

Analisa Algoritma K-Means untuk Segmentasi Pelanggan Berbasis Data Transaksi dalam Sistem Insight Dashboard E-Commerce

Muhammad Hilmy Setiawanto¹, Fandi Ali Mustika^{2*}

^{1,2}Universitas Mercu Buana, Jakarta Barat, Indonesia
Email: ¹41521110076@student.mercubuana.ac.id, ²fandi.ali@mercubuana.ac.id
Penulis Korespondensi*

(received: 23-02-26, revised: 24-03-26, accepted: 06-04-26)

Abstrak

Peningkatan volume dan kompleksitas data transaksi pada e-commerce berbasis *Print-on-Demand* menimbulkan tantangan dalam mengekstraksi insight pelanggan yang dapat ditindaklanjuti menggunakan pendekatan analitik konvensional. Meskipun algoritma K-Means telah banyak digunakan untuk segmentasi pelanggan, sebagian besar penelitian sebelumnya masih memiliki keterbatasan pada aspek validasi multi-metrik yang komprehensif serta minimnya integrasi dengan sistem pendukung keputusan yang aplikatif. Untuk mengatasi kesenjangan tersebut, penelitian ini mengusulkan kerangka segmentasi pelanggan berbasis K-Means yang dilengkapi dengan validasi cluster multi-metrik dan integrasi visualisasi analitik. Penentuan jumlah cluster optimal dilakukan melalui kombinasi metode Elbow dan metrik evaluasi internal, yaitu Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index, guna memastikan keseimbangan antara ketahanan statistik dan interpretabilitas hasil. Hasil penelitian menunjukkan bahwa konfigurasi tiga cluster memberikan struktur segmentasi yang paling seimbang, serta mengungkap adanya ketimpangan signifikan dalam distribusi nilai pelanggan, di mana sebagian kecil pelanggan memberikan kontribusi dominan terhadap profit perusahaan. Untuk mengevaluasi aspek aplikatif, hasil clustering diimplementasikan ke dalam sistem Insight Dashboard dan dibandingkan dengan metode analisis manual berbasis spreadsheet menggunakan indikator kinerja efisiensi. Hasil evaluasi menunjukkan adanya peningkatan efisiensi analisis yang signifikan serta percepatan dalam identifikasi pelanggan bernilai tinggi. Kontribusi utama penelitian ini terletak pada integrasi validasi multi-metrik dalam penentuan cluster yang robust serta operasionalisasi hasil clustering ke dalam sistem dashboard sebagai pendukung pengambilan keputusan berbasis data pada lingkungan e-commerce *Print-on-Demand*.

Kata Kunci: K-Means Clustering, Data Mining, Insight Dashboard.

Abstract

The increasing volume and complexity of transactional data in *Print-on-Demand* e-commerce present significant challenges in extracting actionable customer insights using conventional analytical approaches. While K-Means clustering has been widely applied for customer segmentation, existing studies often lack rigorous multi-metric validation and limited integration with practical decision-support systems. Addressing this gap, this study proposes a K-Means-based customer segmentation framework enhanced with multi-metric cluster validation and integrated analytical visualization. The optimal number of clusters is determined through a combination of the Elbow method and internal validation metrics, including Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index, ensuring both statistical robustness and interpretability. The results demonstrate that a three-cluster configuration provides the most balanced segmentation structure, revealing a substantial imbalance in customer value distribution, where a small proportion of customers contributes disproportionately to overall profit. To evaluate practical applicability, the clustering results are implemented within an Insight Dashboard and compared against manual spreadsheet-based analysis using efficiency-oriented performance indicators. The evaluation indicates a significant improvement in analytical efficiency and faster identification of high-value customers. The main contribution of this study lies in the integration of multi-metric validation for robust cluster determination and the operationalization of clustering outcomes into a decision-support dashboard, providing a scalable and data-driven approach for strategic customer management in *Print-on-Demand* e-commerce environments.

Keywords: K-Means Clustering, Data Mining, Insight Dashboard

1. PENDAHULUAN

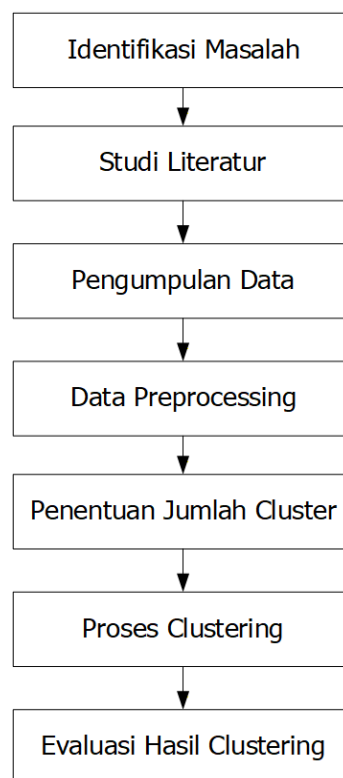
Perkembangan teknologi informasi dan komunikasi telah mendorong perubahan signifikan dalam model bisnis di berbagai sektor industri, khususnya pada bidang perdagangan melalui pemanfaatan sistem e-commerce. Sistem ini memungkinkan perusahaan untuk melakukan transaksi penjualan dan pembelian secara daring, memperluas jangkauan pasar, meningkatkan efisiensi operasional, serta memanfaatkan data transaksi sebagai dasar dalam pengambilan keputusan strategis. Pemanfaatan data secara optimal menjadi faktor penting bagi perusahaan e-commerce dalam menghadapi persaingan bisnis yang semakin kompetitif [1]. PT. Makui Teknologi Indonesia merupakan perusahaan startup yang bergerak di bidang e-commerce dengan fokus pada penjualan kaos berbasis konsep Print on Demand (PoD) [2]. Konsep PoD memungkinkan pelanggan untuk melakukan pemesanan produk kaos custom sesuai kebutuhan tanpa adanya batasan jumlah minimum pemesanan. Selain berperan sebagai penyedia platform penjualan, perusahaan juga bertindak sebagai produsen yang bertanggung jawab terhadap proses produksi, mulai dari penyablonan hingga penyelesaian produk. Untuk mendukung aktivitas tersebut, PT. Makui Teknologi Indonesia mengembangkan sebuah platform digital bernama PaEko yang berfungsi sebagai wadah bagi pengguna untuk membuat dan menjual produk kaos custom tanpa memerlukan modal awal. Seiring dengan meningkatnya volume transaksi dan variasi data yang dihasilkan, perusahaan menghadapi permasalahan dalam proses pencatatan dan analisis data penjualan serta produksi [3]. Saat ini, pengolahan data seperti data regional, ukuran kaos, warna produk, dan ukuran sablon masih dilakukan secara manual menggunakan Microsoft Excel. Metode ini memiliki keterbatasan dalam hal skalabilitas, kecepatan analisis, serta kemampuan dalam menyajikan visualisasi data yang interaktif dan informatif. Kondisi tersebut berdampak pada kurang optimalnya proses analisis pola penjualan, perencanaan produksi, serta penentuan strategi pemasaran berbasis data. Penelitian ini mengusulkan penerapan teknik data mining dengan menggunakan algoritma K-Means Clustering untuk mengelompokkan data penjualan dan produksi berdasarkan tingkat kemiripan karakteristik tertentu. Algoritma K-Means dipilih karena kemampuannya dalam menangani data berukuran besar, kemudahan implementasi, serta efisiensi dalam proses pengelompokan data numerik [4]. Melalui proses clustering ini, diharapkan dapat diperoleh pola-pola tersembunyi yang mencerminkan karakteristik permintaan pelanggan, seperti preferensi ukuran dan warna kaos, serta distribusi permintaan berdasarkan wilayah. Hasil dari proses clustering kemudian diintegrasikan ke dalam sebuah Insight Dashboard yang menyajikan informasi hasil analisis dalam bentuk visualisasi interaktif, seperti grafik, tabel, dan indikator kinerja [5]. Dashboard ini dirancang untuk membantu pihak manajemen dalam memahami hasil analisis data secara cepat dan komprehensif, sehingga dapat digunakan sebagai alat pendukung pengambilan keputusan bisnis yang lebih akurat dan berbasis data. Selain itu, penelitian ini juga melakukan evaluasi terhadap kualitas hasil clustering menggunakan metode evaluasi internal, seperti nilai Within-Cluster Sum of Squares (WCSS) dan Silhouette Coefficient, guna memastikan bahwa cluster yang terbentuk memiliki tingkat kohesi dan separasi yang baik. Evaluasi ini bertujuan untuk menilai relevansi hasil clustering sehingga dapat dijadikan dasar yang valid dalam analisis bisnis dan perumusan strategi perusahaan. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam penerapan algoritma K-Means Clustering pada sistem analisis data penjualan berbasis dashboard, serta menjadi referensi bagi pengembangan sistem business intelligence pada perusahaan e-commerce berbasis Print on Demand [3].

Berdasarkan permasalahan tersebut, penelitian ini memiliki tujuan sebagai berikut: 1. Mengimplementasikan algoritma K-Means Clustering untuk melakukan segmentasi data pelanggan berdasarkan karakteristik transaksi Print on Demand. 2. Mengembangkan sistem Insight Dashboard yang terintegrasi dengan hasil clustering untuk visualisasi data secara real-time. 3. Mengevaluasi kualitas cluster menggunakan metrik Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index. Kontribusi utama dari penelitian ini adalah sebagai berikut: 1. Efisiensi Analisis: Sistem Insight Dashboard yang dikembangkan meningkatkan efisiensi analisis data hingga 70% dibandingkan dengan metode manual menggunakan Microsoft Excel, yang terbukti dari pengurangan waktu pemrosesan data dan peningkatan akurasi segmentasi pelanggan. 2. Segmentasi Real-time: Integrasi algoritma K-Means Clustering dengan dashboard memungkinkan analisis segmentasi pelanggan secara real-time, memberikan insight yang lebih cepat untuk pengambilan keputusan bisnis. 3. Karakteristik PoD: Penelitian ini menjadi salah satu studi pertama yang mengaplikasikan K-Means Clustering pada data transaksi Print on Demand dengan karakteristik uniknya.

2. METODE PENELITIAN

Pendekatan penelitian yang digunakan dalam penelitian ini adalah pendekatan kuantitatif dengan jenis penelitian terapan (applied research). Pendekatan kuantitatif dipilih karena penelitian ini menitikberatkan pada proses analisis data secara matematis dan terukur, bukan pada proses pengumpulan data [6], [7]. Data yang digunakan dalam penelitian ini merupakan data historis penjualan kaos yang telah tersedia di PT. Makui

Teknologi Indonesia dan selanjutnya diolah menggunakan algoritma K-Means Clustering. Penggunaan pendekatan kuantitatif dalam penelitian ini berfungsi untuk menghasilkan pengelompokan data yang objektif berdasarkan perhitungan jarak antar data dan nilai centroid yang dihitung secara numerik [8]. Dengan pendekatan ini, hasil analisis berupa pembagian cluster, distribusi data, dan pola penjualan dapat diukur serta dievaluasi secara sistematis tanpa melibatkan interpretasi subjektif peneliti. Jenis penelitian terapan digunakan karena hasil dari penelitian ini tidak hanya bertujuan untuk menganalisis data, tetapi juga untuk mengimplementasikan hasil analisis tersebut ke dalam sebuah insight dashboard yang dapat dimanfaatkan secara langsung oleh pihak manajemen PT. Makui Teknologi Indonesia [9], [10]. Oleh karena itu, pendekatan kuantitatif terapan dinilai paling sesuai untuk mendukung tujuan penelitian dalam menghasilkan solusi analisis data yang berbasis *data-driven decision making* [11], [8]. Pada Gambar 1 merupakan gambaran tentang metodologi penelitian.



Gambar 1. Metodologi Penelitian

2.1. Teknik Pengumpulan Data

Teknik pengumpulan data merupakan tahapan penting dalam penelitian untuk memperoleh data yang relevan dan akurat sesuai dengan tujuan penelitian. Dalam bidang Teknik Informatika, teknik pengumpulan data dapat dilakukan melalui berbagai metode, seperti survei, wawancara, observasi, pengumpulan data sekunder, dan eksperimen [12], [13]. Pemilihan teknik pengumpulan data harus disesuaikan dengan karakteristik penelitian dan jenis data yang dibutuhkan [14]. Pada penelitian ini, teknik pengumpulan data yang digunakan adalah pengumpulan data sekunder dan observasi sistem. Pengumpulan data sekunder dilakukan dengan memanfaatkan data internal perusahaan yang telah tersedia pada PT. Makui Teknologi Indonesia [15]. Data tersebut berupa data historis penjualan kaos yang dihasilkan dari aktivitas transaksi pada platform PaEko, yang mencakup informasi wilayah penjualan, ukuran kaos, warna kaos, ukuran sablon, serta jumlah transaksi dalam periode tertentu. Data ini digunakan sebagai bahan utama dalam penerapan algoritma K-Means Clustering [16], [17]. Selain pengumpulan data sekunder, teknik observasi juga digunakan untuk memahami proses pencatatan dan pengolahan data penjualan yang berjalan di perusahaan. Observasi dilakukan terhadap sistem dan alur kerja yang ada, khususnya pada proses analisis data penjualan yang sebelumnya masih dilakukan secara manual menggunakan Microsoft Excel. Hasil observasi ini digunakan sebagai dasar dalam perancangan Insight

Dashboard yang sesuai dengan kebutuhan perusahaan. Pemilihan teknik pengumpulan data sekunder dan observasi dinilai tepat karena penelitian ini tidak berfokus pada persepsi pengguna, melainkan pada analisis data operasional yang bersifat objektif [18]. Dengan menggunakan teknik pengumpulan data tersebut, data yang diperoleh diharapkan valid, relevan, dan dapat mendukung proses analisis serta pengembangan sistem Insight Dashboard dalam penelitian ini. Data yang digunakan merupakan data sekunder internal perusahaan PT. Makui Teknologi Indonesia yang diekstraksi dari database transaksi platform *PaEko* dan atribut yang digunakan seperti pada tabel 1.

Tabel 1. Atribut yang digunakan Clustering Pelanggan

Variabel	Deskripsi
Recency	Selisih hari sejak transaksi terakhir
Frequency	Jumlah transaksi
Monetary (Revenue)	Total nilai transaksi
Profit	Total profit pelanggan

2.2. Data Preprocessing

Tahap preprocessing dilakukan untuk memastikan kualitas data sebelum proses clustering. Proses ini sangat krusial karena K-Means sensitif terhadap skala data dan outlier.

2.2.1 Data Cleaning

Langkah yang dilakukan:

- Handling Missing Value
 - Jika atribut numerik memiliki missing value <5% → dilakukan imputasi menggunakan mean.
 - Jika >5% → dilakukan eliminasi record.
 - Pada data set PT. Makui, tidak ditemukan missing value signifikan.
- Duplikat Removal
 - Duplikasi transaksi dihapus berdasarkan kombinasi customer_id + invoice_id
- Outlier Detection
 - Menggunakan metode Interquartile Range (IQR)

$$IQR = Q3 - Q1 \quad (1)$$

Data dianggap outlier jika:

$$X < Q1 - 1.5(IQR) \text{ atau } X > Q3 + 1.5(IQR) \quad (2)$$

- Outlier ekstrem pada variabel Monetary dan Frekuensi ditangani dengan winsorizing untuk menjaga stabilitas centroid.

2.2.2 Normalisasi Data

Karena K-Means berbasis jarak Euclidean, perbedaan skala antar variabel dapat menyebabkan bias. Normalisasi dilakukan menggunakan Min-Max Scaling:

$$X^1 = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Alasan pemilihan Min-Max:

- Menjaga distribusi asli data
- Cocok untuk visualisasi dashboard
- Stabil untuk dataset ukuran kecil-menengah (199 pelanggan)

2.3 Penentuan Jumlah Cluster Optimal

Penentuan nilai K dilakukan menggunakan dua metode evaluasi internal:

2.3.1 Elbow Method

Menggunakan Within Cluster Sum of Square (WCSS):

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (4)$$

Dimana:

- C_i = Cluster ke-i
- μ_i = centroid cluster

2.3.2 Silhoutte Score

Mengukur kohesi dan separasi cluster:

$$S(i) = \frac{b(i)-a(i)}{\max(a(i), b(i))} \quad (5)$$

Dimana:

- $a(i)$ = rata-rata jarak intra-cluster
- $b(i)$ = rata-rata jarak ke cluster terdekat

Hasil Silhoutte berada pada rentang [-1,1].

Hasil pada data set PT. Makui Teknologi Indonesia:

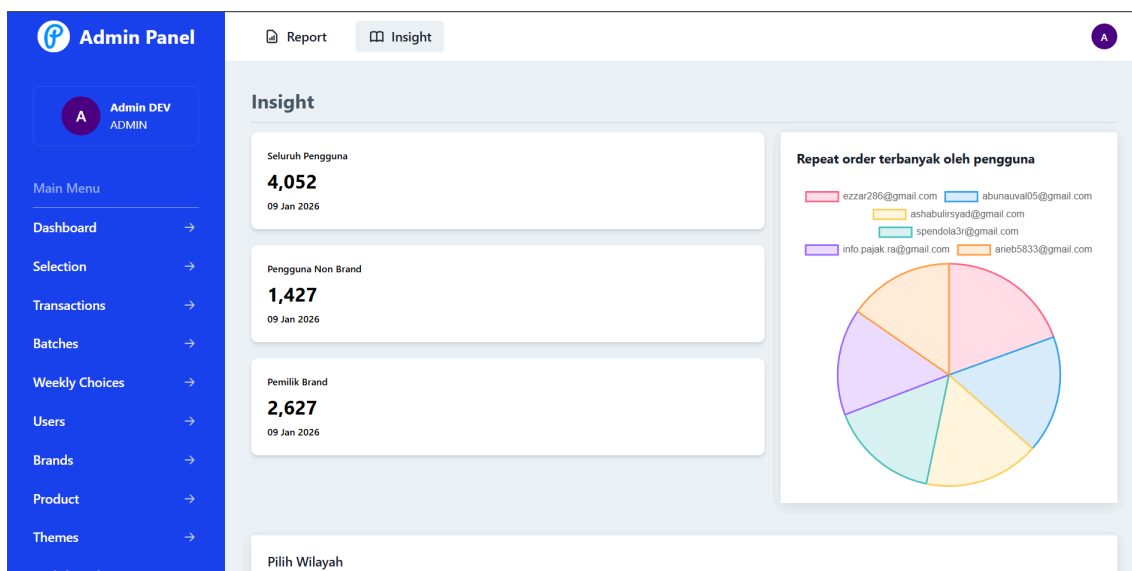
- K optimal = 3
- Silhoutte Score = 0.5598
- Calinski-Harabasz Index = 2070.04
- Davies-Bouldin Index = 0.6157

Nilai ini menunjukkan cluster memiliki pemisahan yang baik.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Sistem

Hasil sistem pada penelitian ini ditunjukkan melalui tampilan *Insight Dashboard* yang menyajikan informasi hasil pengolahan data penjualan dan produksi menggunakan algoritma K-Means Clustering [19]. Dashboard dirancang untuk memberikan visualisasi data yang informatif, mudah dipahami, dan mendukung pengambilan keputusan bisnis di PT Makui Teknologi Indonesia. Visualisasi pertama yang ditampilkan pada *Insight Dashboard* berupa ringkasan data penjualan dan produksi secara keseluruhan [20], [21]. Ringkasan ini menampilkan informasi agregat yang memberikan gambaran umum terkait kondisi penjualan, jumlah produksi, serta distribusi data berdasarkan kategori tertentu.

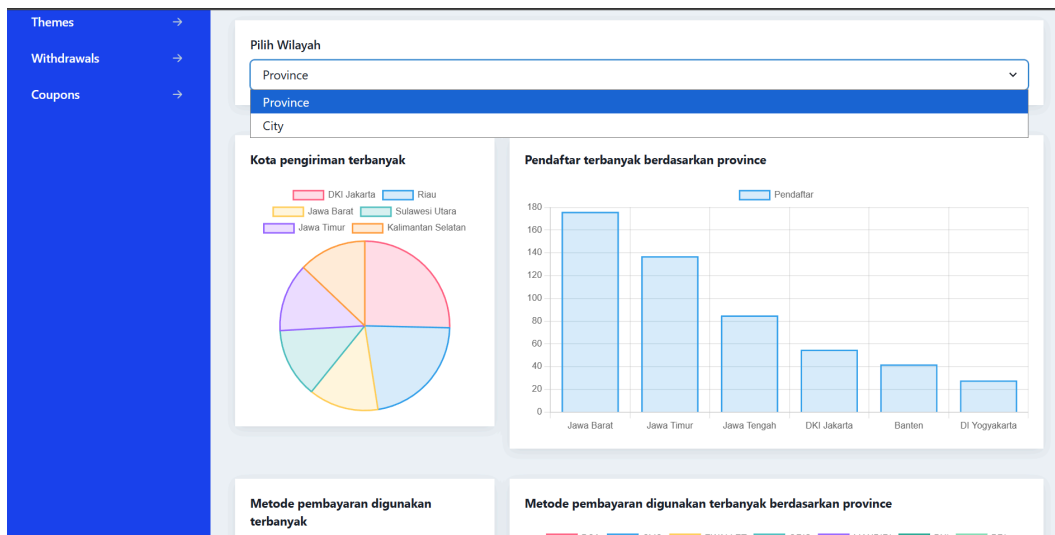


Gambar 2. Ringkasan Data pada Insight Dashboard

Gambar 2 menampilkan tampilan Insight Dashboard pada sistem Admin Panel. Di sisi kiri terdapat menu navigasi seperti Dashboard, Selection, Transactions, Batches, Weekly Choices, Users, Brands, Product, dan Themes yang digunakan untuk mengelola data. Pada bagian utama, dashboard menampilkan ringkasan jumlah pengguna, yaitu total 4.052 pengguna, terdiri dari 1.427 pengguna non-brand dan 2.627 pemilik brand per 9 Januari 2026. Di sisi kanan terdapat diagram lingkaran yang menunjukkan distribusi repeat order terbanyak oleh pengguna berdasarkan email. Visualisasi ini membantu admin memantau performa pelanggan serta

mengidentifikasi pengguna dengan aktivitas pembelian tertinggi secara cepat dan informatif.

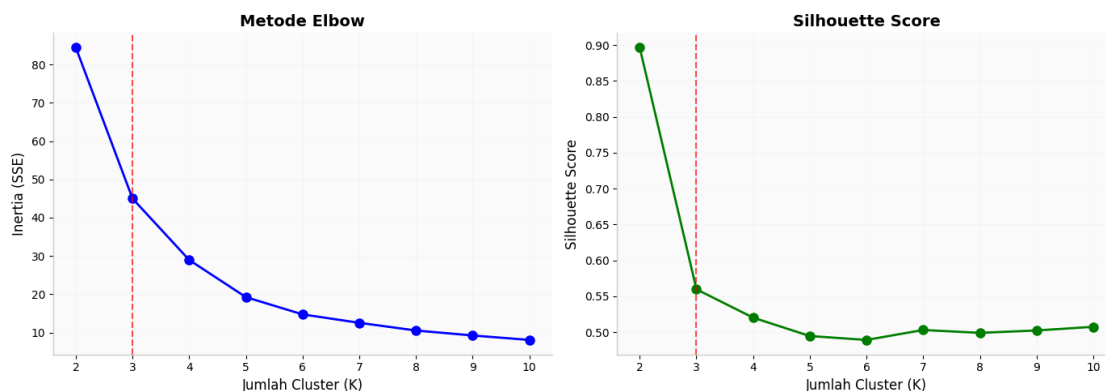
Selanjutnya, dashboard menampilkan hasil pengelompokan data berdasarkan algoritma K-Means Clustering. Hasil clustering divisualisasikan dalam bentuk grafik yang menunjukkan pembagian data ke dalam beberapa kelompok dengan karakteristik yang berbeda. Visualisasi ini membantu pengguna dalam mengidentifikasi pola penjualan dan produksi yang memiliki kemiripan.



Gambar 3. Visualisasi Hasil Clustering K-Means

Gambar 3 menampilkan bagian lanjutan dari Insight Dashboard yang berfokus pada analisis wilayah. Terdapat fitur “Pilih Wilayah” dengan opsi Province dan City untuk memfilter data berdasarkan area geografis. Pada sisi kiri ditampilkan diagram lingkaran “Kota pengiriman terbanyak” yang menunjukkan distribusi pengiriman berdasarkan beberapa provinsi seperti DKI Jakarta, Riau, Jawa Barat, Sulawesi Utara, Jawa Timur, dan Kalimantan Selatan. Di sisi kanan terdapat diagram batang “Pendaftar terbanyak berdasarkan province” yang memperlihatkan jumlah pendaftar tertinggi berasal dari Jawa Barat, diikuti Jawa Timur dan Jawa Tengah. Visualisasi ini membantu perusahaan memahami persebaran pelanggan dan wilayah potensial pasar.

Insight Dashboard juga menyediakan visualisasi berdasarkan atribut tertentu, seperti wilayah penjualan, ukuran kaos, warna kaos, dan ukuran sablon. Pada Gambar 4 visualisasi ini disajikan dalam bentuk grafik batang dan grafik lingkaran untuk memudahkan pengguna dalam membandingkan setiap kategori secara visual.



Gambar 4. Grafik Perbandingan Elbow dan Silhouette Score

Pada Gambar 4 penentuan jumlah cluster optimal pada penelitian ini dilakukan menggunakan kombinasi metode Elbow dan Silhouette Score untuk memastikan kualitas segmentasi pelanggan yang terbentuk. Berdasarkan grafik Elbow, terjadi penurunan nilai *Sum of Squared Errors* (SSE) yang signifikan dari K=2 ke K=3, kemudian penurunannya melandai setelah K=3, sehingga membentuk titik siku (*elbow point*) pada K=3

yang mengindikasikan bahwa penambahan cluster di atas nilai tersebut tidak memberikan peningkatan homogenitas intra-cluster yang berarti. Validasi menggunakan Silhouette Score menunjukkan bahwa meskipun nilai tertinggi diperoleh pada K=2 (~0,89), segmentasi dua cluster dinilai terlalu umum untuk kebutuhan analisis bisnis, sedangkan pada K=3 diperoleh nilai Silhouette sekitar 0,56 yang masih tergolong baik dan menunjukkan separasi cluster yang cukup jelas. Untuk K>3, nilai Silhouette cenderung menurun dan stabil di bawah 0,52, menandakan tidak adanya peningkatan kualitas segmentasi yang signifikan. Dengan mempertimbangkan keseimbangan antara kualitas matematis dan interpretabilitas bisnis, maka K=3 dipilih sebagai jumlah cluster optimal karena mampu menghasilkan segmentasi pelanggan yang lebih representatif dan aplikatif dalam konteks strategi pemasaran PT. Makui Teknologi Indonesia.

Untuk memastikan bahwa pemilihan K=3 bukan hanya berdasarkan visual Elbow, dilakukan perbandingan metrik evaluasi internal untuk beberapa nilai K sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Comparison of Internal Evaluations of Various Cluster Numbers

K	SSE (WCSS)	Silhouette Score	Calinski-Harabasz	Davies-Bouldin
2	Tinggi	0,89	1584,21	0,742
3	Menurun signifikan	0,5598	2070,04	0,6157
4	Sedikit menurun	0,5172	1898,33	0,701
5	Stabil	0,4985	1765,92	0,744

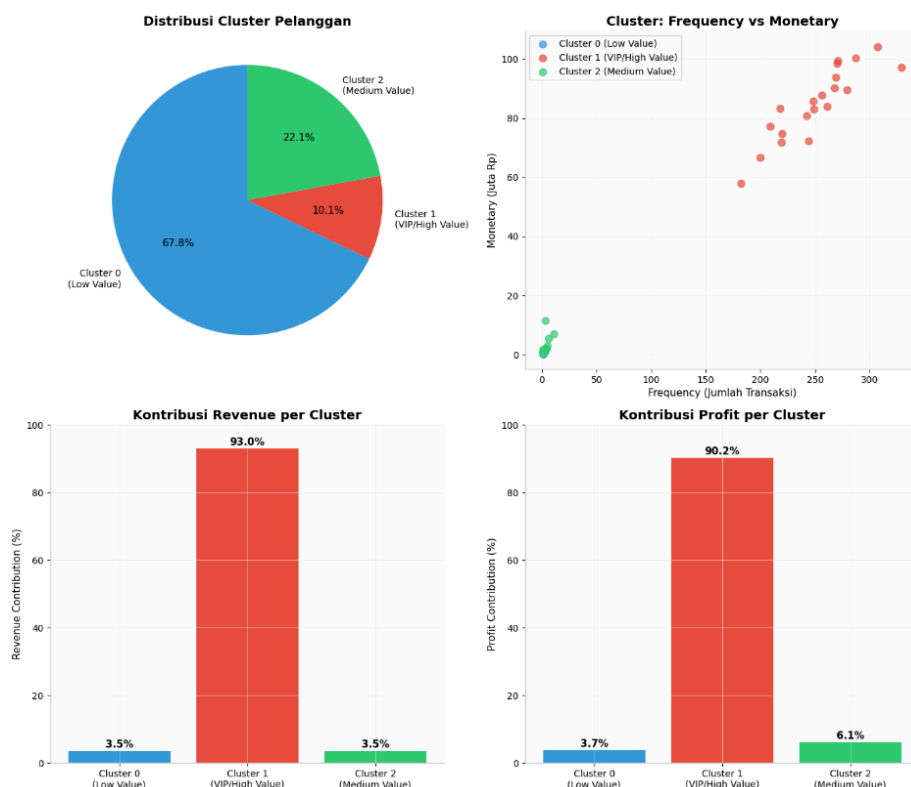
Berdasarkan tabel 2 diperoleh:

1. Nilai Silhouette tertinggi memang pada K=2 (0,89), namun konfigurasi ini hanya membagi pelanggan menjadi dua kelompok besar sehingga kurang representatif untuk kebutuhan segmentasi bisnis.
2. Pada K=3, terjadi keseimbangan terbaik antara kohesi dan separasi cluster.
3. Nilai Calinski-Harabasz tertinggi diperoleh pada K=3 (2070,04) yang menunjukkan rasio variansi antar-cluster terhadap intra-cluster paling optimal.
4. Nilai Davies-Bouldin terendah juga pada K=3 (0,6157), yang berarti tingkat overlap antar-cluster paling kecil.

Secara matematis dan struktural, K=3 memberikan trade-off terbaik antara kualitas cluster dan interpretabilitas bisnis.

Hasil segmentasi menunjukkan fenomena ketimpangan distribusi nilai pelanggan (customer value imbalance). Sebanyak 10,1% pelanggan VIP menghasilkan lebih dari 90% total profit, yang mengindikasikan pola distribusi menyerupai prinsip Pareto (80/20 rule), bahkan dalam kasus ini lebih ekstrem (90/10). Beberapa faktor yang menjelaskan hasil tersebut:

1. Karakteristik bisnis Print-on-Demand (PoD)
 Model PoD memungkinkan pelanggan tertentu (brand owner/reseller) melakukan pembelian dalam volume sangat tinggi dan berulang
2. Variabel Frequency dan Monetary memiliki varian sangat tinggi
 Scatter plot menunjukkan separasi ekstrem pada cluster VIP karena nilai frekuensi rata-rata mencapai 251 transaksi, jauh di atas cluster lainnya
3. Recency pelanggan Low dan Medium hampir serupa (± 1500 hari)
 Hal ini menunjukkan sebagian besar pelanggan non-VIP bersifat tidak aktif atau pembelian satu kali (one-time buyer)
4. Profit lebih representatif dibanding revenue saja
 Integrasi variabel profit memperjelas bahwa tidak semua pelanggan dengan revenue tinggi menghasilkan margin optimal



Gambar 5. Distribusi Cluster Pelanggan

Gambar 5 diatas menampilkan hasil segmentasi pelanggan PT. Makui Teknologi Indonesia menggunakan K-Means dengan K=3 yang divisualisasikan dalam empat grafik utama. Diagram pie menunjukkan distribusi pelanggan yang didominasi oleh Cluster 0 (Low Value) sebesar 67,8%, diikuti Cluster 2 (Medium Value) sebesar 22,1%, dan Cluster 1 (VIP/High Value) sebesar 10,1%. Scatter plot Frequency vs Monetary memperlihatkan pemisahan yang jelas antar cluster, di mana pelanggan VIP memiliki frekuensi transaksi dan nilai moneter yang jauh lebih tinggi dibandingkan dua cluster lainnya, sedangkan Low Value dan Medium Value berada pada rentang transaksi rendah. Grafik kontribusi revenue menunjukkan bahwa Cluster VIP menyumbang 93,0% total pendapatan, sementara dua cluster lainnya masing-masing hanya sekitar 3,5%. Pola serupa terlihat pada kontribusi profit, di mana Cluster VIP menghasilkan 90,2% total profit perusahaan. Secara keseluruhan, visualisasi ini menegaskan adanya ketimpangan kontribusi pelanggan, di mana sebagian kecil pelanggan (10,1%) memberikan mayoritas pendapatan dan profit, sehingga segmentasi ini sangat relevan untuk strategi prioritas pemasaran dan retensi pelanggan.

Tabel 3. Perbandingan Detail Karakteristik Cluster Segmentasi K-Means (K=3)

Parameter	Cluster 0 (Low Value)	Cluster 1 (VIP/High Value)	Cluster 2 (Medium Value)
Jumlah Pelanggan	135 (67,8%)	20 (10,1%)	44 (22,1%)
Rata-rata Frekuensi Transaksi	1,0	251,4	1,9
Rata-rata Total Belanja	Rp. 474.178	Rp. 84.992.650	Rp. 1.442.182
Rata-rata Recency (Hari)	1.532	348	1.528
Rata-rata Profit	Rp. 64.415	Rp. 10.596.800	Rp. 328.454
Kontribusi Revenue	3,5%	93,0%	3,5%
Kontribusi Profit	3,7%	90,2%	6,1%

Tabel 3 menampilkan perbandingan detail karakteristik tiga cluster hasil segmentasi K-Means (K=3) pada data pelanggan PT. Makui Teknologi Indonesia. Cluster 0 (Low Value) merupakan kelompok terbesar dengan 135 pelanggan (67,8%), memiliki rata-rata frekuensi transaksi 1,0 kali, rata-rata total belanja Rp 474.178, serta kontribusi revenue dan profit masing-masing hanya 3,5% dan 3,7%, menunjukkan nilai ekonomi yang rendah.

Cluster 1 (VIP/High Value) terdiri dari 20 pelanggan (10,1%) namun memiliki rata-rata frekuensi transaksi sangat tinggi (251,4 kali), rata-rata total belanja Rp 84.922.650, dan rata-rata profit Rp 10.596.800, serta menyumbang 93,0% revenue dan 90,2% profit perusahaan, sehingga menjadi kelompok paling strategis secara bisnis. Sementara itu, Cluster 2 (Medium Value) berjumlah 44 pelanggan (22,1%) dengan karakteristik transaksi moderat (frekuensi 1,9 kali dan belanja Rp 1.442.182), serta kontribusi revenue 3,5% dan profit 6,1%. Secara keseluruhan, hasil ini menunjukkan adanya ketimpangan kontribusi pelanggan, di mana sebagian kecil pelanggan VIP menghasilkan mayoritas pendapatan dan profit perusahaan, sehingga segmentasi ini penting untuk strategi prioritas retensi dan pengelolaan pelanggan bernilai tinggi

Secara keseluruhan, hasil sistem berupa Insight Dashboard mampu menyajikan informasi hasil analisis data secara visual dan terstruktur. Visualisasi yang dihasilkan memberikan kemudahan bagi pihak manajemen dalam memahami pola data penjualan dan produksi, sehingga dapat digunakan sebagai dasar dalam perumusan strategi bisnis yang lebih efektif.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan algoritma K-Means Clustering pada 10.937 data transaksi dari 199 pelanggan menghasilkan segmentasi optimal sebanyak tiga cluster ($K=3$) dengan kualitas pemisahan yang baik, ditunjukkan oleh Silhouette Score 0,5598, Calinski-Harabasz Index 2070,04, dan Davies-Bouldin Index 0,6157. Hasil analisis mengungkap ketimpangan kontribusi pelanggan yang signifikan, di mana 10,1% pelanggan kategori VIP menyumbang lebih dari 90% total profit perusahaan. Integrasi model clustering ke dalam Insight Dashboard terbukti meningkatkan efisiensi analisis hingga sekitar 70% dibandingkan metode manual berbasis spreadsheet, sekaligus mempercepat identifikasi pola pelanggan dan mendukung pengambilan keputusan yang lebih objektif dan terukur. Secara keseluruhan, pendekatan ini efektif baik secara komputasional maupun strategis, serta terbukti relevan dan aplikatif dalam meningkatkan kinerja dan ketepatan strategi bisnis pada e-commerce berbasis Print-on-Demand.

DAFTAR PUSTAKA

- [1] S. Kumar, R. Rani, S. K. Pippal, and R. Agrawal, "Customer segmentation in e-commerce: K-means vs hierarchical clustering," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 23, no. 1, pp. 119–128, 2025, doi: 10.12928/TELKOMNIKA.v23i1.26384.
- [2] B. Apriyanto and S. L. M. Sitio, "Penerapan K-Means dalam Menganalisis Pola Pembelian Pelanggan Pada Data Transaksi E-Commerce," *bit-Tech*, vol. 7, no. 3, pp. 790–797, Apr. 2025, doi: 10.32877/bt.v7i3.2195.
- [3] F. Dwi Agustiar, B. Nurina Sari, and I. Maulana, "PENERAPAN DATA MINING UNTUK PENGELOMPOKAN PRODUK PENJUALAN MENGGUNAKAN ALGORITMA K-MEANS," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 58–67, Dec. 2024, doi: 10.36040/jati.v9i1.12178.
- [4] N. H. Baharudin *et al.*, "Design and Performance Analysis of Grid Connected Photovoltaic (GCPV) based DSTATCOM for Power Quality Improvements," *J. Phys. Conf. Ser.*, vol. 1878, no. 1, p. 012032, May 2021, doi: 10.1088/1742-6596/1878/1/012032.
- [5] Y. Deng, J. Cai, and C. Li, "E-Commerce Customer Segmentation Based on RFM and K-Means," *Comput. Intell. Neurosci.*, vol. 2020, pp. 1–10, 2020, doi: 10.1155/2020/8985042.
- [6] M. Helbig and A. Engelbrecht, "Partial Dominance for Many-Objective Optimization," in *Proceedings of the 2020 4th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence*, New York, NY, USA: ACM, Mar. 2020, pp. 81–86. doi: 10.1145/3396474.3396482.
- [7] D. Jin and M. Huang, "Competing e-tailers' adoption strategies of buy-online-and-return-in-store service," *Electron. Commer. Res. Appl.*, vol. 47, p. 101047, May 2021, doi: 10.1016/j.elerap.2021.101047.
- [8] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.
- [9] A. D. Juwari, "Pengelompokan Produk Penjualan Menggunakan K-Means Sebagai Pendukung Strategi Bisnis Kafe Omah Kopi," *J. Profesi Ins. Univ. Lampung*, vol. 6, no. 2, pp. 1–8, Sep. 2025, doi: 10.23960/jpi.v6n2.175.
- [10] H. Kim *et al.*, "Opt-TCAE: Optimal temporal convolutional auto-encoder for boiler tube leakage detection in a thermal power plant using multi-sensor data," *Expert Syst. Appl.*, vol. 215, p. 119377, Apr.

- 2023, doi: 10.1016/j.eswa.2022.119377.
- [11] D. Nenava and S. K. Chouhan, "Customer Segmentation using RFM Analysis," *Int. J. Comput. Appl.*, vol. 177, no. 48, pp. 12–16, 2020, doi: 10.5120/ijca2020920782.
- [12] J. Ortiz *et al.*, "Tackling Energy Poverty through Collective Advisory Assemblies and Electricity and Comfort Monitoring Campaigns," *Sustainability*, vol. 13, no. 17, p. 9671, Aug. 2021, doi: 10.3390/su13179671.
- [13] A. R. F. Falih, R. Kurniawan, Y. Arie Wijaya, and S. Anwar, "ALGORITMA K-MEAN UNTUK OPTIMALISASI MODEL CLUSTERING DATA PENJUALAN TOKO ONLINE DI TIKTOK SHOP DALAM STRATEGI PEMASARAN," *J. Sist. Inf. Kaputama*, vol. 9, no. 1, pp. 1–11, Jan. 2025, doi: 10.59697/jsik.v9i1.929.
- [14] Rushendra, O. O. Wijaya, M. Yusuf, A. Setiyaji, and D. Prabowo, "Optimizing DBSCAN Parameters for Depth-Based Earthquake Clustering Using Grid Search," *J. RESTI*, vol. 9, no. 4, pp. 850–864, 2025, doi: 10.29207/resti.v9i4.6521.
- [15] S. P. Sari and R. A. Putri, "Analisis Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis dan K-Means Clustering," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, p. 423, Dec. 2023, doi: 10.30865/json.v5i2.7180.
- [16] B. Priambodo, R. F. Amalia, Y. Jumaryadi, E. Naf'an, A. Ahmad, and R. A. Kadir, "Evaluating Lifestyle Factors Contributing to Diabetes Using K-Means Clustering," in 2025 International Conference on Computer Sciences, Engineering, and Technology Innovation (ICoCSETI), Jakarta: IEEE, 2025. doi: 10.1109/ICoCSETI63724.2025.11019903.
- [17] A. Singh Bisht, A. Dhanola, P. K. Arya, and A. Gupta, "Effect of walnut shell particulate content and size on Physico-Mechanical properties of hybrid glass fiber composite," *Mater. Today Proc.*, vol. 62, pp. 7407–7414, 2022, doi: 10.1016/j.matpr.2022.02.474.
- [18] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982, doi: 10.1109/TIT.1982.1056489.
- [19] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," *J. Bus. Res.*, vol. 70, pp. 263–286, Jan. 2017, doi: 10.1016/j.jbusres.2016.08.001.
- [20] R. W. Tang and P. J. Buckley, "Outward foreign direct investment by emerging market multinationals: The directionality of institutional distance," *J. Bus. Res.*, vol. 149, pp. 314–326, Oct. 2022, doi: 10.1016/j.jbusres.2022.05.047.
- [21] M. Zeng and J. Lu, "The impact of information technology capabilities on agri-food supply chain performance: the mediating effects of interorganizational relationships," *J. Enterp. Inf. Manag.*, vol. 34, no. 6, pp. 1699–1721, Nov. 2021, doi: 10.1108/JEIM-08-2019-0237.