

Analisis Segmentasi Pelanggan Dalam Preferensi Pembelian Produk Menggunakan Metode K-Means: Studi Kasus Industri Sepatu

Muhammad Rafif Hardyanto¹, Ni'matus Sa'adah², Arsyana Oktavia Prameswari³, Michael Rizky Setiawan⁴

¹ Universitas Telkom and rafifhardyanto@student.telkomuniversity.ac.id

² Universitas Telkom and nmsmdn@student.telkomuniversity.ac.id

³ Universitas Telkom and arsyanarsyv@student.telkomuniversity.ac.id

⁴ Universitas Telkom and michaelrizky@student.telkomuniversity.ac.id

ABSTRAK

Industri sepatu telah menjadi salah satu sektor yang berkembang pesat dalam perekonomian global. Meskipun demikian, pemahaman yang komprehensif mengenai preferensi pelanggan dalam pembelian sepatu masih merupakan tantangan yang relevan. Faktor-faktor seperti tren, budaya, dan elemen lainnya memainkan peran penting dalam dinamika ini. Oleh karena itu, tujuan dari penelitian ini adalah untuk mengeksplorasi dan memahami preferensi pelanggan dalam konteks pembelian sepatu. Kami memfokuskan penelitian ini pada dataset penjualan sepatu yang diperoleh dari PT. XYZ, dengan tujuan akhir untuk mengidentifikasi dan memahami segmen pelanggan serta variabilitas yang mendasarinya. Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan metode clustering, khususnya k-means, untuk mengelompokkan data penjualan sepatu berdasarkan karakteristiknya. Penggunaan metode ini memberikan kemampuan untuk mengidentifikasi pola dan hubungan yang mungkin tidak terdeteksi secara langsung. Hasil penelitian ini menunjukkan 3 pembagian kluster pelanggan yang terdiri dari *cluster customers are quite satisfied*, *customers are very satisfied*, dan *customers are not satisfied*, serta nilai *centroid* yang terkait dengan masing-masing *cluster*. Penelitian ini terdapat adanya peluang untuk meningkatkan pengelolaan hubungan pelanggan dan pengembangan strategi pemasaran yang lebih terfokus. Dengan pemahaman yang lebih baik tentang preferensi pelanggan, perusahaan dapat menyesuaikan produk dan layanan dengan lebih efektif, meningkatkan loyalitas pelanggan, dan mengoptimalkan kinerja penjualan.

Kata Kunci: K-Means, Clustering, Preferensi Pelanggan, Customer

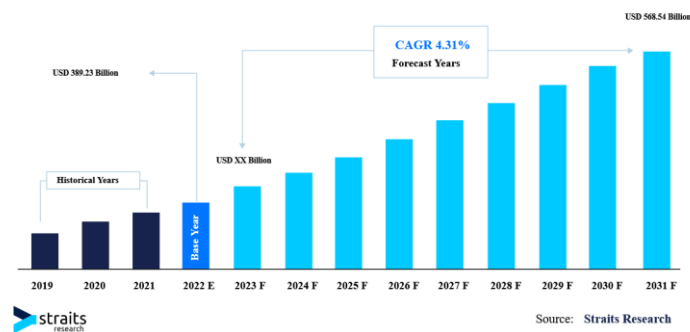
ABSTRACT

The shoe industry has become one of the fastest-growing sectors in the global economy. Nevertheless, understanding customer preferences when purchasing shoes is still a relevant challenge. Factors such as trends, culture, and other elements are essential in this dynamic. Therefore, this research aims to explore and understand customer preferences in shoe purchasing. We focused this research on the shoe sales dataset obtained from PT. XYZ, with the ultimate goal of identifying and understanding customer segments and their underlying variability. This research uses a quantitative approach by applying the clustering method, especially k-means, to group shoe sales data based on their characteristics. This method allows for identifying patterns and relationships that may not be detected directly. The results of this research show 3 divisions of customer clusters consisting of customers who are quite satisfied, customers who are very satisfied, and customers who are not satisfied, as well as the centroid values associated with each cluster. This research provides opportunities to improve customer relationship management and develop more focused marketing strategies. With a better understanding of customer preferences, companies can customize products and services more effectively, increase customer loyalty, and optimize sales performance.

Keywords: K-means, Clustering, Customer preferences, Customer

PENDAHULUAN

Industri fashion menjadi salah satu industri yang termasuk konsisten dalam mempertahankan popularitasnya dari waktu ke waktu. Seiring dengan perkembangan zaman, globalisasi, dan pergeseran budaya, industri ini telah mengalami transformasi besar dalam beberapa dekade terakhir. Karena perkembangannya yang dinamis dan dapat beradaptasi dengan zaman, dampak yang dihasilkan industri fashion begitu besar karena membawa pengaruh pada berbagai aspek, seperti ekonomi, lingkungan, dan sosial. Hal tersebut dibuktikan oleh artikel pers yang dikutip dari website Kementerian Pariwisata dan Ekonomi Kreatif, subsektor fashion masih menjadi andalan ekspor ekonomi kreatif Indonesia dengan nilai kontribusi sebesar 61,6 persen. Latar belakang ini menyoroti pentingnya industri fashion sebagai salah satu motor dalam penggerak ekonomi yang tetap populer. Salah satu industri fashion yang populer adalah industri sepatu. Hal tersebut dibuktikan dengan adanya data yang dikutip dari situs *straits research* yang menyatakan bahwa pasar sepatu di seluruh dunia bernilai USD 389,23 miliar pada tahun 2022. Grafik tersebut ditunjukkan pada Gambar 1.



Gambar 1. Grafik Perkembangan Pasar Sepatu

Pada gambar 1 menjelaskan mengenai peramalan pertumbuhan pasar sepatu di masa yang akan datang. Diperkirakan akan mencapai USD 568,54 miliar pada tahun 2031, meningkat dengan CAGR 4,31% selama periode perkiraan (2023-2031). Fenomena ini tidak hanya dipicu oleh kebutuhan fungsional untuk melindungi kaki, tetapi juga oleh faktor-faktor budaya, mode, dan gaya hidup.

Dalam menghadapi persaingan pasar yang bersifat global, maka perusahaan-perusahaan dituntut untuk terus berinovasi (Achmad dkk., 2024), sehingga eksistensi mereka di mata pelanggan tetap terjaga bahkan diharapkan terus meningkat. Pemahaman menyeluruh tentang perilaku pelanggan sangat penting bagi keberhasilan perusahaan dalam merancang strategi pemasaran yang sukses. Di era digital, pelanggan secara aktif mencari informasi untuk membuat keputusan pembelian dan tidak hanya bergantung pada satu merek atau produk (Santoso, 2018; Abdul Rehman, 2012).

Bagi PT. XYZ yang bergerak di industri sepatu, memahami preferensi pelanggan adalah kunci utama untuk mencapai kesuksesan. Dengan memahami kebutuhan dan keinginan pelanggan, PT. XYZ dapat menyesuaikan produk dan layanannya dengan lebih efektif, sehingga dapat meningkatkan kepuasan dan loyalitas pelanggan. Hal ini ultimately akan meningkatkan penjualan dan profitabilitas perusahaan.

Dalam situasi ini, analisis segmentasi pelanggan menjadi alat yang sangat berguna untuk menentukan preferensi pembelian pelanggan. Oleh karena itu, penting bagi toko sepatu untuk mengetahui preferensi pelanggan terkait jenis sepatu yang mereka inginkan. Salah satu cara pemilik toko sepatu untuk mengetahui minat pelanggan dalam pemilihan sepatu, dapat menggunakan metode *K-Means* sebagai metode untuk mengklasifikasikan minat pelanggan berdasarkan faktor-faktor preferensi yang ada.

K-means merupakan metode klusterisasi mengelompokkan data berdasarkan klasifikasi ke dalam beberapa kelompok berdasarkan karakteristiknya (Yadav, 2013; Abdul-Nasir dkk., 2015). Penggunaan metode *k-means*, membuat data *customer* sepatu dapat dikelompokkan menjadi beberapa *cluster*. Setiap *cluster* akan menunjukkan segmen pasar yang memiliki preferensi yang berbeda. Selain itu, dapat membantu toko dalam proses pengambilan keputusan untuk meningkatkan penjualan dan kepuasan pelanggan.

LANDASAN TEORI

A. Preferensi Pelanggan

Preferensi pelanggan didefinisikan sebagai pilihan baik suka maupun tidak seseorang terhadap suatu produk baik itu jasa yang digunakan atau barang yang dikonsumsi (Erinda, 2016). Preferensi ini terbentuk berdasarkan berbagai faktor, seperti kebutuhan, keinginan, dan harapan konsumen terhadap produk atau layanan tertentu (Kotler & Keller, 2016). Preferensi sering kali dipengaruhi oleh faktor psikologis, budaya, dan sosial yang membentuk persepsi pelanggan terhadap nilai dari produk atau layanan.

B. Clustering

Clustering merupakan metode untuk mengelompokkan data menjadi beberapa kelompok yang sesuai (Nahjan, 2023). Metode ini dapat mengidentifikasi pola tersembunyi atau struktur dalam dataset yang besar dan kompleks (Achmad dkk., 2024). Teori *Clustering* berfokus pada pembentukan kelompok sedemikian rupa sehingga objek dalam satu cluster memiliki tingkat kemiripan yang tinggi, sementara objek antar-cluster memiliki perbedaan yang signifikan (Achmad dkk., 2023). Tujuan digunakannya metode *clustering* adalah untuk mengetahui struktur data yang belum diketahui sebelumnya. Terdapat beberapa jenis *clustering* yang umumnya digunakan, seperti *k-means*, *DBSCAN*, dan *hierarchical clustering*.

C. Data Visualization

Memvisualisasikan data menggunakan *plot* dan grafik adalah cara yang efektif untuk data dan mengidentifikasi pola dan hubungan (Tariq, 2023). Menurut Achmad dkk. (2024), data visualization merupakan proses penyajian data dalam bentuk grafis atau visual, seperti grafik, diagram, peta, dan infografis, dengan tujuan mempermudah pemahaman informasi kompleks secara cepat dan efisien. Visualisasi data bertujuan untuk menyampaikan informasi agar mudah dipahami oleh pengguna. Salah satu prinsip penting dalam data visualization adalah komunikasi efektif yang melibatkan penyajian informasi

yang jelas dan tidak membingungkan (Achmad dkk., 2023). Terdapat beberapa jenis visualisasi yang umumnya dapat digunakan untuk memvisualisasikan data, seperti menggunakan *scatter plot*, *boxplot*, *histogram*, *heat map*, dsb.

D. K-means Clustering

K-means merupakan metode yang digunakan untuk mengelompokkan data dengan mengambil parameter dari beberapa *k-cluster* dan membagi data ke dalam klaster-klaster tersebut berdasarkan kemiripan antara data dalam suatu klaster dan perbedaan antara klaster (Ashari, 2022). Algoritma ini berfungsi dengan membagi data ke dalam sejumlah *K cluster*, di mana setiap data poin dimasukkan ke *cluster* dengan *centroid* terdekat (Sinaga dkk., 2020). Pendekatan ini mengasumsikan bahwa *cluster* berbentuk bulat dan memiliki distribusi yang serupa di sekitar *centroid* masing-masing.

METODE PENELITIAN

Metode yang digunakan untuk melakukan *clustering* dalam penelitian ini adalah menggunakan *k-means*. Objek yang digunakan dalam penelitian ini mengambil dataset penjualan seperti di PT XYZ. Berikut adalah langkah-langkah dalam *k-means clustering*:

1. Menentukan *k* sebagai banyaknya *cluster* yang akan dibentuk, dilakukan metode *elbow* yang dihasilkan dari perbandingan SSE. Berikut merupakan rumus dari SSE:

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} ||x_i - C_k||^2$$

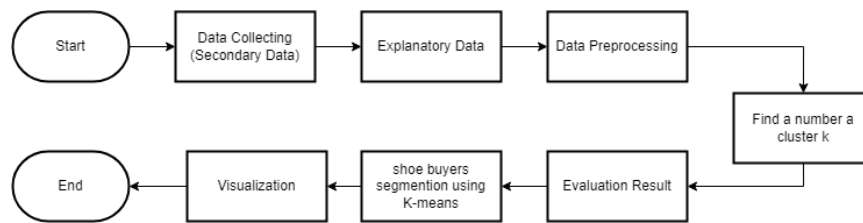
2. Menentukan *k centroid* secara random/acak
3. Menghitung jarak objek dari *centroid*
4. Kelompokkan objek berdasarkan jarak terpendek
5. Hitung *cluster* baru

Atribut dan *data scale* yang digunakan pada studi kasus ini dijelaskan oleh Tabel 1.

Tabel 1- Atribut dan Data Scale

| No | Atribut | Data Scale |
|----|-----------------|--------------------------------------|
| 1 | <i>Brand</i> | Banyaknya <i>brand</i> yang tersedia |
| 2 | <i>Model</i> | Banyaknya model yang tersedia |
| 3 | <i>Type</i> | Banyaknya tipe sepatu yang tersedia |
| 4 | <i>Gender</i> | Rata-rata <i>gender</i> pelanggan |
| 5 | <i>Color</i> | Banyaknya warna yang tersedia |
| 6 | <i>Material</i> | Banyaknya Material yang tersedia |
| 7 | <i>Price</i> | Tingkat harga produk |

Tabel 1 berisi atribut atau variabel apa saja yang akan digunakan untuk membuat segmentasi kepada pelanggan. Segmentasi ini akan diolah melalui metode klasterisasi *k-means*, dimana data yang memiliki karakteristik yang sama akan dikelompokkan ke dalam sebuah *cluster* yang sama. Langkah-langkah pada penelitian ini ditunjukkan pada Gambar 2.



Gambar 2. Tahapan Penelitian

Gambar 2 merupakan diagram alir dari penelitian yang dilakukan, penelitian diawali dengan data collecting yang menggunakan data sekunder, explanatory data, data preprocessing using finding a number of cluster k, evaluation result, segmentation using k-means, dan visualization.

HASIL DAN PEMBAHASAN

A. Data Collecting

Dalam studi kasus ini, data penjualan sepatu di PT XYZ merupakan data sekunder yang diambil dari salah satu situs *dataset* yaitu *kaggle*. Data awalnya berjumlah 1000 data, namun pada studi kasus ini, data yang digunakan berjumlah 85 data dengan 7 atribut yaitu: *brand, model, type, gender, color, material, dan price*. Data penjualan sepatu ditunjukkan pada Tabel 2.

Tabel 2. Data Collection

| No | Brand | Model | Type | Gender | Color | Material | Price |
|-----|----------|-----------------|------------|--------|-----------|-----------|-------|
| 0 | Nike | Air Jordan 1 | Basketball | Men | Red/Black | Leather | 170 |
| 1 | Adidas | Ultra Boost 21 | Running | Men | Black | Primeknit | 180 |
| 2 | Reebok | Classic Leather | Casual | Men | White | Leather | 75 |
| 3 | Converse | Chuck Taylor | Casual | Women | Navy | Canvas | 55 |
| 4 | Puma | Future Rider | Lifestyle | Women | Pink | Mesh | 80 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 80 | Nike | Blazer Low | Lifestyle | Women | White | Leather | 85 |
| 81 | Nike | Air Force 1 | Basketball | Men | White | Leather | 90 |
| 82 | Adidas | NMD_R1 | Lifestyle | Men | Black | Primeknit | 140 |
| 83 | Reebok | Nano X1 | Training | Women | Grey | Flexweave | 130 |
| 84 | Converse | Jack Purcell | Casual | Men | White | Canvas | 65 |

Tabel 2 menjelaskan atribut yang digunakan dalam pengumpulan data pada studi kasus ini. Atribut-atribut tersebut mewakili preferensi yang digunakan pelanggan dalam pembelian sepatu.

B. Exploratory Data

Pada set data yang dipilih dalam studi kasus ini, data diolah dengan menggunakan *python*. Eksplorasi data adalah proses memeriksa dan memahami data untuk mendapatkan wawasan dan mengidentifikasi pola atau hubungan (Tariq, 2023). Setelah *exploratory data* selesai, data dapat diidentifikasi dan dapat dibaca dengan lebih mudah

C. Data Preprocessing

Data preprocessing adalah proses yang digunakan untuk mempersiapkan data mentah untuk selanjutnya akan diproses menggunakan proses lain. Pada tahap *preprocessing data* akan dilakukan normalisasi data yang bertujuan untuk memeriksa apakah didalam data terdapat nilai yang hilang, *noisy data*, dan ketidakkonsistenan lain sebelum mengeksekusinya ke dalam algoritma (Tabianan, 2023). Pada tahap ini, akan dilakukan pemeriksaan untuk nilai tipe data dan pemeriksaan

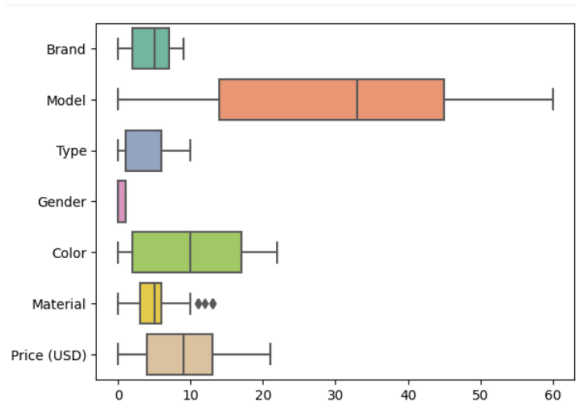
outlier pada data dengan menggunakan bantuan *boxplot*. Hasil pengecekan *missing value* ditunjukkan pada Gambar 3.

```
df.dtypes
df.isnull().sum()

Brand      0
Model      0
Type       0
Gender     0
Color      0
Material   0
Price (USD) 0
dtype: int64
```

Gambar 3. Missing Value

Gambar 3 menjelaskan bahwa semua atribut yang digunakan dalam studi kasus ini memiliki *missing value* yang sama yaitu 0, berarti tidak ada nilai yang hilang dari setiap variabel yang digunakan. Pada *dtype* menunjukkan tipe data yang digunakan adalah data bilangan bulat (*integer*). Jika tidak terdapat *missing value*, maka dilanjutkan ke tahap selanjutnya yaitu pemeriksaan dengan menggunakan *boxplot*. *Boxplot* dapat memeriksa *outlier*, menilai rentang data, dan menunjukkan perbedaan di antara kelompok (Shreffler, 2023). Hasil analisis data *outlier* menggunakan *boxplot* ditunjukkan pada Gambar 4.



Gambar 4. Analisis Boxplot

Boxplot merupakan metode analisis yang ditampilkan secara grafis berdasarkan kuartil untuk membantu menggambarkan sekumpulan data yang ada. Gambar 4 menjelaskan bahwa, berdasarkan hasil pengecekan data *outlier* dengan menggunakan *boxplot* terdapat data *outlier* pada bagian atribut material. Oleh karena itu perlu pada atribut-atributnya akan dilakukan normalisasi/standarisasi *z-score* dengan menggunakan fungsi *StandardScaler*, kemudian hasilnya akan disimpan sebagai sebuah *dataframe* dengan nama "std_atr" atau *standardized attributes*. Hasil dari standarisasi ditunjukkan pada Gambar 5.

```

from sklearn.preprocessing import StandardScaler
scale=StandardScaler()
std_atr=scale.fit_transform(df)
std_atr=pd.DataFrame(std_atr,columns=df.columns)
print(std_atr)

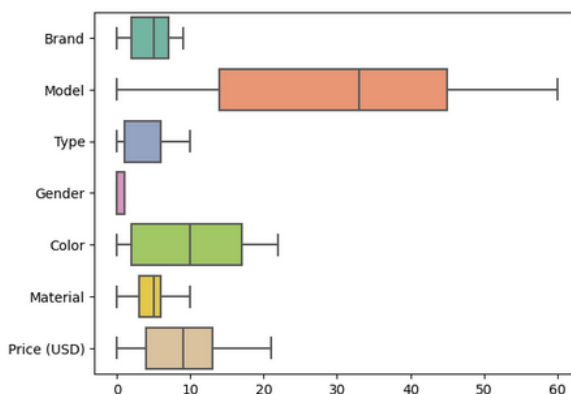
```

| | Brand | Model | Type | Gender | Color | Material | Price (USD) |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-------------|
| 0 | 0.189517 | -1.405498 | -1.593412 | -0.965307 | 0.927106 | -0.583474 | 1.729911 |
| 1 | -1.561457 | 1.467125 | 0.484951 | -0.965307 | -1.320156 | 0.883842 | 2.116864 |
| 2 | 0.889907 | -0.842238 | -1.247018 | -0.965307 | 1.076923 | -0.583474 | -0.591812 |
| 3 | -0.861067 | -0.898564 | -1.247018 | 1.035940 | 0.178018 | -1.463863 | -1.365719 |
| 4 | 0.539712 | 0.002651 | 0.138558 | 1.035940 | 0.477653 | 0.003453 | -0.398335 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 80 | 0.189517 | -1.067542 | 0.138558 | 1.035940 | 1.076923 | -0.583474 | -0.204858 |
| 81 | 0.189517 | -1.461824 | -1.593412 | -0.965307 | 1.076923 | -0.583474 | -0.011381 |
| 82 | -1.561457 | 0.791214 | 0.138558 | -0.965307 | -1.320156 | 0.883842 | 0.956003 |
| 83 | 0.889907 | 0.903866 | 1.524133 | 1.035940 | 0.028201 | -1.170400 | 0.762526 |
| 84 | -0.861067 | 0.622236 | -1.247018 | -0.965307 | 1.076923 | -1.463863 | -0.978765 |

[85 rows x 7 columns]

Gambar 5. Standarisasi Data

Gambar 5 menunjukkan bahwa data telah ternormalisasi dengan baik. Selanjutnya akan dilakukan pengecekan *outlier* kembali dengan menggunakan *boxplot* untuk memastikan tidak ada *outlier* pada data. Hasil analisis data *outlier* setelah data dilakukan normalisasi ditunjukkan pada Gambar 6.

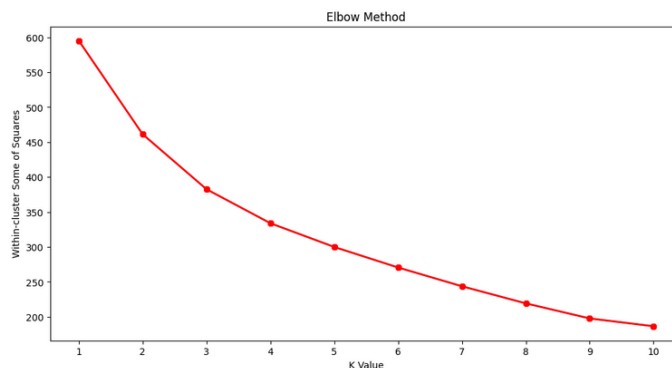


Gambar 6. Analisis Boxplot Setelah Normalisasi

Gambar 6 menunjukkan bahwa, setelah data dilakukan normalisasi, tidak teridentifikasi adanya *outlier* sehingga data dapat dikatakan layak untuk dilakukan pengolahan data.

D. Model Tuning

Proses selanjutnya adalah *model tuning*. *Model tuning* adalah proses yang dapat dilakukan setelah mendapatkan dataset, yang dimulai dengan menentukan tujuan yang ingin dicapai dengan memilih algoritma menggunakan contoh-contoh *model tuning* seperti klasifikasi, pengelompokan, segmentasi, atau uji *statistic* (Achmad, 2023). Pada algoritma *k-means clustering* terdapat nilai parameter *k*, sehingga sebelum diputuskan subjektif nilai *k*, dapat dilakukan *model tuning*, misalnya dengan metode *Elbow*. Metode *elbow* bertujuan untuk mencari tahu berapa kelompok atau jumlah klaster yang optimal (Abdullah, 2023). Hasil analisis dengan menggunakan metode *elbow* ditunjukkan pada Gambar 7.



Gambar 7. Hasil Metode Elbow

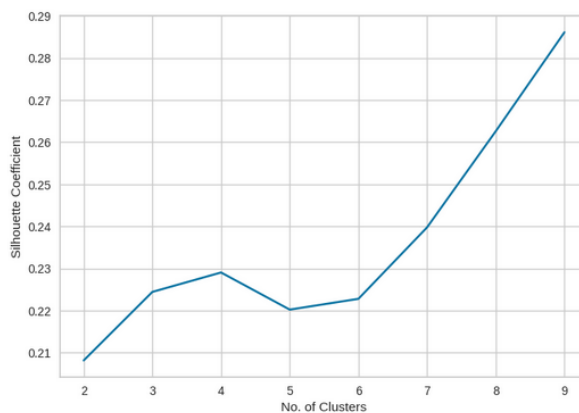
Berdasarkan Gambar 7, hasil analisis metode *elbow* terhadap data ditunjukkan pada tanda panah yang mengarah pada titik *cluster* yang membentuk siku, dimana nilai ($k=3$). Hal tersebut dikarenakan nilai inersia mengalami penurunan yang mulai mengecil. Untuk memperkuat dugaan pada titik ini kondisi k merupakan nilai yang optimal, maka akan dilakukan evaluasi dengan menggunakan metode *silhouette coefficient*. Metode *silhouette coefficient* berguna untuk menilai kualitas dan kekuatan *cluster*, terkait dengan posisi objek ketika ditempatkan dalam suatu *cluster* (Guntara, 2023). Hasil gambar analisis menggunakan metode *silhouette coefficient* dijelaskan pada Gambar 8.

```
from sklearn.metrics import silhouette_score
km = KMeans(n_clusters=3, random_state=42)
km.fit(std_atr)
score = silhouette_score(std_atr, km.labels_)
print('Silhouetter Score: %.3f' % score)

Silhouetter Score: 0.225
```

Gambar 8. Hasil Silhouette Coefficient

Gambar 8 merupakan hasil analisis dengan metode *silhouette coefficient*, dengan *silhouette score* yang terpilih sebesar 0.225. Hasil skor tersebut merupakan skor yang terpilih dari seluruh *cluster silhouette coefficient*. Selanjutnya adalah menentukan jumlah *cluster* berdasarkan hasil dari setiap *silhouette coefficient*. Hasil grafik *silhouette coefficient* ditunjukkan pada Gambar 9.



Gambar 9. Grafik Hasil Silhouette Coefficient

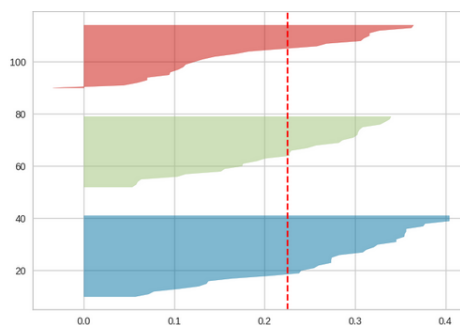
Gambar 9 menunjukkan grafik yang menunjukkan hasil dari nilai *silhouette coefficient* yang diperoleh dari setiap *cluster*. Untuk mengetahui detail dan analisis nilai *silhouette coefficient* pada setiap *cluster* dapat dilihat pada Gambar 10.

```

For n_clusters=2, Silhouette Coefficient = 0.20815496061086444
For n_clusters=3, Silhouette Coefficient = 0.22444123495612595
For n_clusters=4, Silhouette Coefficient = 0.2290508361725455
For n_clusters=5, Silhouette Coefficient = 0.22023167863814835
For n_clusters=6, Silhouette Coefficient = 0.22279201703432325
For n_clusters=7, Silhouette Coefficient = 0.23977902640636886
For n_clusters=8, Silhouette Coefficient = 0.26261616071768007
For n_clusters=9, Silhouette Coefficient = 0.2861288963451525
    
```

Gambar 10. Hasil Nilai Setiap n-cluster

Gambar 10 menunjukkan nilai *silhouette coefficient* yang sesuai dengan *silhouette score* terdapat pada (*n-cluster*=3) dengan nilai 0.2246, hal tersebut karena mendekati nilai dari *silhouette score* yang terpilih sebelumnya yaitu 0.225. Setelah mendapatkan jumlah *cluster* yang dipilih, dilakukan pemeriksaan pada *model tuning* untuk mengetahui apakah objek yang ada pada ketiga *cluster* memiliki *silhouette score* yang positif atau negatif. Hasil pemeriksaan *model tuning* dapat dilihat pada Gambar 11.



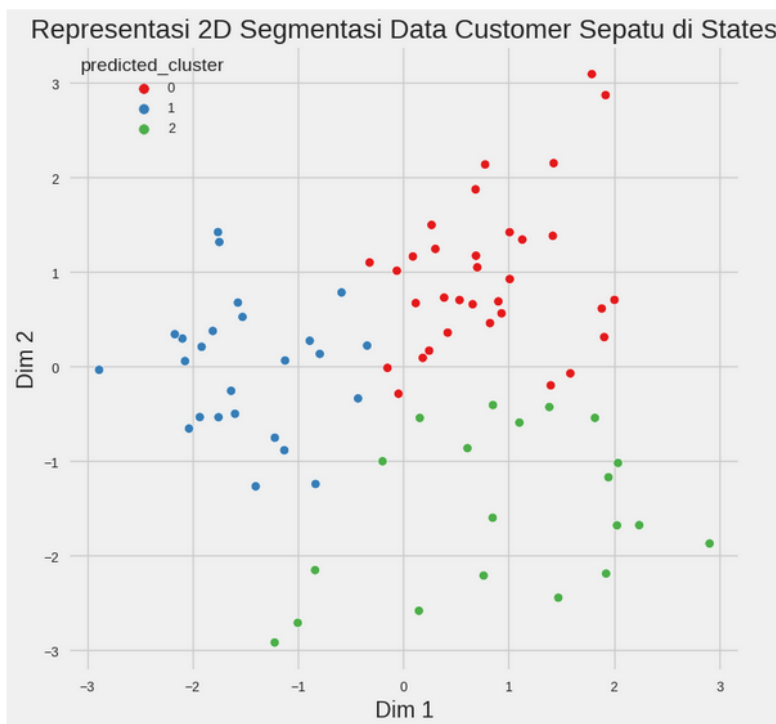
Gambar 11. Hasil Tuning Model

Berdasarkan hasil *tuning model* yang ditunjukkan oleh Gambar 11, ketiga *cluster* yang dipilih telah melewati garis merah dan dapat dikatakan bahwa *silhouette score* yang dihasilkan bernilai positif.

E. Evaluation Result

Setelah didapatkan hasil bahwa *silhouette score* yang dihasilkan bernilai positif dengan jumlah *cluster* adalah 3, langkah selanjutnya adalah hasil pengolahan data tersebut akan divisualisasikan pada Gambar 12. Gambar 12 menunjukkan hasil visualisasi menggunakan *scatter plot*, titik-titik objek visualisasi hasil segmentasi terbagi menjadi 3 *cluster* dengan warna yang berbeda, *cluster* 0 berwarna merah, *cluster* 1 berwarna biru, dan *cluster* 2 berwarna hijau. Tahap ini akan menghitung jarak setiap data ke *centroid* yang telah dibuat dengan menggunakan *Euclidean distance*.

Nilai *centroid* untuk setiap variabel mempengaruhi penilaian terhadap cluster yang terbentuk. Nilai *centroid* untuk variabel variabel pada setiap cluster dijelaskan pada Tabel 3.



Gambar 12. Scatter Plot Customer Segmentation

Tabel 3. Nilai Centroid Atribut

| Centroids | | | | | | | |
|-----------|--------|--------|--------|--------|--------|----------|--------|
| Cluster | Brand | Model | Type | Gender | Color | Material | Price |
| 0 | -0.719 | -0.149 | 0.323 | 0.098 | -0.435 | 0.600 | 0.769 |
| 1 | 1.184 | 0.453 | 0.817 | 0.075 | -0.505 | 0.062 | -0.530 |
| 2 | -0.236 | -0.235 | -1.099 | -0.179 | 0.949 | -0.741 | -0.405 |

Langkah selanjutnya adalah melakukan identifikasi dari setiap cluster yang telah ditemukan. Tabel 3 merupakan hasil centroid dari setiap atribut yang terdiri dari brand, model, type, gender, color, material, dan price. Dari pengelompokan tersebut, dapat diketahui mana cluster yang merupakan cluster terbaik dengan melihat nilai dari centroid masing-masing cluster. Pada cluster 0, nilai centroid yang paling signifikan adalah price, material, dan type dengan nilai 0.769, 0.600, dan 0.323. Untuk cluster 1, nilai centroid yang paling signifikan adalah brand, type, dan model dengan nilai 1.184, 0.817, dan 0.453. Selanjutnya untuk cluster 2, nilai centroid yang paling signifikan adalah color dengan nilai 0.949. Representasi nilai centroid setiap nama cluster ditunjukkan oleh Tabel 4.

Tabel 4. Cluster Value

| Cluster | Cluster Value | Cluster Name |
|---------|------------------|-------------------------------|
| 0 | Moderate (0.069) | Customers are quite satisfied |
| 1 | Best (0.222) | Customers are very satisfied |
| 2 | Poor (-0.278) | Customers are not satisfied |

Tabel 4 menunjukkan bahwa setiap cluster diberikan nama sesuai dengan valuenya masing-masing. Value cluster terbagi menjadi 3, yaitu poor, moderate, dan best. Setiap customer yang masuk dalam cluster 0 termasuk kedalam moderate cluster dengan nilai 0.069 dan diberi nama pelanggan cukup puas, customer pada cluster 1 termasuk kedalam best cluster dengan nilai 0.222 dan diberi nama

pelanggan sangat puas, dan terakhir *customer* pada *cluster* 2 termasuk kedalam *poor cluster* dengan nilai -0.278 dan diberi nama pelanggan kurang puas.

Berdasarkan nilai *centroid* dari variabel penelitian yang diteliti, maka terdapat rekomendasi dari studi kasus ini untuk strategi dalam peningkatan kepuasan pelanggan dalam pembelian sepatu. Strategi dan rekomendasi yang sesuai dapat meningkatkan kinerja dari organisasi atau perusahaan (Achmad dkk., 2023; Achmad dkk., 2024). Strategi yang tepat harus disesuaikan dengan karakteristik dan kebutuhan spesifik dari masing-masing *cluster* untuk memastikan solusi yang efektif dan berdampak langsung pada peningkatan kepuasan serta kinerja organisasi. Berikut adalah rekomendasi strategi dari setiap *cluster*:

1. Klaster 0: Jika toko ingin meningkatkan kepuasan pelanggan yang termasuk dalam *cluster* ini, toko dapat fokus pada 4 bagian yaitu *brand*, *model*, *gender*, dan *color*. Karena variabel-variabel tersebut masih memiliki nilai *centroid* yang belum signifikan bahkan bernilai negatif. Strategi untuk *brand* adalah menambahkan koleksi *brand* sepatu dengan kualitas yang baik. Strategi untuk *model* adalah menambah koleksi *model* yang disesuaikan *trend* pasar terbaru. Strategi *gender* adalah dengan menempatkan jenis sepatu dengan katalog atau etalase terpisah sehingga pelanggan lebih mudah dalam menemukan sepatu yang diinginkan. Strategi *color* dengan menyesuaikan preferensi dan *trend* pasar misalnya untuk laki-laki cenderung memilih warna yang terlihat *casual* namun masih elegan seperti *navy*, hitam, dsb. Sedangkan untuk perempuan cenderung memilih warna cerah atau lebih *soft*.
2. Klaster 1: Jika toko ingin meningkatkan kepuasan pelanggan yang termasuk dalam *cluster* ini, toko dapat fokus pada 2 bagian yaitu *color* dan *price*. Karena variabel-variabel tersebut memiliki nilai *centroid* negatif. Strategi untuk *color* dengan menyesuaikan preferensi dan *trend* pasar misalnya untuk laki-laki cenderung memilih warna yang terlihat *casual* namun masih elegan seperti *navy*, hitam, dsb. Sedangkan untuk perempuan cenderung memilih warna cerah atau lebih *soft*. Strategi *price* dengan memberikan diskon atau pemberian promo kepada pelanggan dengan syarat atau pada saat *event* tertentu.
3. Klaster 2: Jika toko ingin meningkatkan kepuasan pelanggan yang termasuk dalam *cluster* ini, toko dapat fokus pada 6 bagian yaitu *brand*, *model*, *type*, *gender*, *material* dan *price*. Karena variabel-variabel tersebut masih memiliki nilai *centroid* yang bernilai negatif. Strategi untuk *brand* adalah menambahkan koleksi *brand* sepatu dengan kualitas yang baik. Strategi untuk *model* adalah menambah koleksi *model* yang disesuaikan *trend* pasar terbaru. Strategi untuk *type* adalah menyesuaikan dengan *tren* pasar, menyediakan lebih banyak tipe sepatu. Strategi *gender* adalah dengan menempatkan jenis sepatu dengan katalog atau etalase terpisah sehingga pelanggan lebih mudah dalam menemukan sepatu yang diinginkan. Strategi *material* dengan memilih material yang berkualitas, memberikan pedoman bagi pelanggan terkait referensi material yang digunakan pada sepatu terkait kekurangan dan kelebihan sehingga pelanggan dapat melakukan pertimbangan sebelum membeli sepatu. Strategi *price* dengan memberikan diskon atau pemberian promo kepada pelanggan dengan syarat atau pada saat *event* tertentu.

KESIMPULAN

Berdasarkan hasil penelitian, strategi peningkatan kepuasan pelanggan dalam membeli sepatu di PT. XYZ akan lebih efektif jika kita mengetahui preferensi pelanggan dalam memilih sepatu. Untuk mengetahui tingkat kepuasan pelanggan dalam membeli sepatu terbagi menjadi 3 *cluster* yaitu pelanggan cukup puas, pelanggan sangat puas, dan pelanggan kurang puas. Dengan pendekatan ini, perusahaan dapat merancang tindakan yang tepat sasaran untuk meningkatkan kualitas produk dan layanan toko, sehingga kepuasan pelanggan dapat meningkat secara signifikan. Secara keseluruhan, pemahaman tentang preferensi dan segmentasi pelanggan merupakan langkah penting dalam mencapai tujuan peningkatan kepuasan pelanggan dan daya saing di pasar.

Penelitian ini dapat memberikan gambaran bagi toko sepatu agar dapat meningkatkan kepuasan pelanggan. Penelitian pada studi kasus ini memiliki kekurangan yang diharapkan dapat diperbaiki pada penelitian selanjutnya. Kekurangan tersebut yaitu dalam hal pengumpulan data karena hanya mengandalkan referensi dari data sekunder/dataset bukan dari observasi langsung, sehingga mungkin ada data yang terlewat atau kurang relevan.

REFERENSI

- Abdullah, A. &. (2023). Liver Disease Classification Using the Elbow Method to Determine Optimal K in the K-Nearest Neighbor (K-NN) Algorithm. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 12(2), 218-228. DOI: <https://doi.org/10.32736/sisfokom.v12i2.1643>
- Abdul-Nasir, A. S., Mashor, M. Y., Halim, N. H. A., & Mohamed, Z. (2015, May). The cascaded moving k-means and fuzzy c-means clustering algorithms for unsupervised segmentation of malaria images. In *AIP Conference Proceedings* (Vol. 1660, No. 1). AIP Publishing.
- Abdul Rehman, A. (2012). Customer satisfaction and service quality in Islamic banking: A comparative study in Pakistan, United Arab Emirates and United Kingdom. *Qualitative Research in Financial Markets*, 4(2/3), 165-175.
- Achmad, F., & Wiratmadja, I. I. (2024). Driving Sustainable Performance in SMEs through Frugal Innovation: The Nexus of Sustainable Leadership, Knowledge Management, and Dynamic Capabilities. *IEEE Access*, 12, 103329 - 103347. DOI: 10.1109/ACCESS.2024.3433474
- Achmad, F., & Wiratmadja, I. I. (2024). Strategic advancements in tourism development in Indonesia: Assessing the impact of facilities and services using the PLS-SEM approach. *Journal Industrial Servicess*, 10(1), 49-62. DOI: <http://dx.doi.org/10.62870/jiss.v10i1.24494>
- Achmad, F., Prambudia, Y., & Rumanti, A. A. (2023). Improving tourism industry performance through support system facilities and stakeholders: The role of environmental dynamism. *Sustainability*, 15(5), 4103. DOI: <https://doi.org/10.3390/su15054103>
- Achmad, F., Prambudia, Y., & Rumanti, A. A. (2023). Sustainable Tourism Industry Development: A Collaborative Model of Open Innovation, Stakeholders, and Support System Facilities. *IEEE Access*, 11, 83343 - 83363. DOI: 10.1109/ACCESS.2023.3301574
- Achmad, F. A. (2023). Decision-Making Process for Tourism Potential Segmentation. *International Journal of Innovation in Enterprise System*, 7(01), 19-30.
- Achmad, F., Prambudia, Y., & Rumanti, A. A. (2023, December). Examining the Feedback Effects of Support System Facilities on Tourism Industry Performance: A Causal Loop Diagram Modeling Approach. In *2023 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 0768-0772). IEEE.
- Aditya, A. J. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019 . *Jurnal Media Informatika Budidarma*, 4(1), 51-58.

- Ashari, I. F. (2022). Application of data mining with the K-means clustering method and Davies Bouldin index for grouping IMDB movies. *Journal of Applied Informatics and Computing*.
- Erinda, A. &. (2016). Analisis Faktor-Faktor Preferensi Pelanggan Dan Pengaruhnya Terhadap Keputusan Pembelian. Study Terhadap Pelanggan McDonald's di Indonesia dan Malaysia. *Footwear Market Size, Share & Trends Analysis Report By Type (Athletic, Non-Athletic), By End-User (Men, Women, Children) and By Region(North America, Europe, APAC, Middle East and Africa, LATAM) Forecasts, 2023-2031* . (n.d.). Retrieved from strait research: <https://straitresearch.com/report/footwear-market>
- Guntara, M. &. (2023). Optimasi Cacah Klaster pada Klasterisasi dengan Algoritma KMeans Menggunakan Silhouette Coeficient dan Elbow Method. *JuTI" Jurnal Teknologi Informasi"*, 2(1), 43-52.
- Nahjan, M. R. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101-104.
- Santoso, P. Y. (2018). Transformasi Integrated Marketing Communication Di Era Digital. *Jurnal Pustaka Komunikasi*, 1(2), 313-326.
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE access*, 8, 80716-80727. DOI: 10.1109/ACCESS.2020.2988796
- Shreffler, J. &. (2023). *Exploratory data analysis: Frequencies, descriptive statistics, histograms, and boxplots*. . In StatPearls [Internet]. StatPearls Publishing.
- Tabianan, K. V. (2022). K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability*, 14(12), 7243. DOI: <https://doi.org/10.3390/su14127243>
- Tariq, D. T. (2023). Visualization and Explorative Data Analysis. *International Journal of Enhanced Research in Science, Technology & Engineering*, 12(3), 11-21.
- Yadav, J. &. (2013). A Review of K-mean Algorithm. *Int. J. Eng. Trends Technol*, 4(7), 2972-2976.