

## Analisis Profitabilitas Pasar Resale Sneaker Jordan Menggunakan Regresi dan K-Means Clustering untuk Mendukung Strategi Penjualan

Rega Esa Ramadhan<sup>1\*</sup>, Ahmad Al Fajri<sup>2</sup>, Wahyu Omar Gibran<sup>3</sup>

<sup>1,2,3</sup>Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Indonesia

<sup>1\*</sup>regaesa211@gmail.com, <sup>2</sup>ahmadalfajri1999@gmail.com <sup>3</sup>Omarwahyu084@gmail.com  
(regaesa211@gmail.com)

(Email Correspondence: regaesa211@gmail.com)

Received: 4 April 2026 | Revised: 8 April 2026 | Accepted: 30 April 2026

### Abstrak

Pasar resale sneaker mengalami pertumbuhan yang pesat dan menjadi peluang bisnis yang menarik karena adanya selisih antara harga ritel dan harga jual kembali. Namun, perubahan harga resale yang cukup dinamis sering menyulitkan penjual dalam menentukan waktu dan channel penjualan yang mampu menghasilkan keuntungan maksimal. Penelitian ini bertujuan untuk menganalisis profitabilitas pasar resale sneaker Jordan menggunakan metode Random Forest Regression dan K-Means Clustering. Dataset yang digunakan terdiri dari 5.000 data transaksi sneaker yang mencakup harga ritel, harga resale, profit margin, channel penjualan, tanggal penjualan, dan lama penyimpanan produk. Random Forest Regression digunakan untuk memprediksi profit, sedangkan K-Means Clustering digunakan untuk mengelompokkan produk berdasarkan tingkat profitabilitasnya. Hasil penelitian menunjukkan nilai Mean Absolute Error (MAE) sebesar 1,41 dan nilai  $R^2$  sebesar 1,00 yang menunjukkan kinerja model yang sangat baik. Selain itu, terbentuk tiga cluster produk dengan karakteristik profit yang berbeda. Hasil analisis menunjukkan bahwa harga resale merupakan faktor yang paling berpengaruh terhadap profit, Walk-in Retail menjadi channel dengan profit tertinggi, dan bulan Mei merupakan waktu penjualan yang paling menguntungkan.

Kata kunci: Data Mining, Random Forest, K-Means Clustering, Prediksi Profit, Resale Sneaker.

### Abstract

The resale sneaker market has experienced significant growth and become an attractive business opportunity due to the difference between retail and resale prices. However, fluctuations in resale prices often make it difficult for sellers to determine the best selling time and sales channel to maximize profit. This study aims to analyze the profitability of the Jordan sneaker resale market using Random Forest Regression and K-Means Clustering. The dataset used consists of 5,000 sneaker transaction records containing retail price, resale price, profit margin, sales channel, sales date, and inventory duration. Random Forest Regression was applied to predict profit, while K-Means Clustering was used to segment products based on profitability characteristics. The regression model achieved a Mean Absolute Error (MAE) of 1.41 and an  $R^2$  value of 1.00, indicating excellent predictive performance. Clustering results produced three product segments with different profitability levels. The findings show that resale price is the most influential factor affecting profit, Walk-in Retail generates the highest average profit, and May is the most profitable month for selling sneakers. These results can support data-driven decision making in resale sneaker businesses.

Keywords: : Data Mining, Random Forest, K-Means Clustering, Profit Prediction, Sneaker Resale

## 1. PENDAHULUAN

Pasar resale sneaker mengalami perkembangan yang cukup pesat dalam beberapa tahun terakhir. Sneaker tidak lagi hanya digunakan sebagai produk fashion, tetapi juga memiliki nilai investasi karena dapat dijual kembali dengan harga yang lebih tinggi. Perbedaan antara harga ritel dan harga resale menciptakan peluang keuntungan yang menarik bagi penjual maupun kolektor. Kondisi ini membuat persaingan dalam pasar resale semakin meningkat.

Harga resale sneaker bersifat dinamis dan dipengaruhi oleh berbagai faktor, seperti harga ritel, permintaan pasar, lama penyimpanan produk, serta channel penjualan yang digunakan. Perubahan tersebut menyebabkan pelaku bisnis sering mengalami kesulitan dalam menentukan waktu dan media penjualan yang paling menguntungkan. Akibatnya, keputusan bisnis sering kali hanya didasarkan pada pengalaman tanpa dukungan analisis data yang memadai.

Melalui analisis data historis, data mining dapat membantu proses pengambilan keputusan. Sebagai hasil dari metode ini, data yang lebih akurat dapat dihasilkan dengan menemukan pola dan hubungan antar variabel. Pelaku bisnis dapat menggunakan data mining untuk mengetahui apa yang memengaruhi keuntungan dan membuat strategi penjualan yang lebih baik. Penelitian mengenai segmentasi pelanggan dan analisis transaksi penjualan telah dilakukan menggunakan metode data mining dan clustering untuk mendukung pengambilan keputusan bisnis [1], [2]. Penelitian lain juga menunjukkan bahwa segmentasi pelanggan berbasis clustering mampu membantu perusahaan dalam mengidentifikasi karakteristik pelanggan dan menyusun strategi pemasaran yang lebih efektif [3], [4].

Studi sebelumnya menunjukkan bahwa algoritma Random Forest dapat memprediksi data penjualan dengan baik. Metode Random Forest telah banyak digunakan dalam prediksi penjualan karena mampu menghasilkan tingkat akurasi yang baik pada berbagai jenis data penjualan dan forecasting produk [7], [10].

Metode K-Means biasanya digunakan untuk segmentasi data selain regresi. Metode K-Means Clustering terbukti efektif dalam mengelompokkan pelanggan maupun transaksi penjualan berdasarkan karakteristik yang serupa [5], [6]. Pendekatan clustering juga telah diterapkan pada berbagai sektor bisnis untuk menganalisis pola transaksi pelanggan dan menentukan segmentasi pasar [11], [12]. Selain digunakan untuk segmentasi pelanggan, metode K-Means juga telah diterapkan untuk mengelompokkan data penjualan produk berdasarkan karakteristik transaksi yang dimiliki [8], [13].

Penelitian ini menggunakan Random Forest Regression dan K-Means Clustering pada dataset pasar resale sneaker Jordan. Kombinasi kedua metode tersebut dapat memberikan informasi yang lebih komprehensif dalam mendukung pengambilan keputusan bisnis berbasis data [7], [9]. Metode segmentasi pelanggan berbasis K-Means juga terbukti mampu mengelompokkan pelanggan berdasarkan pola transaksi dan tingkat loyalitas yang dimiliki [14], [15].

## 2. METODE PENELITIAN

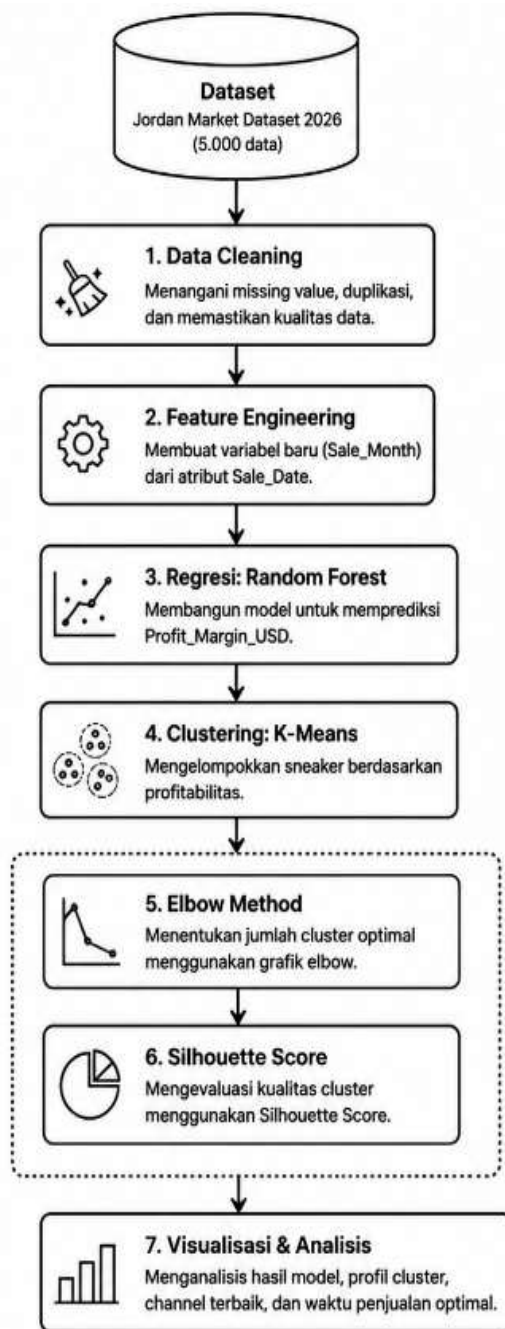
Penelitian ini menggunakan dataset Jordan Market Dataset 2026 yang terdiri dari 5000 data transaksi sneaker Jordan. Dataset tersebut memuat informasi mengenai harga ritel, harga resale, profit margin, channel penjualan, tanggal penjualan, serta lama penyimpanan produk. Data kemudian melalui proses pembersihan untuk menghilangkan nilai kosong dan data duplikat, namun jumlah data tetap 5000 data setelah proses cleaning. Dataset ini digunakan untuk menganalisis faktor-faktor yang memengaruhi profit serta mendukung pengambilan keputusan dalam menentukan strategi penjualan yang lebih optimal.

Pada awal penelitian, langkah pertama yang dilakukan adalah mempersiapkan data dengan membersihkan data yang tidak lengkap dan data yang terduplikasi. Selanjutnya dilakukan proses feature engineering dengan menambahkan atribut `Sale_Month` yang diambil dari tanggal transaksi. Tujuan dari tahap ini adalah agar data memiliki kualitas yang baik sebelum digunakan dalam proses pembuatan model.

Untuk memprediksi profit penjualan sepatu, penelitian ini menggunakan metode Random Forest Regression. Algoritma ini dipilih karena memiliki kemampuan untuk menangani hubungan nonlinier di antara berbagai variabel dalam data yang beragam serta menghasilkan performa prediksi yang baik pada data penjualan dan forecasting [7], [10].

Penelitian ini juga menggunakan K-Means Clustering untuk mengelompokkan sepatu berdasarkan tingkat profitabilitasnya. Metode K-Means telah banyak digunakan untuk segmentasi pelanggan dan transaksi penjualan berdasarkan karakteristik yang serupa [4], [5]. Selain itu, metode ini efektif digunakan

untuk mengidentifikasi kelompok pelanggan berdasarkan perilaku transaksi dan tingkat profitabilitas [6], [14]. Selain itu, pendekatan K-Means berbasis model RFM banyak digunakan untuk mengelompokkan pelanggan berdasarkan nilai transaksi dan loyalitas pelanggan [12], [15]. Jumlah cluster ditentukan menggunakan metode Elbow, sedangkan kualitas cluster dievaluasi menggunakan Silhouette Score.



**Gambar 1.** Alur Penelitian Data Mining pada Pasar Resale Sneaker Jordan.

Gambar 1 menunjukkan tahapan penelitian yang dimulai dari proses pengumpulan dan persiapan data, dilanjutkan dengan pemodelan menggunakan *Random Forest Regression* dan *K-Means Clustering*. Setelah proses pemodelan selesai, dilakukan evaluasi menggunakan *Elbow method* dan *Silhouette Score*, kemudian hasilnya dianalisis melalui visualisasi untuk memperoleh informasi mengenai profitabilitas, segmentasi produk, channel penjualan, dan waktu penjualan yang paling optimal.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil Regresi

**Tabel 1.** Hasil Regresi

| Metrik         | Nilai |
|----------------|-------|
| MAE            | 1.41  |
| R <sup>2</sup> | 1.00  |

Model Random Forest Regression berhasil digunakan untuk memprediksi profit penjualan sneaker. Hasil evaluasi menunjukkan nilai MAE sebesar 1,41 dan nilai R<sup>2</sup> sebesar 1,00. Nilai tersebut menunjukkan bahwa model memiliki tingkat akurasi yang sangat baik dalam mempelajari pola data yang digunakan pada penelitian ini.

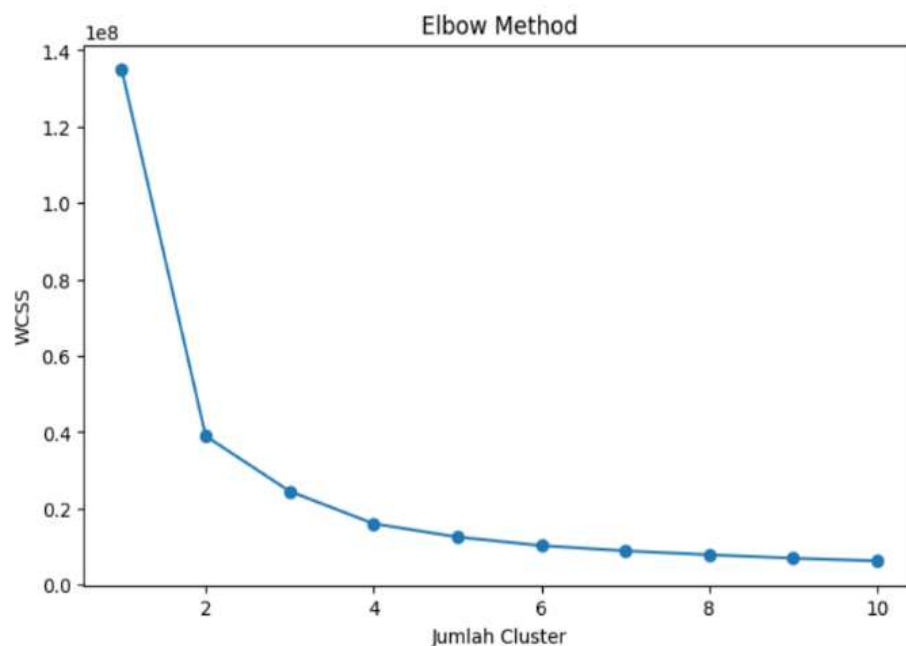
#### 3.2. Analisis Feature Importance

**Tabel 2.** Feature Importance Random Forest Regression

| Variabel       | Importance |
|----------------|------------|
| Resale Price   | 0.9679     |
| Retail Price   | 0.0319     |
| Days Inventory | 0.00007    |
| Sale Month     | 0.00005    |

Berdasarkan hasil feature importance, variabel Resale Price memiliki pengaruh paling besar terhadap profit dengan nilai kepentingan sebesar 96,79%. Sementara itu, Retail Price memberikan pengaruh yang lebih kecil, sedangkan Days in Inventory dan Sale Month memiliki kontribusi yang relatif rendah terhadap profit.

#### 3.3. Hasil Clustering



**Gambar 2.** Hasil Elbow Method

Penentuan jumlah cluster dilakukan menggunakan Elbow Method. Berdasarkan grafik yang diperoleh, titik siku terlihat pada nilai  $K=3$  sehingga dipilih tiga cluster sebagai jumlah cluster yang paling sesuai. Pemilihan tersebut digunakan sebagai dasar dalam proses segmentasi data sneaker.

### 3.4. Evaluasi Clustering

**Tabel 3.** Silhouette Score

| Metrik           | Nilai |
|------------------|-------|
| Silhouette Score | 0.453 |

Evaluasi cluster dilakukan menggunakan Silhouette Score dan menghasilkan nilai sebesar 0,453. Nilai tersebut menunjukkan bahwa cluster yang terbentuk memiliki kualitas pemisahan yang cukup baik. Hasil ini menunjukkan bahwa data berhasil dikelompokkan berdasarkan karakteristik yang dimiliki.

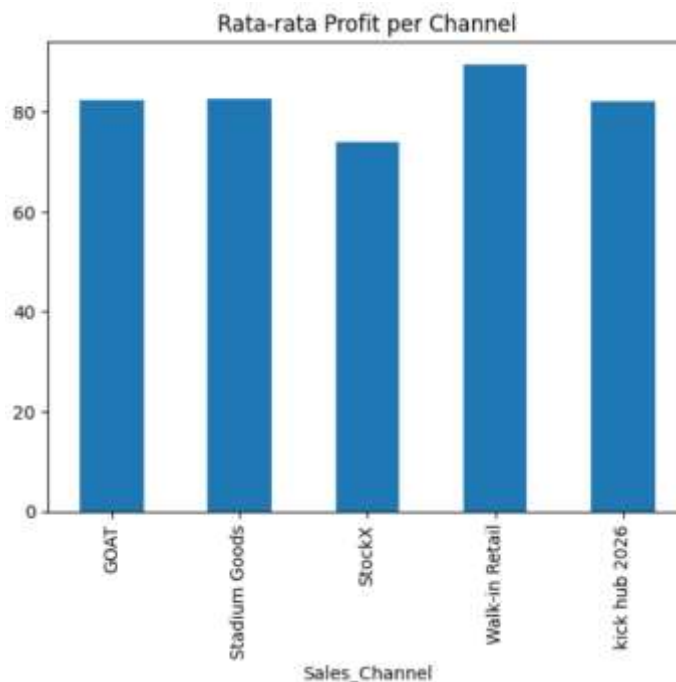
### 3.5. Analisis Karakteristik Cluster



**Gambar 3.** Visualisasi Cluster Produk

Hasil klasterisasi menghasilkan tiga kelompok produk dengan karakteristik yang berbeda. Cluster 2 memiliki rata-rata profit tertinggi, sedangkan Cluster 1 menunjukkan profit yang rendah bahkan cenderung mengalami kerugian. Sementara itu, Cluster 0 berada pada kategori profit menengah dan relatif lebih stabil dibandingkan cluster lainnya.

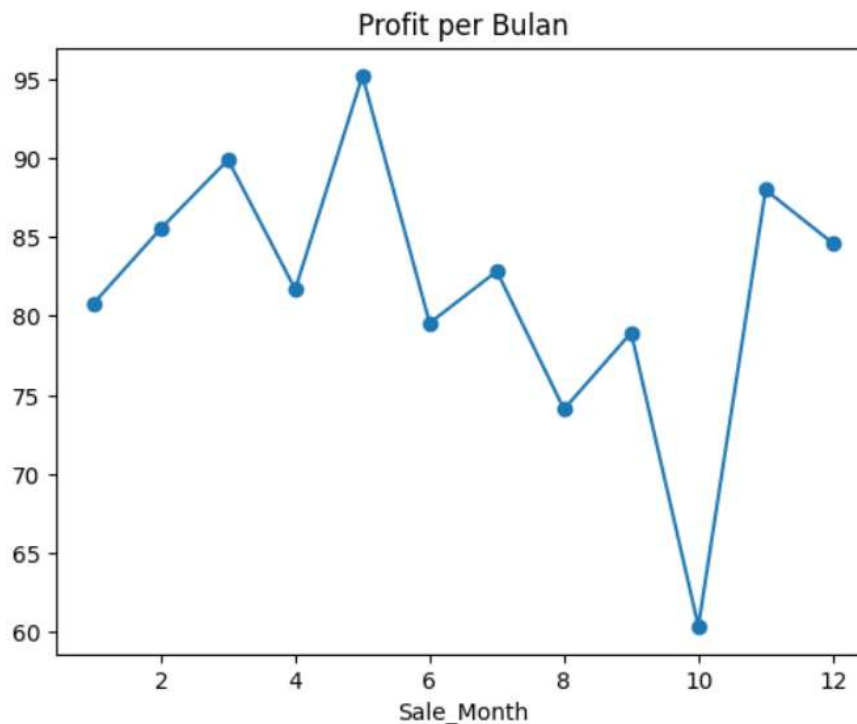
### 3.6. Analisis Profit Berdasarkan Channel Penjualan



**Gambar 5.** Rata-rata Profit Berdasarkan Channel Penjualan

Analisis menunjukkan bahwa Walk-in Retail menghasilkan rata-rata profit tertinggi dibandingkan channel lainnya. Temuan ini menunjukkan bahwa pemilihan channel penjualan memiliki pengaruh terhadap keuntungan yang diperoleh. Oleh karena itu, channel dengan profit tertinggi dapat dijadikan prioritas dalam strategi penjualan.

### 3.7. Analisis Profit Berdasarkan Bulan Penjualan



**Gambar 6.** Rata-rata Profit Berdasarkan Bulan Penjualan



Hasil analisis menunjukkan bahwa bulan Mei memiliki rata-rata profit tertinggi dibandingkan bulan lainnya. Sebaliknya, beberapa bulan menunjukkan nilai profit yang lebih rendah. Hal ini mengindikasikan bahwa waktu penjualan dapat memengaruhi peluang keuntungan yang diperoleh.

### 3.8. Menjawab Rumusan Masalah

Berdasarkan hasil analisis, waktu penjualan yang paling menguntungkan berada pada bulan Mei, Maret, dan November. Channel penjualan yang memberikan profit tertinggi adalah Walk-in Retail. Selain itu, produk yang termasuk dalam Cluster 2 memiliki tingkat profitabilitas paling tinggi sehingga dapat dijadikan prioritas dalam strategi penjualan dan pengelolaan stok.

## 4. KESIMPULAN

Berdasarkan hasil penelitian, metode Random Forest Regression dan K-Means Clustering berhasil diterapkan untuk menganalisis tingkat keuntungan di pasar reseller sepatu Jordan. Model regresi memberikan angka MAE sebesar 1,41 dan  $R^2$  sebesar 1,00, yang menunjukkan model memiliki kemampuan memprediksi yang sangat bagus. Hasil penelitian juga menunjukkan bahwa harga jual kembali merupakan faktor yang paling berpengaruh terhadap laba. Selain itu, Walk-in Retail merupakan saluran penjualan yang memiliki tingkat keuntungan rata-rata tertinggi, dan bulan Mei menjadi masa penjualan yang paling menguntungkan. Dengan menggunakan metode K-Means, data berhasil dibagi menjadi tiga kelompok yang bisa digunakan sebagai acuan dalam merancang strategi penjualan dan pengelolaan stok. Hasil penelitian ini dapat digunakan sebagai dasar pengambilan keputusan bagi pelaku bisnis resale sneaker dalam menentukan strategi penjualan, pengelolaan stok, serta pemilihan channel penjualan yang lebih optimal.

## DAFTAR PUSTAKA

- [1]. Kapti, K., & Astuti, D. (2025). Implementation of the K-Means Clustering Algorithm for Customer Segmentation. *Sistemasi*, 14(6), 2751–2767. <https://doi.org/10.32520/stmsi.v14i6.5573>
- [2]. Nugroho, B. I., Rafhina, A., Ananda, P. S., & Gunawan, G. (2024). Customer Segmentation in Sales Transaction Data Using K-Means Clustering Algorithm. *Journal of Intelligent Decision Support System*, 7(2). <https://doi.org/10.35335/idss.v7i2.236>
- [3]. Mawarni, H., Testiana, G., & Dalafranka, M. L. (2023). Implementasi Algoritma K-Means Untuk Segmentasi Pelanggan Pada PT. Bintang Multi Sarana Cabang Tugumulyo. *J-ICON*, 11(2). <https://doi.org/10.35508/jicon.v11i2.12478>
- [4]. Djun, S. F., Gunadi, I. G. A., & Sariyasa, S. (2024). Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means Clustering pada Model Data RFM. *JTIM*, 5(4), 354–364. <https://doi.org/10.35746/jtim.v5i4.434>
- [5]. Kristanti, B. T., Junaidi, A., & Mandyartha, E. P. (2024). Implementasi K-Means Clustering dalam Segmentasi Pelanggan Berdasarkan Usia, Pendapatan, dan Model RFM. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3). <https://doi.org/10.23960/jitet.v12i3.4677>
- [6]. Yuliasih, B. N., Herman, H., & Sunardi, S. (2024). K-Means Clustering Method for Customer Segmentation Based on Potential Purchases. *ELTIKOM*, 8(1). <https://doi.org/10.31961/eltikom.v8i1.1137>
- [7]. Hidayat, R., Saputra, H. T., Husnah, M., Nabila, N., Hidayatullah, M. B., Nazhmi, M. N., Azra, J., & Rana, A. (2025). Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket. *SIMKOM*, 10(1), 101–109. <https://doi.org/10.51717/simkom.v10i1.703>
- [8]. Azzahra, L., & Yasir, A. (2024). Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food. *Jurnal Ilmu Komputer dan Sistem Informasi*, 3(1). <https://doi.org/10.70340/jirsi.v3i1.88>
- [9]. Putri, Y., Aldo, D., & Ilham, W. (2024). Retail Marketing Strategy Optimization: Customer Segmentation with Artificial Intelligence Integration and K-Means Clustering. *Sinkron : Jurnal dan Penelitian Teknik Informatika*, 8(4), 2155–2163. <https://doi.org/10.33395/sinkron.v8i4.14000>
- [10]. Agustina, A., Tukino, T., Huda, B., & Novalia, E. (2025). Prediksi Volume Penjualan Gadget Berdasarkan Promo dan Channel Penjualan Menggunakan Random Forest. *JUSIFOR*, 4(1). <https://doi.org/10.70609/jusifor.v4i1.6962>

- [11]. Yani, A., Azmi, Z., & Suherdi, D. (2023). Implementasi Data Mining Menganalisa Data Penjualan Menggunakan Algoritma K-Means Clustering. *Jurnal Sistem Informasi Triguna Dharma*, 2(2), 315–323. <https://doi.org/10.53513/jursi.v2i2.6357>
- [12]. Widiyanto, A. T., & Witanti, A. (2021). Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, 1(1), 204–215. <https://doi.org/10.24002/konstelasi.v1i1.4293>
- [13]. Pramudiansyah, A. P. (2021). Segmentasi Pelanggan Menggunakan Algoritma K-Means Berdasarkan Model Recency Frequency Monetary. *Jurnal Ilmiah Ilmu Komputer*, 7(2), 6–19. <https://doi.org/10.35329/jiik.v7i2.201>
- [14]. Kadarsah, D., & Heikal, J. (2024). Customer Segmentation With K-Means Clustering Suzuki Mobil Bandung Customer Case Study. *Jurnal Indonesia Sosial Teknologi*, 5(3), 768–774. <https://doi.org/10.59141/jist.v5i3.935>
- [15]. Mulyadi, I. P. (2022). Clusterisasi Menggunakan Metode Algoritma K-Means dalam Meningkatkan Penjualan Tupperware. *Jurnal Informatika Ekonomi Bisnis*, 4(4), 323–328. <https://doi.org/10.37034/infeb.v4i4.164>