|  |  |
| --- | --- |
|  | **Gameology And Multimedia Expert**  Vol., No., Month, Year, pp. \*\*-\*\*  Journal homepage: http://ojs.unimal.ac.id/game |

|  |  |  |  |
| --- | --- | --- | --- |
| **Penerapan Metode K-Means Clustering Dalam Segmentasi Pelanggan Untuk Strategi Pemasaran Bisnis Yang Efektif** | | |  |
| Cantika Serenita  Universitas Malikussaleh | | |  |
| \*Corresponding Author Email: cantikadaulay813@gmail.com | | | |
|  |  | **ABSTRAK** | |
|  |  |  | |
| **Received:**  **Revised:**  **Accepted:**  **Available online:** |  | Penelitian ini mengeksplorasi penggunaan metode K-Means Clustering untuk segmentasi pelanggan guna meningkatkan efektivitas strategi pemasaran bisnis. Segmentasi pelanggan, yang membagi pelanggan berdasarkan kesamaan geografis, demografis, psikografis, dan perilaku, memungkinkan perusahaan memahami kebutuhan setiap kelompok dan mengidentifikasi pelanggan potensial. Metode K-Means Clustering digunakan untuk mengelompokkan data ke dalam beberapa cluster, dengan teknik Metode Elbow dan Silhouette Score membantu menentukan jumlah cluster optimal. Data dari platform Kaggle, mencakup id pelanggan, usia, jenis kelamin, pendapatan, dan pengeluaran, diolah menggunakan Google Colaboratory dengan pustaka Python seperti Pandas, Numpy, Seaborn, dan Matplotlib. Hasil segmentasi membagi pelanggan menjadi lima cluster berdasarkan variabel usia, pendapatan tahunan, dan pengeluaran, memungkinkan penyesuaian strategi pemasaran. Pelanggan dengan pendapatan dan pengeluaran tinggi ditargetkan dengan produk premium dan layanan VIP, sementara pelanggan dengan pendapatan dan pengeluaran rendah diberikan promosi dan diskon. Penelitian ini menunjukkan bahwa penerapan K-Means Clustering dapat membantu merumuskan strategi pemasaran yang lebih efektif dan efisien. | |
| ***Kata Kunci:*** *Segmentasi pelanggan, K-Means, Elbow,Silhouette Score, Davies-Bouldien* |  |
|  |  | **ABSTRACT** | |
| ***Keywords:*** *Customer Segmentation, K-Means, Elbow, Silhouette Score, Davies-Bouldien* |  | *This research explores the use of the K-Means Clustering method for customer segmentation to increase the effectiveness of business marketing strategies. Customer segmentation, which divides customers based on geographic, demographic, psychographic, and behavioral similarities, allows companies to understand the needs of each group and identify potential customers. The K-Means Clustering method is used to group data into several clusters, with the Elbow and Silhouette Score methods helping to determine the optimal number of clusters. Data from the Kaggle platform, including customer ID, age, gender, income and expenses, is processed using Google Colaboratory with Python libraries such as Pandas, Numpy, Seaborn and Matplotlib. The segmentation results divide customers into five clusters based on the variables of age, annual income and expenses, allowing for adjustments to marketing strategies. Customers with high income and expenditure are targeted with premium products and VIP services, while customers with low income and expenditure are provided with promotions and discounts. This research shows that the application of K-Means Clustering can help formulate more effective and efficient marketing strategies.* | |

# INTRODUCTION

Bisnis dan teknologi adalah dua hal yang saling berkaitan. Semakin berkembangnya zaman, teknologi menjadikan pekerjaan manusia menjadi lebih mudah. Oleh karena itu , untuk berbisnis saat ini tentunya harus bisa memanfaatkan teknologi dengan sebaik mungkin. Apalagi, persaingan bisnis yang semakin ketat, mengharuskan suatu perusahaan lebih inovatif agar dapat menarik perhatian dari *customer*.

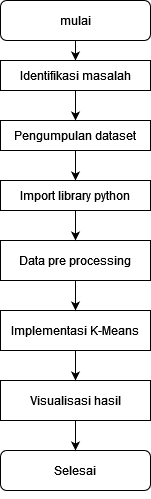
Dalam dunia bisnis tentunya tidak akan lepas dari yang namanya pemasaran, aspek penting ini melibatkan teknologi untuk mencapai hasil yang optimal. Penggunaan teknologi tersebut dapat memberikan keunggulan kompetitif dan signifikan. Tidak hanya membantu dalam mencapai audiens yang lebih luas, teknologi juga memungkinkan perusahaan dapat menganalisa kebutuhan pelanggan. Dibuktikan pada sebuah industri retail yang sedang mengalami permasalahan pada strategi pemasarannya. Dan ini menyebabkan masih ada beberapa produk yang tidak habis terjual karena kurangnya minat pelanggan untuk membeli barang tersebut.

Dimana, produk yang kurang laku bisa memberikan dampak yang buruk apabila perusahaan tersebut tidak mampu dalam mengatur strategi pemasaran dengan baik. Solusi yang diperlukan untuk kasus ini adalah dengan membuat segmentasi pelanggan. Segmentasi pelanggan merupakan proses mengenali profiling pelanggan dengan cara membagi pelanggan menjadi empat bagian yaitu berdasarkan segmentasi geografis, demografis, psikografis, dan perilaku. Tujuan dari segmentasi pelanggan ialah untuk memudahkan bisnis mengetahui kebutuhan dari setiap pelanggan, sehingga dapat menemukan pelanggan potensial.

Proses ini dilakukan dengan menggunakan teknik *clustering* pada algoritma *K-Means* yang merupakan bagian dari *unsupervised learning*. K-Means clustering merupakan salah satu algoritma *non hierarchical clustering*, yang berfungsi untuk membagi dan mengenali pola data dengan menemukan kemiripan dalam satu kelompok dan perbedaan dengan kelompok lainnya. Data akan dibagi berdasarkan jumlah *cluster* terbaik yang didapatkan dengan cara menggunakan *Metode Elbow.* Jumlah dataset yang digunakan sebanyak 200, terdiri dari id pelanggan, usia, jenis kelamin, pendapatan, dan pengeluaran dari pelanggan.

Pemanfaatan segmentasi pelanggan ini diharapkan dapat membantu Perusahaan untuk mengubah strategi pemasaran menjadi lebih baik. Sehingga bisa mengalokasikan sumber daya dan tenaga kepada pelanggan potensial.

**2. METHOD AND MODELS**

****

****

Gambar 1. Tahapan segmentasi

Metode *k-means clustering* yang digunakan merupakan metode yang akurat untuk melakukan pengelompokan data. *Centroid* ditetapkan diawal proses *clustering*, kemudian pengalokasian data ke *centroid* terdekat, dan pembaruan *centroid* didasarkan dari data rata-rata(*mean*) yang dialokasikan.

1. Identifikasi masalah

Identifikasi masalah pada penelitian ini yaitu kurangnya ketepatan pada strategi pemasaran di sebuah mall yang mengakibatkan banyaknya produk yang tidak laku maupun habis terjual. Perusahaan mengalami hal ini, sehingga diperlukan metode pemasaran yang efektif dengan memanfaatkan teknologi yang canggih pada saat sekarang ini.

2. Perumusan masalah

Bagaimana penggunaan metode *k-means clustering* untuk menentukan jumlah *cluster* yang optimal agar dapat memberikan hasil yang akurat pada segmentasi pelanggan.

3. Studi literatur

Studi literatur yang digunakan untuk penelitian ini menggunakan referensi beberapa jurnal mengenai *k-means clustering* yang ada di *Google Schoolar* dengan jurnal dari tahun 2020 sampai dengan tahun 2024. Adapun untuk studi literatur yang berkaitan dengan penelitian ini adalah sebagai berikut :

*A.* *Segmentasi pelanggan*

Segmentasi pelanggan adalah proses membagi pelanggan menjadi kelompok yang lebih kecil berdasarkan kemiripan secara demografis, psikografis, geografis, serta perilaku yang serupa. Dengan adanya segmentasi pelanggan, perusahaan dapat dengan mudah untuk mengenali kebutuhan dari konsumen terhadap produk yang ditawarkan.(Maskanah, 2020)

*B.* *Algoritma K-Means Clustering*

Sebagai salah satu metode dari *unsupervised learning*, *k-means clustering* menjadi algoritma yang sering digunakan untuk mengelompokkan data kedalam sejumlah klaster berdasarkan kemiripan antar data. Algoritma ini efektif dalam menemukan struktur tersembunyi dalam sata dan efektif dalam memberikan hasil klasterisasi yang akurat.

*C.* *Metode Elbow*

*Metode elbow* merupakan teknik yang digunakan untuk mendapatkan jumlah kluster yang optimal pada. Tujuan dari metode ini adalah untuk memudahkan dalam menetukan klaster yang akan digunakan untuk segmentasi.

*D.* *Silhouette score*

*Silhouette score* merupakan teknik untuk menemukan jumlah klaster optimal berdasarkan kecocokan kelompok yang ditentukan dan seberapa berbeda dengan lainnya.

*E.* *Davies-Bouldin Index*

*metode ini* merupakan metode yang digunakan untuk evaluasi cluster yang diperkenalkan oleh David L. Davies dan Donal W. Bouldin.

4. Pengumpulan data

Data pelanggan didapatkan dari kaggle yang merupakan sebuah situs online yang memuat berbagai ilmu mengenai data science dan memiliki lebih dari 6000 *dataset*. Untuk data yang menjadi studi kasus penelitian ini berjumlah 200, yang berisi data id pelanggan, usia, jenis kelamin, pemasukan, dan pengeluaran.

**3. RESULT AND DISCUSSION**

**3.1 Hasil**

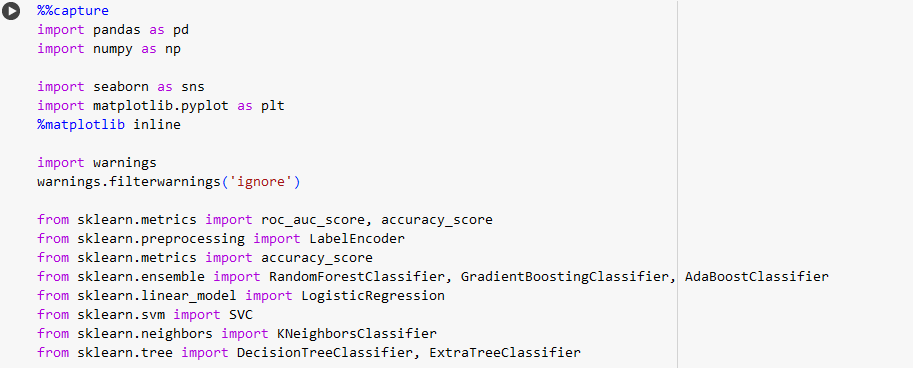
Pada tahapan ini dimulai dengan pengumpulan data untuk segmentasi pelanggan, data yang digunakan diambil dari platform Kaggle yang menyediakan berbagai dataset untuk *data science* dan *machine learning*. Jumlah data sebanyak 200, yang terdiri dari google colaboratory adalah karena produk dari google satu ini memudahkan para programmer untuk berbagi maupun berkolaborasi secara

online. Juga, banyak kumpulan *library machine learning* yang dimuat di dalamnya sehingga pemula tak perlu lagi untuk mengunduh *extention*.

**3.2 Pembahasan**

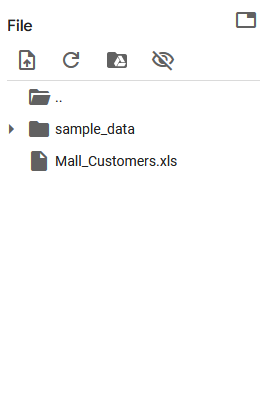
Pada tahapan ini, langkah-langkah dalam pengolahan data akan dijabarkan lebih detail.

A. *Import library python*



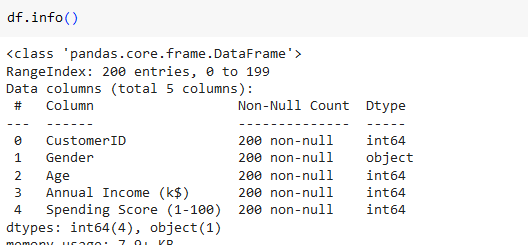
Gambar 2. Pustaka *Python*

Pada gambar ini menunjukkan tentang *library* apa saja yang dibutuhkan untuk memproses *dataset*. Terdiri dari pustaka *pandas* yang digunakan untuk manipulasi dan menganalisi data, Pustaka *numpy* yang memiliki berbagai fungsi matematika. Terdapat juga pustaka *seaborn* untuk membuat visualisasi hasil matrik lebih menarik, dan pustaka matplotlib.



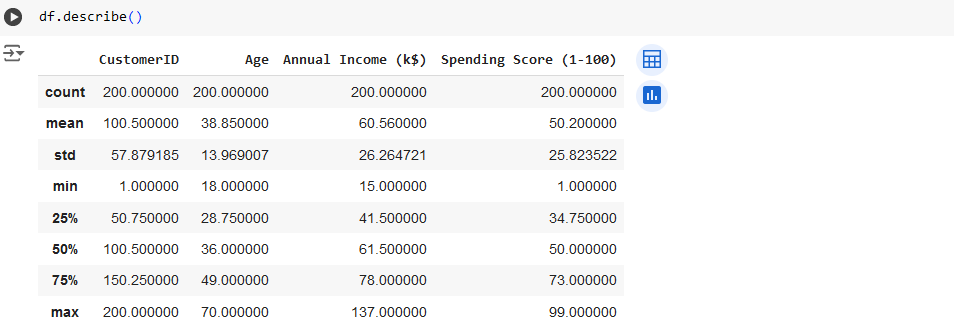
Gambar 3. Dataset

Gambar 3 menunjukkan Lokasi dari dataset pelanggan yang berada didalam folder sampel data, yang mana dataset ini berupa format file excel.



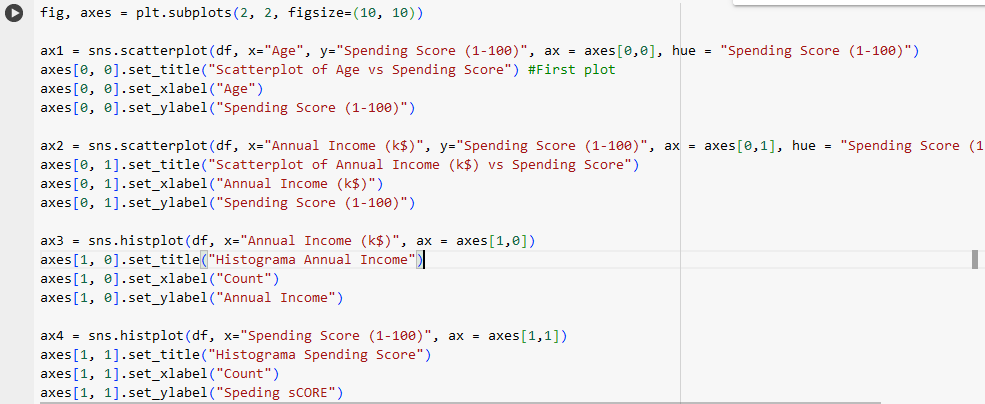
Gambar 4. Info *dataset*

Tahapan ini akan dilakukan eksplorasi data yang bertujuan untuk menganalisis data, mengenali struktur data, mendeteksi pola dan anomali, sebelum kemudian data divisualisasikan kedalam bentuk plot dengan menggunakan *seaborn* dan *matplotlib*.



Gambar 5. Deskripsi data

Deskripsi data bertujuan untuk menjabarkan apa saja yang ada di dalam dataset tersebut, fungsi ini akan dijalankan oleh *library pandas*.kemudian, akan menampilkan statistik dari jumlah data (count), rata-rata(mean), standar deviasi, nilai minimum dan maksimum. pada kolom usia menunjukkan bahwa rata-rata usia pelanggan adalah 38,85 tahun. Dengan standar deviasi 13.96, dan rentang usia mulai dari 18 tahun sampai 70 tahun. Dimana sebanyak dua puluh lima persen pelanggan berada diusia 28.75 tahun, begitu juga untuk bagian pemasukan dan pengeluaran pelanggan.



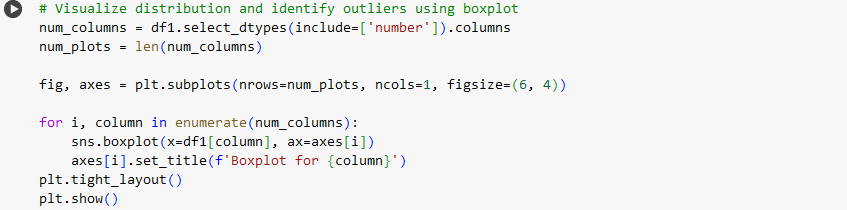
Gambar 6. Hubungan distribusi variabel dataframe

Tujuan dari program ini adalah untuk menampilkan visualisasi dari hubungan antar variabel-variabel yang berada di dalam dataframe ‘df’. Terdapat dua grafik yang digunakan yaitu scatterplot untuk melihat pola hubungan antara usia, pendapatan tahunan, dan pengeluaran. Selanjutnya histogram untuk melihat distribusi frekuensi dari pendapatan tahunan dan skor pengeluaran secara terpisah.



Gambar 7. Jumlah pelanggan

Kode program ini akan menampilkan plot yang berisi jumlah pelanggan laki-laki sebanyak 88 pelanggan dan Perempuan 112 pelanggan.



Gambar 8. Menampilkan *boxplot*

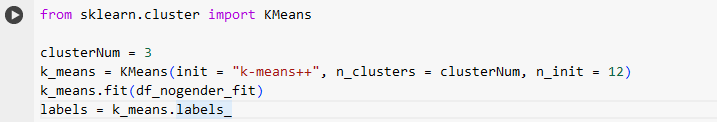
*Boxplot* meruapakan alat visual yang bagus untuk melihat distribusi data secara grafis. Pada *boxplo*t terdapat tampilan median (nilai tengah), kuartil, serta batas atas dan batas bawah data. Boxplot juga dapat mengidentifikasi adanya *outlier* pada data, *outlier* merupakan data yang nilainya jauh dari kumpulan data seharusnya.

C. Deskriptif Analisis Data

Pada tahapan ini, data yang telah dianalisis akan ditampilkan dalam bentuk plot, untuk menampilkan gambaran yang lebih mendalam terhadap data pelanggan yang terdiri dari usia, pendapatan, dan pengeluaran dari pelanggan. Informasi yang dihasilkan dapat menjadi acuan untuk membantu dalam memahami karakteristik dari data.

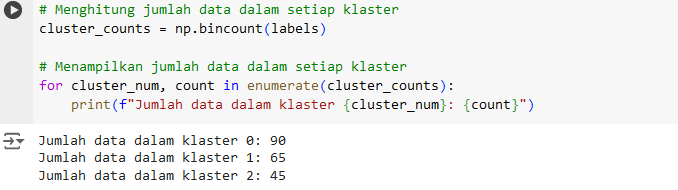
D. Pengaplikasian *K-Means Clustering*

Tahap awal dari *K-Means* adalah menentukan k (cluster) secara random terlebih dahulu. Jumlah klaster yang digunakan adalah sebanyak k=3.



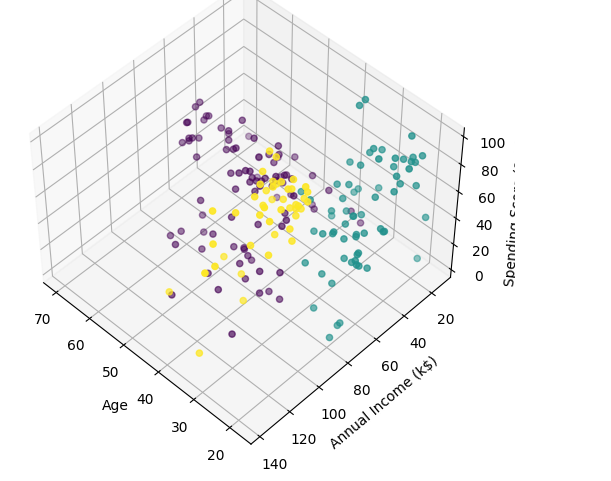
Gambar 9. Proses *K-Means*

Modul *K-Means* diambil dari pustaka *scikit.cluster* hal ini bertujuan untuk melakukan klasterisasi menggunakan algoritma K-Means.



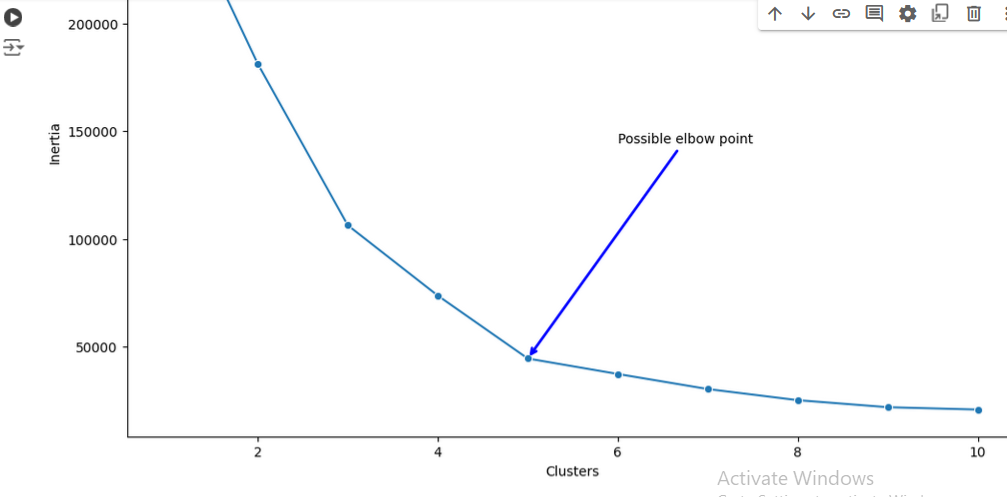
Gambar 10. Jumlah data pada tiga klaster

Pada gambar diatas menunjukkan jumlah data yang tersebar pada setiap klasternya dengan klaster awal yang ditentukan sebanyak 3.



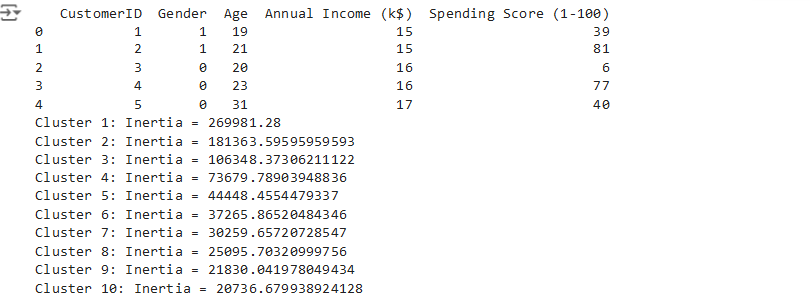
Gambar 11. Plot persebaran 3 klaster

Selanjutnya untuk penentuan jumlah klaster yang optimal, maka akan digunakan metode elbow(siku). Metode *elbow* merupakan pendekatan yang sering digunakan pada algoritma k-means clustering untuk menentukan jumlah klaster terbaik. Berdasarkan plot *elbow* nilai inertia atau WCSS (Whitin Cluster Sum Of Square) akan berkurang seiring dengan bertambahnya jumlah klaster.



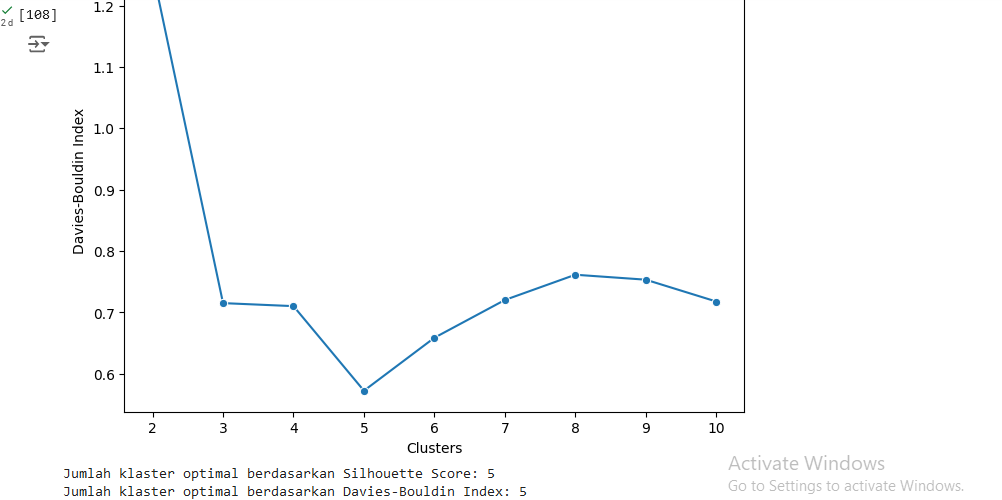
Gambar 12. Grafik *Elbow*

Tanda panah yang tertera diatas menunjukkan jumlah klaster yang optimal berdasarkan metode *elbow*.



Gambar 13. Nilai Inersia

Nilai inersia menunjukkan semakin rendah dengan adanya penambahan klaster, dan klaster lima menunjukkan pengurangan yang signifikan di dalam inersia dibandingkan dengan klaster lainnya.

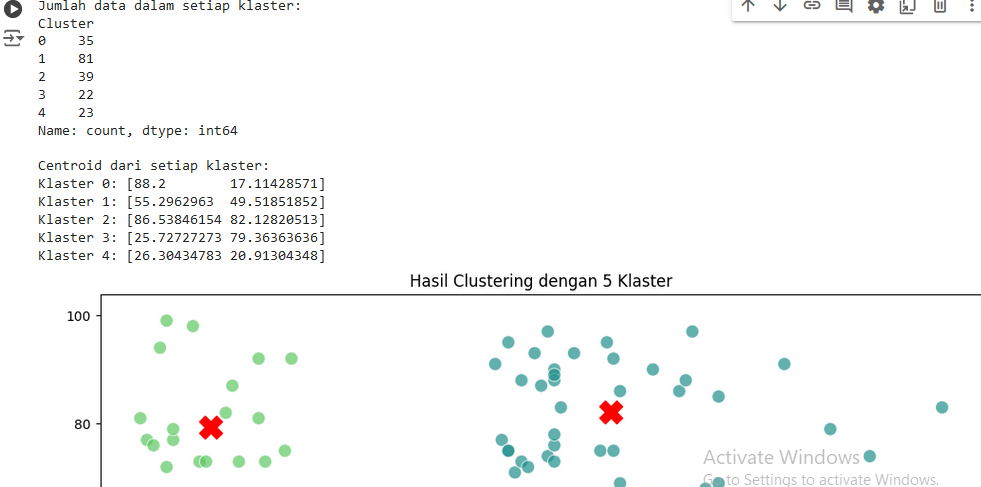


Gambar 14. hasil davies dan silhouette score

*Silhouette Score* dan *Davies-Bouldin Index* merupakan dua metrik yang berfungsi untuk mengevaluasi seberapa bagus kualitas klaster yang akan digunakan pada *K-Means Clustering*.

E. Hasil *Clustering* Segmentasi Pelanggan

Setelah mendapatkan jumlah klaster yang optimal, maka data akan diuji kembali dengan k=5. Sehingga *K-Means* akan membentuk lima klaster yang berbeda. Hasil klaster akan ditampilkan dalam bentuk *plotting* data, yang mana untuk *cluster* satu berjumlah 35 pelanggan, cluster dua 81, klaster tiga sebanyak 39, klaster empat 22, dan klaster lima 23 pelanggan.



Gambar 15. Hasil k=5

Pengelompokan data pelanggan berdasarkan variabel usia (age), pendapatan tahunan (annual income), dan pengeluaran (spending score), dapat disimpulkan sebagai berikut:

1. Cluster 0 : menunjukkan bahwa data dalam klaster ini mempunyai pendapatan tahunan yang tinggi dan score pengeluaran yang rendah. Untuk strategi pemasaran yang dilakukan pada pelanggan klaster ini adalah dengan memberikan produk premium dengan diskon dan juga tawarkan barang khusus yang dapat menarik perhatian pembeli.

2. Cluster 1 : memiliki pendapatan tahunan sedang dan pengeluaran sedang. Untuk pelanggan pada klaster ini gunakan promosi musiman dan tawarkan paket kombo untuk memperoleh nilai lebih.

3. Cluster 2 : untuk pendapatan tahunan tinggi dan skor pengeluaran juga tinggi. Pada pelanggan klaster ini berfokus pada penawaran produk yang berkualitas tinggi dan VIP.

4. Cluster 3 : dengan pendapatan tahunan rendah dan skor pengeluaran tinggi. Rekomendasi yang dapat dilakukan untuk klaster ini adalah dengan menawarkan skema pembayaran yang fleksibel ataupun memberikan opsi cicilan untuk mempermudah pembelian.

5. Cluster 4 : klaster ini memiliki pendapatan tahunan rendah dan skor pengeluaran juga rendah. Tawarkan produk yang terjangkau dan ramah anggaran, kemudian berikan promo diskon untuk menarik pembelian.

Jika halaman sudah tidak muat untuk menulis heading beserta penjelasannya, dianjurkan untuk memindahkan poin heading tersebut ke halaman selanjutnya.

# 4. CONCLUSION

Jurnal ini membahas penggunaan metode *K-Means Clustering* untuk melakukan segmentasi pelanggan dengan tujuan meningkatkan efektivitas strategi pemasaran dalam bisnis. *K-Means Clustering* adalah algoritma yang digunakan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kesamaan. Algoritma ini membantu mengenali pola data dan menentukan jumlah cluster optimal dengan menggunakan teknik seperti Metode *Elbow* dan *Silhouette Score*, yang penting untuk segmentasi pelanggan yang akurat. Segmentasi pelanggan membantu bisnis membagi pelanggan menjadi kelompok berdasarkan kesamaan geografis, demografis, psikografis, dan perilaku. Ini memungkinkan perusahaan untuk memahami kebutuhan setiap kelompok dan mengidentifikasi pelanggan potensial, yang sangat penting untuk merancang strategi pemasaran yang lebih efektif. Data pelanggan yang digunakan dalam penelitian ini diambil dari platform *Kaggle* dan mencakup informasi seperti id pelanggan, usia, jenis kelamin, pendapatan, dan pengeluaran. Pengolahan data dilakukan menggunakan Google Colaboratory dengan pustaka *Python* seperti *Pandas, Numpy, Seaborn*, dan *Matplotlib*, yang memudahkan analisis data secara mendetail. Dengan menggunakan *K-Means Clustering*, data pelanggan dibagi menjadi lima cluster berdasarkan variabel usia, pendapatan tahunan, dan pengeluaran. Setiap *cluster* memiliki karakteristik yang berbeda, sehingga strategi pemasaran dapat disesuaikan. Misalnya, pelanggan dengan pendapatan tinggi dan pengeluaran tinggi ditargetkan dengan produk premium dan layanan VIP, sementara pelanggan dengan pendapatan dan pengeluaran rendah diberikan promosi dan diskon yang menarik.

**REFERENCES**

Maskanah, I. (2020). Segmentasi Pelanggan Toko Purnama dengan Algoritma K-Means dan Model RFM untuk Perancangan Strategi Pemasaran. *INOVTEK Polbeng - Seri Informatika*, *5*(2), 218. https://doi.org/10.35314/isi.v5i2.1443