

Segmentasi Nasabah Bank Pada Data Campuran Menggunakan K-Means Clustering

Joceline Schellenberg W*, Mohammad Andri Budiman, Amalia

Program Studi S2 Sains Data Dan Kecerdasan Buatan, Fakultas Ilmu Komputer Dan Teknologi Informasi, Universitas Sumatera Utara, Medan, Indonesia

Email: ^{1,*}joceline@students.usu.ac.id, ²mandrib@usu.ac.id, ³amalia@usu.ac.id

Email Penulis Korespondensi: joceline@students.usu.ac.id

Submitted: 16/01/2025; Accepted: 23/02/2025; Published: 31/03/2024

Abstrak– Dalam rangka meningkatkan perluasan penggunaan layanan Bank Pemerintah Daerah, segmentasi nasabah sangat penting bagi bank untuk mengembangkan strategi pemasaran yang disesuaikan dengan kelompok nasabah tertentu. Meskipun model RFM umum digunakan, meningkatkan perluasan penggunaan layanan memerlukan data tentang preferensi transaksi pelanggan, yang biasanya bersifat kategoris. Oleh karena itu, penelitian ini mengelompokkan nasabah bank berdasarkan riwayat transaksi mereka, tidak hanya memanfaatkan data numerik tetapi juga data kategoris yang mewakili preferensi transaksi menggunakan K-Means Clustering. Model pengelompokan secara efektif mengelompokkan pelanggan ke dalam empat kluster dengan karakteristik berbeda.

Kata Kunci: Segmentasi Pelanggan; Data Campuran; K-Means Clustering

Abstract– In order to increase the extension of the use of Local Government Banks’s services, customer segmentation is crucial for banks to develop marketing strategies tailored to specific customer groups. While the RFM model is commonly used, enhancing service usage expansion requires data on customer transaction preferences, which are typically categorical in nature. Therefore, this study segments bank customers based on their transaction history, utilizing not only numerical data but also categorical data representing transaction preferences using K-Means Clustering. The clustering model effectively groups customers into four clusters with distinct characteristics.

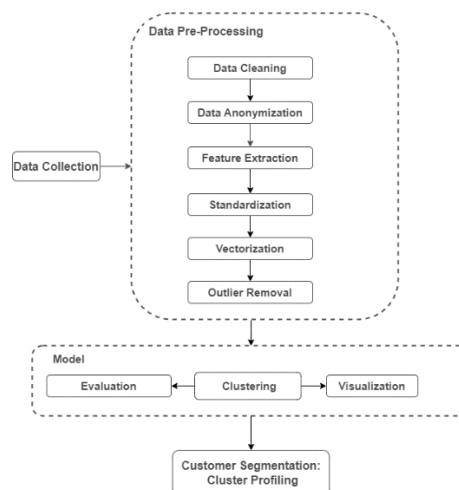
Keywords: Customer Segmentation; Mixed Data; K-Means Clustering;

1. PENDAHULUAN

Jumlah nasabah masing-masing bank pemerintah daerah lebih rendah daripada bank lain karena tanggung jawab utamanya adalah mengembangkan ekonomi dan pembangunan daerah melalui kegiatan perbankan [1]. Bank perlu menerapkan segmentasi nasabah untuk mengembangkan strategi pemasaran yang disesuaikan dengan kelompok nasabah tertentu dalam rangka memperluas penggunaan layanan perbankan. Meskipun model RFM sering digunakan, data tentang preferensi transaksi pelanggan, yang biasanya bersifat kategoris, diperlukan untuk memperluas penggunaan layanan. Oleh karena itu, penelitian ini membagi klien bank mengingat riwayat pertukaran mereka, menggunakan informasi matematis serta informasi langsung yang membahas kecenderungan pertukaran menggunakan K-Means Clustering

2. METODOLOGI PENELITIAN

Metodologi ini menggunakan 3 fase utama: Pengumpulan Data, Pra-pemrosesan Data, dan fase Model. Profil nasabah bank dibuat berdasarkan hasil kluster yang diprediksi, seperti yang ditunjukkan pada gambar 1:



Gambar 1. Arsitektur Umum

2.1 Data Koleksi

Raw Data merupakan laporan transaksi harian Bank Pemerintah pada bulan Januari 2023 dari sistem basis data bank dalam format file TXT yang terdiri dari 19 atribut dengan variabel data berupa numerik, teks kategoris, dan waktu. Terdapat 3 file txt untuk setiap harinya. Penelitian ini hanya menggunakan data rekening bank pribadi, yang terdiri dari ATM, mobile banking, dan transaksi via/channel lainnya.

2.2 Pra-Pemrosesan Data

Langkah ini terdiri dari Pembersihan Data, Anonimisasi Data, Ekstraksi Fitur, Standardisasi, Vektorisasi, dan Penghapusan Outlier.

1. Pembersihan Data: Pada langkah ini, kolom-kolom difilter, lalu baris-baris yang berisi transaksi yang dibatalkan dan biaya atau pungutan yang dicatat dihapus, karena hal tersebut tidak diperlukan dalam studi ini.
2. Anonimisasi Data: kami menggunakan penggantian acak untuk mengubah nomor akun dengan 7 nomor acak untuk memastikan kerahasiaan. Ekstraksi Fitur terdiri dari 2 langkah utama:
 - a. Kategori Transaksi dan Via untuk setiap baris transaksi, menggunakan fungsi kondisional berdasarkan kata kunci yang ada di Rincian Transaksi. Kami menginisialisasi 8 kategori (Tagihan, Utilitas, Telekomunikasi, Transfer, Penarikan, Setoran, dompet elektronik: Top up dompet elektronik) dan 3 via/saluran (ATM, Mobile banking, Lainnya) yang diperoleh dari Rincian Transaksi. Jika kata kunci tertentu ditemukan di Rincian Transaksi, kategorikan transaksi sebagai kategori yang diinisialisasi dan via.
 - b. RFM, Interpurchase Time, dan 4 Preferred Category dan Via untuk setiap pelanggan, menggunakan agregasi dan pemeringkatan padat, dikelompokkan berdasarkan nomor akun. Untuk memperoleh Interpurchase Time, hanya akun dengan nilai F yang digunakan dalam penelitian ini.
3. Standardisasi: Sebelum menerapkan model K-means untuk pengelompokan, penting untuk menstandarisasi atau menormalkan data numerik. Hal ini dilakukan untuk memastikan bahwa semua fitur memiliki pengaruh yang sama dan untuk mencegah fitur skala besar mendominasi, sehingga meningkatkan akurasi, konvergensi, dan stabilitas algoritma K-means. Dalam penelitian ini, kami menggunakan MinMax Scaler untuk menstandarisasi atau menormalkan data numerik. MinMax Scaler mengubah nilai data menjadi rentang tertentu, biasanya antara 0 dan 1, seperti yang ditunjukkan pada Persamaan 1.

$$x_{scaled} = \frac{x - \min}{\max - \min} \quad (1)$$

K-means berkinerja lebih baik dalam menangani data numerik, oleh karena itu data kategoris perlu diubah terlebih dahulu menggunakan pengodean manual, dengan menetapkan label angka pada setiap kategori transaksi dan metode pembayaran untuk merepresentasikan data sebagai angka. Data di setiap baris berisi kategori dan metode pembayaran yang dipisahkan dengan koma. Setiap kategori kemudian dipindahkan ke 4 kolom baru dan diubah menjadi nomor label, dimulai dari 1, 2, dan seterusnya. Pelanggan dengan kurang dari 4 kategori dan metode pembayaran akan diberi label 0.

4. Penghapusan Outlier: kami menggunakan Outlier ECOD, yang diusulkan oleh [12] untuk menghilangkan outlier. Metode ini dipilih karena dapat menjadi metode yang layak untuk mendeteksi outlier dalam data transaksi bank yang divektorkan, karena pemahaman distribusi data yang terbatas, dan penelitian ini memerlukan metode yang kurang sensitif terhadap asumsi. Selain itu, pendekatan ini lebih fleksibel.

2.3 Pemodelan

Tahapan ini terdiri dari penentuan K, Model Clustering, Evaluasi Model dan Visualisasi Cluster.

Penentuan K

Metode Elbow digunakan untuk menemukan jumlah kluster (k) yang optimal menggunakan kumpulan data dan rentang nilai k potensial (2 hingga 8 dalam kasus ini). Pengelompokan K-Means diterapkan untuk setiap nilai k dalam rentang yang ditentukan, dan metrik WCSS dihitung untuk setiap pengelompokan yang dihasilkan. Grafik garis kemudian dibuat, dengan sumbu x yang mewakili jumlah kluster (k) dan sumbu y yang mewakili nilai metrik. Biasanya, plot menunjukkan tren penurunan dengan bentuk "siku" yang khas, yang sering kali menunjukkan jumlah kluster (k) yang optimal. Menambahkan lebih banyak kluster di luar titik ini mungkin tidak menghasilkan manfaat tambahan yang signifikan.

Model Pengelompokan K-means

K-means bekerja secara iteratif dengan mengoptimalkan posisi centroid (titik pusat) setiap cluster. Rumusnya ditunjukkan pada persamaan 2

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (2)$$

Menurut [5], langkah-langkah Algoritma K-means adalah:

1. Inisialisasi: Mulailah dengan menentukan nilai k , yang merupakan jumlah kluster yang akan dibuat. Pilih secara acak k titik data untuk dijadikan centroid awal, yang masing-masing mewakili rata-rata semua titik data dalam klusternya masing-masing.
2. Penetapan Kluster: Hitung jarak Euclidean antara setiap titik data dan semua centroid. Jarak Euclidean digunakan untuk menilai kesamaan antara titik data dan centroid, dihitung sebagai akar kuadrat dari jumlah selisih kuadrat nilai-nilai di setiap dimensi. Selanjutnya, alokasikan setiap titik data ke kluster dengan centroid terdekat.
3. Pembaruan Senroid: Setelah semua titik data ditetapkan ke kluster, lanjutkan dengan memperbarui sentroid setiap kluster berdasarkan anggotanya saat ini.
4. Iterasi: Ulangi langkah 2 dan 3 hingga kriteria konvergensi terpenuhi. Kriteria konvergensi penting karena menentukan kapan algoritma harus berhenti. Kriteria ini penting untuk mencegah algoritma terjebak dalam solusi lokal dan memastikan bahwa algoritma memperoleh hasil yang konsisten. Kriteria konvergensi dapat meliputi: a. Memastikan bahwa perubahan dalam centroid berada di bawah ambang batas tertentu. b. Memeriksa bahwa jumlah iterasi tidak melebihi nilai maksimum. c. Memverifikasi bahwa tidak ada perubahan dalam penugasan kluster
5. Hasil: Setelah konvergensi tercapai, kembalikan pengelompokan data akhir, dengan setiap titik data dalam pengelompokannya yang sesuai.

Evaluasi Model

Model ini dievaluasi menggunakan tiga metrik evaluasi:

1. Indeks Calinski-Harabasz menghitung rasio total varians antar kluster dibagi dengan varians dalam kluster individu. Menurut [13], indeks ini dihitung dengan membandingkan dispersi kluster dengan dispersi intrakluster.
2. Rata-rata kesamaan antara setiap cluster dan cluster yang paling mirip diukur dengan indeks Davies-Buldin. Seperti yang terlihat pada [14], nilai tersebut diperoleh dari rasio rata-rata antara jarak antar node dan jarak antar cluster.
3. Indeks siluet menghitung tingkat kecocokan titik data individual dengan klusternya sendiri dibandingkan dengan kluster lainnya. Skor berkisar dari -1 hingga 1, seperti pada [15].

Visualisasi

Kami memanfaatkan PCA (Principal Component Analysis) dalam langkah visualisasi untuk melihat bagaimana kluster didistribusikan dalam ruang data.

Seperti yang disebutkan dalam [16], PCA mendeteksi pola penting dalam data dengan mengubah variabel asli menjadi komponen baru yang dikenal sebagai komponen utama. Hal ini dilakukan dengan menghitung vektor eigen dan nilai eigen dari matriks kovariansi data. Vektor eigen menunjukkan garis arah komponen utama, dan nilai eigen menunjukkan jumlah varians yang dijelaskan oleh setiap komponen.

PCA secara efektif mengurangi dimensionalitas dan menyediakan representasi visual yang lebih intuitif dalam dimensi yang lebih rendah sambil mempertahankan informasi sebanyak mungkin, membuat visualisasi lebih mudah dipahami.

Pada langkah terakhir, kami membuat profil pelanggan untuk melihat ciri masing-masing kluster, dengan menggabungkan hasil kluster yang diprediksi dan mengelompokkannya berdasarkan kluster.

3. HASIL DAN PEMBAHASAN

NO_REK	KET_TX	JUMLAH_TX	TGL_TX
1,32E+13	TARIK INT VIA ATM-	100000	01/10/2023
1,32E+13	TRF ONLINE VIA MO	1000000	01/10/2023
1,32E+13	TRF ONLINE VIA MO	120000	01/10/2023
1,32E+13	PAY GOPAY VIA MOI	250000	01/10/2023
1,01E+15	PENARIKAN TUNAI	350000	01/10/2023

Gambar 2. Contoh data yang sudah dibersihkan

Gambar 2 menunjukkan contoh data yang telah dibersihkan. Melalui langkah pembersihan data, kami memperoleh 4 kolom yang diperlukan: Nomor Rekening, Nominal Transaksi, Rincian Transaksi, dan Tanggal Transaksi, dengan 1.451.811 baris transaksi dari 254.300 nasabah bank.

Recency	Frequency	Monetary	Interpurchase_Time	Category	Via
0	21	53279515	0	Telecommunication, Transfer, Utilities	m-Banking, m-Banking, m-Banking
1	4	420000	5	Transfer, Withdrawal	m-Banking, ATM
0	12	17300000	1	Transfer, Withdrawal	m-Banking, ATM
13	3	3150000	0	Withdrawal	ATM
1	6	6035000	1	Transfer	m-Banking

Gambar 3. Contoh Data setelah Pra-Pemrosesan

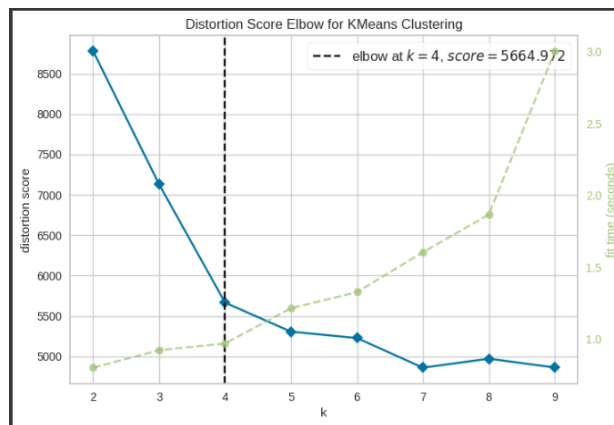
Gambar 3 menunjukkan contoh data setelah langkah pra-pemrosesan. Ada 195044 akun dari 254.300 pelanggan terdaftar yang dipilih untuk langkah berikutnya. Pada kolom Kategori dan Via, kami memperoleh 4 Kategori dan Via teratas untuk setiap pelanggan bank berdasarkan frekuensi, di mana peringkat ditempatkan secara berurutan dalam 1 bidang. Ada baris dengan nilai kurang dari 4, artinya beberapa pelanggan memiliki kurang dari 4 kategori dan via, karena mereka hanya menggunakan sedikit layanan bank. Setelah penghapusan outlier, ada 175.539 akun dari 195044 yang tersisa sebagai input data untuk segmentasi.

Recency	Frequency	Monetary	Interpurchase_Time	cat1	cat2	cat3	cat4	via1	via2	via3	via4
0	0.000000	0.327586	0.029600	0.000000	3	1	2	0	9	9	9
1	0.038462	0.034483	0.000233	0.384615	1	5	0	0	9	10	0
2	0.000000	0.172414	0.009611	0.076923	1	5	0	0	9	10	0
3	0.500000	0.017241	0.001750	0.000000	5	0	0	0	10	0	0
4	0.038462	0.068966	0.003353	0.076923	1	0	0	0	9	0	0
...

Gambar 4. Contoh Data Setelah Transformasi

Seperti yang ditunjukkan pada gambar 4, kategori transaksi dan via ditransformasikan menggunakan Pengodean Manual untuk setiap baris.

Dengan menggunakan Metode Elbow, k=4 ditentukan sebagai n kluster ideal.



Gambar 5. K Ideal

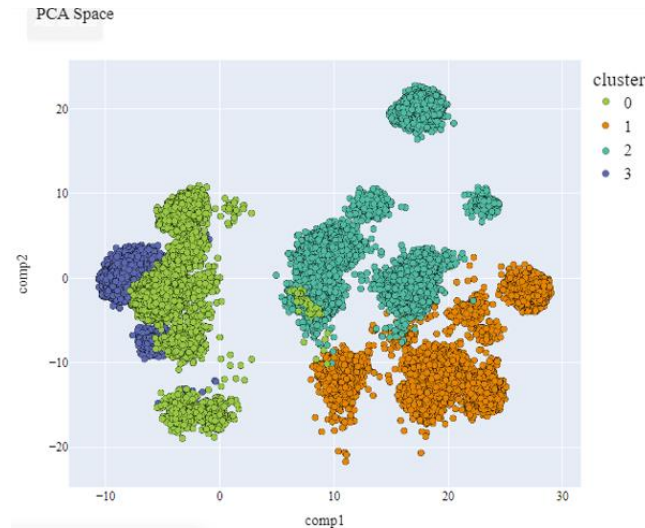
Pada Gambar 5, ditunjukkan bahwa jumlah kluster yang optimal adalah empat. Kesimpulan ini diambil dari diagram, yang menggambarkan penurunan minimal dalam inersia setelah jumlah kluster mencapai tiga. Inersia menunjukkan total jarak kuadrat antara setiap titik data dalam kluster dan pusatnya. Inersia yang lebih rendah menunjukkan bahwa titik data dalam kluster umumnya lebih dekat ke pusatnya. Dengan demikian, nilai k=4 dianggap paling sesuai untuk pemodelan K-Means untuk Pengelompokan dalam studi ini.

Model K-Means mengelompokkan 4 pelanggan. Metrik yang digunakan untuk mengevaluasi model adalah Indeks Calinski-Harabasz, Indeks Davies-Bouldin, dan Skor Silhouette dengan hasil seperti yang terlihat pada gambar 6:

```
Davies bouldin score: 0.23814110982232028
Calinski Score: 1640023.8612840208
Silhouette Score: 0.855598073452313
```

Gambar 6. Skor Metrik Evaluasi

5. Skor Davies-Bouldin (Nilai yang lebih rendah lebih baik): Skor ini menunjukkan pengelompokan yang relatif baik, karena nilai yang mendekati 0 menandakan bahwa pengelompokan tersebut kompak (dispersi internal atau penyebaran data yang rendah). Ini berarti bahwa data dalam pengelompokan terpusat di sekitar satu titik atau centroid dan tidak terlalu tersebar.
6. Skor Calinski-Harabasz (Nilai yang lebih tinggi lebih baik): Skor ini menunjukkan kluster yang terdefinisi dengan baik, dengan pemisahan substansial antara kluster dibandingkan dengan penyebaran dalam kluster.
7. Skor Silhouette (Nilai yang mendekati 1 lebih baik): Skor ini menunjukkan kluster yang terpisah dengan baik
8. Gambar 7 menunjukkan bagaimana cluster didistribusikan dalam ruang data menggunakan PCA. Hal ini dapat membantu kita memahami hubungan antara segmen dan mengidentifikasi outlier atau anomali



Gambar 7. Visualisasi Kluster

Gambar 8 menunjukkan profil pelanggan untuk melihat ciri setiap kluster.

cluster	Recency	Frequency	Monetary	Interpurchase_Time	Category	Via	Count
0	3.462169	4.763648	4.020067e+06	3.302096	Withdrawal	ATM	46985
1	5.715139	7.903971	9.434433e+06	2.048219	Transfer, Withdrawal	ATM, ATM	73457
2	7.166832	4.975358	8.476459e+06	2.768526	Transfer	m-Banking	17125
3	17.005346	3.289476	3.178642e+06	0.763773	Withdrawal	ATM	37972

Gambar 8. Profil Kluster

Kluster 0 (46.985 nasabah) memiliki rata-rata Recency 3, Frequency 5, Monetary Rp. 4.020.067 dan Interpurchase Time 3. Sebagian besar nasabah lebih memilih melakukan penarikan melalui ATM.

Kluster 1 (73.457 nasabah) memiliki rata-rata Recency 6, Frequency 8, Monetary Rp.9.434.433 dan Interpurchase Time 2. Sebagian besar nasabah menggunakan mobile banking untuk berbagai kategori transaksi, namun masih lebih memilih melakukan transfer dan tarik tunai melalui ATM.

Kluster 2 (17.125 nasabah) memiliki rata-rata Recency 7, Frequency 5, Monetary Rp.8.476.459 dan Interpurchase Time 3. Sebagian besar nasabah menggunakan mobile banking hanya untuk transfer dan pembelian/pembayaran biaya telekomunikasi.

Kluster 3 (37.972 nasabah) memiliki rata-rata Recency 17, Frequency 3, Monetary Rp.3.178.642 dan Interpurchase Time 1. Sebagian besar nasabah lebih memilih melakukan penarikan melalui ATM.

Cluster 0 dan Cluster 3 memiliki preferensi yang sama, tetapi Cluster 0 lebih aktif dan menghabiskan lebih banyak daripada Cluster 3. Cluster 1 dan Cluster 2 memiliki aktivitas dan pengeluaran yang sama, tetapi preferensinya berbeda. Secara umum, dapat dilihat bahwa pelanggan hanya menggunakan beberapa layanan, dengan tingkat nominal dan frekuensi interaksi yang sedang. Beberapa strategi yang dapat dipertimbangkan adalah Cross-Selling dan Up-Selling, Program Loyalitas dan Hadiah, Pemasaran yang Dipersonalisasi, Edukasi dan Kesadaran, Promosi dan Diskon Khusus, serta Umpan Balik dan Survei Pelanggan.

4. KESIMPULAN

Dengan melakukan pengodean manual pada data kategorikal, model K-Means dapat mengelompokkan nasabah berdasarkan tidak hanya nilai RFM tetapi juga fitur data kategorikal lainnya secara efektif, yang memungkinkan bank untuk lebih memahami nasabahnya. Hal ini memungkinkan bank untuk merancang strategi pemasaran yang tepat untuk memperluas penggunaan layanan bank oleh nasabah dan meningkatkan loyalitas dan keterlibatan nasabah. Evaluasi dengan metrik yang digunakan menunjukkan pemisahan kluster yang baik, dengan Skor Silhouette sebesar 0,85, indeks Calinski-Harabasz sebesar 1.640.023,86, dan indeks Davies-Bouldin sebesar 0,23. Hasil segmentasi pelanggan menggunakan clustering menunjukkan empat kelompok pelanggan. Secara umum, dapat dilihat bahwa pelanggan hanya menggunakan beberapa layanan, dengan tingkat frekuensi nominal dan interaksi yang moderat. Beberapa strategi yang dapat dipertimbangkan meliputi Cross-Selling dan Up-Selling, Program Loyalitas dan Hadiah, Pemasaran yang Dipersonalisasi, Edukasi dan Kesadaran, Promosi dan Diskon Khusus, serta Umpan Balik dan Survei Pelanggan.

REFERENCES

- [1]. F. Rabbani, “Analisis Perbandingan Efisiensi Bank Pembangunan Daerah dan Unit Usaha Syariah BPD Sebelum dan Sesudah Penerapan BPD Regional Champion (BRC) dengan Metode Data Envelopment Analysis (DEA).,” 2014. [Online]. Tersedia: <https://repository.uinjkt.ac.id/dspace/handle/123456789/25078>
- [2]. K. Tsiptsis dan A. Chorianopoulos, “Teknik Penambangan Data dalam CRM: Segmentasi Pelanggan Internal,” Januari 2010, doi: 10.1002/9780470685815.ch7.
- [3]. AJ Christy, A. Umamakeswari, L. Priyatharsini, dan A. Neyaa, “Peringkat RFM – Pendekatan efektif untuk segmentasi pelanggan,” *J. King Saud Univ. - Ilmu Komputer Inf.*, jilid. 33, tidak. 10, hal. 1251–1257, 2021, doi: <https://doi.org/10.1016/j.jksuci.2018.09.004>.
- [4]. A. Cuce, “Analisis Data dalam Segmentasi Pelanggan dan Metode RFM UNIVERSITAS TEKNIS ISTANBUL □ FAKULTAS MANAJEMEN B.Sc. TESIS Eda TIRYAKI Jurusan Teknik Industri JUNI 2022,” no. September, 2022, doi: 10.13140/RG.2.2.28067.94244.
- [5]. AM Ikotun, AE Ezugwu, L. Abualigah, B. Abuhajja, dan J. Heming, “Algoritma pengelompokan K-means: Tinjauan komprehensif, analisis varian, dan kemajuan di era data besar,” *Inf. Sains (Ny)*, jilid. 622, hlm. 178–210, 2023, doi: <https://doi.org/10.1016/j.ins.2022.11.139>.
- [6]. RJ Kuo, M. Alfareza, dan T. Nguyen, “Algoritma fuzzy c-means kemungkinan puncak kepadatan berbasis genetik untuk analisis kluster- studi kasus pada segmentasi pelanggan,” *Eng. Sci. Technol. dan Int. J.*, vol. 47, hal. 101525, November 2023, doi: 10.1016/j.jestch.2023.101525.
- [7]. H. Chen, L. Zhang, X. Chu, dan B. Yan, “Segmentasi pelanggan ponsel pintar berdasarkan pola penggunaan,” *Teknik Informatika Lanjutan*, jilid. 42, hal. 101000, 2019, doi: <https://doi.org/10.1016/j.aei.2019.101000>.
- [8]. C. Wang, “Segmentasi pelanggan yang efisien dalam pemasaran digital menggunakan pembelajaran mendalam dengan pendekatan kecerdasan kelompok,” *Inf. Proses. Kelola.*, jilid. 59, tidak. 6, hal. 103085, 2022, doi: <https://doi.org/10.1016/j.ipm.2022.103085>.
- [9]. C. Rungruang, P. Riyapan, A. Intarasit, K. Chuarkham, dan J. Muangprathub, “Segmentasi pelanggan model RFM berdasarkan pendekatan hierarkis menggunakan FCA,” *Aplikasi Sistem Pakar*, jilid. 237, hal. 121449, 2024, doi: <https://doi.org/10.1016/j.eswa.2023.121449>.
- [10]. J. Zhou, J. Wei, dan B. Xu, “Segmentasi pelanggan dengan penambahan konten web,” *J. Ritel. Konsumsi. Layanan.*, vol. 61, no. Maret, hal. 102588, 2021, doi: 10.1016/j.jretconser.2021.102588.
- [11]. S. Ghoreishi, “Segmentasi Pelanggan Berdasarkan GRFM: Studi Kasus,” vol. 9, no. 1, hlm. 1–6, 2019.
- [12]. Z. Li, Y. Zhao, X. Hu, N. Botta, C. Ionescu, dan GH Chen, “ECOD: Deteksi Outlier Tanpa Pengawasan Menggunakan Fungsi Distribusi Kumulatif Empiris,” *IEEE Trans. Pengetahuan. Teknik Data.*, jilid. 35, tidak. 12, hal. 12181–12193, 2023, doi: 10.1109/TKDE.2022.3159580.
- [13]. IF Ashari, E. Dwi Nugroho, R. Baraku, I. Novri Yanda, dan R. Liwardana, “Analisis Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, dan Evaluasi Rand-Index pada Algoritma K-Means untuk Klasifikasi Daerah Terdampak Banjir di Jakarta,” *J. Aplikasi Informatika Komputer.*, vol. 7, no. 1, hlm. 89–97, 2023, doi: 10.30871/jaic.v7i1.4947.
- [14]. E. Luthfi dan AW Wijayanto, “Analisis Perbandingan Metode Hierarchy, K-Means, dan K-Medoids Clustering Dalam Pengelompokan Indeks Pembangunan Manusia Indonesia Analisis Komparatif Hierarchy, K-Means, dan K-Medoids Clustering serta Metode Pengelompokan Manusia di Indonesia,” *Inovasi*, vol. 17, no. 4, hal. 761–773, 2021.
- [15]. PJ Rousseeuw, “Siluet: Bantuan grafis untuk interpretasi dan validasi analisis kluster,” *J. Komputasi. Aplikasi. Matematika.*, vol. 20, no. C, hal. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.
- [16]. S. Raschka, “Menerapkan Analisis Komponen Utama.” Diakses: 06 Juni 2024. [Online]. Tersedia: https://sebastianraschka.com/Articles/2014_pca_step_by_step.html