

SEGMENTASI PELANGGAN BERBASIS MEDIA SOSIAL DI LINGKUNGAN UKSW MENGGUNAKAN K-MEANS UNTUK PERSONALISASI PENAWARAN PRODUK FASHION DAN KECANTIKAN

SOCIAL MEDIA-BASED CUSTOMER SEGMENTATION AT UKSW USING K-MEANS FOR PERSONALIZING FASHION AND BEAUTY PRODUCT OFFERS

Cantika Tallo

Universitas Kristen Satya Wacana, Kota Salatiga, Jawa Tengah, Indonesia

cantikatallo@gmail.com

ABSTRACT

This study aims to segment customers in the fashion and beauty industry based on social media interaction data using the K-Means algorithm. The data utilized includes engagement indicators such as likes, comments, shares, saves, hashtags, mentions, and caption length. The research process encompasses data preprocessing, determining the optimal number of clusters using the Elbow and Silhouette methods, and the clustering process itself. The results indicate that the optimal number of clusters is two: a high-engagement group and a low-engagement group. The high-engagement cluster demonstrates more active interaction with content, whereas the low-engagement cluster exhibits a more passive response. These segmentation results can serve as a foundation for formulating more personalized and targeted marketing strategies. This study demonstrates that social media-based data mining is effective for understanding consumer behavior.

Keywords: *Clustering, Data Mining, K-Means, Social Media, Customer Segmentation*

ABSTRAK

Penelitian ini bertujuan untuk melakukan segmentasi pelanggan berdasarkan data interaksi media sosial pada industri fashion dan kecantikan menggunakan algoritma K-Means. Data yang digunakan meliputi indikator engagement seperti likes, comments, shares, saves, hashtag, mention, dan panjang caption. Tahapan penelitian meliputi preprocessing data, penentuan jumlah kluster menggunakan metode Elbow dan Silhouette, serta proses clustering. Hasil penelitian menunjukkan bahwa jumlah kluster optimal adalah dua, yaitu kelompok dengan tingkat keterlibatan tinggi dan rendah. Kluster dengan engagement tinggi menunjukkan interaksi yang lebih aktif terhadap konten, sedangkan kluster dengan engagement rendah menunjukkan respons yang lebih pasif. Hasil segmentasi ini dapat digunakan sebagai dasar dalam menyusun strategi pemasaran yang lebih personal dan tepat sasaran. Penelitian ini membuktikan bahwa data mining berbasis media sosial efektif untuk memahami perilaku konsumen.

Kata Kunci: Clustering, Data Mining, K-Means, Media Sosial, Segmentasi Pelanggan

PENDAHULUAN

Perkembangan teknologi digital telah membawa perubahan signifikan terhadap pola interaksi antara perusahaan dan pelanggan. Meningkatnya literasi digital memungkinkan bisnis berinteraksi dengan pelanggan secara lebih efektif melalui pemahaman dan penggunaan teknologi yang semakin baik ("Peran Literasi Digital," 2024). Media sosial pun tidak lagi berfungsi hanya sebagai sarana komunikasi, tetapi telah berkembang menjadi platform utama untuk kegiatan pemasaran, promosi, hingga transaksi, sekaligus meningkatkan identitas merek serta kepercayaan konsumen (Hulu &

Tantowi, 2023). Industri fashion dan kecantikan menjadi salah satu sektor yang paling dinamis dan kompetitif di era digital, ditandai oleh perubahan cepat dalam tren dan preferensi konsumen yang sangat dipengaruhi oleh gaya hidup, opini publik, serta eksposur konten di media sosial (Akbar dkk., 2024). Dalam konteks kehidupan kampus, khususnya di Universitas Kristen Satya Wacana (UKSW), interaksi civitas akademika terhadap konten komersial dan aktivitas kewirausahaan mahasiswa menunjukkan pola yang berbeda dari pasar umum. Pilihan produk, ketertarikan terhadap tren, serta respons terhadap konten visual sering kali

lebih cepat berubah dan bersifat spesifik terhadap karakteristik lingkungan akademik.

Dalam situasi tersebut, segmentasi pelanggan menjadi langkah strategis bagi pelaku usaha untuk memahami karakteristik audiens secara lebih mendalam. Segmentasi memungkinkan pengelompokan konsumen berdasarkan kesamaan perilaku, preferensi, maupun karakteristik tertentu, sehingga strategi pemasaran dapat diarahkan dengan lebih efektif dan efisien (Asy'ari & Winarno, 2024). Salah satu pendekatan yang relevan untuk mendukung segmentasi ini adalah analisis data media sosial, di mana interaksi pengguna seperti likes, komentar, dan penggunaan hashtag dapat menjadi indikator perilaku konsumen. Algoritma K-Means Clustering merupakan salah satu metode populer dalam data mining yang digunakan untuk melakukan pengelompokan data berdasarkan tingkat kesamaan, dan telah diterapkan dalam berbagai penelitian, termasuk pada segmentasi perilaku generasi Z terhadap konsumsi serangga (Kamenidou dkk., 2023). Dalam konteks pemasaran digital, algoritma ini terbukti membantu perusahaan dalam membuat penawaran produk yang lebih personal, meningkatkan relevansi promosi, serta mendorong keputusan pembelian melalui segmentasi pelanggan berbasis interaksi media sosial ("Analisis Efektivitas," 2024).

Penelitian ini memfokuskan kajian pada lingkungan UKSW atas pertimbangan praktis dan teoretis. Dari sisi praktis, ketersediaan data serta akses langsung ke lapangan memungkinkan verifikasi dan implementasi hasil penelitian dilakukan secara lebih optimal. Dari sisi teoretis, penelitian yang berfokus pada konteks pendidikan tinggi dapat memperkaya literatur mengenai segmentasi audiens di lingkungan kampus yang memiliki karakter demografis dan perilaku unik dibandingkan populasi umum. Penelitian ini disusun untuk menjelaskan dan menjawab sejumlah pertanyaan yang berkaitan dengan proses

segmentasi pelanggan berbasis interaksi media sosial di lingkungan UKSW. Pertanyaan pertama berfokus pada bagaimana pola interaksi pengguna seperti like, komentar, penggunaan hashtag, serta tingkat keterlibatan lainnya terbentuk pada konten yang berkaitan dengan produk fashion dan kecantikan, serta bagaimana algoritma K-Means dapat diterapkan untuk mengelompokkan audiens berdasarkan pola-pola tersebut. Selanjutnya, penelitian ini juga mempertanyakan berbagai tantangan teknis maupun nonteknis yang muncul selama proses pengumpulan data, tahap preprocessing, hingga penerapan algoritma K-Means, termasuk kualitas data, keberadaan noise, isu etika, maupun keterbatasan algoritmik. Selain itu, penelitian ini mengeksplorasi solusi dan pendekatan alternatif yang dapat diterapkan untuk meningkatkan akurasi hasil klasterisasi dan memperkuat efektivitas personalisasi penawaran produk, sehingga segmentasi pelanggan yang terbentuk benar-benar dapat dimanfaatkan secara optimal oleh pelaku usaha maupun unit kewirausahaan kampus.

Sejalan dengan rumusan masalah tersebut, penelitian ini bertujuan untuk mengidentifikasi pola interaksi audiens media sosial yang berkaitan dengan produk fashion dan kecantikan di lingkungan UKSW, menerapkan algoritma K-Means untuk menghasilkan segmentasi berdasarkan data interaksi, serta mengembangkan rekomendasi personalisasi penawaran yang dapat dimanfaatkan oleh pelaku usaha, unit kewirausahaan, maupun pengelola kegiatan pemasaran kampus. Penelitian ini diharapkan memberikan manfaat akademis berupa kontribusi pada pengembangan ilmu data mining, khususnya dalam penerapan algoritma clustering untuk analisis perilaku konsumen, serta menjadi referensi bagi penelitian lanjutan mengenai personalisasi pemasaran digital di lingkungan kampus maupun konteks yang lebih luas. Secara praktis, hasil penelitian ini diharapkan dapat memberikan gambaran berbasis data

yang bermanfaat bagi pelaku usaha fashion dan kecantikan dalam merancang strategi pemasaran yang lebih tepat sasaran, sekaligus meningkatkan efektivitas promosi melalui penyusunan strategi yang sesuai dengan segmen pelanggan yang lebih spesifik. Untuk menjaga fokus penelitian, ruang lingkup dibatasi pada penggunaan data dari platform media sosial tertentu yang relevan dengan pemasaran produk fashion dan kecantikan, penerapan algoritma K-Means Clustering tanpa perbandingan dengan metode lain, serta analisis yang dipusatkan pada segmentasi pelanggan dan implikasinya terhadap personalisasi penawaran, bukan pada pengembangan aplikasi komersial secara menyeluruh.

Berdasarkan uraian latar belakang dan permasalahan yang telah dipaparkan, maka penelitian ini bertujuan untuk mengidentifikasi dan menganalisis pola interaksi pengguna media sosial yang berkaitan dengan produk fashion dan kecantikan di lingkungan Universitas Kristen Satya Wacana (UKSW) sebagai dasar dalam membentuk segmentasi pelanggan yang lebih terstruktur dan berbasis data. Penelitian ini difokuskan pada analisis perilaku audiens media sosial yang tercermin melalui indikator keterlibatan digital, seperti jumlah like, komentar, penggunaan hashtag, mention, serta rasio engagement terhadap konten komersial produk fashion dan kecantikan yang beredar di lingkungan UKSW.

Adapun metode yang digunakan dalam penelitian ini adalah pendekatan kuantitatif dengan penerapan teknik data mining, khususnya algoritma K-Means Clustering, untuk mengelompokkan audiens ke dalam beberapa segmen berdasarkan kesamaan pola interaksi media sosial. Proses penelitian meliputi tahapan pengumpulan data media sosial, preprocessing data, ekstraksi fitur engagement, penentuan jumlah kluster optimal menggunakan metode Elbow dan Silhouette, serta analisis karakteristik setiap kluster yang terbentuk.

Hasil yang ingin dicapai dari penelitian ini adalah terbentuknya segmentasi pelanggan yang jelas dan dapat diinterpretasikan secara praktis, sehingga mampu memberikan rekomendasi personalisasi penawaran produk fashion dan kecantikan yang lebih tepat sasaran di lingkungan kampus. Dengan demikian, hasil penelitian ini diharapkan tidak hanya berkontribusi secara akademis dalam pengembangan kajian data mining dan pemasaran digital, tetapi juga memberikan manfaat praktis bagi pelaku usaha, unit kewirausahaan, serta pengelola kegiatan pemasaran kampus dalam merancang strategi promosi yang lebih efektif dan relevan dengan karakteristik segmen pelanggan yang ada.

KAJIAN PUSTAKA

Landasan Teori

Dalam studinya tentang penggunaan algoritma K-Means dalam pemasaran, Chen menekankan betapa pentingnya penggunaan algoritma K-Means dalam pemasaran berbasis data. Studi ini mengaitkan K-means dengan ide data mining dan penemuan pengetahuan. Ini juga menjelaskan prinsip dasar, keunggulan, dan contoh aplikatif K-means yang relevan dengan praktik bisnis kontemporer. Namun, penelitian ini belum memeriksa masalah teknis seperti sensitivitas pusat awal dan data berdimensi tinggi secara menyeluruh, dan hanya melakukan penelitian teoritis tanpa dukungan studi kasus empiris (Chen, 2024). Akibatnya, untuk mengoptimalkan potensi K-means dalam pemasaran, pengembangan lebih lanjut diperlukan melalui integrasi teknologi baru dan penerapan empiris. Nugraha dan Dwita menemukan bahwa kepuasan konsumen dan kepercayaan merek dipengaruhi positif oleh penggunaan media sosial. Akibatnya, konsumen di Kota Padang menjadi lebih setia pada produk Wardah. Sejalan dengan temuan penelitian sebelumnya, kepuasan dan kepercayaan mediasi pengaruh tersebut, bukan secara langsung. Faktor demografis juga

memengaruhi pola penggunaan media sosial, sehingga strategi digital perusahaan bergantung pada seberapa baik mereka dapat membuat pengalaman pelanggan dan menumbuhkan kepercayaan pelanggan (Nugraha & Dwita, 2023). Supangat dan Mulyani menggunakan metode RFM dan algoritma K-means untuk menganalisis loyalitas pelanggan di toko souvenir Labuan Bajo. Hasil penelitian menunjukkan bahwa ada segmen pelanggan yang berbeda dalam hal loyalitas dan nilai ekonomi. Ini dapat menjadi dasar untuk pendekatan pemasaran yang lebih personal. Kombinasi RFM dan K-means telah terbukti berhasil dalam segmentasi pelanggan dan membantu meningkatkan retensi dan profitabilitas bisnis. Untuk menganalisis dan memprediksi pergerakan pelanggan ritel berdasarkan data transaksi historis, Harish dan Malathy (2023) menggabungkan metode clustering K-Means dan model Markov. Penelitian ini dapat memetakan segmen pelanggan dan memperkirakan perpindahan antar segmen dengan menggunakan metrik RFM. Hasilnya, metode ini masih terbatas pada asumsi Markovian dan kurang mempertimbangkan faktor eksternal, tetapi dinilai efektif dalam mendukung strategi retensi dan pemasaran.

Segmentasi pelanggan merupakan proses membagi konsumen ke dalam kelompok-kelompok yang memiliki karakteristik serupa agar strategi pemasaran dapat lebih terarah. Menurut teori pemasaran, segmentasi bertujuan untuk memahami kebutuhan, preferensi, serta perilaku pelanggan sehingga perusahaan dapat menyusun strategi komunikasi dan penawaran yang lebih relevan. Segmentasi dapat dilakukan berdasarkan berbagai aspek, seperti demografis (usia, jenis kelamin, pendapatan), geografis (wilayah tempat tinggal), psikografis (gaya hidup, minat, nilai), maupun perilaku (pola pembelian, tingkat penggunaan, loyalitas). Surojo dalam penelitiannya pentingnya segmentasi dalam pemasaran, di mana segmentasi bertujuan untuk memahami

kebutuhan, preferensi, serta perilaku pelanggan, sehingga perusahaan dapat menyusun strategi komunikasi dan penawaran yang lebih relevan. Segmentasi dilakukan berdasarkan berbagai aspek, termasuk demografis, geografis, psikografis, dan perilaku, untuk merancang bauran pemasaran yang efektif (Supangat & Mulyani, 2023). Dapat dikatakan bahwa dalam konteks pemasaran digital, segmentasi sangat penting karena tingginya persaingan serta kebutuhan perusahaan untuk menyampaikan pesan yang tepat kepada audiens yang tepat.

Media sosial saat ini menjadi salah satu kanal utama interaksi antara konsumen dan perusahaan. Penelitian oleh Ardani dan Harahap menjelaskan bahwa media sosial telah menjadi salah satu kanal utama bagi interaksi antara konsumen dan perusahaan, dengan peningkatan signifikan dalam jumlah like, share, dan komentar pada konten UMKM setelah penerapan strategi konten kreatif (Surojo et al., 2024). Data yang dihasilkan dari aktivitas pengguna seperti jumlah like, komentar, ulasan, hashtag, hingga konten yang dibagikan dapat memberikan wawasan mendalam mengenai preferensi konsumen. Keunggulan media sosial sebagai sumber data adalah sifatnya yang *real-time*, besar volumenya, serta beragam bentuknya (teks, gambar, video). Namun, tantangan yang dihadapi antara lain adalah keberagaman bahasa, adanya data yang tidak relevan, serta kebutuhan akan teknik analisis khusus agar informasi yang dihasilkan dapat dimanfaatkan dengan baik.

Data mining adalah proses menggali informasi yang berguna dari kumpulan data yang besar dengan bantuan metode statistik, machine learning, maupun algoritma komputasi. Data mining adalah proses penggalian informasi dari sejumlah besar data untuk menemukan pola atau hubungan yang berguna, yang melibatkan metode seperti algoritma klasifikasi (Ardani & Harahap, 2024). Proses ini mencakup beberapa tahapan, yaitu pengumpulan data, *preprocessing*

(pembersihan data dari noise atau duplikasi), transformasi data, pemodelan, hingga evaluasi hasil. Dalam konteks media sosial, data mining berperan penting untuk mengolah big data menjadi informasi yang dapat dimanfaatkan oleh perusahaan dalam memahami perilaku konsumen serta mendukung pengambilan keputusan bisnis.

Clustering merupakan salah satu metode dalam *data mining* yang digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik tertentu tanpa memerlukan label atau target output. Metode ini bertujuan untuk menemukan pola tersembunyi dalam data yang tidak terlihat secara langsung. *Clustering* berbeda dengan *classification* karena tidak menggunakan data yang sudah diberi label, melainkan bekerja dengan data mentah. Dalam analisis data, clustering sering digunakan untuk menentukan area yang homogen atau kelompok yang memiliki kesamaan. Ini memungkinkan pemahaman yang lebih baik tentang struktur data dan memudahkan analisis lebih lanjut (Sofyan dkk., 2023). Terdapat berbagai jenis algoritma clustering, seperti hierarchical clustering, density-based clustering (misalnya DBSCAN), dan partitioning clustering (misalnya K-Means). Pemilihan algoritma sangat bergantung pada tujuan penelitian serta karakteristik data yang dianalisis.

Algoritma K-Means merupakan salah satu metode clustering yang paling populer dan banyak digunakan karena kesederhanaan dan efisiensinya. Prinsip dasar K-Means adalah membagi data ke dalam sejumlah cluster k yang ditentukan di awal, berdasarkan kedekatan data terhadap pusat cluster (centroid). Prosesnya dimulai dengan menentukan titik pusat awal, lalu setiap data dihitung jaraknya terhadap masing-masing pusat. Data akan dikelompokkan ke cluster terdekat, kemudian pusat cluster diperbarui dengan menghitung rata-rata dari anggota cluster tersebut. Proses ini dilakukan berulang hingga posisi pusat cluster stabil. Kelebihan K-Means antara lain sederhana, cepat, dan

cocok untuk dataset berukuran besar. Namun, kelemahannya adalah harus menentukan jumlah cluster sejak awal serta rentan terhadap outlier. Dalam konteks penelitian, algoritma ini sangat efektif untuk mengidentifikasi pola dalam dataset yang sangat besar; ini membantu memahami struktur data dengan cara yang lebih terorganisir (Mulaomerović-Šeta dkk., 2023).

Hasil clustering perlu dievaluasi untuk memastikan kualitas pengelompokan yang dilakukan. Evaluasi hasil clustering merujuk pada proses untuk menilai seberapa baik grup atau cluster yang dihasilkan oleh algoritma clustering, berdasarkan kriteria tertentu (Situmorang & Bintang Fajar, 2023). Evaluasi dapat dilakukan dengan dua pendekatan, yaitu validasi internal dan eksternal. Validasi internal, seperti silhouette coefficient, digunakan untuk mengukur tingkat keseragaman dalam satu cluster dan perbedaan antar cluster. Sementara itu, validasi eksternal digunakan apabila tersedia data pembanding atau label untuk mengecek kesesuaian hasil clustering dengan kategori yang telah diketahui sebelumnya. Dalam penelitian berbasis data media sosial, evaluasi hasil clustering juga harus memperhatikan interpretasi bisnis, sehingga kelompok pelanggan yang terbentuk benar-benar dapat dijadikan dasar dalam strategi pemasaran.

Personalisasi merupakan strategi pemasaran yang menyesuaikan produk, layanan, maupun pesan komunikasi dengan karakteristik dan kebutuhan spesifik setiap pelanggan. Pengertian personalisasi penawaran produk merujuk pada pendekatan yang digunakan untuk menyesuaikan penawaran barang atau jasa sesuai dengan preferensi, kebiasaan, dan kebutuhan unik masing-masing pelanggan (Qirom dkk., 2024). Dalam era digital, personalisasi menjadi sangat penting karena konsumen cenderung lebih merespons positif pada konten yang relevan dengan minat mereka. Segmentasi pelanggan melalui analisis data

memungkinkan perusahaan untuk merancang penawaran yang tepat sasaran, sehingga meningkatkan kepuasan, loyalitas, serta peluang pembelian ulang.

Industri fashion dan kecantikan memiliki karakteristik konsumen yang unik, di mana keputusan pembelian sering kali dipengaruhi oleh tren, gaya hidup, dan citra diri. Berdasarkan penelitian oleh Suhada dkk., (2023) yang membahas perkembangan industri kecantikan di Indonesia, analisis pengaruh kualitas produk dan citra merek terhadap loyalitas pelanggan di klinik kecantikan, menunjukkan bahwa industri kecantikan semakin menjadi gaya hidup di kalangan masyarakat. Media sosial menjadi salah satu platform utama untuk membangun brand awareness di sektor ini, karena konsumen fashion dan kecantikan aktif mencari inspirasi maupun rekomendasi dari influencer, komunitas, maupun konten visual. Sebagai contoh, di pasar Indonesia, keberadaan influencer kecantikan lokal sangat berkontribusi terhadap pertumbuhan industri kecantikan, dengan influencer mampu mengarahkan keputusan pembelian konsumen berkat daya tarik dan relevansi konten mereka (Pramesti dkk., 2025). Oleh karena itu, personalisasi penawaran melalui segmentasi pelanggan berbasis media sosial menjadi strategi yang sangat relevan untuk meningkatkan efektivitas pemasaran dalam industri ini.

Penelitian Terdahulu

Judul Penelitian “*Consumer Segmentation in the Fashion Industry Using Social Media: An Empirical Analysis*” oleh Singhal dkk., (2023) mengelompokkan konsumen industri fashion berdasarkan interaksi di media sosial dengan mengidentifikasi faktor Hubungan Merek (*Consumer Brand Relationship, CBR*) dan Persepsi Merek (*Consumer Brand Perception, CBP*) serta merumuskan model konseptual terkait. Metodologi yang digunakan meliputi survei online (575 responden, 500 data valid), Analisis Faktor Eksploratori untuk

mengekstrak faktor CBR dan CBP, pembentukan indeks FCBRI/FCBPI (*Fashion Consumer Brand Relationship/Perception Index*), lalu pengelompokan kluster menggunakan algoritma K-means. Hasil utama menunjukkan terbentuknya empat kluster konsumen yang berbeda karakteristiknya berdasarkan pola hubungan/persepsi merek melalui media sosial. Misalnya, Kluster 1 dan 3 memiliki tingkat hubungan merek tinggi dan berbeda jauh dari Kluster 2 dan 4. Studi ini juga menunjukkan bagaimana kluster-kluster tersebut dapat digunakan untuk merumuskan strategi penargetan dan positioning pemasaran yang lebih efisien. Relevansinya dengan penelitian ini adalah demonstrasi langsung penggunaan K-means untuk segmentasi konsumen fashion berbasis media sosial, selaras dengan fokus penelitian pada pemodelan segmen pelanggan pada industri fashion/kecantikan.

“Revealing Consumer Preferences in the Fashion Industry Using K-Means Clustering” (Sulianta dkk., 2024). Artikel ini bertujuan mengungkap pola preferensi konsumen produk fashion berdasarkan data penjualan. Studi memanfaatkan data mining dengan algoritma K-means dan metode Elbow pada dataset 1.000 rekaman. Dengan mengelompokkan data berdasarkan preferensi harga dan rating produk, analisis ini berhasil mengidentifikasi empat segmen konsumen utama. Jumlah kluster optimal ($k=4$) dipastikan menggunakan metode WCSS (Within-Cluster Sum of Square). Kontribusi penelitian ini adalah memberikan wawasan penting berupa pembagian segmen konsumen berdasarkan preferensi belanja fashion, yang dapat dijadikan panduan untuk strategi pemasaran produk fashion di masa depan.

“*Analyzing Customer Loyalty Levels through Segmentation in Aesthetic Clinics Using K-Means and RFAM*” (Sinarring dkk., 2024). Penelitian ini mengkaji segmentasi pelanggan klinik estetika berdasarkan tingkat loyalitas dengan

menerapkan algoritma K-means pada model RFAM (*Recency, Frequency, Average Monetary*). Data yang digunakan adalah histori transaksi 493 pelanggan klinik kecantikan. Metodologi meliputi penerapan K-means pada skor RFAM untuk mengelompokkan pelanggan menjadi grup yang homogen. Hasil analisis menunjukkan terbentuknya dua kluster utama: Kluster 1 (156 pelanggan) berkarakter loyalitas rendah, dan Kluster 2 (337 pelanggan) berkarakter loyalitas tinggi. Penelitian ini menemukan bahwa pendekatan K-means berbasis RFAM efektif dalam membedakan pelanggan berdasarkan loyalitas mereka. Kontribusi utamanya adalah model segmentasi yang dapat membantu klinik kecantikan menyesuaikan layanan dan promosi sesuai kebutuhan tiap kluster pelanggan. Hasil ini relevan untuk skripsi karena menggambarkan penggunaan K-means pada data konsumen di industri kecantikan, khususnya untuk analisis loyalitas dan segmentasi pelanggan.

“*Tailoring customer segmentation strategies for luxury brands in the NFT market – The case of SUPERGUCCI*” (Chen dkk., 2025). Studi ini menelusuri segmentasi konsumen di pasar NFT (*Non-Fungible Token*) mode mewah dengan fokus pada koleksi NFT “SUPERGUCCI” oleh Gucci. Metode yang digunakan adalah *web crawling* (OpenSea) dan analisis K-means terhadap perilaku pengguna marketplace NFT. Hasilnya mengidentifikasi tiga kluster konsumen berbeda: *speculators* (spekulan), *casual collectors* (kolektor kasual), dan *cryptocurrency natives* (pengguna kripto). Penelitian ini menunjukkan dampak signifikan NFT terhadap segmentasi dan keterlibatan konsumen di industri mewah, memberikan wawasan strategi bagi merek fashion yang memanfaatkan teknologi baru. Meskipun berbasis NFT, studi ini relevan karena menerapkan K-means dalam konteks fashion modern, memperluas pemahaman tentang segmentasi konsumen digital yang dapat dibandingkan dengan

model segmen tradisional di industri fashion/kecantikan.

“*Clustered for Clarity: Leveraging AIO and K-Means to Decode the Modest Fashion Market for Indonesian Muslimah*” (Mulia dkk., 2025). Penelitian ini menargetkan segmentasi pasar fashion muslimah di Indonesia dengan pendekatan psikografis AIO (*Activities, Interests, Opinions*) dan clustering K-means. Metode yang dipakai adalah survei online kepada 450 perempuan pembeli fashion muslimah (433 data valid), pengukuran indikator psikografis (skala Likert 7 poin), dan analisis K-means (dengan *Elbow* dan *Silhouette*) untuk menentukan jumlah kluster optimal. Hasilnya terbentuk tiga segmen psikografis: *Value-Oriented Pragmatists*, *Trendy Realists*, dan *Aspirational Stylists*, masing-masing dengan preferensi berbeda (kualitas, gaya, nilai aspirasi) dan profil demografi yang berbeda pula. Kontribusi penelitian ini adalah model segmentasi psikografis untuk sektor modest fashion, yang dapat membantu UMKM menyesuaikan strategi pemasaran secara lebih relevan budaya. Studi ini relevan karena menunjukkan aplikasi K-means untuk segmentasi konsumen fashion sesuai demografi/psikografi, memperkuat pendekatan segmentasi yang menjadi fokus skripsi.

“*Implementation of Customer Segmentation Model using K-Means and DBSCAN for Fashion Industry Product Transaction*” (William & Johan, 2025). Artikel ini mengimplementasikan model segmentasi pelanggan untuk data transaksi industri fashion dengan algoritma K-means dan DBSCAN dalam kerangka kerja CRISP-DM. Data berupa transaksi pembelian produk fashion dari sebuah perusahaan, diolah dengan normalisasi (*Standard Scaler*) dan reduksi dimensi (PCA) sebelum *clustering*. Evaluasi hasil kluster menggunakan *Silhouette Score* dan *Davies-Bouldin Index* menunjukkan K-means lebih baik daripada DBSCAN: *optimum* $k=7$ dengan *Silhouette* 0,549

(DBSCAN hanya 0,29). Implementasi akhir berupa dashboard berbasis FastAPI dan React yang mengotomatiskan proses segmentasi pelanggan. Penelitian ini menambah bukti bahwa K-means efektif untuk segmentasi pelanggan *e-commerce fashion*, sesuai konteks skripsi yang membangun model segmen pelanggan berbasis data transaksi.

Sejumlah penelitian terdahulu telah membuktikan efektivitas algoritma K-Means dalam segmentasi konsumen di industri fashion dan kecantikan. Singhal dkk. (2023) melakukan segmentasi konsumen fashion berbasis media sosial dengan mengidentifikasi faktor hubungan dan persepsi merek, menghasilkan empat kluster konsumen yang berbeda. Sulianta dkk. (2024) mengungkap preferensi konsumen fashion berdasarkan data penjualan dengan variabel harga dan rating produk, dan juga menghasilkan empat segmen. Mulia dkk. (2025) menerapkan pendekatan psikografis AIO dikombinasikan dengan K-Means untuk memetakan pasar fashion muslimah di Indonesia, menghasilkan tiga segmen dengan profil yang kaya secara deskriptif. Sementara itu, Sinarring dkk. (2024) menggunakan K-Means berbasis model RFAM pada data transaksi klinik kecantikan untuk menganalisis loyalitas pelanggan. Meskipun penelitian-penelitian tersebut memberikan kontribusi yang signifikan, terdapat beberapa celah yang belum sepenuhnya dijawab.

Pertama, minimnya penelitian yang menggunakan data interaksi media sosial secara langsung sebagai basis segmentasi. Sebagian besar penelitian terdahulu menggunakan data survei, data transaksi, atau indeks merek yang dikonstruksi secara manual. Pendekatan ini membutuhkan waktu dan sumber daya yang lebih besar dalam pengumpulan data. Sementara itu, data interaksi media sosial seperti likes, comments, shares, saves, dan engagement rate tersedia secara real-time dan mencerminkan perilaku konsumen secara organik tanpa bias responden. Penelitian

yang secara khusus memanfaatkan indikator keterlibatan digital mentah sebagai fitur clustering masih terbatas. Kedua, kurangnya penelitian segmentasi konsumen fashion dan kecantikan yang berfokus pada konteks komunitas kampus. Penelitian terdahulu umumnya dilakukan pada pasar umum, pasar *e-commerce*, atau komunitas tertentu seperti pasar muslimah. Lingkungan kampus memiliki karakteristik demografis dan perilaku yang unik, di mana konsumen cenderung lebih terpengaruh oleh tren, komunitas sebaya, dan aktivitas kewirausahaan internal kampus. Segmentasi yang secara spesifik menargetkan ekosistem kampus sebagai unit analisis belum banyak dieksplorasi dalam literatur. Ketiga, penelitian yang menggabungkan validasi kluster ganda yaitu metode *Elbow* dan *Silhouette* secara bersamaan dalam konteks data media sosial kampus masih jarang dilakukan. Beberapa penelitian terdahulu hanya menggunakan salah satu metode validasi, sehingga keputusan penentuan jumlah kluster optimal kurang dapat dipertanggungjawabkan secara metodologis secara menyeluruh.

Berdasarkan ketiga celah tersebut, penelitian ini hadir untuk mengisi gap yang ada dengan menerapkan algoritma K-Means pada data interaksi media sosial secara langsung, berfokus pada lingkungan kampus UKSW, serta menggunakan kombinasi metode *Elbow* dan *Silhouette* untuk memastikan validitas hasil segmentasi yang diperoleh.

METODE PENELITIAN

Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan dukungan metode data mining. Data mining dapat menghemat waktu dan biaya, membantu strategi pemasaran, meningkatkan keamanan, mendukung perencanaan produk, meningkatkan layanan pelanggan, dan mengembangkan sistem rekomendasi. Selain itu, data mining juga dapat membantu strategi pemasaran,

meningkatkan efisiensi dan produktivitas, dan mendukung pengambilan keputusan yang lebih baik. Dengan menerapkan data mining, organisasi dapat meningkatkan layanan pelanggan melalui analisis sentimen dan umpan balik pengguna untuk menyesuaikan layanan yang ditawarkan berdasarkan preferensi dan perilaku pelanggan (Singhal dkk., 2023). Selain itu, dengan menggunakan data besar yang akurat dan dapat dipertanggungjawabkan, teknik ini membuka wawasan baru dan inovasi. Namun, kualitas data dan algoritma yang digunakan sangat memengaruhi metode ini. Pendekatan kuantitatif dipilih karena penelitian ini memerlukan data dalam bentuk angka yang dapat diolah secara statistik untuk menghasilkan temuan yang objektif. Penerapan data mining difokuskan pada penggunaan algoritma K-Means Clustering yang mampu membagi data pelanggan ke dalam kelompok tertentu berdasarkan pola kesamaan perilaku. Pendekatan kuantitatif dapat memberikan data yang diperlukan untuk mengevaluasi dampak dari merek dan *influencer* dalam konteks pemasaran (Alotaibi dkk., 2024). K-Means Clustering dapat mengungkap preferensi konsumen di industri *fashion*, secara efektif melakukan segmentasi berdasarkan perilaku pembelian dan loyalitas merek (Dharmawan dkk., 2024). Dengan pendekatan ini, analisis segmentasi pelanggan dari media sosial dapat dilakukan secara lebih sistematis, terukur, dan dapat dipertanggungjawabkan.

Sumber Data

Sumber data primer adalah posting dan interaksi pada akun media sosial yang berhubungan dengan aktivitas pemasaran di UKSW (akun resmi UKSW, akun kewirausahaan mahasiswa/alumni, dan komunitas penjual yang aktif). Jika diperlukan, sumber sekunder meliputi dokumentasi event kampus dan catatan unit kewirausahaan.

Teknik Sampling

Pemilihan sampel dilakukan secara *purposive* untuk menentukan akun/komunitas yang representatif terhadap aktivitas pemasaran kampus. *Purposive sampling* sering kali digunakan dalam penelitian untuk mengevaluasi situasi secara mendalam, sehingga memberikan informasi yang kaya dan mendalam (Chen dkk., 2024). Dari kumpulan posting yang dipilih, data interaksi diambil pada rentang waktu November-Desember 2025. Jika volume data besar, digunakan simple random sampling untuk memilih sejumlah n posting/interaksi yang memenuhi kriteria kualitas data (mis. tidak spam, bukan bot).

Teknik Pengumpulan Data

Data dikumpulkan melalui metode yang mematuhi kebijakan platform: menggunakan API publik jika tersedia atau *tools* legal seperti Instaloader untuk keperluan penelitian, dilengkapi pernyataan kepatuhan terhadap *Terms of Service*. penggunaan alat yang mematuhi kebijakan hukum dan etika dalam pengumpulan data dapat menghasilkan informasi berharga yang relevan bagi analisis pelanggan (Mulia dkk., 2025). Setiap pengumpulan data akan didampingi dengan permintaan izin atau pemberitahuan kepada pemilik akun jika data berasal dari grup tertutup atau akun pribadi.

Preprocessing dan Ekstraksi Fitur

Tahapan *preprocessing* mencakup pembersihan data, penghapusan duplikasi, penanganan missing value, serta deteksi akun bot atau spam untuk memastikan kualitas data. Implementasi algoritma K-Means untuk segmentasi pelanggan dengan menjelaskan langkah-langkah yang diperlukan untuk mematuhi syarat layanan saat melakukan pengumpulan data, sehingga hasil analisis memberikan pemahaman yang kuat terhadap perilaku pelanggan dalam kerangka hukum. Fitur yang digunakan dalam klasterisasi meliputi frekuensi like, jumlah komentar, rasio

engagement, penggunaan tag, mention, dan hashtag kampus, serta panjang caption dan sentimen teks sebagai opsi tambahan (Hartama & Oktaviani, 2025). Pentingnya konteks dan pengguna dalam mendukung pengumpulan dan pembersihan data, termasuk fitur-fitur yang relevan dalam metrik engagement yang perlu diperhatikan saat melakukan analisis (Ali dkk., 2025). Seluruh data dinormalisasi menggunakan min-max scaling sebelum diberi bobot dan diolah dengan algoritma K-Means agar hasil klusterisasi lebih akurat.

Penentuan jumlah kluster (k) dan Evaluasi

Penentuan k optimal dilakukan dengan metode *Elbow* dan *Silhouette score*. Kinerja klustering diukur menggunakan skor *silhouette* sebagai evaluasi untuk mendapatkan hasil klustering terbaik (Kuswardana dkk., 2025). Skor *silhouette* menghitung perbedaan antara jarak intra-kluster dan inter-kluster (Morais dkk., 2023). Analisa deskriptif merupakan komponen penting dalam pengelompokan pelanggan, di mana mereka mengevaluasi hasil kluster menggunakan metode evaluasi seperti *Silhouette* dan *Elbow* (Sulianta dkk., 2024). Nilai *Silhouette* > 0.25 dianggap menunjukkan kluster yang terpisah secara wajar untuk tujuan praktis, namun keputusan akhir akan mempertimbangkan interpretabilitas kluster bagi pemangku kepentingan kampus.

Analisis Hasil dan Rekomendasi

Setiap kluster dianalisis secara deskriptif untuk mengenali karakteristik utama (misalnya "*Trend-Seekers*", "*Budget-Conscious Campus Shoppers*", "*Event-Driven Buyers*"). Penelitian terdahulu mengidentifikasi bahwa karakteristik perilaku dan preferensi yang mendasari keputusan dan interaksi klien dalam konteks penggunaan layanan kampus (Yildiz, 2023). Rekomendasi personalisasi disusun berdasarkan karakteristik tersebut dan difokuskan pada aplikasi praktis di UKSW: strategi promosi

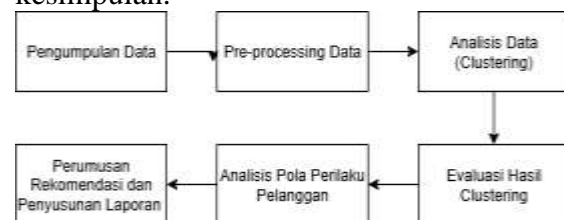
untuk bazar kampus, penyesuaian pesan untuk akun mahasiswa/UMKM, serta segmentasi *audience* untuk kegiatan kolaborasi kampus.

Etika Penelitian

Penelitian ini mengikuti prinsip etika penelitian data digital: data dianonimkan, identitas individu tidak ditampilkan, dan hanya agregat yang disajikan. Perlunya penggunaan model anonimisasi data yang tepat dalam analisis big data untuk memastikan bahwa data yang digunakan tidak mengidentifikasi individu secara langsung, dan hanya hasil agregat yang disajikan (Rajesh dkk., 2024). Ini membantu dalam menjaga keseimbangan antara privasi individu dan utilitas data dalam analisis. Untuk data dari grup tertutup atau akun pribadi, peneliti wajib mendapatkan izin tertulis atau persetujuan pemilik akun. Semua prosedur akan dicantumkan dalam lampiran etika penelitian.

Alur Penelitian

Secara keseluruhan, penelitian ini mengikuti alur yang sistematis mulai dari pengumpulan data hingga penarikan kesimpulan.



Gambar 1. Tahapan Penelitian

Pada gambar 1 dapat dijelaskan bahwa alur penelitian diawali dengan proses pengumpulan data melalui web scraping pada media sosial, kemudian data yang terkumpul diproses melalui tahap pre-processing. Analisis yang dilakukan berdasarkan data yang diperoleh memerlukan proses pre-processing untuk memastikan kualitas dan keakuratan informasi yang dianalisis (Alzami dkk., 2023). Data yang telah bersih selanjutnya dianalisis menggunakan algoritma K-Means *Clustering* untuk menghasilkan pembagian pelanggan ke dalam beberapa

cluster. Setelah hasil *clustering* diperoleh, dilakukan evaluasi jumlah cluster agar sesuai dengan karakteristik data. Hasil evaluasi tersebut kemudian dianalisis untuk menemukan pola dan kecenderungan perilaku pelanggan. Tahap akhir dari penelitian ini adalah merumuskan rekomendasi personalisasi penawaran produk fashion dan kecantikan berdasarkan hasil segmentasi yang telah dilakukan. Dengan alur ini, penelitian dapat dijalankan secara konsisten dan sesuai tujuan yang ditetapkan.

HASIL DAN PEMBAHASAN

Deskripsi Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data interaksi media sosial yang merepresentasikan perilaku pengguna terhadap konten digital, khususnya dalam konteks produk fashion dan kecantikan di lingkungan Universitas Kristen Satya Wacana (UKSW). Data ini bersifat kuantitatif dan terdiri dari sejumlah variabel yang secara langsung maupun tidak langsung mencerminkan tingkat keterlibatan (*engagement*) audiens.

Variabel *likes*, *comments*, *shares*, dan *saves* dikategorikan sebagai indikator utama interaksi eksplisit, karena secara langsung menunjukkan respons pengguna terhadap konten. *Likes* menggambarkan apresiasi awal, *comments* menunjukkan tingkat partisipasi aktif, *shares* merepresentasikan distribusi ulang konten, sedangkan *saves* menunjukkan nilai konten bagi pengguna untuk dikonsumsi kembali di masa mendatang.

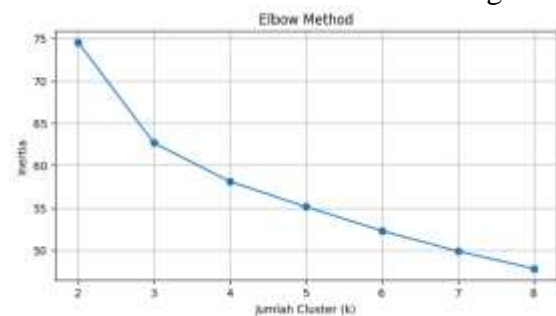
Selain itu, variabel *hashtags_count* dan *mentions_count* berfungsi sebagai indikator strategi distribusi konten. *Hashtag* memungkinkan konten menjangkau audiens yang lebih luas melalui sistem pencarian berbasis topik, sedangkan *mention* berfungsi untuk meningkatkan interaksi langsung dengan akun lain atau komunitas tertentu. Variabel *caption_length* digunakan untuk mengukur kompleksitas dan kedalaman narasi yang disampaikan dalam konten, sedangkan

engagement_rate merupakan indikator komposit yang menggabungkan berbagai bentuk interaksi dalam satu ukuran yang terstandarisasi.

Sebelum dilakukan analisis *clustering*, data melalui tahap *preprocessing* yang mencakup pembersihan data (*data cleaning*), normalisasi, dan transformasi fitur. Normalisasi dilakukan menggunakan teknik skala (misalnya Min-Max atau *Standard Scaling*) untuk memastikan bahwa setiap variabel memiliki kontribusi yang seimbang dalam perhitungan jarak *Euclidean* pada algoritma K-Means. Tanpa normalisasi, variabel dengan skala besar seperti *likes* akan mendominasi proses *clustering* dan menghasilkan bias dalam pembentukan kluster.

Metode *Elbow*

Metode *Elbow* digunakan untuk menentukan jumlah kluster optimal dengan menganalisis nilai *Within Cluster Sum of Squares* (WCSS) atau inerti terhadap jumlah kluster (k). Secara konseptual, inerti mengukur total jarak kuadrat antara setiap titik data dengan *centroid* klusternya. Nilai inerti yang lebih kecil menunjukkan bahwa data dalam kluster lebih homogen.



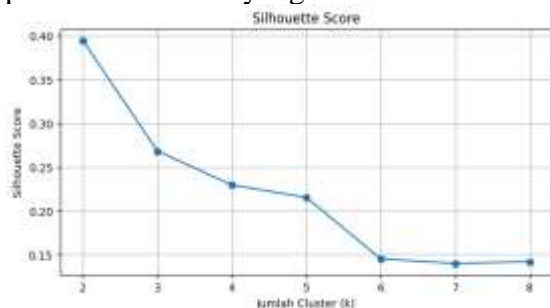
Gambar 2. Grafik *Elbow Method*

Berdasarkan hasil visualisasi pada Gambar 4.1, terlihat bahwa penurunan nilai inerti terjadi secara signifikan dari $k = 1$ ke $k = 2$, yang menunjukkan peningkatan kualitas *clustering* yang cukup besar. Namun, setelah $k = 2$, penurunan inerti menjadi semakin kecil dan cenderung melandai. Fenomena ini menunjukkan bahwa penambahan jumlah kluster setelah titik tersebut tidak memberikan peningkatan kualitas yang signifikan.

Secara teoritis, titik di mana grafik mulai melandai disebut sebagai “*elbow point*”, yang menandakan jumlah kluster optimal. Dalam penelitian ini, *elbow point* teridentifikasi pada $k = 2$, sehingga jumlah kluster yang dipilih adalah dua.

Metode *Silhouette Score*

Untuk memastikan validitas hasil dari metode *Elbow*, digunakan metode tambahan yaitu *Silhouette Score*. Metode ini mengukur seberapa baik suatu data cocok dengan kluster sendiri dibandingkan dengan kluster lain. Nilai *silhouette* berkisar antara -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan pemisahan kluster yang baik.



Gambar 3. Grafik *Silhouette Score*

Berdasarkan Gambar 4.2, nilai *silhouette* tertinggi diperoleh pada $k = 2$. Hal ini menunjukkan bahwa struktur kluster pada $k = 2$ memiliki tingkat kohesi (kedekatan dalam kluster) yang tinggi dan separasi (jarak antar kluster) yang jelas. Dengan demikian, hasil dari *Silhouette Score* mengonfirmasi bahwa dua kluster merupakan pilihan yang paling optimal. Kombinasi hasil dari metode *Elbow* dan *Silhouette* memberikan dasar yang kuat secara metodologis bahwa jumlah kluster yang digunakan dalam penelitian ini sudah tepat dan dapat dipertanggung jawabkan secara ilmiah.

Hasil Clustering dan Interpretasi Kluster

Setelah jumlah kluster ditentukan, proses clustering dilakukan menggunakan algoritma K-Means dengan parameter $k = 2$. Algoritma ini bekerja dengan cara menginisialisasi *centroid* secara acak, kemudian melakukan iterasi untuk

meminimalkan jarak antara data dan *centroid* hingga mencapai konvergensi. Hasil *clustering* menghasilkan dua kelompok utama, yaitu *Cluster 0* dan *Cluster 1*, yang masing-masing memiliki karakteristik berbeda dalam hal interaksi media sosial.

Tabel 1. Ringkasan Kluster

Kluster	Cluster 0	Cluster 1
<i>Likes</i>	129.60	298.32
<i>Comments</i>	17.19	40.97
<i>Shares</i>	8.12	22.09
<i>Saves</i>	11.36	26.27
<i>Hashtags</i>	3.29	7.06
<i>Mentions</i>	1.12	2.69
<i>Caption Length</i>	79.08	124.68
<i>Engagement Rate</i>	0.09	0.11

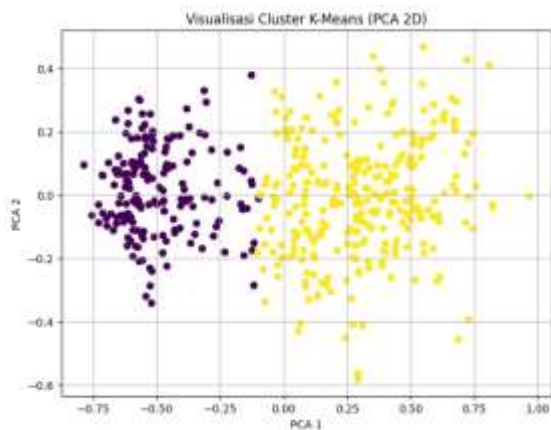
Tabel ringkasan menunjukkan bahwa terdapat perbedaan yang cukup signifikan antara kedua kluster pada hampir seluruh variabel. Cluster 1 memiliki nilai rata-rata yang jauh lebih tinggi pada variabel *likes* (298.32), *comments* (40.97), *shares* (22.09), dan *saves* (26.27) dibandingkan dengan Cluster 0. Hal ini menunjukkan bahwa konten yang dikonsumsi oleh kelompok ini memiliki daya tarik yang lebih tinggi dan mampu mendorong interaksi aktif dari *audiens*. Selain itu, nilai *caption_length* yang lebih panjang pada Cluster 1 menunjukkan bahwa audiens dalam kluster ini cenderung merespons konten dengan narasi yang lebih kompleks dan informatif. Hal ini mengindikasikan bahwa kualitas konten, khususnya dalam hal penyampaian informasi, memiliki pengaruh terhadap tingkat engagement.

Nilai *hashtags_count* dan *mentions_count* yang lebih tinggi pada Cluster 1 juga menunjukkan bahwa strategi distribusi konten yang lebih luas berkontribusi terhadap peningkatan interaksi. Penggunaan hashtag yang tepat memungkinkan konten menjangkau audiens yang lebih luas, sementara mention dapat meningkatkan interaksi langsung. Sebaliknya, Cluster 0 menunjukkan nilai yang lebih rendah pada seluruh indikator,

yang mengindikasikan bahwa kelompok ini memiliki tingkat keterlibatan yang lebih rendah. Hal ini dapat disebabkan oleh berbagai faktor, seperti kurangnya relevansi konten, rendahnya kualitas penyampaian, atau strategi distribusi yang kurang optimal.

Visualisasi dan Validasi Kluster

Untuk memperkuat interpretasi hasil clustering, digunakan teknik *Principal Component Analysis* (PCA) untuk mereduksi dimensi data menjadi dua dimensi.



Gambar 4. Visualisasi Cluster K-Means (PCA 2D)

Visualisasi pada Gambar 4 menunjukkan bahwa data terbagi menjadi dua kelompok yang relatif terpisah dengan batas yang cukup jelas. Meskipun terdapat beberapa titik yang berada di area perbatasan (*borderline*), secara umum distribusi data menunjukkan bahwa kedua kluster memiliki struktur yang berbeda. Secara teknis, PCA bekerja dengan mengubah variabel asli menjadi kombinasi linear baru (*principal components*) yang mempertahankan variansi terbesar dalam data. Dengan demikian, meskipun terjadi reduksi dimensi, informasi utama dalam data tetap terjaga. Keberhasilan pemisahan kluster dalam visualisasi ini memperkuat validitas hasil *clustering*, karena pola yang terbentuk tidak hanya didukung oleh metrik numerik, tetapi juga terlihat secara visual.

Pembahasan Hasil Penelitian

Hasil penelitian menunjukkan bahwa algoritma K-Means mampu

mengelompokkan data interaksi media sosial menjadi dua kluster utama, yaitu kluster dengan tingkat engagement tinggi dan rendah. Kluster dengan engagement tinggi memiliki nilai *likes*, *comments*, *shares*, *saves*, *hashtag*, dan *mention* yang lebih besar, sehingga menunjukkan bahwa konten yang dipublikasikan mampu menarik perhatian serta mendorong partisipasi aktif audiens. Sebaliknya, kluster dengan engagement rendah menunjukkan interaksi yang lebih pasif terhadap konten. Temuan ini memperlihatkan bahwa data interaksi media sosial dapat dimanfaatkan untuk memahami perilaku audiens secara lebih terstruktur dan mendukung strategi pemasaran yang lebih tepat sasaran.

Hasil penelitian ini sejalan dengan penelitian Ardani dan Harahap yang menyatakan bahwa media sosial menjadi kanal penting dalam meningkatkan engagement melalui strategi konten yang tepat. Selain itu, penggunaan algoritma K-Means juga mendukung penelitian Chen yang menunjukkan bahwa K-Means efektif digunakan dalam pemasaran berbasis data karena mampu mengidentifikasi pola perilaku pelanggan secara jelas. Penelitian ini juga memiliki kesamaan dengan penelitian Supangat dan Mulyani serta Sinarring dkk., yang menunjukkan bahwa clustering dapat digunakan untuk membedakan pelanggan berdasarkan tingkat loyalitas dan keterlibatan.

Namun, berbeda dengan beberapa penelitian terdahulu yang menghasilkan lebih dari dua kluster karena menggunakan variabel psikografis, persepsi merek, atau preferensi harga, penelitian ini hanya menghasilkan dua kluster optimal karena variabel yang digunakan berfokus pada interaksi media sosial. Hal ini menunjukkan bahwa jumlah kluster dipengaruhi oleh karakteristik data yang digunakan. Dalam konteks *fashion* dan kecantikan di lingkungan UKSW, hasil segmentasi ini dapat dimanfaatkan sebagai dasar personalisasi strategi pemasaran, di mana kluster dengan engagement tinggi lebih

cocok diberikan konten yang informatif dan mendalam, sedangkan kluster dengan engagement rendah memerlukan pendekatan visual yang lebih sederhana dan menarik.

Implikasi dan Kontribusi Penelitian

Secara metodologis, penelitian ini menggunakan kombinasi metode *Elbow* dan *Silhouette Score* untuk menentukan jumlah kluster optimal, sehingga keputusan penetapan $k = 2$ dapat dipertanggungjawabkan secara ilmiah. Hasil visualisasi PCA juga memperkuat validitas kluster yang terbentuk, karena pemisahan antar kelompok terlihat secara visual, bukan hanya berdasarkan metrik numerik. Secara praktis, hasil segmentasi yang menghasilkan dua kluster, kelompok dengan tingkat keterlibatan tinggi dan rendah dapat langsung dimanfaatkan sebagai dasar pengambilan keputusan pemasaran yang berbasis data. Pelaku usaha fashion dan kecantikan di lingkungan UKSW, termasuk unit kewirausahaan mahasiswa, dapat menggunakan hasil ini untuk merancang strategi konten yang berbeda bagi setiap segmen. Kluster dengan engagement tinggi dapat didekati dengan konten yang lebih informatif dan mendalam, sedangkan kluster dengan engagement rendah dapat didekati dengan konten visual yang lebih menarik dan pesan yang lebih sederhana.

Keterbatasan Penelitian

Pertama, keterbatasan variabel. Penelitian ini hanya menggunakan variabel interaksi kuantitatif dari media sosial, seperti *likes*, *comments*, *shares*, *saves*, *hashtags*, *mentions*, dan panjang *caption*. Variabel-variabel ini belum mencakup dimensi yang lebih kaya seperti aspek psikografis, persepsi merek, atau data demografis pengguna. Akibatnya, segmentasi yang dihasilkan hanya mampu membedakan *audiens* berdasarkan tingkat keterlibatan secara umum, yaitu tinggi dan rendah, tanpa menggambarkan motivasi, nilai, atau karakteristik konsumen yang

lebih mendalam. Penambahan variabel yang lebih beragam di masa mendatang berpotensi menghasilkan jumlah kluster yang lebih banyak dan segmentasi yang lebih terperinci.

Kedua, keterbatasan ruang lingkup data. Data yang digunakan bersumber dari satu lingkungan spesifik, yaitu ekosistem media sosial di sekitar Universitas Kristen Satya Wacana (UKSW). Karakteristik audiens di lingkungan kampus cenderung homogen secara demografis, karena mayoritas terdiri dari mahasiswa dengan rentang usia dan latar belakang yang relatif serupa. Kondisi ini menyebabkan pola yang dihasilkan belum tentu mewakili perilaku konsumen pada populasi yang lebih luas. Oleh karena itu, generalisasi hasil penelitian ini ke konteks pasar yang lebih umum perlu dilakukan dengan hati-hati.

Ketiga, keterbatasan algoritmik K-Means. Algoritma K-Means memiliki beberapa kelemahan yang bersifat inheren. Algoritma ini sensitif terhadap pemilihan centroid awal yang dilakukan secara acak, sehingga hasil clustering berpotensi tidak konsisten apabila tidak dikontrol dengan baik. Selain itu, K-Means mengasumsikan bahwa setiap kluster memiliki bentuk yang bulat dan ukuran yang relatif seimbang. Dalam kondisi di mana distribusi data tidak seragam atau terdapat outlier yang signifikan, algoritma ini cenderung menghasilkan pembagian kluster yang kurang optimal. Keterbatasan ini menjadi relevan mengingat data interaksi media sosial sering kali memiliki distribusi yang tidak normal, di mana sebagian kecil konten mendapatkan interaksi yang jauh lebih tinggi dibandingkan mayoritas konten lainnya.

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma K-Means clustering mampu mengelompokkan data interaksi media sosial menjadi dua kluster utama dengan karakteristik yang berbeda. Penentuan jumlah kluster menggunakan metode *Elbow*

dan *Silhouette* menunjukkan bahwa dua kluster merupakan jumlah yang paling optimal. Kluster pertama menunjukkan tingkat keterlibatan yang rendah, sedangkan kluster kedua menunjukkan tingkat keterlibatan yang tinggi. Perbedaan ini mencerminkan bahwa audiens media sosial memiliki perilaku yang beragam dalam merespons konten. Hasil penelitian juga menunjukkan bahwa variabel interaksi seperti *likes*, *comments*, *shares*, dan *saves*, serta variabel pendukung seperti panjang *caption*, penggunaan *hashtag*, dan *mention* memiliki keterkaitan dengan tingkat engagement pengguna. Hal ini menunjukkan bahwa strategi konten dan distribusi memiliki peran penting dalam meningkatkan keterlibatan audiens. Penelitian ini membuktikan bahwa pendekatan *data mining*, khususnya *clustering*, dapat digunakan untuk memahami perilaku konsumen secara lebih mendalam. Segmentasi yang dihasilkan dapat menjadi dasar dalam penerapan strategi pemasaran yang lebih personal dan tepat sasaran, terutama dalam industri *fashion* dan kecantikan yang sangat dipengaruhi oleh media sosial.

Disarankan bagi pelaku usaha di bidang *fashion* dan kecantikan untuk memanfaatkan segmentasi pelanggan dalam menyusun strategi pemasaran digital. Pendekatan yang berbeda perlu diterapkan pada setiap kluster agar strategi yang digunakan lebih efektif. Untuk kluster dengan tingkat keterlibatan tinggi, strategi pemasaran dapat difokuskan pada penyampaian informasi yang lebih detail, promosi yang lebih intensif, serta penggunaan konten yang mendorong interaksi lebih lanjut. Sementara itu, untuk kluster dengan tingkat keterlibatan rendah, disarankan untuk menggunakan pendekatan yang lebih sederhana, visual yang menarik, serta meningkatkan frekuensi komunikasi untuk membangun keterlibatan secara bertahap. Bagi peneliti selanjutnya, disarankan untuk menambahkan variabel lain seperti data demografis, psikografis, atau data transaksi

agar hasil segmentasi menjadi lebih komprehensif. Selain itu, penggunaan metode *clustering* lain dapat dilakukan sebagai pembandingan untuk memperoleh hasil yang lebih optimal.

DAFTAR PUSTAKA

- Ali, S. M. S., Deivanai, P., Roy, S., Singh, N., Vishwakarma, V., Parida, P. K., & Dhiman, K. (2025). *Optimizing marketing strategies through customer segmentation and visual analytics. International Journal of Computational and Experimental Science and Engineering*, 11(3). <https://doi.org/10.22399/ijcesen.3221>.
- Akbar, M. R. F., Farid, M. S., Pambudining, Z. C., & Putra, A. (2024). Analisis efektivitas kepemimpinan kharismatik terhadap kinerja dan keberhasilan organisasi: Studi kasus industri manufaktur. *PAR*, 1(3), 16. <https://doi.org/10.47134/par.v1i3.2549>
- Alotaibi, O., Tomy, S., & Pardede, E. (2024). A framework for cleaning streaming data in healthcare: A context and user-supported approach. *Computers*, 13(7), 175. <https://doi.org/10.3390/computers13070175>
- Alzami, F., Sambasri, F., Nabila, M., Megantara, R., Akrom, A., & Pramunendar, R. (2023). Implementation of RFM method and k-means algorithm for customer segmentation in e-commerce with Streamlit. *ILKOM Jurnal Ilmiah*, 15(1), 32–44. <https://doi.org/10.33096/ilkom.v15i1.1524.32-44>
- Analisis efektivitas dan pengaruh aktivitas pemasaran melalui media sosial pada merek. (2024). *BISMA Cendekia*, 4(2), 40–51.
- Ardani, S. P., & Harahap, W. L. (2024). Strategi konten kreatif untuk meningkatkan engagement UMKM di social media. *Jurnal DKV*, 2(2),

- 136–143.
<https://doi.org/10.62375/jdkv.v2i2.228>
- Asy'ari, T., & Winarno, A. (2024). Implementation of segmenting targeting and positioning strategies in improving marketing performance. *BIMA Journal of Business and Innovation Management*, 6(2), 280–292.
<https://doi.org/10.33752/bima.v6i2.5939>
- Chen, Q. (2024). Application of k-means algorithm in marketing. *Advances in Economics, Management and Political Sciences*, 71(1), 178–184.
<https://doi.org/10.54254/2754-1169/71/20241485>
- Chen, Q., Choi, B.-J., & Lee, S.-J. (2024). Tailoring customer segmentation strategies for luxury brands in the NFT market—The case of SUPERGUCCI. *Journal of Retailing and Consumer Services*.
<https://doi.org/10.1016/j.jretconser.2024.104121>
- Dharmawan, T., Candramaya, C., & Widharta, V. (2024). Forming dataset of the undergraduate thesis using simple clustering methods. *International Journal of Innovation in Enterprise System*, 7(1), 31–40.
<https://doi.org/10.25124/ijies.v7i01.187>
- Gürel, M., Bilgiseven, E., Aktürk, İ., & Yaman, Ö. (2024). A qualitative research on the process of drug addict individuals accessing substances. *OPUS Toplum Araştırmaları Dergisi*, 21(1), 17–31.
<https://doi.org/10.26466/opusjsr.1407945>
- Harish, S., & Malathy, C. (2023). Customer segment prediction on retail transactional data using k-means and Markov model. *Intelligent Automation & Soft Computing*, 36(1), 589–600.
<https://doi.org/10.32604/iasc.2023.032030>
- Hartama, D., & Oktaviani, S. (2025). Optimization of k-means and k-medoids clustering using DBI silhouette elbow on student data. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 11(2), 289–296.
<https://doi.org/10.33330/jurteksiv11i2.3531>
- Hulu, M. S., & Tantowi, I. (2023). Pengaruh media sosial dan persepsi merek terhadap strategi komunikasi pemasaran (studi kasus pelaku UMKM Cibubur Jakarta Timur). *Jurnal Syntax Imperatif: Jurnal Ilmu Sosial dan Pendidikan*, 4(3), 241–252.
<https://doi.org/10.36418/syntax-imperatif.v4i3.246>
- Implementation of customer segmentation model using K-means and DBSCAN for fashion industry product transaction. (2025). *International Journal on Informatics Visualization*, 9(6).
<https://doi.org/10.62527/joiv.9.6.2978>
- Kamenidou, E., Mamalis, S., Gkitsas, S., Mylona, I., & Stavrianea, A. (2023). Is generation Z ready to engage in entomophagy? A segmentation analysis study. *Nutrients*, 15(3), 525.
- Kuswardana, D., Prasetya, D., Trimono, T., & Masdiyasa, I. (2025). Comparison of elbow and silhouette methods in optimizing k-prototype clustering for customer transactions. *Jurnal Ilmiah Edutic: Pendidikan Dan Informatika*, 12(1), 43–48.
<https://doi.org/10.21107/edutic.v12i1.29744>
- Morais, S., Esteves, D., & Raposo, R. (2023). Social media and digital influencers on Instagram: A case study. *European Conference on Social Media*, 10(1), 60–67.
<https://doi.org/10.34190/ecsm.10.1.993>
- Mulaomerović-Šeta, B., Blagojević, V., Mihailović, V., & Petroselli, A.

- (2023). A silhouette-width-induced hierarchical clustering for defining flood estimation regions. *Hydrology*, *10*(6), 126. <https://doi.org/10.3390/hydrology10060126>
- Mulia, D. R. M., Rohayati, Y., & Yastica, T. V. (2025). Clustered for clarity: Leveraging AIO and K-means to decode the modest fashion market for Indonesian Muslimah. *Journal of Management and Entrepreneurship Research*, *6*(3), 235–254. <https://doi.org/10.34001/jmer.2025.906.3-72>
- Nugraha, M. D., & Dwita, V. (2023). The influence of social media usage on loyalty with customer satisfaction & brand trust as mediation variable. *Banking & Management Review*, *11*(2), 1616–1640.
- Peran literasi digital dalam meningkatkan kompetensi teknologi siswa sekolah dasar. (2024). *Jurnal Ilmiah Pendidikan Citra Bakti*, *11*(4), 1255–1267.
- Pramesti, E., Hidayat, A., Pringgabayu, D., Pradana, M., Prasetyo, B., & Rizaldi, F. (2025). The influence of social media influencers on purchase intention with brand image and customer engagement as intervening variables. *International Conference on Medical Imaging, Electronic Imaging, Information Technologies, and Sensors (MIEITS 2025)*, 23. <https://doi.org/10.1117/12.3058642>
- Qirom, D. S., Faqih, A., & Dwilestari, G. (2024). Implementasi algoritma k-means untuk klasterisasi pasien hipertensi berdasarkan karakteristik pasien. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 2056–2063.
- Rajesh, M., Chintalapudi, S., & Prasad, M. (2024). A comparative analysis of RFM-based customer segmentation with K-means and BIRCH clustering techniques. *Advances in Computer Science Research*, 977–989. https://doi.org/10.2991/978-94-6463-471-6_94
- Riyana, S., Sasujit, K., & Homdoun, N. (2023). Privacy-enhancing data aggregation for big data analytics. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, *17*(3), 452–468. <https://doi.org/10.37936/ecti-cit.2023173.252952>
- Sinarring, Hermansyah, D., Rithmaya, C. L., Zainuddin, M. Z., Aji, G. A. I. P., & Mukhlis, I. R. (2024). Analyzing customer loyalty levels through segmentation in aesthetic clinics using K-means and RFAM. *International Journal on Informatics for Development (IJID)*, *13*(2), 473–484. <https://doi.org/10.14421/ijid.2024.4841>
- Singhal, V., Singh, A. B., Ahuja, V., & Gera, R. (2023). Consumer segmentation in the fashion industry using social media: An empirical analysis. *Journal of Information and Organizational Sciences*, *47*(2), 401–421. <https://doi.org/10.31341/jios.47.2.9>
- Situmorang, H., & Bintang Fajar, G. A. (2023). Klasterisasi penjual Facebook Live di Thailand menggunakan metode k-means. *Jurnal Sains dan Teknologi (JSIT)*, *3*(3), 284–292.
- Sofyan, F. M. A., Riyandoro, A. P., Maulana, D. F., & Jaman, J. H. (2023). Penerapan data mining dengan algoritma C5.0 untuk prediksi penyakit stroke. *J-SISKO Tech (Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD)*, *6*(2), 619. <https://doi.org/10.53513/jsk.v6i2.8578>
- Suhada, S., Bahri, S., Nugraha, S. B., Hidayatulloh, T., & Wintana, D. (2023). Sistem rekomendasi produk menggunakan metode user-based collaborative filtering pada digital

- marketing. *J-INTECH*, 11(1), 159–167. <https://doi.org/10.32664/j-intech.v11i1.866>
- Sulianta, F., Ulfah, K., & Amalia, E. (2024). Revealing consumer preferences in the fashion industry using K-means clustering. *International Journal of Engineering Continuity*, 3(2), 34–53. <https://doi.org/10.58291/ijec.v3i2.280>
- Supangat, & Mulyani, Y. (2023). Customer loyalty analysis using recency, frequency, monetary (RFM) and k-means cluster for Labuan Bajo souvenirs in online store. *Journal of Information Systems and Informatics*, 5(1), 285–299. <https://doi.org/10.51519/journalisi.v5i1.421>
- Surojo, H., Indradewa, R., Negoro, D. A., & Hamdi, E. (2024). Strategi pemasaran BP Active Mobile Fuel Station PT Whez Energi Sejahtera di bawah lisensi British Petroleum–AKR. *Jurnal Pendidikan Indonesia*, 5(11), 1453–1466. <https://doi.org/10.59141/japendi.v5i11.6288>
- Yildiz, P. (2023). An evaluation of professional sense of belonging in master's theses, dissertations and articles in Turkey. *Akademik Platform Eğitim Ve Değişim Dergisi*, 6(2), 364–387. <https://doi.org/10.55150/apjec.1292708>