

Clustering Data Konsumen E-Commerce Menggunakan Algoritma K-Means Dan Dataset Kaggle

E-Commerce Consumer Data Clustering Using K-Means Algorithm And Kaggle Dataset

Ade Eka Sabrina^{*1}, Muhammad Yasin²,
^{1,2}Teknik Informatika, Universitas Asahan
E-mail: ^{1,2}adeeka59@gmail.com

Received: 22-11-2024 | Revised: 24-3-2025 | Accepted: 30-04-2025

Abstrak

Belanja *online* telah menjadi bagian gaya hidup masyarakat. Hal tersebut dikarenakan banyaknya kemudahan yang didapat untuk memenuhi kebutuhan primer sampai dengan tersier. Kondisi saat ini, data riwayat transaksi pembelian konsumen belum dimanfaatkan dengan maksimal sehingga kurang berdaya guna. Hal tersebut membuat perusahaan dirasa belum maksimal dalam memenuhi harapan pelanggan dalam peningkatan loyalitas konsumen. Selain itu, strategi pemasaran saat ini juga dirasa belum efektif karena penawaran yang dilakukan oleh perusahaan kepada setiap konsumen masih bersifat umum, perusahaan belum menawarkan produk ataupun promosi yang benar-benar dibutuhkan oleh konsumen. Melalui penelitian ini, penulis berusaha memberikan solusi kepada perusahaan untuk meningkatkan efisiensi strategi pemasaran yang sangat berpengaruh terhadap peningkatan loyalitas konsumen dengan cara melakukan analisis clustering data konsumen e-commerce. Tujuan penelitian ini adalah untuk merancang dan membuat aplikasi aplikasi clustering data konsumen e-commerce menggunakan algoritma k-means dan dataset kaggle. Data yang digunakan dalam penelitian ini adalah data konsumen *e-commerce*. Hasil *output* penyelesaian masalah *clustering* data konsumen *e-commerce*. Dapat dilihat berdasarkan hasil pengelompokan data pembayaran untuk Cluster 1 (Pembayaran Elektronik): Total Pembayaran: 72,000,000.00, Cluster 2 (Pembayaran Sepatu): Total Pembayaran: 6,600,000.00, Cluster 3 (Pembayaran Baju): Total Pembayaran: 2,940,000.00 dan Cluster 4 (Pembayaran Kosmetik): Total Pembayaran: 105,000,000.00.

Kata kunci: Clustering, Data Konsumen, E-Commerce, Algoritma K-Means, Dataset Kaggle, Web.

Abstract

Online shopping has become a part of people's lifestyle. This is because of the many conveniences obtained to meet primary to tertiary needs. The current condition, consumer purchase transaction history data has not been utilized optimally so that it is less effective. This makes the company feel that it has not been optimal in meeting customer expectations in increasing consumer loyalty. In addition, the current marketing strategy is also considered ineffective because the offers made by the company to each consumer are still general, the company has not offered products or promotions that are really needed by consumers. Through this study, the author tries to provide solutions to companies to increase the efficiency of marketing strategies that greatly influence increasing consumer loyalty by conducting clustering analysis of e-commerce consumer data. The purpose of this study is to design and create an application for clustering e-commerce consumer data using the k-means algorithm and the kaggle dataset. The data used in this study is e-commerce consumer data. The output results of solving the problem of clustering e-commerce

consumer data. It can be seen based on the results of the payment data grouping for Cluster 1 (Electronic Payment): Total Payment: 72,000,000.00, Cluster 2 (Shoe Payment): Total Payment: 6,600,000.00, Cluster 3 (Clothing Payment): Total Payment: 2,940,000.00 and Cluster 4 (Cosmetic Payment): Total Payment: 105,000,000.00.

Keywords Clustering, Consumer Data, E-Commerce, K-Means Algorithm, Kaggle Dataset, Web.

1. PENDAHULUAN

Penggunaan teknologi sangat memberikan manfaat yang besar terhadap dunia bisnis. Perusahaan yang menggunakan teknologi merupakan perusahaan yang mampu bersaing dalam kompetisi. Salah satu teknologi yang dapat meningkatkan persaingan bisnis adalah penggunaan *electronic commerce* (*e-commerce*), dimana *e-commerce* digunakan untuk memasarkan berbagai macam produk maupun jasa baik dalam bentuk fisik maupun digital [1].

E-commerce merupakan cara untuk menjual suatu produk secara *online* dengan memanfaatkan internet. Internet merupakan suatu jaringan komputer yang saling terhubung dan digunakan untuk keperluan informasi dan komunikasi.[1]. *E-Commerce* atau juga dikenal sebagai perdagangan elektronik merupakan salah satu bisnis *online* sebagai penyebaran, pembelian, serta pemasaran barang dan jasa dengan sistem eletronik melalui jaringan.[2]

Belanja *online* telah menjadi bagian gaya hidup masyarakat. Hal tersebut dikarenakan banyaknya kemudahan yang didapat untuk memenuhi kebutuhan primer sampai dengan tersier. Terdapat beberapa alasan masyarakat yang lebih memilih melakukan pembelian secara *online* yaitu karena adanya kemudahan untuk berbelanja yang dapat dilakukan dimana saja, lebih cepat dan praktis hanya dengan memanfaatkan *smartphone* sebagai media dalam melakukan transaksi pembelian. Konsumen tidak perlu lagi datang langsung ke toko, namun semua dapat dilakukan dimanapun dengan cara yang sangat mudah yaitu mengunduh aplikasi belanja *online*.

Kondisi saat ini, data riwayat transaksi pembelian konsumen belum dimanfaatkan dengan maksimal sehingga kurang berdaya guna. Hal tersebut membuat perusahaan dirasa belum maksimal dalam memenuhi harapan pelanggan dalam peningkatan loyalitas konsumen. Selain itu, strategi pemasaran saat ini juga dirasa belum efektif karena penawaran yang dilakukan oleh perusahaan kepada setiap konsumen masih bersifat umum, perusahaan belum menawarkan produk ataupun promosi yang benar-benar dibutuhkan oleh konsumen.

Clustering banyak digunakan diberbagai bidang seperti biologi, psikologi, dan ekonomi. Hasil pengelompokan bervariasi karena jumlah perubahan parameter *cluster* maka tantangan utama analisis *cluster* adalah jumlah *cluster* atau jumlah parameter model jarang diketahui, dan harus ditentukan sebelum pengelompokan.[3]. *Clustering* atau analisis kelompok merupakan proses mengelompokkan data (objek) berdasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya.[4]

Melalui penelitian ini, penulis berusaha memberikan solusi kepada perusahaan untuk meningkatkan efisiensi strategi pemasaran yang sangat berpengaruh terhadap peningkatan loyalitas konsumen dengan cara melakukan analisis *clustering* data konsumen *e-commerce*. Analisis ini mampu memberikan gambaran tentang karakteristik konsumen di beberapa *cluster* yang berbeda. Konsumen dikelompokkan menjadi beberapa *cluster* yang dibedakan berdasarkan perilakunya dalam melakukan transaksi pembelian.

Platform Kaggle.com menyediakan berbagai sumber data publik untuk data penelitian. Dengan menyediakan akses terbuka terhadap berbagai jenis dataset untuk digunakan dalam proyek penelitian dan kompetisi, Kaggle adalah komunitas *online* terkemuka untuk para praktisi data science. [5] dalam penelitian ini data bersumber dari Platform Kaggle.com.

Penelitian terdahulu yang pernah dilakukan oleh [6] dengan judul (Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada CV Digital Dimensi). Hasil pengujian klaster menggunakan Davies Bouldin Index mempunyai nilai 0.5021462639132475. Dari percobaan klaster tiga sampai dengan delapan, klaster lima adalah yang nilainya paling mendekati angka 0.

Penelitian terdahulu yang pernah dilakukan oleh [7] dengan judul (Analisis Data Produk Elektronik Di *E-Commerce* Dengan Metode Algoritma K-Means Menggunakan Python). Hasil dari

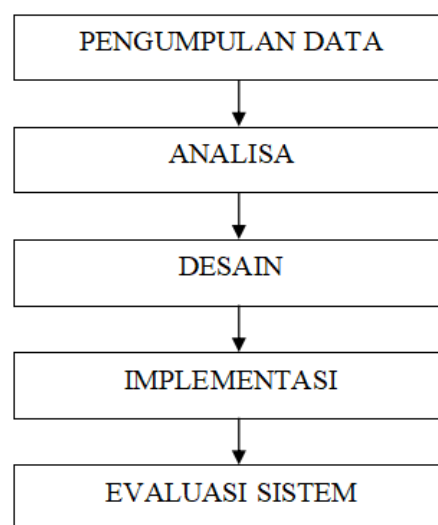
penelitian didapatkan pada analisis teks komentar produk dengan *wordcloud* produk *Smartphone low cost* menunjukkan data komentar *marketplace shopee* Indonesia dapat bahwa baik di *smartphone low cost* maupun yang *high cost* cenderung memiliki pola *wordcloud* yang sama dimana kata-kata yang dominan muncul cenderung netral dan positif, sedang kata-kata yang bermakna negatif cenderung tidak dominan. Sementara kata yang sering muncul yaitu barang, mantap, cepat, kirim dan bagus dengan nilai akurasi presentase sebesar 92%. Sedangkan hasil proses *wordcloud medium high cost* diperoleh kata yang sering muncul ialah kata (kirim, cepat, dan bagus) dengan nilai akurasi dengan presentase 94%. Serta berdasarkan hasil grafik dari proses *clustering* data k-means menunjukkan bahwa angka penjualan 0 sampai 1000 mendapatkan skor rating bintang tertinggi dan penjualan dengan skor rating bintang terendah ialah antara 1500 sampai 2000 ke atas.

Penelitian terdahulu yang pernah dilakukan oleh [8] dengan judul (Analisis Segmentasi Pelanggan Menggunakan Kmeans Clustering Studi Kasus Aplikasi Alfagift). Algoritma K-Means digunakan untuk membentuk kluster dan metode *Elbow* digunakan untuk mengevaluasi kluster-kluster yang terbentuk. Algoritma K-means digunakan untuk pembentukan kluster dan metode *Elbow* digunakan untuk mengevaluasi kluster-kluster yang terbentuk. Hasil k optimum yang terbentuk dari hasil perhitungan SSE (*Sum Squared Error*) dengan metode *Elbow* adalah tiga *cluster* dengan selisih nilai SSE terbesar yaitu 1119.28644. Hasil pengelompokan pelanggan sebanyak tiga *cluster*; *cluster* pertama berjumlah 7.219 pelanggan, *cluster* kedua sebanyak 6.902 pelanggan, dan *cluster* ketiga sebanyak 5.371 pelanggan. Hasil K-Means *clustering* tersebut selanjutnya diinterpretasikan sebagai data yang akan digunakan dalam menentukan strategi pemasaran yang tepat.

K-Means adalah salah satu metode analisis data yang digunakan untuk mengelompokkan data menjadi beberapa kelompok yang saling berhubungan berdasarkan kesamaan atribut. Metode ini bekerja dengan cara mendefinisikan pusat kluster (*centroid*) secara acak, kemudian menghitung jarak antara setiap data dengan *centroid* tersebut. [9]. K-Means merupakan metode non-hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat *cluster* awal [4]. K-means merupakan metode non-hirarki yang paling banyak digunakan untuk pengelompokan karena kemudahan dan kemampuannya untuk mengelompokkan dalam jumlah data yang besar dengan waktu komputasi yang cepat dan efisien. K-means melihat jarak antar objek terhadap nilai *centroid cluster* [10].

Melihat permasalahan di atas, perlu dikembangkan sebuah sistem *clustering* data konsumen *e-commerce* yang dapat menyelesaikan permasalahan yang terjadi. Dengan adanya sistem ini diharapkan perusahaan dapat mengetahui dan dapat segera mengambil tindakan terhadap strategi pemasaran. Melalui penelitian ini, penulis berinisiatif untuk membuat penelitian dengan judul

2. METODE PENELITIAN



Gambar 1. Kerangka Kerja Penelitian

2.1 Kerangka Kerja Penelitian

Untuk menyusun penelitian ini, dilakukan metode penelitian dalam memperoleh data-data yang dibutuhkan sehingga penyusunan penelitian ini dapat diselesaikan dengan baik dan sistematis.

2.2 Metode K-Means

Dalam pembuatan aplikasi ini, Metode yang digunakan penelitian ini adalah K-Means *clustering*, yang merupakan salah satu teknik data mining yang dapat digunakan untuk mengelompokkan berdasarkan kemiripan karakteristik tertentu dimana data-data yang memiliki kemiripan akan berada pada *cluster* yang sama.

Langkah-langkah yang dilakukan dalam *Algoritma K-Means* adalah:

- a. Tentukan jumlah *cluster* (K).
Dalam hal ini, peneliti tetapkan bahwa $K=3$
- b. Pilih titik acak sebanyak K.
Titik ini merupakan titik *seed* dan akan menjadi titik *centroid* proses pertama.
Titik ini tidak harus titik data kita.
- c. Label semua data berdasarkan titik *centroid* terdekat.
Semua data diberikan label mengikuti titik *centroid* dari setiap *klaster*.
Perhitungan jarak ini bisa menggunakan algoritma jarak tertentu, secara *default* dilakukan dengan *Euclidean distance*.
- d. Tentukan titik *centroid* baru berdasarkan *cluster* yang terbentuk.
Titik *centroid* selanjutnya “berpindah” ke lokasi *centroid* setiap *cluster* yang telah terbentuk.
- e. Label ulang data berdasarkan jarak terdekat terhadap *centroid* baru
Langkah ini merupakan langkah yang sama dengan langka ke 3. Pada tahap ini akan ada perubahan titik.
- f. Ulangi langka 4 dan langka 5 sampai tidak pergerakan lagi.
Ditahap ini algoritma akan mencari lokasi *centroid* baru dan melabel data berdasarkan *centroid* tersebut sampai mendapatkan hasil akhir, yaitu tidak ada lagi perpindahan *centroid* di setiap *cluster*.

2.3 Objek Penelitian

Objek penelitian merupakan permasalahan yang diteliti. Objek dari penelitian ini ialah yang berkaitan dengan *clustering* data konsumen *e-commerce* dengan pemanfaatan sistem *clustering* ini nantinya akan dapat membantu pihak perusahaan dalam melakukan pengelompokan data konsumen.

2.4 Metode Analisis Data

Metode analisis data yang dilakukan peneliti dalam bentuk analisis data kualitatif melalui pengolahan data sebagai prosedur penelitian yang bersikap deskriptif. Yaitu data yang diperoleh kemudian dianalisa dan dibandingkan dengan teori-teori dan kemudian dievaluasi. Hasil evaluasi tersebut yang akan ditarik sebagai kesimpulan untuk menjawab permasalahan yang muncul. Proses analisis data pada penelitian ini adalah sebagai berikut

1. Mengumpulkan seluruh data yang tersedia dari berbagai sumber, yaitu dari hasil dataset kaggle konsumen *E-Commerce*.
2. Reduksi data, yaitu dilakukan dengan jalan melakukan abstraksi. Abtraksi merupakan usaha membuat rangkuman inti, proses, dan pernyataan-pernyataan yang perlu dijaga

- sehingga tetap berada didalamnya. Reduksi data yang digunakan peneliti adalah merangkum data konsumen.
3. Menyusun data hasil reduksi, data tersebut disusun menjadi satuan-satuan yang kemudian dilakukan proses pengujian yang menggunakan algoritma K-Means.
 4. Menyajikan sistem dalam bentuk aplikasi clustering data konsumen *E-Commerce*, dan sebagainya serta uraian singkat yang menjelaskan hubungan antar masing-masing kategori.
 5. Penarikan kesimpulan dan verifikasi. Kesimpulan awal yang dikemukakan masih bersifat sementara dan akan berubah bila tidak ditemukan bukti-bukti yang kuat yang mendukung pada tahap pengumpulan data berikutnya. Akan tetapi, jika kesimpulan yang dikemukakan pada tahap awal didukung oleh bukti-bukti yang valid dan konsisten saat peneliti kembali mengumpulkan data, maka kesimpulan yang dikemukakan merupakan kesimpulan yang sudah kredibel atau dapat dipertanggungjawabkan.

3. HASIL DAN PEMBAHASAN

3.1 Analisa Masalah

Belanja *online* atau biasa disebut *e-commerce* telah menjadi bagian gaya hidup masyarakat. Hal tersebut dikarenakan banyaknya kemudahan yang didapat untuk memenuhi kebutuhan primer sampai dengan tersier. Terdapat beberapa alasan masyarakat yang lebih memilih melakukan pembelian secara *online* yaitu karena adanya kemudahan untuk berbelanja yang dapat dilakukan dimana saja, lebih cepat dan praktis hanya dengan memanfaatkan *smartphone* sebagai media dalam melakukan transaksi pembelian.

Data konsumen mencakup semua informasi yang diperoleh dari pelanggan atau calon pelanggan, yang meliputi preferensi, perilaku, usia, jenis kelamin, lokasi geografis pelanggan, riwayat pembelian, interaksi dengan merek, serta umpan balik. Data ini dapat berasal dari berbagai sumber, seperti data transaksi, yang mencatat riwayat pembelian dan interaksi pelanggan dengan produk atau layanan, serta data konsumen yang digunakan dalam penelitian ini berdasarkan usia, jenis kelamin dan lokasi geografis pelanggan. Dengan mayoritas pengguna berusia 25-34 tahun.

Melalui penelitian ini, peneliti melakukan klasifikasi produk yang dibeli oleh konsumen melalui website *e-commerce* yang bertujuan untuk melihat produk yang laris dibeli. Produk yang akan diklasifikasi seperti data pembelian baju, sepatu, bahan elektronik dan kosmetik kecantikan. Data pembelian di dapat dari situs kaggle, Platform Kaggle.com yang menyediakan berbagai sumber data publik untuk data penelitian. Dengan menyediakan akses terbuka terhadap berbagai jenis dataset untuk digunakan dalam proyek penelitian dan kompetisi. Analisis ini mampu memberikan gambaran tentang hasil klasifikasi transaksi pembelian produk yang paling diminati konsumen.

3.2 Analisa Kebutuhan Sistem

Analisa kebutuhan dilakukan untuk mengidentifikasi kebutuhan-kebutuhan dan komponen-komponen yang diperlukan untuk aplikasi clustering data konsumen *E-Commerce*, yang menggunakan algoritma K-Means. Dari analisa kebutuhan nantinya akan di gunakan untuk tahap perancangan sistem.

3.3 Kebutuhan Perangkat Keras (Hardware)

Perangkat keras yang digunakan dalam sistem ini memerlukan spesifikasi khusus yang harus dipenuhi. Hal ini bertujuan agar program dalam aplikasi yang dihasilkan dapat dipresentasikan dengan maksimal. Selain itu spesifikasi ini akan berpengaruh terhadap kinerja aplikasi yang akan digunakan. Spesifikasi *hardware* yang digunakan perlu memperhatikan beberapa hal sebagai berikut :

1. Kegiatan komputer di kantor dalam tahap awal tidak perlu adanya penyesuaian dalam pengembangan.
2. Pemilihan *hardware* dan *software* memperhatikan kebutuhan sekarang dan yang akan datang.
3. Adanya pertimbangan biaya yang minimal tetapi mencapai hasil yang relatif optimal. Untuk menjalankan aplikasi ini memang membutuhkan sebuah perangkat keras yang mendukung agar program tersebut berjalan sesuai dengan yang kita harapkan.

Perangkat keras tersebut meliputi :

- a. Processor Intel Core i3 4130/4150/4160
- b. Harddisk 500 GB
- c. Memory 4 GB
- d. LCD 4.1.2

3.4 Analisa Metode K-Means Clustering

Data yang akan dianalisis terdiri dari 48 sampel data pembayaran yang difilter dari data set kaggle pada link <https://www.kaggle.com/code/fitriarahmadani/e-commerce-customer-segmentation> dengan detail sebagai berikut:

Tabel 1. Data Transaksi Pembayaran

ID	Tanggal	Pelanggan	Kategori	Harga	Jlh	Total
1	28/01/2025	pelanggan 1	Bahan Elektronik	1000000.00	6	6000000.00
2	28/01/2025	pelanggan 2	Bahan Elektronik	1000000.00	6	6000000.00
3	28/01/2025	pelanggan 3	Bahan Elektronik	6500000.00	1	6500000.00
4	28/01/2025	pelanggan 4	Bahan Elektronik	6500000.00	1	6500000.00
5	28/01/2025	pelanggan 5	Bahan Elektronik	6500000.00	1	6500000.00
6	28/01/2025	pelanggan 6	Bahan Elektronik	6700000.00	1	6700000.00
7	28/01/2025	pelanggan 7	Bahan Elektronik	6700000.00	1	6700000.00
8	28/01/2025	pelanggan 8	Bahan Elektronik	6700000.00	1	6700000.00
9	28/01/2025	pelanggan 9	Bahan Elektronik	6700000.00	1	6700000.00

10	28/01/2025	pelanggan 10	Bahan Elektronik	6700000.00	1	6700000.00
11	28/01/2025	pelanggan 11	Bahan Elektronik	6700000.00	1	6700000.00
12	28/01/2025	pelanggan 12	Bahan Elektronik	6700000.00	1	6700000.00
13	28/01/2025	pelanggan 13	Sepatu	550000.00	1	550000.00
14	28/01/2025	pelanggan 14	Sepatu	510000.00	1	510000.00
15	28/01/2025	pelanggan 15	Sepatu	500000.00	1	500000.00
16	28/01/2025	pelanggan 16	Sepatu	520000.00	1	520000.00
17	28/01/2025	pelanggan 17	Sepatu	430000.00	1	430000.00
18	28/01/2025	pelanggan 18	Sepatu	480000.00	1	480000.00
19	28/01/2025	pelanggan 19	Sepatu	470000.00	1	470000.00
20	28/01/2025	pelanggan 20	Sepatu	574000.00	1	574000.00
22	13/03/2025	Pelanggan 21	Sepatu	554000.00	1	554000.00
23	13/03/2025	Pelanggan 22	Sepatu	554000.00	1	554000.00
24	13/03/2025	Pelanggan 23	Sepatu	584000.00	1	584000.00
25	13/03/2025	Pelanggan 24	Sepatu	534000.00	1	534000.00
26	13/03/2025	Pelanggan 25	Baju	245000.00	1	245000.00
27	13/03/2025	Pelanggan 26	Baju	275000.00	1	275000.00
28	13/03/2025	Pelanggan 27	Baju	320000.00	1	320000.00
29	13/03/2025	Pelanggan 28	Baju	320000.00	1	320000.00
30	13/03/2025	Pelanggan 29	Baju	360000.00	1	360000.00
31	13/03/2025	Pelanggan 30	Baju	430000.00	1	430000.00
32	13/03/2025	Pelanggan 31	Baju	420000.00	1	420000.00

33	13/03/2025	Pelanggan 32	Baju	470000.00	1	470000.00
34	13/03/2025	Pelanggan 33	Baju	410000.00	1	410000.00
35	13/03/2025	Pelanggan 34	Baju	412000.00	1	412000.00
36	13/03/2025	Pelanggan 35	Baju	418000.00	1	418000.00
37	13/03/2025	Pelanggan 36	Baju	450000.00	1	450000.00
38	13/03/2025	Pelanggan 37	Kosmetik Kecantikan	8750000.00	1	8750000.00
39	13/03/2025	Pelanggan 38	Kosmetik Kecantikan	8750000.00	1	8750000.00
40	13/03/2025	Pelanggan 39	Kosmetik Kecantikan	8850000.00	1	8850000.00
41	13/03/2025	Pelanggan 40	Kosmetik Kecantikan	8850000.00	1	8850000.00
42	13/03/2025	Pelanggan 41	Kosmetik Kecantikan	8950000.00	1	8950000.00
43	13/03/2025	Pelanggan 42	Kosmetik Kecantikan	8350000.00	1	8350000.00
44	13/03/2025	Pelanggan 43	Kosmetik Kecantikan	8750000.00	1	8750000.00
45	13/03/2025	Pelanggan 44	Kosmetik Kecantikan	8850000.00	1	8850000.00
46	13/03/2025	Pelanggan 45	Kosmetik Kecantikan	8880000.00	1	8880000.00
47	13/03/2025	Pelanggan 46	Kosmetik Kecantikan	8820000.00	1	8820000.00
48	13/03/2025	Pelanggan 47	Kosmetik Kecantikan	8620000.00	1	8620000.00
49	13/03/2025	Pelanggan 48	Kosmetik Kecantikan	8520000.00	1	8520000.00

3.5 Langkah-langkah K-Means Clustering

Data yang akan dianalisis terdiri dari 48 sampel data pembayaran yang difilter dari data set kaggle dengan detail sebagai berikut:

Berikut adalah langkah-langkah rinci dalam proses K-Means Clustering:

1. Inisialisasi Centroid:

Pilih centroid awal secara acak. Misalkan kita memilih empat centroid awal berikut:

- a. Centroid 1: 6,000,000.00
- b. Centroid 2: 550,000.00
- c. Centroid 3: 245,000.00
- d. Centroid 4: 8,750,000.00

2. Menghitung Jarak

Setiap data dalam dataset dihitung jaraknya ke setiap centroid menggunakan rumus jarak Euclidean:

$$\text{Jarak} = |x - c|$$

Di mana x adalah jumlah pengeluaran dan c adalah centroid.

1. Inisialisasi Centroid: Pilih centroid awal secara acak. Misalkan kita memilih empat centroid awal berikut:
 - o a. Centroid 1: 6,000,000.00
 - o b. Centroid 2: 550,000.00
 - o c. Centroid 3: 245,000.00
 - o d. Centroid 4: 8,750,000.00
2. Menghitung Jarak: Setiap data dalam dataset dihitung jaraknya ke setiap centroid menggunakan rumus jarak Euclidean:

Cluster	Centeroid
Cluster 1	6,000,000.00
Cluster 2	550,000.00
Cluster 3	245,000.00
Cluster 4	8,750,000.00

Jarak ke Centroid:

- o Jarak ke Centroid 1:
 $|6,000,000.00 - 6,000,000.00| = 0$
 $|6,000,000.00 - 6,000,000.00| = 0$
- o Jarak ke Centroid 2 : $|6,000,000.00 - 550,000.00| = 5,450,000.00$
 $|6,000,000.00 - 550,000.00| = 5,450,000.00$
- o Jarak ke Centroid 3:
 $|6,000,000.00 - 245,000.00| = 5,755,000.00$
 $|6,000,000.00 - 245,000.00| = 5,755,000.00$
- o Jarak ke Centroid 4:
 $|6,000,000.00 - 8,750,000.00| = 2,750,000.00$
 $|6,500,000.00 - 8,750,000.00| = 2,250,000.00$

Jarak Antar Cluster:

Setelah menghitung jarak dari ID 48, kita juga perlu menghitung jarak antar centroid cluster untuk mendapatkan informasi lebih lanjut.

- o Jarak antara Cluster 1 dan Cluster 2:
 $|6,000,000.00 - 550,000.00| = 5,450,000.00$
 $|6,000,000.00 - 550,000.00| = 5,450,000.00$
- o Jarak antara Cluster 1 dan Cluster 3:
 $|6,000,000.00 - 245,000.00| = 5,755,000.00$
 $|6,000,000.00 - 245,000.00| = 5,755,000.00$

- o Jarak antara Cluster 1 dan Cluster 4:
 $|6,000,000.00 - 8,750,000.00| = 2,750,000.00$
 $|6,000,000.00 - 8,750,000.00| = 2,750,000.00$
- o Jarak antara Cluster 2 dan Cluster 3:
 $|550,000.00 - 245,000.00| = 305,000.00$
 $|550,000.00 - 245,000.00| = 305,000.00$
- o Jarak antara Cluster 2 dan Cluster 4:
 $|550,000.00 - 8,750,000.00| = 8,200,000.00$
 $|550,000.00 - 8,750,000.00| = 8,200,000.00$
- o Jarak antara Cluster 3 dan Cluster 4:
 $|245,000.00 - 8,750,000.00| = 8,505,000.00$
 $|245,000.00 - 8,750,000.00| = 8,505,000.00$

Hasil Jarak Antar Cluster:

Berdasarkan perhitungan di atas, kita mendapatkan hasil jarak antar cluster sebagai berikut:

- o a. Jarak antara Cluster 1 dan Cluster 2: 5,450,000.00
- o b. Jarak antara Cluster 1 dan Cluster 3: 5,755,000.00
- o c. Jarak antara Cluster 1 dan Cluster 4: 2,750,000.00
- o d. Jarak antara Cluster 2 dan Cluster 3: 305,000.00
- o e. Jarak antara Cluster 2 dan Cluster 4: 8,200,000.00
- o f. Jarak antara Cluster 3 dan Cluster 4: 8,505,000.00

3. Hasil Cluster: Setelah proses K-Means Clustering, data pembayaran berhasil dikelompokkan ke dalam empat cluster berdasarkan jumlah pembayaran. Berikut adalah detail lebih lengkap tentang masing-masing cluster.

a. Data dalam Setiap Cluster

Cluster 1:

- o Centroid: 6,000,000.00
- o Data:
 - 1. ID 1: 6,000,000.00 (Bahan Elektronik)
 - 2. ID 2: 6,000,000.00 (Bahan Elektronik)
 - 3. ID 3: 6,500,000.00 (Bahan Elektronik)
 - 4. ID 4: 6,500,000.00 (Bahan Elektronik)
 - 5. ID 5: 6,500,000.00 (Bahan Elektronik)
 - 6. ID 6: 6,700,000.00 (Bahan Elektronik)
 - 7. ID 7: 6,700,000.00 (Bahan Elektronik)
 - 8. ID 8: 6,700,000.00 (Bahan Elektronik)
 - 9. ID 9: 6,700,000.00 (Bahan Elektronik)
 - 10. ID 10: 6,700,000.00 (Bahan Elektronik)
 - 11. ID 11: 6,700,000.00 (Bahan Elektronik)
 - 12. ID 12: 6,700,000.00 (Bahan Elektronik)

Jumlah Total Pembayaran dalam Cluster 1: Total = $12 \times 6,000,000.00 = 72,000,000.00$

Cluster 2:

- o Centroid: 550,000.00
- o Data:
 - 1. ID 13: 550,000.00 (Pembayaran Sepatu)
 - 2. ID 14: 510,000.00 (Pembayaran Sepatu)

3. ID 15: 500,000.00 (Pembayaran Sepatu)
4. ID 16: 520,000.00 (Pembayaran Sepatu)
5. ID 17: 430,000.00 (Pembayaran Sepatu)
6. ID 18: 480,000.00 (Pembayaran Sepatu)
7. ID 19: 470,000.00 (Pembayaran Sepatu)
8. ID 20: 574,000.00 (Pembayaran Sepatu)
9. ID 21: 554,000.00 (Pembayaran Sepatu)
10. ID 22: 554,000.00 (Pembayaran Sepatu)
11. ID 23: 584,000.00 (Pembayaran Sepatu)
12. ID 24: 534,000.00 (Pembayaran Sepatu)

Jumlah Total Pembayaran dalam Cluster 2: Total = $12 \times 550,000.00 = 6,600,000.00$
Cluster 3:

- Centroid: 245,000.00
- Data:
 1. ID 25: 245,000.00 (Pembayaran Baju)
 2. ID 26: 275,000.00 (Pembayaran Baju)
 3. ID 27: 320,000.00 (Pembayaran Baju)
 4. ID 28: 320,000.00 (Pembayaran Baju)
 5. ID 29: 360,000.00 (Pembayaran Baju)
 6. ID 30: 430,000.00 (Pembayaran Baju)
 7. ID 31: 420,000.00 (Pembayaran Baju)
 8. ID 32: 470,000.00 (Pembayaran Baju)
 9. ID 33: 410,000.00 (Pembayaran Baju)
 10. ID 34: 412,000.00 (Pembayaran Baju)
 11. ID 35: 418,000.00 (Pembayaran Baju)
 12. ID 36: 450,000.00 (Pembayaran Baju)

Jumlah Total Pembayaran dalam Cluster 3: Total = $12 \times 245,000.00 = 2,940,000.00$
Cluster 4:

- Centroid: 8,750,000.00
- Data:
 1. ID 37: 8,750,000.00 (Pembayaran Kosmetik)
 2. ID 38: 8,750,000.00 (Pembayaran Kosmetik)
 3. ID 39: 8,850,000.00 (Pembayaran Kosmetik)
 4. ID 40: 8,850,000.00 (Pembayaran Kosmetik)
 5. ID 41: 8,850,000.00 (Pembayaran Kosmetik)
 6. ID 42: 8,800,000.00 (Pembayaran Kosmetik)
 7. ID 43: 8,750,000.00 (Pembayaran Kosmetik)
 8. ID 44: 8,700,000.00 (Pembayaran Kosmetik)
 9. ID 45: 8,650,000.00 (Pembayaran Kosmetik)
 10. ID 46: 8,900,000.00 (Pembayaran Kosmetik)
 11. ID 47: 8,950,000.00 (Pembayaran Kosmetik)
 12. ID 48: 8,700,000.00 (Pembayaran Kosmetik)

Jumlah Total Pembayaran dalam Cluster 4: Total = $12 \times 8,750,000.00 = 105,000,000.00$

3.6 Hasil Pengelompokan Data Pembayaran

Setelah proses K-Means Clustering, kita dapat meringkas hasil pengelompokan data Pembayaran sebagai berikut:

- Cluster 1 (Pembayaran Elektronik): Total Pembayaran: 72,000,000.00
- Cluster 2 (Pembayaran Sepatu): Total Pembayaran: 6,600,000.00
- Cluster 3 (Pembayaran Baju): Total Pembayaran: 2,940,000.00
- Cluster 4 (Pembayaran Kosmetik): Total Pembayaran: 105,000,000.00

3.6 Implementasi Aplikasi

Dalam menjalankan aplikasi yang penulis rancang, berikut langkah-langkah yang harus dilaksanakan :

1. Membuka aplikasi yang sudah dipersiapkan. Maka akan tampil halaman menu awal dan kemudian admin memilih menu login maka akan tampil menu login seperti gambar 2.



Gambar 2 Halaman Tampilan Login

2. Setelah proses login selesai. Maka akan tampil halaman menu utama admin seperti gambar 3.



Gambar 3 Halaman Tampilan Utama Admin

3. Klik menu hasil yang terdapat pada menu utama admin kemudian sistem akan menampilkan hasil cluster pembelian seperti terlihat pada gambar 4.



- | Data Akun Pengguna | | | | | | |
|--------------------|----------|--|---------|----------|----------|--|
| No | Username | Password | Nama | Jabatan | Level | Aksi |
| 1 | operator | \$2y\$10\$xnPUDRBj6LEQcZnD41.VLeFisjwdbQr0lfcv2PoD3qZyN04x/XOm | umu eka | Operator | operator | Edit Hapus |
- [Tambah User](#)
[Kembali](#)

Gambar 5 Halaman Menu Akun

Sistem kerja dari pada Algoritma K-Means dalam *clustering* data konsumen *e-commerce* memiliki beberapa tahapan untuk menghasilkan sebuah informasi kepada pengguna. Tentukan nilai k sebagai jumlah kluster yang diinginkan. Pilih k dari set data X sebagai *centroid*. Nilai *centroid* merupakan rata-rata objek dalam kluster tersebut. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing kluster. Hitung menggunakan jarak *Euclidean*. Hitung kembali *centroid* berdasarkan data yang mengikuti kluster masing-masing. Dalam merancang dan membuat aplikasi *clustering* data konsumen *e-commerce* peneliti melakukan tahap awal seperti perancangan sistem kerja aplikasi menggunakan UML. Selanjutnya melakukan rancangan desain menu aplikasi menggunakan Microsoft Visio. Setelah tahapan perancangan sistem kerja aplikasi dan rancangan desain menu aplikasi selesai peneliti melakukan pembuatan aplikasi dengan bahasa pemrograman PHP yang ditampilkan dalam bentuk web. Hasil *output* penyelesaian masalah *clustering* data konsumen *e-commerce*. Dapat dilihat berdasarkan hasil pembayaran beberapa iterasi, hasil pengelompokan data pembayaran untuk Cluster 1 (Pembayaran Elektronik): Total Pembayaran: 72,000,000.00, Cluster 2 (Pembayaran Sepatu): Total Pembayaran: 6,600,000.00, Cluster 3 (Pembayaran Baju): Total Pembayaran: 2,940,000.00 dan Cluster 4 (Pembayaran Kosmetik): Total Pembayaran: 105,000,000.00.

DAFTAR PUSTAKA

- [1] A. S. Wijaya And J. F. Andry, “Perancangan Aplikasi E-Commerce Berbasis Android Pada Ud Hoky Celluler Shop,” *J. Teknoinfo*, Vol. 15, No. 2, P. 97, 2021, Doi: 10.33365/Jti.V15i2.1065.
- [2] J. P. Sutrisno And N. Anwar, “Rancang Bangun Aplikasi E-Commerce Berbasis Web Pada Toko Vapein,” *Ikra-Ith Inform. J. Komput. Dan Inform.*, Vol. 7, No. 3, Pp. 32–39, 2023, Doi: 10.37817/Ikraith-Informatika.V7i3.3048.
- [3] A. Adiyanto And Y. Arie Wijaya, “Penerapan Algoritma K-Means Pada Pengelompokan Data Set Bahan Pangan Indonesia Tahun 2022-2023,” *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 7, No. 2, Pp. 1344–1350, 2023, Doi: 10.36040/Jati.V7i2.6849.
- [4] M. Norshahlan, H. Jaya, And R. Kustini, “Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Data Calon Siswa Baru,” *J. Sist. Inf. Triguna Dharma (Jursi Tgd)*, Vol. 2, No. 6, P. 1042, 2023, Doi: 10.53513/Jursi.V2i6.9148.
- [5] Z. F. Jalil Abdul, Ahmad Homaiddi, “Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita,” Vol. 8, No. 3, Pp. 2070–2079, 2024.
- [6] D. A. Manalu, G. Gunadi, And T. Informatika, “Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada Cv Digital Dimensi,” Vol. 8, No. 1, Pp. 45–54, 2022.
- [7] A. Rahman And H. Suroyo, “Analisis Data Produk Elektronik Di E-Commerce Dengan Metode Algoritma K-Means Menggunakan Python,” Vol. 3, No. 2, Pp. 11–18, 2021.
- [8] S. F. Florentin And A. Santoso, “Analisis Segmentasi Pelanggan Menggunakan K- Means,” Vol. 26, No. 2, Pp. 446–457, 2022, Doi: 10.46984/Sebatik.V26i2.2134.
- [9] R. Maoulana, B. Irawan, And A. Bahtiar, “Data Mining Dalam Konteks Transaksi Penjualan Hijab Dengan Menggunakan Algoritma Clustering K-Means,” *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 8, No. 1, Pp. 515–521, 2024, Doi: 10.36040/Jati.V8i1.8504.
- [10] N. M. A. K. Lenama And I. G. Pasangka, “K-Means Clustering Analysis Pada Pengelompokan Kabupaten/Kota Di Provinsi Nusa Tenggara Timur Berdasarkan Indikator Pendidikan,” vol. 08, no. 9, pp. 3365–3376, 2023.