

Penerapan Algoritma Fuzzy C-Means untuk Segmentasi Pelanggan Berdasarkan Profil dan Perilaku Pembelian

Eka Oktaviani¹, Rismawati², Yusfi Jauhari³

¹²³Program Studi : Sistem Informasi, Universitas Darwan Ali

Email : eka99oktavia@gmail.com¹, risma.yusmin@gmail.com², usjauhari@gmail.com³

ABSTRACT— Customer segmentation is required to understand differences in customer characteristics and support more targeted marketing strategies. This study aims to apply the Fuzzy C-Means algorithm to group customers based on their demographic characteristics and purchasing behavior. The study used the *Customer Segmentation: Clustering* dataset from Kaggle, comprising 2,240 records and 29 attributes. The research stages included data examination, missing-value handling, categorical attribute transformation, data standardization, Fuzzy C-Means clustering into three clusters, and evaluation using the Fuzzy Partition Coefficient and Silhouette Score. The clustering process produced three customer profiles: middle-tier, premium, and budget customers. The evaluation yielded a Fuzzy Partition Coefficient value of 0.4016 and a Silhouette Score of 0.1697. These values indicate that customer membership remains distributed across clusters and that the separation between clusters is relatively weak because several customer characteristics overlap. Nevertheless, the findings provide an initial overview of customer profile differences based on income and spending patterns. Further optimization of the number of clusters and feature selection is required to achieve clearer customer segmentation.

Keywords — Customer Segmentation, Fuzzy C-Means, Clustering, Fuzzy Partition Coefficient, Silhouette Score.

ABSTRAK— Segmentasi pelanggan diperlukan untuk memahami perbedaan karakteristik pelanggan sehingga strategi pemasaran dapat disusun secara lebih tepat. Penelitian ini bertujuan menerapkan algoritma Fuzzy C-Means untuk mengelompokkan pelanggan berdasarkan karakteristik demografi dan perilaku pembelian. Dataset yang digunakan adalah *Customer Segmentation: Clustering* dari Kaggle, yang terdiri atas 2.240 data dan 29 atribut. Tahapan penelitian meliputi pemeriksaan kondisi data, penanganan nilai kosong, transformasi atribut kategorikal, standarisasi data, penerapan Fuzzy C-Means dengan tiga kluster, serta evaluasi menggunakan Fuzzy Partition Coefficient dan Silhouette Score. Hasil pengelompokan menghasilkan tiga profil pelanggan, yaitu pelanggan menengah, pelanggan premium, dan pelanggan hemat. Evaluasi menghasilkan nilai Fuzzy Partition Coefficient sebesar 0,4016 dan Silhouette Score sebesar 0,1697. Nilai tersebut menunjukkan bahwa derajat keanggotaan pelanggan masih tersebar dan pemisahan antarkluster relatif lemah karena sebagian karakteristik pelanggan saling beririsan. Meskipun demikian, hasil penelitian dapat memberikan gambaran awal mengenai perbedaan profil pelanggan berdasarkan pendapatan dan pola pengeluaran. Optimalisasi jumlah kluster dan pemilihan fitur masih diperlukan untuk memperoleh segmentasi yang lebih jelas.

Kata kunci— Segmentasi Pelanggan, Fuzzy C-Means, *Clustering*, Fuzzy Partition Coefficient, Silhouette Score.

I. PENDAHULUAN

Data pelanggan merupakan salah satu sumber informasi penting dalam mendukung pengambilan keputusan pemasaran berbasis data. Melalui data pelanggan, organisasi dapat memahami karakteristik demografi, perilaku belanja, saluran pembelian yang digunakan, serta respons pelanggan terhadap kampanye pemasaran. Apabila seluruh pelanggan diperlakukan dengan strategi yang sama, perusahaan berisiko memberikan penawaran yang kurang relevan karena setiap pelanggan memiliki kebutuhan, nilai, dan pola konsumsi yang berbeda. Oleh karena itu, segmentasi pelanggan diperlukan untuk membentuk kelompok pelanggan yang memiliki kemiripan karakteristik sehingga strategi pemasaran dapat diarahkan secara lebih tepat.

Segmentasi pelanggan dapat dilakukan menggunakan pendekatan data mining, salah satunya melalui metode clustering. Clustering merupakan teknik pembelajaran tanpa pengawasan yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan

antarobjek tanpa memerlukan label kelas sebagai target [1]. Dalam konteks pemasaran, clustering dapat membantu menemukan kelompok pelanggan bernilai tinggi, pelanggan dengan aktivitas belanja sedang, dan pelanggan dengan aktivitas belanja rendah berdasarkan pola yang muncul dari data. Hasil segmentasi tersebut dapat digunakan sebagai dasar dalam menyusun strategi promosi, mempertahankan pelanggan potensial, serta meningkatkan efektivitas kampanye pemasaran.

Penelitian ini menggunakan algoritma Fuzzy C-Means (FCM) untuk melakukan segmentasi pelanggan. FCM memiliki karakteristik keanggotaan lunak, yaitu setiap data pelanggan dapat memiliki derajat keanggotaan pada lebih dari satu kluster. Karakteristik ini sesuai dengan kondisi perilaku pelanggan yang tidak selalu dapat dipisahkan secara tegas ke dalam satu kelompok tertentu. Sebagai contoh, seorang pelanggan dapat memiliki pendapatan tinggi, tetapi tingkat pengeluaran sedang, sehingga posisinya dapat berada di antara dua segmen pelanggan. Dengan pendekatan fuzzy, proses segmentasi dapat memberikan gambaran yang lebih fleksibel dibandingkan metode hard clustering yang hanya

menetapkan satu data ke dalam satu kluster secara mutlak [2], [3].

Dataset yang digunakan dalam penelitian ini adalah dataset segmentasi pelanggan yang terdiri dari 2.240 data pelanggan dengan 29 atribut. Atribut tersebut mencakup informasi demografi, seperti tahun lahir, pendidikan, status pernikahan, dan pendapatan, serta informasi perilaku belanja, seperti recency, pengeluaran pada beberapa kategori produk, jumlah pembelian melalui berbagai kanal, penerimaan kampanye, komplain, dan respons pelanggan. Kombinasi atribut demografi dan perilaku belanja menjadikan dataset ini relevan untuk dianalisis menggunakan metode clustering karena mampu menggambarkan profil pelanggan dari beberapa sudut pandang [4].

Tahapan penelitian dilakukan melalui pemeriksaan kualitas data, pembersihan data, transformasi atribut, standarisasi data numerik, penerapan algoritma Fuzzy C-Means, serta evaluasi hasil kluster. Jumlah kluster yang digunakan dalam penelitian ini ditetapkan sebanyak tiga kluster dengan nilai fuzziness sebesar $m = 2$. Selain implementasi menggunakan Python, penelitian ini juga menyajikan perhitungan manual dua iterasi menggunakan sampel data agar proses pembentukan pusat kluster dan pembaruan derajat keanggotaan dapat dipahami secara bertahap.

Hasil implementasi Fuzzy C-Means menghasilkan tiga kelompok pelanggan. Cluster 0 terdiri dari 391 pelanggan, Cluster 1 terdiri dari 705 pelanggan, dan Cluster 2 terdiri dari 1.144 pelanggan. Evaluasi model menghasilkan nilai Fuzzy Partition Coefficient (FPC) sebesar 0,4016 dan Silhouette Score sebesar 0,1697. Nilai tersebut menunjukkan bahwa hasil segmentasi masih memiliki batas antarkluster yang saling tumpang tindih, sehingga interpretasi hasil perlu dilakukan secara hati-hati. Meskipun demikian, hasil segmentasi tetap dapat digunakan untuk menggambarkan kecenderungan profil pelanggan berdasarkan karakteristik demografi dan perilaku belanja.

Penelitian ini bertujuan untuk menerapkan algoritma Fuzzy C-Means dalam membentuk segmentasi pelanggan berdasarkan data demografi dan perilaku belanja, menganalisis karakteristik setiap kluster, serta mengevaluasi kualitas hasil clustering menggunakan FPC dan Silhouette Score. Hasil penelitian diharapkan dapat memberikan gambaran awal mengenai pemanfaatan Fuzzy C-Means dalam analisis segmentasi pelanggan dan menjadi dasar dalam penyusunan strategi pemasaran yang lebih terarah.

II. METODOLOGI PENELITIAN

A. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk melakukan segmentasi pelanggan menggunakan algoritma Fuzzy C-Means. Metode ini digunakan karena mampu mengelompokkan data berdasarkan tingkat kemiripan, tetapi tetap memberikan derajat keanggotaan pada setiap data terhadap masing-masing kluster. Dengan pendekatan tersebut, pelanggan tidak hanya dikelompokkan secara

tegas ke dalam satu kluster, tetapi juga dapat diketahui tingkat kedekatannya terhadap kluster lain.

B. Dataset dan Sumber Data

Dataset yang digunakan dalam penelitian ini adalah dataset publik Customer Segmentation: Clustering yang diperoleh dari Kaggle [4]. Dataset ini berisi data pelanggan yang dapat digunakan untuk analisis segmentasi, karena memuat informasi demografi, perilaku pembelian, respons terhadap kampanye pemasaran, serta atribut lain yang berkaitan dengan karakteristik pelanggan.

Dataset terdiri dari 2.240 data pelanggan dan 29 atribut. Beberapa atribut yang digunakan antara lain Year_Birth, Education, Marital_Status, Income, Kidhome, Teenhome, Recency, jumlah pengeluaran pada beberapa kategori produk, jumlah pembelian melalui berbagai kanal, respons kampanye, komplain, dan status respons pelanggan. Kombinasi atribut tersebut memungkinkan proses clustering dilakukan untuk menemukan kelompok pelanggan yang memiliki kemiripan karakteristik.

Dalam penelitian ini, dataset tidak digunakan untuk menerapkan algoritma K-Means sebagaimana konteks umum pada sumber dataset, tetapi diolah kembali menggunakan algoritma Fuzzy C-Means. Hal ini dilakukan karena Fuzzy C-Means mampu memberikan derajat keanggotaan pada setiap pelanggan terhadap masing-masing kluster, sehingga hasil segmentasi dapat menggambarkan hubungan pelanggan dengan kluster secara lebih fleksibel.

C. Teknik Preprocessing Data

Preprocessing data dilakukan sebelum proses clustering agar data siap dianalisis. Tahapan ini meliputi pemeriksaan data, pembersihan data, transformasi data, dan standarisasi data. Preprocessing penting dilakukan karena kualitas data sangat memengaruhi hasil analisis data mining [5].

Pemeriksaan data dilakukan untuk mengetahui jumlah data, jumlah atribut, tipe data, missing value, dan data duplikat. Selanjutnya, data dibersihkan dengan menangani nilai kosong dan menghapus atribut yang tidak digunakan dalam proses clustering, seperti atribut identitas pelanggan.

Atribut kategorikal, seperti pendidikan dan status pernikahan, diubah menjadi bentuk numerik melalui proses encoding. Setelah itu, data numerik distandarisasi agar setiap atribut memiliki skala yang seimbang. Standarisasi diperlukan karena atribut seperti pendapatan memiliki rentang nilai yang lebih besar dibandingkan atribut jumlah pembelian atau respons kampanye.

D. Metrik Evaluasi

Evaluasi hasil clustering dilakukan menggunakan Fuzzy Partition Coefficient (FPC) dan Silhouette Score. FPC digunakan untuk mengukur tingkat kejelasan derajat keanggotaan data pada kluster fuzzy. Nilai FPC berada pada rentang 0 sampai 1. Semakin mendekati 1, maka hasil clustering menunjukkan keanggotaan data yang semakin jelas terhadap kluster tertentu [6].

Rumus FPC dapat ditulis sebagai berikut.

$$FPC = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^c (u_{ij})^2$$

Silhouette Score digunakan untuk mengukur kualitas pemisahan antar kluster. Nilai Silhouette Score yang semakin tinggi menunjukkan bahwa data dalam satu kluster memiliki kemiripan yang baik dan berbeda dari kluster lain [7].

Rumus Silhouette Score dapat ditulis sebagai berikut.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

E. Prosedur Analisis

Prosedur analisis dilakukan secara bertahap mulai dari dataset pelanggan sampai penarikan kesimpulan.

Tahap pertama adalah membaca dataset pelanggan. Setelah itu dilakukan pemeriksaan data untuk mengetahui struktur dan kualitas awal dataset. Tahap berikutnya adalah pembersihan data, yaitu menangani missing value, duplikasi, dan atribut yang tidak diperlukan.

Setelah data bersih, dilakukan transformasi data agar atribut kategorikal dapat diproses oleh algoritma. Data kemudian distandardisasi agar seluruh atribut memiliki skala yang seimbang. Selanjutnya, algoritma Fuzzy C-Means diterapkan dengan jumlah kluster yang telah ditentukan.

Hasil clustering kemudian dievaluasi menggunakan FPC dan Silhouette Score. Setelah evaluasi dilakukan, setiap kluster diinterpretasikan berdasarkan karakteristik pelanggan, seperti pendapatan, total pengeluaran, jumlah pembelian, recency, dan respons kampanye. Tahap akhir adalah menarik kesimpulan berdasarkan hasil segmentasi pelanggan.

III. HASIL DAN PEMBAHASAN

A. Kondisi Awal dan Preprocessing Data

Dataset `customer_segmentation.csv` terdiri atas 2.240 data pelanggan dan 29 atribut. Atribut tersebut mencakup informasi demografi, pendapatan, jumlah anak, waktu pembelian terakhir, pengeluaran produk, kanal pembelian, penerimaan kampanye, komplain, dan respons pelanggan.

Pemeriksaan awal menemukan 24 nilai kosong pada atribut `Income`. Tidak ditemukan baris yang terduplikasi secara identik. Selain itu, atribut `Z_CostContact` dan `Z_Revenue` memiliki nilai konstan masing-masing 3 dan 11 sehingga tidak memberikan variasi untuk membedakan pelanggan.

Nilai maksimum `Income` mencapai 666.666, sedangkan nilai minimum `Year_Birth` adalah 1893. Kedua nilai tersebut menunjukkan adanya kemungkinan outlier atau data yang tidak realistis. Dalam implementasi yang digunakan pada laporan, nilai tersebut tetap dipertahankan, tetapi dicatat sebagai keterbatasan analisis.

Nilai kosong pada `Income` diisi menggunakan median karena median lebih tahan terhadap pengaruh nilai ekstrem dibandingkan rata-rata. Atribut `ID` tidak digunakan karena hanya berfungsi sebagai penanda pelanggan. Atribut kategorikal seperti `Education` dan

`Marital_Status` digunakan untuk membantu interpretasi profil, sedangkan proses Fuzzy C-Means menggunakan matriks numerik yang telah distandardisasi.

Penelitian ini menggunakan dataset pelanggan sebanyak 2.240 data dengan 29 atribut. Sebelum proses clustering dilakukan, data terlebih dahulu diperiksa untuk mengetahui kondisi awal dataset. Hasil pemeriksaan menunjukkan bahwa terdapat nilai kosong pada atribut `Income`, sedangkan atribut lain dapat digunakan setelah melalui proses transformasi dan standardisasi. Nilai kosong tersebut ditangani agar tidak mengganggu proses pembentukan kluster.

Tabel I
RINGKASAN DATASET DAN HASIL PREPROCESSING

Informasi	Kondisi Awal	Setelah Preprocessing
Jumlah data	2.240	2.240
Jumlah atribut awal	29	26 fitur pada matriks hasil notebook
Missing value	24 pada <code>Income</code>	0
Data duplikat	0	0
Atribut ID	Tersedia	Tidak digunakan

Hasil preprocessing menunjukkan bahwa seluruh pelanggan dapat dipertahankan karena missing value ditangani melalui imputasi. Standardisasi dilakukan menggunakan `StandardScaler` agar fitur berukuran besar, seperti `Income` dan pengeluaran produk, tidak mendominasi perhitungan jarak.

Matriks yang ditampilkan dalam laporan memiliki ukuran 2.240×26 . Namun, laporan juga mencatat kemungkinan adanya kolom label `Cluster` dari proses sebelumnya yang masih masuk ke matriks kerja. Hal ini perlu diperiksa kembali karena label hasil metode lain seharusnya tidak digunakan sebagai fitur dalam Fuzzy C-Means.

Histogram `Income` menunjukkan distribusi yang tidak sepenuhnya merata dan memiliki nilai ekstrem pada bagian kanan. Distribusi `Recency` berada pada rentang 0 sampai 99 hari dengan rata-rata sekitar 49,11 hari. Sementara itu, `Total_Spending` menunjukkan perbedaan pengeluaran yang lebar karena sebagian pelanggan memiliki transaksi jauh lebih tinggi daripada pelanggan lainnya. Kondisi tersebut mendukung penggunaan standardisasi sebelum clustering dilakukan.

Setelah proses preprocessing, atribut kategorikal seperti pendidikan dan status pernikahan diubah menjadi bentuk numerik. Atribut numerik kemudian distandardisasi agar seluruh fitur memiliki skala yang seimbang. Standardisasi penting karena atribut seperti pendapatan memiliki rentang nilai yang lebih besar

dibandingkan atribut jumlah pembelian atau respons kampanye. Jika tidak distandardisasi, atribut dengan nilai besar dapat lebih dominan dalam proses clustering.

Penerapan algoritma Fuzzy C-Means dilakukan dengan jumlah kluster sebanyak tiga. Hasil clustering menunjukkan bahwa data pelanggan terbagi ke dalam tiga kelompok dengan jumlah anggota yang berbeda. Cluster 0 terdiri dari 391 pelanggan, Cluster 1 terdiri dari 705 pelanggan, dan Cluster 2 terdiri dari 1.144 pelanggan.

B. Penerapan Fuzzy C-Means

Fuzzy C-Means diterapkan menggunakan tiga kluster dengan parameter fuzziness $m=2$. Nilai error ditetapkan sebesar 0,005 dan jumlah iterasi maksimum sebesar 1.000. Matriks data ditransposisikan menjadi X_{scaled} untuk mengikuti format masukan pustaka scikit-fuzzy, yaitu fitur sebagai baris dan sampel sebagai kolom.

Tabel II
PARAMETER DAN HASIL EVALUASI FUZZY C-MEANS

Parameter atau hasil	Nilai	Keterangan
Jumlah kluster (c)	3	Kluster yang dibentuk
Fuzziness (m)	2	Tingkat keaburan keanggotaan
Error	0,005	Batas perubahan untuk berhenti
Iterasi maksimum	1.000	Batas maksimum proses
Inisialisasi	Acak	init=None
FPC	0,4016	Keanggotaan masih menyebar
Silhouette Score	0,1697	Pemisahan kluster lemah
Jumlah kluster (c)	3	Kluster yang dibentuk

Dalam perhitungan manual menggunakan lima pelanggan, fungsi objektif turun dari 1,9252 pada iterasi pertama menjadi 1,3465 pada iterasi kedua. Penurunannya dapat dihitung sebagai berikut.

$$\frac{1,9252 - 1,3465}{1,9252} \times 100\% = 30,06\%$$

Penurunan sekitar 30,06% menunjukkan bahwa pembaruan pusat kluster dan derajat keanggotaan berhasil memperkecil jarak berbobot antara data dan pusat kluster. Walaupun contoh manual hanya menggunakan dua iterasi,

hasil tersebut telah memperlihatkan cara kerja proses optimasi pada Fuzzy C-Means [8].

C. Hasil Pembentukan Kluster

Label akhir pelanggan ditentukan berdasarkan nilai keanggotaan tertinggi pada setiap pelanggan. Hasil implementasi menghasilkan tiga kelompok dengan ukuran yang berbeda. Cluster 0 terdiri atas 391 pelanggan atau 17,46%, Cluster 1 sebanyak 705 pelanggan atau 31,47%, dan Cluster 2 sebanyak 1.144 pelanggan atau 51,07%.

Cluster 2 merupakan kelompok terbesar dan mencakup lebih dari separuh jumlah pelanggan. Kondisi ini menunjukkan bahwa mayoritas pelanggan memiliki karakteristik yang relatif dekat dengan kelompok berpendapatan dan pengeluaran rendah. Cluster 1 menempati posisi kedua, sedangkan Cluster 0 merupakan kelompok terkecil.

D. Profil Setiap Kluster

Nilai rata-rata Recency ketiga kluster berada pada kisaran 48–49 hari. Perbedaannya relatif kecil sehingga Recency bukan pembeda utama dalam hasil clustering. Pembeda yang lebih kuat terlihat pada pendapatan, total pengeluaran, jenis produk yang dibeli, dan kanal pembelian.

Tabel III
PROFIL DAN INTERPRETASI SETIAP KLUSTER

Kluster	Jumlah, n (%)	Income / Total Spending	Interpretasi
0	391 (17,46 %)	58.016,48 / 711,52	Pelanggan menengah
1	705 (31,47 %)	74.905,17 / 1.331,05	Pelanggan premium
2	1.144 (51,07 %)	36.294,13 / 122,72	Pelanggan hemat

1) Cluster 0—Pelanggan Menengah

Cluster 0 terdiri atas 391 pelanggan dengan rata-rata pendapatan sebesar 58.016,48 dan total pengeluaran sebesar 711,52. Pengeluaran terbesar berasal dari produk wine dengan rata-rata 465,95, diikuti produk daging sebesar 122,24.

Perbandingan rata-rata pendapatan dan total pengeluaran digunakan untuk melihat perbedaan tingkat ekonomi dan aktivitas belanja setiap kluster.

Pelanggan pada kelompok ini melakukan pembelian melalui toko rata-rata 7,75 kali. Nilai Teenhome sebesar 0,94 merupakan yang tertinggi di antara ketiga kluster. Kelompok ini juga lebih sering menggunakan pembelian berbasis diskon dibandingkan Cluster 1 dan Cluster 2.

Berdasarkan karakteristik tersebut, Cluster 0 dapat diinterpretasikan sebagai pelanggan menengah. Strategi yang dapat digunakan adalah bundling produk wine,

program loyalitas toko, dan pemberian diskon terarah berdasarkan riwayat pembelian.

2) Cluster 1—Pelanggan Premium

Cluster 1 terdiri atas 705 pelanggan dengan rata-rata pendapatan tertinggi, yaitu 74.905,17. Total pengeluarannya juga paling tinggi, yaitu 1.331,05. Pengeluaran wine mencapai 611,10, sedangkan pengeluaran produk daging mencapai 417,42.

Kelompok ini memiliki rata-rata pembelian melalui katalog sebesar 5,78 kali dan pembelian melalui toko sebesar 8,56 kali. Kedua nilai tersebut merupakan yang tertinggi dibandingkan kluster lainnya. Nilai Response sebesar 0,2596 menunjukkan bahwa pelanggan Cluster 1 juga lebih responsif terhadap kampanye pemasaran.

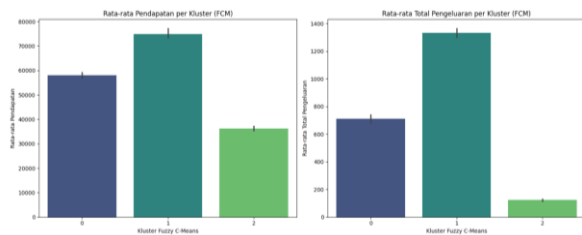
Cluster 1 dapat dikategorikan sebagai pelanggan premium atau pelanggan bernilai tinggi. Kelompok ini dapat diprioritaskan melalui penawaran eksklusif, rekomendasi produk bernilai tinggi, akses awal terhadap produk baru, dan program retensi pelanggan.

3) Cluster 2—Pelanggan Hemat

Cluster 2 beranggotakan 1.144 pelanggan dan merupakan kelompok terbesar. Rata-rata pendapatannya sebesar 36.294,13, sedangkan total pengeluarannya hanya 122,72. Kedua nilai tersebut merupakan yang terendah di antara ketiga kluster.

Nilai Kidhome pada kelompok ini mencapai 0,78 dan merupakan yang tertinggi. Rata-rata pembelian melalui web hanya 2,44 kali, pembelian katalog 0,63 kali, dan pembelian toko 3,41 kali. Meskipun kunjungan web mencapai 6,51 kali per bulan, jumlah pembeliannya masih rendah.

Cluster 2 dapat dikategorikan sebagai pelanggan hemat. Strategi yang sesuai mencakup paket produk keluarga, kupon dengan nilai terjangkau, promosi berbasis harga, dan perbaikan konversi kunjungan web menjadi transaksi.

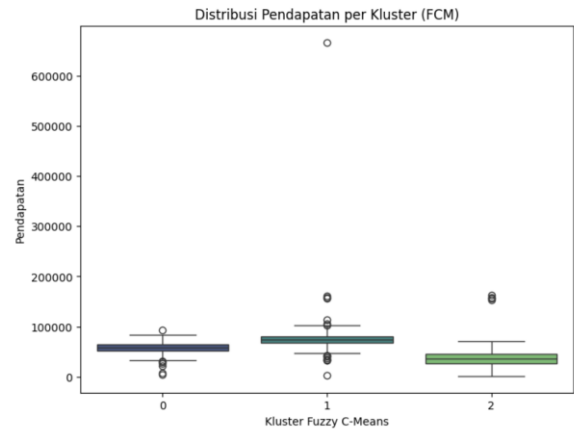


Gambar 1. Perbandingan Rata-Rata Pendapatan dan Total Pengeluaran Setiap Kluster

Gambar 1 menunjukkan bahwa Cluster 1 memiliki rata-rata pendapatan dan total pengeluaran tertinggi sehingga dikategorikan sebagai pelanggan premium. Cluster 0 berada pada tingkat menengah, sedangkan Cluster 2 memiliki rata-rata pendapatan dan pengeluaran terendah sehingga dikategorikan sebagai pelanggan hemat. Hasil ini menunjukkan bahwa pendapatan dan total pengeluaran merupakan pembeda utama karakteristik ketiga kluster.

E. Distribusi Pendapatan Setiap Kluster

Boxplot digunakan untuk melihat median, rentang data, dan outlier pendapatan pada setiap kluster.



Gambar 2. Distribusi Pendapatan pada Setiap Kluster

Boxplot menunjukkan bahwa Cluster 1 memiliki pendapatan yang cenderung lebih tinggi dibandingkan Cluster 0 dan Cluster 2. Namun, terdapat outlier Income yang mendekati 666.666 pada Cluster 1. Nilai ekstrem tersebut dapat meningkatkan rata-rata pendapatan Cluster 1 dan memengaruhi posisi pusat kluster.

Oleh karena itu, rata-rata pendapatan tidak boleh ditafsirkan sebagai nilai yang mewakili seluruh pelanggan Cluster 1. Pengembangan selanjutnya dapat mempertimbangkan validasi data, winsorization, RobustScaler, atau penghapusan nilai ekstrem berdasarkan aturan bisnis.

F. Perbandingan Karakteristik Kluster

Perbandingan profil kluster dilakukan berdasarkan nilai rata-rata atribut Income, Total_Spending, jumlah pembelian, Recency, dan respons kampanye pada setiap kluster. Hasil perbandingan menunjukkan bahwa Cluster 1 memiliki nilai tertinggi pada sebagian besar indikator ekonomi dan transaksi. Cluster 0 berada pada tingkat menengah, sedangkan Cluster 2 memiliki nilai terendah. Perbedaan tersebut terutama terlihat pada atribut Income, MntWines, MntMeatProducts, pembelian melalui katalog, dan pembelian di toko.

Seluruh kategori produk memperlihatkan pola yang serupa. Cluster 1 memiliki rata-rata pengeluaran tertinggi, diikuti oleh Cluster 0, sedangkan Cluster 2 memiliki rata-rata pengeluaran terendah. Produk wine menjadi komponen pengeluaran terbesar pada Cluster 0 dan Cluster 1.

Atribut Education dan Marital_Status tidak menjadi pembeda utama antarkluster karena ketiga kluster memiliki modus yang sama, yaitu Graduation untuk pendidikan dan Married untuk status pernikahan. Dengan demikian, perbedaan profil kluster lebih banyak dipengaruhi oleh pendapatan dan perilaku transaksi daripada karakteristik pendidikan dan status pernikahan.

G. Evaluasi Kualitas Clustering

Nilai Fuzzy Partition Coefficient yang diperoleh adalah 0,4016. Untuk tiga kluster, batas minimum teoretis FPC mendekati 1/3 atau 0,333. Nilai 0,4016 hanya sedikit lebih tinggi daripada batas tersebut. Hal ini menunjukkan bahwa derajat keanggotaan sebagian pelanggan masih tersebar pada beberapa kluster dan belum terkonsentrasi secara tegas pada satu kelompok.

Silhouette Score sebesar 0,1697 bernilai positif, tetapi dekat dengan nol. Nilai tersebut menunjukkan bahwa sebagian data berada di dekat batas antar-kluster dan pemisahan ketiga kelompok masih lemah [7]. Hasil ini sejalan dengan karakteristik Fuzzy C-Means yang mengakomodasi pelanggan dengan profil yang saling beririsan.

Hasil clustering masih dapat digunakan untuk eksplorasi segmentasi dan penyusunan strategi pemasaran awal. Namun, hasilnya belum cukup kuat untuk digunakan sebagai satu-satunya dasar keputusan otomatis, terutama untuk keputusan yang memiliki risiko tinggi.

Jumlah kluster dalam implementasi langsung ditetapkan sebanyak tiga. Oleh karena itu, hasil penelitian belum membuktikan bahwa tiga merupakan jumlah kluster paling optimal. Pengujian selanjutnya perlu membandingkan $c=2$ sampai $c=6$ menggunakan FPC, Silhouette Score, Partition Entropy, atau Xie-Beni Index[9], [10].

H. Implikasi Hasil Segmentasi

Hasil penelitian menunjukkan bahwa pendapatan dan pengeluaran menjadi pembeda paling jelas di antara ketiga kluster. Cluster 1 merupakan pelanggan premium yang layak dipertahankan melalui program loyalitas dan layanan eksklusif. Cluster 0 merupakan pelanggan menengah yang memiliki peluang untuk ditingkatkan melalui promosi dan bundling produk. Cluster 2 merupakan pelanggan hemat yang memerlukan pendekatan berbasis harga dan peningkatan konversi pembelian.

Fuzzy C-Means memberikan kelebihan karena pelanggan tidak dibatasi secara mutlak pada satu segmen. Pelanggan dengan derajat keanggotaan yang hampir seimbang pada dua kluster dapat diberikan strategi transisi. Sebagai contoh, pelanggan Cluster 0 yang memiliki kedekatan tinggi dengan Cluster 1 dapat menjadi target peningkatan menuju segmen premium.

Secara keseluruhan, penerapan Fuzzy C-Means berhasil membentuk tiga profil pelanggan, yaitu pelanggan menengah, pelanggan premium, dan pelanggan hemat. Walaupun kualitas pemisahan kluster masih rendah, hasil tersebut tetap memberikan gambaran awal mengenai perbedaan daya beli, pengeluaran, kanal pembelian, dan respons pelanggan terhadap kampanye.

IV. KESIMPULAN

A. Kesimpulan

Penerapan metode Fuzzy C-Means pada 2.240 data pelanggan berhasil membentuk tiga segmen berdasarkan karakteristik pendapatan, pengeluaran, pola pembelian, dan respons terhadap kampanye pemasaran. Tahap preprocessing menghasilkan data tanpa nilai kosong melalui imputasi median pada 24 data Income, penghapusan atribut identitas, serta standarisasi fitur agar perbedaan skala tidak mendominasi proses clustering.

Hasil segmentasi membentuk Cluster 0 sebanyak 391 pelanggan atau 17,46% sebagai pelanggan menengah, Cluster 1 sebanyak 705 pelanggan atau 31,47% sebagai pelanggan premium, dan Cluster 2 sebanyak 1.144

pelanggan atau 51,07% sebagai pelanggan hemat. Cluster 1 memiliki rata-rata pendapatan dan total pengeluaran tertinggi, Cluster 0 berada pada tingkat menengah, sedangkan Cluster 2 memiliki pendapatan dan aktivitas pembelian paling rendah. Pendapatan, total pengeluaran, jenis produk yang dibeli, serta kanal pembelian menjadi pembeda utama antarkluster, sedangkan Recency, pendidikan, dan status pernikahan tidak menunjukkan perbedaan yang kuat.

Kualitas clustering memperoleh nilai FPC sebesar 0,4016 dan Silhouette Score sebesar 0,1697. Nilai tersebut menunjukkan bahwa keanggotaan pelanggan masih cukup tersebar dan pemisahan ketiga kluster relatif lemah. Meskipun demikian, hasil segmentasi dapat digunakan sebagai gambaran awal untuk menyusun strategi pemasaran yang berbeda bagi pelanggan premium, menengah, dan hemat. Karakteristik fuzzy juga memungkinkan pelanggan yang berada di antara dua segmen dikenali melalui derajat keanggotaannya.

B. Saran

Penelitian selanjutnya disarankan membandingkan beberapa jumlah kluster, misalnya $c=2$ sampai $c=6$, menggunakan FPC, Silhouette Score, Partition Entropy, dan Xie-Beni Index agar jumlah kluster dipilih berdasarkan hasil evaluasi yang lebih kuat.

Nilai ekstrem pada Income dan Year_Birth perlu diperiksa berdasarkan aturan bisnis sebelum clustering dilakukan. Penanganannya dapat menggunakan penghapusan data tidak realistis, winsorization, atau RobustScaler agar pusat kluster tidak terlalu dipengaruhi oleh outlier. Matriks fitur juga perlu diperiksa untuk memastikan bahwa atribut identitas, atribut konstan, dan label kluster dari proses sebelumnya tidak ikut digunakan sebagai masukan.

Hasil segmentasi sebaiknya divalidasi bersama pihak yang memahami karakteristik pelanggan. Selanjutnya, derajat keanggotaan Fuzzy C-Means dapat dimanfaatkan untuk menyusun promosi transisi, terutama bagi pelanggan menengah yang memiliki kedekatan tinggi dengan segmen premium. Efektivitas strategi tersebut perlu dievaluasi menggunakan indikator seperti peningkatan jumlah transaksi, total pengeluaran, tingkat respons kampanye, dan retensi pelanggan.

V. REFERENSI

- [1] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.
- [2] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973, doi: 10.1080/01969727308546046.
- [3] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, NY, USA: Plenum Press, 1981, doi: 10.1007/978-1-4757-0450-1.
- [4] Vishakh, "Customer Segmentation: Clustering," *Kaggle*. [Online]. Available:

<https://www.kaggle.com/datasets/vishakhdapat/customer-segmentation-clustering>. [Accessed: Jul. 23, 2026].

- [5] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*. Cham, Switzerland: Springer, 2015, doi: 10.1007/978-3-319-10247-4.
- [6] J. C. Bezdek, "Cluster validity with fuzzy sets," *Journal of Cybernetics*, vol. 3, no. 3, pp. 58–73, 1973, doi: 10.1080/01969727308546047.
- [7] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.
- [8] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2–3, pp. 191–203, 1984, doi: 10.1016/0098-3004(84)90020-7.
- [9] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841–847, Aug. 1991, doi: 10.1109/34.85677.
- [10] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognition*, vol. 46, no. 1, pp. 243–256, Jan. 2013, doi: 10.1016/j.patcog.2012.07.021.
- [11] M. Wedel and W. A. Kamakura, *Market Segmentation: Conceptual and Methodological Foundations*, 2nd ed. Boston, MA, USA: Kluwer Academic Publishers, 2000, doi: 10.1007/978-1-4615-4651-1.
- [12] M.-S. Yang, "A survey of fuzzy clustering," *Mathematical and Computer Modelling*, vol. 18, no. 11, pp. 1–16, 1993, doi: 10.1016/0895-7177(93)90202-A.
- [13] G. W. Milligan and M. C. Cooper, "A study of standardization of variables in cluster analysis," *Journal of Classification*, vol. 5, pp. 181–204, 1988, doi: 10.1007/BF01897163.
- [14] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *Journal of Intelligent Information Systems*, vol. 17, no. 2–3, pp. 107–145, 2001, doi: 10.1023/A:1012801612483.
- [15] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis*, 5th ed. Chichester, UK: John Wiley & Sons, 2011, doi: 10.1002/9780470977811.
- [16] A. Banerjee and R. N. Davé, "Robust clustering," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 29–59, 2012, doi: 10.1002/widm.49.