

Implementasi Algoritma K-Means Clustering untuk Segmentasi Pelanggan

Implementation of K-Means Clustering Algorithm for Customer Segmentation

¹Kapti*, ²Dwi Astuti

¹Program Studi Teknik Informatika, STMIK Bina Patria

²Program Studi Sistem Informasi, STMIK Bina Patria

^{1,2}Jl. Raden Saleh No.7, Potrobangsari, Kec. Magelang Utara, Kota Magelang,
Jawa Tengah, Indonesia

*e-mail: tensmart18@stmikbinapatria.ac.id

(received: 20 July 2025, revised: 13 August 2025, accepted: 15 August 2025)

Abstrak

Seiring dengan berkembangnya teknologi digital, perusahaan dihadapkan pada volume data pelanggan yang semakin besar dan kompleks. Untuk itu, diperlukan pendekatan yang lebih efektif dalam segmentasi pelanggan agar dapat memahami pola perilaku dan preferensi konsumen dengan lebih baik. Banyaknya pelanggan yang melakukan transaksi setiap harinya, sehingga menyulitkan manager dalam mengelompokkan segmentasi antara pelanggan rajin belanja dan yang kurang rajin belanja berdasarkan Data Transaksi dan Preferensi Pembelian. Untuk itu peran algoritma k-means clustering sangat dibutuhkan dalam menyelesaikan permasalahan ini. Tujuan dari Penelitian ini adalah untuk mengimplementasikan algoritma K-Means Clustering dalam segmentasi pelanggan dengan membagi dalam tiga kluster yaitu pelanggan sangat rajin, cukup rajin dan kurang rajin belanja. Metode yang digunakan dalam penelitian ini seperti: Pengumpulan Data Transaksi dan Preferensi Pelanggan, Pra Pemrosesan Data, Implementasi Algoritma, Validasi dan Pengujian. Parameter yang digunakan dalam klusterisasi ini berdasarkan jumlah pembelian, jumlah transaksi dan jumlah barang. Hasil pengujian fungsionalitas menunjukkan sistem dinyatakan baik karena seluruh skenario uji berjalan sukses. Hasil Pengujian dengan metode *Silhouette Coefficient* (SC) menghasilkan status *strong struktur* dengan nilai rata SC sebesar 0,97 hal ini menunjukkan bahwa dataset dinyatakan kuat untuk dijadikan sebagai role model dalam segmentasi pelanggan.

Kata kunci: k-means clustering, segmentasi pelanggan, pengelompokan, pemasaran digital

Abstract

Along with the development of digital technology, companies are faced with increasingly large and complex volumes of customer data. For this reason, a more effective approach in customer segmentation is needed to better understand consumer behavior patterns and preferences. The large number of customers who make transactions every day makes it difficult for managers to segment customers between frequent shoppers and those who are less frequent shoppers based on Transaction Data and Purchase Preferences. For this reason, the role of the k-means clustering algorithm is needed to solve this problem. The purpose of this study is to implement the K-Means Clustering algorithm in customer segmentation by dividing them into three clusters, namely very frequent customers, quite frequent customers, and less frequent shoppers. The methods used in this study are: Collection of Transaction Data and Customer Preferences, Data Preprocessing, Algorithm Implementation, Validation, and Testing. The parameters used in this clustering are based on the number of purchases, the number of transactions, and the number of items. The results of the functional testing show that the system is well stated because all test scenarios run successfully. The test results using the *Silhouette Coefficient* (SC) method produced a strong structure status with an average SC value of 0.97. This indicates that the dataset is declared strong to be used as a role model in customer segmentation.

Keywords: clustering, grouping, drugs, k-means clustering, health center

<http://sistemasi.ftik.unisi.ac.id>

1 Pendahuluan

Perkembangan teknologi informasi yang pesat telah membawa perubahan signifikan dalam berbagai sektor bisnis, termasuk dalam pengelolaan dan analisis data pelanggan. Data pelanggan yang dihasilkan dari transaksi e-commerce, aplikasi, dan platform digital lainnya semakin besar dan kompleks, menuntut perusahaan untuk mengadopsi teknik analisis data yang lebih canggih untuk memahami perilaku pelanggan. Segmentasi pelanggan menjadi salah satu metode penting dalam pemasaran yang memungkinkan perusahaan untuk mengelompokkan pelanggan berdasarkan karakteristik atau perilaku tertentu sehingga dapat memberikan layanan yang lebih personal dan efektif.

Segmentasi pelanggan tradisional sering kali mengandalkan data demografis atau geografis yang terbatas. Namun, dalam era digital ini, segmentasi yang lebih tepat harus didasarkan pada data transaksi dan preferensi pembelian pelanggan yang lebih dinamis. Salah satu teknik yang populer digunakan untuk segmentasi pelanggan adalah algoritma K-Means Clustering. Algoritma ini memungkinkan pengelompokan pelanggan berdasarkan kemiripan data mereka dalam ruang fitur yang multidimensional. Peran Algoritma K-Means Clustering sangatlah penting karena masalah utama dari penelitian ini adalah banyak transaksi yang terjadi namun manager kesulitan dalam mengelompokkan segmentasi pelanggan secara sistematis [1].

Permasalahan ini dapat menyulitkan bagi perusahaan dalam pengambilan keputusan dalam menentukan pengelompokan segmentasi pelanggan sesuai klasternya. Untuk itu, diperlukan Implementasi dari Algoritma K-Means sebagai solusi dalam klasterisasi dan segmentasi pelanggan. Pengembangan Model dan Implementasi sistem ini dapat membantu stakeholder dalam pengambilan keputusan. Dengan mengintegrasikan teknik-teknik ini, diharapkan dapat diperoleh hasil segmentasi yang lebih stabil dan akurat, yang pada gilirannya dapat meningkatkan efektivitas strategi pemasaran dan pengalaman pelanggan.

Algoritma K-Means Clustering telah menjadi salah satu metode yang banyak digunakan untuk segmentasi pelanggan [2]. Namun, algoritma ini menghadapi beberapa keterbatasan, di antaranya adalah ketergantungannya terhadap inisialisasi centroid awal dan kerentanan terhadap outlier, yang dapat menyebabkan hasil segmentasi yang tidak stabil dan tidak akurat [3]-[4]. Hal ini dapat membahayakan keputusan strategis perusahaan dan berdampak negatif pada pengalaman pelanggan. Dengan mengalami salah satu pengelompokan, perusahaan dapat membuang sumber daya dalam upaya pemasaran yang tidak efektif, dan berpotensi kehilangan loyalitas pelanggan dan pendapatan [5]-[6].

Untuk itu riset ini dirancang untuk menjawab tantangan tersebut dengan mengimplementasikan Algoritma K-Means untuk melakukan klasterisasi dan segmentasi pelanggan [7]. Penelitian ini bertujuan tidak hanya untuk meningkatkan kualitas segmentasi tetapi juga untuk memberikan panduan praktik bagi perusahaan dalam menerapkan strategi pemasaran yang lebih efektif dan personal [8]-[9].

Urgensi penelitian ini dapat dilihat dari beberapa dimensi kritis. Secara ekonomi, penelitian ini berpotensi mencegah pemborosan sumber daya pemasaran dan mengoptimalkan strategi target pasar [10]-[11]. Dari perspektif teknologi, penelitian ini memberikan kontribusi metodologis dalam bidang data mining dan menghadirkan solusi inovatif bagi keterbatasan algoritma eksisting [12]. Secara strategis, penelitian ini sejalan dengan Prioritas Riset Nasional dan mendukung transformasi digital yang sedang berlangsung [13].

Penelitian ini mengusulkan pendekatan komprehensif untuk mengimplementasikan Algoritma K-Means clustering dalam segmentasi pelanggan [14]. Strategi utama difokuskan pada perbaikan inisialisasi centroid awal, yang memungkinkan pemilihan titik pusat awal yang lebih cerdas dan tersebar, secara signifikan mengurangi bias inisialisasi tradisional [15].

Pendekatan inovatif ini akan dikembangkan pada suatu pengembangan sistem yang dinamis sehingga mampu mempermudah dalam proses klasterisasi dan segmentasi pelanggan. Dengan menggunakan teknik deteksi berbasis jarak untuk mengelola data ekstrim tanpa menghilangkan informasi penting [16].

Penelitian ini bertujuan untuk mengembangkan dan mengimplementasikan algoritma K-Means Clustering yang lebih adaptif untuk segmentasi pelanggan, dengan fokus pada penggunaan data transaksi dan preferensi pembelian. Penelitian ini juga akan melakukan evaluasi komparatif terhadap algoritma K-Means yang telah dimodifikasi dan K-Means konvensional untuk mengukur perbaikan dalam akurasi dan stabilitas segmentasi. Hasil dari penelitian ini diharapkan dapat memberikan

kontribusi signifikan dalam penerapan algoritma K-Means untuk strategi pemasaran yang lebih personal dan berbasis data [17].

2 Tinjauan Literatur

Penelitian ini memiliki keterkaitan dengan penelitian terdahulu seperti penelitian dengan judul, "Applying Machine Learning in Marketing: An Analysis Using the NMF and k-Means Algorithms", Penelitian ini menggunakan NMF untuk topic modeling dan k-means untuk analisis komparatif dalam pemasaran. Fokusnya adalah meningkatkan segmentasi pelanggan dan kampanye periklanan. Penelitian ini menggunakan k-means namun lebih menekankan pada penerapan machine learning secara umum, sedangkan penelitian pengusul fokus pada pengembangan k-means untuk segmentasi berdasarkan data transaksi dan preferensi pembelian [18].

Penelitian dengan judul, "Customer Segmentation in E-commerce: A Context-Aware Quality Framework for Comparing Clustering Algorithms", Penelitian ini membandingkan beberapa algoritma clustering, termasuk k-means, dalam konteks e-commerce. Meskipun menggunakan k-means, penelitian ini juga memperkenalkan kerangka kerja kualitas dan membandingkan dengan algoritma lain. Penelitian pengusul fokus pada pengembangan algoritma k-means untuk segmentasi pelanggan berdasarkan data yang lebih spesifik [19].

Penelitian dengan judul, "Enhancing K-Means Clustering with Post-Redistribution", Penelitian ini mengusulkan penyempurnaan algoritma k-means dengan redistribusi pasca-clustering untuk meningkatkan kualitas cluster. Fokusnya pada perbaikan algoritma k-means berbeda dengan penelitian pengusul yang tidak mengusulkan modifikasi serupa tetapi tetap berfokus pada pengembangan k-means untuk segmentasi pelanggan [20].

Penelitian dengan judul, "Applied Density-Based Clustering Techniques for Classifying High-Risk Customers: A Case Study of Commercial Banks in Vietnam", Penelitian ini membandingkan k-means dengan algoritma berbasis kepadatan (DBSCAN, HDBSCAN) dalam mengklasifikasikan pelanggan berisiko tinggi. Penelitian pengusul lebih terfokus pada k-means untuk segmentasi pelanggan berdasarkan kriteria jumlah pembelian, jumlah transaksi dan jumlah barang [21].

Penelitian ini menghadirkan lima kebaruan utama dalam pengembangan algoritma K-Means untuk segmentasi pelanggan. Pertama, algoritma dibuat lebih adaptif dengan mengintegrasikan K-Means++ untuk inisialisasi centroid, sehingga meningkatkan kualitas cluster. Kedua, diterapkan metode robust untuk mendeteksi dan menangani outlier secara efektif guna menjaga integritas data. Ketiga, penelitian ini memfokuskan pada data transaksi dan preferensi pembelian pelanggan, bukan hanya data demografis atau perilaku online seperti pada penelitian sebelumnya, sehingga memberikan segmentasi yang lebih relevan untuk strategi pemasaran. Keempat, algoritma dikembangkan secara spesifik untuk konteks industri ritel atau e-commerce agar lebih aplikatif dalam dunia bisnis nyata. Kelima, digunakan metrik evaluasi yang disesuaikan dengan tujuan bisnis, bukan hanya metrik standar, agar penilaian kinerja algoritma lebih tepat dan kontekstual.

3 Metode Penelitian

Metodologi penelitian ini dirancang untuk menjelaskan alur penelitian dalam implementasi Algoritma K-Means Clustering untuk segmentasi pelanggan berdasarkan data transaksi dan preferensi pembelian. Metode penelitian ini mengikuti empat tahapan utama seperti yang disebutkan dalam ringkasan, yaitu:

1. Pengumpulan Data Transaksi dan Preferensi Pelanggan

1.1 Sumber Data

- Data transaksional diambil dari perusahaan ritel/e-commerce di Kota Magelang, data ini mencakup data pembelian pelanggan minimal 1 tahun terakhir. Data diambil di salah satu cabang Perusahaan PT Indomart di Kota Magelang melalui kerjasama mitra secara resmi.
- Data preferensi pembelian dari survei pelanggan dan rekam jejak digital
- Data demografis pelanggan sebagai data pendukung

1.2 Strategi Pengumpulan Data

- Pengumpulan data dilakukan melalui kegiatan FGD dengan survei untuk identifikasi data
- Pengambilan data menggunakan teknik API integration dan database querying

<http://sistemasi.ftik.unisi.ac.id>

- c. Pengumpulan data preferensi melalui survei online dengan mengambil sampel sebanyak 20 data.
- d. Pengumpulan data juga dilakukan secara langsung melalui wawancara

2. Pra Pemrosesan Data

- a. Pembersihan data untuk menangani missing values dan data yang tidak konsisten
- b. Transformasi data melalui normalisasi dan standardisasi
- c. Identifikasi awal outlier menggunakan metode statistik standar

3. Implementasi Algoritma K-Means Clustering

3.1 Menentukan Jumlah Cluster (K)

- a. Penentuan Jumlah Cluster K dilakukan secara acak untuk sebagai dasar perhitungan.
- b. Nilai Centroid awal diambil secara acak serta disesuaikan dengan nilai rata-rata dari dataset[22].
- c. Pengujian stabilitas inisialisasi pada berbagai skenario dataset

3.2 Inisialisasi Centroid

- a. Setelah jumlah cluster ditentukan, langkah selanjutnya adalah memilih titik awal untuk setiap cluster, yang disebut centroid. Titik-titik ini biasanya dipilih secara acak dari data yang ada.
- b. Pemilihan secara acak serta disesuaikan dengan nilai rata-rata dataset.

3.3 Penentuan Anggota Cluster

- a. Setiap titik data dalam dataset kemudian proses ke cluster yang centroidnya terdekat. Jarak antara titik data dan centroid diukur menggunakan metrik jarak, seperti Euclidean distance.
- b. Setelah itu hasilnya digunakan sebagai acuan dalam proses iterasi.

3.4 Penentuan Anggota Cluster

- a. Setelah semua titik data anggota ke cluster, centroid dari masing-masing cluster dihitung ulang. Centroid baru adalah rata-rata dari semua titik data yang termasuk dalam cluster tersebut.
- b. Hasil cluster ini dilanjutkan pada tahap iterasi selanjutnya.

3.5 Iterasi hingga konvergensi

- a. Setiap titik Langkah 3 dan 4 diulang secara iteratif. Titik data diassign ulang ke cluster berdasarkan centroid baru, dan centroid baru dihitung kembali. Proses ini terus berlanjut hingga tidak ada perubahan yang signifikan pada posisi centroid (konvergensi), atau hingga jumlah iterasi maksimum tercapai.
- b. Tahapan iterasi berakhir ketika hasil cluster sebelum dan setelahnya nilainya sama.

4. Evaluasi Komparatif

4.1 Metrik Evaluasi Internal

- a. Pengujian Evaluasi sistem ini digunakan sebagai bentuk penentuan kualitas model dataset.
- b. Silhouette Index untuk mengukur kohesi dan separasi cluster dengan batas nilai $SC \geq 0,71$ (struktur kuat)[24].

4.2 Uji Fungsionalitas Model Sistem

- a. Pengujian fungsionalitas model sistem digunakan sebagai alat dalam mendeteksi kinerja dari suatu sistem.
- b. Apabila seluruh form sistem dapat berjalan dengan lancar, maka sistem dinyatakan normal atau berfungsi dengan baik.

4.3 Uji Validasi

- a. Pengujian Validasi dilakukan dengan membandingkan antara hasil dari perhitungan manual dengan hasil dari program aplikasi.

- b. Apabila hasil perbandingan diperoleh hasil yang sama, maka model sistem ini dinyatakan valid[25].

5. Validasi melalui Studi Kasus Riil

5.1 Implementasi Strategi Pemasaran

- a. Pengembangan strategi pemasaran yang dipersonalisasi berdasarkan segmen pelanggan hasil clustering
- b. Implementasi strategi pada kelompok eksperimen di perusahaan mitra
- c. Pelaksanaan melakukan testing untuk mengukur efektivitas strategi pemasaran berbasis segmentasi baru [26].

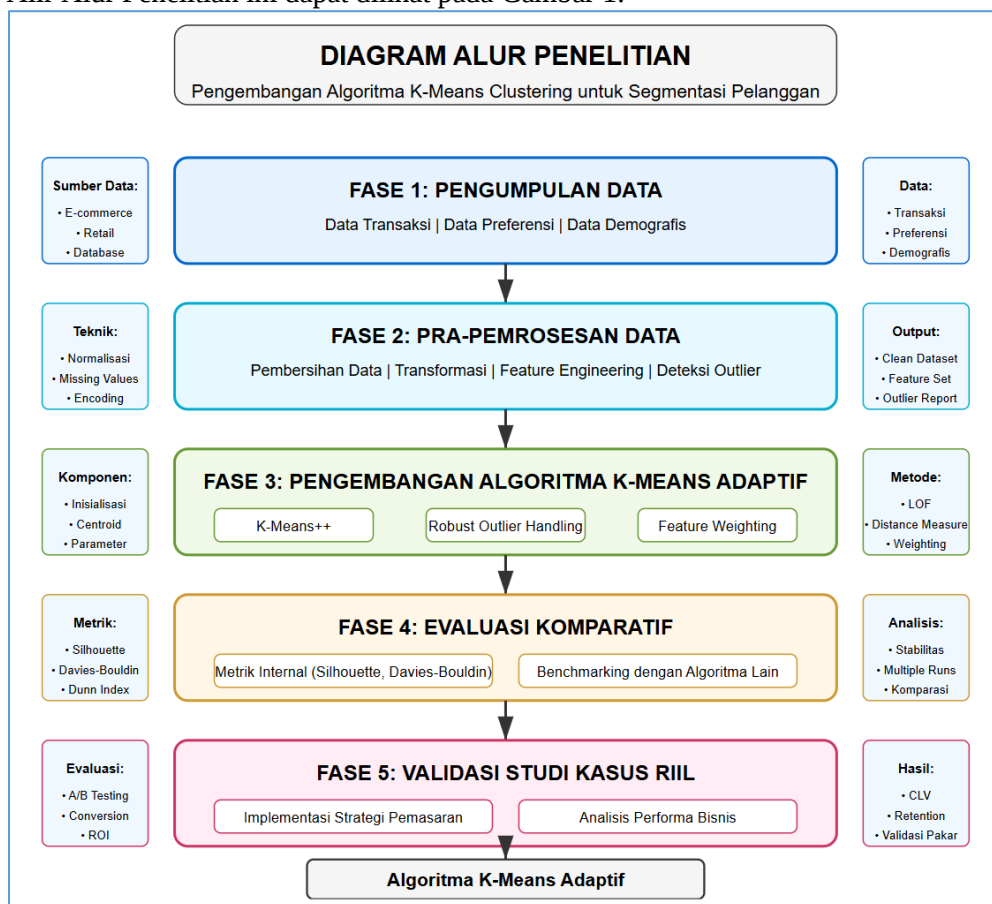
5.2 Pengukuran Performa Bisnis

- a. Evaluasi Sistem juga dilakukan berdasarkan performa bisnis.
- b. Pengukuran dilakukan dengan memberikan testing sistem kepada stake holder dan pengguna sistem.
- c. Analisis Nilai Kuisioner diberikan setelah pengguna selesai menggunakan sistem[27].

5.3 Validasi dengan Domain Expert

- a. Evaluasi interpretabilitas segmen oleh pakar pemasaran
- b. Penilaian kegunaan segmentasi dalam pengambilan keputusan strategis
- c. Analisis kualitatif terhadap manfaat segmentasi dalam personalisasi layanan
- d. Implementasi dan pembuktian hasil dilakukan melalui uji validasi dengan membandingkan antara hasil perhitungan manual dengan hasil prototipe.
- e. Validasi eksternal dilakukan melalui penilaian langsung oleh stakeholder dan konsumen menggunakan kuesioner.

Adapun Alir Alur Penelitian ini dapat dilihat pada Gambar 1.



Gambar 1 Diagram alur penelitian

4 Hasil dan Pembahasan

Pembahasan dalam penelitian ini difokus dalam proses implementasi algoritma k-means clustering dalam melakukan klasterisasi dan segmentasi pelanggan dengan membagi kedalam 3 klaster yaitu pelanggan kurang rajin belanja, cukup rajin belanja dan sangat rajin belanja.

A. Implementasi Algoritma K-Means Clustering

Pada penelitian ini peneliti membatasi dan fokus pada implementasi dari Algoritma K-Means Clustering serta dikembangkan pada suatu sistem untuk mengelompokan segmentasi pelanggan. Data yang digunakan dalam penelitian ini mengambil sebanyak 20 data sampel secara acak dari PT Indomart Magelang. Data yang diambil berdasarkan prameter seperti jumlah pembelian, jumlah transaksi dan jumlah barang. Tahapan Implementasi Algoritma K-Means Clustering sebagai berikut:

1. Menentukan Dataset Pelanggan

Pada penelitian ini dari data pelanggan yang telah dikumpulkan kemudian dibuat dalam bentuk dataset seperti disajikan pada Tabel 1.

Tabel 1 Dataset pelanggan

No	Nama Pelanggan	Alamat	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang
1	Ahmad	Dukun	Rp 87.625	4	60
2	Ardian	Kaliangkrik	Rp 34.250	3	15
3	Bagas	Kajoran	Rp 40.000	1	5,5
4	Budianto	Dukun	Rp 37.000	2	22
5	Cerly	Kajoran	Rp 256.750	5	36
6	Farid	Bandongan	Rp 139.000	3	40
7	Fitriani	Dukun	Rp 271.750	5	142
8	Galih	Kaliangkrik	Rp 41.750	4	30
9	Gading	Ngablak	Rp 49.250	5	20
10	Herianto	Bandongan	Rp 145.750	5	27
11	Hermawan	Secang	Rp 312.500	7	129
12	Ikah	Salam	Rp 36.500	3	64
13	Joko	Bandongan	Rp 49.250	4	39
14	Kamal	Tegalrejo	Rp 49.250	3	41
15	Kurnia	Ngablak	Rp 40.750	2	23
16	Lutfy	Pakis	Rp 49.250	4	45
17	Mona	Sawangan	Rp 149.250	6	65
18	Nasir	Ngluwar	Rp 254.250	6	45
19	Opta	Pakis	Rp 45.250	4	32
20	Prabu	Ngluwar	Rp 17.500	2	9
MIN :			Rp 17.500	1	5,5
MAX :			Rp 312.500	7	142

2. Menghitung Normalisasi

Berikut ini contoh perhitungan Normalisasi pada baris pertama.

$$\begin{aligned}\text{Jumlah Pembelian} &= (\text{Nilai}-\text{Min}(\text{nilai})) / (\text{Max}(\text{nilai})-(\text{Min}(\text{nilai}))) \\ &= (87625-17500) / (312500-17500) \\ &= 0,24\end{aligned}$$

$$\begin{aligned}\text{Jumlah Transaksi} &= (\text{Nilai}-\text{Min}(\text{nilai})) / (\text{Max}(\text{nilai})-(\text{Min}(\text{nilai}))) \\ &= ((4-1) / (7-1)) \\ &= 0,5\end{aligned}$$

$$\begin{aligned}\text{Jumlah Barang} &= (\text{Nilai}-\text{Min}(\text{nilai})) / (\text{Max}(\text{nilai})-(\text{Min}(\text{nilai}))) \\ &= ((60-5,5) / (142-5,5)) \\ &= 0,40\end{aligned}$$

Adapun rekap proses perhitungan normalisasi ditunjukkan pada Tabel 2.

Tabel 2 Normalisasi

No	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang
1	0,24	0,50	0,40
2	0,06	0,33	0,07
3	0,08	0,00	0,00
4	0,07	0,17	0,12
5	0,81	0,67	0,22
6	0,41	0,33	0,25
7	0,86	0,67	1,00
8	0,08	0,50	0,18
9	0,11	0,67	0,11
10	0,43	0,67	0,16
11	1,00	1,00	0,90
12	0,06	0,33	0,43
13	0,11	0,50	0,25
14	0,11	0,33	0,26
15	0,08	0,17	0,13
16	0,11	0,50	0,29
17	0,45	0,83	0,44
18	0,80	0,83	0,29
19	0,09	0,50	0,19
20	0,00	0,17	0,03

Cluster Centroid Awal			
	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang
Cluster 1	0,06	0,17	0,03
Cluster 2	0,11	0,33	0,25
Cluster 3	0,86	0,83	1,00

$$d(x_i, x_j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2)}$$

Cluster Centroid Awal ditentukan dengan cara diambil nilai secara acak yang ditandai dengan tanda warna dari tabel 2 normalisasi.

3. Menghitung Iterasi 1

Pada perhitungan iterasi ke 1 dilakukan dengan menghitung nilai parameter berdasarkan dibandingkan dengan centroid awal. Berikut ini penulis sajikan pada contoh baris pertama yaitu

Cluster 1 = $\sqrt{(0,24-0,06)^2 + ((0,50-0,17)^2) + ((0,40-0,03)^2)} = 0,53$

Cluster 2 = $\sqrt{(0,24-0,11)^2 + ((0,50-0,33)^2) + ((0,40-0,25)^2)} = 0,26$

Cluster 3 = $\sqrt{(0,24-0,86)^2 + ((0,50-0,83)^2) + ((0,40-1)^2)} = 0,93$

Berdasarkan proses perhitungan normalisasi yang telah dilakukan sehingga menghasilkan iterasi ke 1 seperti disajikan pada Tabel 3.

Tabel 3 Hasil perhitungan iterasi ke 1

No	Cluster 1	Cluster 2	Cluster 3	Posisi Cluster	Jarak Terdekat	Pelanggan
1	0,53	0,26	0,93	2	0,26	Ahmad
2	0,17	0,19	1,33	1	0,19	Ardian
3	0,17	0,42	1,52	1	0,17	Bagas
4	0,09	0,21	1,36	1	0,09	Budianto
5	0,92	0,78	0,80	2	0,78	Cerly
6	0,45	0,30	1,00	2	0,30	Farid
7	1,35	1,11	0,16	3	0,16	Fitriani
8	0,36	0,19	1,18	2	0,19	Galih
9	0,50	0,37	1,18	2	0,37	Gading
10	0,64	0,48	0,96	2	0,48	Herianto
11	1,53	1,29	0,24	3	0,24	Hermawan
12	0,43	0,18	1,10	2	0,18	Ikah
13	0,40	0,17	1,12	2	0,17	Joko
14	0,29	0,01	1,17	2	0,01	Kamal
15	0,10	0,21	1,35	1	0,10	Kurnia
16	0,42	0,17	1,09	2	0,17	Lutfy
17	0,87	0,63	0,70	2	0,63	Mona
18	1,03	0,86	0,71	3	0,71	Nasir
19	0,37	0,18	1,16	2	0,18	Opta
20	0,06	0,30	1,46	1	0,06	Prabu

Berdasarkan hasil iterasi ke 1 maka diperoleh jumlah Cluster C1 = 5, C2 = 12 dan C3 = 3

4. Menentukan Centroid Baru

Nilai Centroid baru ini digunakan sebagai salah satu tahapan dalam perhitungan iterasi dari masing-masing cluster. Adapun hasil centroid baru seperti disajikan Gambar 2.

Centeroid Baru			
	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang
C1	0,056	0,167	0,0689
C2	0,251	0,528	0,2643
C3	0,888	0,833	0,7314

0,056 = Jumlah Pembelian/C1 = 0,28/5
0,167 = Jumlah Transaksi/C1 = 0,83/5
0,0689 = Jumlah Transaksi/C1 = 0,34/5

Gambar 2 Proses perhitungan centroid baru – iterasi ke 2

Pada perhitungan iterasi ke 2 dilakukan dengan menghitung nilai parameter berdasarkan dibandingkan dengan centroid baru. Berikut ini penulis sajikan pada contoh baris pertama yaitu

$$\text{Cluster 1} = \sqrt{(0,24-0,056)^2 + ((0,50-0,167)^2) + ((0,40-0,0689)^2)} = 0,50$$

$$\text{Cluster 2} = \sqrt{(0,24-0,251)^2 + ((0,50-0,528)^2) + ((0,40-0,2643)^2)} = 0,14$$

$$\text{Cluster 3} = \sqrt{(0,24-0,888)^2 + ((0,50-0,833)^2) + ((0,40-0,7314)^2)} = 0,80$$

Setelah perhitungan centroid baru selesai, maka hasil dari iterasi ke 2 seperti tampil pada Tabel 4.

Tabel 4 Hasil perhitungan iterasi ke 2

No	Cluster 1	Cluster 2	Cluster 3	Posisi Cluster	Jarak Terdekat	Pelanggan
1	0,50	0,14	0,80	2	0,14	Ahmad
2	0,17	0,34	1,17	1	0,17	Ardian
3	0,18	0,62	1,37	1	0,18	Bagas
4	0,05	0,43	1,22	1	0,05	Budianto
5	0,92	0,58	0,54	3	0,58	Cerly
6	0,43	0,25	0,84	2	0,25	Farid
7	1,33	0,97	0,32	3	0,32	Fitriani
8	0,35	0,19	1,03	2	0,19	Galih
9	0,50	0,25	1,01	2	0,25	Gading
10	0,63	0,25	0,75	2	0,25	Herianto
11	1,51	1,09	0,27	3	0,27	Hermawan
12	0,40	0,32	1,01	2	0,32	Ikah
13	0,38	0,15	0,98	2	0,15	Joko
14	0,26	0,24	1,04	2	0,24	Kamal
15	0,06	0,42	1,21	1	0,06	Kurnia
16	0,40	0,15	0,96	2	0,15	Lutfy
17	0,86	0,40	0,53	2	0,40	Mona
18	1,03	0,63	0,45	3	0,45	Nasir
19	0,36	0,17	1,02	2	0,17	Opta
20	0,07	0,50	1,32	1	0,07	Prabu

Berdasarkan hasil iterasi ke 1 maka diperoleh jumlah Cluster C1 = 5, C2 = 11 dan C3 = 4 (Iterasi ke 2 ini belum valid dengan dibandingkan hasil iterasi ke 1, untuk itu dilanjutkan perhitungan iterasi ke 3 dengan cara menentukan nilai centroid baru seperti disajikan pada Gambar 3.

Centeroid Baru			
	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang
C1	0,056	0,167	0,0689
C2	0,200	0,515	0,2681
C3	0,869	0,792	0,6044

0,056 => Jumlah Pembelian/C1 = 0,28/5
0,167 => Jumlah Transaksi/C1 = 0,83/5
0,0689=> Jumlah Transaksi/C1 = 0,34/5

Gambar 3 Proses perhitungan centroid baru – iterasi ke 3

Pada perhitungan iterasi ke 2 dilakukan dengan menghitung nilai parameter berdasarkan dibandingkan dengan centroid baru. Berikut ini penulis sajikan pada contoh baris pertama yaitu

$$\text{Cluster 1} = \sqrt{(0,24-0,056)^2 + ((0,50-0,167)^2) + ((0,40-0,0689)^2)} = 0,50$$

$$\text{Cluster 2} = \sqrt{(0,24-0,200)^2 + ((0,50-0,515)^2) + ((0,40-0,2681)^2)} = 0,14$$

$$\text{Cluster 3} = \sqrt{(0,24-0,869)^2 + ((0,50-0,792)^2) + ((0,40-0,6044)^2)} = 0,72$$

<http://sistemasi.ftik.unisi.ac.id>

Setelah perhitungan centroid baru selesai, maka hasil dari iterasi ke 3 seperti tampil pada Tabel 4.

Tabel 4 Hasil perhitungan iterasi 3

No	Cluster 1	Cluster 2	Cluster 3	Posisi Cluster	Jarak Terdekat	Pelanggan
1	0,50	0,14	0,72	2	0,14	Ahmad
2	0,17	0,30	1,07	1	0,30	Ardian
3	0,18	0,59	1,27	1	0,18	Bagas
4	0,05	0,40	1,13	1	0,05	Budianto
5	0,92	0,63	0,41	3	0,63	Cerly
6	0,43	0,28	0,74	2	0,28	Farid
7	1,33	1,00	0,41	3	0,41	Fitriani
8	0,35	0,15	0,94	2	0,15	Galih
9	0,50	0,24	0,92	2	0,24	Gading
10	0,63	0,30	0,64	2	0,30	Herianto
11	1,51	1,13	0,39	3	0,39	Hermawan
12	0,40	0,28	0,94	2	0,28	Ikah
13	0,38	0,10	0,89	2	0,10	Joko
14	0,26	0,20	0,95	2	0,20	Kamal
15	0,06	0,39	1,11	1	0,06	Kurnia
16	0,40	0,10	0,87	2	0,10	Lutfy
17	0,86	0,44	0,46	2	0,44	Mona
18	1,03	0,68	0,32	3	0,32	Nasir
19	0,36	0,13	0,92	2	0,13	Opta
20	0,07	0,47	1,22	1	0,07	Prabu

Berdasarkan hasil iterasi ke 1 maka diperoleh jumlah Cluster C1 = 5, C2 = 11 dan C3 = 4 (Iterasi ke 3 ini telah sama atau sesuai dengan iterasi ke 2) sehingga dengan demikian dataset ini telah valid. Hasil Klasterisasi Segementasi pelanggan seperti disajikan pada Tabel 5.

Tabel 5 Hasil akhir klasterisasi sagmentasi pelanggan

No	No. ID	Nama Pelanggan	Alamat	Klaster	Keterangan
1	2	Ardian	Kaliangkrik	C1	Kurang Rajin Belanja
2	3	Bagas	Kajoran	C1	Kurang Rajin Belanja
3	4	Budianto	Dukun	C1	Kurang Rajin Belanja
4	15	Kurnia	Ngablak	C1	Kurang Rajin Belanja
5	20	Prabu	Ngluwar	C1	Kurang Rajin Belanja
No	No. ID	Nama Pelanggan	Alamat	Klaster	Keterangan
1	1	Ahmad	Dukun	C2	Cukup Rajin Belanja
2	6	Farid	Bandongan	C2	Cukup Rajin Belanja
3	8	Galih	Kaliangkrik	C2	Cukup Rajin Belanja
4	9	Gading	Ngablak	C2	Cukup Rajin Belanja
5	10	Herianto	Bandongan	C2	Cukup Rajin Belanja
6	12	Ikah	Salam	C2	Cukup Rajin Belanja
7	13	Joko	Bandongan	C2	Cukup Rajin Belanja
8	14	Kamal	Tegalrejo	C2	Cukup Rajin Belanja
9	16	Lutfy	Pakis	C2	Cukup Rajin Belanja
10	17	Mona	Sawangan	C2	Cukup Rajin Belanja
11	19	Opta	Pakis	C2	Cukup Rajin Belanja
No	No. ID	Nama Pelanggan	Alamat	Klaster	Keterangan
1	5	Cerly	Kajoran	C3	Sangat Rajin Belanja
2	7	Fitriani	Dukun	C3	Sangat Rajin Belanja
3	11	Hermawan	Secang	C3	Sangat Rajin Belanja
4	18	Nasir	Ngluwar	C3	Sangat Rajin Belanja

Berdasarkan hasil klasterisasi dari Tabel 5 maka dapat diringkas bahwa C1 (Kurang Rajin Belanja) yaitu 4 orang sedangkan C2 (Cukup Rajin Belanja) ada 11 orang dan C3 Sangat Rajin Belanja yaitu 4 orang.

B. Implementasi Sistem

Pada Implementasi sistem merupakan bagian dari pengembangan sistem kedalam suatu program aplikasi. Sistem dibuat dengan menggunakan bahasa Pemrograman PHP dan Database MySQL. Adapun tampilan program Aplikasi dari Sistem Klasterisasi Sagmentasi Pelanggan dengan menggunakan Metode K-Means Clustering sebagai berikut:

1. Tampilan Halaman Utama

Program Aplikasi Klasterisasi Sagmentasi Pelanggan ini ketika dijalankan maka akan tampil halaman awal sistem seperti disajikan pada Gambar 4.



Gambar 4 Tampilan halaman utama aplikasi

2. Tampilan Data Pelanggan

Pada aplikasi ini memiliki menu untuk mengelola data pelanggan yang digunakan sebagai objek dalam klasterisasi sagmentasi pelanggan. Adapun tampilan data pelanggan di program aplikasi seperti disajikan pada Gambar 5.

No	Nama Pelanggan	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang	Jenis Pelanggan	Alamat
1	Ahmad	350500.00	4	120	Online	Dukun
2	Ardian	68500.00	3	30	Offline	Kaliangkrik
3	Bagas	80000.00	1	11	Offline	Kajoran
4	Budianto	74000.00	2	45	Online	Dukun
5	Cerly	513500.00	5	73	Offline	Kajoran
6	Farid	278000.00	3	80	Online	Bandongan
7	Fitriani	543500.00	5	285	Offline	Dukun
8	Galih	83500.00	4	60	Online	Kaliangkrik
9	Gading	98500.00	5	45	Online	Ngablak
10	Herianto	291500.00	5	55	Offline	Bandongan

Gambar 5 Tampilan data pelanggan

3. Tampilan Halaman Login Sistem

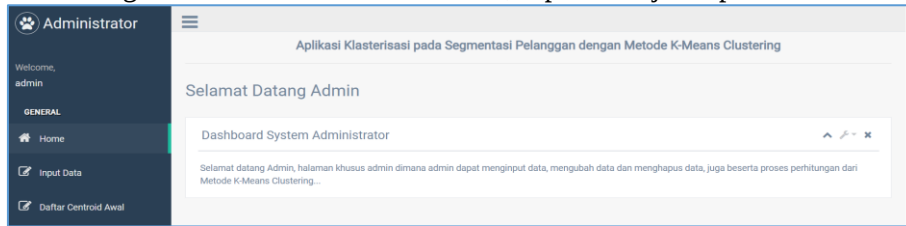
Pada Aplikasi ini telah dilengkapi halaman login aplikasi yang digunakan untuk masuk kedalam sistem. Tampilan halaman login aplikasi seperti disajikan pada Gambar 6.

Gambar 6 Tampilan halaman login aplikasi

<http://sistemasi.ftik.unisi.ac.id>

4. Tampilan Halaman Administrator

Apabila telah berhasil login, maka akan tampil langsung dihalaman dashboard sistem yang berfungsi untuk mengelola sistem secara keseluruhan, seperti disajikan pada Gambar 7.



Gambar 7 Tampilan halaman administrator

5. Tampilan Halaman Input Data Pelanggan

Pada Aplikasi ini telah dilengkapi fitur untuk menginputkan data pelanggan kedalam sistem. Input data pelanggan ini berguna sebagai bagian data dasar untuk melakukan proses klasterisasi pada segmentasi pelanggan, adapun tampilan dari halaman input data pelanggan disajikan di Gambar 8.

Gambar 8 Tampilan halaman input data pelanggan

Setelah data input data pelanggan selesai diinputkan kedalam aplikasi maka data pelanggan tersebut akan tersimpan kedalam sistem seperti daftar pelanggan pada Gambar 9.

#	Nama Pelanggan	Alamat	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang	Jenis Pelanggan	
1	Ahmad	Dukun	350500.00	4	120	Online	Edit Delete
2	Ardian	Kaliangkrik	68500.00	3	30	Offline	Edit Delete
3	Bagas	Kajoran	80000.00	1	11	Offline	Edit Delete
4	Budianto	Dukun	74000.00	2	45	Online	Edit Delete
5	Cerly	Kajoran	513500.00	5	73	Offline	Edit Delete
6	Farid	Bandongan	278000.00	3	80	Online	Edit Delete
7	Fitriani	Dukun	543500.00	5	285	Offline	Edit Delete
8	Galih	Kaliangkrik	83500.00	4	60	Online	Edit Delete
9	Gading	Ngablak	98500.00	5	45	Online	Edit Delete
10	Herianto	Bandongan	291500.00	5	55	Offline	Edit Delete
11	Hermawan	Secang	625000.00	7	258	Online	Edit Delete
12	Ikah	Salam	73000.00	3	135	Online	Edit Delete

Gambar 9 Tampilan daftar data pelanggan

6. Tampilan Halaman Setting Centroid

Pada sistem ini terdapat menu untuk memberikan nilai centroid awal yang digunakan sebagai pendukung dalam proses perhitungan algoritma k-means clustering. Tampilan setting nilai centroid awal seperti disajikan pada Gambar 10.

#	Kode Data	Centroid	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang	
1	11	1	0.06	0.170	0.03	Edit
2	15	2	0.11	0.330	0.25	Edit
3	19	3	0.86	0.830	1.00	Edit

Gambar 10 Tampilan daftar data pelanggan

7. Proses Klaster Pelanggan dengan Algoritma K-Means Clustering

Pada menu inti dari aplikasi ini adalah proses clusterisasi dengan menggunakan metode Algoritma K-Means Clustering, caranya klik menu K-Means Clustering sehingga tampil di Gambar 11.

Proses

Ini section hasil

C1 - Pelanggan Kurang Rajin Belanja

C2 - Pelanggan Cukup Rajin Belanja

C3 - Pelanggan Rajin Belanja

© 2025 STMIK Batira Magelang | Kapiti, S.Kom., M.Kom | NIDN : 0615018103 | Dwi Astuti, S.Kom, M.Kom | NIDN : 0604038903

Gambar 11 Tampilan halaman proses cluster data pelanggan

Setelah mengklik Proses maka secara otomatis aplikasi akan melakukan proses clusterisasi dengan perhitungan algoritma k-means cluster melalui beberapa iterasi seperti disajikan pada Gambar 12.

Proses

Iterasi ke 1

Centroid Awal :

Cluster	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang
1	0.06	0.179	0.33
2	0.11	0.330	0.25
3	0.86	0.830	1.00

Jarak cluster

id	Nama Pelanggan	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang	C1	C2	Jarak Terdekat
1	Ahmad	0.53	0.50	0.40	0.67	0.33	0.590 29
2	Jordan	0.06	0.33	0.07	0.16	0.33	1.170 16
3	Bagas	0.08	0.00	0.00	0.18	0.61	3.70 18
4	Budiarto	0.07	0.17	0.12	0.05	0.44	1.220 05
5	Carly	0.81	0.67	0.23	0.92	0.58	0.80 23
6	Fenri	0.41	0.33	0.25	0.43	0.24	0.640 26
7	Fritani	0.86	0.67	0.10	1.33	0.95	0.84 10
8	Guth	0.08	0.50	0.08	0.18	0.50	1.00 22
9	Gading	0.11	0.67	0.12	0.19	0.33	1.00 26
10	Herianto	0.43	0.67	0.16	0.63	0.24	0.750 24
11	Hermawan	1.00	1.00	0.90	1.51	1.01	0.260 26
12	Ikhak	0.06	0.33	0.45	0.41	0.31	0.10 34
13	Joko	0.11	0.50	0.24	0.38	0.17	0.980 17
14	Kamal	0.11	0.33	0.27	0.39	0.24	1.030 26
15	Kurnia	0.08	0.17	0.12	0.06	0.43	1.10 06
16	Luffy	0.11	0.50	0.29	0.40	0.17	0.950 17
17	Mona	0.43	0.83	0.43	0.85	0.33	0.530 39
18	Nasir	0.80	0.83	0.29	1.02	0.81	0.540 45
19	Opta	0.09	0.50	0.19	0.36	0.20	1.020 20
20	Prabu	0.00	0.17	0.03	0.07	0.51	1.310 07

Iterasi ke 2

Centroid Awal :

Cluster	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang
1	0.05800000000000001	0.168	0.06799999999999999
2	0.27499999999999997	0.5275	0.2675
3	0.8866666666666667	0.8333333333333334	0.73

Jarak cluster

id	Nama Pelanggan	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang	C1	C2	Jarak Terdekat
1	Ahmad	0.53	0.50	0.40	0.67	0.33	0.490 33
2	Jordan	0.06	0.33	0.07	0.16	0.33	1.170 16
3	Bagas	0.08	0.00	0.00	0.18	0.61	3.70 18
4	Budiarto	0.07	0.17	0.12	0.05	0.44	1.220 05
5	Carly	0.81	0.67	0.23	0.92	0.58	0.80 23
6	Fenri	0.41	0.33	0.25	0.43	0.24	0.640 26
7	Fritani	0.86	0.67	0.10	1.33	0.95	0.84 10
8	Guth	0.08	0.50	0.08	0.18	0.50	1.00 22
9	Gading	0.11	0.67	0.12	0.19	0.33	1.00 26
10	Herianto	0.43	0.67	0.16	0.63	0.24	0.750 24
11	Hermawan	1.00	1.00	0.90	1.51	1.01	0.260 26
12	Ikhak	0.06	0.33	0.45	0.41	0.31	0.10 34
13	Joko	0.11	0.50	0.24	0.38	0.17	0.980 17
14	Kamal	0.11	0.33	0.27	0.39	0.24	1.030 26
15	Kurnia	0.08	0.17	0.12	0.06	0.43	1.10 06
16	Luffy	0.11	0.50	0.29	0.40	0.17	0.950 17
17	Mona	0.43	0.83	0.43	0.85	0.33	0.530 39
18	Nasir	0.80	0.83	0.29	1.02	0.81	0.540 45
19	Opta	0.09	0.50	0.19	0.36	0.20	1.020 20
20	Prabu	0.00	0.17	0.03	0.07	0.51	1.310 07

Iterasi ke 3

Centroid Awal :

Cluster	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang
1	0.05800000000000001	0.168	0.06799999999999999
2	0.22636363636363638	0.5145454545454545	0.27909090909090909
3	0.8674999999999999	0.7925	0.605

Jarak cluster

id	Nama Pelanggan	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang	C1	C2	Jarak Terdekat
1	Ahmad	0.53	0.50	0.40	0.67	0.33	0.490 33
2	Jordan	0.06	0.33	0.07	0.16	0.33	1.170 16
3	Bagas	0.08	0.00	0.00	0.18	0.61	3.70 18
4	Budiarto	0.07	0.17	0.12	0.05	0.41	1.270 18
5	Carly	0.81	0.67	0.23	0.92	0.61	0.490 40
6	Fenri	0.41	0.33	0.25	0.43	0.24	0.760 26
7	Fritani	0.86	0.67	0.10	1.33	0.98	0.84 10
8	Guth	0.08	0.50	0.08	0.18	0.50	1.00 22
9	Gading	0.11	0.67	0.12	0.19	0.33	1.00 26
10	Herianto	0.43	0.67	0.16	0.63	0.28	0.640 28
11	Hermawan	1.00	1.00	0.90	1.51	1.11	0.360 38
12	Ikhak	0.06	0.33	0.45	0.41	0.31	0.10 34
13	Joko	0.11	0.50	0.24	0.38	0.12	0.890 12
14	Kamal	0.11	0.33	0.27	0.39	0.23	0.920 23
15	Kurnia	0.08	0.17	0.12	0.06	0.41	1.10 06
16	Luffy	0.11	0.50	0.29	0.40	0.15	0.970 15
17	Mona	0.43	0.83	0.43	0.85	0.33	0.530 39
18	Nasir	0.80	0.83	0.29	1.02	0.80	0.540 42
19	Opta	0.09	0.50	0.19	0.36	0.16	0.930 16
20	Prabu	0.00	0.17	0.03	0.07	0.48	1.210 07

Gambar 12 Tampilan halaman proses iterasi clustering

8. Hasil Clustering

Pada proses akhir dari sistem ini terbentuknya 3 klaster yaitu kategori C1 (Kurang Rajin Belanja) yaitu 4 orang sedangkan C2 (Cukup Rajin Belanja) ada 11 orang dan C3 Sangat Rajin Belanja yaitu 4 orang. Tampilan hasil klaster ini dibentuk berdasarkan proses perhitungan dengan algoritma k-means clustering seperti disajikan pada Gambar 13.

C1 - Pelanggan Kurang Rajin Belanja			C2 - Pelanggan Cukup Rajin Belanja		
Nama Pelanggan	Keanggotaan	Keterangan	Nama Pelanggan	Keanggotaan	Keterangan
Ardian	C1	KECIL	Ahmad	C2	SEDANG
Bagas	C1	KECIL	Farid	C2	SEDANG
Budiarto	C1	KECIL	Galih	C2	SEDANG
Kurnia	C1	KECIL	Herianto	C2	SEDANG
Prabu	C1	KECIL	Ikah	C2	SEDANG
			Joko	C2	SEDANG
			Kamal	C2	SEDANG
			Lutfy	C2	SEDANG
			Mona	C2	SEDANG
			Opta	C2	SEDANG
C3 - Pelanggan Rajin Belanja					
Nama Pelanggan	Keanggotaan	Keterangan			
Cerly	C3	BESAR			
Fitriani	C3	BESAR			
Hermawan	C3	BESAR			
Nasir	C3	BESAR			

Gambar 13 Tampilan hasil klasterisasi data pelanggan

C. Pengujian Sistem

Pengujian sistem merupakan salah satu tahapan dalam proses pengembangan perangkat lunak yang bertujuan untuk memastikan bahwa sistem telah berjalan sesuai dengan desain yang direncanakan. Tahapan ini difokuskan pada evaluasi menyeluruh terhadap sistem setelah seluruh komponen terintegrasi, guna menilai kelayakan sistem secara keseluruhan. Dalam penelitian ini, dilakukan tiga jenis pengujian, yaitu uji fungsionalitas sistem, uji validitas, dan uji silhouette coefficient.

1. Uji Fungsionalitas Sistem

Uji fungsionalitas sistem merupakan jenis pengujian yang bertujuan untuk memverifikasi bahwa setiap fitur dalam sistem beroperasi sesuai dengan spesifikasi yang telah ditentukan. Pengujian ini dilakukan untuk memastikan bahwa sistem dapat berjalan dengan baik tanpa mengalami kegagalan saat diimplementasikan, sebagaimana ditunjukkan pada Tabel 6.

Tabel 6 Rekapitulasi hasil uji fungsionalitas

No	Jenis Uji	Komponen Sistem yang diuji	Skenario Uji	Hasil yang diharapkan	Hasil yang dihasilkan	Satus Uji	Hasil Pengujian
1	Uji Normal	Form Login Admin	• Masukan username dan password dan dengan benar	• Tampil halaman menu utama admin	• Muncul pesan “Login Sukses, Selamat Datang admin” • Tampil halaman admin	Normal	Diterima
	Uji Salah	Form Login Admin	• Masukan username dan password dan dengan salah	• Muncul pesan kesalahan	• Muncul pesan “Code Salah!” • Tidak masuk admin	Normal	
2	Uji Normal	Form Input Data Pelanggan	• Masukan data obat secara lengkap dan benar	• Data tersimpan dengan baik dan benar	• Muncul pesan “Penyimpanan Berhasil” • Tersimpan dengan baik	Normal	Diterima
	Uji Salah	Form Input Data Pelanggan	• Masukan data obat secara tidak lengkap	• Tidak bisa menyimpan	• Tidak bisa disimpan	Normal	
3	Uji Normal	Form Input Data Pelanggan	• Masukan data pelanggan secara lengkap dan benar	• Data siswa tersimpan dengan baik dan benar	• Muncul pesan “Penyimpanan Berhasil” • Tersimpan dengan baik	Normal	Diterima
	Uji Salah	Form Input Data Pelanggan	• Masukan data pelanggan secara tidak lengkap	• Tidak bisa menyimpan	• Tidak bisa disimpan	Normal	
4	Uji Normal	Form Input Centroid Awal	• Masukan data Centroid Awal secara lengkap dan benar	• Data periode tersimpan dengan baik dan benar	• Muncul pesan “Penyimpanan Berhasil” • Tersimpan dengan baik	Normal	Diterima
	Uji Salah	Form Input Centroid Awal	• Masukan data Centroid Awal secara tidak lengkap	• Tidak bisa menyimpan	• Tidak bisa disimpan	Normal	
5	Uji Normal	Proses Cluster	• Masukan data cluster secara lengkap dan benar	• Data kuota tersimpan dengan baik dan benar	• Muncul pesan “Penyimpanan Berhasil” • Tersimpan dengan baik	Normal	Diterima
	Uji Salah	Proses Cluster	• Masukan data cluster secara tidak lengkap	• Tidak bisa menyimpan	• Tidak bisa disimpan	Normal	

Berdasarkan hasil pengujian uji fungsionalitas yang telah dilakukan maka dapat diambil kesimpulan bahwa seluruh scenario uji menghasilkan hasil “Diterima” artinya sistem telah diterima berfungsi dengan baik.

2. Uji Validitas

Pengujian Validitas digunakan untuk menguji sistem dengan cara membandingkan antara hasil perhitungan manual dengan hasil dari program aplikasi. Tahap Perhitungan Algoritma K-Means Clustering sebagai berikut:

- 1) Menentukan Dataset
- 2) Menentukan Nilai MIN dan MAX pada masing-masing Variabel berikut:
MIN => Jumlah Pembelian: 17500 || Jumlah Transaksi: 1 || Jumlah Barang: 5,5
MIN => Jumlah Pembelian: 312500 || Jumlah Transaksi: 7 || Jumlah Barang: 142
- 3) Menghitung Nilai Normalisasi

Berikut ini contoh perhitungan Normalisasi pada baris pertama.

$$\text{Jumlah Pembelian} = (87,625 - 17500) / (7 - 1) = 0,24$$

$$\text{Jumlah Transaksi} = (4 - 1) / (7 - 1) = 0,50$$

$$\text{Jumlah Barang} = (60 - 5,5) / (142 - 5,5) = 0,40$$

Adapun rekam proses perhitungan normalisasi ditunjukkan pada Tabel 7.

Tabel 7 Data normalisasi

No	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang
1	0,24	0,50	0,40
2	0,06	0,33	0,07
3	0,08	0,00	0,00
4	0,07	0,17	0,12
5	0,81	0,67	0,22
6	0,41	0,33	0,25
7	0,86	0,67	1,00
8	0,08	0,50	0,18
9	0,11	0,67	0,11
10	0,43	0,67	0,16
11	1,00	1,00	0,90
12	0,06	0,33	0,43
13	0,11	0,50	0,25
14	0,11	0,33	0,26
15	0,08	0,17	0,13
16	0,11	0,50	0,29
17	0,45	0,83	0,44
18	0,80	0,83	0,29
19	0,09	0,50	0,19
20	0,00	0,17	0,03

Pada pembahasan ini peneliti langsung berfokus pada hasil klasterisasi agar dapat lebih fokus pada hasil penelitian.

- 4) Hasil Klastering
- Berdasarkan proses perhitungan algoritma k-means clustering untuk mengelompokkan data pelanggan maka dapat direkap hasil clustering dari data pelanggan disajikan di Tabel 8.

Tabel 8 Hasil klaster data pelanggan

No	Cluster 1	Cluster 2	Cluster 3	Posisi Cluster	Jarak Terdekat	Nama Pelanggan
1	0,50	0,14	0,72	2	0,14	Ahmad
2	0,17	0,30	1,07	1	0,30	Ardian
3	0,18	0,59	1,27	1	0,18	Bagas
4	0,05	0,40	1,13	1	0,05	Budianto
5	0,92	0,63	0,41	3	0,63	Cerly
6	0,43	0,28	0,74	2	0,28	Farid
7	1,33	1,00	0,41	3	0,41	Fitriani

8	0,35	0,15	0,94	2	0,15	Galih
9	0,50	0,24	0,92	2	0,24	Gading
10	0,63	0,30	0,64	2	0,30	Herianto
11	1,51	1,13	0,39	3	0,39	Hermawan
12	0,40	0,28	0,94	2	0,28	Ikah
13	0,38	0,10	0,89	2	0,10	Joko
14	0,26	0,20	0,95	2	0,20	Kamal
15	0,06	0,39	1,11	1	0,06	Kurnia
16	0,40	0,10	0,87	2	0,10	Lutfy
17	0,86	0,44	0,46	2	0,44	Mona
18	1,03	0,68	0,32	3	0,32	Nasir
19	0,36	0,13	0,92	2	0,13	Opta
20	0,07	0,47	1,22	1	0,07	Prabu

Berdasarkan hasil klasterisasi dari implementasi algoritma k-means clustering maka dapat diambil kesimpulan bahwa C1 : Pelanggan Kurang Rajin Belanja = 5 data, C2 Pelanggan Cukup Rajin Belanja = 11 data dan Pelanggan Rajin Belanja = 4 data. Hasil cluster ini telah sama dan sesuai program aplikasi maka model ini dinyatakan valid.

3. Uji Silhouette Coefficient

Silhouette coefficient merupakan metrik evaluasi yang digunakan dalam analisis kluster untuk menentukan jumlah kluster yang optimal dalam proses data mining. Metrik ini menilai seberapa baik suatu objek berada dalam kluster yang sesuai dibandingkan dengan kedekatannya terhadap kluster lainnya. Rekapitulasi hasil perhitungan uji Silhouette coefficient disajikan pada Tabel 9.

Tabel 9 Rekapitulasi uji silhouette coefficient

No	Jumlah Pembelian	Jumlah Transaksi	Jumlah Barang	ai	bi	si
1	0,5034	0,1373	0,7249	0,4815	0,4552	0,9737
2	0,1667	0,3050	1,0750	0,5593	0,5155	0,9562
3	0,1815	0,5938	1,2729	0,7238	0,6827	0,9589
4	0,0531	0,4014	1,1264	0,5684	0,5269	0,9585
5	0,9190	0,6309	0,4051	0,6792	0,6517	0,9725
6	0,4342	0,2795	0,7366	0,5090	0,4834	0,9744
7	1,3293	0,9982	0,4149	0,8709	0,9142	1,0433
8	0,3522	0,1483	0,9404	0,5251	0,4803	0,9552
9	0,5041	0,2402	0,9183	0,5909	0,5542	0,9633
10	0,6337	0,3003	0,6354	0,5459	0,5232	0,9773
11	1,5116	1,1314	0,3884	0,9565	1,0105	1,0540
12	0,3965	0,2780	0,9424	0,5741	0,5390	0,9649
13	0,3808	0,0965	0,8907	0,5020	0,4560	0,9540
14	0,2589	0,2042	0,9529	0,5135	0,4720	0,9585
15	0,0637	0,3946	1,1142	0,5653	0,5242	0,9589
16	0,4030	0,0962	0,8739	0,5027	0,4577	0,9550
17	0,8556	0,4360	0,4565	0,5960	0,5827	0,9867
18	1,0252	0,6816	0,3246	0,7021	0,6771	0,9750
19	0,3582	0,1302	0,9239	0,5158	0,4708	0,9550
20	0,0704	0,4693	1,2168	0,6289	0,5855	0,9566
					SC =>	0,9726

Dari hasil uji Silhouette Coefficient (SC) menghasilkan nilai SC rata-rata sebesar 0,97 (strong structure) artinya model struktur tersebut memiliki tingkat struktur yang kuat.

5 Kesimpulan

Kesimpulan dari penelitian ini adalah Implementasi Algoritma K-Means Clustering mampu mengelompokkan Data Pelanggan kedalam tiga cluster yaitu pada cluster yaitu C1: Pelanggan Kurang Rajin Belanja = 5 data, C2 Pelanggan Cukup Rajin Belanja = 11 data dan Pelanggan Rajin Belanja = 4 data. Hasil uji validitas telah dinyatakan valid karena berdasarkan perbandingan antara perhitungan manual dengan program aplikasi telah diperoleh hasil yang sama. Hasil pengujian sistem yang telah dilakukan pada uji fungsionalitas diperoleh diperoleh hasil "Diterima" pada seluruh skenario uji hal ini memberikan tanda bahwa sistem telah berfungsi dengan baik dan terakhir pada uji model Silhouette Coefficient (SC) maka diperoleh hasil rata-rata yang baik yakni 0,97 hampir mendekati angka 1 sehingga sistem ini dinyatakan strong structure. Hal ini menunjukkan bahwa dataset dinyatakan kuat untuk dijadikan sebagai role model dalam segmentasi pelanggan berdasarkan Data Transaksi dan Preferensi Pembelian.

Ucapan Terima Kasih

Pada kesempatan ini Tim Peneliti mengucapkan terimakasih banyak pada Direktorat Penelitian dan Pengabdian Masyarakat (DPPM) di Lingkungan Kemendikti Saintek yang telah memberikan pendanaan penelitian dari program HIBAH Dikti Tahun 2025, Pada Skema Penelitian Dosen Pemula (PDP), semoga hasil penelitian ini dapat memberikan manfaat untuk masyarakat.

Referensi

- [1] D. Suprayitno, N. D. Irmadiani, M. Munizu, M. Muchayatin, I. Mawarni, S. Saktisyahputra, and E. Erwin, Manajemen pemasaran: Teori dan strategi. PT. Green Pustaka Indonesia, 2024.
- [2] E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe, "Comprehensive review of K-means clustering algorithms," *International Journal of Advanced Scientific Research and Engineering*, vol. 7, no. 8, pp. 64–69, 2021. [Online]. Available: <https://doi.org/10.31695/IJASRE.2021.34050>
- [3] Y. Zhou, R. Wang, R. Ding, D. Shi, and Q. Ye, "Investigation on hierarchical control for driving stability and safety of intelligent HEV during car-following and lane-change process," *Science China Technological Sciences*, vol. 65, no. 1, pp. 53–76, 2022. [Online]. Available: <https://doi.org/10.1007/s11431-021-1891-8>
- [4] V. Duarte, S. Zuniga-Jara, and S. Contreras, "Machine learning and marketing: A systematic literature review," *IEEE Access*, vol. 10, pp. 93273–93288, Aug. 2022. [Online]. Available: <http://doi.org/10.1109/ACCESS.2022.3202896>
- [5] X. H. Vázquez, "Adapting for sustainable success: Navigating change in strategic management," Doctoral Dissertation, 2023. [Online]. Available: <https://www.investigobiblioteca.uvigo.es/xmlui/handle/11093/5523>
- [6] L. Zhang and Z. Lu, "Advances in InSAR imaging and data processing—A review," *Remote Sensing*, vol. 14, no. 17, pp. 1–7, 2022. [Online]. Available: <https://doi.org/10.3390/rs14174307>
- [7] X. Sáez-de-Cámara, J. L. Flores, C. Arellano, A. Urbieta, and U. Zurutuza, "Clustered federated learning architecture for network anomaly detection in large scale heterogeneous IoT networks," *Computers & Security*, vol. 131, 2023. [Online]. Available: <https://doi.org/10.1016/j.cose.2023.103299>
- [8] P. Singh, L. Khoshaim, B. Nuwisher, and I. Alhassan, "How information technology (IT) is shaping consumer behavior in the digital age: A systematic review and future research directions," *Sustainability*, vol. 16, no. 4, 2024. [Online]. Available: <https://doi.org/10.3390/su16041556>
- [9] D. O. Hassan and B. A. Hassan, "A comprehensive systematic review of machine learning in the retail industry: Classifications, limitations, opportunities, and challenges," *Neural Computing and Applications*, vol. 37, no. 4, pp. 2035–2070, 2025. [Online]. Available: <https://doi.org/10.1007/s00521-024-10869-w>
- [10] E. Dritsas and M. Trigka, "Exploring the intersection of machine learning and big data: A survey," *Machine Learning and Knowledge Extraction*, vol. 7, no. 1, 2025. [Online]. Available: <https://doi.org/10.3390/make7010013>
- [11] S. Jabeen, "The landscape of CRM research: A bibliometric analysis of key trends and future

- directions,” *Journal of Relationship Marketing*, pp. 1–41, 2025. [Online]. Available: <https://doi.org/10.1080/15332667.2025.2462884>
- [12] S. Zhou, S. Bi, and G. Qi, “Rule mining trends from 1987 to 2022: A bibliometric analysis and visualization,” *Data Intelligence*, pp. 1–44, 2023. [Online]. Available: https://doi.org/10.1162/dint_a_00239
- [13] H. Liu, Y. Luo, J. Geng, and P. Yao, “Research hotspots and frontiers of product R&D management under the background of the digital intelligence era—Bibliometrics based on CiteSpace and HistCite,” *Applied Sciences*, vol. 11, no. 15, 2021. [Online]. Available: <https://doi.org/10.3390/app11156759>
- [14] J. Yuan, C. Chen, W. Yang, M. Liu, J. Xia, and S. Liu, “A survey of visual analytics techniques for machine learning,” *Computational Visual Media*, vol. 7, no. 1, pp. 3–36, 2021. [Online]. Available: <http://doi.org/10.1007/s41095-020-0191-7>
- [15] J. Yang, Y. K. Wang, X. Yao, and C. T. Lin, “Adaptive initialization method for K-means algorithm,” *Frontiers in Artificial Intelligence*, vol. 4, pp. 1–13, Nov. 2021. [Online]. Available: <https://doi.org/10.3389/frai.2021.740817>
- [16] J. Yang and S. Rahardja, “Dimensional outlier detection,” 2024. [Online]. Available: <https://doi.org/10.36227/techrxiv.171172948.85055456/v1>
- [17] A. Kumar, R. Saini, and R. Kumar, “A comparative analysis of machine learning algorithms for breast cancer detection and identification of key predictive features,” *Traitement du Signal*, vol. 41, no. 1, pp. 127–140, 2024. [Online]. Available: <https://doi.org/10.18280/ts.410110>
- [18] V. Gallego, A. Freixes, and J. Ligan, “Applying machine learning in marketing: An analysis using the NMF and K-means algorithms,” *Lecture Notes in Computer Science*, vol. 14778, pp. 14–26, 2024. [Online]. Available: <https://doi.org/10.3390/info15070368>
- [19] A. Wasilewski, “Customer segmentation in e-commerce: A context-aware quality framework for comparing clustering algorithms,” *Journal of Internet Services and Applications*, vol. 15, no. 1, pp. 160–178, 2024. [Online]. Available: <https://doi.org/10.5753/jisa.2024.3851>
- [20] A. Takie, E. Selmi, M. F. Zerarka, and A. Cheriet, “Enhancing K-means clustering with post-redistribution,” *Ingénierie des Systèmes d’Information*, vol. 29, no. 2, pp. 429–436, 2024. [Online]. Available: <https://doi.org/10.18280/isi.290204>
- [21] N. M. Nhat, “Applied density-based clustering techniques for classifying high-risk customers: A case study of commercial banks in Vietnam,” *Journal of Applied Data Science*, vol. 5, no. 4, pp. 1639–1653, 2024. [Online]. Available: <https://doi.org/10.47738/jads.v5i4.344>
- [22] B. I. Nugroho, A. Raffhina, P. S. Ananda, and G. Gunawan, “Customer segmentation in sales transaction data using K-means clustering algorithm,” *Journal of Intelligent Decision Support System*, vol. 7, no. 2, pp. 130–136, 2024. [Online]. Available: <https://doi.org/10.35335/idss.v7i2.236>
- [23] S. Grünewälder, “Machine learning: What is machine learning?,” in *Proc. 2019 3rd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, vol. 11, no. 113, pp. 13–21, 2015. [Online]. Available: <https://doi.org/10.28945/5086>
- [24] A. Saxena, A. Agarwal, B. K. Pandey, and D. Pandey, “Examination of the criticality of customer segmentation using unsupervised learning methods,” *Circular Economy and Sustainability*, vol. 4, no. 2, pp. 1447–1460, 2024. [Online]. Available: <https://doi.org/10.1007/s43615-023-00336-4>
- [25] S. Hidayati, A. T. Darmaliana, and R. Riski, “Comparison of K-means, fuzzy C-means, fuzzy Gustafson Kessel, and DBSCAN for village grouping in Surabaya based on poverty indicators,” *Jurnal Pendidikan Matematika*, vol. 5, no. 2, p. 185, 2022. [Online]. Available: <https://doi.org/10.1007/s43615-023-00336-4>
- [26] C. H. Ardana, A. A. A. A. Khoyum, and M. Faisal, “Segmentasi pelanggan penjualan online menggunakan metode K-means clustering,” *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 1, pp. 1–9, 2024. [Online]. Available: <https://doi.org/10.14421/jiska.2024.9.1.1-9>
- [27] E. B. Firmansyah, M. R. Machado, and J. L. R. Moreira, “How can artificial intelligence (AI) be used to manage customer lifetime value (CLV)—A systematic literature review,” *International Journal of Information Management Data Insights*, vol. 4, no. 2, p. 100279, 2024. [Online]. Available: <https://doi.org/10.1016/j.ijimei.2024.100279>