

## Analisis Segmentasi Pelanggan Mall Menggunakan Algoritma K-Means

### *Mall Customer Segmentation Analysis Using the K-Means Algorithm*

Yustia Hapsari<sup>1</sup>, Deddy Prihadi<sup>2</sup>, Agnes Dwita Susilawati<sup>3</sup>, Eka Matriayani<sup>4</sup>, Kafaby Juweriyah<sup>5</sup>

<sup>1,2,3,4,5</sup> Universitas Pancasakti Tegal, Kota Tegal

Corresponding author : [yustia.hapsari@gmail.com](mailto:yustia.hapsari@gmail.com)

### Abstrak

Segmentasi pelanggan adalah tindakan mengelompokkan konsumen ke dalam beberapa kategori berdasarkan karakteristik umum seperti usia, jenis kelamin, pendapatan tahunan, lokasi tempat tinggal, frekuensi pembelian, dan faktor lainnya. Hasil dari proses segmentasi ini digunakan sebagai panduan untuk merumuskan strategi pemasaran serta memperkenalkan produk baru kepada setiap kelompok pelanggan. Selain itu, hasil segmentasi juga membantu dalam mengembangkan produk khusus untuk kelompok pelanggan yang memiliki nilai ekonomi tinggi. Dengan menggunakan algoritma clustering seperti K-Means, pelaku bisnis dapat memastikan bahwa kebijakan yang diambil memiliki dasar yang sistematis dan terukur. Terdapat 5 kelompok pelanggan mall berdasarkan hasil analisis data, yakni: kelompok pertama adalah pelanggan yang berpenghasilan rata-rata dengan skor pengeluaran rata-rata. Kelompok kedua adalah pelanggan yang berpenghasilan tinggi dan memiliki skor pembelanjaan. Kelompok ketiga adalah pelanggan yang memiliki pendapatan lebih tinggi namun tidak menghabiskan lebih banyak uang di mall. Kelompok keempat adalah pelanggan yang berpenghasilan rendah dengan skor pengeluaran rendah. Kelompok kelima adalah pelanggan berpenghasilan rendah dengan skor pembelanjaan tinggi.

**Kata Kunci** : segmentasi, pelanggan, mall, k-mean

### PENDAHULUAN

Segmentasi pelanggan melibatkan proses mengelompokkan konsumen ke dalam kategori-kategori berbeda berdasarkan ciri-ciri umum seperti usia, jenis kelamin, pendapatan tahunan, lokasi tempat tinggal, pola pembelian, dan faktor-faktor lainnya. [1]. Hasil dari segmentasi pelanggan ini dapat menjadi pedoman untuk mengembangkan strategi pemasaran dan *cross selling* produk baru untuk setiap grup, dan pengembangan produk untuk kelompok pelanggan yang paling berharga [2]. Melalui proses ini setiap kebijakan yang dihasilkan dari pelaku bisnis memiliki kerangka kerja sistematis yang mapan.

Tempat berbelanja masyarakat seperti mall harus terus berinovasi agar laju penjualan setiap produknya bisa memenuhi target. Pihak pengelola mall harus mengetahui kebutuhan apa saja yang diperlukan oleh setiap calon pelanggannya. Jangan sampai sebuah mall menjual aneka ragam produk yang sebenarnya tidak begitu diminati oleh para pelanggan. Pengelola mall juga harus mengetahui tingkat pendapatan masyarakat sekitar, rentang usia dominan pelanggan yang aktif berbelanja, pengeluaran pelanggan setiap kali bertransaksi di mall, dan juga jenis kelaminnya. Apabila semua

kebutuhan tersebut dapat disegmentasikan akan sangat membantu pengambilan kebijakan pihak pengelola mall.

Penelitian ini bertujuan melakukan analisis segmentasi pelanggan sebuah mall. *Dataset* penelitian diperoleh dari Kaggle [3]. Terdapat empat atribut pada dataset yaitu *Customer ID*, *Gender* (Jenis Kelamin), *Age* (Usia), *Annual Income* (Pendapatan Tahunan), dan *Spending Score* (Skor Belanja). *Customer ID* adalah nomor unik yang digunakan untuk mewakili masing-masing pelanggan, sehingga setiap pelanggan memiliki ID yang unik dan berbeda dari yang lain. *Gender* dibagi menjadi dua kategori, yaitu *Male* (laki-laki) dan *Female* (perempuan). *Usia* diukur dalam tahun. *Pendapatan tahunan* diukur dalam ribuan dolar. *Spending score* adalah angka bulat (1-100) yang ditetapkan oleh mal berdasarkan perilaku dan kebiasaan belanja pelanggan. Semakin tinggi nilai *spending score*, semakin baik penilaian yang diberikan oleh mal terhadap pelanggan tersebut. Proses segmentasi menggunakan algoritma clustering, K-Means.

Algoritma K-Means adalah teknik pengelompokan non-hierarki di mana data dibagi menjadi satu atau beberapa kluster. Data dengan ciri-ciri serupa ditempatkan dalam kluster yang sama, sementara data yang memiliki perbedaan ciri-ciri dikelompokkan ke kluster yang berbeda [4]. K-Means sangat cocok digunakan untuk melakukan segmentasi pelanggan. Setelah dilakukan segmentasi akan dilakukan analisis untuk mendapatkan *insight* pengetahuan terkait strategi penjualan produk.

## METODE

### 1.1 Bahan Penelitian

Penelitian ini membutuhkan bahan penelitian berupa :

1. Artikel Penelitian Terbaru

Referensi utama dari penelitian ini menggunakan jurnal penelitian pada rentang periode publikasi tahun 2020-2022.

2. *Dataset*

*Dataset* yang akan digunakan adalah *customer segmentation dataset* yang diambil dari Kaggle.

### 1.2 Alat Penelitian

1. *Hardware*

Dalam merancang dan mengimplementasikan perangkat lunak, beberapa perangkat keras diperlukan. Perangkat keras utama meliputi komputer dengan spesifikasi tinggi, termasuk processor Core i7 3.4 GHz (8 CPUs), RAM 16 GB, Bios Version 5.5, VGA 8 GB, dan webcam Logitech 720p.

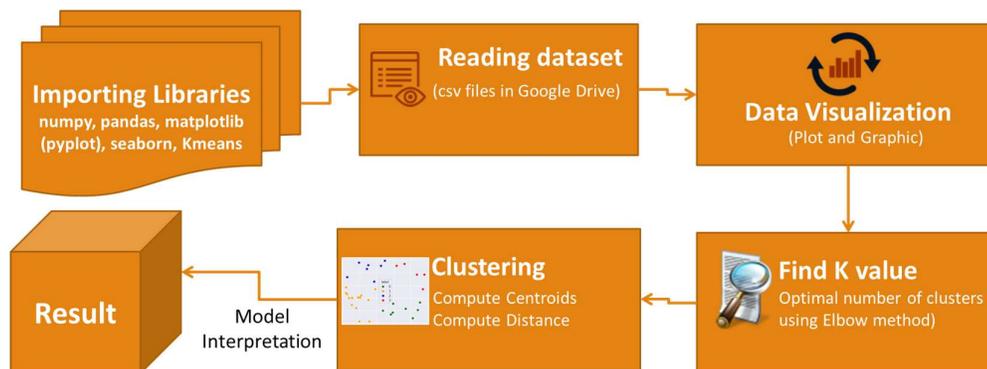
## 2. Software

Selain dukungan perangkat keras, dibutuhkan pula perangkat lunak yang digunakan dalam melakukan pengembangan sistem, diantaranya :

- a. *Anaconda 3.8*
- b. *Pycharm*
- c. *Sublime Editor*
- d. *Apache Web Server*
- e. *Google Chrome*
- f. *Mendeley Desktop*

### 1.3 Prosedur Penelitian

Tahap pertama melakukan *importing* berbagai *library* yang diperlukan. Setelah itu dilanjutkan dengan membaca *dataset* segmentasi pelanggan. *Dataset* yang sudah di-*import* akan divisualisasikan dalam bentuk plot dan grafik. Selanjutnya dilakukan pencarian nilai k optimal menggunakan metode Elbow yang nantinya akan diverifikasi dengan metode Silhouette. Tahap terakhir setelah melakukan segmentasi produk adalah melakukan interpretasi terhadap hasil analisis.



Gambar 1. Tahapan Prosedur Penelitian

## HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dataset segmentasi pelanggan yang di-download dari Kaggle (<https://www.kaggle.com/code/obrunet/customer-segmentation-k-means-analysis>). *Dataset* tersebut memiliki 5 atribut yaitu *CustomerID* (ID Pelanggan), *Gender* (Usia Gender), *Annual Income (k\$)* (Pendapatan Tahunan) dan *Spending Score (1-100)*

(Skor Pengeluaran). Suatu *mall* dapat mengetahui seluruh nilai pada tiap atribut tersebut berdasarkan data pada *member card*. Ketika pihak manajemen *mall* mampu melakukan analisis segmentasi pelanggan maka dapat merumuskan berbagai langkah strategis di awal untuk mendongkrak tingkat penjualan.

Keuntungan yang didapatkan dari hasil segmentasi pelanggan di antaranya menentukan harga produk yang sesuai, melakukan kampanye pemasaran yang lebih strategis, memilih fitur produk tertentu untuk diterapkan dan memprioritaskan upaya pengembangan produk baru.

### 1.3.1 *Import Library dan Dataset*

Tahap awal ini sangat penting karena berkaitan dengan penggunaan *library* pendukung untuk melakukan seluruh proses segmentasi. Jika ada *library* yang kurang atau tidak di-*import* akan berimbas ke hasil segmentasi. *Dataset* juga akan di-*import* ke dalam model yang selanjutnya akan diolah.

```

1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5
6 df = pd.read_csv("../input/customer-segmentation-tutorial-in-python/Mall_Customers.csv")
7 df.head()

```

*Dataset* yang berhasil di-*import* akan ditampilkan dalam bentuk *data frame* sebagai berikut:

Tabel 1. Dataset Pelanggan Mall

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Jika telah muncul tabel di atas maka proses *import* dataset berhasil.

### 1.3.2 *Visualisasi Data*

#### 1. Frekuensi Usia Pelanggan *Mall*

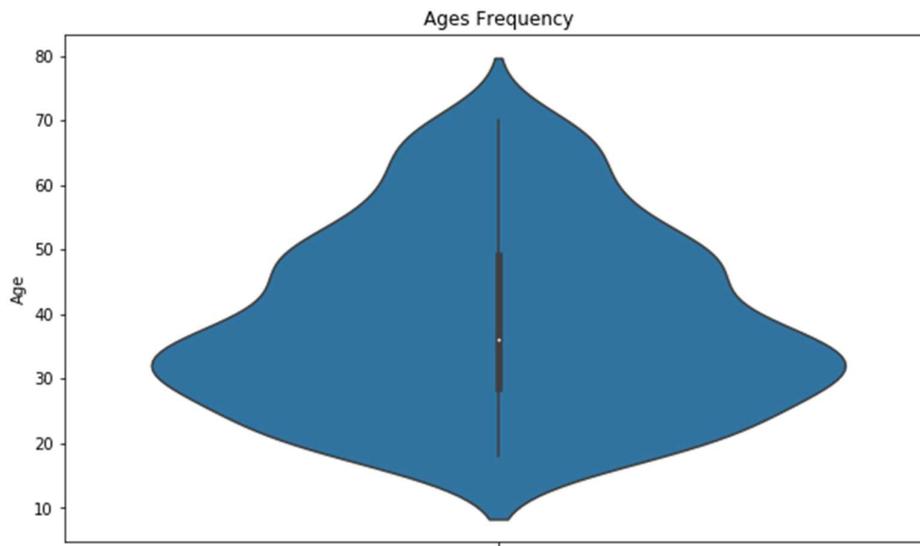
Sebelum dilakukan visualisasi terlebih dahulu dilakukan penghapusan kolom CustomerID. Penghapusan tersebut dilakukan karena kolom atau atribut CustomerID tidak begitu relevan.

```

1 df.drop(["CustomerID"], axis = 1, inplace=True)
2
3 plt.figure(figsize=(10,6))
4 plt.title("Ages Frequency")
5 sns.axes_style("dark")
6 sns.violinplot(y=df["Age"])
7 plt.show()

```

Visualisasi yang pertama kali dilakukan adalah menampilkan frekuensi sebaran usia. Tujuannya adalah untuk mengetahui pada *range* usia berapa pelanggan paling banyak berbelanja di *mall*.



Gambar 2. Frekuensi Usia Pelanggan *Mall*

Hasil visualisasi menunjukkan bahwa pelanggan *mall* paling banyak berada pada range usia 30 – 40 tahun.

## 2. Pengeluaran dan Pendapatan Tahunan

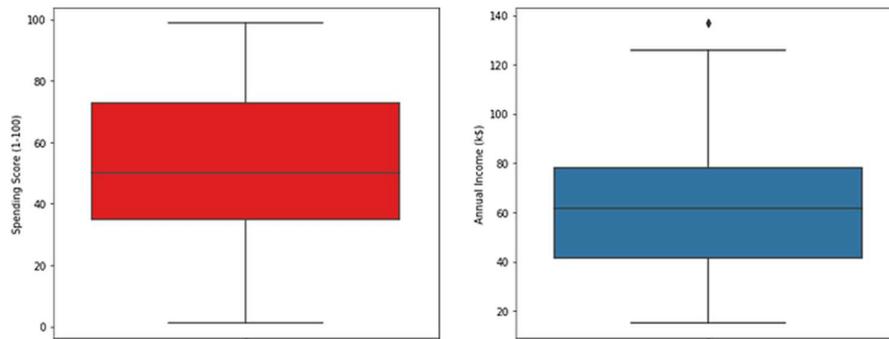
Visualisasi akan ditampilkan dalam bentuk *boxplot*.

```

1 plt.figure(figsize=(15,6))
2 plt.subplot(1,2,1)
3 sns.boxplot(y=df["Spending Score (1-100)"], color="red")
4 plt.subplot(1,2,2)
5 sns.boxplot(y=df["Annual Income (k$)"])
6 plt.show()

```

Komparasi antara Pengeluaran dan Pendapatan Tahunan perlu untuk divisualisasikan. Hal ini digunakan untuk melihat sejauh mana tingkat kemampuan belanja seorang pelanggan dibandingkan dengan pendapatannya.



Gambar 3. Distribusi Data Pengeluaran dengan Pendapatan Tahunan

### 3. Sebaran Data Jenis Kelamin

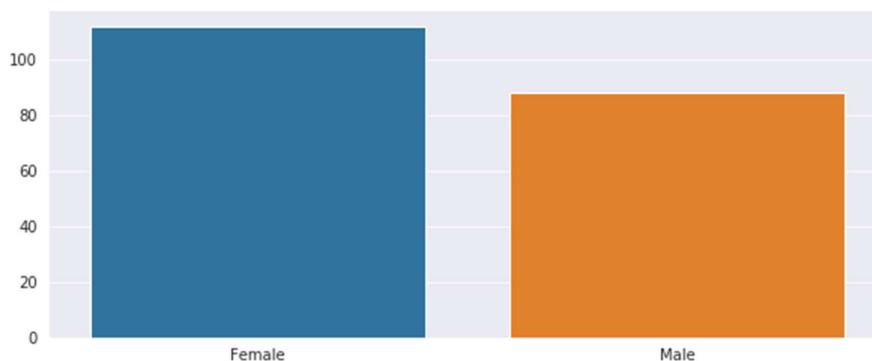
Sebaran frekuensi data jenis kelamin pelanggan suatu *mall* sangat penting untuk diketahui. Hal ini dapat digunakan oleh pihak *mall* dalam memberikan diskon, produk rekomendasi pria atau wanita, dan beragam program menarik lainnya.

```

1 genders = df.Gender.value_counts()
2 sns.set_style("darkgrid")
3 plt.figure(figsize=(10,4))
4 sns.barplot(x=genders.index, y=genders.values)
5 plt.show()

```

Hasil visualisasinya ditampilkan pada Gambar 4.



Gambar 4. Frekuensi Pengunjung Mall Berdasarkan Gender

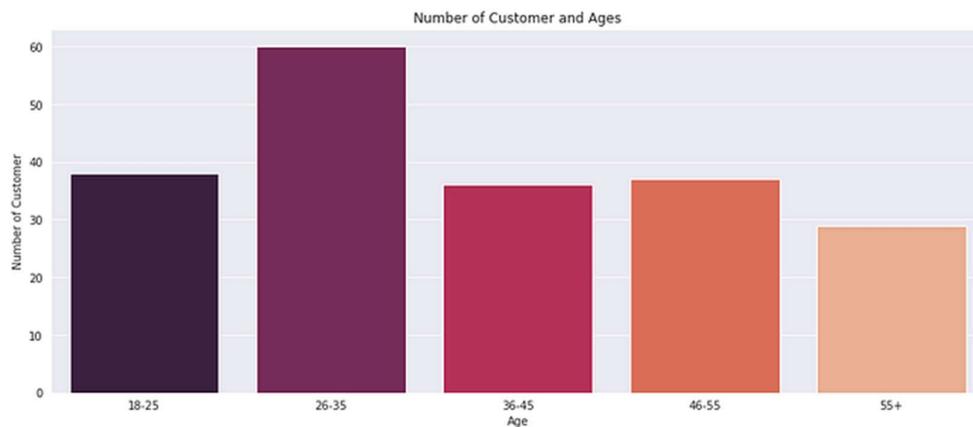
Selain itu pembagian rentang usia pelanggan *mall* secara spesifik juga perlu divisualisasikan agar ke depannya program-program *marketing* dari pihak *mall* dapat lebih tepat sasaran.

```

1 age18_25 = df.Age[(df.Age <= 25) & (df.Age >= 18)]
2 age26_35 = df.Age[(df.Age <= 35) & (df.Age >= 26)]
3 age36_45 = df.Age[(df.Age <= 45) & (df.Age >= 36)]
4 age46_55 = df.Age[(df.Age <= 55) & (df.Age >= 46)]
5 age55above = df.Age[df.Age >= 56]
6
7 x = ["18-25", "26-35", "36-45", "46-55", "55+"]
8 y = [len(age18_25.values), len(age26_35.values), len(age36_45.values), len(age46_55.values), len(age55above.values)]
9
10 plt.figure(figsize=(15,6))
11 sns.barplot(x=x, y=y, palette="rocket")
12 plt.title("Number of Customer and Ages")
13 plt.xlabel("Age")
14 plt.ylabel("Number of Customer")
15 plt.show()

```

Adapun tampilan visualnya ditunjukkan pada Gambar 5.



Gambar 5. Kategorisasi Rentang Usia Pelanggan

#### 4. Frekuensi Jumlah Skor Pengeluaran

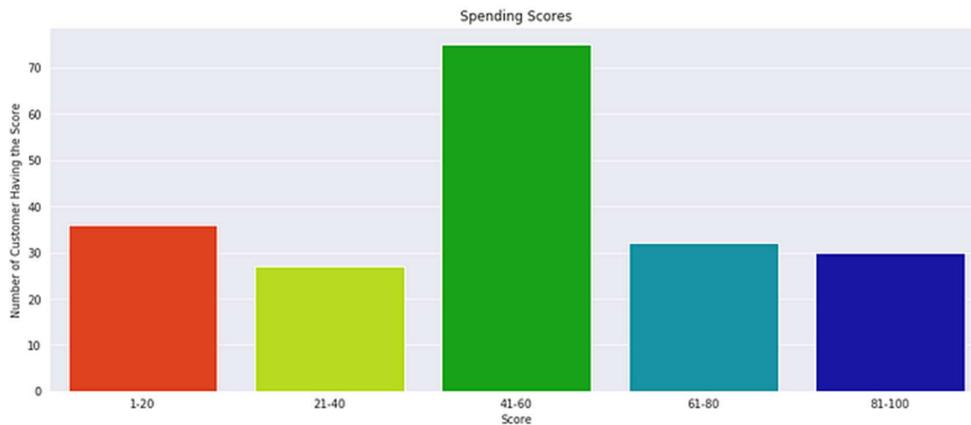
Visualisasi ini digunakan untuk melihat seberapa banyak jumlah pelanggan dengan skor pengeluaran belanjanya.

```

1  ss1_20 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 1) & (df["Spending S
2  ss21_40 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 21) & (df["Spending
3  ss41_60 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 41) & (df["Spending
4  ss61_80 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 61) & (df["Spending
5  ss81_100 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 81) & (df["Spending
6
7  ssx = ["1-20", "21-40", "41-60", "61-80", "81-100"]
8  ssy = [len(ss1_20.values), len(ss21_40.values), len(ss41_60.values), len(ss61_80.values), len(ss81_100.values)]
9
10 plt.figure(figsize=(15,6))
11 sns.barplot(x=ssx, y=ssy, palette="nipy_spectral_r")
12 plt.title("Spending Scores")
13 plt.xlabel("Score")
14 plt.ylabel("Number of Customer Having the Score")
15 plt.show()

```

Tampilan visualisasinya ditampilkan pada Gambar 6.



Gambar 6. Sebaran Jumlah Skor Pengeluaran

Dari hasil visualisasi pada Gambar 6 dapat dilihat bahwa mayoritas Skor Pengeluaran berada pada rentang 41-60.

#### 5. Frekuensi Jumlah Pendapatan Tahunan

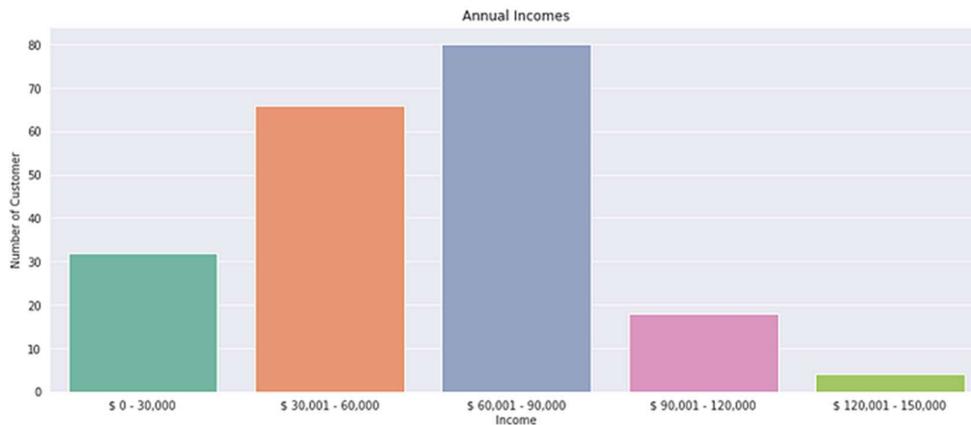
Jika pihak manajemen *mall* dapat mengetahui besaran pendapatan tahunan dari pengunjunnya maka hal ini dapat digunakan untuk menyetok barang-barang yang dapat disesuaikan dengan tingkat pendapatan pengunjung.

```

1  ai0_30 = df["Annual Income (k$)"][df["Annual Income (k$)"] >= 0] & (df["Annual Income (k$)"] >= 0) & (df["Annual Income (k$)"] <= 30)
2  ai31_60 = df["Annual Income (k$)"][df["Annual Income (k$)"] >= 31] & (df["Annual Income (k$)"] >= 31) & (df["Annual Income (k$)"] <= 60)
3  ai61_90 = df["Annual Income (k$)"][df["Annual Income (k$)"] >= 61] & (df["Annual Income (k$)"] >= 61) & (df["Annual Income (k$)"] <= 90)
4  ai91_120 = df["Annual Income (k$)"][df["Annual Income (k$)"] >= 91] & (df["Annual Income (k$)"] >= 91) & (df["Annual Income (k$)"] <= 120)
5  ai121_150 = df["Annual Income (k$)"][df["Annual Income (k$)"] >= 121] & (df["Annual Income (k$)"] >= 121) & (df["Annual Income (k$)"] <= 150)
6
7  aix = ["$ 0 - 30,000", "$ 30,001 - 60,000", "$ 60,001 - 90,000", "$ 90,001 - 120,000", "$ 120,001 - 150,000"]
8  aiy = [len(ai0_30.values), len(ai31_60.values), len(ai61_90.values), len(ai91_120.values), len(ai121_150.values)]
9
10 plt.figure(figsize=(15,6))
11 sns.barplot(x=aix, y=aiy, palette="Set2")
12 plt.title("Annual Incomes")
13 plt.xlabel("Income")
14 plt.ylabel("Number of Customer")
15 plt.show()

```

Tampilan visualisasinya ditampilkan pada Gambar 7.



Gambar 7. Sebaran Skor Pengeluaran Pelanggan *Mall*

### 1.3.3 Segmentasi Pelanggan Menggunakan K-Means

Sebelum segmentasi dilakukan, langkah pertama adalah menentukan jumlah *cluster*-nya atau diistilahkan WCSS (Within Clusters Sum of Squares). WCSS didefinisikan sebagai jumlah kuadrat jarak antara masing-masing anggota cluster dan pusat massanya.

$$WCSS = \sum_{i \in N} (X_i - Y_i)^2$$

```

1  wcss = []
2
3  for i in range(1,11):
4      kmeans= KMeans(n_clusters = i, init = 'k-means++')
5      kmeans.fit(X) #Finds the Kmeans to our data
6      wcss.append(kmeans.inertia_) #This gives us the wcss values for each clusters

```

Metode

yang digunakan untuk menentukan nilai k optimal adalah Elbow.



Nilai di atas mewakili nomor cluster untuk setiap titik data. Selanjutnya adalah mencari titik pusat *cluster* atau *centroid* dari masing-masing *cluster* tersebut.

```
1 print(kmeans.cluster_centers_)

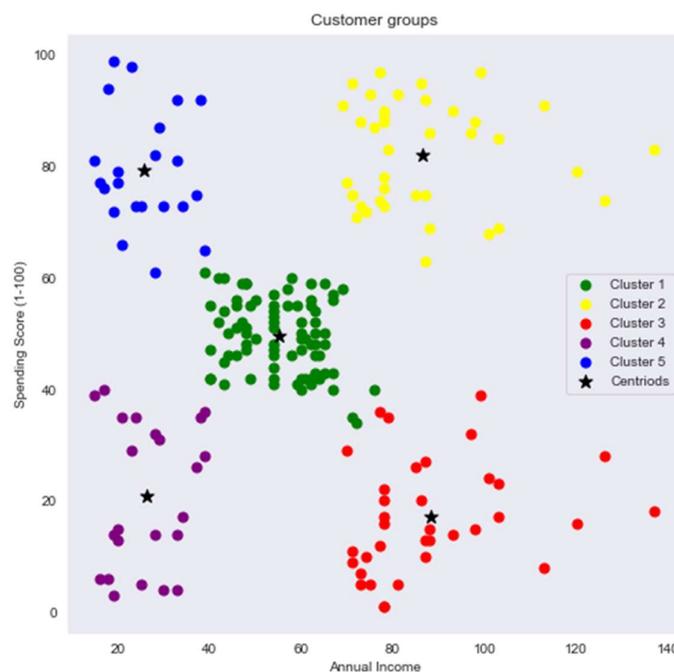
[[55.2962963  49.51851852]
 [86.53846154  82.12820513]
 [88.2         17.11428571]
 [26.30434783  20.91304348]
 [25.72727273  79.36363636]]
```

## 2. Visualisasi Cluster

Agar dapat dilakukan penarikan *insight* maka dibutuhkan visualisasi *cluster*.

```
1 plt.figure(figsize=(8,8))
2 plt.scatter(X[label == 0,0], X[label== 0,1], s=50, c='green', label='Cluster 1')
3 plt.scatter(X[label == 1,0], X[label== 1,1], s=50, c='yellow', label='Cluster 2')
4 plt.scatter(X[label == 2,0], X[label== 2,1], s=50, c='red', label='Cluster 3')
5 plt.scatter(X[label == 3,0], X[label== 3,1], s=50, c='purple', label='Cluster 4')
6 plt.scatter(X[label == 4,0], X[label== 4,1], s=50, c='blue', label='Cluster 5')
7 plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1], s= 100, c='black')
8 plt.title('Customer groups')
9 plt.xlabel('Annual Income')
10 plt.ylabel('Spending Score (1-100)')
11 plt.legend()
12 plt.show()
```

Hasil visualisasi *cluster* dapat dilihat pada Gambar 9.



Gambar 9. Visualisasi Cluster Pelanggan

### 3. Penarikan *Insight* Bisnis

Hasil analisis menunjukkan bahwa pelanggan *mall* dapat dikelompokkan menjadi 5 *cluster* atau segmen untuk pemasaran yang ditargetkan.

*Cluster 1* (hijau) adalah pelanggan yang berpenghasilan rata-rata dengan skor pengeluaran rata-rata. Kelompok ini berhati-hati dengan pengeluaran mereka ketika berbelanja.

*Cluster 2* (kuning) adalah pelanggan yang berpenghasilan tinggi dan memiliki skor pembelanjaan tinggi. Kelompok ini sangat potensial mendatangkan keuntungan. Diskon dan penawaran lain yang ditargetkan pada kelompok ini akan meningkatkan skor pembelanjannya dan memaksimalkan keuntungan.

*Cluster 3* (merah) adalah pelanggan yang memiliki pendapatan lebih tinggi namun tidak menghabiskan lebih banyak uang di *mall*. Salah satu asumsinya adalah karena tidak puas dengan pelayanan yang diberikan oleh pihak *mall*. Kelompok ini adalah kelompok ideal untuk dibidik oleh tim pemasaran karena sangat berpotensi mendatangkan peningkatan keuntungan bagi *mall*.

*Cluster 4* (ungu) adalah pelanggan yang berpenghasilan rendah dengan skor pengeluaran rendah. Fenomena ini terjadi karena orang-orang dengan pendapatan rendah akan cenderung membeli lebih sedikit barang di *mall*.

*Cluster 5* (biru) adalah pelanggan berpenghasilan rendah dengan skor pembelanjaan tinggi. Kelompok ini rela menghabiskan pendapatannya untuk berbelanja di *mall* adalah karena menikmati dan puas dengan layanan yang diberikan selama berbelanja.

### **KESIMPULAN**

Terdapat 5 kelompok pelanggan mall berdasarkan hasil analisis data. Lima kelompok pelanggan tersebut adalah:

1. Kelompok 1 adalah pelanggan yang berpenghasilan rata-rata dengan skor pengeluaran rata-rata. Kelompok ini berhati-hati dengan pengeluaran mereka ketika berbelanja.
2. Kelompok 2 adalah pelanggan yang berpenghasilan tinggi dan memiliki skor pembelian tinggi. Kelompok ini sangat potensial mendatangkan keuntungan. Diskon dan penawaran lain yang ditargetkan pada kelompok ini akan meningkatkan skor pembelajarannya dan memaksimalkan keuntungan.
3. Kelompok 3 adalah pelanggan yang memiliki pendapatan lebih tinggi namun tidak menghabiskan lebih banyak uang di *mall*. Salah satu asumsinya adalah karena tidak puas dengan pelayanan yang diberikan oleh pihak *mall*. Kelompok ini adalah kelompok ideal untuk dibidik oleh tim pemasaran karena sangat berpotensi mendatangkan peningkatan keuntungan bagi *mall*.
4. Kelompok 4 adalah pelanggan yang berpenghasilan rendah dengan skor pengeluaran rendah. Fenomena ini terjadi karena orang-orang dengan pendapatan rendah akan cenderung membeli lebih sedikit barang di *mall*.
5. Kelompok 5 adalah pelanggan berpenghasilan rendah dengan skor pembelian tinggi. Kelompok ini rela menghabiskan pendapatannya untuk berbelanja di *mall* adalah karena menikmati dan puas dengan layanan yang diberikan selama berbelanja.

#### DAFTAR PUSTAKA

- [1] S. Hanoatubun, "Dampak Covid - 19 terhadap Perekonomian Indonesia," *J. Educ. Psychology Couns.*, vol. 2, no. 2, 2020.
- [2] F. R. Yamali and R. N. Putri, "Dampak Covid-19 Terhadap Ekonomi Indonesia," *Ekon. J. Econ. Bus.*, vol. 4, no. 2, p. 384, 2020.
- [3] D. Junaedi and F. Salistia, "Dampak Pandemi COVID-19 Terhadap Pertumbuhan Ekonomi Negara-Negara Terdampak," *Simposium Nasional Keuangan Negara*. pp. 995-1115, 2020.
- [4] F. Nidaul Khasanah *et al.*, "Pemanfaatan Media Sosial dan Ecommerce Sebagai Media Pemasaran Dalam Mendukung Peluang Usaha Mandiri Pada Masa Pandemi Covid 19," *J. Sains Teknol. dalam Pemberdaya. Masy.*, vol. 1, no. 1, pp. 51-62, 2020.

- [5] E. Maria, S. Suharyadi, and R. K. Hudiono, "Implementasi pemasaran digital berbasis website sebagai strategi kenormalan baru Dusun Srumbung Gunung pasca Covid-19," *Riau J. Empower.*, vol. 4, no. 1, pp. 1–10, 2021.
- [6] N. A. Rakhmawati, A. E. Permana, A. M. Reyhan, and H. Rafli, "Analisa Transaksi Belanja Online Pada Masa Pandemi Covid-19," *J. Teknoinfo*, vol. 15, no. 1, p. 32, 2021.
- [7] I. Atmaja, Hanung Eka; Novitaningtyas, "Analisis Aspek Pemasaran UMKM di Masa Resesi Global Dampak dari Pandemi Covid-19," *J. Ilm. Poli Bisnis*, vol. 13, no. 1, pp. 60–69, 2021.
- [8] N. I. Purnama and L. P. Putri, "Analisis Penggunaan E-Commerce Di Masa Pandemi," *Semin. Nas. Teknol. Edukasi Sos. dan Hum.*, vol. 1, no. 1, pp. 556–561, 2021.
- [9] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020.
- [10] C. D. Manning, 2008, *Introduction to Information Retrieval*, Cambridge University Press, Cambridge.
- [11] W. Xiaohong, W. Bin, S. Jun, Q. Shengwei och a. L. Xiang, 2014, A hybrid fuzzy K-harmonic means clustering algorithm, *Applied Mathematical Modelling* 39 (12), pp 3398-3409.
- [12] D. T. Larose, 2014, *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley-Interscience, Canada.
- [13] E. Prasetyo, 2014, *Data Mining, Mengolah Data Menjadi Informasi Menggunakan Matlab, ANDI*, Yogyakarta.
- [14] J. Oyelade, O. Oladipupo, 2010, Application of k Means Clustering algorithm for prediction of Students Academic Performance, *International Journal of Computer Science and Information Security* 7, pp 292–295.
- [15] T. R. Delgado, 2020, *Hands-on Data Analytics for Beginners with Google Colaboratory*, Packt Publishing, Birmingham.
- [16] M. Grinberg, 2018, *Flask Web Development: Developing Web Applications with Python*, O'Reilly Media, California.