

PENERAPAN ALGORITMA CLUSTERING UNTUK SEGMENTASI PELANGGAN BERDASARKAN PERILAKU PEMBELIAN

Raflesia Mukhlis Juliyanto¹, Fito Patria², Havid Danggo Pamungkas³
Febrianta Surya Nugraha⁴

¹²³⁴Prodi S1 Informatika STMIK Amikom Surakarta

¹²³⁴Sukoharjo Indonesia

Email: 1raflesia.10428@mhs.amikomsolo.ac.id,
2fito.10427@mhs.amikomsolo.ac.id, 3havid.10424@mhs.amikomsolo.ac.id,
ubingg@gmail.com

Abstract

This study explores the application of the K-Means clustering algorithm for customer segmentation based on purchasing behavior. Using a dataset from Kaggle containing consumer shopping trends, the research focuses on seven key attributes: Age, Gender, Purchase Amount, Category, Purchase Frequency, Previous Purchases, and Preferred Payment Method. The K-Means algorithm successfully identified six distinct customer clusters, each with unique characteristics reflecting specific purchasing patterns. The analysis provides valuable insights into product preferences, buying habits, and demographic characteristics of customers. These findings have significant implications for marketing strategies and business decision-making, enabling companies to develop more targeted and personalized approaches to customer engagement and retention. The study demonstrates the effectiveness of the K-Means algorithm in customer segmentation and highlights its potential for optimizing marketing efforts and improving customer satisfaction in various business contexts.

Keywords: K-Means algorithm, customer segmentation, purchasing behavior, clustering, data mining

Abstraksi

Penelitian ini mengkaji penerapan algoritma clustering K-Means untuk segmentasi pelanggan berdasarkan perilaku pembelian. Data sheet yang digunakan diperoleh dari kaggle, yang memuat data trend belanja konsumen. Studi ini fokus pada tujuh atribut utama yaitu usia, jenis kelamin, jumlah pembelian, kategori, frekuensi pembelian sebelumnya, dan metode pembayaran yang disukai. Algoritma k-means berhasil mengidentifikasi enam kelompok pelanggan dengan karakteristik unik yang mencerminkan pola pembelian tertentu. Temuan ini memberikan preferensi produk, kebiasaan belanja, dan karakteristik demografis pelanggan. Implikasi dari hasil ini dapat dimanfaatkan dalam strategi pemasaran yang lebih terarah dan personal, meningkatkan keterlibatan pelanggan, serta potensi penggunaannya untuk optimalisasi penggunaannya untuk optimalisasi pelanggan. Menggunakan dataset dari Kaggle yang berisi tren belanja konsumen, penelitian ini berfokus pada tujuh atribut utama: Usia, Jenis Kelamin, Jumlah Pembelian, Kategori, Frekuensi Pembelian, Pembelian Sebelumnya, dan Metode Pembayaran yang Disukai. Algoritma K-Means berhasil mengidentifikasi enam cluster pelanggan yang berbeda, masing-masing dengan karakteristik unik yang mencerminkan pola pembelian tertentu. Analisis ini memberikan

wawasan berharga tentang preferensi produk, kebiasaan pembelian, dan karakteristik demografis pelanggan. Temuan ini memiliki implikasi penting untuk strategi pemasaran dan pengambilan keputusan bisnis, memungkinkan perusahaan untuk mengembangkan pendekatan yang lebih terarah dan personal dalam keterlibatan dan retensi pelanggan. Studi ini mendemonstrasikan efektivitas algoritma K-Means dalam segmentasi pelanggan dan menyoroti potensinya untuk mengoptimalkan upaya pemasaran dan meningkatkan kepuasan pelanggan dalam berbagai konteks bisnis.

Kata Kunci: Algoritma K-Means, segmentasi pelanggan, perilaku pembelian, clustering, penambangan data

1. PENDAHULUAN

Dalam era digital saat ini, pengelolaan data merupakan menjadi faktor penting bagi perusahaan untuk tetap kompetitif. salah satu metode yang umum digunakan dalam pemrosesan data adalah data mining, yang memungkinkan perusahaan untuk mengekstrak pola atau informasi penting dari data besar. salah satu teknik utama dalam data mining adalah clustering, dimana data dikelompokkan berdasarkan kemiripan karakteristik. algoritma k-means sering digunakan dalam clustering karena kesederhanaan dan efisiensinya. segmentasi pelanggan melalui k-means sangat relevan dalam dunia bisnis karena memungkinkan perusahaan memahami pelanggan dengan lebih baik merancang strategi pemasaran yang lebih tepat sasaran dan meningkatkan efektivitas kampanye [1].

Penelitian ini bertujuan untuk mengidentifikasi pola dalam perilaku pelanggan menggunakan k-means, yang diharapkan dapat membantu perusahaan dalam mengoptimalkan strategi pemasaran mereka untuk meningkatkan loyalitas pelanggan. segmentasi ini memungkinkan perusahaan untuk mengetahui preferensi, frekuensi, dan jenis produk yang sering dibeli pelanggan. Dengan pemahaman ini, perusahaan dapat menawarkan promosi yang lebih personal dan meningkatkan retensi pelanggan [2].

Penerapan algoritma K-Means dalam segmentasi pelanggan menjadi sangat relevan di dunia bisnis. Segmentasi pelanggan adalah proses pembagian pelanggan menjadi subkelompok yang berbeda dan homogen berdasarkan perilaku, kebutuhan, atau karakteristik mereka. Dengan menggunakan segmentasi, perusahaan dapat lebih memahami pelanggan mereka, merancang strategi pemasaran yang lebih tepat sasaran, dan meningkatkan efektivitas kampanye pemasaran [2].

Dalam penelitian sebelumnya yang dilakukan oleh [N. Rofiqo dkk.], K-Means telah diaplikasikan pada berbagai kasus, termasuk pengelompokan data penderita penyakit, pemanfaatan sarana kesehatan, serta pengelompokan data berdasarkan karakteristik lainnya [4]. Keberhasilan penerapan K-Means dalam berbagai domain ini menunjukkan potensi algoritma ini untuk diaplikasikan dalam segmentasi pelanggan, di mana data yang dihasilkan dapat memberikan wawasan yang berguna bagi pengambilan keputusan bisnis.

Secara keseluruhan, penggunaan algoritma clustering, khususnya K-Means, dalam segmentasi pelanggan memberikan pendekatan yang efisien dan efektif dalam mengelompokkan pelanggan berdasarkan perilaku pembelian mereka. Dengan demikian, bisnis dapat mengoptimalkan strategi pemasaran dan meningkatkan retensi pelanggan melalui pendekatan yang lebih terarah dan personal [3][4].

2. TINJAUAN PUSTAKA

2.1 Data Mining dan Clustering

Data mining adalah proses untuk menemukan pola atau informasi penting dari kumpulan data besar yang dapat dimanfaatkan untuk pengambilan keputusan bisnis. Salah satu teknik utama dalam data mining adalah clustering, yaitu metode yang mengelompokkan data berdasarkan kemiripan karakteristiknya [1]. Teknik ini memungkinkan perusahaan untuk mengidentifikasi segmen pelanggan dengan pola tertentu yang mendukung strategi pemasaran yang lebih efektif [2]

2.2 Algoritma K-Means

Algoritma K-Means adalah salah satu metode clustering yang sering digunakan karena efisiensi dan kesederhanaannya. K-Means bekerja dengan membagi data ke dalam k kelompok atau cluster yang mirip berdasarkan jarak antar data. Algoritma ini telah diaplikasikan pada berbagai penelitian, seperti segmentasi data obat-obatan [3] dan analisis keluhan kesehatan [4]. Dalam penelitian tersebut, K-Means terbukti efektif untuk memproses data dengan berbagai karakteristik dan aplikasi.

2.3 Segmentasi Pelanggan Berdasarkan Perilaku Pembelian

Segmentasi pelanggan adalah proses mengelompokkan pelanggan ke dalam beberapa segmen yang memiliki kesamaan karakteristik, seperti perilaku pembelian. Segmentasi ini memungkinkan perusahaan untuk mengidentifikasi preferensi pelanggan dan menyusun strategi pemasaran yang lebih personal dan efektif. Menurut Kotler dan Keller (2009), segmentasi pelanggan memainkan peran penting dalam menyusun strategi pemasaran yang tepat sasaran [5]. Segmentasi ini memungkinkan perusahaan untuk mengidentifikasi preferensi pelanggan dan merancang strategi pemasaran yang lebih personal dan efektif.

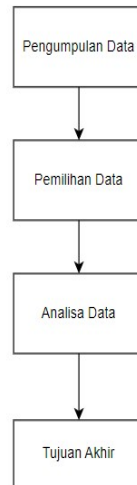
2.4 Penerapan K-Means dalam Segmentasi Pelanggan

Penelitian terkait penerapan algoritma K-Means untuk segmentasi pelanggan telah menunjukkan hasil positif. Studi oleh Perdana dan Florentin [2] menggunakan K-Means untuk menganalisis segmentasi pelanggan pada aplikasi Alfagift. Hasilnya, K-Means mampu mengelompokkan pelanggan dengan baik berdasarkan pola penggunaan aplikasi dan frekuensi pembelian. Penelitian lain oleh Saputra et al. [5] menggunakan algoritma clustering pada deteksi penyakit tanaman. Walaupun berbeda dalam aplikasi, penelitian ini menunjukkan keunggulan K-Means dalam pengelompokan data berdasarkan kemiripan atribut.

2.5 Penelitian Terkait Lainnya

Penelitian lain menunjukkan efektivitas K-means dalam berbagai aplikasi, seperti analisis pelanggan di industri e-commerce[7]. Identifikasi kelompok pelanggan berdasarkan nilai transaksi[8], segmentasi pelanggan berbasis perilaku pembelian untuk efektivitas pemasaran[9], dan analisis pelanggan di industri ritel[10].

3. METODE PENELITIAN

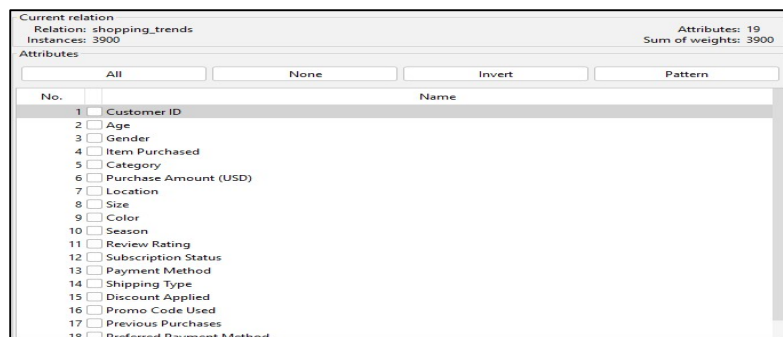


Gambar 1. Metode Penelitian

Dalam melakukan pengolahan data, Peneliti melakukan berbagai metode-metode agar menghasilkan sebuah data yang dapat memberikan informasi seperti pada gambar 1. Adapun metode yang Peneliti lakukan diantaranya:

1. Pengumpulan Data

Dataset diambil dari kaggle, yang terdiri dari dua jenis data yaitu, shopping_behavior_updated, dan shopping_trends. Dataset ini memuat tren belanja pelanggan yang menjadi dasar analisis seperti gambar 2.



Gambar 2. Dataset Tren Pelanggan

2. Pemilihan Data

Dari dataset yang diperoleh penelitian fokus pada shopping-trends dengan atribut kunci: usia, jenis kelamin, jumlah pembelian(USD), kategori, frekuensi pembelian, pembelian sebelumnya dan metode pembayaran yang disukai.

3. Analisa Data

Data diolah dengan K-Means clustering untuk menemukan pola pembelian dalam kelompok pelanggan. proses clustering ini menghasilkan enam kelompok pelanggan dengan karakteristik yang berbeda.

4. Evaluasi dan interpretasi

Setelah cluster dianalisis untuk memahami karakteristik pembelian, yang akan dihubungkan dengan strategi pemasaran yang disesuaikan untuk setiap segmen pelanggan..

4. HASIL DAN PEMBAHASAN

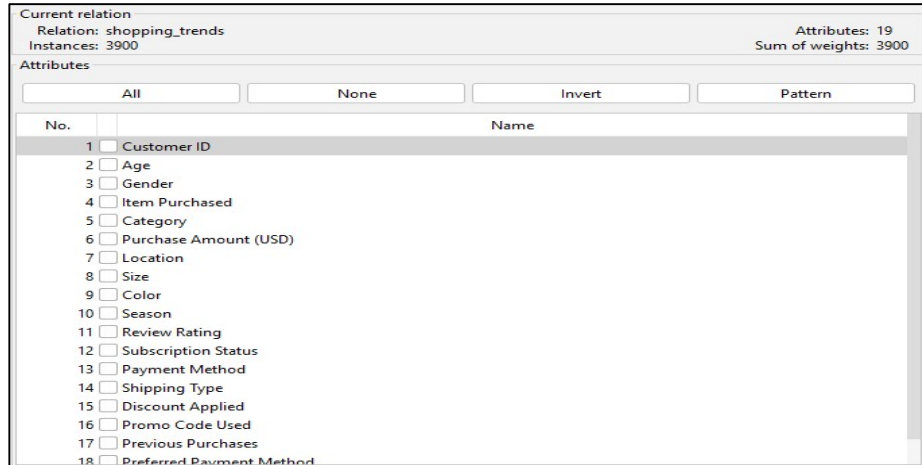
Dalam penelitian ini, peneliti menerapkan algoritma K-Means clustering untuk melakukan segmentasi pelanggan berdasarkan perilaku pembelian mereka. Dataset yang digunakan berasal dari Kaggle, khususnya dataset "shopping_trends" yang berisi informasi tentang tren belanja konsumen.

Proses analisis dimulai dengan pemilihan dan penyaringan atribut yang relevan dari dataset awal. Dari 19 atribut yang tersedia, peneliti memfokuskan pada 7 atribut kunci yang mencerminkan perilaku pembelian pelanggan: Usia, Jenis Kelamin, Jumlah Pembelian (USD), Kategori, Frekuensi Pembelian, Pembelian Sebelumnya, dan Metode Pembayaran yang Disukai.

Setelah melakukan clustering menggunakan algoritma K-Means, peneliti berhasil mengidentifikasi 6 cluster pelanggan yang berbeda. Setiap cluster memiliki karakteristik unik yang mencerminkan pola perilaku pembelian tertentu. Analisis mendalam terhadap masing-masing cluster ini memberikan wawasan berharga tentang preferensi produk, kebiasaan pembelian, dan karakteristik demografis pelanggan.

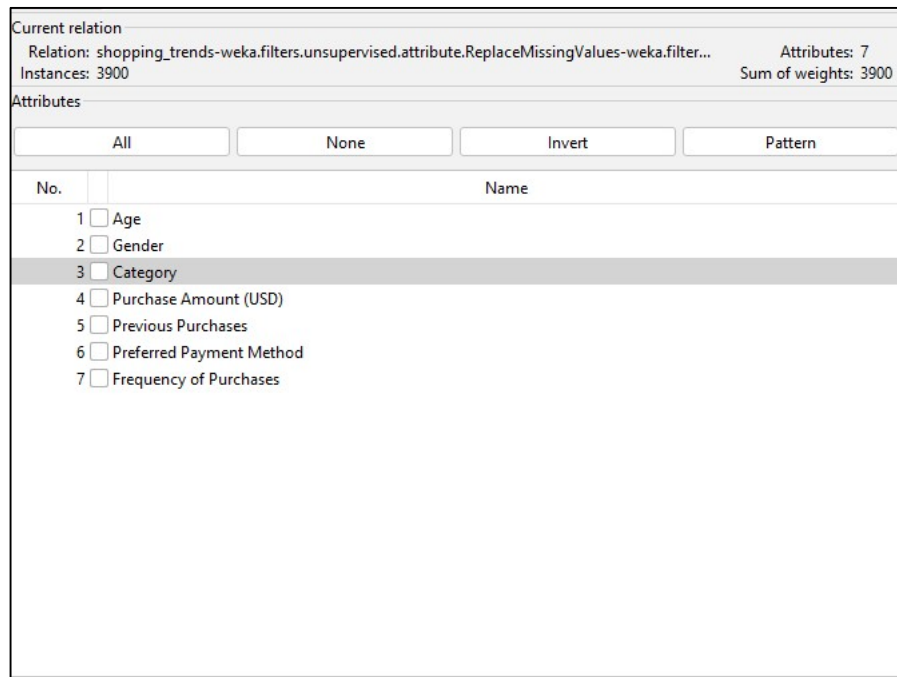
Dalam bagian berikut menyajikan hasil clustering secara rinci, menguraikan profil masing-masing cluster, dan mendiskusikan implikasi temuan ini untuk strategi pemasaran

dan pengambilan keputusan bisnis. Berikut ini dataset yang belum diproses tersaji pada gambar 3.



Gambar 3. Dataset Sebelum Diproses

Dataset shopping_trends sebelum dilakukan proses ada 19 atribut didalam dataset tersebut yang berisikan tentang trend belanja tersaji pada gambar 4.



Gambar 4. Dataset Setelah Diproses

Pada dataset shopping_trends yang awalnya terdapat 19 atribut dilakukan filter atribut menjadi 7 atribut yaitu Age, Gender, Purchase Amount (USD), Category, Frequency of Purchases, Previous Purchases, Preferred Payment Method. Pemilihan atribut

disesuaikan dengan tujuan untuk mencari Segmentasi Pelanggan Berdasarkan Perilaku Pembelian tersaji pada gambar 5.

```
Cluster 0: 27,Female,Accessories,56,33,'Credit Card',Monthly
Cluster 1: 44,Male,Footwear,55,47,Cash,Weekly
Cluster 2: 38,Male,Footwear,21,1,'Bank Transfer',Weekly
Cluster 3: 38,Female,Clothing,39,36,PayPal,Annually
Cluster 4: 69,Male,Clothing,90,23,'Bank Transfer',Monthly
Cluster 5: 57,Female,Clothing,79,24,Venmo,Quarterly

Missing values globally replaced with mean/mode

Final cluster centroids:
Attribute          Full Data          Cluster#
                   (3900.0)          0           1           2           3           4           5
-----
Age                44.0685           46.5285     43.7771     40.1337     41.7561     48.2029     41.469
Gender             Male              Female      Male         Male         Female      Male         Female
Category           Clothing          Accessories Footwear     Clothing     Clothing     Clothing     Clothing
Purchase Amount (USD) 59.7644           61.3742     55.8312     61.6066     48           62.7475     70.1032
Previous Purchases  25.3515           19.2304     29.7887     21.6822     23.3478     24.6549     29.6679
Preferred Payment Method PayPal             Credit Card  Cash Bank Transfer PayPal Bank Transfer Venmo
Frequency of Purchases Every 3 Months     Monthly      Weekly       Weekly       Annually     Monthly     Quarterly

Time taken to build model (full training data) : 0.3 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      473 ( 12%)
1      942 ( 24%)
2      516 ( 13%)
3      529 ( 14%)
4      907 ( 23%)
5      533 ( 14%)
```

Gambar 5. Hasil Cluster

Hasil clustering yang telah dilakukan menggunakan algoritma K-Means menghasilkan 6 cluster yang berbeda berdasarkan perilaku pembelian pelanggan.

a. Cluster 0 (12% dari data)

1. Usia: 46,53 tahun (Mayoritas dewasa).
2. Gender: Female (Wanita).
3. Kategori: Accessories (Aksesori).
4. Purchase Amount: \$61,37.
5. Previous Purchases: 19,23 kali pembelian sebelumnya.
6. Preferred Payment Method: Credit Card (Kartu Kredit).
7. Frequency of Purchases: Monthly (Bulanan).

Pelanggan dewasa mayoritas wanita sering membeli aksesoris dengan metode pembayaran kartu kredit dan memiliki frekuensi pembelian bulanan.

b. Cluster 1 (24% dari data)

1. Usia: 43,78 tahun.
2. Gender: Male (Pria).
3. Kategori: Accessories (Aksesori).
4. Purchase Amount: \$55,83.

5. Previous Purchases: 29,79 kali pembelian sebelumnya.

6. Preferred Payment Method: Cash (Tunai).

7. Frequency of Purchases: Weekly (Mingguan).

pria dewasa yang sering membeli aksesoris mingguan dengan preferensi pembayaran tunai, menunjukkan pelanggan setia dengan metode pembayaran tradisional.

c. Cluster 2 (13% dari data)

1. Usia: 40,13 tahun.

2. Gender: Male (Pria).

3. Kategori: Footwear (Sepatu).

4. Purchase Amount: \$61,61.

5. Previous Purchases: 21,68 kali pembelian sebelumnya.

6. Preferred Payment Method: Bank Transfer.

7. Frequency of Purchases: Weekly (Mingguan).

pria dewasa yang sering membeli sepatu dengan transfer bank dan frekuensi mingguan, menunjukkan kecenderungan pada metode pembayaran non-tunai.

d. Cluster 3 (14% dari data)

1. Usia: 41,76 tahun.

2. Gender: Female (Wanita).

3. Kategori: Clothing (Pakaian).

4. Purchase Amount: \$48.

5. Previous Purchases: 23,35 kali pembelian sebelumnya.

6. Preferred Payment Method: PayPal.

7. Frequency of Purchases: Annually (Tahunan).

wanita yang membeli pakaian dengan harga rendah secara tahunan, lebih sering menggunakan paypal sebagai metode pembayaran.

e. Cluster 4 (23% dari data)

1. Usia: 48,20 tahun.

2. Gender: Male (Pria).

3. Kategori: Clothing (Pakaian).

4. Purchase Amount: \$62,75.

5. Previous Purchases: 24,65 kali pembelian sebelumnya.

6. Preferred Payment Method: Bank Transfer.

7. Frequency of Purchases: Monthly (Bulanan).

pria dewasa yang membeli pakaian dengan frekuensi bulanan dan transfer bank, menunjukkan loyalitas pada produk pakaian.

f. Cluster 5 (14% dari data)

1. Usia: 41,47 tahun.
2. Gender: Female (Wanita).
3. Kategori: Clothing (Pakaian).
4. Purchase Amount: \$70,10.
5. Previous Purchases: 29,67 kali pembelian sebelumnya.
6. Preferred Payment Method: Venmo.
7. Frequency of Purchases: Quarterly (Triwulanan).

wanita dengan pola pembelian triwulanan pada pakaian dengan preferensi venmo sebagai metode pembayaran menunjukkan adaptasi terhadap metode pembayaran modern.

Analisis cluster menunjukkan diversifikasi signifikan dalam perilaku pembelian pelanggan. temuan ini menjadi dasar dalam strategi pemasaran yang lebih terarah dan personal untuk meningkatkan keterlibatan pelanggan sesuai dengan karakteristik dan preferensi setiap kelompok.

5. KESIMPULAN

Penerapan algoritma K-Means berhasil mengidentifikasi enam kelompok pelanggan dengan karakteristik yang berbeda dalam hal usia, jenis kelamin, kategori produk, frekuensi pembelian, dan metode pembayaran. analisis ini memberikan wawasan penting tentang pola pembelian yang beragam dan dapat dimanfaatkan untuk strategi pemasaran yang lebih efektif. temuan ini memungkinkan perusahaan untuk memberikan pelayanan yang lebih sesuai dengan kebutuhan pelanggan yang akhirnya dapat meningkatkan loyalitas pelanggan dan efisiensi dalam alokasi sumber daya.

DAFTAR PUSTAKA

- [1] Fujiyama, "Pengertian Data Mining dan Implementasinya di Berbagai Sektor," <https://surabaya.telkomuniversity.ac.id/>. Accessed: Aug. 15, 2024. [Online]. Available: <https://surabaya.telkomuniversity.ac.id/pengertian-data-mining-dan-implementasinya-di-berbagai-sektor/>
- [2] A. S. Satria Ardi Perdana, Sara Famayla Florentin, "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.
- [3] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [4] N. Rofiqo, A. P. Windarto, and D. Hartama, "Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 216–223, 2018, doi: 10.30865/komik.v2i1.929.
- [5] D. F. S. Rizal Amegia Saputra, Sri Wasyianti, Adi Supriyatna, "Penerapan Algoritma Convolutional Neural Network Dan Arsitektur MobileNet Pada Aplikasi Deteksi Penyakit Daun Padi," *J. SWABUMI*, vol. 9, no. 2, 2021, [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Rice>
- [6] A. Rahman, Y. Wijaya, "Penerapan Algoritma K-Means untuk ClusterData Penjualan Produk Fashion," *J. Teknologi Informasi dan Komunikasi*, vol. 10, no. 1, pp. 44-50, 2020. doi: 10.30587/jtic.v10i1.1367.
- [7] I. Sari, R. Prasetyo, "Analisis Segmentasi Pelanggan Menggunakan Metode K-Means Clustering dalam Ritel," *Jurnal Ilmu Komputer dan Informatika*, vol. 8, no. 3, pp. 159-165, 2021. doi: 10.32498/jiki.v8i3.2160.
- [8] F. Saputra, M. J. Ananta, "Analisis Data Mining untuk Segmentasi Pelanggan dengan Algoritma K-Means," *Prosiding Seminar Nasional Teknologi dan Informasi*, 2021, pp. 88-95. doi: 10.26905/snati.v8i1.115

- [9] H. Lestari, D. Nugroho, "Clustering Data Konsumen dengan Algoritma K-Means untuk Analisis Preferensi Produk," *Jurnal Sistem Informasi*, vol. 13, no. 4, pp. 124-130, 2019. doi: 10.14710/jsi.13.4.124-130.
- [10] J. Widodo, B. Astuti, "Implementasi K-Means Clustering untuk Segmentasi Pelanggan pada Industri Kuliner," *Jurnal Teknologi Informasi*, vol. 5, no. 2, pp. 22-29, 2023. doi: 10.32388/jti.v5i2.2751.