

Pendekatan Data Mining dalam Optimalisasi Margin Penjualan Adidas: Studi Klasterisasi dengan K-Means dan Fuzzy C-Means

Lutfiah Firlan dkk.



Volume 13, Issue 2, Pages 229–237, Aug. 2025

Diterima 4 Mei 2025, Direvisi 8 Juli 2025, Disetujui 12 Juli 2025, Diterbitkan 20 Juli 2025

To Cite this Article : L. Firlan, R. Fadila, M. K. Ridho, E. Sunandi, dan U. Aflah, "Pendekatan Data Mining dalam Optimalisasi Margin Penjualan Adidas: Studi Klasterisasi dengan K-Means dan Fuzzy C-Means", *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 13, no. 2, pp. 229–237, 2025, <https://doi.org/10.37905/euler.v13i2.32417>

© 2025 by author(s)

JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI



	Homepage	:	http://ejournal.ung.ac.id/index.php/euler/index
	Journal Abbreviation	:	Euler J. Ilm. Mat. Sains dan Teknol.
	Frequency	:	Three times a year
	Publication Language	:	English (preferable), Indonesia
	DOI	:	https://doi.org/10.37905/euler
	Online ISSN	:	2776-3706
	License	:	Creative Commons Attribution-NonCommercial 4.0 International License
	Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
	Country	:	Indonesia
	OAI Address	:	http://ejournal.ung.ac.id/index.php/euler/oai
	Google Scholar ID	:	QF_r-gAAAAJ
	Email	:	euler@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



Jambura Journal of Biomathematics



Jambura Journal of Mathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics

Pendekatan Data Mining dalam Optimalisasi Margin Penjualan Adidas: Studi Klasterisasi dengan K-Means dan Fuzzy C-Means

Lutfiah Firlia¹, Risfa Fadila¹, Muhammad Kevyn Ridho¹, Etis Sunandi^{1,*}, Ukasyah Aflah¹

¹Program Studi Statistika, Universitas Bengkulu, Bengkulu 28122, Indonesia

ARTICLE HISTORY

Diterima 4 Mei 2025
Direvisi 8 Juli 2025
Disetujui 12 Juli 2025
Diterbitkan 20 Juli 2025

KATA KUNCI

Klaster
Fuzzy C-Means
K-Means
Adidas

KEYWORDS

Clustering
Fuzzy C-Means
K-Means
Adidas

ABSTRAK. Strategi penjualannya untuk mempertahankan daya saing, terutama di pasar Amerika Serikat. Penelitian ini memiliki tujuan untuk mengelompokkan data penjualan Adidas di AS dengan membandingkan metode clustering K-Means dan Fuzzy C-Means, guna memperoleh segmentasi pasar yang efektif. Pengelolaan data dilakukan dengan proses standarisasi data terlebih dahulu dari variabel unit terjual, total penjualan, laba operasional dan margin keuntungan. Analisis cluster dilakukan dengan menentukan cluster yang optimal dengan elbow plot, dan didapatkan hasil yang paling optimal adalah empat cluster. Evaluasi performa cluster dilakukan dengan membandingkan nilai Davies-Bouldin Index (DBI), Sum of Squares Within (SSW), dan Sum of Squares Between (SSB) dari masing-masing metode. Hasil analisis menunjukkan bahwa K-Means menghasilkan DBI sebesar 0.98 dan SSW sebesar 8645.74, sementara Fuzzy C-Means memiliki DBI sebesar 0.983 dan SSW sebesar 8678.49. Berdasarkan hasil tersebut, K-Means dinilai lebih optimal karena menghasilkan cluster yang lebih kompak dan terpisah secara jelas. Temuan ini dapat menjadi dasar dalam menyusun strategi penjualan yang lebih terarah dan efisien, sesuai dengan karakteristik masing-masing segmen pasar.

ABSTRACT. Adidas is trying to stay competitive, especially when it comes to selling in the U.S. market. This study aims to cluster Adidas sales data in the US by comparing K-Means and Fuzzy C-Means clustering methods to achieve effective market segmentation. Data management is performed by first standardizing the data from the variables of units sold, total sales, operating profit, and profit margin. Cluster analysis is performed by determining the optimal cluster with the elbow plot, and the most optimal result is four clusters. Cluster performance evaluation was performed by comparing the Davies-Bouldin Index (DBI), Sum of Squares Within (SSW), and Sum of Squares Between (SSB) values of each method. The analysis results show that K-Means has a DBI of 0.98 and an SSW of 8645.74, while Fuzzy C-Means has a DBI of 0.983 and an SSW of 8678.49. Based on these results, K-Means is considered more optimal because it produces clusters that are more compact and separated. This finding can be the basis for developing a more targeted and efficient sales strategy according to the characteristics of each market segment.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. *Editorial of EULER:* Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

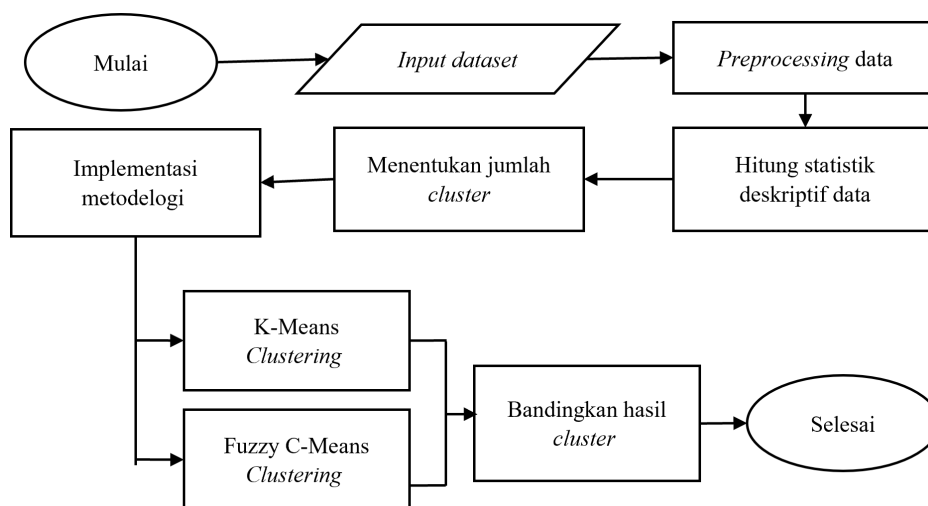
1. Pendahuluan

Dominasi pasar Adidas di Amerika Serikat (AS) sedang menghadapi tantangan signifikan di tengah lanskap industri olahraga dan *fashion* yang kian dinamis. Dengan pangsa pasar yang substansial, merek global ini harus terus berinovasi dalam strategi penjualan untuk mempertahankan posisinya di tengah preferensi konsumen yang berubah cepat dan persaingan merek yang semakin intens. Untuk mempertahankan daya saingnya, Adidas yang merupakan pemimpin pasar dalam pakaian dan perlengkapan olahraga, harus terus mengoptimalkan strategi penjualannya. Terutama berlaku di pasar Amerika Serikat, di mana merek ini memiliki pangsa pasar terbesar. Segmentasi pasar adalah salah satu strategi yang dapat digunakan oleh bisnis untuk menyesuaikan strategi pemasaran mereka untuk memenuhi kebutuhan dan preferensi target pasar. Sebagai bagian dari upaya untuk menyipkan produk dengan merek dan citra perusahaan tertentu, se-

gmentasi pasar memungkinkan perusahaan untuk menganggap posisi perusahaan sebagai penting dibandingkan pesaing lainnya [1]. Untuk menghadapi kompleksitas data konsumen yang masif untuk menentukan strategi segmentasi yang tepat, pendekatan konvensional seringkali tidak memadai. Dengan demikian data mining adalah metode untuk menganalisis pola dan karakteristik di masa depan serta untuk mengumpulkan informasi tak terduga yang belum pernah terlihat sebelumnya dari *database* yang besar [2].

Salah satu pendekatan penting dalam data mining adalah teknik *clustering*, yang bertujuan untuk mengelompokkan data ke dalam kelompok yang homogen. Teknik *clustering*, seperti K-Means dan Fuzzy C-Means, sering digunakan dalam segmentasi untuk mengelompokkan data kedalam sejumlah *cluster* berdasarkan atribut tertentu. Algoritma K-Means membagi data ke dalam *k cluster* berdasarkan jarak terdekat antara data dan pusat *cluster*. K-Means *Clustering* adalah algoritma yang mengelompokkan data ke dalam beberapa *cluster* berdasarkan kesamaan karakte-

*Penulis Korespondensi.



Gambar 1. Prosedur penelitian

ristik tertentu [3]. Metode ini sederhana dan efisien dalam mengelompokkan data dalam jumlah besar. Namun, K-Means memiliki kelemahan dalam menangani data dengan batas *cluster* yang tidak jelas atau tumpang tindih [4]. Sebaliknya, Fuzzy C-Means memungkinkan setiap data memiliki derajat keanggotaan dalam beberapa *cluster*, sehingga lebih fleksibel dalam menangani data dengan batas yang kabur.

Beberapa penelitian melakukan perbandingan metode K-Means dan Fuzzy C-Means diantaranya pada [5–7]. Hasil penelitian menunjukkan bahwa Fuzzy C-Means (FCM) merupakan salah satu algoritma *clustering* yang paling populer. Algoritma ini lebih baik daripada K-Means, karena memberikan informasi yang lebih kaya dalam himpunan data yang kompleks maupun sederhana berkat konsep derajat keanggotaannya [8]. Meskipun demikian, klaim superioritas FCM sangat bergantung pada karakteristik spesifik data dan tujuan analisis. Di sisi lain, K-Means umumnya lebih efisien dalam hal waktu komputasi. Oleh karena itu, pemilihan algoritma yang tepat harus mempertimbangkan karakteristik data dan tujuan analisis secara cermat.

Dengan mempertimbangkan keunggulan dan keterbatasan masing-masing metode *clustering*, penelitian ini bertujuan untuk mengoptimalkan strategi penjualan Adidas di AS melalui penerapan analisis K-Means dan Fuzzy C-Means. Dengan menggunakan kedua metode ini, diharapkan segmentasi pasar dapat dilakukan secara lebih akurat, memungkinkan perusahaan untuk memahami pola pembelian pelanggan dengan lebih baik serta merancang strategi pemasaran yang lebih efektif. Hasil dari penelitian ini diharapkan dapat memberikan wawasan berharga bagi Adidas dalam meningkatkan daya saingnya di pasar yang semakin kompetitif serta menjadi referensi bagi industri serupa dalam mengoptimalkan strategi pemasaran berbasis data mining.

2. Metode

2.1. Data dan Variabel Penelitian

Pada penelitian ini data yang digunakan adalah *dataset* penjualan Adidas yaitu kumpulan data yang mencakup informasi tentang penjualan produk Adidas di Amerika Serikat. Data ini diperoleh dari *website kaggle* (<https://www.kaggle.com/datasets/>). Terdapat beberapa variabel pada *dataset* ini, yang disajikan pada Tabel 1.

Tabel 1. Variabel penelitian

No	Variabel	Keterangan	Tipe data/Satuan
1	Kota	Nama kota tempat penjualan terjadi.	Nominal
2	Unit terjual	Total unit produk terjual	Jumlah unit
3	Total penjualan	Laba operasional dari hasil penjualan produk	Dolar AS
4	Laba operasional	Keuntungan operasional	Dolar AS
5	Margin Keuntungan	Persentase keuntungan operasional terhadap total penjualan	Persentase

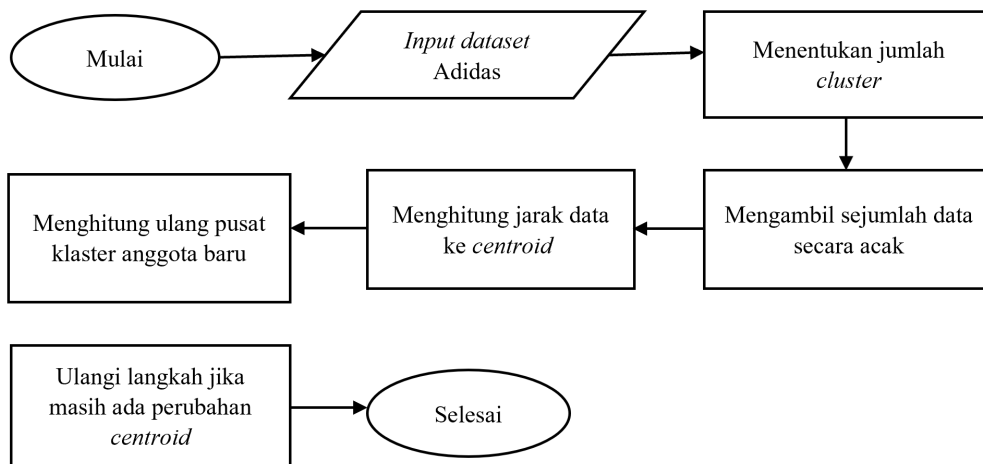
2.2. Prosedur Penelitian

2.2.1. Alur Penelitian

Alur penelitian yang dilakukan pada penelitian ini yaitu diawali dengan pengumpulan data. Data tersebut terlebih dahulu diolah pada tahap *preprocessing data* meliputi mendeteksi missing value, menghapus outlier, dan standarisasi data. Setelah itu dilanjutkan pada analisis selanjutnya yaitu K-Means dan Fuzzy C-Means. Alur penelitian diberikan dalam bentuk diagram alir pada Gambar 1.

2.2.2. Algoritma K-Means

Algoritma K-means adalah metode yang memerlukan parameter input berupa jumlah kluster, yaitu k , untuk mengelompokkan sekumpulan objek sebanyak n ke dalam k kluster. Tujuan utama algoritma ini adalah memastikan bahwa objek-objek dalam satu kluster memiliki tingkat kemiripan yang tinggi, sedangkan kemiripan dengan objek di kluster lain seminimal mungkin. Kemiripan ini diukur berdasarkan kedekatan objek terhadap nilai rata-rata dalam kluster tersebut, yang dikenal sebagai *centroid* atau pusat massa kluster. Algoritma ini bekerja secara iteratif dengan memperbarui posisi *centroid* hingga diperoleh kluster yang optimal. K-means banyak digunakan dalam berbagai bidang seperti segmentasi pelanggan, pengelompokan data citra, dan analisis pola dalam data besar [9]. Algoritma K-Means adalah salah satu algoritma terpenting dalam bidang data mining karena memiliki keunggulan dalam kemudahan implementasi, efisiensi waktu



Gambar 2. Algoritma K-Means

komputasi, serta penerapannya yang luas dalam berbagai permasalahan komputasi. Algoritma ini dikenal sebagai metode pengelompokan yang paling populer, karena mampu membagi data ke dalam kluster secara efektif dan telah digunakan dalam berbagai aplikasi, mulai dari analisis data hingga kecerdasan buatan [10]. Algoritma K-Means ini sederhana dan bekerja dengan baik untuk kumpulan data besar jika dibandingkan dengan pengelompokan hierarkis sehingga bagus untuk inisialisasi parameter. Evaluasi kinerja yang dikembangkan telah dilakukan dengan melakukan analisis berbagai kategori gambar yang berbeda sebagai studi kasus [11]. Selanjutnya, algoritma K-Means Clustering [12] ditunjukkan pada Gambar 2.

Adapun jarak data ke centroid dapat diperoleh menggunakan persamaan berikut:

$$D_{ij} = \sqrt{(x_{1j} - v_{1j})^2 + \dots + (x_{1j} - v_{kj})^2}, \quad (1)$$

dimana D merupakan jarak cluster, X_{ik} adalah nilai data pada i dan j , C_{jk} merupakan nilai centroid pada j dan k . Kemudian, menentukan data berdasarkan pusat cluster yang paling dekat. Untuk menghitung pusat cluster yang baru dapat menggunakan rumus:

$$C_1 = \frac{1}{M} \sum_{j=1}^M x_j, \quad (2)$$

dimana C_i merupakan pusat cluster ke $-i$ dan d adalah titik data ke- i , n banyaknya cluster k .

2.2.3. Algoritma Fuzzy C-Means

Fuzzy C-Means (FCM) adalah sebuah metode pengelompokan data yang merupakan salah satu bentuk dari *soft clustering*. Berbeda dengan *hard clustering* yang menetapkan data ke dalam satu cluster secara eksklusif, metode ini bekerja dengan memberikan setiap titik data sebuah nilai derajat keanggotaan (*degree of membership*) untuk setiap cluster yang ada. Nilai keanggotaan ini berkisar antara 0 hingga 1, yang menunjukkan seberapa kuat hubungan sebuah data dengan suatu cluster. Dengan demikian, satu titik data dapat menjadi bagian dari beberapa cluster sekaligus dengan tingkat keanggotaan yang berbeda-beda [13, 14]. Berikut adalah algoritma Fuzzy C-Means Clustering yang diperoleh [15]:

1. Masukkan data menjadi matriks berukuran $n \times m$.
2. Tetapkan beberapa parameter yaitu:
 - (a) Jumlah cluster (c)
 - (b) Pangkat (w)
 - (c) Maksimum iterasi ($MaxIter$)
 - (d) Error terkecil (ϵ)
 - (e) Fungsi Objektif ($P_O = 0$)
 - (f) Iterasi awal ($t = 1$)

Pada parameter *fuzzifier* (m) diatur pada nilai standar 2 untuk menjamin keseimbangan terbaik antara kejelasan dan tumpang tindih antar cluster [16].

3. Tentukan bilangan random μ_{ik} ($i = 1, 2, \dots, n$ dan $k = 1, 2, \dots, c$) sebagai elemen matriks U . Kemudian, hitung jumlah kolom dengan persamaan berikut:

$$Q_i = \sum_{k=1}^c \mu_{ik}. \quad (3)$$

Matriks random untuk $= 1, 2, \dots, n$ menggunakan persamaan berikut:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i}. \quad (4)$$

4. Pusat cluster ke $-k$ dapat dicari dengan persamaan berikut:

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n ((\mu_{ik})^w)}. \quad (5)$$

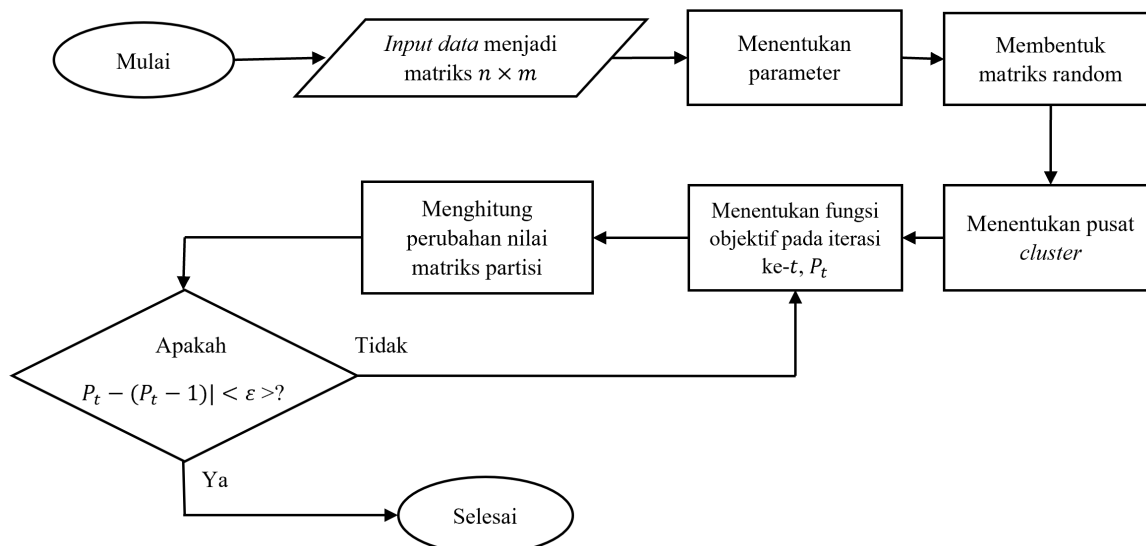
5. Menentukan fungsi objektif pada iterasi ke- t , P_t dengan persamaan sebagai berikut:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (X_{ij} - V_{Kj})^2] (\mu_{ik})^w). \quad (6)$$

6. Menghitung perubahan nilai matriks partisi dengan persamaan sebagai berikut:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{Kj})^2]^{\frac{-1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{Kj})^2]^{\frac{-1}{w-1}}}, \quad (7)$$

untuk nilai $i = 1, 2, \dots, n$ dan nilai $k = 1, 2, \dots, c$.



Gambar 3. Algoritma Fuzzy C-Means

Tabel 2. Data bersih

Nomor	Unit Terjual	Total Penjualan	Laba Operasional	Margin Keuntungan
1	4,401685	3,570609	4,900526	0,792292
2	3,468205	2,865967	2,132647	-1,26538
3	3,468205	2,161325	1,948121	-0,75096
4	2,768095	2,038012	1,8351	-0,75096
5	3,001465	3,147823	2,354077	-1,26538
6	3,468205	2,865967	1,671333	-1,77979
7	4,635055	3,746769	5,131183	0,792292
⋮	⋮	⋮	⋮	⋮
9398	-1,012499	-0,6468152	-0,6254013	-0,6480758

7. Cek kondisi hingga berhenti dengan persamaan:

$$P_t - (P_t - 1) < \epsilon,$$

atau ($t > MaxIter$) maka berhenti. Jika tidak $t = t + 1$, maka ulangi Langkah ke-4.

Algoritma Fuzzy C-Means Clustering dapat dilihat pada Gambar 3.

2.2.4. Evaluasi Cluster

Para peneliti melakukan evaluasi cluster untuk menentukan jumlah cluster yang terbaik dan optimal. Salah satu teknik evaluasi cluster adalah Davies Bouldin Index (DBI), yang dapat menyempurnakan hasil kinerja dari metode cluster atau kelompok optimalisasi untuk mendapatkan jumlah cluster yang terbaik. Selanjutnya akan dihitung matriks kohesi dalam sebuah cluster ke- i menggunakan *Sum of Square Within Cluster* (SSW), yang rumusnya adalah sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_i, c_i), \tag{8}$$

dengan:

- m = jumlah data dalam cluster,
- c = centroid cluster,
- d = jarak.

Sum of Square Between Cluster (SSB) diperlukan untuk menghitung separasi antar cluster. Persamaan untuk *Sum of Square Between Cluster* (SSB) adalah sebagai berikut:

$$SSB_i = d(c_i, c_j). \tag{9}$$

Setelah memperoleh nilai kohesi dan separasi. Langkah selanjutnya yaitu melakukan pengukuran rasio (R) untuk mengetahui nilai perbandingan antar cluster ke- i dan ke- j . Cluster yang memiliki nilai kohesi terkecil dan separasi yang terbesar adalah cluster yang paling baik. Berikut persamaan dalam menghitung nilai rasio:

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i=j} (R_{ij}), \tag{10}$$

dimana k merupakan jumlah cluster yang digunakan. Cluster yang diperoleh dapat dikatakan semakin baik jika nilai DBI semakin kecil yaitu ≥ 0 .

3. Hasil dan Pembahasan

3.1. Preprocessing Data

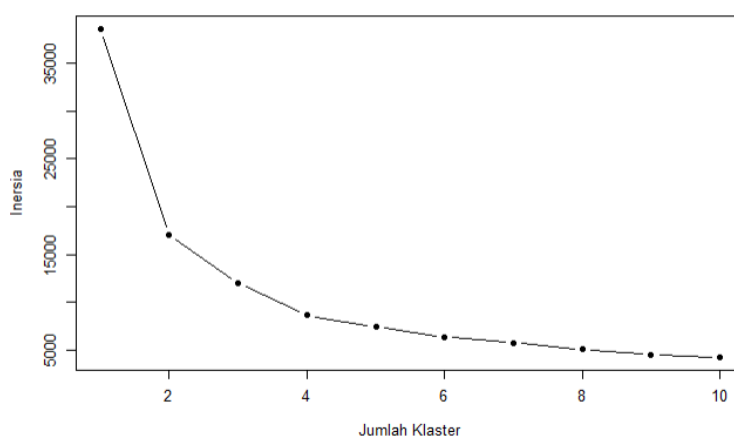
Pada penelitian kali ini digunakan data penjualan dari Perusahaan Adidas di AS, dengan variabel yang digunakan adalah unit terjual, total penjualan, laba operasional dan margin keuntungan. Sebelum dilakukan analisis pada data akan dilakukan *preprocessing* data terlebih dahulu. *Preprocessing* yang dilakukan kali

Tabel 3. Statistik deskriptif

	Unit Terjual	Total Penjualan	Laba Operasional	Margin Keuntungan
Minimum	-1,1992	-0,6572	-0,6352	-3,3230
Kuartil 1	-0,7045	-0,6273	-0,5998	-0,7510
Median	-0,37777	-0,5898	-0,5546	-0,1337
Mean	0	0	0	0
Kuartil 3	0,4344	0,3997	0,3255	0,6894
Maximum	4,7517	5,1561	6,5613	3,8788

Tabel 4. Pusat cluster K-Means

	Unit Terjual	Total Penjualan	Laba Operasional	Margin Keuntungan
Cluster 1	162,7482	32491,94	11190,79	0,365133
Cluster 2	764,9535	454495,7	177176,6	0,39621
Cluster 3	131,5675	7307,661	3828,496	0,514931
Cluster 4	463,4976	223175,1	77838,43	0,356563

**Gambar 4.** Pemilihan jumlah cluster optimal

ini dengan melakukan *scaling* data. *Scaling* data atau standarisasi data digunakan agar dapat meningkatkan performa algoritma. Hasil data yang telah di *preprocessing* dapat dilihat pada **Tabel 2**

Sebelum melakukan analisis lanjut diperlukan statistik deskriptif, karena sebelum dilakukan pengelompokan data menjadi beberapa *cluster*, dicari terlebih dahulu ciri-ciri data dari setiap variabel. Maka pada penelitian ini, hasil statistik deskriptif dari data yang telah di *preprocessing*, dapat dilihat pada **Tabel 3**.

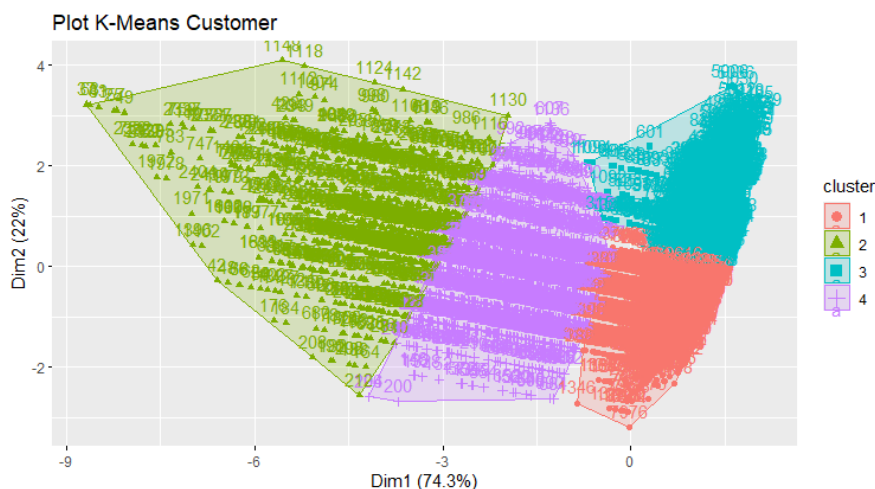
Pada **Tabel 3** dapat dilihat bahwa rata-rata terpusat di sekitar nol, yang memastikan bahwa skala seimbang untuk analisis, sehingga tidak atiribut yang mendominasi setelah dinormalisasi. Meskipun sebagian besar data tersebar di sekitar nilai nol dengan rentang antar kuartil yang relatif sempit, interpretasi bisnis yang lebih dalam mengungkap dinamika krusial. Secara khusus, sebaran "ke dua arah" pada Margin Keuntungan dengan nilai median dan kuartil pertama yang negatif bukan sekadar 'keragaman data', melainkan sebuah indikasi kuat adanya tantangan profitabilitas sistemik, di mana lebih dari separuh transaksi berpotensi tidak menguntungkan. Selanjutnya, nilai ekstrem yang teridentifikasi bukan sekadar data yang jauh dari rata-rata, melainkan representasi dari dua kelompok strategis, nilai maksimum pada Laba Operasional menyoroti keberadaan segmen pelanggan yang sangat profitabel, sementara nilai minimum pada Margin Keuntungan menandakan adanya transaksi yang sangat merugikan. Oleh karena itu, temuan gabungan ini menetapkan tujuan yang jelas

untuk analisis segmentasi selanjutnya: mengisolasi secara akurat kelompok pelanggan yang paling menguntungkan dari kelompok yang merugikan, agar dapat dirancang strategi bisnis yang spesifik dan efektif untuk masing-masing.

3.2. Implementasi Algoritma K-Means

Pada penelitian kali dalam menentukan jumlah *cluster* dilakukan dengan melihat *Elbow plot*. Maka didapatkan hasil dari *Elbow plot* pada data kali ini dapat dilihat pada **Gambar 4**.

Berdasarkan analisis Metode Siku (*Elbow Method*) pada **Gambar 4**, jumlah *cluster* yang paling optimal untuk dataset ini adalah empat. Grafik tersebut memetakan hubungan antara jumlah *cluster* (k) pada sumbu horizontal dengan nilai Inertia yang mengukur kepadatan data dalam *cluster* pada sumbu vertikal. Terlihat jelas adanya penurunan nilai inertia yang sangat tajam saat jumlah *cluster* bertambah dari satu hingga empat, yang menandakan bahwa setiap penambahan *cluster* pada rentang ini secara signifikan meningkatkan kualitas pengelompokan data. Namun, setelah mencapai empat *cluster*, kurva mulai melandai secara drastis. Hal ini menunjukkan bahwa penambahan *cluster* lebih lanjut (misalnya menjadi lima atau enam) tidak lagi memberikan perbaikan yang sepadan dengan kompleksitas yang ditambahkan. Oleh karena itu, titik "siku" yang terbentuk pada $k = 4$ diidentifikasi sebagai titik optimal, karena menawarkan keseimbangan terbaik antara kepadatan *cluster* yang tinggi dan jumlah segmen yang efi-



Gambar 5. Visualisasi hasil K-Means

Tabel 5. Hasil cluster K-Means

No	Kota	Unit Terjual	Total Penjualan	Laba Operasional	Margin Keuntungan	Cluster
1	New York	900	225000	78750	0,35	1
2	New York	700	245000	85750	0,35	1
3	New York	700	245000	110250	0,45	1
4	Chicago	275	123750	61875	0,5	2
5	Chicago	275	123750	61875	0,5	2
6	Chicago	225	112500	56250	0,5	2
7	Los Angeles	275	123750	43312,5	0,35	3
8	Los Angeles	275	96250	33687,5	0,35	3
9	Los Angeles	125	50000	20000	0,4	3
10	New York	1200	600000	300000	0,5	4
11	New York	1000	500000	150000	0,3	4
12	New York	1000	400000	140000	0,35	4

sien. Setelah dilakukan penentuan jumlah cluster optimal, selanjutnya dilakukan analisis dengan menggunakan cluster K-Means, yaitu nilai rata-rata dari beberapa variabel pada data penjualan Adidas yang menjadi titik pusat setiap cluster. Rekapitan hasil yang diperoleh disajikan pada Tabel 4.

Hasil clustering menggunakan metode K-Means dengan 4 cluster menunjukkan bahwa setiap cluster memiliki karakteristik yang berbeda. Cluster 1 terdiri dari rata-rata penjualan rendah dengan laba operasional hanya 11190,79 dengan rata-rata margin keuntungan hanya 0,365 maka, cluster 1 dapat dikategorikan sebagai unit dengan performa rendah. Cluster 2 memiliki rata-rata penjualan tertinggi dengan total penjualan mencapai 454406,7. Serta, rata-rata laba operasionalnya 177176 dengan margin keuntungan yang menengah yaitu 0,396 maka, cluster 2 dapat dikategorikan sebagai cluster unggulan (top perfoma). Pada cluster 3 memiliki rata-rata penjualan yang terendah dengan total penjualan hanya 7308 akan tetapi, rata-rata margin keuntungannya sangat tinggi yaitu 0,5149. Hal ini menunjukkan bahwa produk yang terjual di Cluster 3 merupakan produk premium yang berkualitas sehingga harganya sangat tinggi per unit. Maka, target pasarnya merupakan kalangan atas cluster ini dapat dikategorikan sebagai unit premium. Cluster 4, memiliki penjualan yang bagus rata-rata unit terjualnya yaitu 463 unit dan total penjualan 223175, cluster ini menunjukkan performa sedang di semua aspek tetapi margin keuntungannya paling rendah. Hal ini me-

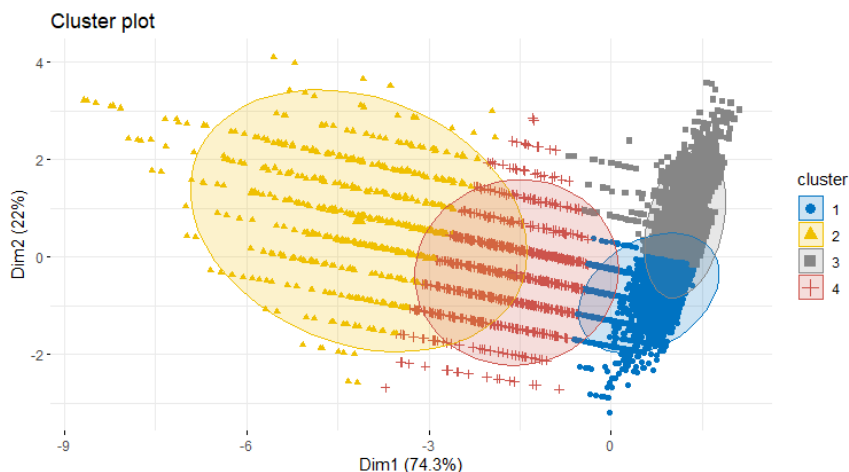
nunjukkan bahwa segmen pada kota di cluster 4 ini fokus pada volume penjualan, tapi kurang efisien dalam profit.

Gambar 5 menunjukkan hasil pengelompokan pelanggan dengan metode K-Means. Data pelanggan dibagi menjadi empat cluster berbeda, yang ditandai dengan warna merah, hijau, ungu, dan biru. Setiap cluster terdiri dari pelanggan dengan karakteristik yang mirip, misalnya pola pembelian atau kebiasaan tertentu. Cluster 2 (hijau) memiliki penyebaran data yang lebih luas, yang bisa berarti variasi perilaku pelanggan di dalamnya lebih besar dibandingkan kelompok lain, sehingga strategi pemasaran bisa disesuaikan untuk masing-masing cluster agar lebih efektif. Hasil clustering dengan metode K-Means pada data penjualan di berbagai kota di Amerika Serikat terlihat pada Tabel 5.

Hasil clustering menggunakan metode K-Means pada data penjualan di berbagai kota di Amerika Serikat menunjukkan setiap cluster memiliki kategori masing-masing. Misalnya, Kota New York dengan penjualan tinggi dan laba besar dikelompokkan dalam cluster 2, sementara kota lainnya seperti Los Angeles dan Chicago dengan volume penjualan lebih rendah masuk ke cluster lain. Hal ini membantu dalam memahami pola penjualan dan profitabilitas di setiap kota, sehingga strategi bisnis dapat disesuaikan untuk masing-masing kelompok pelanggan atau wilayah. Peninjauan segmentasi penjualan Adidas di Amerika Serikat berdasarkan kota dapat dilihat dari catatan penjualan setiap kota tersebut ke dalam 4 cluster yang terlihat pada Tabel 6.

Tabel 6. Segmentasi 7 kota besar dengan K-Means

No	Kota	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	Chicago	44	0	74	26
2	Dallas	53	25	97	41
3	Houston	59	43	85	29
4	Los Angeles	68	33	79	36
5	New York	92	69	52	3
6	Philadelphia	114	5	78	19
7	Phoenix	95	1	58	62

**Gambar 6.** Visualisasi hasil Fuzzy C-Means

Melalui hasil *clustering*, perusahaan Adidas dapat membuat strategi yang optimal untuk beberapa kota yang berbeda. Seperti Philadelphia memiliki jumlah yang sangat tinggi di *cluster 1* menunjukkan bahwa sebagian besar data dari kota ini memiliki kesamaan dengan karakteristik yang ditentukan untuk *cluster*, sehingga akan lebih bagus jika Perusahaan Adidas di Philadelphia melakukan optimasi berdasarkan karakteristik dari *cluster 1*, seperti dengan dilakukannya promosi yang besar-besaran sehingga dapat meningkatkan penjualan dan keuntungan. Pada kota Chicago banyak *store* yang masuk ke dalam *cluster 2*, maka bisa dilakukan optimasi dengan cara efisiensi biaya dan melakukan *up selling* ke beberapa barang yang memiliki nilai tinggi. Begitu juga kota-kota lainnya, Perusahaan dapat menyesuaikan cara atau strategi optimasi tergantung dari karakteristik *cluster* yang telah dikelompokkan.

3.3. Implementasi Algoritma Fuzzy C-Means

Pada penelitian ini penggunaan algoritma Fuzzy C-means untuk *clustering* sama dengan algoritma metode K-Means sebelumnya yaitu terdapat 4 *cluster*. Penentuan jumlah *cluster* dilakukan menggunakan elbow dimana diperoleh *cluster* terbaik yaitu 4. Sehingga, dengan menerapkan algoritma Fuzzy C-means pada data penjualan Adidas di Amerika Serikat dengan pembagian 4 *cluster* diperoleh nilai pusat setiap *cluster* dapat dilihat pada Tabel 7.

Berdasarkan hasil nilai rata-rata dari beberapa variabel pada data penjualan Adidas yang menjadi titik pusat setiap *cluster* tersebut dapat disimpulkan bahwa penjualan Adidas paling banyak terjual berada pada *cluster 2* selama periode tahun 2020-2021 yaitu total penjualannya 429653,6 akan tetapi margin keuntungan

atau rasio laba operasional-nya menengah yaitu hanya 0,38898, hal ini menunjukkan bahwa penjualan pada *cluster* ini memiliki biaya operasional yang tinggi. *Cluster 2* ini tetap bisa dikategorikan sebagai *cluster* dengan kinerja terbaik. *Cluster 3* dapat dikategorikan sebagai unit produk premium, dilihat dari unit terjual yang rendah dan laba operasionalnya sangat rendah yaitu 4183,2 menunjukkan biaya operasionalnya sangat tinggi dengan penjualan yang sedikit. Namun, margin keuntungannya sangat tinggi 0,513 artinya biaya produk perunitnya tinggi. *Cluster 4* merupakan unit dengan performa yang stabil untuk keempat aspek seimbang berada di kelas rata-rata, dengan total penjualan yang cukup tinggi rata-rata-nya mencapai 212496,5. *Cluster* yang paling buruk untuk semua aspek yaitu *cluster 1* dengan penjualan terburuk, penjualan unit dan laba operasional sangat rendah serta margin keuntungannya yang kecil.

Hasil *cluster* ini dapat divisualisasikan untuk melihat sebaran *clusternya* dengan lebih jelas seperti yang dapat dilihat pada Gambar 6. Melalui gambar tersebut terlihat bahwa *cluster 2* memiliki cakupan yang lebih luas karena data-datanya menyebar dibandingkan *cluster 1* dan 3 yang datanya sangat rapat. Dapat diketahui bahwa *cluster 1* memiliki sebanyak 3322 observasi, *cluster 2* sebanyak 838 observasi, *cluster 3* sebanyak 3586 observasi, dan *cluster 4* sebanyak 1902 observasi. Data penjualan Adidas ini paling banyak berada pada 2 yaitu dengan kategori penjualan dan profit rendah namun efisiensi keuangan sangat baik. Kemudian disusul dengan *cluster 1* dimana terdapat sebanyak 3322 data penjualan yang termasuk kategori penjualan buruk. Data dari masing-masing *cluster* beserta dengan nilai dari variabel yang digunakan, disajikan pada Tabel 8.

Dengan menggunakan data sebenarnya tanpa standarisasi

Tabel 7. Pusat cluster Fuzzy C-Means

	Unit Terjual	Total Penjualan	Laba Operasional	Margin Keuntungan
Cluster 1	153,8971	25944,67	9146,401	0,375148
Cluster 2	740,8587	429653,6	165219,4	0,388981
Cluster 3	131,3605	8624,622	4183,166	0,513
Cluster 4	444,5378	212496,5	74438,69	0,354854

Tabel 8. Hasil cluster Fuzzy C-Means

No	Kota	Unit Terjual	Total penjualan	Laba operasional	Margin Keuntungan	Cluster
1	Los Angeles	275	123750	43312,5	0,35	1
2	Los Angeles	275	96250	33687,5	0,35	1
3	Los Angeles	125	50000	20000	0,4	1
4	New York	1200	600000	300000	0,5	2
5	New York	1000	500000	150000	0,3	2
6	New York	1000	400000	140000	0,35	2
7	Chicago	275	123750	61875	0,5	3
8	Chicago	275	123750	61875	0,5	3
9	Chicago	225	112500	56250	0,5	3
10	New York	900	225000	78750	0,35	4
11	New York	700	245000	85750	0,35	4
12	New York	700	245000	110250	0,45	4

terlihat dengan jelas perbedaan antar cluster untuk setiap variabel. Data penjualan ini merupakan catatan harian sehingga, penjualan pada setiap kota dapat berubah setiap harinya. Oleh karena itu kota-kota di Amerika Serikat ini dapat berada di berbagai Cluster, karena hasil penjualan tiap harinya dapat berubah kadang baik kadang buruk. Selanjutnya, dilakukan perbandingan pada 7 kota besar di Amerika Serikat berdasarkan jumlah penduduk, seperti yang disajikan pada Tabel 9.

Tabel 9. Segmenetasi 7 kota besar AS dengan Fuzzy C-Means

No	Kota	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	Chicago	44	0	74	26
2	Dallas	51	27	97	41
3	Houston	59	49	85	23
4	Los Angeles	68	44	79	25
5	New York	92	69	52	3
6	Philadelphia	115	5	75	21
7	Phoenix	97	4	55	60

Peninjauan segmentasi penjualan Adidas di Amerika Serikat berdasarkan kota dapat dilihat dari catatan penjualan setiap kota lebih banyak berada di cluster 1,2,3, atau 4. Sehingga perusahaan Adidas dapat membuat strategi yang optimal untuk beberapa kota yang berbeda. Philadelphia dan Phoenix merupakan kota yang paling banyak berada di cluster 1 dibandingkan 5 kota besar lainnya hal ini menunjukkan bahwa penjualan Adidas di kota ini rendah sehingga diperlukannya strategi penjualan misalnya membuat harga produk yang lebih murah dengan memberikan diskon. New York paling banyak berada di cluster 1 akan tetapi kota ini juga mendominasi cluster 2 yang memiliki tingkat penjualan tinggi dan laba operasional yang cukup stabil, sehingga kota New York cocok sebagai pasar premium dengan meningkatkan branding produk dan strategi diskon di event tertentu. Houston dan Los Angeles banyak berada di cluster 3 namun sebagian besar juga berada di cluster 1 hal ini menunjukkan bahwa perlu dilakukan royalty program untuk mempertahankan pelanggan serta

memberikan diskon kepada pelanggan. Kemudian, kota Phoenix dan Dallas mendominasi cluster 4 yang menunjukkan penjualan tinggi tetapi margin laba sangat rendah yang berarti perlu dilakukan efisiensi biaya produksi pada produk yang dijual di kota Phoenix dan Dallas.

3.4. Pemilihan Model Terbaik

Penelitian ini dilakukan dengan dua metode cluster yaitu K-Means dan Fuzzy C-Means, untuk mengevaluasi kualitas hasil clustering dengan kedua metode ini yaitu menggunakan Davies-Bouldin Index (DBI), (Within Sum of Squares) SSW, dan Between Sum of Squares (SSB). Algoritma K-Means dan Fuzzy C-Means mempunyai nilai validitas cluster yang hampir sama, seperti yang ditunjukkan pada Tabel 10.

Tabel 10. Evaluasi hasil K-Means dan Fuzzy C-Means

Clustering	DBI	SSW	SSB
K-Means	0,98	8645,738	29942,26
Fuzzy C-Means	0,983	8678,486	28476,79

Nilai DBI dengan kedua metode ini masih cukup tinggi, nilai DBI yang lebih kecil menunjukkan bahwa kluster kompak dan terpisah dengan baik. Algoritma K-Means memiliki nilai DBI sedikit lebih rendah dibandingkan Fuzzy C-Means, Ini berarti kluster-kluster yang dihasilkan oleh K-Means cenderung lebih kompak di dalamnya dan lebih terpisah dengan baik dari kluster lain. Selain itu, K-Means memiliki nilai SSW yang lebih rendah (8645,738) dibandingkan FCM (8678,486). Nilai SSW yang lebih rendah menunjukkan bahwa titik-titik data di dalam setiap kluster lebih dekat ke pusat kluster, yang menunjukkan bahwa kluster yang dihasilkan oleh K-Means memiliki homogenitas internal yang lebih tinggi. Berdasarkan nilai SSD, K-Means juga memiliki nilai yang lebih tinggi (29942,26) dibandingkan FCM (28476,79). SSB yang lebih tinggi menunjukkan bahwa kluster-kluster ini memiliki perbedaan karakteristik yang jelas antara satu sama lain, yang membuatnya lebih mudah untuk kategori segmen yang berbeda. Al-

goritma K-Means memiliki *runtime* yang lebih cepat karena algoritmanya yang lebih sederhana cenderung membutuhkan lebih sedikit iterasi dibandingkan dengan FCM. Sehingga, K-Means lebih efisien secara komputasi. Hasil ini juga sejalan dengan penelitian sebelumnya [6, 17]. Berdasarkan hasil evaluasi serta pertimbangan komputasi, algoritma K-Means dianggap lebih optimal untuk penelitian ini. Namun, perlu dicatat bahwa Fuzzy C-Means juga memberikan hasil yang sangat baik dan menawarkan fleksibilitas dalam memahami keanggotaan klaster yang tidak mutlak, yang dapat menjadi pertimbangan penting dalam konteks lain.

4. Kesimpulan

Dari hasil kedua metode *cluster* K-Means dan Fuzzy C-Means, didapatkan kedua metode *cluster* yang terbentuk masuk dalam kategori baik, dan hasil pengclusteran memiliki karakteristik yang kurang lebih sama. Tetapi dari hasil perbandingan dengan DBI, SSW, dan SSB didapatkan metode yang paling baik adalah *clustering* dengan K-Means. Maka dapat disimpulkan, pada *cluster* 1 dimana seluruh variabel berada pada urutan ketiga yang berarti stabil tetapi tidak unggul, sehingga beberapa hal yang harus dilakukan untuk kota yang termasuk *cluster* 1 adalah mendorong pertumbuhan penjualan dengan melalui promosi dan bisa dilakukan efisiensi dalam biaya distribusi atau logistik, sehingga dapat meningkatkan *margin*. Pada *cluster* 2 seluruh penjualan, pendapatan, dan total penjualan berada di posisi pertama tetapi *margin* keuntungan berada di posisi kedua, maka kota yang termasuk segmentasi ini diperlukan untuk melakukan efisiensi biaya, agar bisa menekan angka total penjualan yang dikeluarkan dan menaikkan *margin* keuntungan, dan bisa dilakukan *upselling* untuk produk yang memiliki harga yang tinggi, agar *margin* naik secara optimal. *Cluster* 3, didapatkan penjualan terendah tetapi *margin* tertinggi, maka untuk segmentasi kota ini diperlukan strategi seperti, perkuat sebagai *brand premium*, memperluas jangkauan dari produk, dan meminimalkan biaya promosi karena customer yang berada di kota yang termasuk *cluster* 3 sudah loyal terhadap *brand*. Pada *cluster* 4 memiliki penjualan yang tinggi tetapi *margin* yang sangat rendah, maka diperlukan optimasi dengan mengecek kembali harga dan diskon pada produk, evaluasi biaya apakah biaya operasional yang dikeluarkan sangat besar, dan difokuskan untuk melakukan optimasi *margin* dengan melakukan *upselling*, *bundling* atau pengurangan diskon.

Kontribusi Penulis. Lutfiah Firlan: Konseptualisasi, metodologi, analisis formal, perangkat lunak, validasi, penulisan—persiapan draft awal. Risfa Fadila: Metodologi, analisis formal, perangkat lunak, kurasi data, visualisasi, penulisan—persiapan draft awal. Muhammad Keyvn Ridho: Konseptualisasi, sumber daya, administrasi proyek, penulisan—persiapan draft awal. Etis Sunandi: Supervisi. Ukasyah Aflah: Metodologi, investigasi, penulisan—persiapan draft awal. Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

Ucapan Terima Kasih. Para penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan selama proses penelitian ini berlangsung. Kami sangat menghargai editor dan reviewer atas masukan serta dukungannya dalam menyempurnakan karya ini.

Pembiayaan. Penelitian ini tidak menerima pendanaan eksternal.

Konflik Kepentingan. Para penulis menyatakan tidak ada konflik kepentingan yang terkait dengan artikel ini.

Ketersediaan Data. Data yang digunakan merupakan data sekunder berupa data penjualan produk Adidas di Amerika Serikat sebanyak 9652 data dengan 14 variabel. Data ini diperoleh dari *website kaggle*: <https://www.kaggle.com/datasets/>.

Referensi

- [1] S. Gea, "Pengaruh Segmentasi Pasar Terhadap Peningkatan Volume Penjualan," *J. Akuntansi, Manaj. dan Ekon.*, vol. 1, no. 1, hal. 48–54, 2022, doi: 10.56248/jamane.v1i1.12.
- [2] S. Setyaningtyas, B. I. Nugroho, dan Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, hal. 52–61, 2022, doi: 10.21063/jtif.2022.v10.2.52-61.
- [3] A. A. A. Putri dan S. A. Rahmah, "Implementasi Data Mining Dengan Algoritma K-Means Clustering Untuk Analisis Bisnis Pada Perusahaan Asuransi," *Djtechno J. Teknol. Inf.*, vol. 5, no. 1, hal. 139–152, 2024, doi: 10.46576/djtechno.v5i1.4537.
- [4] A. F. Khairati, A. Adlina, G. Hertono, dan B. Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA," *Prism. Pros. Semin. Nas. Mat.*, vol. 2, hal. 161–170, 2019.
- [5] I. W. A. W. Kusuma dan A. Kusumadewi, "Analysis of Comparison of Segmentation Results Using K-Means and Fuzzy C-Means Methods on Compressed Input Images," *Elektrika*, vol. 13, no. 2, hal. 63, 2021.
- [6] A. L. R. Putri dan N. Dwidayati, "Analisa Perbandingan K-Means dan Fuzzy C-Means dalam Pengelompokan Daerah Penyebaran Covid-19 Indonesia," *UNNES J. Math.*, vol. 10, no. 2, hal. 50–55, 2021.
- [7] H. S. Firdaus, A. L. Nugraha, B. Sasmito, dan M. Awaluddin, "Perbandingan Metode Fuzzy C-Means dan K-Means Untuk Pemetaan Daerah Rawan Kriminalitas Di Kota Semarang," *Elipsoida J. Geod. dan Geomatika*, vol. 4, no. 1, hal. 58–64, 2021, doi: 10.14710/elipsoida.2021.9219.
- [8] A.-Y. Al-Yasir, M. Afdal, Z. Zarnelly, dan A. Marsal, "Analisis Loyalitas Pelanggan Business To Business Berdasarkan Model RFM Menggunakan Algoritma Fuzzy C-Means," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, hal. 359–365, 2024, doi: 10.57152/malcom.v4i1.1163.
- [9] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, hal. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [10] D. Triyansyah dan D. Fitriana, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *J. Telekomun. dan Komput.*, vol. 8, no. 3, hal. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [11] J. Hutagalung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, hal. 606–620, 2022, doi: 10.35957/jatisi.v9i1.1516.
- [12] N. A. Kurniawati, "Pemodelan Tingkat Pengangguran Terbuka di Nonparametrik Spline Truncated," *J. Sains Dan Seni ITS*, vol. 8, no. 2, hal. 2–8, 2019.
- [13] S. Dwididanti dan D. A. Anggoro, "Analisis Perbandingan Algoritma Bisectioning K-Means dan Fuzzy C-Means pada Data Pengguna Kartu Kredit," *Emit. J. Tek. Elektro*, vol. 22, no. 2, hal. 110–117, 2022, doi: 10.23917/emitor.v22i2.15677.
- [14] H. Syukron, M. F. Fayyad, F. J. Fauzan, Y. Ikhsani, dan U. R. Gurning, "Perbandingan K-Means K-Medoids dan Fuzzy C-Means untuk Pengelompokan Data Pelanggan dengan Model LRFM," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, hal. 76–83, 2022, doi: 10.57152/malcom.v2i2.442.
- [15] A. Rizal, D. C. R. Novitasari, dan M. Hafiyusholeh, "Pengelompokan Karyawan Berdasarkan Kesalahan Menggunakan Perbandingan Fuzzy C-Means, K-Means, dan Probabilistic Distance Clustering," *J. Fourier*, vol. 11, no. 2, hal. 69–77, 2022, doi: 10.14421/fourier.2022.112.69-77.
- [16] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*. New York: Plenum Press, 1981, doi: 10.1007/978-1-4757-0450-1.
- [17] T. G. Manik, W. I. Rahayu, dan R. N. S. Fathonah, "Perbandingan Metode Fuzzy C-Means dan K-Means Clustering Pada Data Penggunaan Obat Di R.S National Hospital Surabaya," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, hal. 591–597, 2023, doi: 10.36040/jati.v7i1.6430.