

Segmentasi Pelanggan Produk Agribisnis Berdasarkan Pola Belanja Menggunakan *K-Means Clustering*

Customer Segmentation of Agribusiness Products Based On Purchasing Patterns Using K-Means Clustering

Syifa Fauziah¹, Imas Wildan Rafiqah^{1*}, Alfian Destha Joanda², Nurhana Jafaruddin³

¹Prodi Agribisnis, Universitas Singaperbangsa Karawang, Karawang

²Program Profesi Insinyur, Universitas Bina Nusantara, Jakarta

³Prodi Agribisnis, Universitas Borobudur, Jakarta

*Email: imas.wildan@faperta.unsika.ac.id

(Diterima 30-05-2026; Disetujui 30-07-2026)

ABSTRAK

Penelitian ini bertujuan untuk menganalisis segmentasi pelanggan produk agribisnis berdasarkan pola belanja menggunakan pendekatan *machine learning*. Permasalahan utama dalam sektor agribisnis adalah heterogenitas perilaku pelanggan yang menyebabkan strategi pemasaran kurang optimal. Data yang digunakan merupakan dataset wholesale customers sebanyak 440 observasi. Metode analisis menggunakan K-Means Clustering dengan tahapan eksplorasi data, penanganan outlier, penentuan jumlah cluster optimal menggunakan metode elbow, serta evaluasi menggunakan silhouette score. Hasil penelitian menunjukkan bahwa terbentuk lima cluster pelanggan dengan karakteristik yang berbeda secara signifikan. Segmentasi ini mampu memberikan insight strategis dalam penentuan target pasar dan pengambilan keputusan berbasis data.

Kata kunci: Segmentasi Pelanggan, Agribisnis, *K-Means*, *Machine Learning*, *Clustering*

ABSTRACT

This study aims to analyze customer segmentation of agribusiness products based on purchasing patterns using a machine learning approach. The main issue in agribusiness is the heterogeneity of customer behavior, which leads to suboptimal marketing strategies. The dataset used consists of 440 observations from wholesale customers. The analysis method employs K-Means Clustering, including data exploration, outlier handling, determining the optimal number of clusters using the elbow method, and evaluation using silhouette score. The results show that five distinct customer clusters were formed with significantly different characteristics. This segmentation provides strategic insights for market targeting and data-driven decision-making.

Keywords: Customer Segmentation, Agribusiness, Clustering, Machine Learning, K-Means

PENDAHULUAN

Sektor agribisnis merupakan salah satu pilar penting dalam perekonomian, khususnya di negara berkembang seperti Indonesia. Peran agribisnis tidak hanya terbatas pada produksi, tetapi juga mencakup distribusi, pemasaran, hingga konsumsi produk pertanian. Pada perkembangannya, sektor ini menghadapi tantangan yang semakin kompleks, terutama dalam memahami perilaku konsumen yang dinamis dan beragam. Tantangan utama yang dihadapi pada usaha di sektor agribisnis adalah rendahnya efisiensi dan produktivitas usaha, yang seringkali dipengaruhi oleh keterbatasan dalam pengelolaan pasar dan pemahaman terhadap perilaku konsumen (Fauziah et al., 2020).

Perubahan perilaku konsumen tersebut berdampak pada meningkatnya kebutuhan akan pendekatan berbasis data dalam pengambilan keputusan. Pendekatan tradisional yang mengandalkan intuisi dan pengalaman menjadi kurang relevan dalam menghadapi kompleksitas data saat ini. Oleh karena itu, pemanfaatan Machine Learning menjadi solusi yang semakin banyak digunakan dalam analisis perilaku konsumen. Teknologi ini memungkinkan pengolahan data dalam jumlah besar untuk menemukan pola tersembunyi yang tidak dapat diidentifikasi secara manual (Zheng, 2023).

Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan tersebut adalah segmentasi pelanggan. Segmentasi pelanggan merupakan proses pengelompokan konsumen berdasarkan karakteristik tertentu, seperti perilaku pembelian, preferensi produk, dan tingkat konsumsi (Kotler & Keller, 2016). Dengan adanya segmentasi yang jelas, perusahaan dapat menyusun strategi pemasaran yang lebih terarah dan efisien.

Segmentasi pelanggan bertujuan untuk mengelompokkan konsumen ke dalam kelompok homogen berdasarkan karakteristik tertentu, seperti pola pembelian dan preferensi produk. Penelitian terbaru menunjukkan bahwa segmentasi berbasis data mampu meningkatkan efektivitas strategi pemasaran dan memberikan nilai tambah dalam pengambilan keputusan bisnis (Awaliyah et al., 2024; Nugroho et al., 2024).

Seiring dengan perkembangan teknologi, pendekatan segmentasi pelanggan mengalami pergeseran dari metode tradisional ke pendekatan berbasis data. Pemanfaatan Machine Learning dalam analisis perilaku konsumen memungkinkan pengolahan data dalam skala besar untuk menemukan pola tersembunyi yang tidak dapat diidentifikasi secara manual. Pendekatan ini menjadi semakin relevan dalam era big data, di mana volume, variasi, dan kecepatan data terus meningkat.

Salah satu teknik yang banyak digunakan dalam segmentasi berbasis data adalah K-Means Clustering. Metode ini termasuk dalam kategori unsupervised learning yang mampu mengelompokkan data tanpa label dengan cara meminimalkan jarak antar data dalam satu kelompok dan memaksimalkan jarak antar kelompok (Han et al., 2012). Keunggulan metode ini terletak pada kemampuannya dalam mengidentifikasi pola secara efisien serta kemudahan dalam implementasi.

Pada implementasinya, teknik clustering menjadi metode yang paling umum digunakan dalam segmentasi pelanggan. Salah satu algoritma yang populer adalah K-Means Clustering, yang mampu mengelompokkan data berdasarkan kedekatan karakteristik antar objek. Metode ini dinilai efektif karena memiliki kemampuan komputasi yang relatif sederhana serta hasil yang mudah diinterpretasikan (Li et al., 2023). Selain itu, penelitian oleh Zhao (2024) menunjukkan bahwa K-Means mampu menghasilkan segmentasi pelanggan yang jelas dan dapat digunakan sebagai dasar dalam penyusunan strategi pemasaran.

Meskipun demikian, beberapa penelitian menunjukkan bahwa penerapan segmentasi pelanggan masih memiliki keterbatasan, terutama dalam konteks implementasi praktis. Sebagian besar studi masih berfokus pada tahap analisis tanpa mengintegrasikan hasil ke dalam sistem yang dapat digunakan secara langsung oleh pelaku usaha (Awaliyah et al., 2024; Nugroho et al., 2024). Selain itu, dalam konteks agribisnis, penelitian terkait segmentasi pelanggan berbasis Machine Learning masih relatif terbatas dibandingkan sektor ritel atau e-commerce yang telah lebih dahulu mengadopsi pendekatan berbasis data (Zheng, 2023; Li et al., 2023).

Berbagai penelitian terdahulu menunjukkan bahwa penggunaan metode clustering dalam segmentasi pelanggan dapat meningkatkan efektivitas strategi pemasaran. Indika dan Jovita (2017) menyatakan bahwa pemanfaatan data dalam pemasaran mampu meningkatkan minat beli konsumen. Selain itu, Tan et al. (2006) menjelaskan bahwa teknik data mining, termasuk clustering, sangat efektif dalam mengidentifikasi pola perilaku konsumen yang kompleks. Namun demikian, penerapan metode ini dalam konteks agribisnis masih relatif terbatas, terutama dalam mengkaji pola belanja pelanggan secara spesifik.

Keterbatasan penelitian sebelumnya terletak pada kurangnya integrasi antara analisis data dan implementasi hasil dalam bentuk sistem yang aplikatif. Sebagian besar penelitian hanya berhenti pada tahap analisis tanpa mengembangkan model yang dapat digunakan secara langsung oleh pelaku usaha. Oleh karena itu, penelitian ini tidak hanya berfokus pada proses segmentasi, tetapi juga mengintegrasikan hasil analisis ke dalam bentuk visualisasi dan dashboard interaktif untuk mendukung pengambilan keputusan.

Penelitian ini menjadi penting karena menawarkan pendekatan yang lebih komprehensif dalam memahami perilaku pelanggan agribisnis melalui pemanfaatan teknologi Data Science. Dengan mengelompokkan pelanggan berdasarkan pola belanja, diharapkan pelaku usaha dapat memperoleh insight yang lebih mendalam mengenai karakteristik konsumen, sehingga strategi pemasaran yang diterapkan menjadi lebih efektif dan efisien.

Berdasarkan uraian tersebut, permasalahan dalam penelitian ini difokuskan pada bagaimana mengidentifikasi pola belanja pelanggan, menentukan jumlah cluster optimal, serta menganalisis karakteristik masing-masing segmen pelanggan. Tujuan penelitian ini adalah untuk menghasilkan segmentasi pelanggan yang dapat digunakan sebagai dasar dalam penyusunan strategi pemasaran berbasis data di sektor agribisnis.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis berbasis *Data Science*. Pendekatan ini dipilih untuk menemukan pola tersembunyi dari data melalui teknik statistik dan komputasi. Data yang digunakan adalah data sekunder dari Kaggle, yaitu *Wholesale Customers Dataset*, yang terdiri atas 440 observasi. Variabel yang diteliti mencakup delapan (8) jenis produk rumah tangga dan grosir yang dibeli oleh konsumen, seperti sayuran, susu, produk grosir, dan makanan beku. Seluruh proses pengolahan data dilakukan menggunakan Google Collab dengan bahasa pemrograman Python.

Secara sistematis, tahapan analisis data dalam penelitian ini dibagi menjadi lima langkah utama:

1. Analisis Eksplorasi Data (EDA) dan Visualisasi

Tahap awal dimulai dengan *Exploratory Data Analysis* (EDA) dan statistik deskriptif untuk memahami karakteristik dasar dan persebaran data (Geron, 2022). Hubungan atau korelasi antarvariabel pengeluaran konsumen dianalisis menggunakan grafik *heatmap*, sementara pola distribusinya dilihat secara visual melalui *pairplot* (Waskom, 2021).

2. Pra-pemrosesan Data (*Data Preprocessing*)

Sebelum masuk ke tahap pemodelan, data dibersihkan terlebih dahulu agar hasilnya akurat. Langkah pertama adalah memastikan tidak ada data yang hilang (*missing values*), karena data yang kosong dapat mengganggu perhitungan algoritma (Han et al., 2012). Langkah kedua adalah mendeteksi dan menangani data ekstrem (*outlier*) menggunakan metode *Interquartile Range* (IQR). Data yang berada di luar jangkauan rumus berikut akan dihapus atau disesuaikan:

$$\text{Batas Bawah} = Q_1 - 1.5 \times \text{IQR}$$

$$\text{Batas Atas} = Q_3 + 1.5 \times \text{IQR}$$

Penanganan *outlier* ini penting karena algoritma K-Means sangat sensitif terhadap nilai-nilai ekstrem yang dapat menggeser pusat kelompok (klaster) secara signifikan (Indriyani dan Irfiani, 2019).

3. Reduksi Dimensi dengan PCA

Data ini memiliki delapan variabel yang cukup kompleks, dilakukan penyederhanaan data menggunakan *Principal Component Analysis* (PCA). Metode PCA berfungsi untuk mengurangi jumlah dimensi (variabel) dengan cara mengubahnya menjadi variabel baru yang saling tegak lurus (*orthogonal*) tanpa membuang informasi penting di dalamnya (Bruce et al., 2020). Selain itu, PCA memudahkan kita untuk memvisualisasikan hasil pengelompokan ke dalam grafik dua dimensi (2D).

4. Pemodelan Klasterisasi K-Means dan Metode Elbow

Data yang telah disederhanakan kemudian dikelompokkan menggunakan algoritma *K-Means Clustering*. Algoritma ini bekerja dengan cara membagi data ke dalam beberapa kelompok berdasarkan jarak terdekat dengan pusat klaster (*centroid*) (James et al., 2021). Untuk menentukan jumlah kelompok (k) yang paling ideal, digunakan *Metode Elbow*. Jumlah kelompok terbaik dipilih dengan melihat titik sudut (siku) pada grafik penurunan *Within-Cluster Sum of Squares* (WCSS) yang mulai melandai (Bholowalia & Kumar, 2014).

5. Evaluasi Klaster dengan Silhouette Score

Langkah terakhir adalah mengevaluasi kualitas pengelompokan menggunakan nilai *Silhouette Score*. Metode ini menguji seberapa akurat suatu data masuk ke dalam klasternya sendiri dan seberapa jelas pemisahan antara satu kelompok dengan kelompok lainnya (Moudak, 2023). Nilai *Silhouette Score* berkisar antara -1 hingga +1; semakin mendekati nilai +1, maka pengelompokan yang dihasilkan dianggap semakin baik dan terpisah secara jelas.

HASIL DAN PEMBAHASAN

Tahap awal dalam menganalisis data adalah dengan persiapan dan analisis eksplorasi data (EDA), kemudian dilanjutkan dengan pemodelan dan evaluasi (*unsupervised learning*) dan tahap terakhir yaitu *deployment* serta implementasi.

Eksplorasi Data dan Pola Awal Pelanggan (EDA)

Hasil eksplorasi data pada Gambar 1 menunjukkan bahwa seluruh variabel dalam dataset tidak memiliki *missing value*, sehingga data dinyatakan layak untuk dianalisis lebih lanjut. Analisis

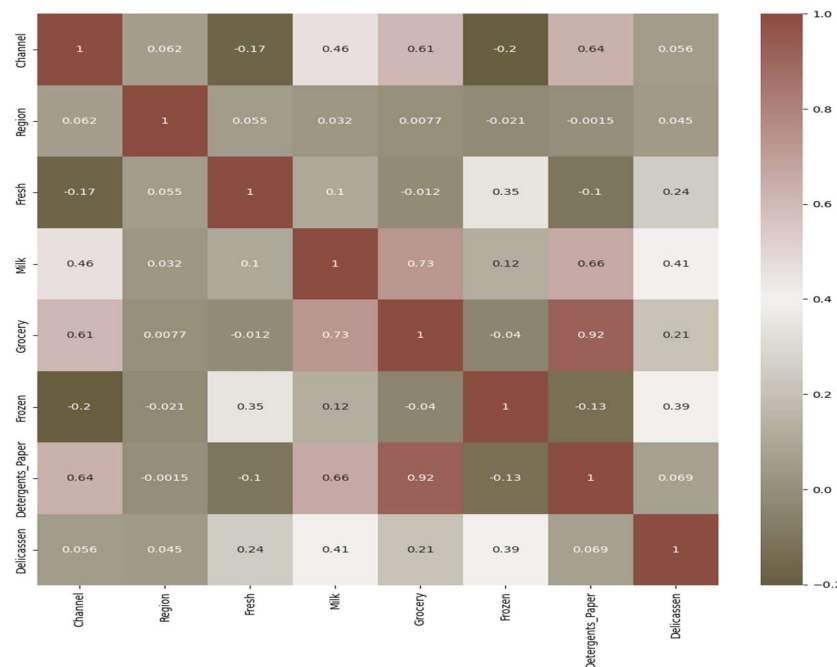
statistik deskriptif menunjukkan adanya variasi nilai yang cukup tinggi antar variabel, yang mengindikasikan heterogenitas pola belanja pelanggan.



Gambar 1. Explorasi *Data Analysis* (EDA)

Menurut Geron (2022), keberagaman pola data merupakan indikator penting dalam proses clustering karena semakin tinggi variasi data, maka semakin besar peluang terbentuknya kelompok pelanggan dengan karakteristik yang unik. Selain itu, hasil heatmap korelasi yang ditunjukkan pada Gambar 2 menunjukkan adanya hubungan antara beberapa kategori produk. Khususnya, pembelian produk segar yang memiliki hubungan dengan produk grocery, yang menunjukkan adanya pola pembelian komplementer. Hal ini mengindikasikan bahwa pelanggan cenderung membeli produk secara komplementer sesuai kebutuhan usaha atau konsumsi rumah tangga.

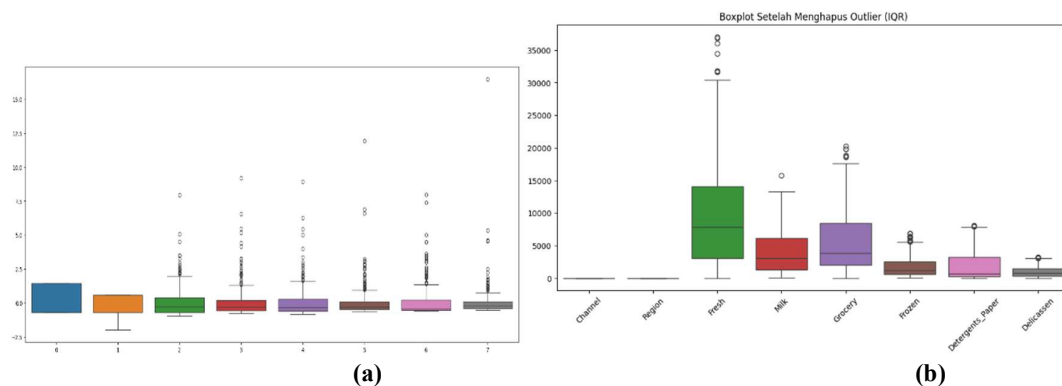
Temuan ini sejalan dengan konsep perilaku konsumen dalam pemasaran yang menyatakan bahwa konsumen cenderung membeli produk yang saling berkaitan (Kotler & Keller, 2016).



Gambar 2. Korelasi *Heatmap*

Penelitian Zheng (2023) juga menyatakan bahwa analisis *clustering* mampu mengidentifikasi keterkaitan pola pembelian konsumen yang sebelumnya sulit dikenali melalui analisis konvensional. Dalam konteks agribisnis, pola pembelian komplementer menunjukkan bahwa pelanggan tidak hanya membeli satu jenis produk, tetapi melakukan kombinasi pembelian sesuai kebutuhan operasional bisnis mereka. Kondisi ini penting bagi pelaku usaha agribisnis karena dapat dimanfaatkan untuk menyusun strategi cross-selling maupun *bundling* produk.

Selain itu, hasil boxplot menunjukkan adanya outlier pada beberapa variabel, yang kemudian ditangani menggunakan metode IQR. Penanganan ini penting untuk meningkatkan akurasi model clustering karena outlier dapat memengaruhi pembentukan centroid pada *K-Means Clustering*.



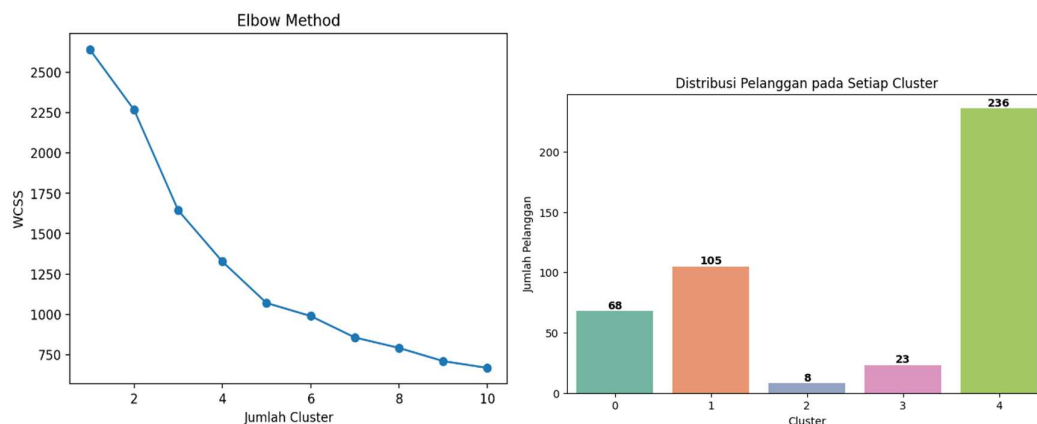
Gambar 3. Hasil Boxplot Sebelum *Outlier* Ditangani (a), dan Hasil Boxplot Setelah Menghapus *Outlier* (b)

Tahap *preprocessing* data menjadi langkah penting dalam penelitian ini, terutama pada proses penanganan outlier menggunakan metode *Interquartile Range* (IQR). Hasil *boxplot* menunjukkan adanya beberapa nilai ekstrem pada variabel pembelian pelanggan. Nilai ekstrem tersebut berpotensi memengaruhi pembentukan *centroid* pada algoritma K-Means sehingga hasil clustering menjadi kurang optimal. Menurut Li et al. (2023), algoritma K-Means memiliki sensitivitas tinggi terhadap outlier karena proses pembentukan *cluster* didasarkan pada perhitungan jarak antar data terhadap centroid. Oleh karena itu, penghapusan atau penyesuaian *outlier* diperlukan agar pusat *cluster* dapat merepresentasikan karakteristik kelompok secara lebih akurat.

Hasil Clustering

Setelah *preprocessing* dilakukan, data direduksi menggunakan Principal Component Analysis (PCA). Penggunaan PCA membantu menyederhanakan dimensi data tanpa menghilangkan informasi penting yang terkandung dalam dataset. Bruce et al. (2020) menjelaskan bahwa PCA efektif digunakan dalam analisis data berdimensi tinggi karena mampu mengurangi kompleksitas data dan meningkatkan efisiensi komputasi. Pada penelitian ini, PCA juga mempermudah visualisasi hasil *clustering* ke dalam bentuk dua dimensi sehingga distribusi pelanggan antar cluster dapat diamati dengan lebih jelas.

Penentuan jumlah *cluster* optimal berdasarkan hasil uji metode elbow yang ditunjukkan pada Gambar 4, diperoleh jumlah *cluster* optimal sebanyak lima kelompok pelanggan. Hasil ini menunjukkan adanya heterogenitas pola belanja dalam data pelanggan produk agribisnis.



Gambar 4. Hasil Uji Metode Elbow dan Distribusi Pelanggan pada Setiap *Cluster*

Pelanggan produk agribisnis dapat dibedakan berdasarkan pola belanja dan intensitas konsumsi produk. Menurut Bholowalia dan Kumar (2014), metode elbow merupakan pendekatan yang efektif dalam menentukan jumlah cluster optimal karena mampu menunjukkan titik penurunan signifikan pada nilai *Within Cluster Sum of Squares* (WCSS). Titik siku (*elbow point*) pada penelitian ini menunjukkan bahwa penambahan jumlah *cluster* di atas lima tidak lagi memberikan penurunan WCSS yang signifikan.

Karakteristik dan Interpretasi Cluster

Cluster 1: High Fresh Product Buyers

Cluster ini menunjukkan pelanggan dengan tingkat pembelian tinggi pada produk segar. Karakteristik ini mengindikasikan pelanggan yang berorientasi pada kualitas dan kemungkinan besar terdiri atas pelaku usaha kuliner, restoran, hotel, atau katering yang membutuhkan pasokan bahan baku segar secara rutin. Karakteristik pelanggan pada cluster ini menunjukkan orientasi terhadap kualitas produk dan kontinuitas distribusi. Penelitian Zhao (2024) menjelaskan bahwa pelanggan dengan tingkat konsumsi tinggi pada produk tertentu umumnya memiliki loyalitas yang lebih tinggi apabila kualitas dan ketersediaan produk dapat dipertahankan. Oleh karena itu, strategi yang dapat diterapkan pada *cluster* ini adalah menjaga kualitas produk, mempercepat distribusi, serta memberikan layanan prioritas.

Cluster 2: Low Value Customers

Kelompok ini memiliki tingkat pembelian rendah di semua kategori produk. Kelompok pelanggan ini cenderung memiliki sensitivitas tinggi terhadap harga dan tingkat loyalitas yang rendah. Kondisi tersebut menunjukkan bahwa pelanggan lebih mudah berpindah ke kompetitor apabila terdapat penawaran harga yang lebih murah. Menurut Nugroho et al. (2024), pelanggan dengan nilai transaksi rendah memerlukan pendekatan promosi yang lebih agresif, seperti diskon, *cashback*, maupun bundling produk untuk meningkatkan frekuensi pembelian. Dalam konteks agribisnis, strategi promosi yang tepat dapat membantu meningkatkan keterlibatan pelanggan dan mendorong peningkatan transaksi.

Cluster 3: Mixed Buyers

Cluster ini memiliki pola pembelian yang relatif seimbang pada berbagai kategori produk. Kelompok ini memiliki kebutuhan yang beragam sehingga berpotensi menjadi target utama dalam strategi pemasaran berbasis personalisasi. Pelanggan dalam *cluster* ini cenderung fleksibel dan terbuka terhadap berbagai jenis produk. Penelitian Awaliyah et al. (2024) menyebutkan bahwa pelanggan dengan pola pembelian campuran memiliki potensi tinggi untuk ditingkatkan nilai transaksinya melalui strategi *cross-selling* dan rekomendasi produk berbasis data. Oleh karena itu, pelaku usaha dapat memanfaatkan hasil segmentasi untuk menawarkan kombinasi produk yang relevan sesuai pola konsumsi pelanggan.

Cluster 4: Loyal High Volume Customers

Kelompok ini merupakan kelompok pelanggan dengan volume pembelian tertinggi dan konsisten. *Cluster* ini dapat dikategorikan sebagai *core customer* karena memberikan kontribusi besar terhadap pendapatan usaha. Menurut Kotler dan Keller (2016), mempertahankan pelanggan loyal jauh lebih menguntungkan dibandingkan memperoleh pelanggan baru karena biaya retensi lebih rendah dibandingkan biaya akuisisi pelanggan. Temuan ini menunjukkan bahwa pelaku usaha agribisnis perlu membangun hubungan jangka panjang dengan pelanggan utama melalui program loyalitas, pemberian reward, dan pelayanan khusus.

Cluster 5: Specialized Buyers

Cluster ini menunjukkan pelanggan yang fokus pada kategori produk tertentu. Pola ini mengindikasikan adanya spesialisasi kebutuhan sesuai jenis usaha pelanggan. Misalnya, pelanggan yang hanya membeli *frozen food* dalam jumlah besar kemungkinan merupakan reseller atau usaha makanan tertentu. Zheng (2023) menjelaskan bahwa segmentasi pelanggan spesifik memungkinkan perusahaan menyusun strategi diferensiasi produk dan layanan secara lebih efektif. Dalam konteks agribisnis, strategi yang dapat diterapkan adalah menyediakan produk yang lebih spesifik, menjaga konsistensi kualitas, serta memberikan layanan sesuai kebutuhan pelanggan.

Evaluasi Model

Hasil evaluasi menggunakan *silhouette score* menunjukkan bahwa model memiliki kualitas pemisahan *cluster* yang cukup baik. Hal ini menandakan bahwa metode K-Means Clustering efektif dalam mengidentifikasi pola belanja pelanggan. Nilai *silhouette score* yang mendekati +1 menunjukkan bahwa objek dalam satu *cluster* memiliki tingkat kemiripan yang tinggi dan memiliki perbedaan yang jelas dengan *cluster* lainnya. Hal ini menunjukkan bahwa metode *K-Means Clustering* efektif digunakan dalam segmentasi pelanggan agribisnis. Temuan ini sejalan dengan penelitian Li *et al.* (2023) dan Zhao (2024) yang menyatakan bahwa K-Means merupakan algoritma *clustering* yang mampu menghasilkan segmentasi pelanggan secara jelas dan mudah diinterpretasikan.

Pembahasan

Hasil penelitian menunjukkan bahwa pendekatan berbasis *Machine Learning* mampu memberikan *insight* yang lebih mendalam dibandingkan metode konvensional. Segmentasi yang dihasilkan dapat digunakan untuk meningkatkan efisiensi pemasaran dan distribusi dalam sektor agribisnis.

Secara praktis, hasil penelitian ini memberikan implikasi penting bagi pelaku usaha agribisnis, khususnya UMKM. Segmentasi pelanggan berbasis data memungkinkan pelaku usaha mengambil keputusan pemasaran secara lebih objektif dan efisien. Fauziah *et al.* (2020) menjelaskan bahwa peningkatan produktivitas UMKM sangat dipengaruhi oleh kemampuan usaha dalam mengelola sumber daya dan strategi pemasaran secara tepat. Dengan memahami karakteristik pelanggan, pelaku usaha dapat menentukan target pasar yang lebih spesifik, mengurangi biaya pemasaran yang tidak efektif, serta meningkatkan kepuasan pelanggan.

Selain itu, penelitian ini menunjukkan bahwa pemanfaatan *Machine Learning* dalam sektor agribisnis memiliki potensi besar untuk mendukung transformasi digital. Selama ini, penerapan teknologi data science lebih banyak digunakan pada sektor *e-commerce* dan retail modern, sedangkan sektor agribisnis masih relatif terbatas. Padahal, sektor agribisnis memiliki volume data pelanggan yang cukup besar dan beragam sehingga sangat potensial untuk dianalisis menggunakan pendekatan berbasis data. Oleh karena itu, integrasi teknologi *clustering* dan *dashboard* interaktif dapat membantu pelaku usaha dalam memantau perilaku pelanggan secara *real-time* dan mendukung pengambilan keputusan berbasis data.

Meskipun penelitian ini memberikan hasil yang cukup baik, terdapat beberapa keterbatasan yang perlu diperhatikan. Pertama, dataset yang digunakan masih berasal dari data sekunder sehingga belum sepenuhnya merepresentasikan kondisi pelanggan agribisnis di Indonesia. Kedua, penelitian ini hanya menggunakan algoritma *K-Means Clustering* sehingga hasil segmentasi mungkin berbeda apabila dibandingkan dengan algoritma lain seperti DBSCAN atau *Hierarchical Clustering*. Ketiga, variabel yang digunakan masih terbatas pada pola pembelian produk tanpa mempertimbangkan faktor demografis maupun perilaku digital pelanggan.

Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan data primer dari sektor agribisnis lokal agar hasil segmentasi lebih relevan terhadap kondisi nyata di lapangan. Selain itu, penelitian lanjutan dapat mengkombinasikan metode *clustering* dengan analisis prediktif atau sistem rekomendasi berbasis *Artificial Intelligence* sehingga hasil analisis dapat digunakan secara lebih komprehensif dalam pengembangan strategi bisnis agribisnis.

KESIMPULAN

Penelitian ini dilakukan untuk mengatasi permasalahan heterogenitas perilaku pelanggan pada sektor agribisnis yang menyebabkan strategi pemasaran kurang optimal. Tujuan penelitian adalah mengidentifikasi pola belanja pelanggan, menentukan jumlah *cluster* optimal, serta menganalisis karakteristik setiap segmen pelanggan menggunakan metode *K-Means Clustering*.

Hasil penelitian menunjukkan bahwa terbentuk lima *cluster* pelanggan dengan karakteristik yang berbeda, yaitu pelanggan produk segar, pelanggan bernilai rendah, pelanggan dengan pembelian campuran, pelanggan loyal dengan volume tinggi, dan pelanggan spesifik pada produk tertentu. Metode *K-Means Clustering* terbukti efektif dalam mengelompokkan pelanggan berdasarkan pola pembelian sehingga dapat membantu pelaku usaha agribisnis menyusun strategi pemasaran yang lebih tepat, efisien, dan berbasis data.

DAFTAR PUSTAKA

- Awaliyah, D. A., Prasetyo, B., Muzayanah, R., & Lestari, A. D. (2024). Optimizing customer segmentation in online retail transactions through the implementation of K-Means clustering algorithm. *Scientific Journal of Informatics*, 11(2). <https://doi.org/10.15294/sji.v11i2.6137>
- Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications*, 105 (9).
- Bruce, P., Bruce, A., & Gedeck, I. (2020). *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python (2nd ed.)*. O'Reilly Media Inc.
- Fauziah, S., Rifin, A., & Adhi, A. K. (2020). Produktivitas usaha mikro dan kecil industri makanan di Indonesia. *Agriekonomika*, 9(2), 157–170. <https://doi.org/10.21107/agriekonomika.v9i2.8232>
- Geron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.)*. O'Reilly Media Inc.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Indriyani, F., & Irfiani, E. (2019). Pengaruh penanganan outlier terhadap kinerja algoritma K-Means dalam pengelompokan data. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 3(2), 211-217. <https://doi.org/10.29207/resti.v3i2.985>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R/Python*. Springer Science and Business Media. New York
- Moudak, S. (2023). Finding Groups in Data: An Introduction to Cluster Analysis. *Journal of Applied Statistic*. 4;51(8):1618–1620. <https://doi.org/10.1080/02664763.2023.2220087>
- Kotler, P., & Keller, K. L. (2016). *Marketing management* (15th ed.). Pearson Education.
- Li, Y., Qi, J., & Mu, W. (2023). Customer segmentation using K-Means clustering and hybrid particle swarm optimization algorithm. *The Computer Journal*, 66(4), 941–962. <https://doi.org/10.1093/comjnl/bxab206>
- Nugroho, B. I., Rafhina, A., Ananda, P. S., & Gunawan. (2024). Customer segmentation in sales transaction data using K-Means clustering algorithm. *Journal of Intelligent Decision Support System*, 7(2), 130–136. <https://doi.org/10.35335/idss.v7i2.236>
- Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021>
- Zhao, J. (2024). Customer segmentation application based on K-Means. *Applied and Computational Engineering*, 47(1), 242–247.
- Zheng, Y. (2023). Customer segmentation research in marketing through clustering algorithm analysis. *Journal of Intelligent & Fuzzy Systems*, 45(4), 1–15. <https://doi.org/10.3233/JIFS-232589>