IMPLEMENTASI POWERTRANSFORMER DAN PRINCIPAL COMPONENT ANALYSIS (PCA) UNTUK OPTIMALISASI CLUSTERING PADA SEGMENTASI PELANGGAN USIA DEWASA MENGGUNAKAN ALGORITMA K-MEANS

Cindy Avitaselly Bambang Saputri, Arie Nugroho, Anita Sari Wardani
Sistem Informasi, Universitas Nusantara PGRI Kediri
Jalan Ahmad Dahlan No. 76 Kota Kediri, Indonesia
cindyavitaselly@gmail.com

ABSTRAK

Perkembangan teknologi informasi menyebabkan peningkatan volume data sehingga diperlukan metode analisis data yang mampu membantu proses pengambilan keputusan secara efektif. Salah satu teknik yang banyak digunakan adalah clustering untuk segmentasi pelanggan. Namun, algoritma K-Means memiliki kelemahan karena sensitif terhadap distribusi data, perbedaan skala fitur, serta noise yang dapat mempengaruhi kualitas cluster. Penelitian ini bertujuan menganalisis pengaruh preprocessing dan reduksi dimensi terhadap kualitas clustering pada segmentasi pelanggan usia dewasa menggunakan algoritma K-Means. Dataset yang digunakan adalah Mall Customer Segmentation Data dengan atribut Age, Annual Income, dan Spending Score. Penelitian dilakukan menggunakan pendekatan CRISP-DM dengan beberapa skenario preprocessing, yaitu MinMaxScaler, RobustScaler, PowerTransformer, PCA, serta kombinasi preprocessing dan PCA. Evaluasi dilakukan menggunakan Silhouette Score dan Davies Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa metode PowerTransformer dan kombinasi PowerTransformer + PCA menghasilkan kualitas clustering terbaik dengan nilai Silhouette Score sebesar 0,5558 dan DBI sebesar 0,5618. Hasil tersebut menunjukkan bahwa preprocessing dan reduksi dimensi mampu meningkatkan kualitas segmentasi pelanggan secara lebih optimal.

Kata kunci : *Clustering, K-Means, PCA, PowerTransformer, Segmentasi Pelanggan*

1. PENDAHULUAN

Perkembangan teknologi informasi yang diiringi dengan meningkatnya jumlah data di berbagai bidang membuat kebutuhan terhadap teknik analisis data semakin tinggi. Salah satu metode dalam data mining yang sering digunakan adalah clustering, yaitu proses pengelompokan data berdasarkan tingkat kemiripan antar objek tanpa menggunakan label tertentu [1]. Clustering banyak diterapkan pada segmentasi pelanggan karena mampu membantu perusahaan memahami pola perilaku konsumen berdasarkan karakteristik tertentu [2].

Algoritma K-Means merupakan salah satu metode clustering yang banyak digunakan karena memiliki proses perhitungan yang relatif sederhana serta mampu mengolah data secara efisien [3]. Meskipun demikian, algoritma K-Means memiliki kelemahan karena sensitif terhadap distribusi data, perbedaan skala fitur, serta keberadaan noise dan outlier [4]. Kondisi tersebut dapat mempengaruhi proses pembentukan centroid sehingga kualitas cluster menjadi kurang optimal. Permasalahan tersebut menunjukkan bahwa kualitas distribusi data memiliki pengaruh penting terhadap performa algoritma K-Means dalam proses pembentukan cluster. Data yang memiliki perbedaan rentang nilai, distribusi yang tidak stabil, serta keberadaan noise dapat menyebabkan hasil clustering menjadi kurang representatif. Oleh karena itu, diperlukan metode preprocessing dan reduksi dimensi yang mampu memperbaiki kualitas data sebelum proses clustering dilakukan agar hasil segmentasi pelanggan menjadi lebih optimal.

Dataset Mall Customer Segmentation yang digunakan dalam penelitian ini memiliki atribut numerik berupa Age, Annual Income (k$), dan Spending Score (1–100). Karakteristik data yang memiliki rentang berbeda dan distribusi yang belum stabil menyebabkan proses clustering memerlukan tahapan preprocessing sebelum dilakukan pengelompokan data. Beberapa penelitian sebelumnya menunjukkan bahwa preprocessing dan reduksi dimensi mampu meningkatkan performa algoritma clustering [5].

PowerTransformer digunakan untuk memperbaiki distribusi data agar lebih mendekati distribusi normal, sedangkan Principal Component Analysis (PCA) digunakan untuk mereduksi dimensi data dan mengurangi noise [6]. Penelitian ini bertujuan untuk menganalisis pengaruh preprocessing dan reduksi dimensi terhadap kualitas clustering pada algoritma K-Means menggunakan beberapa skenario eksperimen. Evaluasi clustering dilakukan menggunakan Silhouette Score dan Davies Bouldin Index (DBI) untuk menentukan metode preprocessing yang paling optimal pada segmentasi pelanggan usia dewasa. Penelitian ini dilakukan dengan menerapkan beberapa skenario preprocessing dan reduksi dimensi pada algoritma K-Means untuk mengetahui metode yang memberikan performa clustering terbaik. Hasil penelitian diharapkan dapat membantu meningkatkan kualitas segmentasi pelanggan serta menjadi referensi dalam pengembangan penelitian clustering berbasis preprocessing pada bidang data mining.

2. TINJAUAN PUSTAKA

2.1. Clustering dan K-Means

Clustering merupakan salah satu teknik dalam data mining yang digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik antar objek [1]. Salah satu metode clustering yang cukup banyak digunakan adalah K-Means karena memiliki proses pengelompokan data yang sederhana dan efisien [7].

K-Means bekerja dengan menentukan centroid awal kemudian mengelompokkan data berdasarkan jarak terdekat terhadap centroid. Perhitungan jarak pada algoritma K-Means umumnya menggunakan Euclidean Distance dengan persamaan sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

a. Xi = nilai data pertama
b. yi = nilai data kedua
c. n = jumlah atribut

Algoritma K-Means memiliki kelemahan karena sensitif terhadap distribusi data dan perbedaan skala fitur sehingga diperlukan preprocessing sebelum proses clustering dilakukan [4]. Kondisi distribusi data yang tidak stabil dapat mempengaruhi proses pembentukan centroid pada algoritma K-Means.

Apabila terdapat perbedaan rentang nilai antar atribut, maka atribut dengan nilai yang lebih besar cenderung lebih dominan dalam proses perhitungan jarak. Selain itu, data yang memiliki skewness tinggi juga dapat menyebabkan hasil clustering menjadi kurang representatif karena posisi centroid tidak mampu menggambarkan kelompok data secara optimal. Oleh karena itu, proses preprocessing menjadi tahapan penting sebelum proses clustering dilakukan [4], [8].

2.2. Preprocessing Data

Preprocessing menjadi salah satu tahap penting dalam proses data mining yang dilakukan untuk memperbaiki dan meningkatkan kualitas data sebelum tahap clustering dijalankan [5]. Pada penelitian ini digunakan beberapa metode preprocessing, yaitu MinMaxScaler, RobustScaler, dan PowerTransformer.

PowerTransformer digunakan untuk memperbaiki distribusi data dan mengurangi skewness sehingga data menjadi lebih stabil untuk proses clustering berbasis jarak seperti K-Means[9]. Pada algoritma clustering berbasis jarak, kualitas distribusi data sangat mempengaruhi hasil pengelompokan. Data yang memiliki distribusi lebih stabil akan membantu proses perhitungan jarak antar data menjadi lebih seimbang sehingga cluster yang terbentuk dapat memiliki tingkat pemisahan yang lebih baik. Selain itu, preprocessing juga membantu mengurangi pengaruh noise dan ketidakseimbangan data yang dapat mempengaruhi kualitas segmentasi pelanggan [4], [5]. Selain preprocessing, penelitian ini juga menggunakan Principal Component Analysis (PCA) untuk mereduksi dimensi data dan mengurangi noise.

2.3. Penelitian Terdahulu

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma K-Means banyak digunakan dalam proses segmentasi pelanggan karena mampu melakukan pengelompokan data secara efisien. Penelitian yang dilakukan oleh Febrianty menerapkan algoritma K-Means pada segmentasi pelanggan transaksi online retail dan menunjukkan bahwa clustering dapat membantu perusahaan dalam menentukan strategi pemasaran yang lebih tepat sasaran [1].

Penelitian lain yang dilakukan oleh Kristanti menggunakan K-Means Clustering untuk segmentasi pelanggan berdasarkan usia, pendapatan, dan model RFM. Hasil penelitian tersebut menunjukkan bahwa proses segmentasi pelanggan mampu memberikan gambaran perilaku konsumen berdasarkan karakteristik tertentu sehingga dapat dimanfaatkan dalam pengambilan keputusan bisnis [2].

Selain itu, penelitian oleh Hilmy Naufan menunjukkan bahwa penerapan Principal Component Analysis (PCA) pada algoritma K-Means mampu meningkatkan kualitas clustering berdasarkan nilai Davies Bouldin Index (DBI) [6]. Berdasarkan beberapa penelitian tersebut, penelitian ini berfokus pada penerapan preprocessing dan reduksi dimensi menggunakan PowerTransformer dan PCA untuk meningkatkan kualitas clustering pada segmentasi pelanggan usia dewasa.

2.4. Evaluasi Clustering

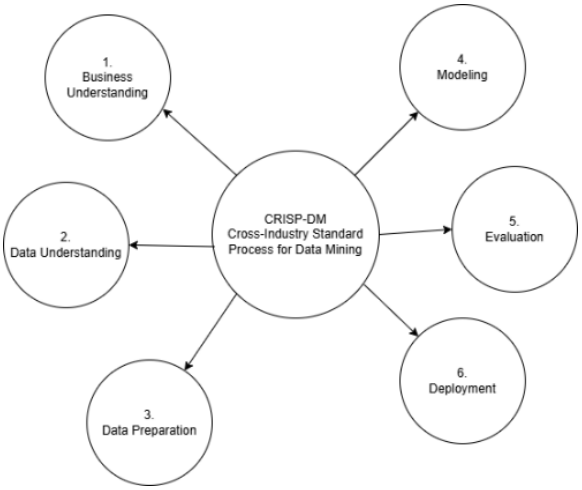
Evaluasi hasil clustering pada penelitian ini dilakukan menggunakan Silhouette Score dan Davies Bouldin Index (DBI). Silhouette Score digunakan untuk melihat seberapa baik suatu data berada pada cluster-nya dibandingkan dengan cluster lain. Sementara itu, DBI digunakan untuk menilai tingkat kedekatan antar cluster yang terbentuk. Semakin tinggi nilai Silhouette Score dan semakin rendah nilai DBI, maka kualitas clustering yang dihasilkan dinilai semakin optimal [2].

3. METODE PENELITIAN

Penelitian ini menerapkan metode CRISP-DM yang meliputi beberapa tahapan, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment sebagai alur utama dalam proses penelitian. Dataset yang digunakan adalah Mall Customer Segmentation Data yang diperoleh dari Kaggle. Penelitian menggunakan atribut numerik berupa Age, Annual Income (k$), dan Spending Score (1–100). Tahap preprocessing dilakukan menggunakan MinMaxScaler, RobustScaler, dan PowerTransformer [5].

Selain tahap preprocessing, penelitian ini juga menerapkan reduksi dimensi menggunakan Principal Component Analysis (PCA) [6]. Proses clustering dilakukan menggunakan improved K-Means dengan parameter init='k-means++', n_init=100, dan max_iter=500. Penentuan jumlah cluster optimal

dilakukan menggunakan Elbow Method dengan pengujian nilai K dari 2 hingga 9. Selanjutnya, evaluasi hasil clustering dilakukan menggunakan Silhouette Score dan Davies Bouldin Index (DBI) [2]. Seluruh proses pengolahan dan pengujian data dilakukan menggunakan Google Colaboratory dengan bahasa pemrograman Python.



Gambar 1. Alur penelitian menggunakan metode CRISP-DM

Gambar 1 menunjukkan alur penelitian menggunakan metode CRISP-DM mulai dari proses pengumpulan dataset hingga evaluasi hasil clustering. Tahapan preprocessing dan reduksi dimensi dilakukan sebelum proses clustering untuk meningkatkan kualitas segmentasi pelanggan.

3.1. Skenario Eksperimen

Tabel 1. Skenario eksperimen penelitian

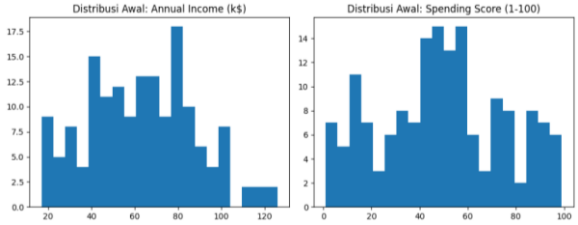
No	Metode	Preprocessing	PCA
1	Tanpa Preprocessing	Tidak	Tidak
2	MinMaxScaler	Ya	Tidak
3	MinMaxScaler + PCA	Ya	Ya
4	PowerTransformer	Ya	Tidak
5	PowerTransformer + PCA	Ya	Ya
6	PCA	Tidak	Ya
7	RobustScaler	Ya	Tidak
8	RobustScaler + PCA	Ya	Ya

Tabel 1 menunjukkan beberapa skenario eksperimen yang digunakan untuk mengetahui pengaruh preprocessing dan reduksi dimensi terhadap kualitas clustering pada algoritma K-Means.

4. HASIL DAN PEMBAHASAN

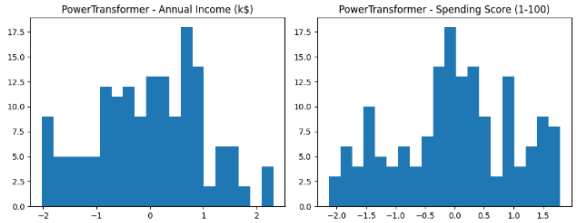
4.1. Hasil Preprocessing Data

Sebelum proses clustering dilakukan, dataset melalui tahap preprocessing untuk memperbaiki distribusi data dan mengurangi perbedaan skala antar atribut. Berdasarkan hasil preprocessing menggunakan PowerTransformer, distribusi data menjadi lebih stabil dibandingkan kondisi awal.



Gambar 2. Distribusi data sebelum preprocessing

Berdasarkan Gambar 2, distribusi data pada beberapa atribut masih terlihat belum stabil dan memiliki perbedaan persebaran antar fitur. Kondisi tersebut menunjukkan bahwa dataset masih memiliki karakteristik distribusi yang kurang optimal untuk proses clustering menggunakan algoritma K-Means. Selain itu, beberapa atribut masih menunjukkan kecenderungan skewness sehingga diperlukan preprocessing untuk memperbaiki kualitas distribusi data sebelum proses clustering dilakukan.

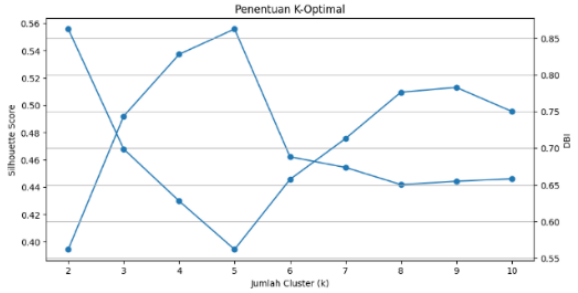


Gambar 3. Distribusi data setelah PowerTransformer

Berdasarkan Gambar 3, distribusi data setelah preprocessing menggunakan PowerTransformer terlihat lebih stabil dan mendekati distribusi normal. Kondisi tersebut membantu proses clustering menjadi lebih optimal karena perhitungan jarak antar data menjadi lebih representatif.

4.2. Penentuan Jumlah Cluster

Penentuan jumlah cluster optimal dilakukan menggunakan Elbow Method dengan melakukan pengujian pada beberapa nilai K. Berdasarkan hasil pengujian diperoleh titik elbow pada nilai K = 5 sehingga jumlah cluster tersebut digunakan pada seluruh eksperimen clustering.



Gambar 4. Hasil Elbow Method untuk menentukan jumlah cluster optimal

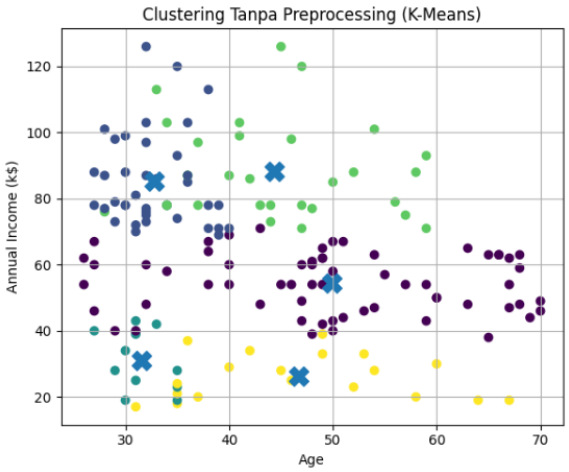
Berdasarkan Gambar 4, titik elbow terlihat pada nilai K = 5 sehingga jumlah cluster tersebut dipilih sebagai cluster optimal pada penelitian ini karena

memberikan keseimbangan terbaik antara kualitas cluster dan kompleksitas segmentasi.

4.3. Hasil Eksperimen Clustering

Hasil eksperimen menunjukkan bahwa preprocessing memberikan pengaruh signifikan terhadap kualitas clustering pada algoritma K-Means. Metode tanpa preprocessing menghasilkan nilai Silhouette Score sebesar 0,4632 dan DBI sebesar 0,7754. Hasil tersebut menunjukkan bahwa cluster masih memiliki tingkat overlap yang cukup tinggi. Penggunaan MinMaxScaler menghasilkan peningkatan kualitas clustering dengan nilai Silhouette Score sebesar 0,5565 dan DBI sebesar 0,5698. Sementara itu, metode RobustScaler menghasilkan performa yang lebih rendah dibandingkan metode preprocessing lainnya.

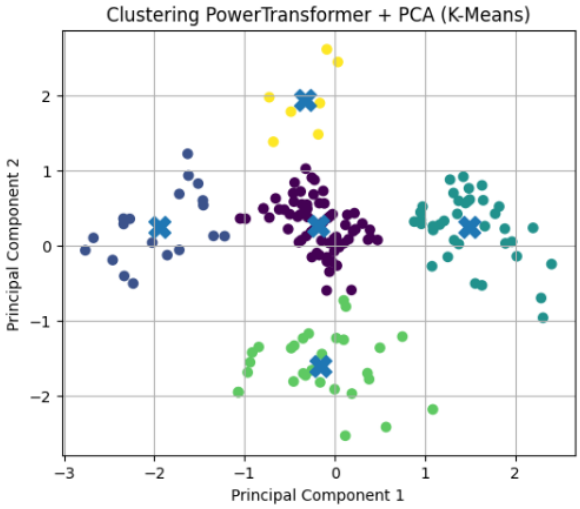
Hasil tersebut menunjukkan bahwa dataset penelitian ini tidak terlalu dipengaruhi oleh outlier ekstrem sehingga penggunaan RobustScaler belum memberikan peningkatan kualitas clustering yang signifikan. Selain itu, metode RobustScaler lebih berfokus pada normalisasi berbasis median dan Interquartile Range (IQR), namun tidak secara langsung memperbaiki distribusi data seperti PowerTransformer. Akibatnya, proses pembentukan cluster masih belum optimal dibandingkan metode preprocessing lainnya[4], [5].



Gambar 5. Hasil clustering tanpa preprocessing

Metode PowerTransformer menghasilkan kualitas clustering yang lebih baik karena mampu memperbaiki distribusi data dan mengurangi skewness. Perbaikan distribusi data menggunakan PowerTransformer membantu proses pembentukan centroid menjadi lebih stabil dibandingkan metode preprocessing lainnya. Distribusi data yang lebih mendekati normal menyebabkan perhitungan jarak Euclidean pada algoritma K-Means menjadi lebih representatif sehingga cluster yang terbentuk memiliki tingkat overlap yang lebih rendah. Kondisi tersebut menunjukkan bahwa kualitas distribusi data memiliki pengaruh besar terhadap performa algoritma clustering berbasis centroid[4], [8]. Berdasarkan hasil evaluasi, diperoleh nilai Silhouette Score sebesar 0,5558 dan DBI sebesar 0,5618. Hasil tersebut menunjukkan bahwa cluster yang dihasilkan memiliki kualitas pemisahan yang lebih baik dibandingkan metode lainnya.

Berdasarkan Gambar 5, hasil clustering tanpa preprocessing menunjukkan tingkat overlap antar cluster yang masih cukup tinggi. Kondisi tersebut menunjukkan bahwa distribusi data awal belum mampu menghasilkan pemisahan cluster yang optimal pada algoritma K-Means.



Gambar 6. Hasil clustering menggunakan PowerTransformer dan Principal Component Analysis (PCA)

Berdasarkan Gambar 6, cluster yang terbentuk menggunakan PowerTransformer dan PCA terlihat lebih jelas dan memiliki tingkat overlap yang rendah. Hal tersebut menunjukkan bahwa preprocessing dan reduksi dimensi membantu meningkatkan kualitas segmentasi pelanggan.

4.4. Perbandingan Hasil Seluruh Metode

Tabel 2. Perbandingan hasil evaluasi clustering

Metode	K	Silhouette	DBI
Tanpa Preprocessing	5	0,4632	0,7754
MinMaxScaler	5	0,5565	0,5698
MinMax + PCA	5	0,5565	0,5698
PCA	5	0,5531	0,5712
PowerTransformer	5	0,5558	0,5618
PowerTransformer + PCA	5	0,5558	0,5618
RobustScaler	5	0,4278	0,8596
RobustScaler + PCA	5	0,5478	0,5702

Berdasarkan Tabel 2, metode PowerTransformer dan kombinasi PowerTransformer + PCA menghasilkan kualitas clustering terbaik dibandingkan metode lainnya. Hasil tersebut menunjukkan bahwa kualitas distribusi data berpengaruh terhadap performa algoritma K-Means dalam membentuk cluster yang lebih optimal. Penggunaan Principal Component Analysis (PCA) membantu menyederhanakan

representasi data dengan mempertahankan informasi utama yang memiliki variansi terbesar.

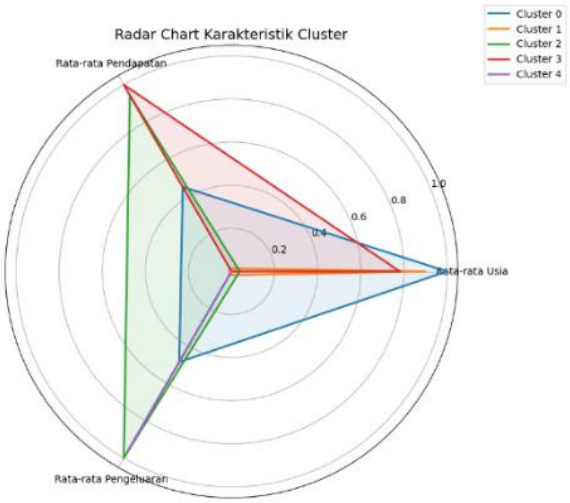
Selain membantu mengurangi kompleksitas dimensi, PCA juga membantu mengurangi noise pada dataset sehingga proses clustering menjadi lebih stabil. Meskipun peningkatan nilai evaluasi tidak terlalu jauh dibandingkan PowerTransformer tanpa PCA, penggunaan PCA tetap memberikan manfaat dalam efisiensi representasi data[6].

Berdasarkan seluruh hasil eksperimen, kombinasi PowerTransformer dan Principal Component Analysis (PCA) memberikan performa paling optimal dibandingkan metode lainnya. Kombinasi tersebut mampu menghasilkan distribusi data yang lebih stabil serta membantu proses reduksi noise sehingga kualitas cluster menjadi lebih baik dan lebih mudah dipisahkan antar kelompok data[6].

4.5. Analisis Karakteristik Cluster

Hasil clustering membentuk lima kelompok pelanggan dengan karakteristik yang berbeda berdasarkan usia, pendapatan, dan tingkat pengeluaran pelanggan. Cluster dengan pendapatan dan tingkat pengeluaran tinggi menunjukkan kelompok pelanggan potensial yang dapat dijadikan target utama dalam strategi pemasaran. Hasil segmentasi pelanggan dapat dimanfaatkan sebagai dasar dalam penyusunan strategi pemasaran yang lebih tepat sasaran.

Kelompok pelanggan dengan tingkat pengeluaran tinggi dapat dijadikan target utama dalam program loyalitas pelanggan maupun promosi premium. Sementara itu, pelanggan dengan tingkat pendapatan tinggi namun pengeluaran rendah dapat diberikan strategi pemasaran yang lebih personal untuk meningkatkan aktivitas pembelian[2], [10]. Sementara itu, cluster dengan pendapatan tinggi namun pengeluaran rendah menunjukkan pelanggan yang lebih selektif dalam melakukan pembelian.



Gambar 7. Perbandingan karakteristik pelanggan pada setiap cluster

Gambar 7 menunjukkan perbedaan karakteristik pelanggan pada setiap cluster berdasarkan rata-rata usia, pendapatan, dan tingkat pengeluaran pelanggan. Hasil segmentasi menunjukkan bahwa setiap cluster memiliki pola perilaku pelanggan yang berbeda.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa preprocessing dan reduksi dimensi memberikan pengaruh signifikan terhadap kualitas clustering pada algoritma K-Means. Metode PowerTransformer dan kombinasi PowerTransformer + PCA menghasilkan kualitas clustering terbaik dengan nilai Silhouette Score tertinggi dan DBI terendah dibandingkan metode lainnya. PowerTransformer mampu memperbaiki distribusi data sehingga proses pembentukan centroid menjadi lebih stabil dan representatif. Penelitian selanjutnya disarankan menggunakan dataset dengan jumlah data yang lebih besar serta melakukan perbandingan dengan algoritma clustering lain, seperti DBSCAN dan Agglomerative Clustering, agar hasil segmentasi pelanggan dapat diperoleh secara lebih optimal. Selain itu, hasil penelitian ini menunjukkan bahwa kualitas preprocessing memberikan pengaruh yang cukup besar terhadap performa algoritma clustering, terutama pada metode berbasis centroid seperti K-Means.

DAFTAR PUSTAKA

[1] E. Febrianty, L. Awalina, and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions," *Jurnal Teknologi dan Informasi (JATI)*, vol. 13, 2023, doi: 10.34010/jati.v13i2.

[2] B. T. Kristanti, A. Junaidi, and E. P. Mandyartha, "IMPLEMENTASI K-MEANS CLUSTERING DALAM SEGMENTASI PELANGGAN BERDASARKAN USIA, PENDAPATAN, DAN MODEL RFM (STUDI KASUS: LANTIKYA STORE JOMBANG)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4677.

[3] R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," 2020. doi: 10.51519/journalsea.v1i1.10.

[4] G. T. Reddy *et al.*, "Analysis of Dimensionality Reduction Techniques on Big Data," *IEEE Access*, vol. 8, pp. 54776–54788, 2020, doi: 10.1109/ACCESS.2020.2980942.

[5] U. Surapati and M. Jannah, "Penerapan Data Mining Menggunakan Metode K-Means Untuk Mengetahui Minat Customer Dalam Pembelian Merchandise Kpop," *Jurnal Sains dan*

Teknologi, vol. 5, no. 3, 2024, doi: 10.55338/saintek.v5i1.2739.

[6] M. Hilmy Naufan, R. Kurniawan, and T. Suprapti, "OPTIMASI NILAI DAVIES BOULDIN INDEX PADA PROGRAM PENDAFTARAN TANAH SISTEMATIS LENGKAP (PTSL) MENGGUNAKAN ALGORITMA K-MEANS DAN PCA," *Jurnal Teknik Elektro dan Informatika*, vol. 20, pp. 17–28, 2025, doi: 10.30587/e-link.v20i1.9063.

[7] D. Marcelina, A. Kurnia, and T. Terttiaavini, "Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 293–301, Nov. 2023, doi: 10.57152/malcom.v3i2.952.

[8] T. M. Ghazal *et al.*, "Performances of k-means clustering algorithm with different distance metrics," *Intelligent Automation and Soft Computing*, vol. 30, no. 2, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.

[9] D. Irawan, G. Wijaya, and T. T. Warisaji, "Penerapan Algoritma K-Means Clustering untuk Segmentasi Nasabah Bank," *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 6, no. 1, pp. 47–53, Mar. 2025, doi: 10.37148/bios.v6i1.162.

[10] F. D. Wahyuningtyas, A. Arafat, A. Stiawan, and D. Rolliawati, "Komparasi Algoritma Hierarchical, K-Means, dan DBSCAN pada Analisis Data Penjualan Melalui Facebook," *Explore: Jurnal Sistem Informasi dan Telematika*, vol. 14, no. 1, p. 7, Jun. 2023, doi: 10.36448/jsit.v14i1.2931