

Pengelompokkan Toko Kaus Termurah E-Commerce Shopee berdasarkan Reputasi Toko Menggunakan Metode Clustering K-Medoids dan K-Means

Lalu Riza Singrapati^{a1}, Rika Dora^{b2}, Robert Kurniawan^{a,b3}

^aProgram Studi Statistika, Politeknik Statistika STIS
Jl. Otto Iskandardinata 64C, Jakarta 1330

¹1212011673@stis.ac.id

²1212011339@stis.ac.id

³robertk@stis.ac.id

Abstrak

Kaus merupakan salah satu jenis pakaian yang paling diminati saat ini. Terutama setelah hadirnya toko *e-commerce* yang memudahkan pembeli untuk bertransaksi dengan cepat tanpa harus pergi ke tokonya secara langsung. Banyak toko online yang menawarkan kaus dengan harga yang terjangkau. Namun, pembeli harus selektif dalam melakukan transaksi jual beli melalui *e-commerce* karena banyaknya risiko yang bisa timbul. Untuk mengatasi hal tersebut, salah satu hal yang dapat dilakukan yaitu mengelompokkan toko pada platform *e-commerce* Shopee berdasarkan reputasi menggunakan metode K-Means dan K-Medoids. Penelitian ini menggunakan data dari tiga ratus akun toko kaus dengan harga termurah di Shopee. Tahapan dalam penelitian ini meliputi pengumpulan data, preprocessing data, penentuan jumlah *cluster* optimum, analisis cluster menggunakan K-Means dan K-Medoids, evaluasi model, interpretasi output, dan penarikan Kesimpulan. Berdasarkan hasil evaluasi, diperoleh metode terbaik ialah K-Means dengan *k* optimum sebanyak tiga *cluster*. Kemudian, *cluster* yang direkomendasikan kepada *customer* ialah *cluster* pertama.

Kata kunci: K-Means, K-Medoids, Kaus, *Cluster*

Clustering Shopee E-Commerce Cheapest Sweatshirt Stores based on Store Reputation Using K-Medoids and K-Means Clustering Methods

Abstract

T-shirts are one of the most sought-after types of clothing today, especially with the advent of e-commerce platforms that make it convenient for buyers to transact quickly without having to visit physical stores. Many online stores offer affordable t-shirts. However, buyers need to be selective when engaging in online transactions due to potential risks. To address this, one approach is to categorize stores on the Shopee e-commerce platform based on reputation using the K-Means and K-Medoids clustering methods. This study utilized data from three hundred accounts of stores selling the cheapest t-shirts on Shopee. The research process included data collection, data preprocessing, determining the optimal number of clusters, cluster analysis using K-Means and K-Medoids, model evaluation, output interpretation, and drawing conclusions. Based on the evaluation results, the best method obtained was K-Means with an optimum *k* value of three clusters. Furthermore, the recommended cluster for customers was the first cluster.

Keywords: K-Means, K-Medoids, Sweatshirts, Cluster

I. PENDAHULUAN

Pertumbuhan *e-commerce* atau kegiatan jual beli barang dan jasa secara *online* saat ini berlangsung sangat cepat. Hal ini disebabkan oleh mudahnya transaksi barang dan jasa tanpa batasan jarak dan waktu selama memiliki akses internet [1]. Salah satu platform *e-commerce* dengan jumlah pengunjung terbanyak di Indonesia adalah Shopee.

Pada kuartal kedua tahun 2022, Shopee berada di urutan kedua dengan kunjungan yang mencapai 131 juta pengunjung. Selain itu, Shopee menempati peringkat pertama pada *Appstore* dan *playstore* [2]. Platform ini didukung oleh sistem logistik dan pembayaran yang handal sehingga dapat memberikan pengalaman berbelanja *online* yang lebih mudah, aman, dan cepat.

Pada aplikasi ini, disediakan berbagai kategori produk, mulai dari *fashion*, perlengkapan rumah tangga, elektronik, kesehatan, hingga kecantikan.

Dari banyaknya kategori barang, industri *fashion* merupakan salah satu yang paling diminati di dunia. Selain itu, cepatnya perkembangan tren *fashion* bersamaan dengan banyaknya pakaian murah yang diproduksi oleh merek *fashion* global membuat barang tersebut menjadi pembelian sehari-hari [3]. Begitu pula di Indonesia, kategori ini digunakan untuk memenuhi kebutuhan dasar dan mencerminkan status sosial ekonomi serta popularitas seseorang. Salah satu jenis *fashion* yang mengalami perkembangan paling cepat di industri *fashion* secara global adalah kaus. Hal ini disebabkan oleh pergantian model kaus yang lebih cepat dibandingkan dengan jenis *fashion* lainnya. Selain itu, kaus juga merupakan produk yang paling diminati dan dibeli oleh masyarakat jika dibandingkan dengan produk *fashion* lainnya [4].

Di Shopee, terdapat banyak pilihan kaus dengan berbagai harga yang tersedia. Harga barang yang akan dibeli merupakan salah satu aspek yang diperhatikan oleh pembeli saat melakukan transaksi di toko online [5]. Biasanya, pembeli suka memilih kaus dengan harga yang murah yang sesuai dengan budget namun tetap mengharapkan kualitas yang bagus. Dari situ, timbul beberapa masalah yang sering terjadi, seperti kaus dengan harga murah namun kualitasnya buruk, kaus yang tidak sesuai dengan pesanan, kaus yang tidak sesuai dengan deskripsi, kemasan yang rusak, dan sebagainya. Dalam memilih toko kaus yang dapat dipercaya, pengguna harus selektif terhadap pilihan yang ada untuk menghindari risiko, terutama jika toko tersebut menawarkan produk dengan harga murah yang diklaim memiliki kualitas bagus. Selain itu, diperlukan ketelitian untuk dapat melihat spesifikasi dan deskripsi barang pada toko. Salah satu faktor yang perlu diperhatikan dalam memilih toko kaus adalah reputasinya, sebab reputasi ini mencerminkan kemampuan penjual dalam memberikan pelayanan yang baik dan sarana untuk membangun kepercayaan pada calon pelanggan [6].

Pada umumnya *platform e-commerce*, terdapat fitur *customer rating* yang merupakan bagian dari *customer review*. Fitur ini ditampilkan dalam bentuk bintang kepada pengguna yang bertujuan untuk memberikan apresiasi kepada penjual atau toko. Kepercayaan pelanggan yang semakin tinggi ditandainya dengan tingginya *rating* yang diberikan [7]. Reputasi toko kaus juga dapat dinilai berdasarkan banyaknya pengikut atau followers. Toko online dengan basis pengikut yang besar di media sosial dapat memiliki interaksi yang lebih kuat dengan pelanggan dan meningkatkan kesetiaan pelanggan. Oleh karena itu, banyaknya pengikut atau *followers* dapat meningkatkan daya saing toko dibandingkan dengan toko lainnya, serta menunjukkan bahwa toko tersebut dipercaya oleh pengikutnya [8]. Selain itu, hal lain yang mempengaruhi kualitas toko adalah kinerja *chat* sebab pelayanan yang diberikan kepada pelanggan dapat tergambar melalui hal ini. Semakin tinggi kinerja *chat* suatu toko, menunjukkan bahwa semakin baik toko

tersebut. Hal ini disebabkan oleh respons yang cepat dalam membantu pelanggan mampu meningkatkan kepercayaan terhadap toko tersebut [9]. Selain fitur-fitur tersebut, variabel lain seperti waktu bergabung juga mempengaruhi kualitas dari toko yang bersangkutan karena semakin lama suatu toko bergabung dengan platform *online*, biasanya semakin berpengalaman [10]. Dari faktor-faktor tersebut, perlu adanya suatu sistem pengelompokan yang dapat membantu pelanggan untuk mengetahui kualitas dari tiap-tiap toko pada *marketplace*.

Pengelompokan atau *clustering* merupakan proses menggabungkan sampel ke dalam beberapa kelompok berdasarkan atribut atau fitur tertentu. Prinsip pada pengelompokan ini ialah menjadikan objek dalam kelompok yang sama memiliki kesamaan, sementara objek dalam kelompok yang berbeda memiliki perbedaan. Semakin tinggi kesamaan di dalam kelompok dan semakin rendah kesamaan antar-kelompok, hasil pengelompokan akan semakin baik [11]. Terdapat dua jenis metode *clustering* yang umum digunakan pada data yang tidak berlabel, yaitu K-Means dan K-Medoids. Kesamaan dari kedua metode ini adalah membagi himpunan data menjadi *k cluster*. Namun, penentuan titik *centroid* untuk metode K-Means berdasarkan nilai rata-rata data dari tiap cluster, sedangkan dalam metode K-Medoids, titik pusat ditentukan secara acak [12].

Berdasarkan penelitian sebelumnya, diperoleh kesimpulan bahwa pengelompokan dengan pendekatan K-medoids menjadi metode terbaik [13]. Namun, pada penelitian lainnya yang menggunakan metode *clustering*, menyimpulkan bahwa metode K-Means yang terbaik dalam pengelompokan data [14]. Oleh karena itu, dalam kasus ini perlu dibandingkan metode mana yang lebih baik untuk dapat mengelompokkan toko berdasarkan beberapa karakteristik yang telah dijelaskan sebelumnya.

Penelitian ini bertujuan untuk mengelompokkan toko kaus termurah pada platform *e-commerce* Shopee berdasarkan faktor banyak pengikut, rating, performa chat (%), dan tahun bergabung. Dengan melakukan pengelompokan, akan terbentuk cluster toko kaus dengan reputasi yang berbeda-beda berdasarkan faktor-faktor tersebut. Hal ini dapat menjadi sumber informasi bagi calon konsumen untuk memilih toko kaus yang berkualitas dengan harga yang lebih terjangkau serta dapat membangun kepercayaan konsumen terhadap penjual. Meskipun terdapat tipe penjual yang telah ditentukan di Shopee, terkadang terdapat beberapa toko yang tidak terkelompokkan dengan benar dan terlihat seperti tidak berada dalam kategori yang tepat. Terkadang, toko-toko dengan banyak pengikut (*follower*) tidak selalu ditempatkan dalam pengelompokan yang baik. Selanjutnya, penelitian ini bertujuan untuk membantu pembeli dalam menemukan toko yang sesuai dengan preferensi dan kebutuhan mereka.

II. METODOLOGI

A. Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah dataset karakteristik pada tiga ratus toko kaus termurah

yang diambil melalui aplikasi Shopee per tanggal 6 Juni 2023 yang terdiri atas banyak pengikut, *rating*, performa *chat*, dan waktu bergabung.

B. *Preprocessing Data*

Pada penelitian ini, masih terdapat *data noise* dan inkonsisten pada data. Hal ini dapat mengurangi hasil akurasi yang diperoleh sehingga perlu dilakukan *preprocessing data* untuk mempersiapkan dataset tersebut sebelum digunakan. *Preprocessing data* yang dilakukan berupa pengecekan *missing value*, non-multikolinieritas, dan *outlier*, serta standarisasi data. Melalui *software R-Studio*, pada dataset, tidak terdapat indikasi adanya *missing value* dan multikolinieritas sehingga tidak perlu menghilangkan observasi maupun variabel. Namun, setelah dilakukan pemeriksaan data menggunakan *boxplot*, masih terdapat *outlier* atau pencilan. Metode yang digunakan untuk mengatasi masalah ini adalah *winsorizing*.

Selain itu, terdapat perbedaan satuan pada nilai variabel data. Perbedaan nilai yang sangat jauh dapat menyebabkan perhitungan menjadi tidak valid [15]. Oleh karena itu, diperlukan standarisasi data yang disajikan dalam persamaan berikut [16].

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij} \tag{1}$$

$$\sigma_j^2 = \frac{1}{N-1} \sum_{i=1}^N (x_{ij} - \bar{x}_j)^2 \tag{2}$$

$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \tag{3}$$

Dimana:

N = Banyak Data

X_{ij} = Data ke- i pada variabel ke- j dengan $i = 1, 2, \dots, N$ dan $j = 1, 2, \dots, n$

\bar{x}_j = Rata-rata pada variabel ke- j

S_j^2 = Varians

S_j = Standar Deviasi

\hat{x}_{ij} = Standarisasi data ke- i variabel ke- j

C. *Penentuan Jumlah Cluster Optimum*

Banyaknya *cluster* yang ditentukan sendiri oleh peneliti, pada umumnya dapat berpengaruh pada hasil yang kurang optimal sehingga dalam penelitian ini nilai k ditentukan menggunakan suatu metode agar lebih akurat. Metode yang digunakan adalah metode *silhouette*. Metode ini termasuk ukuran yang populer dan digunakan untuk mengukur kedekatan antar observasi pada *cluster* yang sama dan seberapa jauh jarak observasi tersebut pada *cluster* yang berbeda. Atau dengan kata lain, metode ini mengukur kemiripan observasi pada suatu *cluster* dibandingkan dengan *cluster* lain yang berbeda [17].

D. *K-Means*

K-Means merupakan salah satu metode *non-hierarchical clustering* yang setiap observasi data diukur jaraknya terhadap titik tengah (rata-rata) dari suatu *cluster*

(*centroid*). Observasi data yang memiliki jarak terdekat terhadap *centroid* pada sebuah *cluster* akan menjadi anggota *cluster* tersebut [11]. Adapun tahapan-tahapan yang dilakukan dalam metode *K-Means* adalah [18].

1. Menentukan banyaknya *cluster* data.
2. Menentukan nilai *centroid* secara acak.
3. Mengukur jarak observasi terhadap masing-masing *centroid* dengan *euclidean distance*. Adapun persamaan *euclidean distance* adalah sebagai berikut:

$$d_{(x,y)} = |x - y| = \sqrt{\left(\sum_{i=1}^n (x_i - y_i)\right)^2} \tag{4}$$

Dimana:

$d_{(x,y)}$ = Jarak *Euclidean* antara data dan titik pusat *cluster*

x_i = Data Observasi ke- i

y_i = Titik pusat pada *cluster* ke- i

4. Berdasarkan pengukuran pada tahap sebelumnya, distribusikan observasi ke dalam *cluster* terdekat.
5. Pilih observasi menggunakan rumus di bawah ini untuk menentukan calon *medoid* yang baru.

$$\bar{v}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \tag{5}$$

Dimana:

v_{ij} = titik *centroid cluster* ke- i variabel ke- j

N_i = Jumlah data pada *cluster* ke- i

x_{kj} = Nilai data ke- k di *cluster* tersebut pada variabel ke- i

6. Hitung kembali jarak observasi terhadap nilai *centroid* yang baru.
7. Observasi dimasukkan ke dalam *cluster* tertentu berdasarkan jarak terpendek yang dihasilkan.
8. Iterasi akan terus dilakukan sampai hasil yang diperoleh sama dengan iterasi sebelumnya.

E. *K-Medoids*

K-Medoids adalah salah satu metode partisi karena metode ini membagi data sebanyak k *cluster* dan menggunakan salah satu objek secara acak sebagai titik pusat atau yang biasa disebut dengan *medoid* pada tiap *cluster*-nya [19]. Metode ini mirip dengan *k-means*, tetapi dalam penentuan titik pusatnya tidak didasarkan pada mean dari tiap *clusternya*, melainkan ditentukan secara acak. Berikut adalah tahapan-tahapan dalam algoritma *K-Medoids* [12].

1. Menentukan jumlah *cluster* (k) yang diinginkan dan secara acak pilih k titik data sebagai *centroid* awal.
2. Hitung jarak setiap titik data terhadap *centroid*, dan menempatkan titik observasi ke kelompok terdekat dengan *centroid* menggunakan jarak yang ditentukan. Pada umumnya, jarak yang digunakan adalah jarak *Euclidean*, tetapi dapat juga menggunakan jarak lain seperti jarak *Manhattan* atau jarak *Mahalanobis*.
3. Pilih titik observasi *non centroid* dari setiap kelompok secara acak yang menghasilkan total jarak minimum dalam kelompok tersebut. Titik

data baru ini akan menjadi medoid (*centroid*) baru untuk kelompok tersebut.

4. Ulangi langkah tersebut sampai tidak ada perubahan medoid atau perubahan total jarak yang signifikan antara iterasi yang berurutan.

F. Evaluasi Model

Data yang sudah dikelompokkan ke dalam beberapa *cluster* perlu dilakukan evaluasi untuk mengetahui *cluster* terbaik yang dihasilkan. Dalam penentuan model dengan *cluster* terbaik, peneliti menggunakan beberapa metode evaluasi sebagai berikut.

1. Silhouette Index

Silhouette Index merupakan salah satu metode yang digunakan untuk menentukan model dengan *cluster* terbaik. Nilai *Silhouette Index* berkisar antara -1 hingga +1, semakin mendekati +1 maka semakin baik pengelompokkan tersebut. Sebaliknya, semakin mendekati -1 maka semakin buruk pengelompokkan yang dilakukan. Adapun persamaan untuk nilai *Silhouette Index* adalah sebagai berikut [20].

$$S_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \quad (6)$$

Dengan:

$a(i)$ = Rata-rata jarak antara titik ke- i terhadap seluruh titik yang ada pada *cluster* yang sama dengan titik i tersebut.

$b(i)$ = Rata-rata jarak antara titik ke- i terhadap seluruh titik yang ada pada selain *cluster* yang sama dengan titik i tersebut.

2. Davies-Bouldin Index

Davies-Bouldin Index merupakan salah satu metode yang digunakan untuk mengukur validitas *cluster*. Konsep dalam pengukuran menggunakan indeks ini yaitu meminimalkan jarak antartitik dalam sebuah *cluster* dan memaksimalkan jarak antar-*cluster*. Apabila jarak antar-*cluster* maksimal dan intra-*cluster* minimal, artinya perbedaan antar-*cluster* semakin terlihat jelas dan observasi intra-*cluster* memiliki kemiripan karakteristik yang tinggi (homogen). Nilai indeks DB yang semakin rendah, menunjukkan semakin baik pengelompokan yang dihasilkan. Metode ini menggunakan *Sum of Squared Error* dalam perhitungannya [21].

3. Dunn Index

Selain DB Indeks dan *silhouette index*, metode ini juga dapat digunakan untuk mengukur kedekatan dalam *cluster* dan tingkat pemisahan antar-*cluster* [22]. Nilai indeks yang semakin besar menunjukkan semakin baik jumlah *cluster* yang terbentuk. Adapun persamaan Indeks Dunn adalah sebagai berikut [23].

$$DN = \quad (7)$$

$$\min_{p=1, \dots, k} \left\{ \min_{q=i+1, \dots, k} \left(\frac{d(c_p, c_q)}{\max_{r=1, \dots, k} \text{diam}(c_r)} \right) \right\}$$

$$d(c_p, c_q) = \min_{x_i \in c_p, y_j \in c_q} d(x_i, y_j) \quad (8)$$

$$\text{diam}(c_r) = \max_{z_i, z_j \in c_r} d(z_i, z_j) \quad (9)$$

Dengan:

k = Banyaknya *cluster*

N = Banyaknya Observasi

C_l = *cluster* ke- l

x_i = titik ke- i pada *cluster* ke- l

x_l = titik pusat pada *cluster* ke- l

4. Calinski-Harabasz Index (CH Index)

Indeks evaluasi ini didasarkan pada tingkat dispersi antar klaster. Nilai Indeks CH yang semakin besar, menunjukkan bahwa semakin baik pengelompokkan yang dihasilkan. Adapun persamaannya adalah sebagai berikut [24].

$$CH(K) = \frac{B(K)(N-K)}{W(K)(K-1)} \quad (10)$$

$$B(K) = (\sum_{k=1}^K \alpha_k ||\bar{x}_k - \bar{x}||^2) \quad (11)$$

$$W(K) = (\sum_{k=1}^K \sum_{c(j)=k} ||x_j - \bar{x}_k||^2) \quad (12)$$

Dengan:

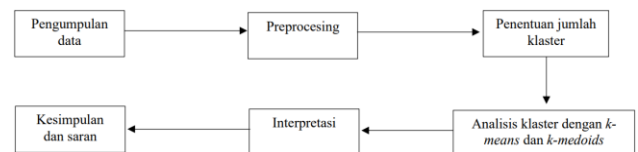
$W(K)$ = Perbedaan intraklaster

$B(K)$ = Perbedaan antarklaster

5. Entropy

Entropy digunakan untuk mengukur tingkat ketidakpastian atau keacakan dalam suatu sistem. Secara matematis, *entropy* dinyatakan sebagai ukuran rata-rata dari jumlah informasi yang dibutuhkan untuk mengidentifikasi suatu kejadian atau karakteristik dalam distribusi probabilitas. Nilai *entropy* berkisar antara 0 dan 1, dengan nilai terbaiknya adalah nol. Atau dengan kata lain, semakin kecil nilai yang diperoleh, akan menghasilkan model yang lebih baik [13].

G. Tahapan Penelitian



Gambar 1. Diagram Alir

III. HASIL DAN PEMBAHASAN

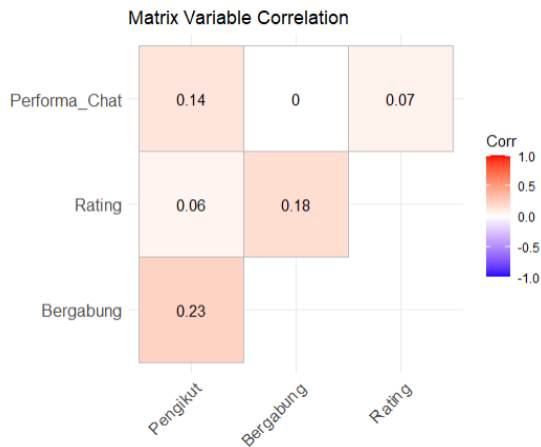
A. Standarisasi Data

Pada dataset yang digunakan, terdapat variabel dengan satuan yang berbeda, misalnya performa chat dalam persentase (%) dan waktu bergabung dalam tahun. Untuk menyamakan satuan tersebut, perlu dilakukan standarisasi data menggunakan *z-score*.

B. Pengecekan Non Multikolinieritas

Pengecekan asumsi ini menggunakan *matrix variable correlation*. Berdasarkan hasil yang diperoleh, tidak terdapat indikasi adanya hubungan antara dua variabel yang disajikan dalam bentuk korelasi pearson. Nilai korelasi tersebut tidak ada yang melebihi 0.8 yang mengindikasikan bahwa asumsi ini terpenuhi [25]. Oleh

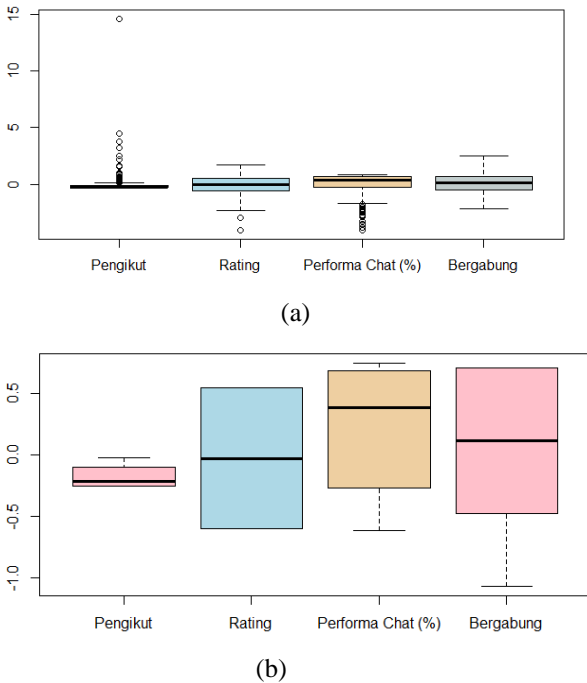
karena itu, dalam hal ini tidak perlu ada variabel yang dihilangkan.



Gambar 2. Pengecekan Asumsi Non-multikolinieritas

C. Pengecekan Outlier

Setelah melakukan standarisasi data, eksplorasi lebih lanjut menggunakan *boxplot*, ditemukan bahwa terdapat *outlier* pada variabel jumlah pengikut, *rating*, dan performa *chat*, sedangkan untuk variabel waktu bergabung dengan aplikasi Shopee tidak menunjukkan adanya *outlier*. Untuk mengatasi *outlier* ini, digunakan metode *winsorizing*.

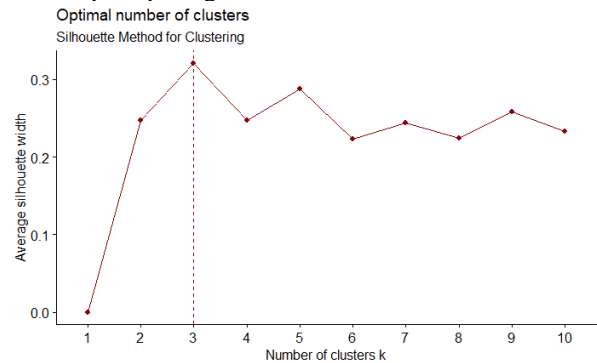


Gambar 3. (a) Pengecekan dan (b) Penanganan Outlier

Berdasarkan gambar 2(b), pencilan atau *outlier* pada keempat variabel amatan sudah berhasil diatasi. Hal ini dapat dilihat dari ketiadaan observasi yang melebihi batas yang ditentukan untuk mengidentifikasi *outlier* dalam data. Dengan demikian, hasil data dapat digunakan untuk melakukan proses analisis lebih lanjut.

D. Jumlah Cluster Optimum

Sebelum dilakukan klusterisasi toko kaus, perlu ditentukan jumlah *k* yang optimum dengan menggunakan metode *Average silhouette*. Metode ini mengukur kualitas dan kecocokan seberapa baik setiap sampel data cocok dalam klusternya. Dari metode tersebut diperoleh hasil seperti pada grafik berikut.

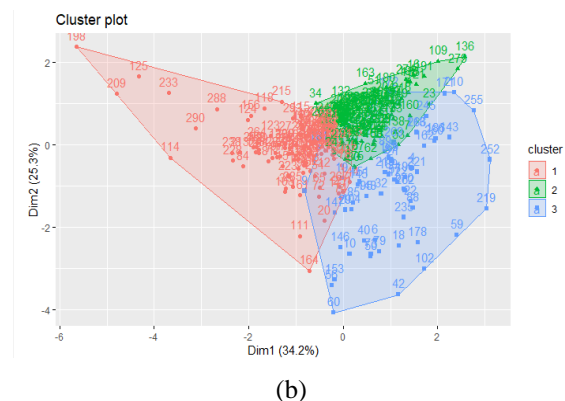
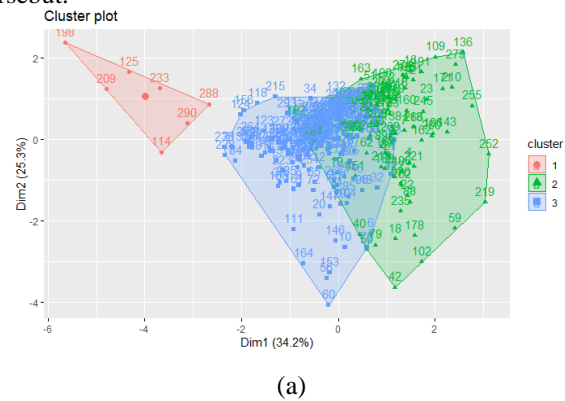


Gambar 4. Jumlah cluster optimum (*Average silhouette*)

Berdasarkan grafik di atas, jumlah kluster optimum ini ditandai dengan nilai *average silhouette* yang paling tinggi atau memiliki tingkat kohesi yang baik di dalam kluster dan tingkat pemisahan yang baik antara kluster. Dalam hal ini, nilai *Average Silhouette* paling tinggi dicapai ketika *k=3* sehingga hal ini menunjukkan bahwa klusterisasi data menjadi tiga kluster memberikan hasil yang lebih baik dibandingkan dengan jumlah kluster lainnya.

E. Analisis Cluster dengan K-Means dan K-Medoids

Dengan menggunakan metode K-Means dan K-Medoids, data akan dikelompokkan sebanyak tiga *cluster*. Berikut ditampilkan hasil visualisasi dari kedua metode tersebut.



Gambar 5. Visualisasi cluster plot dari metode (a) k-means dan (b) k-medoids

Berdasarkan kedua gambar di atas, terdapat perbedaan yang cukup signifikan di antara kedua metode tersebut. Apabila dilihat dari segi observasi, klusterisasi menggunakan metode k-means memiliki lebih banyak observasi pada *cluster* ketiga, yakni sebanyak 200 observasi, 92 observasi dalam *cluster* kedua, dan 7 observasi pada *cluster* pertama. Sedangkan untuk metode k-medoids, pada *cluster* pertama menghasilkan sebanyak 117 observasi, sebanyak 125 observasi pada *cluster* kedua, dan 57 observasi pada *cluster* ketiga. Perbedaan jumlah observasi pada tiap *cluster* dapat disebabkan oleh penentuan titik *centroid* masing-masing metode.

F. Evaluasi

Cluster yang dikatakan baik dalam mengelompokkan data apabila variasi di dalam *cluster* rendah tetapi variasi antar-*cluster* yang tinggi. Untuk menentukan metode yang terbaik di antara kedua metode yang sebelumnya digunakan, perlu adanya evaluasi model. Pada penelitian ini, ukuran yang digunakan yaitu *Davies-Bouldin Index* (DBI), *Calinski-Hazzez Index* (CH Index), *silhouette coefficient*, *entropy*, dan *dunn index*. Dengan menggunakan beberapa ukuran evaluasi ini, kita dapat membandingkan metode pengelompokkan yang berbeda dan memilih metode yang memberikan hasil terbaik dalam memisahkan dan mengelompokkan data dengan akurat.

TABEL I
EVALUASI MODEL TERBAIK

	DB	Silhouette	Dunn	Entropy	CH Index
K-Means	1.358	0.278	0.055	0.719	85.846
K-Medoids	1.633	0.228	0.037	1.048	75.531

Berdasarkan output di atas, disajikan perbandingan antara metode K-Means dan K-Medoids menggunakan beberapa ukuran evaluasi. Pada metode evaluasi tersebut, diperoleh bahwa nilai DBI dan *entropy* K-means lebih kecil dibandingkan K-Medoids. Kemudian, jika dilihat dari nilai evaluasi sisanya, yaitu DBI, *Silhouette*, dan CH Indeks, nilai K-Means lebih besar dibandingkan K-Medoids. Dengan demikian, dapat disimpulkan bahwa secara keseluruhan, metode K-Means lebih baik dibandingkan K-Medoids karena memenuhi seluruh kriteria dalam ukuran evaluasi yang digunakan. Hal ini menunjukkan bahwa performa K-Means lebih baik dalam hal klusterisasi data yang digunakan. Namun, perlu diingat bahwa pemilihan metode terbaik juga bergantung pada karakteristik dan tujuan analisis yang spesifik, serta pentingnya faktor-faktor lain yang perlu dipertimbangkan dalam konteks penggunaannya.

G. Analisis Kluster berdasarkan Metode Terbaik

Berdasarkan hasil evaluasi, metode terbaik ialah K-Means dengan *k* optimum sebanyak tiga *cluster*. Karakteristik tiap *cluster* tersebut, ditampilkan pada tabel berikut ini.

TABEL II
OUTPUT K-MEANS CLUSTERING

Cluster	Pengikut	Rating	Performa Chat (%)	Bergabung
1	552571	4.757143	92.0000	5.571429
2	15579	4.515217	73.4674	2.809351
3	32875	4.789500	85.9600	4.201137

Berdasarkan tabel di atas, *cluster* pertama mempunyai rata-rata pengikut yang lebih banyak, performa *chat* yang lebih tinggi, serta waktu bergabung yang lebih lama dibandingkan dengan *cluster* kedua dan ketiga. Namun, *cluster* ketiga mempunyai rata-rata *rating* tertinggi di antara ketiga *cluster*. Sementara itu, *cluster* kedua memiliki rata-rata pengikut yang lebih sedikit, *rating* dan performa *chat* yang lebih rendah, serta waktu bergabung yang lebih pendek dibandingkan dengan *cluster* pertama dan ketiga.

Berdasarkan hasil tersebut, *cluster* yang paling banyak memenuhi kriteria di atas adalah *cluster* pertama. Meskipun *rating* pada *cluster* ketiga paling tinggi, tetapi nilai ini tidak berbeda jauh dibandingkan *cluster* pertama yang sama-sama masih berkisar pada angka 4.7-an. Oleh karena itu, *cluster* pertama lebih direkomendasikan untuk membeli kaus jika didasarkan pada harga termurah. Ditinjau dari segi jumlah pengikut atau *followers* menunjukkan seberapa banyak pelanggan yang percaya terhadap toko kita sehingga jumlah yang semakin banyak akan semakin meningkatkan kepercayaan terhadap toko yang bersangkutan [8]. Kemudian, waktu bergabung yang lebih lama menunjukkan bahwa toko tersebut lebih berpengalaman. Seperti dalam suatu penelitian yang dilakukan oleh Gomez et al (2022), bahwa urutan masuk merupakan salah satu hal penting yang berpengaruh terhadap kinerja, semakin lama suatu perusahaan memasuki pasar, semakin sedikit *customer* yang mengikuti toko tersebut, dan akan membuat kinerjanya lebih buruk karena perlu bersaing dengan para pelopor [26]. Selain itu, performa *chat* yang lebih tinggi yang menunjukkan kualitas layanan yang lebih baik. Hal ini sesuai dengan penelitian sebelumnya mengenai pengaruh *chat* terhadap keputusan konsumen dalam membeli suatu barang, hasilnya menunjukkan bahwa konsumen yang berkomunikasi melalui *chat* cenderung lebih besar peluangnya untuk melakukan pembelian barang. Peluang pembelian akan semakin tinggi apabila penjual memberikan balasan yang cepat dalam merespons *customer* [9]. Sebaliknya, *cluster* yang paling tidak direkomendasikan ialah *cluster* kedua karena memiliki rata-rata pengikut yang lebih sedikit, *rating* dan performa *chat* yang lebih rendah, serta waktu bergabung yang lebih pendek dibandingkan dengan *cluster* pertama dan ketiga.

Hal ini dapat mengindikasikan bahwa cluster kedua terdiri dari toko-toko yang relatif baru, masih dalam tahap pengembangan, atau belum memiliki basis pengikut yang besar.

Adapun berdasarkan jumlah observasi pada *cluster* pertama ini sangat sedikit, yang hanya berjumlah tujuh observasi. Hal ini menandakan bahwa hanya sedikit toko dengan penjualan barang murah, dalam hal ini kaus, tetapi memiliki kualitas yang lebih baik. Hal ini sesuai dengan penelitian yang dilakukan oleh Guizzardi et al. (2022), yang menyatakan bahwa hubungan antara kualitas dan harga ialah positif, artinya secara umum harga yang lebih tinggi akan memberikan kualitas yang lebih baik atau dengan kata lain semakin rendah harga yang ditawarkan maka semakin sedikit produk dengan kualitas yang baik [27].

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis untuk mengelompokkan tiga ratus toko dengan harga kaus termurah pada *platform online Shopee* diperoleh bahwa berdasarkan evaluasi yang dilakukan, metode terbaik ialah K-Means dengan *k* optimum sebanyak tiga *cluster*. Dari ketiga *cluster* ini, *cluster* pertama lebih direkomendasikan karena mempunyai rata-rata pengikut yang lebih banyak, performa chat yang lebih tinggi, serta waktu bergabung yang lebih lama dibandingkan dengan dua *cluster* lainnya. Meskipun *rating* pada *cluster* ketiga lebih tinggi, tetapi angka ini tidak terlalu berbeda jauh. Sebaliknya, *cluster* yang tidak direkomendasikan bagi *customer* ialah *cluster* kedua karena mempunyai pengikut yang lebih sedikit, performa *chat* dan *rating* yang lebih rendah, serta waktu bergabung yang lebih singkat dibandingkan dua *cluster* lainnya. Apabila ditinjau dari segi observasi, *cluster* pertama memiliki observasi yang jauh lebih sedikit karena pada umumnya harga memiliki hubungan positif dengan kualitas sehingga hanya sedikit toko dengan produk yang memiliki kualitas lebih tinggi tetapi dengan harga yang lebih murah. Kemudian, pada penelitian selanjutnya diharapkan dapat menggunakan variabel lain yang lebih spesifik sehingga dapat lebih menjelaskan reputasi pada toko kaus termurah. Selain itu, dapat digunakan metode klastering lain untuk meningkatkan nilai evaluasi yang dihasilkan.

DAFTAR PUSTAKA

- [1] İsoraité, M., & Miniotiené, N. (2018). *Electronic Commerce: Theory and Practice*. 2(2), 73–79. <https://doi.org/10.5281/zenodo.1258437>
- [2] iPrice. 2023. "The Map of E-Commerce in Indonesia". Diakses 13 Juni 2023. <https://iprice.co.id/insights/mapofecommerce/en/>
- [3] Brydges, T. (2021). Closing the loop on take, make, waste: Investigating circular economy practices in the Swedish fashion industry. *Journal of Cleaner Production*, 293. <https://doi.org/10.1016/j.jclepro.2021.126245>
- [4] Lu, Y. H., & Zhang, W. T. (2018). The Impact of Topic Factors on Commodity Reviews. *Procedia Computer Science*, 154, 806–816. <https://doi.org/10.1016/j.procs.2019.06.079>
- [5] Abdelsalam, S., Salim, N., Alias, R. A., & Husain, O. (2020). Understanding Online Impulse Buying Behavior in Social Commerce: A Systematic Literature Review. *IEEE Access*, 8, 89041–89058. <https://doi.org/10.1109/ACCESS.2020.2993671>
- [6] Oghazi, P., Karlsson, S., Hellström, D., Mostaghel, R., & Sattari, S. (2021). From Mars to Venus: Alteration of trust and reputation in online shopping. *Journal of Innovation and Knowledge*, 6(4), 197–202. <https://doi.org/10.1016/j.jik.2020.06.002>
- [7] Nanda, P., & Patnaik, S. (2023). A multi-agent coalition-based approach for order fulfillment in e-commerce. *Decision Analytics Journal*, 7. <https://doi.org/10.1016/j.dajour.2023.100227>
- [8] Hennig-Thurau, T., Malthouse, E. C., Frieger, C., Gensler, S., Lobschat, L., Rangaswamy, A., & Skiera, B. (2010). The impact of new media on customer relationships. *Journal of Service Research*, 13(3), 311–330. <https://doi.org/10.1177/1094670510375460>
- [9] Lv, Z., Jin, Y., & Huang, J. (2018). How do sellers use live chat to influence consumer purchase decision in China? *Electronic Commerce Research and Applications*, 28, 102–113. <https://doi.org/10.1016/j.elerap.2018.01.003>
- [10] Chen, N., & Yang, Y. (2021). The impact of customer experience on consumer purchase intention in cross-border E-commerce—Taking network structural embeddedness as mediator variable. *Journal of Retailing and Consumer Services*, 59. <https://doi.org/10.1016/j.jretconser.2020.102344>
- [11] Liu, G., Ji, F., Sun, W., & Sun, L. (2023). Optimization design of short-circuit test platform for the distribution network of integrated power system based on improved K-means clustering. *Energy Reports*, 9, 716–726. <https://doi.org/10.1016/j.egvr.2023.04.319>
- [12] Wu, Z., Jin, L., Zhao, J., Jing, L., & Chen, L. (2022). Research on Segmenting E-Commerce Customer through an Improved K-Medoids Clustering Algorithm. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/9930613>
- [13] Sureja, N., Chawda, B., & Vasant, A. (2022). An improved K-medoids clustering approach based on the crow search algorithm. *Journal of Computational Mathematics and Data Science*, 3, 100034. <https://doi.org/10.1016/j.jcmds.2022.100034>
- [14] El-Khatib, S. A., Skobtsov, Y. A., & Rodzin, S. I. (2021). Comparison of Hybrid ACO-k-means algorithm and Grub cut for MRI images segmentation. *Procedia Computer Science*, 186, 316–322. <https://doi.org/10.1016/j.procs.2021.04.150>
- [15] Peña, M. (2018). Robust clustering methodology for multi-frequency acoustic data: A review of standardization, initialization and cluster geometry. In *Fisheries Research* (Vol. 200, pp. 49–60). Elsevier B.V. <https://doi.org/10.1016/j.fishres.2017.12.013>
- [16] Bouguettaya, A., Yu, Q., Liu, X., Zhou, X., & Song, A. (2015). Efficient agglomerative hierarchical clustering. *Expert Systems with Applications*, 42(5), 2785–2797. <https://doi.org/10.1016/j.eswa.2014.09.054>
- [17] Batool, F., & Hennig, C. (2021). Clustering with the Average Silhouette Width. *Computational Statistics and Data Analysis*, 158. <https://doi.org/10.1016/j.csda.2021.107190>
- [18] Qi, J., Yu, Y., Wang, L., Liu, J., & Wang, Y. (2017). An effective and efficient hierarchical K-means clustering algorithm. *International Journal of Distributed Sensor Networks*, 13(8), 1–17. <https://doi.org/10.1177/1550147717728627>
- [19] Zhang, Q., & Couloigner, I. (2005). A New and Efficient K-Medoid Algorithm for Spatial Clustering. In *LNCS* (Vol. 3482). http://dx.doi.org/10.1007/11424857_20
- [20] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. In *Journal of Computational and Applied Mathematics* (Vol. 20).
- [21] Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- [22] Ben Ncir, C. E., Hamza, A., & Bouaguel, W. (2021). Parallel and scalable Dunn Index for the validation of big data clusters. *Parallel Computing*, 102. <https://doi.org/10.1016/j.parco.2021.102751>
- [23] Azhir, E., Jafari Navimipour, N., Hosseinzadeh, M., Sharifi, A., & Darwesh, A. (2021). An automatic clustering technique for query plan recommendation. *Information Sciences*, 545, 620–632. <https://doi.org/10.1016/j.ins.2020.09.037>
- [24] Wang, X., & Xu, Y. (2019). An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. *IOP Conference Series: Materials Science and Engineering*, 569(5). <https://doi.org/10.1088/1757-899X/569/5/052024>

- [25] Baig, M. A., Zaman, Q., Baig, S. A., Qasim, M., Khalil, U., Khan, S. A., Ismail, M., Muhammad, S., & Ali, S. (2021). Regression analysis of hydro-meteorological variables for climate change prediction: A case study of Chitral Basin, Hindukush region. *Science of the Total Environment*, 793. <https://doi.org/10.1016/j.scitotenv.2021.148595>
- [26] Gómez, J., Pérez-Aradros, B., & Salazar, I. (2022). How to beat early movers: The role of competitive strategy and industry dynamism on followers' performance in the telecommunications industry. *Long Range Planning*, 55(5). <https://doi.org/10.1016/j.lrp.2022.102244>
- [27] Guizzard, A., Ballestra, L. V., & D'Innocenzo, E. (2022). Hotel dynamic pricing, stochastic demand and covid-19. *Annals of Tourism Research*, 97. <https://doi.org/10.1016/j.annals.2022.103495>