

ANALISIS SEGMENTASI PELANGGAN BERDASARKAN POLA PEMBELIAN MENGGUNAKAN METODE CLUSTERING

Muhammad Ibnu Rayyan¹, Eka Ramadanta Sitepu², Sofy Ertika Dewi³, Suci Pratiwi⁴
Sekolah Tinggi Manajemen Informatika dan Komputer KAPUTAMA, Sumatra Utara
E-mail : * ibnurayyan2004@gmail.com¹, dhanta07@gmail.com², sofyerikadewi2003@gmail.com³,
sucitiwi2612@gmail.com⁴

ABSTRAK

Segmentasi pelanggan merupakan strategi penting untuk memahami keberagaman karakteristik pelanggan dan perilaku pembelian guna meningkatkan efektivitas pemasaran serta manajemen hubungan pelanggan. Penelitian ini bertujuan untuk melakukan segmentasi pelanggan berdasarkan pola pembelian menggunakan metode *K-Means clustering*. Penelitian ini menggunakan dataset pelanggan publik yang berisi atribut demografis dan perilaku, termasuk usia, pendapatan tahunan, dan skor pengeluaran (*spending score*). Prapemrosesan data dilakukan untuk meningkatkan kualitas data, yang mencakup penanganan nilai hilang (*missing values*) serta normalisasi data. Jumlah kluster optimal ditentukan menggunakan Metode Elbow dan *Silhouette Score*, yang menghasilkan empat kluster sebagai segmentasi yang paling representatif. Hasil pengelompokan menunjukkan adanya profil pelanggan yang berbeda dengan karakteristik perilaku pembelian dan tingkat pendapatan yang beragam. Salah satu kluster merepresentasikan pelanggan dengan pendapatan tinggi dan skor pengeluaran yang tinggi, yang menunjukkan segmen pelanggan premium. Sementara itu, kluster lainnya mencerminkan pelanggan dengan tingkat pengeluaran sedang hingga rendah meskipun memiliki tingkat pendapatan yang relatif tinggi. Temuan ini menunjukkan bahwa metode *K-Means clustering* efektif dalam mengidentifikasi segmen pelanggan yang bermakna dan mampu mendukung pengambilan keputusan bisnis berbasis data. Model segmentasi yang diusulkan memberikan wawasan praktis dalam pengembangan strategi pemasaran yang lebih terarah serta peningkatan manajemen hubungan pelanggan dalam lingkungan bisnis digital.

Kata kunci

Segmentasi Pelanggan, Klasterisasi K-Means, Penambangan Data, Perilaku Pembelian

ABSTRACT

Customer segmentation is a critical strategy for understanding heterogeneous customer characteristics and purchasing behavior in order to enhance marketing effectiveness and customer relationship management. This study aims to segment customers based on their purchasing patterns using the K-Means clustering method. A public customer dataset containing demographic and behavioral attributes, including age, annual income, and spending score, was employed in this research. Data preprocessing was conducted to improve data quality, including handling missing values and normalization. The optimal number of clusters was determined using the Elbow Method and Silhouette Score, resulting in four clusters as the most representative segmentation. The clustering results reveal distinct customer profiles with different purchasing behaviors and income characteristics. One cluster represents high-income customers with high spending scores, indicating a premium customer segment, while other clusters reflect customers with moderate to low spending behavior despite having relatively high income levels. These findings demonstrate that K-Means clustering is effective in identifying meaningful customer segments that can support data-driven business decision-making. The proposed segmentation model provides practical insights for developing targeted marketing strategies and improving customer relationship management in digital business environments.

Keywords

Customer Segmentation, K-Means Clustering, Data Mining, Purchasing Behavior

1. PENDAHULUAN

Perkembangan teknologi informasi dan pemanfaatan *big data* telah mendorong organisasi untuk mengadopsi pendekatan berbasis data (*data-driven decision making*) dalam pengambilan keputusan strategis. Data tidak lagi dipandang sebagai arsip pasif, melainkan sebagai aset strategis yang memiliki nilai tinggi apabila dianalisis menggunakan metode yang tepat. Dalam konteks pemasaran dan manajemen pelanggan, segmentasi pelanggan menjadi langkah penting untuk memahami karakteristik, kebutuhan, dan perilaku pembelian yang heterogen. Berbagai penelitian menunjukkan bahwa penerapan teknik data mining dan *machine learning* mampu meningkatkan akurasi segmentasi pelanggan serta mendukung perumusan strategi pemasaran yang lebih efektif (Nguyen, 2021; Wang, 2022; Kasem, Hamada and Taj-Eddin, 2024).

Teknik clustering merupakan salah satu pendekatan utama dalam segmentasi pelanggan karena mampu mengelompokkan data tanpa label berdasarkan tingkat kemiripan karakteristik. Berbagai metode telah dikembangkan, mulai dari K-Means, RFM-based clustering, hingga pendekatan *deep learning* dan *multi-criteria decision making* (Christy et al., 2021; Yadegaridehkordi et al., 2021; Barrera, Segura and Maroto, 2024). Meskipun pendekatan berbasis *deep learning* menunjukkan performa yang tinggi, metode tersebut sering kali memerlukan sumber daya komputasi yang besar serta tingkat interpretabilitas yang rendah, sehingga sulit diterapkan pada organisasi berskala kecil dan menengah (Dong, Wang and Abbas, 2021; Joung and Kim, 2023). Hal ini menunjukkan adanya kebutuhan akan metode segmentasi yang lebih sederhana, efisien, dan mudah diinterpretasikan.

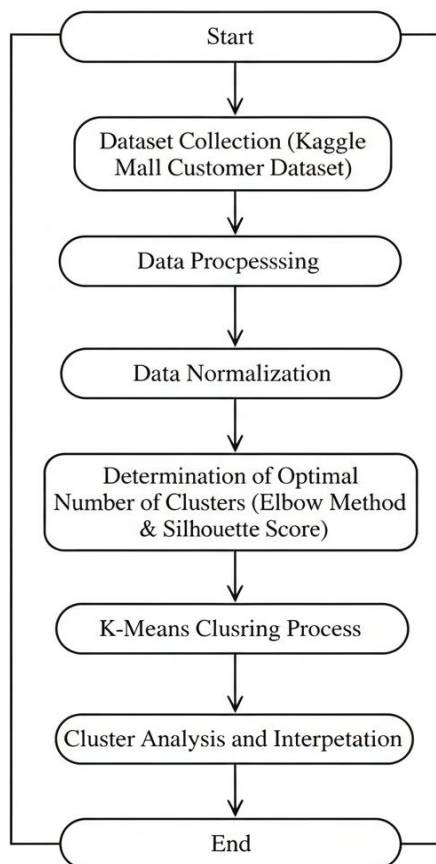
Metode K-Means tetap menjadi algoritma clustering yang populer karena kesederhanaan, efisiensi komputasi, serta kemudahan implementasinya (Syakur et al., 2018; Likas, Vlassis and Verbeek, 2003). Namun, tantangan utama dalam penerapan K-Means terletak pada penentuan jumlah kluster yang optimal serta keberagaman skala atribut data yang dapat memengaruhi kualitas hasil segmentasi. Beberapa penelitian sebelumnya masih terbatas pada penggunaan satu teknik evaluasi kluster atau belum mengombinasikan proses validasi kluster dengan analisis karakteristik pelanggan secara komprehensif (Chen et al., 2019; Dolnicar et al., 2012). Oleh karena itu, diperlukan penelitian yang mengintegrasikan proses penentuan jumlah kluster yang optimal dengan analisis karakteristik setiap segmen pelanggan secara sistematis.

Berdasarkan celah penelitian tersebut, penelitian ini bertujuan untuk menerapkan metode K-Means dalam melakukan segmentasi pelanggan berdasarkan atribut usia, pendapatan tahunan, dan skor pengeluaran dengan mengombinasikan Elbow Method dan Silhouette Score untuk menentukan jumlah kluster yang optimal. Kontribusi utama penelitian ini terletak pada penyajian model segmentasi pelanggan yang sederhana, efisien, dan mudah diinterpretasikan, serta memberikan analisis karakteristik setiap kluster sebagai dasar perumusan strategi pemasaran berbasis data. Hasil penelitian diharapkan dapat menjadi referensi praktis bagi organisasi dalam mengembangkan strategi pemasaran yang lebih tepat sasaran dan berorientasi pada kebutuhan pelanggan.

2. METODE PENELITIAN

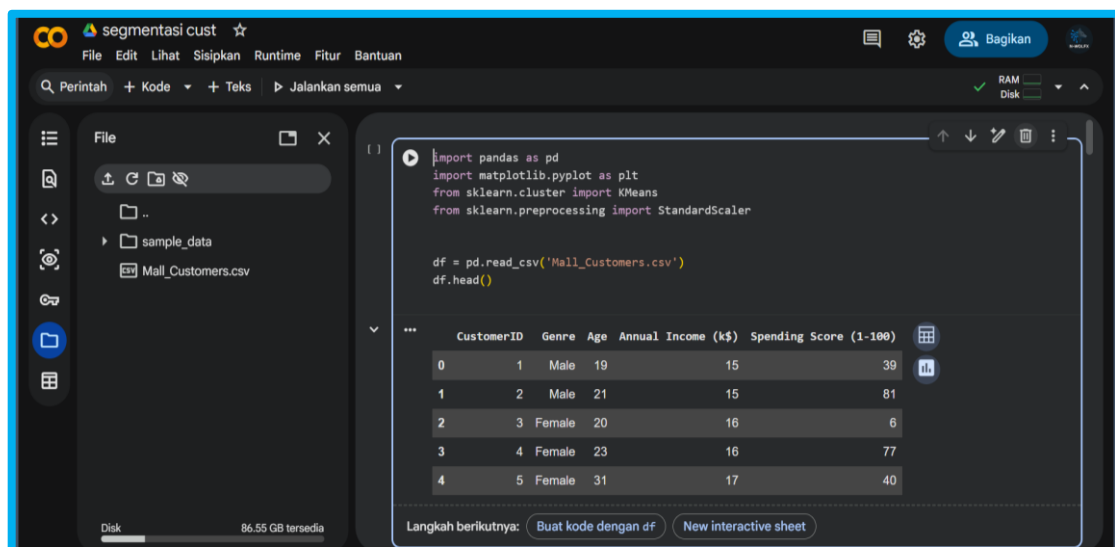
Penelitian ini menggunakan pendekatan kuantitatif dengan metode *data mining* untuk melakukan segmentasi pelanggan berdasarkan pola pembelian. Pendekatan kuantitatif dipilih karena mampu mengolah data numerik secara objektif dan menghasilkan pola segmentasi yang dapat diukur secara statistik (Mukhamediev et al., 2022). Dataset yang digunakan merupakan dataset publik pelanggan pusat perbelanjaan yang diperoleh dari platform Kaggle, yang memuat atribut usia (*age*), pendapatan tahunan (*annual income*), dan skor pengeluaran (*spending score*). Pemilihan atribut ini didasarkan pada penelitian sebelumnya yang menunjukkan bahwa kombinasi atribut demografis dan perilaku belanja efektif dalam membentuk segmentasi pelanggan yang representatif (Christy et al., 2021; Chen et al., 2019).

Tahapan penelitian meliputi pra-pemrosesan data, normalisasi, penentuan jumlah kluster optimal, proses clustering, serta analisis hasil. Pra-pemrosesan data dilakukan untuk memastikan kualitas data, termasuk pemeriksaan *missing values* dan konsistensi data, sehingga data berada dalam kondisi layak dianalisis. Selanjutnya, normalisasi data diterapkan untuk menyamakan skala antaratribut agar tidak terjadi dominasi nilai tertentu dalam perhitungan jarak, sebagaimana direkomendasikan dalam penelitian clustering berbasis K-Means (Syakur et al., 2018; Likas, Vlassis and Verbeek, 2003). Alur keseluruhan tahapan penelitian ditunjukkan pada Gambar 1, mulai dari pengumpulan data hingga analisis hasil clustering.



Gambar 1. Tahapan penelitian

Penentuan jumlah kluster optimal dilakukan menggunakan dua metode evaluasi, yaitu Elbow Method dan Silhouette Score. Elbow Method digunakan untuk mengamati perubahan nilai *within-cluster sum of squares* (WCSS) terhadap jumlah kluster, sedangkan Silhouette Score digunakan untuk mengukur tingkat kohesi dan separasi antar kluster. Berdasarkan hasil evaluasi dari kedua metode tersebut, diperoleh jumlah kluster optimal sebanyak empat kluster. Selanjutnya, proses pengelompokan pelanggan dilakukan menggunakan algoritma K-Means secara iteratif hingga pusat kluster (*centroid*) mencapai kondisi konvergen dan stabil. Lingkungan pengolahan data menggunakan bahasa Python pada platform Google Colaboratory ditampilkan pada Gambar 2 sebagai media komputasi yang digunakan dalam penelitian ini.



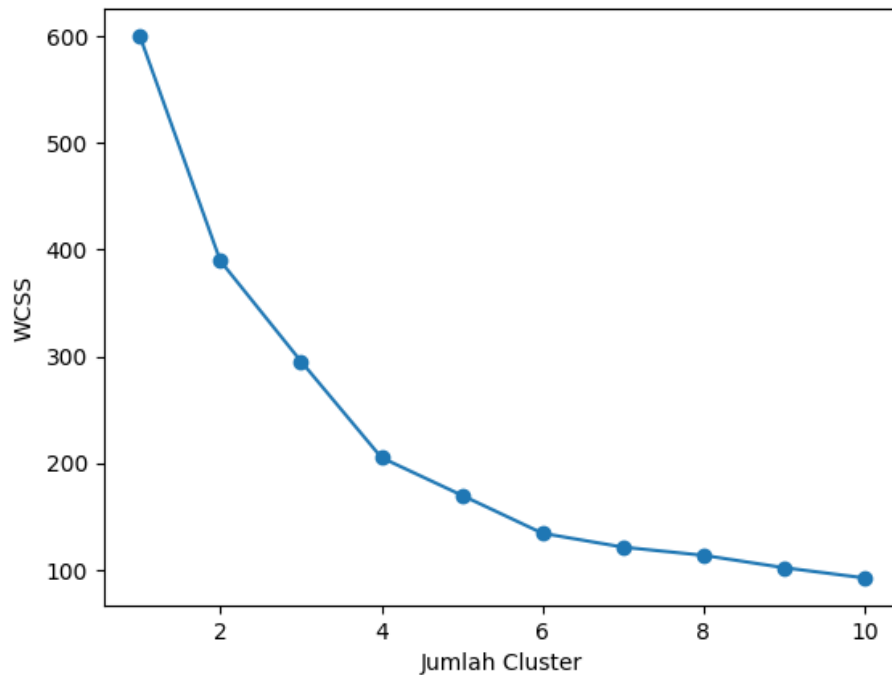
Gambar 2. Lingkungan pengolahan data

Analisis hasil clustering dilakukan dengan menghitung nilai rata-rata setiap atribut pada masing-masing kluster untuk mengidentifikasi karakteristik pelanggan. Setiap kluster kemudian diinterpretasikan berdasarkan kombinasi atribut usia, pendapatan tahunan, dan skor pengeluaran guna memperoleh gambaran segmentasi pelanggan yang lebih jelas. Hasil analisis ini digunakan sebagai dasar dalam pembahasan segmentasi pelanggan serta interpretasi pola pembelian yang dapat mendukung pengambilan keputusan strategis di bidang pemasaran dan manajemen pelanggan.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Hasil penelitian diperoleh melalui proses clustering menggunakan algoritma K-Means terhadap data pelanggan yang telah melalui tahap pra-pemrosesan dan normalisasi. Penentuan jumlah kluster optimal dilakukan menggunakan Elbow Method dan Silhouette Score untuk memastikan kualitas pengelompokan yang dihasilkan. Berdasarkan hasil evaluasi kedua metode tersebut, jumlah kluster optimal yang digunakan dalam penelitian ini adalah empat kluster, sebagaimana ditunjukkan pada Gambar 1.



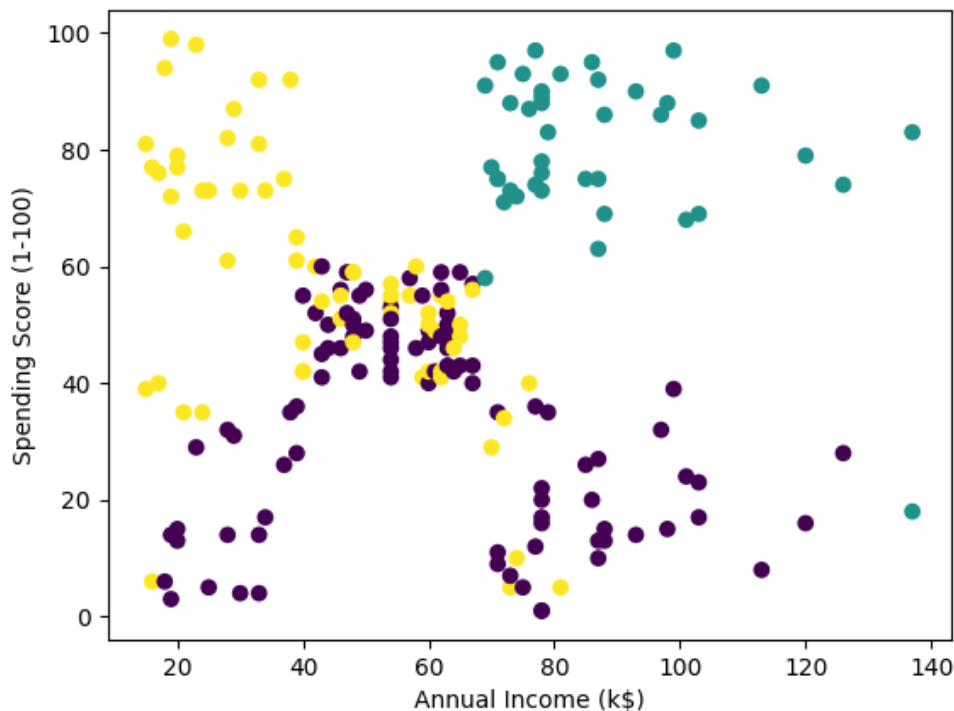
Gambar 1. Grafik Elbow Method

Proses clustering dengan empat kluster menghasilkan pengelompokan pelanggan berdasarkan atribut usia, pendapatan tahunan, dan skor pengeluaran. Ringkasan nilai rata-rata setiap atribut pada masing-masing kluster disajikan pada Tabel 1.

Tabel 1. Rata-rata Atribut Setiap Kluster

Kluster	Usia	Pendapatan Tahunan (k\$)	Skor Pengeluaran
0	53.98	47.71	39.97
1	32.88	86.10	81.53
2	25.44	40.00	60.30
3	39.37	86.50	19.58

Untuk memperjelas hasil pengelompokan, dilakukan visualisasi segmentasi pelanggan menggunakan diagram sebar berdasarkan atribut pendapatan tahunan dan skor pengeluaran, sebagaimana ditunjukkan pada Gambar 2. Visualisasi tersebut menunjukkan pemisahan kluster yang cukup jelas, yang mengindikasikan bahwa algoritma K-Means mampu membentuk kelompok pelanggan dengan karakteristik yang berbeda secara signifikan.



Gambar 2. Visualisasi Segmentasi Pelanggan Menggunakan Metode K-Means

3.2 Pembahasan

Hasil clustering menunjukkan bahwa setiap klaster memiliki karakteristik pelanggan yang berbeda berdasarkan kombinasi atribut usia, pendapatan tahunan, dan skor pengeluaran. Klaster 0 didominasi oleh pelanggan dengan usia relatif lebih tua, tingkat pendapatan menengah, serta skor pengeluaran yang rendah. Karakteristik ini mengindikasikan bahwa pelanggan pada klaster ini cenderung memiliki pola pembelian yang konservatif dan kurang responsif terhadap aktivitas promosi, sebagaimana juga ditemukan pada penelitian segmentasi pelanggan sebelumnya (Christy *et al.*, 2021).

Klaster 1 merepresentasikan pelanggan dengan pendapatan dan skor pengeluaran tertinggi dibandingkan klaster lainnya. Kelompok ini dapat dikategorikan sebagai pelanggan bernilai tinggi (*high-value customers*) dengan tingkat aktivitas pembelian yang intensif. Temuan ini sejalan dengan penelitian Wang (2022) dan Kasem, Hamada and Taj-Eddin (2024) yang menyatakan bahwa pelanggan dengan kombinasi pendapatan dan pengeluaran tinggi merupakan target utama dalam strategi pemasaran berbasis data dan program loyalitas pelanggan.

Klaster 2 terdiri dari pelanggan berusia relatif muda dengan tingkat pendapatan rendah hingga menengah, namun memiliki skor pengeluaran yang cukup tinggi. Kondisi ini menunjukkan adanya kecenderungan perilaku belanja yang aktif meskipun daya beli belum optimal. Segmentasi ini mengindikasikan potensi pengembangan pelanggan melalui strategi promosi, diskon, atau kampanye pemasaran digital yang tepat sasaran, sebagaimana disarankan oleh Nguyen (2021) dalam studi segmentasi pelanggan berbasis perilaku.

Klaster 3 menunjukkan pelanggan dengan tingkat pendapatan tinggi tetapi skor pengeluaran yang rendah. Pola ini mengindikasikan adanya peluang bisnis

yang belum dimanfaatkan secara optimal. Pelanggan pada klaster ini berpotensi untuk ditingkatkan nilai transaksinya melalui pendekatan pemasaran yang lebih personal dan berbasis preferensi pelanggan, sebagaimana direkomendasikan oleh Joung and Kim (2023) dalam konteks interpretasi segmentasi pelanggan.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan pelanggan berdasarkan karakteristik dan pola pembelian secara efektif. Segmentasi yang dihasilkan tidak hanya memberikan gambaran struktural mengenai profil pelanggan, tetapi juga memberikan implikasi praktis bagi pengambilan keputusan bisnis dan perumusan strategi pemasaran yang lebih tepat sasaran dan berbasis data.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Penelitian ini bertujuan untuk melakukan segmentasi pelanggan berdasarkan karakteristik usia, pendapatan tahunan, dan pola pengeluaran menggunakan metode *K-Means clustering*. Berdasarkan hasil pengujian jumlah cluster menggunakan *Silhouette Score*, diperoleh jumlah cluster optimal sebanyak empat cluster yang mampu merepresentasikan perbedaan karakteristik pelanggan secara cukup jelas.

Hasil pengelompokan menunjukkan bahwa setiap cluster memiliki ciri yang berbeda, baik dari sisi usia rata-rata, tingkat pendapatan, maupun skor pengeluaran. Temuan ini mengindikasikan bahwa metode *K-Means* dapat digunakan secara efektif untuk mengelompokkan pelanggan berdasarkan pola pembelian, sehingga memberikan gambaran segmentasi yang dapat dimanfaatkan sebagai dasar pengambilan keputusan strategis, khususnya dalam perencanaan pemasaran dan penentuan target pelanggan.

Dengan demikian, metode *K-Means* terbukti mampu menghasilkan segmentasi pelanggan yang relevan dan mudah diinterpretasikan, terutama ketika diterapkan pada data pelanggan dengan atribut numerik yang jelas.

4.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian selanjutnya. Pertama, penelitian berikutnya dapat menambahkan atribut lain seperti frekuensi transaksi, jenis produk yang dibeli, atau data historis pembelian agar hasil segmentasi menjadi lebih komprehensif. Kedua, penggunaan metode clustering lain seperti *Hierarchical Clustering* atau *DBSCAN* dapat dilakukan sebagai pembandingan untuk mengevaluasi efektivitas metode *K-Means*.

Selain itu, hasil segmentasi yang diperoleh dalam penelitian ini dapat dikembangkan lebih lanjut dengan mengintegrasikannya ke dalam sistem pendukung keputusan atau strategi pemasaran berbasis data, sehingga dapat memberikan manfaat yang lebih aplikatif bagi pihak pengelola atau pelaku bisnis.

5. DAFTAR REFERENSI

- Abdi, H. and Williams, L.J. (2010) 'Principal component analysis', *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), pp. 433–459.
<https://doi.org/10.1002/wics.101>

- Barrera, F., Segura, M. and Maroto, C. (2024) 'Multiple criteria decision support system for customer segmentation using a sorting outranking method', *Expert Systems with Applications*, 238, 122310. <https://doi.org/10.1016/j.eswa.2023.122310>
- Chen, H., Zhang, L., Chu, X. and Yan, B. (2019) 'Smartphone customer segmentation based on the usage pattern', *Advanced Engineering Informatics*, 42, 101000. <https://doi.org/10.1016/j.aei.2019.101000>
- Christy, A.J., Umamakeswari, A., Priyatharsini, L. and Neyaa, A. (2021) 'RFM ranking-An effective approach to customer segmentation', *Journal of King Saud University - Computer and Information Sciences*, 33(10), pp. 1251-1257. <https://doi.org/10.1016/j.jksuci.2018.09.004>
- Dong, S., Wang, P. and Abbas, K. (2021) 'A survey on deep learning and its applications', *Computer Science Review*, 40, 100379. <https://doi.org/10.1016/j.cosrev.2021.100379>
- Joung, J. and Kim, H. (2023) 'Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews', *International Journal of Information Management*, 70, 102641. <https://doi.org/10.1016/j.ijinfomgt.2023.102641>
- Kasem, M.S., Hamada, M. and Taj-Eddin, I. (2024) 'Customer profiling, segmentation, and sales prediction using AI in direct marketing', *Neural Computing and Applications*, 36(9), pp. 4995-5005. <https://doi.org/10.1007/s00521-023-09192-5>
- Mukhamediev, R.I. et al. (2022) 'Review of artificial intelligence and machine learning technologies: classification, restrictions, opportunities and challenges', *Mathematics*, 10(15), 2552. <https://doi.org/10.3390/math10152552>
- Nguyen, S.P. (2021) 'Deep customer segmentation with applications to a Vietnamese supermarkets' data', *Soft Computing*, 25(12), pp. 7785-7793. <https://doi.org/10.1007/s00500-020-05559-5>
- Rungruang, C. et al. (2024) 'RFM model customer segmentation based on hierarchical approach using FCA', *Expert Systems with Applications*, 237, 121449. <https://doi.org/10.1016/j.eswa.2023.121449>
- Syakur, M.A., Khotimah, B.K., Rochman, E.M.S. and Satoto, B.D. (2018) 'Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster', *IOP Conference Series: Materials Science and Engineering*, 336(1), 012017. <https://doi.org/10.1088/1757-899X/336/1/012017>
- Wang, C. (2022) 'Efficient customer segmentation in digital marketing using deep learning with swarm intelligence approach', *Information Processing & Management*, 59(6), 103085. <https://doi.org/10.1016/j.ipm.2022.103085>
- Yadegaridehkordi, E. et al. (2021) 'Customers segmentation in eco-friendly hotels using multi-criteria and machine learning techniques', *Technology in Society*, 65, 101528. <https://doi.org/10.1016/j.techsoc.2021.101528>
- Dolnicar, S., Kaiser, S., Lazarevski, K. and Leisch, F. (2012) 'Biclustering: Overcoming data dimensionality problems in market segmentation', *Journal of Travel Research*, 51(1), 41-59. <https://doi.org/10.1177/0047287510394192>

Likas, A., Vlassis, N. and Verbeek, J.J. (2003) 'The global k-means clustering algorithm', *Pattern Recognition*, 36(2), pp. 451–461.
[https://doi.org/10.1016/S0031-3203\(02\)00060-2](https://doi.org/10.1016/S0031-3203(02)00060-2)