

Penerapan Metode K-Means Clustering Dalam Analisis Profil Konsumen Untuk Strategi Manajemen Pemasaran

Wira Apriani¹, Yuda Perwira, Refin Herissandi³, Septian Jordan Purba⁴

^{1,2} Teknik Informatika, STMIK Pelita Nusantara

Email: ¹wiraapriani@gmail.com, ²yudaperwira@gmail.com, ³refintarigan2910@gmail.com

⁴septianjordanpurba63@gmail.com

Email Penulis Korespondensi: wiraapriani@gmail.com

Abstrak

Segmentasi konsumen merupakan pendekatan penting dalam manajemen pemasaran untuk memahami karakteristik pelanggan dan merumuskan strategi yang tepat sasaran. Namun, tidak semua organisasi memiliki data transaksi yang lengkap, sehingga pemanfaatan data profil pelanggan non-transaksional menjadi alternatif yang relevan. Penelitian ini bertujuan untuk melakukan segmentasi konsumen berdasarkan profil pelanggan menggunakan metode K-Means Clustering sebagai dasar penyusunan strategi manajemen pemasaran. Penelitian ini menggunakan pendekatan kuantitatif dengan teknik data mining. Dataset yang digunakan adalah dataset publik Customers Purchase Behavior dari Kaggle yang terdiri dari 72.637 data pelanggan. Variabel yang dianalisis meliputi lifestage dan premium customer. Data kategorik ditransformasikan melalui proses encoding dan dinormalisasi sebelum diterapkan algoritma K-Means Clustering. Hasil penelitian menunjukkan bahwa konsumen dapat dikelompokkan ke dalam tiga segmen utama dengan karakteristik yang berbeda berdasarkan tahapan kehidupan dan preferensi harga. Segmentasi ini mencerminkan perbedaan sensitivitas harga dan potensi nilai pelanggan pada masing-masing kelompok. Penelitian ini menyimpulkan bahwa metode K-Means Clustering efektif digunakan untuk segmentasi konsumen berbasis data profil pelanggan non-transaksional. Hasil segmentasi dapat dimanfaatkan sebagai dasar pengambilan keputusan dalam penyusunan strategi manajemen pemasaran yang lebih terarah dan berbasis data.

Kata Kunci: Segmentasi Konsumen, Data Mining, K-Means Clustering, Profil Pelanggan, Manajemen Pemasaran.

Abstract

Customer segmentation is an important approach in marketing management to understand customer characteristics and to formulate targeted marketing strategies. However, not all organizations have complete transactional data; therefore, the utilization of non-transactional customer profile data becomes a relevant alternative. This study aims to segment customers based on profile characteristics using the K-Means Clustering method as a basis for developing marketing management strategies. This research employs a quantitative approach using data mining techniques. The dataset used is the public Customers Purchase Behavior dataset obtained from Kaggle, consisting of 72,637 customer records. The analyzed variables include lifestage and premium customer. Categorical data are transformed through an encoding process and normalized before applying the K-Means Clustering algorithm. The results show that customers can be grouped into three main segments with distinct characteristics based on life stages and price preference levels. This segmentation reflects differences in price sensitivity and potential customer value across segments. The study concludes that the K-Means Clustering method is effective for customer segmentation based on non-transactional customer profile data. The segmentation results can be utilized as a basis for decision-making in developing more targeted and data-driven marketing management strategies.

Keywords: customer segmentation, data mining, K-Means clustering, customer profile, marketing management.

1. PENDAHULUAN

Perkembangan bisnis dan pemasaran modern menuntut perusahaan untuk memahami karakteristik konsumennya secara lebih mendalam. Segmentasi konsumen merupakan salah satu pendekatan strategis dalam manajemen pemasaran untuk mengelompokkan pelanggan ke dalam segmen yang homogen sehingga strategi pemasaran dapat disusun secara lebih efektif dan efisien [1]

Segmentasi konsumen tidak selalu harus berbasis data transaksi. Informasi profil pelanggan seperti tahapan kehidupan (lifestage) dan preferensi harga (premium customer) juga mampu merepresentasikan perilaku konsumen dalam konteks pengambilan keputusan pembelian. Preferensi harga mencerminkan sensitivitas konsumen terhadap nilai dan kualitas produk, sedangkan tahapan kehidupan erat dengan kebutuhan dan pola konsumsi pelanggan [2]

Pemanfaatan teknik data mining memungkinkan pengolahan data pelanggan dalam jumlah besar untuk menemukan pola tersembunyi yang sulit diidentifikasi secara manual. Salah satu teknik data mining yang banyak digunakan adalah clustering, yaitu proses pengelompokan data berdasarkan tingkat kemiripan karakteristik tanpa menggunakan label kelas [3].

Algoritma K-Means Clustering merupakan metode partitional clustering yang bertujuan meminimalkan jarak antara data dan pusat cluster (centroid). K-Means banyak digunakan dalam segmentasi konsumen karena memiliki kompleksitas komputasi yang rendah dan menghasilkan cluster yang mudah diinterpretasikan oleh manajemen [4]. Berdasarkan kondisi tersebut, penelitian ini menerapkan metode K-Means Clustering untuk melakukan segmentasi konsumen berbasis profil pelanggan menggunakan data publik Customers Purchase Behavior dari Kaggle, dengan tujuan mendukung penyusunan strategi manajemen pemasaran yang lebih terarah.

Penelitian mengenai segmentasi konsumen menggunakan teknik clustering telah banyak dilakukan [1]. Rahim et al. (2021) menunjukkan bahwa pendekatan segmentasi pelanggan berbasis karakteristik perilaku mampu menghasilkan kelompok pelanggan yang bermakna bagi pengambilan keputusan pemasaran. Studi lain oleh Zhou et al. (2021) [5] menegaskan bahwa teknik clustering efektif dalam mengelompokkan pelanggan berdasarkan kemiripan karakteristik digital dan perilaku konsumsi.

Dalam konteks metode, K-Means masih menjadi algoritma yang dominan digunakan untuk segmentasi pelanggan karena efisiensinya dalam menangani data berukuran besar serta kemudahan interpretasi hasil cluster [4] [2]. Namun demikian, penelitian terkini menyoroti pentingnya validasi hasil clustering menggunakan pendekatan evaluasi yang komprehensif serta interpretasi berbasis konteks bisnis agar hasil segmentasi tidak hanya kuat secara teknis tetapi juga relevan secara manajerial[3].

Kebaruan penelitian ini terletak pada penerapan metode K-Means Clustering menggunakan data profil pelanggan non-transaksional yang bersifat replikatif dan terbuka, serta pada penerjemahan langsung hasil segmentasi ke dalam rekomendasi strategi manajemen pemasaran. Pendekatan ini memberikan alternatif segmentasi konsumen yang tetap relevan meskipun tanpa data transaksi, serta dapat dijadikan dasar pengambilan keputusan pemasaran berbasis data. Kebaruan Penelitian ini diharapkan memberikan kontribusi Menyediakan model segmentasi konsumen berbasis profil pelanggan menggunakan metode K-Means Clustering. Memberikan pemetaan karakteristik konsumen berdasarkan tahapan kehidupan dan preferensi harga. Menjadi dasar penyusunan strategi manajemen pemasaran yang lebih terarah dan berbasis data.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Data yang digunakan adalah dataset publik Customers Purchase Behavior yang diperoleh dari Kaggle, berjumlah 72.637 data pelanggan selanjutnya diolah dengan data mining yaitu proses eksplorasi dan analisis data berukuran besar untuk menemukan pola, hubungan, dan pengetahuan baru yang berguna dalam pengambilan keputusan. Dalam konteks pemasaran, data mining digunakan untuk memahami perilaku konsumen, memprediksi kebutuhan pelanggan, serta mengidentifikasi segmen pasar yang potensial [6] [7]. Berikut adalah kerangka kerja penelitian

1. Pengumpulan Dataset

Tahap ini merupakan proses pengambilan data penelitian yang digunakan sebagai objek analisis. Dataset yang digunakan adalah dataset publik Customers Purchase Behavior yang diperoleh dari platform Kaggle. Dataset tersebut berisi data profil pelanggan yang mencakup variabel lifestage dan premium customer. Data dikumpulkan dalam bentuk file digital dan diverifikasi kelengkapannya sebelum dilakukan pengolahan lebih lanjut [8][9].

2. Pra-Pemrosesan Data

Pra-pemrosesan data dilakukan untuk memastikan kualitas data yang digunakan dalam penelitian. Tahapan ini meliputi pemeriksaan nilai kosong, duplikasi data, dan konsistensi atribut. Data yang tidak relevan atau bersifat identifikasi, seperti nomor kartu pelanggan, dipisahkan dan tidak digunakan sebagai variabel dalam proses clustering [10] [11].

3. Encoding Variabel Kategorik

Karena variabel lifestage dan premium customer bersifat kategorik, maka dilakukan proses encoding untuk mengubah data kategorik menjadi data numerik. Teknik yang digunakan adalah label encoding, di mana setiap kategori diberikan nilai numerik tertentu. Proses ini bertujuan agar data dapat diproses oleh algoritma K-Means yang berbasis perhitungan jarak numerik.

4. Normalisasi Data

Setelah proses encoding, data dinormalisasi menggunakan metode standardisasi agar setiap variabel memiliki skala yang sebanding. Normalisasi diperlukan untuk mencegah dominasi satu variabel terhadap variabel lainnya dalam proses perhitungan jarak pada algoritma K-Means Clustering.

5. Penerapan K-Means Clustering

Pada tahap ini, algoritma K-Means Clustering diterapkan pada data yang telah melalui proses encoding dan normalisasi. Proses clustering dilakukan dengan menentukan jumlah cluster tertentu, menghitung jarak antara data dan centroid, serta mengelompokkan data ke dalam cluster terdekat. Proses ini dilakukan secara iteratif hingga posisi centroid stabil dan hasil clustering konvergen.

$$d(x - y) = \sqrt{\sum_{j=1}^n (x_j - y_j)^2} \quad (1)$$

Dimana :

d : titik dokumen

x : Pusat Cluster

y : Data

Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centorid (jarak terkecil).

Memperbarui nilai centorid. Nilai centorid baru di peroleh dari rata-rata. cluster yang bersangkutan dengan menggunakan rumus:

$$c_k = \frac{1}{n_k} \sum d_1 \quad (2)$$

Dimana :

n_k = jumlah dokumen dalam cluster

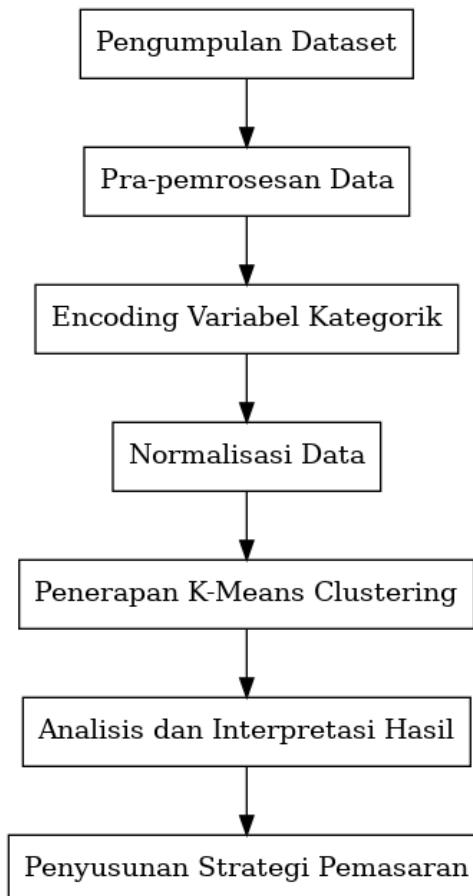
k dan d_1 = dokumen dalam cluster k

6. Analisis dan Interpretasi Hasil

Hasil clustering yang diperoleh dianalisis untuk mengetahui karakteristik masing-masing cluster. Analisis dilakukan dengan mengamati distribusi variabel pada setiap cluster serta menginterpretasikan makna bisnis dari hasil segmentasi. Setiap cluster kemudian diberi label sesuai dengan karakteristik dominannya, seperti pelanggan sensitif harga atau pelanggan bernilai tinggi.

7. Penyusunan Strategi Pemasaran

Tahap akhir penelitian adalah penyusunan strategi manajemen pemasaran berdasarkan hasil segmentasi konsumen. Strategi dirancang dengan menyesuaikan karakteristik masing-masing cluster, sehingga setiap segmen konsumen mendapatkan pendekatan pemasaran yang berbeda dan lebih tepat sasaran.



Gambar. 1 Diagram Alir

3. HASIL DAN PEMBAHASAN

1. Data

Dataset yang digunakan dalam penelitian ini merupakan dataset publik Customers Purchase Behavior yang diperoleh dari Kaggle. Dataset terdiri dari 72.637 data pelanggan dengan tiga atribut utama, yaitu lyly_card_nbr, lifestage, dan premium_customer. atribut lyly_card_nbr berfungsi sebagai identitas pelanggan, sedangkan lifestage dan premium_customer digunakan sebagai variabel utama dalam proses segmentasi. Variabel life stage merepresentasikan tahapan kehidupan pelanggan, sedangkan premium_customer menunjukkan tingkat preferensi harga pelanggan yang diklasifikasikan ke dalam kategori Budget, Mainstream, dan Premium. Dataset ini tidak memuat informasi transaksi, sehingga segmentasi dilakukan berdasarkan profil dan preferensi pelanggan.

2. Pra-Pemrosesan dan Encoding Data

Sebelum dilakukan clustering, data melalui tahapan pra-pemrosesan berupa: Pemeriksaan kelengkapan data (tidak ditemukan nilai kosong). Transformasi data kategorik menjadi data numerik menggunakan teknik label encoding. Hasil encoding ditunjukkan pada Tabel berikut.

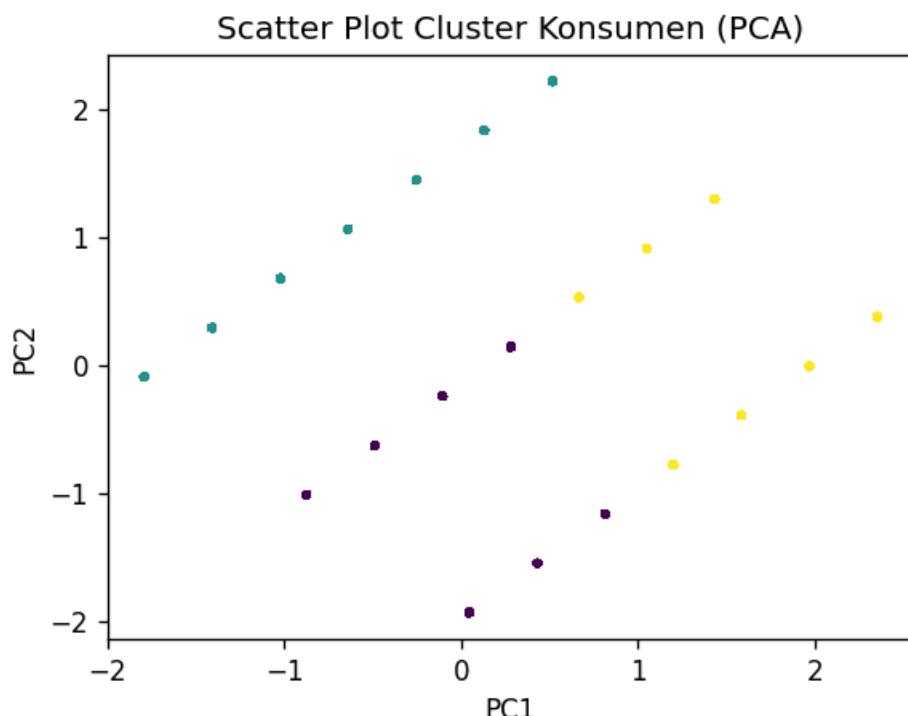
Tabel 1 Hasil Encoding Variabel

Variabel	Kategori	Nilai Encoding
LIFESTAGE	Young Singles/Couples	0
	Midage Singles/Couples	1
	Young Families	2
	Older Families	3
	Retirees	4
PREMIUM_CUSTOMER	Budget	0
	Mainstream	1
	Premium	2

Selanjutnya, data hasil encoding dinormalisasi menggunakan Standardisasi (Z-Score) agar setiap variabel memiliki skala yang sebanding dalam proses perhitungan jarak pada algoritma K-Means.

3 Hasil Clustering Menggunakan Metode K-Means

Pengujian clustering dilakukan dengan beberapa nilai jumlah cluster (K). Berdasarkan kestabilan hasil dan kemudahan interpretasi dalam konteks manajemen pemasaran, jumlah cluster yang digunakan dalam penelitian ini adalah K = 3. Pemilihan ini mempertimbangkan prinsip segmentasi pelanggan yang umum digunakan, yaitu pemisahan pelanggan berdasarkan sensitivitas harga dan tahapan kehidupan. Hasil clustering menghasilkan tiga kelompok pelanggan dengan karakteristik yang berbeda sebagai berikut:



Gambar 2 Scatter plot hasil clustering konsumen menggunakan metode K-Means setelah reduksi dimensi PCA.

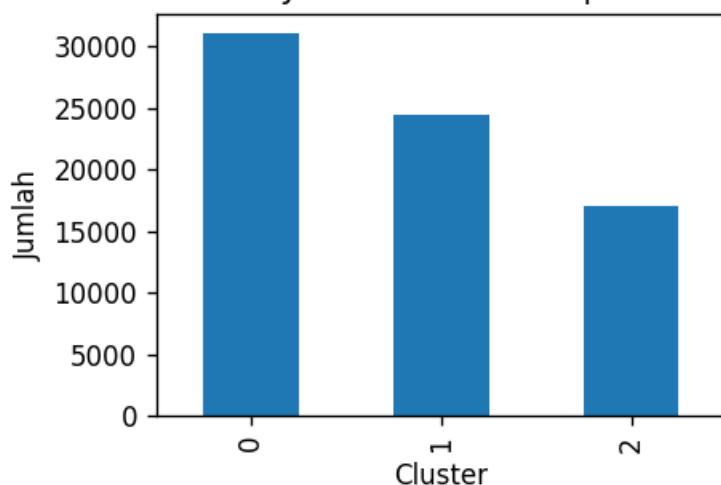
Setelah dilakukan reduksi dimensi dengan Principal Component Analysis (PCA). Setiap titik pada grafik merepresentasikan satu konsumen, sedangkan warna yang berbeda menunjukkan cluster yang berbeda. Visualisasi ini memperlihatkan bahwa konsumen dapat dikelompokkan ke dalam beberapa cluster yang relatif terpisah, sehingga menunjukkan bahwa metode K-Means mampu membentuk segmentasi konsumen berdasarkan kemiripan karakteristik profil pelanggan.

Tabel 2 Hasil Segmentasi Pelanggan

Cluster	Karakteristik Dominan	Interpretasi Segmen
Cluster 1	<i>Budget – Families – Retirees</i>	Pelanggan sensitif harga
Cluster 2	<i>Mainstream – Beragam lifestage</i>	Pelanggan reguler
Cluster 3	<i>Premium – Singles/Couples</i>	Pelanggan bernilai tinggi

Hasil ini menunjukkan bahwa kombinasi variabel lifestage dan premium customer mampu membentuk segmen pelanggan yang jelas dan mudah diinterpretasikan.

Distribusi Jumlah Konsumen per Cluster



Gambar 3 Distribusi jumlah konsumen pada masing-masing cluster.

Gambar 3 menunjukkan distribusi jumlah konsumen pada setiap cluster yang dihasilkan oleh metode K-Means. Perbedaan jumlah anggota pada masing-masing cluster mengindikasikan adanya variasi dominasi segmen konsumen dalam dataset. Informasi ini penting untuk memahami proporsi setiap segmen, sehingga manajemen dapat menentukan prioritas strategi pemasaran berdasarkan ukuran dan potensi masing-masing cluster.

5. Contoh 1 Iterasi Algoritma K-Means

Untuk memberikan gambaran proses kerja algoritma K-Means, berikut ditunjukkan satu iterasi awal berdasarkan data yang telah di-encoding.

Langkah 1: Penentuan Centroid Awal

Misalkan dipilih tiga centroid awal secara acak sebagai berikut:

Tabel 2 centroid awal

Cluster	LIFESTAGE_ENC	PREMIUM_ENC
C1	4	0
C2	2	1
C3	0	2

Centroid tersebut masing-masing mewakili:

C1: Retirees – Budget

C2: Families – Mainstream

C3: Young Singles – Premium

Langkah 2: Perhitungan Jarak

Sebagai contoh, satu data pelanggan dengan:

LIFESTAGE_ENC = 0 (Young Singles/Couples)

PREMIUM_ENC = 2 (Premium)

Dihitung jaraknya ke masing-masing centroid menggunakan Euclidean Distance:

$$d(C1) = \sqrt{(0 - 4)^2 + (2 - 0)^2} = 20$$

$$d(C2) = \sqrt{(0 - 2)^2 + (2 - 1)^2} = 5$$

$$d(C3) = \sqrt{(0 - 0)^2 + (2 - 2)^2} = 0$$

Karena jarak terdekat adalah ke Cluster 3, maka data pelanggan tersebut dimasukkan ke Cluster 3.

Langkah 3: Pembaruan Centroid

Setelah seluruh data dialokasikan ke cluster masing-masing, centroid baru dihitung berdasarkan rata-rata nilai variabel dalam setiap cluster. Proses ini diulang hingga posisi centroid tidak berubah secara signifikan (konvergen).

3.1 Implementasi

Hasil penelitian menunjukkan bahwa metode K-Means mampu mengelompokkan pelanggan ke dalam segmen yang bermakna secara manajerial. Cluster pelanggan dengan kategori premium didominasi oleh kelompok singles/couples yang memiliki potensi nilai tinggi dan loyalitas yang kuat. Sebaliknya, pelanggan dengan kategori budget lebih banyak berasal dari kelompok keluarga dan pensiunan yang memiliki sensitivitas harga tinggi. Temuan ini memperkuat konsep segmentasi pelanggan dalam manajemen pemasaran, di mana pemahaman terhadap profil dan preferensi pelanggan menjadi dasar utama dalam penyusunan strategi pemasaran yang efektif dan efisien.

Implikasi Manajerial Segmentasi yang dihasilkan dapat digunakan sebagai dasar strategi pemasaran, antara lain:

1. Cluster Budget: promosi harga, diskon, dan bundling produk
2. Cluster Mainstream: promosi musiman dan cross-selling
3. Cluster Premium: program loyalitas dan penawaran eksklusif



Gambar 4 Peta strategi manajemen pemasaran berdasarkan hasil segmentasi konsumen.

Gambar 4 menggambarkan pemetaan strategi manajemen pemasaran yang disusun berdasarkan hasil segmentasi konsumen. Setiap cluster dihubungkan dengan strategi pemasaran yang berbeda sesuai dengan karakteristik dominan konsumen dalam segmen tersebut. Pendekatan ini menunjukkan bagaimana hasil analisis data mining dapat diterjemahkan secara langsung ke dalam rekomendasi strategi pemasaran yang lebih terarah dan tepat sasaran.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut. Metode K-Means Clustering berhasil diterapkan untuk melakukan segmentasi konsumen berdasarkan data profil pelanggan yang terdiri dari variabel lifestage dan premium customer. Proses encoding dan normalisasi data kategorik memungkinkan algoritma K-Means mengelompokkan konsumen ke dalam segmen-semen yang memiliki karakteristik homogen. Hasil clustering menghasilkan beberapa segmen konsumen yang memiliki perbedaan karakteristik yang jelas, khususnya dalam hal tahapan kehidupan dan preferensi harga. Segmentasi konsumen berbasis profil pelanggan terbukti relevan untuk mendukung pengambilan keputusan dalam manajemen pemasaran, meskipun penelitian ini tidak menggunakan data transaksi. Hasil segmentasi dapat digunakan sebagai dasar dalam penyusunan strategi manajemen pemasaran yang lebih terarah, efektif, dan sesuai dengan karakteristik masing-masing segmen konsumen.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada LPPM STMIK Pelita Nusantara atas dukungan moril dan materil sehingga penelitian ini dapat dilaksanakan dengan sebaik baiknya. Terimakasih juga kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

DAFTAR PUSTAKA

- [1] M. Abdul, M. Mushafiq, S. Khan, and Z. Ali, “Journal of Retailing and Consumer Services RFM -based repurchase behavior for customer classification and segmentation,” *J. Retail. Consum. Serv.*, vol. 61, no. March, p. 102566, 2021, doi: 10.1016/j.jretconser.2021.102566.
- [2] K. Tabianan and S. Velu, “K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data,” pp. 1–15, 2022.
- [3] M. Gagolewski, M. Bartoszuk, and A. Cena, “Are cluster validity measures (in) valid ?,” vol. 581, p. 3220, 2021.
- [4] P. Anitha and M. M. Patil, “RFM model for customer purchase behavior using K-Means algorithm,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1785–1792, 2022, doi: 10.1016/j.jksuci.2019.12.011.
- [5] J. Zhou, J. Wei, and B. Xu, “Journal of Retailing and Consumer Services,” *J. Retail. Consum. Serv.*, vol. 61, no. March, p. 102588, 2021, doi: 10.1016/j.jretconser.2021.102588.
- [6] A. Ullah, M. I. Mohmand, H. Hussain, S. Johar, and I. Khan, “Customer Analysis Using Machine Learning-Based,” 2023.
- [7] D. Fitriani, T. N. Padilah, and B. N. Sari, “Penerapan Algoritma K-Means Dalam Pengelompokan Kesejahteraan Rakyat Berdasarkan Kecamatan di Kabupaten Karawang,” *Progresif J. Ilm. Komput.*, vol. 17, no. 2, p. 73, 2021, doi: 10.35889/progresif.v17i2.649.
- [8] M. A. Mahendra, M. Firmansyah, and G. Triyono, “Design and Construction of Employee Recruitment System Application using Profile Matching Method,” *INOVTEK Polbeng-Seri Inform.*, vol. 10, no. 2, pp. 762–771, 2025.
- [9] N. Padilla and E. Ascarza, “Overcoming the Cold Start Problem of Customer Relationship Management Using a Probabilistic Machine Learning Approach,” vol. 58, no. 5, pp. 981–1006, 2021, doi: 10.1177/00222437211032938.
- [10] M. Rezaei, P. Fränti, and S. Member, “Set Matching Measures for External Cluster Validity,” vol. 4347, no. c, pp. 1–14, 2016, doi: 10.1109/TKDE.2016.2551240.
- [11] J. Joung and H. Kim, “International Journal of Information Management Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews,” *Int. J. Inf. Manage.*, vol. 70, no. January, p. 102641, 2023, doi: 10.1016/j.ijinfohmgt.2023.102641.