

Segmentasi Pelanggan Berdasarkan Kebutuhan Primer Skunder dan Tersier Menggunakan K-Means Clustering

Caesaliana Indah Mu'assaroh^{1*}, M. Rizieq Sultan Arkan², Galuh Sitoresmi³,
Sabrina Rizkiya⁴, Zida Nadhifah Aulia Kencana⁵

^{1,3,4,5}Program Studi Bisnis Digital, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung

²Program Studi Sains Data, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung

Email: ¹caesaliana.2312140015@mail.darmajaya.ac.id, ²zieqsultan.2311080022@mail.darmajaya.ac.id,

³galuh.2312140027@mail.darmajaya.ac.id, ⁴sabrina25.2312140028@mail.darmajaya.ac.id,

⁵zidandhfa.2312140030@mail.darmajaya.ac.id

* corresponding author: caesaliana.2312140015@mail.darmajaya.ac.id



Article History:

Received : 10-09-2025

Revised : 23-10-2025

Accepted : 30-11-2025

Online : 30-11-2025



This is an open access article under the
CC-BY-SA license

ABSTRACT

Dalam era pemasaran digital, perusahaan dituntut untuk memahami perilaku pelanggan secara mendalam agar dapat menyusun strategi yang tepat sasaran. Segmentasi pelanggan merupakan teknik yang umum digunakan untuk mengelompokkan pelanggan berdasarkan kemiripan karakteristik dan perilaku konsumsi mereka. Penelitian ini bertujuan untuk mengidentifikasi segmen pelanggan berdasarkan data perilaku pembelian menggunakan teknik pembelajaran tanpa pengawasan (unsupervised learning) dengan algoritma K-Means Clustering. Dataset yang digunakan diambil dari Kaggle dan mencakup 2.240 data pelanggan dengan atribut demografis dan perilaku belanja. Enam fitur utama yang dianalisis adalah Income, Age, TotalChildren, MntMeatProducts, NumCatalogPurchases, dan Recency. Hasil segmentasi menunjukkan keberadaan kelompok pelanggan dengan karakteristik dan kecenderungan konsumsi yang berbeda, yang dapat dimanfaatkan untuk menyusun strategi pemasaran yang lebih personal dan efisien.

Kata Kunci: Segmentasi Pelanggan, Clustering, Perilaku Konsumen, Unsupervised Learning, Analisis Data.

ABSTRACT

In the digital marketing era, companies are required to deeply understand customer behavior in order to develop targeted strategies. Customer segmentation is a common technique used to group customers based on similarities in their characteristics and consumption behaviors. This study aims to identify customer segments using unsupervised learning techniques with the K-Means clustering algorithm. The dataset, obtained from Kaggle, contains 2,240 customer records with demographic and purchase behavior attributes. The six primary features analyzed include Income, Age, TotalChildren, MntMeatProducts, NumCatalogPurchases, and Recency. The clustering results reveal distinct customer groups with different characteristics and purchasing tendencies, which can be used to develop more personalized and efficient marketing strategies.

Keywords: Customer segmentation; clustering; consumer behavior; unsupervised learning; data analysis

1. PENDAHULUAN

Perkembangan teknologi informasi telah mengubah secara fundamental cara perusahaan berinteraksi dengan pelanggan. Di tengah persaingan bisnis yang semakin kompetitif, pemahaman terhadap perilaku konsumen menjadi kunci utama dalam menyusun strategi pemasaran yang efektif dan relevan. Salah satu pendekatan yang banyak digunakan dalam memahami karakteristik pelanggan adalah segmentasi pasar, yaitu proses membagi konsumen ke dalam kelompok-kelompok berdasarkan kesamaan atribut seperti usia, pendapatan, preferensi belanja, dan perilaku konsumsi.

Segmentasi memungkinkan perusahaan untuk menargetkan pelanggan secara lebih tepat, baik dalam hal produk, promosi, maupun pelayanan. Contohnya, pelanggan dapat dibagi ke dalam kelompok seperti pembeli loyal, pemburu diskon, dan pembeli musiman. Penerapan segmentasi pasar yang tepat membantu

perusahaan mengenali peluang serta mengembangkan strategi yang sesuai dengan kebutuhan setiap kelompok pelanggan [10].

Seiring dengan pertumbuhan data digital, perusahaan kini memiliki akses terhadap data pelanggan yang sangat besar meliputi informasi demografis, riwayat pembelian, dan pola konsumsi. Data tersebut dapat dimanfaatkan secara lebih efisien menggunakan metode data mining, yang merupakan teknik dalam ilmu komputer untuk menggali informasi dan pola tersembunyi dari kumpulan data besar [8].

Dalam konteks segmentasi pelanggan, salah satu teknik data mining yang paling umum digunakan adalah *clustering*, yaitu metode untuk mengelompokkan pelanggan berdasarkan kesamaan tertentu tanpa perlu label atau klasifikasi awal. Algoritma yang sering digunakan dalam clustering adalah K-Means karena kemampuannya dalam mengelompokkan data besar secara efisien. Selain itu, terdapat juga alternatif seperti DBSCAN dan Agglomerative Hierarchical Clustering yang digunakan dalam beberapa kasus tergantung struktur data yang dihadapi.

Di Indonesia, segmentasi pelanggan berbasis algoritma clustering telah menjadi topik penelitian yang cukup banyak dibahas. Misalnya Nabila Dea Azahra dkk. meneliti penggunaan algoritma DBSCAN dan K-Means untuk mengelompokkan pelanggan berdasarkan data demografis dan konsumsi seperti pengeluaran untuk daging, buah, dan produk mewah. Hasilnya, DBSCAN mampu membentuk empat kluster utama dan mendeteksi data pencilan. Setiap kluster menunjukkan pola konsumsi yang khas dan dapat digunakan untuk strategi pemasaran yang lebih terarah [3].

Furqan M. Pranata dkk. menganalisis performa K-Means dan DBSCAN dalam segmentasi pelanggan dengan pendekatan RFM menggunakan lima *dataset* dari sektor retail. Evaluasi menggunakan Silhouette Score dan Davies-Bouldin Index menunjukkan bahwa performa algoritma tergantung pada karakteristik dataset. K-Means unggul pada beberapa dataset, sementara DBSCAN lebih baik pada dataset lain, menunjukkan pentingnya pemilihan algoritma yang sesuai dengan pola data [4].

Nita Mirantika dan Estiko Rijanto membandingkan K-Means dan K-Medoids untuk segmentasi pelanggan pada data penjualan produk peternakan menggunakan pendekatan RFM. K-Means menghasilkan tiga kluster pelanggan aktif, rata-rata, dan pasif dengan performa lebih baik berdasarkan Davies-Bouldin Index. K-Medoids dianggap kurang optimal karena kluster yang dihasilkan tumpang tindih dan tidak jelas [11].

Elmira Faustina Achmal dkk. mengeksplorasi Kernel K-Means dalam segmentasi pelanggan Smartlegal.id. Algoritma ini mampu menangkap struktur kluster non-linear dan menghasilkan dua kluster utama dengan nilai Silhouette Score 0,9035, yang menunjukkan segmentasi sangat baik. Kernel K-Means terbukti lebih unggul dibanding K-Means biasa dalam hal kejelasan pemisahan antar kluster [2].

Keempat penelitian tersebut menunjukkan bahwa pemilihan algoritma clustering sangat memengaruhi keberhasilan dalam segmentasi pelanggan. K-Means tetap menjadi metode yang efisien dan banyak digunakan, terutama pada dataset dengan distribusi homogen, sedangkan algoritma seperti DBSCAN dan Kernel K-Means menunjukkan keunggulan pada data yang lebih kompleks atau non-linier.

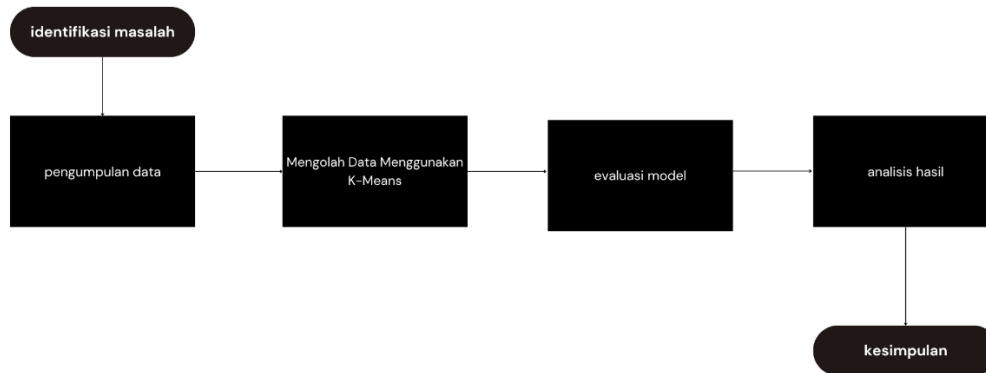
Berangkat dari latar belakang tersebut, penelitian ini bertujuan untuk membangun model segmentasi pelanggan berdasarkan kebutuhan konsumsi yang diklasifikasikan menjadi kebutuhan primer, sekunder, dan tersier. Pendekatan ini membantu perusahaan memahami pola konsumsi secara lebih mendalam dan menyusun strategi pemasaran yang disesuaikan dengan karakteristik setiap segmen.

Penelitian ini menggunakan metode unsupervised learning dengan algoritma K-Means Clustering. Dataset yang digunakan adalah Customer Personality Analysis dari Kaggle, yang berisi informasi demografis, perilaku pembelian, serta tingkat keterlibatan pelanggan. Dataset ini menjadi dasar kuat untuk segmentasi karena mencakup atribut-atribut yang mencerminkan kebutuhan konsumsi secara nyata.

Berbeda dengan pendekatan berbasis atribut umum seperti RFM yang digunakan pada sebagian besar penelitian terdahulu, studi ini secara eksplisit mengklasifikasikan fitur berdasarkan kategori kebutuhan. Misalnya, pengeluaran untuk produk daging masuk ke dalam kebutuhan primer, sementara pembelian katalog mencerminkan kebutuhan sekunder atau tersier. Dengan pendekatan ini, hasil segmentasi menjadi lebih aplikatif dalam menyusun strategi promosi dan pengelolaan anggaran pemasaran yang lebih terarah.

2. METODE

Penelitian ini menggunakan pendekatan unsupervised learning dengan algoritma K-Means Clustering untuk melakukan segmentasi pelanggan berdasarkan kebutuhan konsumsi, yang diklasifikasikan menjadi kebutuhan primer, sekunder, dan tersier.



Gambar 1. Flowchart Metodologi

2.1. Akuisisi dan Pemahaman Data

Langkah awal dalam proses penelitian ini adalah melakukan akuisisi dan pemahaman data. Akuisisi data merujuk pada proses pengumpulan data dari sumber yang relevan, sedangkan pemahaman data bertujuan untuk menelaah isi, struktur, serta karakteristik data tersebut agar dapat digunakan secara tepat dalam tahap analisis lanjutan.

Pada penelitian ini, data diambil dari platform Kaggle, sebuah situs populer yang menyediakan beragam dataset terbuka untuk keperluan analisis data dan pembelajaran mesin. Dataset yang digunakan berjudul *Customer Personality Analysis*, yang memuat informasi mengenai karakteristik demografis dan perilaku pembelian dari 2.240 pelanggan. Dataset ini dapat diakses melalui tautan berikut: <https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis>.

Setelah proses pengunduhan data dilakukan, tahap berikutnya adalah memahami secara menyeluruh setiap atribut atau kolom yang tersedia dalam dataset tersebut. Hal ini dilakukan untuk mengetahui jenis data yang digunakan, cakupan informasi yang dimuat, serta relevansinya terhadap tujuan utama penelitian, yaitu segmentasi pelanggan berdasarkan kebutuhan primer, sekunder, dan tersier. Beberapa atribut bersifat numerik, seperti pendapatan tahunan dan jumlah pengeluaran untuk produk tertentu, sementara lainnya bersifat kategorikal, seperti status pernikahan dan tingkat pendidikan.

Dengan melakukan pemahaman data sejak awal, peneliti dapat mengidentifikasi fitur-fitur mana yang memiliki kontribusi besar terhadap pembentukan kluster pelanggan, serta menentukan fitur mana yang perlu diolah, digabungkan, atau dihapus. Tahap ini menjadi pondasi penting sebelum memasuki proses praproses data, karena pemahaman yang baik terhadap struktur dan isi dataset akan sangat berpengaruh terhadap ketepatan dan keberhasilan analisis selanjutnya.

Setelah pemahaman data dilakukan, semua fitur dalam dataset didokumentasikan dalam bentuk tabel yang memuat nama fitur, tipe data, serta penjelasan singkat dari masing-masing atribut, sebagaimana ditunjukkan pada bagian berikut.

Tabel 1. Fitur-Fitur dalam Dataset

Nama Fitur	Tipe Data	Penjelasan
ID	Integer	ID unik untuk setiap pelanggan.
Year_Birth	Integer	Tahun kelahiran pelanggan.
Education	String	Tingkat pendidikan pelanggan (misalnya: Graduation, PhD).
Marital_Status	String	Status pernikahan pelanggan (misalnya: Single, Married).
Income	Float	Pendapatan tahunan pelanggan (dalam satuan mata uang yang tidak ditentukan).
Kidhome	Integer	Jumlah anak kecil di rumah pelanggan.
Teenhome	Integer	Jumlah remaja di rumah pelanggan.
Dt_Customer	String	Tanggal pelanggan mulai menjadi pelanggan (format: dd-mm-yyyy).
Recency	Integer	Berapa hari sejak terakhir kali pelanggan melakukan pembelian.
MntWines	Integer	Jumlah uang yang dihabiskan untuk produk wine dalam dua tahun terakhir.
MntFruits	Integer	Jumlah uang yang dihabiskan untuk buah-buahan.
MntMeatProducts	Integer	Jumlah uang yang dihabiskan untuk produk daging.

MntFishProducts	Integer	Jumlah uang yang dihabiskan untuk produk ikan.
MntSweetProducts	Integer	Jumlah uang yang dihabiskan untuk makanan manis.
MntGoldProds	Integer	Jumlah uang yang dihabiskan untuk produk emas.
NumDealsPurchases	Integer	Jumlah pembelian yang dilakukan selama promosi/penawaran.
NumWebPurchases	Integer	Jumlah pembelian melalui website.
NumCatalogPurchases	Integer	Jumlah pembelian melalui katalog.
NumStorePurchases	Integer	Jumlah pembelian langsung di toko.
NumWebVisitsMonth	Integer	Jumlah kunjungan ke website dalam sebulan.
AcceptedCmp1	Integer	Apakah pelanggan menerima kampanye 1 (0 = tidak, 1 = ya).
AcceptedCmp2	Integer	Sama seperti di atas untuk kampanye 2.
AcceptedCmp3	Integer	Sama seperti di atas untuk kampanye 3.
AcceptedCmp4	Integer	Sama seperti di atas untuk kampanye 4.
AcceptedCmp5	Integer	Sama seperti di atas untuk kampanye 5.
Complain	Integer	Apakah pelanggan pernah mengajukan keluhan (0 = tidak, 1 = ya).
Z_CostContact	Integer	Nilai tetap (semua 3); bisa diabaikan.
Z_Revenue	Integer	Nilai tetap (semua 11); bisa diabaikan.
Response	Integer	Apakah pelanggan merespons kampanye pemasaran terakhir (0 = tidak, 1 = ya).

2.2. Praproses Data

Praproses Data merupakan tahapan penting dalam penelitian ini yang bertujuan untuk menyiapkan data mentah menjadi data yang siap dianalisis dengan algoritma machine learning, khususnya K-Means Clustering. Praproses data sendiri dapat diartikan sebagai proses awal yang dilakukan terhadap data sebelum masuk ke tahap analisis inti, dengan tujuan utama untuk memastikan bahwa data yang digunakan bersih, lengkap, dan relevan terhadap tujuan penelitian. Tanpa tahapan ini, hasil segmentasi yang diperoleh bisa menjadi tidak akurat karena terganggu oleh *noise*, data yang tidak lengkap, atau atribut yang tidak berkontribusi terhadap pemodelan.

Dalam penelitian ini, tahapan praproses data dimulai dengan menangani nilai-nilai yang hilang atau missing values. Kehadiran data yang kosong dapat mengganggu proses perhitungan dalam algoritma clustering, sehingga langkah awal yang dilakukan adalah mengidentifikasi data mana saja yang memiliki nilai kosong, lalu menggantinya dengan pendekatan tertentu seperti rerata (mean), median, atau modus tergantung pada tipe datanya. Dengan begitu, data yang digunakan menjadi lebih utuh dan tidak mengganggu perhitungan jarak antar data dalam proses klusterisasi.

Langkah berikutnya adalah melakukan kalkulasi usia pelanggan. Data awal yang tersedia hanya memberikan informasi tahun lahir (*Year_Birth*), sehingga perlu diolah terlebih dahulu untuk mendapatkan usia aktual dengan mengurangkan tahun saat ini dengan tahun lahir tersebut. Usia pelanggan merupakan variabel yang penting karena berkaitan erat dengan kebiasaan konsumsi dan preferensi produk, yang akan menjadi dasar dalam segmentasi berdasarkan kebutuhan primer, sekunder, dan tersier.

Selanjutnya, peneliti melakukan proses penghapusan terhadap fitur-fitur yang dianggap tidak relevan untuk analisis. Beberapa fitur seperti ID, *Z_CostContact*, dan *Z_Revenue* dihilangkan dari dataset karena tidak memberikan kontribusi apapun terhadap perilaku konsumsi pelanggan. Fitur ID hanya merupakan penanda unik yang tidak memiliki makna analitis, sedangkan *Z_CostContact* dan *Z_Revenue* memiliki nilai yang konstan di seluruh data sehingga tidak memberikan informasi variasi yang dapat digunakan oleh algoritma clustering.

Dengan rangkaian proses praproses ini—yang meliputi penanganan nilai kosong, transformasi data (penghitungan usia), serta penghapusan atribut yang tidak relevan—data yang digunakan dalam penelitian ini menjadi lebih terstruktur, bersih, dan siap untuk digunakan dalam tahap selanjutnya yaitu seleksi fitur dan analisis clustering. Keberhasilan tahapan praproses ini sangat menentukan kualitas segmentasi pelanggan yang dihasilkan, karena algoritma K-Means sangat sensitif terhadap input data yang tidak konsisten atau tidak bermakna.

2.3. Seleksi Fitur

Langkah penting selanjutnya yang perlu dilakukan adalah seleksi fitur. Seleksi fitur merupakan proses pemilihan variabel atau atribut dalam dataset yang dianggap paling relevan dan berpengaruh terhadap tujuan analisis. Dalam konteks penelitian ini, yaitu segmentasi pelanggan berdasarkan kebutuhan konsumsi, pemilihan fitur yang tepat sangat krusial untuk menghasilkan kluster yang bermakna dan dapat ditindaklanjuti dalam strategi pemasaran.

Proses seleksi fitur diawali dengan telaah terhadap keseluruhan atribut dalam dataset *Customer Personality Analysis* yang diambil dari platform Kaggle. *Dataset* ini terdiri atas 29 atribut yang mencakup

informasi demografis, perilaku pembelian, hingga respons terhadap kampanye pemasaran. Tidak semua fitur dalam dataset digunakan, karena sebagian di antaranya bersifat konstan (seperti *Z_CostContact* dan *Z_Revenue*), merupakan identifier (seperti ID), atau tidak relevan dengan tujuan segmentasi konsumsi (seperti *Dt_Customer* atau fitur terkait kampanye pemasaran).

Untuk itu, peneliti melakukan proses seleksi berdasarkan dua pendekatan utama yaitu relevansi teoretis terhadap perilaku konsumsi pelanggan, serta hasil eksplorasi awal terhadap data. Berdasarkan pendekatan tersebut, enam fitur dipilih karena memiliki hubungan erat dengan konsumsi dan pola pembelian pelanggan.

Pertama adalah *Income*, yaitu pendapatan tahunan pelanggan. Fitur ini sangat penting karena menunjukkan daya beli seseorang. Pelanggan dengan pendapatan tinggi cenderung memiliki pola konsumsi yang berbeda dibandingkan pelanggan berpendapatan rendah, baik dalam hal jumlah maupun jenis produk yang dibeli.

Kedua adalah *Age*, yang dihitung dari atribut *Year_Birth*. Usia menjadi faktor penting dalam mempengaruhi preferensi produk dan kebiasaan belanja. Misalnya, pelanggan muda cenderung lebih impulsif dan menyukai produk gaya hidup, sedangkan pelanggan berusia lebih tua mungkin lebih fokus pada kebutuhan pokok dan kestabilan.

Ketiga adalah *TotalChildren*, yaitu hasil penjumlahan dari *Kidhome* dan *Teenhome*. Fitur ini menunjukkan jumlah anak yang dimiliki pelanggan. Keberadaan anak dalam rumah tangga berdampak langsung pada pola pengeluaran, terutama pada kebutuhan primer seperti makanan dan produk rumah tangga.

Fitur keempat adalah *MntMeatProducts*, yang mencerminkan jumlah pengeluaran pelanggan untuk produk daging dalam dua tahun terakhir. Atribut ini dipilih sebagai representasi dari kebutuhan primer, karena makanan merupakan kebutuhan dasar yang pasti dikonsumsi oleh setiap individu atau keluarga.

Fitur kelima adalah *NumCatalogPurchases*, yaitu jumlah pembelian melalui katalog. Perilaku ini dianggap mencerminkan konsumsi terhadap kebutuhan sekunder atau tersier, karena biasanya pembelian melalui katalog berkaitan dengan produk non-esensial seperti pakaian, aksesoris, atau barang mewah.

Terakhir, fitur *Recency* digunakan untuk mengukur seberapa aktif pelanggan dalam berinteraksi dengan perusahaan, yaitu dengan melihat jumlah hari sejak terakhir kali mereka melakukan pembelian. Pelanggan yang baru saja melakukan pembelian diasumsikan masih memiliki ketertarikan terhadap produk, sementara yang sudah lama tidak bertransaksi mungkin mulai tidak aktif atau berpaling ke kompetitor.

Dengan mempertimbangkan relevansi fitur terhadap kebutuhan dasar, pelengkap, dan mewah, serta peran fitur tersebut dalam menggambarkan perilaku konsumsi, keenam atribut tersebut dipilih untuk dijadikan dasar dalam proses clustering. Pemilihan fitur ini bertujuan agar hasil segmentasi tidak hanya menggambarkan pola yang matematis, tetapi juga mencerminkan perilaku nyata yang dapat dijadikan dasar strategi pemasaran yang tepat sasaran.

2.4. Penjelasan K-Means

Pada tahapan analisis segmentasi pelanggan dalam penelitian ini, digunakan algoritma K-Means Clustering sebagai metode utama. K-Means merupakan salah satu algoritma *unsupervised learning* yang paling banyak digunakan untuk mengelompokkan data berdasarkan kemiripan antar objek. Dalam konteks segmentasi pelanggan, algoritma ini berperan penting dalam mengelompokkan pelanggan ke dalam beberapa segmen berdasarkan karakteristik konsumsi dan demografis tertentu.

Kinerja K-Means didasarkan pada proses iteratif. Algoritma ini dimulai dengan menentukan jumlah kluster yang diinginkan, kemudian secara acak memilih titik awal sebagai pusat kluster atau yang disebut dengan centroid. Setiap data dalam dataset kemudian dihitung jaraknya ke seluruh centroid, dan akan dikelompokkan ke dalam kluster dengan centroid terdekat. Setelah semua data dikelompokkan, posisi centroid akan diperbarui berdasarkan rata-rata posisi semua data yang berada dalam kluster tersebut. Proses ini diulang terus-menerus sampai posisi centroid tidak mengalami perubahan yang signifikan, atau dengan kata lain telah mencapai konvergensi.

Dalam penelitian ini, proses klusterisasi K-Means dilakukan terhadap enam fitur utama yang dipilih dari dataset, yaitu *Income*, *Age*, *TotalChildren*, *MntMeatProducts*, *NumCatalogPurchases*, dan *Recency*. Keenam fitur ini dipilih karena merepresentasikan gabungan antara data demografis dan perilaku pembelian pelanggan. *Income* digunakan untuk melihat daya beli, *Age* merepresentasikan perubahan preferensi konsumsi berdasarkan usia, *TotalChildren* menggambarkan tanggungan keluarga, *MntMeatProducts* sebagai representasi pengeluaran untuk kebutuhan pokok, *NumCatalogPurchases* mencerminkan minat terhadap barang sekunder atau tersier, dan *Recency* menunjukkan seberapa aktif pelanggan tersebut dalam melakukan pembelian terbaru.

Untuk menentukan jumlah kluster yang paling optimal, digunakan metode Elbow yang didasarkan pada nilai inerti atau within-cluster sum of square. Melalui visualisasi grafik Elbow, titik tekuk atau "elbow" diamati untuk mengetahui nilai K yang memberikan efisiensi terbaik tanpa menyebabkan overfitting. Berdasarkan hasil analisis, nilai K yang paling tepat adalah tiga, karena pada titik ini penurunan inerti mulai melambat, menandakan bahwa menambah jumlah kluster selanjutnya tidak memberikan peningkatan signifikan dalam kualitas segmentasi.

Dengan menggunakan K-Means, pelanggan berhasil dikelompokkan ke dalam tiga segmen berbeda yang memiliki karakteristik konsumsi dan profil demografis yang bervariasi. Segmentasi ini diharapkan mampu menjadi dasar dalam menyusun strategi pemasaran yang lebih tepat sasaran dan personal bagi masing-masing kelompok pelanggan.

2.5 Penjelasan PCA

Sebelum menjalankan proses klusterisasi dengan algoritma K-Means, dilakukan tahap reduksi dimensi menggunakan metode Principal Component Analysis (PCA). PCA adalah teknik analisis statistik yang bertujuan untuk menyederhanakan data berdimensi tinggi menjadi dimensi yang lebih rendah, namun tetap mempertahankan sebanyak mungkin informasi penting dari data asli. Dalam konteks ini, PCA sangat membantu dalam meningkatkan efisiensi dan akurasi dari proses clustering, terutama ketika data memiliki banyak fitur yang saling berkorelasi.

PCA bekerja dengan mentransformasikan fitur-fitur asli ke dalam sekumpulan komponen baru yang disebut dengan principal components. Komponen-komponen ini merupakan kombinasi linier dari fitur asli yang disusun berdasarkan kontribusinya terhadap total variasi data. Komponen pertama (PC1) memuat informasi paling banyak, diikuti oleh komponen kedua (PC2), dan seterusnya. Dengan memilih beberapa komponen utama pertama saja, kita dapat menyederhanakan data tanpa kehilangan pola-pola penting yang terkandung di dalamnya.

Dalam penelitian ini, PCA digunakan untuk mereduksi data yang telah dinormalisasi sebelum dimasukkan ke dalam algoritma K-Means. Proses ini tidak hanya membantu mempercepat komputasi, tetapi juga memudahkan dalam visualisasi hasil klusterisasi. Hasil reduksi PCA divisualisasikan dalam bentuk scatter plot dua dimensi, yang memperlihatkan distribusi setiap pelanggan dalam ruang dua komponen utama. Visualisasi ini memberikan gambaran yang jelas mengenai pemisahan antar kluster, serta mempermudah interpretasi hasil segmentasi.

Selain meningkatkan visualisasi, PCA juga membantu meningkatkan performa K-Means dengan menghilangkan fitur-fitur yang memiliki kontribusi rendah terhadap variasi data. Dalam studi yang dilakukan oleh Gandomi dan Haider [6], penggunaan PCA terbukti dapat memperbaiki kualitas segmentasi pada kasus-kasus data pemasaran yang kompleks. Hal ini selaras dengan hasil yang diperoleh dalam penelitian ini, di mana hasil clustering menunjukkan pola pemisahan kluster yang cukup jelas dan mendukung interpretasi segmen pelanggan berdasarkan kebutuhan primer, sekunder, dan tersier.

Dengan demikian, penerapan PCA dalam penelitian ini berfungsi ganda, yakni sebagai alat bantu untuk meningkatkan efektivitas algoritma clustering dan sebagai pendekatan visualisasi untuk memperjelas hasil segmentasi pelanggan.

3. HASIL DAN PEMBAHASAN

Pada tahap ini, hasil dari proses segmentasi pelanggan menggunakan algoritma K-Means Clustering akan dibahas secara terperinci. Selain itu, akan dijelaskan juga karakteristik setiap kelompok pelanggan (kluster) yang terbentuk, serta implikasi dari hasil segmentasi tersebut terhadap strategi pemasaran.

3.1. Evaluasi Kinerja Clustering

Dalam penelitian ini, proses segmentasi dilakukan menggunakan algoritma K-Means Clustering, yang merupakan salah satu metode unsupervised learning paling umum digunakan untuk mengelompokkan data. Sebelum dilakukan proses segmentasi pelanggan, tahap awal yang dilakukan adalah analisis statistik deskriptif terhadap fitur-fitur utama yang digunakan. Analisis ini bertujuan untuk memahami karakteristik dasar dari data pelanggan, serta memastikan bahwa terdapat keragaman atau variasi yang cukup untuk mendukung pembentukan kluster.

Penelitian ini menganalisis sejumlah fitur utama yang mencerminkan karakteristik demografis dan perilaku konsumsi pelanggan. Fitur Income merepresentasikan pendapatan tahunan pelanggan yang mencerminkan daya beli mereka, sementara Age menunjukkan usia pelanggan yang dapat memengaruhi preferensi konsumsi. Total Children menggambarkan jumlah anak dalam rumah tangga yang berpengaruh terhadap kebutuhan dasar keluarga. Fitur Kebutuhan Primer mencakup total pengeluaran untuk produk daging, ikan, dan buah sebagai kebutuhan pokok, sedangkan Kebutuhan Sekunder meliputi konsumsi makanan manis, pembelian melalui toko fisik, dan katalog yang mencerminkan kebutuhan pelengkap.

Kebutuhan Tersier menunjukkan minat pelanggan terhadap barang mewah dan gaya hidup, seperti pembelian *online* dan produk emas. Terakhir, Recency mengukur jumlah hari sejak transaksi terakhir pelanggan, yang digunakan untuk menilai tingkat keterlibatan mereka dengan perusahaan.

Analisis deskriptif ini bertujuan untuk memberikan gambaran awal mengenai sebaran data, rentang nilai, dan variabilitas masing-masing fitur. Tabel berikut merupakan ringkasan statistik dari ketujuh fitur tersebut:

Tabel 2. Statistik Deskriptif

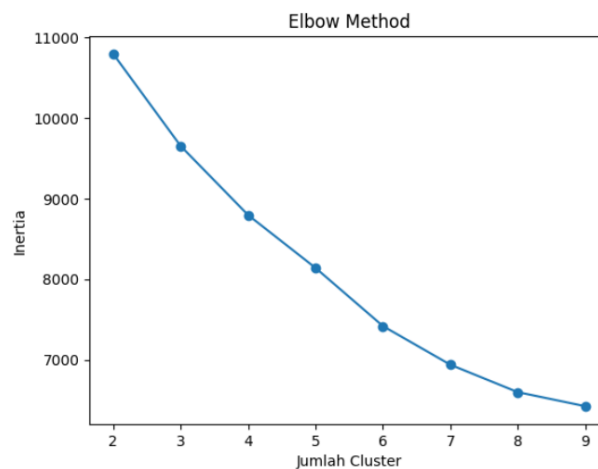
Fitur	Min	Max	Rata-rata	Standar Deviasi
Income	1.730	666.666	52.247,25	25.173,08
Age	29	132	56,18	11,99
Total Children	0	3	0,95	0,75
Kebutuhan Primer	1	1.727	230,99	285,19
Kebutuhan Sekunder	0	262	35,5	44,23
Kebutuhan Tersier	0	346	48,05	52,99
Recency	0	99	49,01	28,95

Dari tabel di atas, terlihat bahwa semua fitur memiliki sebaran nilai yang cukup lebar. Misalnya, pendapatan pelanggan berkisar antara Rp1.730 hingga lebih dari Rp666.000, dengan rata-rata sekitar Rp52.000. Selain itu, fitur-fitur konsumsi juga menunjukkan variasi besar, seperti kebutuhan primer yang berkisar antara Rp1 hingga lebih dari Rp1.700.

Untuk mendapatkan hasil yang optimal, data terlebih dahulu distandardisasi (distandardisasi) menggunakan StandardScaler. Hal ini dilakukan karena fitur-fitur seperti pendapatan dan usia memiliki skala nilai yang sangat berbeda. Jika tidak distandardisasi, fitur dengan skala lebih besar dapat mendominasi proses clustering.

Setelah data distandardisasi, peneliti menggunakan metode Elbow untuk menentukan jumlah kluster optimal. Metode ini bekerja dengan menghitung nilai *inertia* (jumlah kuadrat jarak data ke pusat kluster) untuk berbagai nilai K, dan kemudian memplotnya dalam grafik. Dari grafik Elbow yang dihasilkan, terlihat bahwa nilai *inertia* menurun tajam hingga K = 3, setelah itu penurunannya melambat.

Dengan demikian, nilai K = 3 dipilih sebagai jumlah kluster optimal. Pemilihan nilai ini memungkinkan segmentasi pelanggan yang efisien dan tetap informatif, tanpa membebani model dengan kluster yang terlalu banyak.



Gambar 2. Grafik Elbow

3.2. Analisis Karakteristik Setiap Klaster

Setelah proses clustering dilakukan, pelanggan terbagi menjadi tiga klaster berdasarkan kemiripan perilaku konsumsi dan atribut demografis. Karakteristik masing-masing klaster dianalisis berdasarkan tujuh fitur utama, yaitu pendapatan (Income), usia (Age), jumlah anak (TotalChildren), pengeluaran untuk kebutuhan primer (gabungan dari MntMeatProducts, MntFishProducts, dan MntFruits), kebutuhan sekunder (gabungan dari MntSweetProducts, NumStorePurchases, dan NumCatalogPurchases), kebutuhan tersier (gabungan dari MntGoldProds dan NumWebPurchases), serta Recency.

3.2.1. Klaster 0 – Pelanggan Kebutuhan Primer Tinggi

Klaster ini terdiri dari pelanggan yang memiliki total pengeluaran tinggi pada kebutuhan primer seperti daging, ikan, dan buah-buahan. Mereka cenderung memiliki jumlah anak yang lebih banyak dan nilai Recency yang rendah, yang berarti mereka masih tergolong pelanggan aktif karena baru saja melakukan pembelian. Pendapatan di klaster ini bervariasi, namun kecenderungan konsumsi mereka lebih berorientasi pada pemenuhan kebutuhan rumah tangga.

Strategi Pemasaran yang tepat adalah dengan menawarkan program loyalitas keluarga, promosi bahan pokok bulanan, serta diskon untuk pelanggan yang aktif berbelanja kebutuhan sehari-hari.

3.2.2. Klaster 1 – Pelanggan Berorientasi Tersier

Pelanggan dalam klaster ini memiliki pendapatan yang relatif tinggi dan menunjukkan minat besar terhadap kebutuhan tersier seperti produk emas dan pembelian online. Mereka juga cenderung aktif berbelanja melalui katalog, menunjukkan preferensi terhadap produk premium dan gaya hidup mewah. Namun, nilai Recency mereka tinggi, menandakan bahwa mereka sudah cukup lama tidak melakukan pembelian, dan mungkin perlu ditarik kembali ke dalam siklus pembelian.

Strategi Pemasaran yang dapat dilakukan adalah Promosi eksklusif dan penawaran premium seperti early access atau membership card bisa menjadi pendekatan efektif untuk menarik kembali perhatian mereka.

3.2.3. Klaster 2 – Pelanggan Konsumsi Moderat

Klaster ini menunjukkan pola konsumsi yang relatif seimbang, namun sedikit lebih condong ke arah kebutuhan sekunder. Pengeluaran mereka pada produk makanan manis dan pembelian di toko/katalog berada pada level menengah, begitu juga dengan pengeluaran kebutuhan primer dan tersier. Mereka juga memiliki nilai Recency sedang, menandakan tingkat keterlibatan yang moderat. Dari segi demografi, mereka umumnya memiliki lebih sedikit tanggungan anak.

Strategi Pemasarannya bisa dengan Pendekatan fleksibel seperti bundling berbagai jenis produk, promosi tematik musiman, serta kampanye pengingat bisa diterapkan untuk menjaga keterlibatan dan mendorong frekuensi pembelian mereka.

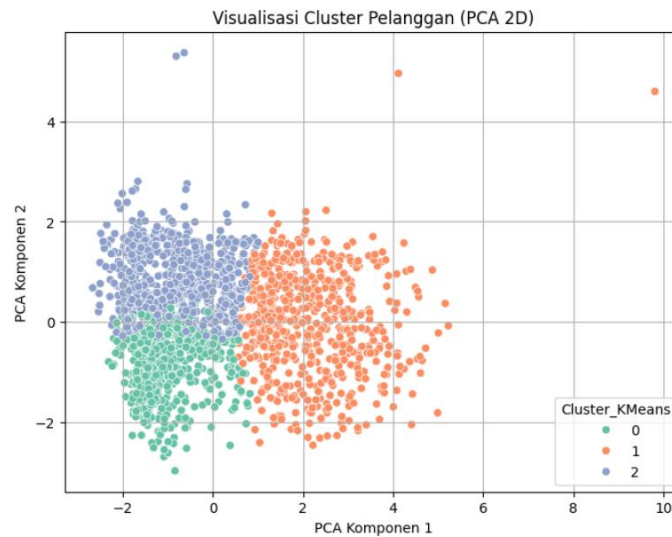
3.3. Visualisasi Hasil Klasterisasi Menggunakan PCA

Setelah proses klasterisasi dilakukan menggunakan algoritma K-Means, langkah selanjutnya adalah memvisualisasikan hasil segmentasi pelanggan agar lebih mudah dipahami dan diinterpretasikan. Untuk tujuan ini, digunakan metode Principal Component Analysis (PCA), yaitu teknik reduksi dimensi yang mentransformasikan data berdimensi tinggi ke dalam dimensi yang lebih rendah sambil tetap mempertahankan informasi penting dari data asli.

Dalam konteks penelitian ini, PCA diterapkan pada tujuh fitur utama yang digunakan dalam proses clustering, yaitu Income, Age, TotalChildren, Kebutuhan Primer, Kebutuhan Sekunder, Kebutuhan Tersier, dan Recency. Data yang telah distandardisasi terlebih dahulu kemudian direduksi menjadi dua komponen utama, yaitu PC1 (Principal Component 1) dan PC2 (Principal Component 2), yang bersama-sama menjelaskan proporsi terbesar dari variasi data pelanggan.

Visualisasi hasil reduksi dimensi dengan PCA ditampilkan dalam bentuk scatter plot dua dimensi. Setiap titik dalam visualisasi tersebut merepresentasikan satu pelanggan, sedangkan warna titik menunjukkan klaster tempat pelanggan tersebut tergolong berdasarkan hasil K-Means. Dari grafik terlihat bahwa ketiga klaster yang terbentuk (label 0, 1, dan 2) memiliki distribusi spasial yang cukup jelas, meskipun terdapat sedikit tumpang tindih di antara beberapa klaster. Hal ini menunjukkan bahwa segmentasi yang dilakukan memiliki dasar yang kuat secara struktural dan dapat digunakan untuk interpretasi lanjutan.

Visualisasi ini memberikan manfaat penting, yaitu memungkinkan pemahaman intuitif terhadap pemisahan klaster dan memperkuat validasi bahwa kelompok pelanggan memang memiliki karakteristik yang berbeda secara signifikan.



Gambar 3. Visualisasi Klaster Pelanggan menggunakan PCA

3.4. Implikasi Strategis dari Hasil Segmentasi

Hasil segmentasi pelanggan yang diperoleh melalui algoritma K-Means dan divisualisasikan menggunakan PCA memberikan wawasan penting bagi perusahaan dalam merancang strategi pemasaran yang lebih terarah dan berbasis data. Tiga klaster pelanggan yang terbentuk menunjukkan perbedaan signifikan dalam hal pola konsumsi, tingkat kebutuhan, dan karakteristik demografis.

Klaster 0, yang didominasi oleh pelanggan dengan tanggungan keluarga dan tingkat konsumsi tinggi terhadap kebutuhan primer (seperti daging, buah, dan ikan), merupakan target ideal untuk strategi pemasaran yang berfokus pada diskon kebutuhan pokok, program loyalitas keluarga, serta bundling produk esensial. Karena pelanggan dalam klaster ini memiliki recency rendah (baru berbelanja), menjaga keterlibatan mereka secara rutin menjadi prioritas untuk meningkatkan retensi.

Klaster 1 menunjukkan orientasi yang lebih kuat pada kebutuhan tersier, ditandai dengan pengeluaran yang tinggi untuk produk mewah seperti emas dan frekuensi pembelian online atau melalui katalog. Dengan karakteristik recency yang tinggi, pelanggan dalam klaster ini cenderung sudah lama tidak bertransaksi. Maka, strategi seperti kampanye eksklusif berbasis personalisasi, flash sale terbatas, atau premium loyalty program dapat digunakan untuk menarik kembali minat mereka sekaligus menciptakan kesan eksklusif dan personal.

Klaster 2 memiliki karakteristik konsumsi yang moderat dan relatif seimbang antara kebutuhan primer, sekunder, dan tersier. Klaster ini mencerminkan konsumen umum dengan perilaku belanja yang fleksibel. Oleh karena itu, strategi pemasaran yang tepat mencakup pendekatan promosi tematik musiman, pemberitahuan ulang secara berkala, serta paket bundling lintas kategori produk yang dapat mendorong keputusan pembelian spontan.

Dengan pemahaman yang lebih dalam terhadap setiap klaster, perusahaan dapat mengalokasikan sumber daya pemasaran secara lebih efisien. Tidak semua pelanggan memerlukan pendekatan yang sama; pesan, media promosi, hingga timing kampanye dapat diatur berdasarkan segmentasi yang telah dihasilkan. Selain itu, hasil ini juga membuka peluang untuk mengembangkan personalisasi berbasis kebutuhan di platform digital perusahaan, seperti sistem rekomendasi produk atau email marketing yang disesuaikan dengan tipe klaster pelanggan.

Secara keseluruhan, pendekatan segmentasi berbasis kebutuhan primer, sekunder, dan tersier ini memberikan kerangka kerja yang lebih aplikatif dalam pengambilan keputusan strategis, baik dalam aspek promosi, distribusi, maupun pengembangan produk, sehingga perusahaan dapat lebih adaptif terhadap preferensi dan prioritas pelanggan yang beragam.

4. KESIMPULAN

Penelitian ini berhasil melakukan segmentasi pelanggan berdasarkan tingkat kebutuhan konsumsi mereka dengan menggunakan pendekatan *unsupervised learning*, khususnya algoritma K-Means Clustering. Segmentasi dilakukan dengan menganalisis kombinasi atribut demografis dan perilaku pembelian pelanggan, yang diklasifikasikan ke dalam tiga kategori kebutuhan, yaitu kebutuhan primer,

sekunder, dan tersier. Melalui proses klusterisasi yang didukung oleh teknik reduksi dimensi Principal Component Analysis (PCA), pelanggan dikelompokkan ke dalam tiga klaster yang memiliki karakteristik konsumsi dan profil yang berbeda.

Klaster pertama terdiri dari pelanggan yang menunjukkan kecenderungan tinggi terhadap pemenuhan kebutuhan primer, seperti pengeluaran besar untuk makanan pokok dan memiliki tanggungan keluarga yang lebih banyak. Klaster kedua mencerminkan kelompok pelanggan dengan orientasi konsumsi terhadap kebutuhan tersier, ditandai oleh pendapatan yang tinggi serta minat pada produk-produk mewah dan gaya hidup, seperti pembelian emas dan belanja katalog. Sementara itu, klaster ketiga merupakan kelompok dengan karakteristik konsumsi yang lebih moderat, menunjukkan pola belanja yang seimbang antara ketiga kategori kebutuhan serta keterlibatan yang stabil dalam aktivitas pembelian.

Visualisasi hasil clustering dengan PCA memperjelas struktur segmentasi yang terbentuk dan menunjukkan bahwa pelanggan dalam dataset memang memiliki perbedaan signifikan dalam hal perilaku konsumsi. Pemisahan antar klaster dalam ruang dua dimensi memperkuat validitas segmentasi, serta memudahkan interpretasi bagi pengambil keputusan dalam konteks strategi pemasaran.

Secara strategis, hasil penelitian ini memberikan kontribusi nyata dalam membantu perusahaan menyusun kampanye pemasaran yang lebih personal, efisien, dan berbasis data. Segmentasi pelanggan berbasis kebutuhan memungkinkan pendekatan yang lebih tepat sasaran, mulai dari penyesuaian konten promosi hingga pengelolaan saluran distribusi yang sesuai dengan preferensi masing-masing kelompok pelanggan. Dengan demikian, perusahaan dapat mengoptimalkan alokasi anggaran pemasaran, meningkatkan loyalitas pelanggan, dan mendorong konversi secara lebih efektif.

Sebagai arah pengembangan ke depan, penelitian ini membuka peluang untuk eksplorasi lebih lanjut dengan membandingkan kinerja algoritma clustering lain seperti DBSCAN, Hierarchical Clustering, atau metode berbasis density dan kernel. Penambahan dimensi psikografis dan pemantauan perilaku pelanggan secara dinamis dari waktu ke waktu juga dapat meningkatkan kedalaman analisis. Lebih jauh lagi, integrasi hasil segmentasi ini ke dalam sistem personalisasi atau rekomendasi produk dapat menjadi implementasi praktis yang bernilai tinggi bagi pengembangan strategi pemasaran digital di masa depan.

DAFTAR PUSTAKA

- [1] Imakash3011. (2021). *Customer Personality Analysis* [Data set]. Kaggle. <https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis>.
- [2] Achmal, E. F., Cholissodin, I., & Adikara, P. P. (2022). Segmentasi pelanggan menggunakan metode Kernel K-Means (Studi kasus: Smartlegal.id). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(6), 2835–2843.
- [3] Azahra, N. D., Maulida, Z. L., Farizqi, M. F., & Kharisudin, I. (2025, Maret). Segmentasi pelanggan dengan algoritma DBSCAN dan K-Means: Mengungkap pola dalam analisis kepribadian pelanggan. *Prosiding Seminar Nasional Matematika PRISMA*, 370–376.
- [4] Brawijaya, F. M. P., Wijoyo, S. H., & Setiawan, N. Y. (2017). Analisis performa algoritma K-Means dan DBSCAN dalam segmentasi pelanggan dengan pendekatan model RFM. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 8(7).
- [5] Chitra, J., & Heikal, J. (2024). Customer segmentation using the K-Means clustering algorithm in foreign banks in Indonesia. *Indonesia Accounting Research Journal*, 11(4), 230–241.
- [6] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- [7] Gomathi Nayagam, S. P., & Kirubakaran, V. S. (2025). Customer segmentation using clustering techniques: Data-driven approach to enhance marketing strategy. In *2025 International Conference on Data Science*.
- [8] Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques* (3rd ed.). Elsevier.
- [9] John, J. M., Shobayo, O., & Ogunleye, B. (2023). An exploration of clustering algorithms for customer segmentation in the UK retail market. *Analytics*, 2(4), 809–823.
- [10] Kotler, P., & Keller, K. L. (2016). *Marketing management* (15th ed.). Pearson Education.
- [11] Mirantika, N., & Rijanto, E. (2024). Implementasi metode clustering partisi dalam menentukan segmentasi pelanggan. *Jurnal Tata Kelola dan Kerangka Kerja Teknologi Informasi*, 10(1), 8–14.
- [12] Nugroho, R., Rafhina, A., Ananda, E., & Gunawan. (2024). Comparative analysis of clustering techniques for customer segmentation evaluating KMeans, Hierarchical, and DBSCAN models alongside RFM frameworks. *International Journal of Advanced Engineering and Management*.

- [13] Pranata, F. M., Wijoyo, S. H., & Setiawan, N. Y. (2024). Analisis performa algoritma K-Means dan DBSCAN dalam segmentasi pelanggan dengan pendekatan model RFM. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 8(7).
- [14] Saputra, A., & Yusuf, R. (2024). Perbandingan algoritma DBSCAN dan KMeans dalam segmentasi pelanggan pengguna transportasi publik Transjakarta menggunakan metode RFM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1346–1361.
- [15] Saragih, H. (2024). Customer segmentation analysis using DBSCAN method in marketing research of retail company. *Jurnal Teknologi Informasi*.
- [16] Sufairoh, I., Rani, A. C., Amalia, K., & Rolliawati, D. (2023). Perbandingan hasil analisis clustering metode KMeans, DBSCAN dan Hierarchical pada data marketplace electronic phone. *Jurnal Sistem Informasi*, 8(1), 97–105.
- [17] Wang, G. (2025). Customer segmentation in digital marketing using a Q-learning-based differential evolution algorithm integrated with KMeans clustering. *PLOS ONE*, 20(2), e0318519. <https://doi.org/10.1371/journal.pone.0318519>
- [18] Wang, S., Sun, L., & Yu, Y. (2024). A dynamic customer segmentation approach by combining LRFMS and multivariate time series clustering. *Scientific Reports*, 14, Article 17491. <https://doi.org/10.1038/s41598-024-17491-0>
- [19] Bansal, A. (2023). Optimizing customer segmentation for enhanced recommendation systems through comparative analysis of KMeans, Hierarchical Clustering, and DBSCAN algorithms. *International Journal of Core Engineering and Management*, 7(6).
- [20] Ufeli, C. P., Sattar, M. U., Hasan, R., & Mahmood, S. (2025). Enhancing customer segmentation through FAMD-based approach using KMeans and hierarchical clustering algorithms. *International Journal of Data Analytics*.
- [21] Salman, F., & Fauziah. (2023). Comparison analysis of KMeans and DBSCAN algorithms for improving budget absorption efficiency in EIS. *Brilliance: Research of Artificial Intelligence*, 3(2), 378–383.
- [22] Ramadhan, H., Kamaludin, M. R. A., Nasrullah, M. A., & Rolliawati, D. (2023). Comparison of hierarchical, KMeans, and DBSCAN clustering methods for credit card customer segmentation analysis based on expenditure level. *Journal of Artificial Intelligence and Computing*, 7(2), 246–251.
- [23] Komatsu, S., & Kimura, O. (2023). Customer segmentation based on smart meter data analytics: Behavioral similarities with manual categorization for building types. *Energy and Buildings*, 284, 113885. <https://doi.org/10.1016/j.enbuild.2023.113885>
- [24] Gopalakrishnan, K. (2024). Customer segmentation using KMeans clustering for targeted marketing in banking. *International Journal of Artificial Intelligence and Machine Learning*, 3(2), 89–94.
- [25] Pabel, B. H., Bhattacharjee, B., Dey, S., Jamee, S., Obaid, M., Mia, M., Khan, S. I., & Sharif, M. (2025). Business analytics for customer segmentation: A comparative study of machine learning algorithms in personalized banking services. *International Journal of Economics, Finance and Management Sciences*, 10(3), 1–13.