

Penerapan K-Means Clustering Untuk Segmentasi Penjualan Di Minimarket Mardi Dengan Pendekatan Machine Learning

Imam Frandika^{1*}, Sofana Bayor Hud², Wiwin Handoko³

^{1,2,3}Sistem Informasi, Universitas Royal

^{1*}prandikaimam@gmail.com, ²Sofanabayorhud123@gmail.com, ³win.van.handoko@gmail.com

Abstrak

Peningkatan efektivitas strategi pemasaran menjadi kebutuhan mendesak di era kompetitif seperti sekarang. Penelitian ini bertujuan untuk membantu Minimarket Mardi dalam memahami pola pembelian pelanggan melalui penerapan algoritma K-Means untuk segmentasi penjualan. Dengan metode ini, data transaksi dan penjualan diolah untuk menghasilkan kluster pelanggan berdasarkan pola perilaku pembelian. Klusterisasi dilakukan menggunakan perangkat lunak anaconda dengan bahasa pemrograman Python, memanfaatkan pustaka seperti Pandas, NumPy, dan Scikit-learn. Evaluasi kualitas kluster dilakukan dengan Davies-Boulding Index (DBI). Hasil penelitian menunjukkan pembagian pelanggan ke dalam 3 kluster utama, mencakup kategori rendah, sedang, dan tinggi. Temuan ini menawarkan peluang bagi Minimarket Mardi untuk meningkatkan kinerja bisnis secara signifikan.

Kata kunci : K-Means Clustering, segmentasi pelanggan, Python, data mining, machine learning.

Abstract

Increasing the effectiveness of marketing strategies is an urgent need in the current competitive era. This research aims to help Minimarket Mardi understand customer purchasing patterns through the application of the K-Means algorithm for sales segmentation. With this method, transaction and sales data are processed to produce customer clusters based on purchasing behavior patterns. Clustering was carried out using Anaconda software with the Python programming language, utilizing libraries such as Pandas, NumPy, and Scikit-learn. Evaluation of cluster quality is carried out using the Davies-Boulding Index (DBI). The research results show the division of customers into 3 main clusters, including low, medium and high categories. These findings offer opportunities for Mardi Minimarket to improve business performance significantly.

Keywords: K-Means Clustering, customer segmentation, Python, data mining, machine learning.

PENDAHULUAN

Dalam era globalisasi dan digitalisasi yang semakin pesat, persaingan di sektor ritel, khususnya minimarket, menjadi semakin kompetitif. Untuk mempertahankan dan meningkatkan pangsa pasar, pemahaman yang mendalam mengenai perilaku dan preferensi pelanggan menjadi sangat krusial. Salah satu pendekatan yang efektif untuk mencapai pemahaman tersebut adalah melalui analisis data transaksi pelanggan menggunakan teknik *data mining*. Teknik ini memungkinkan perusahaan untuk menggali informasi berharga dari data yang dimiliki, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan strategis[1][2].

Salah satu metode *data mining* yang banyak digunakan untuk segmentasi pelanggan adalah algoritma *K-Means Clustering*. Algoritma ini bekerja dengan mengelompokkan data ke dalam

beberapa kluster berdasarkan kesamaan karakteristik tertentu. Dalam konteks ritel, penerapan *K-Means Clustering* dapat membantu mengidentifikasi kelompok pelanggan dengan pola pembelian yang serupa, sehingga perusahaan dapat merancang strategi pemasaran yang lebih tepat sasaran. *K-Means* efektif dalam mengelompokkan pelanggan berdasarkan pola pembelian, yang pada gilirannya dapat meningkatkan efektivitas promosi dan kepuasan pelanggan[3][4][5].

Segmentasi pelanggan menggunakan metode *K-Means Clustering* dapat membantu perusahaan ritel dalam memahami karakteristik masing-masing segmen pelanggan. Dengan demikian, perusahaan dapat menyusun program loyalitas dan promosi yang lebih sesuai dengan kebutuhan dan preferensi setiap segmen, yang pada akhirnya dapat meningkatkan loyalitas pelanggan dan profitabilitas perusahaan[6].

Penelitian ini bertujuan untuk mengimplementasikan algoritma *K-Means Clustering* dalam menganalisis pola transaksi di Minimarket Mardi. Dengan menggunakan data penjualan historis, pelanggan akan dikelompokkan ke dalam beberapa kluster yang memiliki karakteristik serupa. Hasil dari segmentasi ini diharapkan dapat memberikan wawasan bagi manajemen dalam merancang strategi pemasaran yang lebih terarah dan sesuai dengan kebutuhan masing-masing segmen pelanggan. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan metode analisis data yang efektif untuk sektor ritel, khususnya dalam konteks pasar Indonesia[7][8][9].

METODE

1. Pengumpulan Data

Langkah pertama dalam penelitian ini adalah mengumpulkan data transaksi pelanggan dari Minimarket Mardi selama periode satu tahun. Data yang dikumpulkan mencakup informasi seperti jumlah transaksi (frekuensi), total penjualan. Pengumpulan data yang komprehensif ini bertujuan untuk memastikan bahwa analisis yang dilakukan mencerminkan perilaku pembelian pelanggan secara akurat. Data transaksi penjualan yang lengkap dan terstruktur sangat penting dalam analisis pola penjualan di minimarket. Berikut adalah data penjualan yang didapatkan dari pemilik minimarket Mardi:

no	barang	jenis	harga	total_transaksi	total_penjualan
0 1	Chitato	Snack	7000	680	720
1 2	Qtella	Snack	7000	595	660
2 3	Siip	Snack	2000	870	1242
3 4	Garuda Kacang	Snack	500	1800	1800
4 5	Mogidon	Snack	22000	115	120
...
105 106	Nugget Ayam	Produk Makanan	50000	43	50
106 107	Bakso Ikan (Frozen)	Produk Makanan	40000	393	400
107 108	Kopi gula aren	Minuman	1000	2323	2589
108 109	Kerupuk Udang	Camilan	12000	778	890
109 110	Sambal Sachet	Bumbu Dapur	1000	2390	2432

110 rows x 6 columns

Gambar 1 : Tabel dataset penjualan Minimarket mardi

Selain itu, pengumpulan data yang tepat memungkinkan identifikasi produk dengan penjualan tertinggi, sehingga manajemen stok dapat dioptimalkan untuk meningkatkan penjualan. Analisis data transaksi juga membantu dalam memahami preferensi pelanggan, yang dapat digunakan untuk

merancang strategi pemasaran yang lebih efektif. Dengan demikian, pengumpulan data yang akurat dan lengkap menjadi dasar penting dalam penelitian ini.

2. Pra-pemrosesan Data

Setelah data terkumpul, langkah selanjutnya adalah pra-pemrosesan data, yang meliputi pembersihan data dan normalisasi. Pembersihan data bertujuan untuk menghapus nilai yang hilang (*missing values*) dan duplikasi yang dapat mempengaruhi hasil analisis. Proses ini memastikan bahwa data yang digunakan dalam analisis adalah valid dan reliabel. Pra-pemrosesan data yang tepat sangat penting untuk meningkatkan akurasi model klasifikasi[10][11].

```
]: df.isnull().sum()
]: no          0
   barang      0
   jenis       0
   harga       0
   total_transaksi  0
   total_penjualan  0
dtype: int64
```

Gambar 2: Kode untuk menghilangkan missing values

3. Penentuan Jumlah Kluster Optimal

Penentuan jumlah kluster optimal merupakan langkah krusial dalam penerapan algoritma *K-Means Clustering*. Metode *Elbow* digunakan untuk menentukan jumlah kluster yang tepat dengan menganalisis grafik *Within-Cluster Sum of Squares* (WCSS). Titik di mana penurunan WCSS mulai melambat secara signifikan dianggap sebagai jumlah kluster optimal. Metode *Elbow* efektif dalam menentukan jumlah kluster yang optimal dalam analisis pola penjualan di minimarket[12].

Dengan menentukan jumlah kluster yang optimal, analisis dapat menghasilkan segmentasi pelanggan yang lebih akurat, sehingga strategi pemasaran dapat disesuaikan dengan karakteristik masing-masing segmen. Hal ini penting untuk meningkatkan efektivitas promosi dan kepuasan pelanggan.

Jumlah kluster optimal ditentukan menggunakan **Metode Elbow**, yang mengacu pada grafik *Within-Cluster Sum of Squares* (WCSS). WCSS didefinisikan sebagai total jarak kuadrat antara setiap titik data ke centroid kluster yang sesuai, dengan rumus:

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

Di mana:

- K = jumlah kluster
- C_i = kluster ke- i
- μ_i = centroid kluster C_i
- x = data dalam kluster C_i

Jumlah kluster optimal adalah titik di mana penurunan WCSS mulai melambat secara signifikan, membentuk "siku" pada grafik. Metode ini efektif untuk mengidentifikasi jumlah kluster optimal dalam analisis pola pembelian pelanggan.

4. Implementasi Algoritma K-Means

Setelah jumlah kluster optimal ditentukan, algoritma *K-Means Clustering* diterapkan pada data yang telah dipra-pemroses. Algoritma ini bekerja dengan mengelompokkan data ke dalam kluster berdasarkan kesamaan karakteristik. Implementasi dilakukan menggunakan pustaka Python seperti Scikit-learn, yang menyediakan fungsi-fungsi untuk melakukan klusterisasi dengan efisien.

Algoritma *K-Means* efektif dalam mengelompokkan data penjualan di minimarket, sehingga dapat membantu dalam pengambilan keputusan strategis[13][14].

Penerapan algoritma ini memungkinkan identifikasi kelompok pelanggan dengan perilaku pembelian yang serupa, sehingga manajemen dapat merancang strategi pemasaran yang lebih tepat sasaran. Selain itu, segmentasi pelanggan yang dihasilkan dapat digunakan untuk mengoptimalkan penataan produk dan layanan yang ditawarkan.

Algoritma *K-Means Clustering* diterapkan untuk mengelompokkan data pelanggan ke dalam beberapa kluster. Proses algoritma ini melibatkan langkah-langkah berikut:

1. Menentukan jumlah kluster (kkk) berdasarkan hasil metode Elbow.
2. Memilih secara acak kkk centroid awal.
3. Menghitung jarak setiap data ke centroid menggunakan rumus **Euclidean Distance**:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Di mana:

- xxx dan yyy adalah dua data
 - nnn adalah jumlah fitur
4. Menetapkan data ke centroid terdekat.
 5. Memperbarui posisi centroid berdasarkan rata-rata data dalam kluster:

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$$

6. Mengulangi langkah 3–5 hingga centroid tidak lagi berubah atau perubahan menjadi sangat kecil.

Implementasi algoritma ini dilakukan menggunakan pustaka Python seperti Scikit-learn. Hasil klusterisasi memberikan segmentasi pelanggan berdasarkan perilaku belanja mereka.

5. Evaluasi Kualitas Kluster

Evaluasi kualitas kluster dilakukan untuk memastikan bahwa hasil klusterisasi memiliki kekompakan dan pemisahan yang baik antar kluster. Salah satu metrik yang digunakan adalah *Davies-Bouldin Index* (DBI), yang mengukur rata-rata kesamaan antara setiap kluster dengan kluster yang paling mirip. Nilai DBI yang lebih rendah menunjukkan kluster yang lebih baik, dengan nilai ideal mendekati nol. Evaluasi menggunakan DBI penting untuk memastikan kualitas kluster yang dihasilkan dalam analisis data.

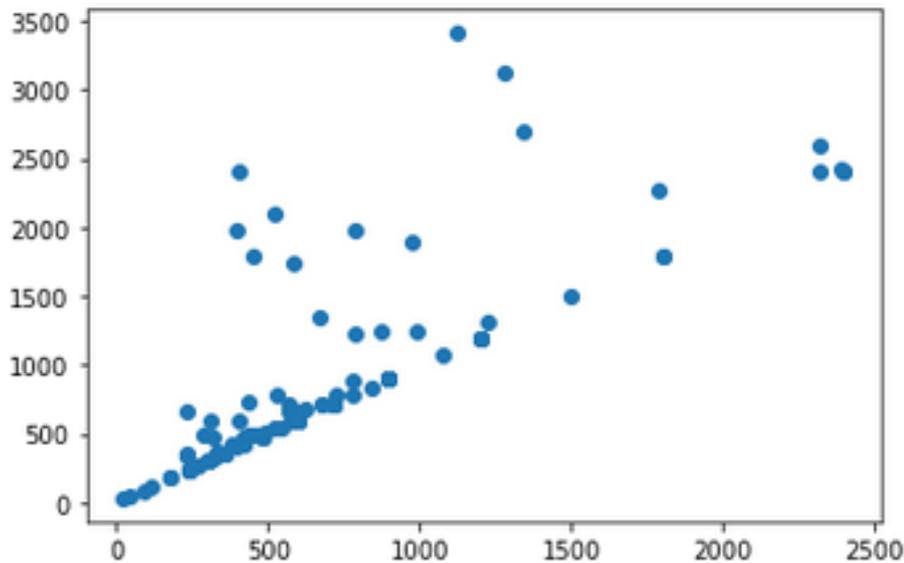
Dengan evaluasi yang tepat, hasil klusterisasi dapat diandalkan untuk digunakan dalam perencanaan strategi bisnis. Evaluasi juga membantu dalam mengidentifikasi area yang memerlukan perbaikan dalam proses klusterisasi, sehingga dapat dilakukan penyesuaian yang diperlukan untuk meningkatkan akurasi dan efektivitas analisis.

HASIL DAN PEMBAHASAN

1. Penentuan Kluster Optimal

Dalam penelitian ini, metode **Elbow** digunakan untuk menentukan jumlah kluster optimal. Metode Elbow memanfaatkan grafik hubungan antara jumlah kluster dan nilai *Within-Cluster Sum of Squares* (WCSS). Grafik tersebut menunjukkan adanya titik "siku" atau penurunan drastis pada nilai WCSS ketika jumlah kluster adalah tiga. Oleh karena itu, dipilih tiga kluster sebagai jumlah kluster optimal. Pemilihan ini didasarkan pada keseimbangan antara kekompakan dalam setiap kluster (*intra-cluster compactness*) dan perbedaan yang jelas antar kluster (*inter-cluster separation*).

Jumlah kluster ini memberikan struktur yang logis dan relevan untuk memahami pola pelanggan yang ada.



Gambar 3 : Visualisasi data yang membentuk siku

```
from sklearn.cluster import KMeans  
  
km = KMeans(n_clusters=3, random_state=0)  
km
```

KMeans|

```
KMeans(n_clusters=3, random_state=0)
```

Grafik tersebut menunjukkan adanya titik "siku" atau penurunan drastis pada nilai WCSS ketika jumlah kluster adalah tiga. Oleh karena itu, dipilih tiga kluster sebagai jumlah kluster optimal. Pemilihan ini didasarkan pada keseimbangan antara kekompakan dalam setiap kluster (*intra-cluster compactness*) dan perbedaan yang jelas antar kluster (*inter-cluster separation*). Jumlah kluster ini memberikan struktur yang logis dan relevan untuk memahami pola pelanggan yang ada.

2. Membuat Kluster

$y_{prediksi}$ adalah hasil klasterisasi yang mengelompokkan data Anda ke dalam kluster tertentu. Nilai ini penting untuk memahami pola dalam data, memvisualisasikan hasil klasterisasi, dan membuat strategi berbasis kluster.

```
]: #membuat dataframe hasil klaster agar bisa dilihat pada dataset
df['cluster'] = y_prediksi
df
```

	no	barang	jenis	harga	total_transaksi	total_penjualan	cluster
0	1	Chitato	Snack	7000	680	720	0
1	2	Qtella	Snack	7000	595	660	0
2	3	Siip	Snack	2000	870	1242	2
3	4	Garuda Kacang	Snack	500	1800	1800	1
4	5	Mogidon	Snack	22000	115	120	0
...
105	106	Nugget Ayam	Produk Makanan	50000	43	50	0
106	107	Bakso Ikan (Frozen)	Produk Makanan	40000	393	400	0
107	108	Kopi gula aren	Minuman	1000	2323	2589	1
108	109	Kerupuk Udang	Camilan	12000	778	890	2
109	110	Sambal Sachet	Bumbu Dapur	1000	2390	2432	1

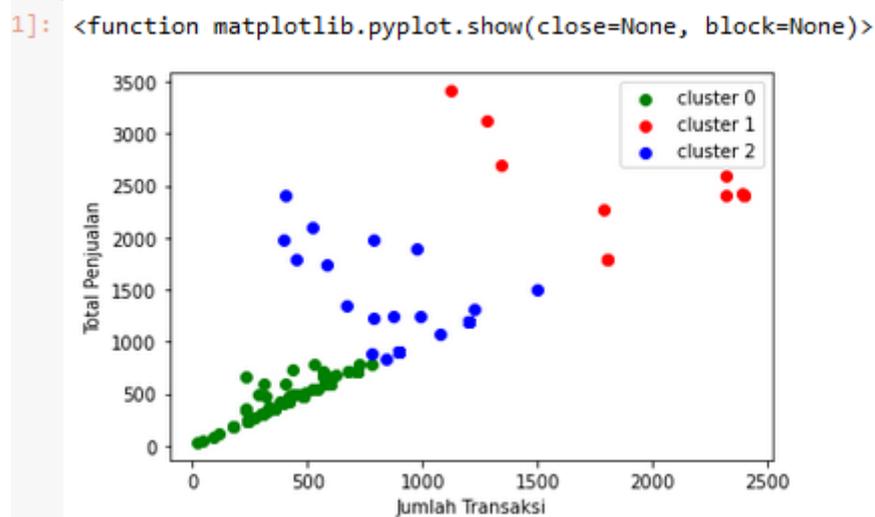
110 rows × 7 columns

Gambar 4 : Hasil Klaster

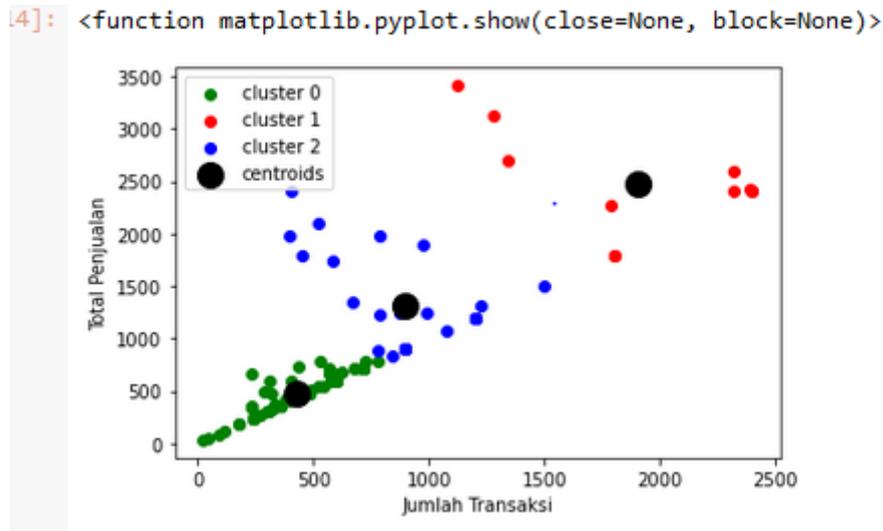
3. Visualisasi Hasil Klaster

Visualisasi hasil klaster adalah cara untuk merepresentasikan data yang telah dikelompokkan melalui algoritma klasterisasi, seperti **K-Means**, dalam bentuk grafis yang mudah dipahami. Setiap klaster dalam visualisasi biasanya diwakili dengan warna, simbol, atau bentuk yang berbeda untuk membedakan satu kelompok data dari kelompok lainnya. Tujuan utama dari visualisasi ini adalah untuk membantu memahami pola, distribusi, dan hubungan antar kelompok dalam data[15]¹.

Ketika data telah diproses menggunakan metode klasterisasi, hasilnya berupa pengelompokan berdasarkan kesamaan karakteristik tertentu. Namun, hasil ini sering kali berupa angka atau label klaster yang sulit dipahami secara intuitif. Di sinilah visualisasi menjadi penting, karena dapat mengubah data numerik atau kategorikal ini menjadi bentuk yang lebih visual, sehingga lebih mudah diinterpretasikan.



Gambar 5 : Visualisasi Hasil Klaster



Gambar 6 : Visualisasi Hasil Klaster & Centroid

4. Transformasi Hasil Klasterisasi ke dalam Label Deskriptif untuk Analisis Lebih Lanjut

Proses klasterisasi adalah teknik yang umum digunakan dalam analisis data untuk mengelompokkan objek atau entitas yang memiliki kemiripan dalam satu grup atau klaster. Hasil klasterisasi biasanya berupa angka atau kode yang mewakili setiap klaster, seperti 0, 1, atau 2, yang dihasilkan oleh algoritma seperti K-Means. Namun, hasil ini sering kali sulit dipahami oleh orang yang tidak terbiasa dengan analisis data, terutama bagi pemangku kepentingan atau tim bisnis yang lebih memerlukan informasi yang mudah dimengerti.

Untuk mempermudah pemahaman dan analisis lebih lanjut, hasil klasterisasi dapat ditransformasikan menjadi label deskriptif yang lebih intuitif. Misalnya, klaster dengan angka 0 bisa diberi label "*Penjualan Rendah*", klaster dengan angka 1 bisa diberi label "*Penjualan Sedang*", dan klaster dengan angka 2 diberi label "*Penjualan Tinggi*". Langkah ini dilakukan dengan menggunakan fungsi dalam Python, yang memungkinkan untuk memeriksa setiap nilai klaster dan menggantinya dengan label yang sesuai. Seperti yang pada gambar berikut ini :

```
: conditions = [  
    (df['cluster']== 0),  
    (df['cluster']== 1),  
    (df['cluster']== 2)]  
choices = ['Penjualan Rendah','Penjualan Tinggi','Penjualan Sedang']  
df['cluster'] = np.select(conditions, choices)  
df
```

Gambar 7 : Tranformasi klaster

	kode_barang	nama_barang	jumlah_transaksi	total_penjualan	rata_rata	cluster
0	2 TANG BLACK TEA 1 RENCENG (ISI 10)	2 TANG BLACK TEA 1 RENCENG (ISI 10)	1	1	1.0000	Penjualan Rendah
1	2 TANG MELATI 1 RENCENG (ISI 10)	2 TANG MELATI 1 RENCENG (ISI 10)	1	1	1.0000	Penjualan Rendah
2	AQUA 1500 ML 1 DUS	AQUA 1500 ML 1 DUS	10	11	1.1000	Penjualan Rendah
3	RIBUT KILOAN	RIBUT KILOAN	77	83	1.0779	Penjualan Rendah
4	7916248823	MINYAK TAWON FF	8	9	1.1250	Penjualan Rendah
...
7398	YOU YOU STICK HUAU 1 PACK	YOU YOU STICK HUAU 1 PACK	28	30	1.0714	Penjualan Rendah
7399	YOU YOU STICK MERAH 1 PACK	YOU YOU STICK MERAH 1 PACK	27	28	1.0370	Penjualan Rendah
7400	YULIM SNACK ALL VARIANT	YULIM SNACK ALL VARIANT	3	4	1.3333	Penjualan Rendah
7401	YUPI KISS STRAWBERRY 2PCS	YUPI KISS STRAWBERRY 2PCS	45	91	2.0222	Penjualan Rendah
7402	ZEGAR 1 DUS	ZEGAR 1 DUS	84	92	1.0952	Penjualan Rendah

7373 rows × 6 columns

Gambar 8 : Klaster yang sudah diubah

Transformasi ini membawa sejumlah manfaat, antara lain:

1. **Mempermudah Pemahaman:** Label deskriptif memudahkan tim bisnis untuk memahami hasil klasterisasi tanpa harus memahami angka atau kode yang digunakan oleh algoritma.
2. **Mendukung Pengambilan Keputusan:** Dengan label yang jelas, perusahaan dapat lebih mudah menentukan langkah-langkah strategis berdasarkan segmentasi pelanggan, seperti memberikan diskon khusus untuk pelanggan dalam klaster "*Penjualan Rendah*" atau program loyalitas untuk pelanggan dalam klaster "*Penjualan Tinggi*"[16].
3. **Meningkatkan Komunikasi:** Dengan label deskriptif, hasil klasterisasi dapat lebih mudah disampaikan kepada pemangku kepentingan, baik dalam laporan maupun presentasi.

KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma K-Means Clustering untuk segmentasi penjualan di Minimarket Mardi, dengan tujuan untuk meningkatkan efektivitas strategi pemasaran melalui pemahaman pola pembelian pelanggan. Berdasarkan data transaksi pelanggan, penelitian ini mengelompokkan pelanggan ke dalam tiga klaster utama: rendah, sedang, dan tinggi, menggunakan perangkat lunak Anaconda dan bahasa pemrograman Python. Evaluasi kualitas klaster dilakukan dengan Davies-Bouldin Index (DBI) untuk memastikan kualitas klaster yang dihasilkan. Hasil penelitian ini memberikan wawasan yang berharga bagi Minimarket Mardi dalam merancang strategi pemasaran yang lebih terarah, seperti pengoptimalan promosi dan program loyalitas berdasarkan karakteristik setiap segmen pelanggan. Penerapan metode ini diharapkan dapat meningkatkan kinerja bisnis dan mengoptimalkan keputusan strategis dalam pemasaran.

DAFTAR PUSTAKA

- [1] C. H. Ardana, A. A. A. A. Khoyum, and M. Faisal, "Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 1, pp. 1–9, 2024, doi: 10.14421/jiska.2024.9.1.1-9.
- [2] N. Firdaus, S. Defit, and G. Widi Nurcahyo, "Penerapan Algoritma K-Means Clustering Dalam Pengelompokan Data Penjualan Pada Pabrik Mobil Toyota Indonesia," *Fountain Informatics J.*, vol. 8, no. 2, pp. 2548–5113, 2023, [Online]. Available: <http://dx.doi.org/10.21111/fij.v8i2.10798>
- [3] U. Semarang and U. Clustering, "Implementasi Algoritma K-means untuk Segmentasi Pelanggan pada Mall Menggunakan Clustering," pp. 1–12.
- [4] W. T. Pambudi and A. Witanti, "Penerapan Algoritma K-Means Clustering Untuk Menganalisis Penjualan Pada Toko Ayu Collection Barbasis Web," *J. Inform. Univ. Pamulang*, vol. 6, no. 3, pp. 645–650, 2021, [Online]. Available:

<https://www.neliti.com/publications/466160/penerapan-algoritma-k-means-clustering-untuk-menganalisis-penjualan-pada-toko-ay>

- [5] S. F. Djun, I. G. A. Gunadi, and S. Sariyasa, “Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means Clustering pada Model Data RFM,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 5, no. 4, pp. 354–364, 2024, doi: 10.35746/jtim.v5i4.434.
- [6] P. H. Suharti, A. S. Suryandari, and R. N. Amalia, “ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN K-MEANS CLUSTERING STUDI KASUS APLIKASI ALFAGIFT,” *Sebatik*, vol. 26, no. 2, pp. 420–427, 2022, doi: 10.46984/sebatik.v26i2.2134.
- [7] T. Wahyudi and T. Silfia, “Implementation of Data Mining Using K-Means Clustering Method To Determine Sales Strategy in S&R Baby Store,” *J. Appl. Eng. Technol. Sci.*, vol. 4, no. 1, pp. 93–103, 2022, doi: 10.37385/jaets.v4i1.913.
- [8] C. Udokwu, P. Brandtner, F. Darbanian, and T. Falatouri, “Improving Sales Prediction for Point-of-Sale Retail Using Machine Learning and Clustering,” pp. 55–73, 2022, doi: 10.1007/978-3-031-18483-3_4.
- [9] M. Di, T. Erremy, and I. Nawangsih, “Analisa Penjualan Produk Kosmetik Dengan Metode Algoritma K- Means Di Toko Erremy,” vol. 4, no. 1, pp. 140–145, 2023.
- [10] L. H. Annisa and D. Rusvinasari, “Segmentasi Pembelian Produk Menggunakan Algoritma K-Means Berdasarkan Clusterisasi pada pemilihan menu yang ada diUMKM Kuliner,” vol. 9, no. 3, pp. 203–212, 2024, doi: 10.30591/jpit.v9i3.6556.
- [11] Z. G. Wei *et al.*, “Comparison of Methods for Biological Sequence Clustering,” *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 20, no. 5, pp. 2874–2888, 2023, doi: 10.1109/TCBB.2023.3253138.
- [12] S. P. Putra, A. Hoar, L. T. Rahmawati, S. P. Ersida, and E. Auliana, “Jurnal Hilirisasi Penelitian Masyarakat Analisis Strategi Pemasaran Dengan Sistem Online Melalui Aplikasi Whatsapp Pada Penjualan Produk Jeans Drink JURNAL JIHAPENMAS Jurnal Hilirisasi Penelitian Masyarakat,” vol. 1, no. 1, pp. 32–39, 2024.
- [13] A. Saputra and R. Yusuf, “Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1346–1361, 2024, doi: 10.57152/malcom.v4i4.1516.
- [14] W. Lu and D. Ren, “Commodity Sales Forecast Based on Cluster Analysis and Time Series,” 2024, doi: 10.4108/eai.27-10-2023.2341974.
- [15] J. Duo, P. Zhang, and L. Hao, “A K-means Text Clustering Algorithm Based on Subject Feature Vector,” *J. Web Eng.*, vol. 20, no. 6, pp. 1935–1946, 2021, doi: 10.13052/jwe1540-9589.20612.
- [16] F. A. Dewa and M. T. Jatipaningrum, “SEGMENTASI E-COMMERCE DENGAN CLUSTER K-MEANS DAN FUZZY C-MEANS (Studi Kasus: Media Sosial di Indonesia yang diunduh di Play Store),” *J. Stat. Ind. dan Komputasi*, vol. 4, no. 1, pp. 53–67, 2019.