P49 | 20 | 1980 |
| 50 | P50 | 21 | 2000 |

Begitu variabel data ditetapkan, langkah berikutnya adalah mencari tahu berapa jumlah cluster yang paling ideal. Untuk keperluan ini, metode *Elbow* dipilih sebagai alat penentunya.


Submitted: **17/06/2026**; Accepted: **29/06/2026**; Published:**08/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli Desember 2026 | Page -9-18
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT*
Penerbit : Utiliti Project Solution




**Gambar 1.** Jumlah Cluster

Python membantu menerjemahkan data ke dalam grafik *Elbow*, dan di sanalah ceritanya menjadi jelas: penurunan *SSE* terjun paling curam saat k = 4. Begitu melewati angka itu, kurva mulai melandai ibarat jalan menurun yang tiba-tiba berubah jadi datar. Artinya, menambah klaster lagi hanya akan membuang tenaga tanpa hasil yang sepadan. Oleh karena itu, jumlah cluster optimal pada dataset pelanggan minimarket ini adalah k = 4, yang merepresentasikan dua segmen utama yaitu pelanggan reguler dan pelanggan premium. Selanjutanya menenetukan *cluster* setiap pelanggan, dengan menggunakan phyton maka didapat hasil sebagai berikut:


Submitted: **17/06/2026**; Accepted: **29/06/2026**; Published:**08/07/2026**

**Jurnal Kolaborasi**
**Sains dan Ilmu Terapan**
Volume 5 Nomor 1 Juli Desember 2026 | Page -9-18
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT*
Penerbit : Utiliti Project Solution

| | Frekuensi | Pengeluaran | Cluster |
|---|---|---|---|
| 0 | 2 | 350 | 3 |
| 1 | 2 | 380 | 3 |
| 2 | 3 | 400 | 3 |
| 3 | 3 | 450 | 3 |
| 4 | 3 | 420 | 3 |
| 5 | 3 | 440 | 3 |
| 6 | 4 | 480 | 3 |
| 7 | 4 | 520 | 3 |
| 8 | 4 | 490 | 3 |
| 9 | 5 | 500 | 3 |
| 10 | 5 | 550 | 3 |
| 11 | 5 | 570 | 3 |
| 12 | 5 | 560 | 3 |
| 13 | 6 | 520 | 3 |
| 14 | 6 | 680 | 1 |
| 15 | 6 | 630 | 3 |
| 16 | 7 | 600 | 3 |
| 17 | 7 | 650 | 3 |
| 18 | 7 | 720 | 1 |
| 19 | 7 | 710 | 1 |
| 20 | 8 | 780 | 1 |
| 21 | 8 | 790 | 1 |
| 22 | 9 | 700 | 1 |
| 23 | 9 | 850 | 1 |
| 24 | 9 | 880 | 1 |
| 25 | 10 | 800 | 1 |
| 26 | 10 | 950 | 1 |
| 27 | 10 | 970 | 1 |
| 28 | 11 | 900 | 1 |
| 29 | 11 | 1050 | 1 |
| 30 | 11 | 1080 | 1 |
| 31 | 12 | 1200 | 2 |
| 32 | 12 | 1300 | 2 |
| 33 | 12 | 1150 | 2 |
| 34 | 12 | 1180 | 2 |
| 35 | 13 | 1250 | 2 |
| 36 | 13 | 1280 | 2 |
| 37 | 14 | 1400 | 2 |
| 38 | 14 | 1350 | 2 |
| 39 | 14 | 1380 | 2 |
| 40 | 15 | 1480 | 2 |
| 41 | 15 | 1480 | 2 |
| 42 | 16 | 1450 | 2 |
| 43 | 16 | 1380 | 2 |
| 44 | 17 | 1450 | 2 |
| 45 | 17 | 1680 | 0 |
| 46 | 18 | 1780 | 0 |
| 47 | 19 | 1680 | 0 |
| 48 | 20 | 1980 | 0 |
| 49 | 21 | 2000 | 0 |

**Gambar 2.** Hasil Pengelompokkan Cluster

Hasil eksekusi algoritma *K-Means Clustering* menunjukkan bahwa basis pelanggan dapat dipilah ke dalam empat segmen utama, yang masing-masing menampilkan profil perilaku belanja yang saling berbeda. *Cluster* 3 tergolong pada kelompok pelanggan dengan aktivitas rendah. *Cluster* ini didominasi oleh pelanggan dengan frekuensi pembelian yang relative rendah yaitu berkisar antara 2 hingga 6 kali, serta tingkat pengeluaran yang berada pada rentang sekitar Rp


Submitted: **17/06/2026**; Accepted: **29/06/2026**; Published:**08/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli Desember 2026 | Page -9-18
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT*
Penerbit : Utiliti Project Solution

350.000 hingga Rp 680.000. Secara umum mereka dapat dikategorikan sebagai pelanggan yang belum aktif seecara konsisten atau hanya melakukan pembelian dalam skala kecil.

*Cluster* 1 tergolong pada kelompok pelanggan dengan aktivitas menengah. *Cluster* ini terdiri dari pelanggan dengan frekuensi pembelian antara 6 hingga 11 kali dengan pengeluaran berkisar Rp 600.000 hingga Rp 1.080.000. Secara strategis kelompok ini merupakan target yang penting untuk dipertahankan melalui program loyalitas, promosi berkala, serta penawaran bundling produk agar frekuensi pembelian dapat meningkat.

Sementara itu, *Cluster* 2 merepresentasikan segmen pelanggan yang sangat aktif. Mereka yang berada di kelompok ini mencatatkan frekuensi transaksi antara 12 hingga 17 kali, dengan total pengeluaran berkisar dari Rp 1.150.000 hingga Rp 1.480.000. Profil semacam ini menandakan adanya keterikatan yang kuat terhadap produk yang ditawarkan. Tidak berlebihan jika mereka disebut sebagai pelanggan loyal merupakan sebuah aset berharga yang menyumbang porsi pendapatan signifikan bagi toko. Maka dari itu, mempertahankan kelompok ini menjadi sebuah keniscayaan. Strategi retensi seperti program keanggotaan eksklusif, akumulasi poin reward, serta layanan khusus bagi pelanggan tetap dapat menjadi benteng agar loyalitas mereka tidak pudar.

Di puncak piramida pelanggan, terdapat *Cluster* 0 yang menyandang predikat sebagai segmen premium. Kelompok ini mencatatkan performa tertinggi pada kedua variabel yang diukur: frekuensi pembelian berkisar antara 17 hingga 21 kali, sementara total pengeluarannya menembus angka Rp 1.680.000 hingga Rp 2.000.000. Pelanggan dalam cluster ini dapat dikategorikan sebagai palanggan utama (premium customer) yang memberikan kontribusi terbesar tehadap total penjualan. Secara strategis, kelompok ini sangat penting untuk diprioritaskan melalui pendekatan eksklusif seperti layanan VIP, prioritas distribusi, serta program penghargaan pelanggan terbaik. Perbedaan karakteristik antar cluster tersebut menunjukkan adanya heterogenitas dalam pola konsumsi pelanggan. Kelompok pada *Cluster* 0 dan 2 dapat diidentifikasi sebagai pelanggan bernilai tinggi (*high value customers*) yang memiliki kontribusi lebih besar terhadap pendapatan, sedangkan *Cluster* 1 dan 3 merupakan segmen pelanggan reguler dengan potensi peningkatan nilai transaksi melalui strategi pemasaran yang tepat. Temuan ini dapat dijadikan pijakan dalam merancang strategi segmentasi pasar, terutama untuk menyusun program promosi yang lebih terfokus dan selaras dengan karakter unik setiap segmen pelanggan.

Berikut hasil perhitungan *K-Means clustering*, dengan menggunakan phyton

**Tabel 2.** Data Centroid Cluster

| Cluster | Frekuensi | Pengeluaran |
|---|---|---|
| 0 | 19.000000 | 1824.000000 |
| 1 | 9.000000 | 847.142857 |
| 2 | 13.928571 | 1337.857143 |
| 3 | 4.352941 | 500.588235 |

Hasil perhitungan *K-Means clustering* menunjukkan terbentuknya dua pusat klaster (*centroid*) yang masing-masing merepresentasikan karakteristik rata-rata anggota dalam setiap kelompok. *Cluster* 3 (*Centroid*: 4.35; 500.59) memiliki nilai rata-rata frekuensi transaksi sebesar 4 kali dengan rata-rata pengeluaran sebesar Rp 500.000, yang mengindikasikan bahwa kelompok ini menggambarkan pelanggan dengan intensitas pembelian rendah. *Cluster* 1 (*Centroid*: 9.00 ; 847.14), cluster ini berada pada tingkat menengah.dengan frekuensi pembelian rata-rata sekitar 9 kali dan pengeluaran rata-rata sekitar Rp 847.000. *Cluster* 2 (*Centroid*: 13.93 ; 1337.86), cluster ini menunjukkan nilai yang lebih tinggi. Frekuensi pembelian mendekati 14 kali pengeluaran sekitar Rp 1.330.000. Kelompok ini merupakan pelanggan loyal dengan intensitas transaksi tinggi. Kelompok ini menyumbang porsi yang cukup signifikan terhadap keseluruhan pendapatan penjualan. *Cluster* 0 (*Centroid*: 19.00 ; 1824.00), cluster ini memiliki nilai tertinggi di antara seluruh cluster. Frekuensi pembelian sekitar 19 kali, pengeluaran mencapai sekitar Rp 1.820.000 juta. Kelompok ini merupakan pelanggan utama atau pelanggan premium. Mereka memiliki tingkat loyalitas dan kontribusi ekonomi paling tinggi dibanding kelompok lainnya.

## 4. KESIMPULAN

Melalui penerapan algoritma *K-Means Clustering* pada data pelanggan dengan dua variabel utama yang frekuensi transaksi dan total pengeluaran serta analisis ini berhasil mengidentifikasi empat kelompok berbeda, masing-masing


Submitted: **17/06/2026**; Accepted: **29/06/2026**; Published:**08/07/2026**

**Jurnal Kolaborasi**
**Sains dan Ilmu Terapan**
Volume 5 Nomor 1 Juli Desember 2026 | Page -9-18
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT*
Penerbit : Utiliti Project Solution

dengan corak perilaku belanja yang khas (K = 4).pelanggan pada *Cluster* 0 dengan frekuensi pembelian dan jumlah pengeluaran tinggi perlu dipertahankan melalui program loyalitas dan layanan yang lebih baik. *Cluster* 2 dengan frekuensi pembelian tinggi tetapi jumlah pengeluaran lebih rendah dapat didorong untuk meningkatkan nilai transaksi melalui promosi atau penawaran produk pelengkap. *Cluster* 1 dengan frekuensi pembelian rendah namun jumlah pengeluaran tinggi perlu dipertahankan melalui penawaran yang lebih personal agar melakukan pembelian lebih rutin. Sementara itu, *Cluster* 3 dengan frekuensi pembelian dan jumlah pengeluaran rendah memerlukan strategi promosi yang lebih intensif untuk meningkatkan aktivitas pembelian atau menjadi bahan evaluasi dalam penyusunan strategi pemasaran. Temuan *clusterisasi* ini membuktikan bahwa algoritma *K-Means* sanggup menyingkap pola-pola tersembunyi yang ada dalam data pelanggan. Dengan kemampuan tersebut, metode ini dapat diandalkan sebagai fondasi dalam menjalankan segmentasi pasar.Informasi yang dihasilkan dari proses klasterisasi ini memiliki potensi untuk mendukung pengambilan keputusan strategis, khususnya dalam merancang pendekatan pemasaran yang lebih adaptif dan berbasis karakteristik masing-masing kelompok pelanggan. Dengan penerapan strategi yang sesuai pada setiap klaster, Toko Susu Asia diharapkan dapat meningkatkan loyalitas pelanggan dan mengoptimalkan kinerja penjualan.

## UCAPAN TERIMAKASIH

Ucapan terima kasih yang sebesar-besarnya penulis sampaikan kepada STMIK Pelita Nusantara, khususnya kepada seluruh jajaran pimpinan dan staf, yang telah memberikan dukungan penuh, baik secara moril maupun materiil, sehingga penelitian ini dapat terlaksana dengan baik.

## REFERENCES

[1] A. Adio, A. Orobosade, and O. Temitope, “An Improvement on Initial Centre Points of K-Means Clustering Algorithm,” *Int. J. Res. Sci. Innov.*, vol. XII, no. 2321, 2025, doi: 10.51244/IJRSI.

[2] S. Zhang, S. Chen, X. Yu, and S. Mei, “Research on collaborative filtering algorithm based on improved K-means algorithm for user attribute rating and co-rating,” 2025.

[3] M. A. Aldi *et al.*, “IMPLEMENTASI K-MEANS CLUSTER ING DALAM PENGELOMPOKAN DATA KUNJUNGAN WISATAWAN ASING DI INDONESIA,” *J. Ilm. Multidisiplin Ilmu*, vol. 2, no. 1, pp. 13–19, 2025.

[4] A. Fatkhudin, A. Khambali, F. A. Artanto, and N. A. P. Zade, “Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa Studi Kasus ( Prodi Manajemen Informatika ),” *J. Minfo Polgan*, vol. 12, no. 2, pp. 777–783, 2023.

[5] M. Adjie Pahlevi, A. Abimanyu, M. Hafizh Arrizky, and A. Ryaldi, “Analisis Distribusi Tenaga Kesehatan Di Indonesia Menggunakan K-Means Clustering,” *J. Tenik Inform. dan Sist. Inf.*, vol. 11, no. 3, pp. 287–298, 2024.

[6] A. F. Zabidi, “Penerapan Algoritma K-Means untuk Pengelompokan Koleksi Perpustakaan dengan Data Mining,” *Media J. Inform.*, vol. 16, no. 2, 2024.

[7] W. Permata Sari and T. Sutabri, “ANALISA CLUSTER DENGAN K-MEAN CLUSTERING UNTUK PENGELOMPOKAN DATA CYBERCRIME,” *J. Inform. Teknol. dan Sains*, vol. 5, no. 1, pp. 49–53, 2023.

[8] R. Buaton and Solikhun, “Application of Numerical Measure Variations in K-Means Clustering for Grouping Data,” *Matrik J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 23, no. 1, pp. 103–112, 2023, doi: 10.30812/matrik.v23i1.3269.

[9] N. A. Maori, “METODE ELBOW DALAM OPTIMASI JUMLAH CLUSTER PADA K-MEANS CLUSTERING,” *J. Simetris*, vol. 14, no. 2, pp. 277–287, 2023.

[10] E. Wira Liyanto, A. Homaidi, and A. Lutfi, “IMPLEMENTASI K-MEANS CLUSTERING MENGGUNAKAN RAPIDMINER DALAM PENGELOMPOKAN DATA KUNJUNGAN WISATAWAN ASING DI PROVINSI JAWA TIMUR,” *J. Tek. Elektro dan Inform.*, vol. 19, no. September, pp. 205–216, 2024.

[11] M. Amelia, A. Faqih, A. R. Rinaldi, and K. Sosial, “PENERAPAN METODE K-MEANS CLUSTERING DALAM PEMETAAN KEMISKINAN KABUPATEN / KOTA DI TEPAT,” *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, pp. 437–444, 2025.

[12] B. Hartono and V. Lusiana, “Analisis Metode Elbow SSE , Silhouette Score , dan Jaccard Stability dalam Pemilihan Jumlah Klaster Data yang Optimal TIN : Terapan Informatika Nusantara,” *TIN Terap. Infomatika Nusant.*, vol. 6, no. 8, pp. 1521–1532, 2026, doi: 10.47065/tin.v6i8.9271.

[13] B. A. Hendriansyah, A. Tri, J. Harjanta, and K. Latifah, “IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA SISTEM INFORMASI GEOGRAFIS FASILITAS KESEHATAN BPJS KESEHATAN KOTA SEMARANG,” *J. Inform. Teknol. dan Sains*, vol. 7, no. 1, pp. 438–448, 2025.

[14] A. Fikri, B. F. Hutabarat, and U. Khaira, “Komparasi K-Means Clustering Dan Complete Linkage Dalam Pengelompokan Penyaluran Pinjaman Financial Technology,” *J. Ilm. MEDIA ISFO*, vol. 17, no. 2, pp. 228–239,


Submitted: **17/06/2026**; Accepted: **29/06/2026**; Published:**08/07/2026**

**Jurnal Kolaborasi**
**Sains dan Ilmu Terapan**
Volume 5 Nomor 1 Juli Desember 2026 | Page -9-18
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT*
Penerbit : Utiliti Project Solution

2023.
[15] D. Anggraini and M. Siddik Hasibuan, "Initial Centroid Pada Algoritma K-Means Dan K-Medoids," *J. Multimed. dan Teknol. Inf.*, vol. 07, no. 03, pp. 487–500, 2025.


Submitted: **17/06/2026**; Accepted: **29/06/2026**; Published:**08/07/2026**