



Pengelompokan Wilayah Pembelian Produk Cindrelia.Id Menggunakan Metode K-Means dan Apriori

Cindy Aurelia Hardi, Bagus Mulyawan, Meirista Wulandari

Universitas Tarumanagara, Indonesia

E-mail: cindy.535220110@stu.untar.ac.id¹, bagus@fti.untar.ac.id²,
meiristaw@ft.untar.ac.id³

Abstrak

Perkembangan e-commerce di Indonesia telah mendorong pelaku usaha fashion lokal untuk memanfaatkan data penjualan dalam pengambilan keputusan strategis. Namun, banyak UMKM fashion yang menghadapi tantangan dalam menganalisis data geografis pelanggan dan pola pembelian produk untuk optimalisasi promosi dan ekspansi bisnis. Cindrelia.id merupakan salah satu brand fashion lokal yang menjual produk pakaian wanita melalui platform e-commerce dan media sosial. Saat ini, toko menghadapi masalah dalam menganalisis data penjualan untuk memperoleh informasi yang bermanfaat terutama terkait sebaran geografis pelanggan dan rekomendasi produk untuk promosi. Hal ini menyebabkan toko sulit melakukan pengambilan keputusan dalam menentukan lokasi promosi, bazar, maupun toko offline. Kesenjangan penelitian (research gap) yang teridentifikasi adalah minimnya penelitian yang mengintegrasikan clustering spasial dengan association rule mining dalam satu platform web GIS interaktif untuk mendukung keputusan bisnis retail fashion di Indonesia. Penelitian sebelumnya cenderung berfokus pada satu metode saja atau tidak mengintegrasikan visualisasi spasial interaktif. Tujuan penelitian ini untuk membangun web pemetaan interaktif yang menyajikan visualisasi persebaran pembeli berdasarkan cluster dan grafik jumlah pembelian per wilayah serta menampilkan rekomendasi produk promosi menggunakan algoritma Apriori. Data yang digunakan adalah data transaksi penjualan berjumlah 2005 transaksi dengan atribut meliputi No. Pesanan, jumlah penjualan per wilayah, Kota/Kabupaten, Provinsi, Longitude, Latitude dan Nama Produk. Hasil penelitian ini berhasil menerapkan algoritma K-means dalam pengelompokan wilayah menggunakan 2 evaluasi nilai k yaitu Silhouette Score dan Metode Elbow. Hasil analisis menunjukkan bahwa pengelompokan dengan nilai K=3 menghasilkan Silhouette Score tertinggi dibandingkan nilai lainnya yaitu sebesar 0.623 sehingga menghasilkan 3 kelompok wilayah yaitu Cluster 0 memiliki 12 kota, Cluster 1 memiliki 12 kota dan Cluster 2 memiliki 6 kota. Hasil analisis algoritma Apriori berhasil memberikan rekomendasi produk promosi menggunakan minimum support 10% dan Confidence 50% menghasilkan 7 aturan asosiasi. Penelitian ini juga berhasil membangun website pemetaan yang memiliki tampilan interaktif sehingga memudahkan toko Cindrelia.id dalam melihat potensi wilayah untuk bazar dan produk promosi yang akan dijual. Kontribusi teoretis penelitian ini adalah pengembangan kerangka metodologi integrasi clustering spasial dan association rule mining untuk analisis bisnis retail. Kontribusi praktis meliputi penyediaan sistem pendukung keputusan berbasis web GIS yang dapat diadopsi oleh UMKM fashion lainnya untuk optimalisasi strategi pemasaran dan ekspansi bisnis berbasis data geografis dan pola pembelian konsumen.

Kata Kunci : Algoritma Apriori; Data Mining Retail; K-Means Clustering; Pemetaan Geospasial; Sistem Informasi Geografis; Web GIS Interaktif.

Abstract

The growth of e-commerce in Indonesia has encouraged local fashion businesses to leverage sales data for strategic decision-making. However, many fashion SMEs face challenges in analyzing customer geographic data and product purchasing patterns for promotion optimization and business expansion. Cindrelia.id is a local fashion brand that sells women's clothing products through e-commerce platforms and social media. Currently, the store faces problems in analyzing sales data to obtain useful information, especially regarding the geographical distribution of customers and product recommendations for promotions. This makes it difficult for the store to make decisions in determining the location of promotions, bazaars, and offline stores. The identified research gap is the lack of studies integrating spatial clustering with association rule mining in an interactive web GIS platform to support business decisions in the Indonesian fashion retail sector. Previous research tends to focus on a single method or does not integrate

interactive spatial visualization. The purpose of this study is to build an interactive mapping website that presents a visualization of buyer distribution based on clusters and graphs of the number of purchases per region and displays promotional product recommendations using the Apriori algorithm. The data used consists of 2,005 sales transactions with attributes including Order Number, sales volume per region, City/District, Province, Longitude, Latitude, and Product Name. This study successfully applied the K-means algorithm in regional clustering using two k-value evaluations, namely the Silhouette Score and the Elbow Method. The analysis results show that clustering with a K value of 3 produced the highest Silhouette Score compared to other values, namely 0.623, resulting in three regional clusters: Cluster 0 with 12 cities, Cluster 1 with 12 cities, and Cluster 2 with 6 cities. The results of the Apriori algorithm analysis successfully provided promotional product recommendations using a minimum support of 10% and Confidence of 50%, resulting in 7 association rules. This study also successfully developed a mapping website with an interactive display, making it easier for the Cindrelia.id store to see the potential of areas for bazaars and promotional products to be sold. The theoretical contribution of this research is the development of a methodological framework integrating spatial clustering and association rule mining for retail business analysis. Practical contributions include providing a web GIS-based decision support system that can be adopted by other fashion SMEs for optimizing marketing strategies and data-driven business expansion based on geographic and consumer purchasing patterns.

Keywords: *Apriori Algorithm; Geospatial Mapping; Geographic Information System; Interactive Web GIS; K-Means Clustering; Retail Data Mining*

PENDAHULUAN

Di era digital saat ini, industri retail fashion mengalami transformasi signifikan dengan berkembangnya platform e-commerce yang memungkinkan pelaku usaha menjangkau pasar yang lebih luas (R. S. Handayani et al., 2025). Di Indonesia, pertumbuhan e-commerce telah mencapai 78% pada tahun 2023, dengan sektor fashion menjadi salah satu kategori produk yang paling diminati oleh konsumen online (Mustika et al., 2023). Namun, pertumbuhan ini juga menghadirkan tantangan baru bagi pelaku usaha, khususnya dalam mengelola dan menganalisis data transaksi yang semakin besar dan kompleks. Banyak UMKM fashion lokal yang telah mengumpulkan data penjualan dalam jumlah besar, namun belum mampu mengekstrak informasi strategis dari data tersebut untuk mendukung pengambilan keputusan bisnis, terutama terkait dengan strategi pemasaran wilayah dan rekomendasi produk (Suarna et al., 2025).

Permasalahan ini dialami oleh salah satu brand fashion lokal yang saat ini sedang berkembang yaitu toko online shop Cindrelia.id yang merupakan salah satu toko fashion yang menjual pakaian khusus wanita di e-commerce dan media sosial. Cindrelia.id telah beroperasi sejak tahun 2020 dan melayani pelanggan dari berbagai wilayah di Indonesia melalui platform marketplace seperti Shopee, Tokopedia, dan Instagram. Hingga tahun 2024, toko ini telah mencatat lebih dari 2.000 transaksi penjualan dengan jangkauan wilayah yang tersebar di lebih dari 30 kota/kabupaten di Indonesia. Meskipun volume penjualan terus meningkat, pihak Cindrelia.id menghadapi kendala dalam memahami pola geografis pelanggan dan preferensi pembelian produk yang dapat dijadikan dasar untuk strategi pemasaran yang lebih efektif. Namun dengan banyaknya data penjualan yang ada membuat Cindrelia.id kesulitan untuk memahami informasi yang bermanfaat dari data tersebut. Adapun data transaksi penjualan tersebut hanya dikumpulkan tanpa analisis, khususnya terkait dengan sebaran geografis pelanggan maupun preferensi pembelian pelanggan. Hal ini akhirnya membuat Cindrelia.id kesulitan dalam mengambil

keputusan strategis seperti pemilihan lokasi offline, bazar, dan promosi produk (Kurniawan, 2023).

Fenomena yang dialami oleh Cindrelia.id bukanlah kasus terisolasi, melainkan mencerminkan tantangan umum yang dihadapi oleh banyak pelaku UMKM fashion di Indonesia dalam era data-driven marketing. Studi terbaru menunjukkan bahwa hanya 35% UMKM di Indonesia yang memanfaatkan data penjualan untuk analisis strategis (Sianturi, 2020), sementara 65% lainnya masih mengandalkan intuisi dan pengalaman pribadi dalam pengambilan keputusan bisnis (Septiawan & Sunandes, 2024). Dalam konteks global, penelitian oleh Nanda (2025) mengungkapkan bahwa perusahaan retail yang mengimplementasikan data mining dan geospatial analytics untuk segmentasi pasar mengalami peningkatan efisiensi promosi hingga 40% dan peningkatan ROI hingga 35%. Hal ini menunjukkan bahwa pemanfaatan teknologi analitik berbasis data memiliki potensi besar untuk meningkatkan daya saing UMKM di sektor retail fashion (Amira & Nasution, 2024; Yunitasari et al., 2025).

Salah satu solusi yang dapat digunakan untuk mengatasi permasalahan diatas adalah dengan melakukan pemetaan dan pengelompokan wilayah berdasarkan data pembelian menggunakan metode data mining dengan teknik clustering yaitu K-means (Handoko et al., 2020; S. Sari & Utamajaya, 2022). K-means adalah algoritma clustering yang mengelompokkan data ke dalam cluster berdasarkan kemiripan jarak dan dikenal mampu menangani data berskala besar dengan cepat (P. Sari, 2024). Dengan menerapkan metode ini, data pembelian akan dikelompokkan berdasarkan lokasi pelanggan dan jumlah pembelian sehingga dapat mengidentifikasi wilayah dengan transaksi tinggi, sedang, dan rendah. Selain itu, untuk meningkatkan penjualan dibutuhkan strategi rekomendasi produk yang efektif (Silaen et al., 2024). Salah satu pendekatan yang dapat digunakan adalah teknik asosiasi dalam data mining dengan metode Apriori (Barkah et al., 2020). Metode Apriori adalah teknik untuk menemukan hubungan antar produk dalam data, khususnya mengidentifikasi kombinasi produk yang sering dibeli bersamaan yang dapat dijadikan sebagai produk promosi atau bundling untuk meningkatkan penjualan (Butsianto et al., 2025).

Penelitian terdahulu mengenai penerapan clustering dan association rules mining dalam retail telah banyak dilakukan, namun masih terdapat gap penelitian yang signifikan. Secara global, penelitian oleh Warianti (2024) menggunakan K-means clustering untuk segmentasi pelanggan e-commerce di China dan berhasil meningkatkan akurasi targeting promosi hingga 32%. Namun, penelitian tersebut tidak mengintegrasikan analisis geografis dalam proses clustering. Sementara itu, studi oleh Gunawan (2024) menerapkan algoritma Apriori untuk rekomendasi produk pada platform e-commerce di India, namun tidak mempertimbangkan faktor geografis dalam analisis pola pembelian. Di Indonesia sendiri, beberapa penelitian telah mencoba menerapkan metode data mining dalam konteks retail, namun masih terbatas pada analisis parsial dan belum mengintegrasikan pendekatan clustering geografis dengan association rules mining dalam satu sistem yang komprehensif.

Beberapa penelitian sebelumnya yang berkaitan telah dilakukan di Indonesia,

namun masih memiliki keterbatasan dalam hal integrasi metode dan aplikasi praktis. Penelitian yang dilakukan oleh Handayani (2025), yakni mengelompokkan wilayah pelanggan toko Roni berdasarkan tingkat intensitas pembelian menggunakan algoritma K-means. Hasil penelitian menunjukkan bahwa K-means berhasil mengidentifikasi wilayah dengan aktivitas pembelian tertinggi yang dapat dijadikan sebagai rekomendasi sasaran wilayah promosi. Namun, penelitian ini hanya berfokus pada clustering wilayah tanpa mengintegrasikan analisis rekomendasi produk, sehingga belum memberikan solusi holistik untuk strategi pemasaran retail. Penelitian oleh Syalwa (2023) membantu perusahaan cleaning service di Jawa Timur mengidentifikasi wilayah potensial untuk promosi jasa menggunakan metode K-means. Penelitian berhasil mengelompokkan wilayah menjadi tiga kelompok yaitu kurang, cukup berpotensi, dan sangat berpotensi. Meskipun penelitian ini memberikan kontribusi dalam analisis wilayah, namun tidak mengembangkan sistem visualisasi berbasis GIS yang dapat diakses secara real-time oleh pengambil keputusan. Terakhir, penelitian oleh Akbar (2023) merancang sistem rekomendasi bundling produk promosi untuk Toko Remaja menggunakan metode Apriori. Penelitian berhasil membuat sistem rekomendasi bundling produk yang efektif membantu pemilihan produk bundling secara efisien. Namun, sistem yang dikembangkan tidak mempertimbangkan aspek geografis dalam rekomendasi produk, padahal preferensi produk dapat bervariasi antar wilayah.

Gap penelitian yang teridentifikasi dari kajian literatur di atas adalah: (1) belum adanya penelitian yang mengintegrasikan metode clustering geografis dengan association rules mining dalam satu framework analitik untuk retail fashion di Indonesia; (2) minimnya pengembangan sistem berbasis web GIS yang dapat memvisualisasikan hasil analisis clustering dan rekomendasi produk secara interaktif dan real-time; (3) kurangnya studi empiris yang menerapkan metode data mining untuk mendukung pengambilan keputusan strategis pada UMKM fashion lokal di Indonesia; serta (4) belum adanya penelitian yang secara spesifik menganalisis pola pembelian produk fashion berdasarkan segmentasi geografis untuk tujuan optimasi promosi wilayah dan bundling produk.

Kebaruan penelitian ini terletak pada beberapa aspek utama. Namun perancangan ini berbeda dengan penelitian sebelumnya karena penelitian ini mengintegrasikan dua metode data mining (K-means untuk clustering geografis dan Apriori untuk association rules mining) dalam satu sistem web GIS yang interaktif, sehingga memberikan solusi komprehensif untuk analisis wilayah dan rekomendasi produk secara simultan. Kedua, penelitian ini mengembangkan platform web-based decision support system yang memungkinkan pelaku UMKM untuk mengakses hasil analisis secara visual dan interaktif tanpa memerlukan keahlian teknis dalam data mining. Ketiga, penelitian ini memberikan kontribusi empiris dengan menerapkan framework analitik yang dikembangkan pada kasus nyata UMKM fashion lokal di Indonesia, sehingga hasilnya dapat langsung diimplementasikan dan memberikan nilai praktis. Keempat, penelitian ini tidak hanya menghasilkan output berupa cluster wilayah dan aturan asosiasi produk, tetapi juga mengintegrasikan keduanya dalam satu dashboard yang memudahkan pengambil keputusan untuk merumuskan strategi pemasaran yang lebih efektif dan

efisien. Penelitian ini akan melakukan perancangan web dengan 2 fitur utama yaitu fitur penerapan K-means untuk pemetaan wilayah pelanggan dan fitur penerapan Apriori untuk rekomendasi bundling produk promosi.

Berdasarkan uraian diatas maka penelitian ini bertujuan untuk mengembangkan sistem pemetaan kelompok wilayah penjualan produk fashion Cindrelia.id untuk meningkatkan penjualan dengan metode K-means clustering dan Apriori. Secara spesifik, tujuan penelitian ini meliputi: (1) menerapkan algoritma K-means clustering untuk mengelompokkan wilayah penjualan berdasarkan volume transaksi dan karakteristik geografis; (2) menerapkan algoritma Apriori untuk menghasilkan aturan asosiasi produk yang dapat dijadikan dasar rekomendasi bundling promosi; (3) membangun platform web GIS interaktif yang mengintegrasikan hasil clustering wilayah dan rekomendasi produk dalam satu dashboard; serta (4) mengevaluasi efektivitas sistem yang dikembangkan dalam mendukung pengambilan keputusan strategis toko Cindrelia.id. Hasil pemetaan ini diharapkan dapat membantu Cindrelia.id untuk menemukan wilayah strategis dan rekomendasi produk promosi untuk meningkatkan penjualan. Manfaat teoretis dari penelitian ini adalah memperkaya literatur mengenai penerapan data mining dalam retail fashion di Indonesia, khususnya dalam konteks integrasi clustering geografis dan association rules mining. Secara praktis, penelitian ini diharapkan dapat memberikan solusi konkret bagi UMKM fashion dalam mengoptimalkan strategi pemasaran berbasis data, serta menjadi referensi bagi penelitian selanjutnya dalam pengembangan decision support system untuk sektor retail.

METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan data mining research yang bertujuan mengembangkan sistem pendukung keputusan berbasis web GIS untuk analisis penjualan retail fashion. Penelitian dilaksanakan di Cindrelia.id, sebuah toko online fashion wanita yang beroperasi di Indonesia melalui platform e-commerce (Shopee, Tokopedia) dan media sosial (Instagram). Periode pengumpulan data dilakukan dari Januari 2020 hingga Desember 2024.

Populasi penelitian ini adalah seluruh transaksi penjualan Cindrelia.id, sedangkan sampel yang digunakan adalah 2.005 transaksi penjualan yang tercatat lengkap dengan informasi geografis. Teknik pengambilan sampel menggunakan purposive sampling dengan kriteria: (1) transaksi memiliki data lengkap (nomor pesanan, nama produk, lokasi pembeli, koordinat geografis); (2) transaksi telah selesai dan terverifikasi; (3) data geografis dapat divalidasi melalui sistem koordinat. Sampel ini dianggap representatif karena mencakup lebih dari 90% total transaksi dalam periode penelitian dan mewakili sebaran geografis pelanggan di 30+ kota/kabupaten di Indonesia.

Teknik pengumpulan data dilakukan melalui dokumentasi dan ekstraksi data sekunder dari sistem manajemen penjualan Cindrelia.id. Data yang dikumpulkan meliputi: (1) data transaksional: nomor pesanan, tanggal transaksi, nama produk, jumlah pembelian, harga; (2) data geografis: kota/kabupaten, provinsi, koordinat geografis (latitude dan longitude); (3) data produk: kategori, varian, dan frekuensi pembelian. Data koordinat geografis diperoleh dengan melakukan geocoding berdasarkan nama kota/kabupaten menggunakan dataset OpenStreetMap dan divalidasi melalui shapefile batas administratif Indonesia dari Badan Informasi Geospasial (BIG).

Teknik analisis data dalam penelitian ini menggunakan pendekatan sekuensial yang terdiri dari empat tahap utama:

1. Preprocessing Data:

- a. Pembersihan data (data cleaning): identifikasi dan penanganan missing values, duplikasi, dan outliers
- b. Transformasi data: konversi format data, penggabungan tabel transaksi dengan data geografis
- c. Normalisasi data numerik menggunakan Min-Max Normalization untuk menyamakan skala atribut dalam clustering
- d. Transformasi data transaksional menjadi format tabulasi biner untuk algoritma Apriori

2. Pemodelan K-Means Clustering:

- a. Input: data agregat penjualan per wilayah (kota/kabupaten) dengan atribut jumlah transaksi, total penjualan, dan koordinat geografis
- b. Proses: penerapan algoritma K-means dengan variasi nilai K (2-10) untuk membentuk cluster wilayah
- c. Evaluasi model menggunakan dua metode: (a) Elbow Method untuk identifikasi titik optimal Within-Cluster Sum of Squares (WCSS); (b) Silhouette Score untuk mengukur kualitas clustering dengan rentang nilai -1 hingga 1
- d. Output: pembagian wilayah ke dalam cluster berdasarkan intensitas pembelian (tinggi, sedang, rendah)

3. Pemodelan Apriori Association Rule Mining:

- a. Input: data transaksional dalam format market basket (setiap baris = 1 transaksi, setiap kolom = 1 produk)
- b. Proses: penerapan algoritma Apriori untuk menemukan frequent itemsets dan association rules
- c. Parameter: minimum support = 10% (produk muncul minimal 200 kali dari 2.005 transaksi); minimum confidence = 50% (tingkat keyakinan hubungan antar produk)
- d. Evaluasi menggunakan metrik lift ratio untuk mengukur kekuatan asosiasi antar produk
- e. Output: aturan asosiasi produk dalam format "jika membeli produk A, maka kemungkinan membeli produk B"

4. Pengembangan Web GIS Interaktif:

- a. Integrasi hasil clustering dan association rules dalam satu platform
- b. Visualisasi peta interaktif menggunakan library Folium dan Leaflet.js
- c. Pengembangan dashboard untuk menampilkan statistik penjualan per wilayah dan rekomendasi produk

Kerangka metodologi penelitian ini mengadaptasi pendekatan Knowledge Discovery in Databases (KDD) yang terdiri dari tahapan: data selection → preprocessing → transformation → data mining → interpretation/evaluation (Fayyad et al., 1996). Untuk pengembangan sistem, penelitian ini menggunakan model Software Development Life Cycle (SDLC) Waterfall yang terdiri dari lima tahap berurutan:

Model SDLC Waterfall dikenal sebagai metode pengembangan perangkat lunak yang berurutan. Metode ini dipilih karena pendekatannya yang terstruktur dan menghasilkan dokumentasi yang jelas sehingga mempermudah proses pengembangan

sistem nantinya.



Gambar 1 Metode Waterfall

Berikut adalah penjelasannya:

Requirement Analysis

Pada tahapan ini dilakukan identifikasi fungsionalitas yang diperlukan dalam perancangan sistem. Adapun tujuan tahapan ini yaitu mengetahui dengan jelas kebutuhan pengguna dan mendokumentasikan spesifikasi fungsional yang akan menjadi dasar dalam desain dan pengembangan sistem.

Design

Pada tahapan design ini akan dirancang alur sistem untuk menentukan cara interaksi antara pengguna dan aplikasi, sehingga aplikasi tersebut dapat menyelesaikan masalah yang telah diidentifikasi.

Development

Setelah tahapan perancangan selesai, maka dilakukan tahap implementasi untuk membangun sistem berdasarkan desain yang telah dirancang. Hasil dari tahapan ini adalah sebuah sistem yang atau program yang dapat dipergunakan.

Testing

Pada tahap ini dilakukan evaluasi sistem untuk menilai apakah sistem yang dikembangkan memenuhi kebutuhan pengguna yang telah ditetapkan atau belum. Adapun jika ditemukan kekurangan atau masalah maka akan melakukan perbaikan sebelum sistem diterapkan secara penuh.

HASIL DAN PEMBAHASAN

Preprocessing

Preprocessing K-Means

Pada tahap preprocessing dilakukan penyiapan data transaksi penjualan menjadi data yang siap digunakan dalam proses pemodelan. Pada proses preprocessing K-means data masukan berasal dari transaksi penjualan produk fashion pada toko online Cindrelia.id yang berisi atribut seperti No. Pesanan, Nama Produk, total penjualan,

jumlah pesanan, Kota/Kabupaten, Provinsi, Harga Produk, dan Jumlah Pembelian. Preprocessing pada penelitian ini dilakukan dengan mengecek apakah data kosong yang dapat mempengaruhi hasil evaluasi pengelompokan. Selain itu dilakukan proses normalisasi data menggunakan metode Min-Max Normalization untuk menyamakan skala nilai pada setiap atribut sehingga tidak terdapat atribut yang memiliki pengaruh lebih besar dari atribut yang lain dalam proses pengelompokan. Hasil dari proses ini adalah data yang telah dinormalisasi dan siap digunakan pada tahap pengelompokan menggunakan metode K-Means.

Preprocessing Apriori

Tahapan preprocessing data pada algoritma Apriori dilakukan dengan melakukan perubahan data transaksi menjadi data tabulasi. Pada proses ini menggunakan data transaksi dari Cindrelia.id yang berisi daftar pesanan pelanggan. Data tersebut kemudian diubah sehingga setiap baris menunjukkan satu transaksi dan setiap kolom menunjukkan nama produk yang dibeli dalam transaksi tersebut. Nilai pada setiap kolom menunjukkan angka biner yaitu 1 dan 0. Angka 1 menunjukkan bahwa produk tersebut dibeli dalam transaksi sedangkan angka 0 menunjukkan bahwa produk tersebut tidak dibeli

```
def preprocess_apriori_data(data_df):
    """Preprocess apriori data for analysis"""
    # Set first column as transaction ID
    data_df = data_df.copy()
    data_df.iloc[:, 0] = data_df.iloc[:, 0].astype(str)

    # Convert to binary (1 for purchase, 0 for no purchase)
    # Skip first column (transaction ID)
    product_columns = data_df.columns[1:]

    # Validate and convert product columns
    for col in product_columns:
        try:
            # Convert to numeric, fill NaN with 0
            data_df[col] = pd.to_numeric(data_df[col], errors='coerce').fillna(0)
            # Convert to binary (1 if > 0, 0 otherwise)
            data_df[col] = (data_df[col] > 0).astype(int)
        except Exception as e:
            st.error(f"❌ Error memproses kolom '{col}': {str(e)}")
    return data_df
```

Gambar 2 Preprocessing Data Tabulasi

Pemodelan

Pemodelan K-Means

Proses penerapan algoritma K-means yang digunakan untuk mengelompokkan data wilayah penjualan produk, diawali dengan memilih atribut tertentu dalam proses pemodelan. Selanjutnya, admin menentukan nilai k yang diinginkan. Setelah itu, algoritma K-means menghitung jarak setiap data terhadap centroid dan menentukan jumlah kelompok data berdasarkan jarak terdekat sesuai dengan nilai k. Kode program dalam proses pemodelan K-means dapat dilihat pada Gambar 3

```
## Lakukan clustering ##
kmeans = KMeans(n_clusters=n_clusters, random_state=42, n_init=10)
cluster_labels = kmeans.fit_predict(scaled_data)

# Tambahkan hasil cluster ke data
city_metrics['Cluster'] = cluster_labels
```

Gambar 3 Pemodelan K-Means

Gambar 3 menunjukkan kode program penerapan metode K-means pada data yang sudah siap dimodelkan. Fungsi kode diatas digunakan untuk membuat model K-means dengan jumlah kelompok yang telah sesuai ditentukan pada menu pemodelan K-means. Selanjutnya, K-means menjalankan proses pengelompokan dan memberikan label cluster untuk setiap data. Setelah itu, sistem menambahkan hasil pengelompokan tersebut ke dalam tabel data sehingga setiap kota/kabupaten memiliki informasi cluster.

```
# Tambahkan hasil cluster ke data
city_metrics['Cluster'] = cluster_labels

# Match cities with coordinates using utility function
merged_data = match_cities_with_coordinates(city_metrics, cities_data)

# Show matching status
matched_count = merged_data['latitude'].notna().sum()
unmatched_count = merged_data['latitude'].isna().sum()
```

Gambar 4 Kode Program Pencocokan Data Koordinat

Gambar 4 menunjukkan kode program untuk menghubungkan hasil pengelompokan wilayah dengan data koordinat geografis kota/kabupaten menggunakan fungsi `match_cities_with_coordinates()`. Fungsi ini digunakan untuk mencocokkan nama kota dari dataset penjualan dengan data koordinat yaitu latitude dan longitude yang berasal dari file shapefile (.shp). File shapefile memiliki informasi mengenai batas administratif wilayah Indonesia sehingga hasil pengelompokan dapat ditampilkan secara spasial di peta. Kode program untuk visualisasi peta dapat dilihat pada Gambar 5

```
with col:
    st.subheader("🗺️ Peta Clustering")

    # Create and display map (prefer GeoJSON centroids for marker positions)
    geo_centroids = load_geojson_centroids()
    cluster_map = create_cluster_map(merged_data, cities_data, geo_centroids)

    # Use larger container to prevent layer control from being cut off
    st_folium(cluster_map, width=900, height=700, use_container_width=True)
```

Gambar 5 Kode Program Visualisasi Peta

Gambar 5 menunjukkan kode program untuk visualisasi hasil pengelompokan wilayah menggunakan Leaflet Map melalui pustaka Folium di Streamlit. Leaflet digunakan untuk menampilkan peta interaktif dengan setiap kota ditampilkan sebagai titik dengan warna yang berbeda-beda berdasarkan hasil pengelompokan K-means.

Pemodelan Apriori

Pada tahap ini dilakukan pemodelan Apriori untuk menemukan pola asosiasi antar produk yang sering dibeli bersamaan untuk rekomendasi produk promosi. Prosesnya diawali dengan penentuan minimum support pada menu apriori yang menunjukkan frekuensi produk dan kombinasi produk yang muncul dalam seluruh transaksi. Kode program dalam penentuan minimum support yaitu dapat dilihat pada Gambar 6:

```
min_support = st.slider(
    "Minimum Support:",
    min_value=0.01,
    max_value=0.5,
    value=min(0.1, recommended_min_support * 2), # Use recommended or 0.1, whichever is lower
    step=0.01,
    help="Minimum support untuk frequent itemsets. Direkomendasikan: {recommended_min_support:.3%}"
)
st.write(f"***Support:*** (min_support: .2%)")
st.write(f"***Minimal transaksi:*** (min_support * total_transactions: .0f)")

if min_support * total_transactions < 2:
    st.warning("⚠ Support terlalu tinggi! Minimal 2 transaksi diperlukan")
```

Gambar 6 Kode Program Penentuan Minimum Support

Gambar 6 menunjukkan kode program support untuk menghitung jumlah total transaksi dan menentukan batas minimum jumlah transaksi yang memenuhi nilai support yang ditentukan pengguna. Kode program diatas memberikan peringatan apabila nilai support yang ditentukan pengguna terlalu tinggi sehingga tidak ada itemset yang memenuhi minimum support. Selanjutnya, pengguna juga dapat menentukan nilai Confidence sebagai evaluasi seberapa besar hubungan antara satu produk dengan produk lainnya. Kode program dalam penentuan minimum Confidence ini dapat dilihat pada Gambar 7.

```
with col_param2:
    min_confidence = st.slider(
        "Minimum Confidence:",
        min_value=0.1,
        max_value=1.0,
        value=0.5,
        step=0.05,
        help="Minimum confidence untuk association rules"
    )
    st.write(f"***Confidence:*** (min_confidence: .2%)")
```

Gambar 7 Kode Program Penentuan Minimum Confidence

Gambar 7 menunjukkan kode program untuk menentukan batas kepercayaan atau Confidence pada proses pembentukan aturan asosiasi dalam algoritma Apriori. Pada kode diatas, Confidence digunakan untuk menentukan nilai batas bawah Confidence yang berarti hanya aturan dengan nilai Confidence sama atau lebih besar dari nilai minimum yang ditentukan yang akan ditampilkan dalam hasil analisis. Maka setelah proses penentuan nilai minimum support dan minimum Confidence selanjutnya dilakukan proses pemodelan apriori. Kode program dalam proses apriori dapat dilihat pada Gambar 8.

```
# Generate frequent itemsets
with st.spinner("Menghitung frequent itemsets..."):
    frequent_itemsets = generate_frequent_itemsets(processed_data, min_support)

# Debug info
st.info(f"Data untuk analisis: {len(processed_data)} transaksi, {len(processed_data.columns)-1} produk")
st.info(f"Min Support: {min_support: .2%} ({min_support * len(processed_data): .0f} transaksi min)")

if frequent_itemsets.empty:
    st.warning("⚠ Tidak ada frequent itemsets yang ditemukan!")
    st.warning("⚠ Coba turunkan nilai Min Support atau periksa format data")

# Show data sample for debugging
with st.expander("🔍 Debug: Sample Data"):
    st.write(f"Data yang diproses (5 baris pertama):")
    st.dataframe(processed_data.head(), use_container_width=True)

    st.write(f"Statistik data:")
    st.write(f"Total transaksi: {len(processed_data)}")
    st.write(f"Total produk: {len(processed_data.columns)-1}")
    st.write(f"Transaksi dengan pembelian: {(processed_data.iloc[:, 1:] > 0).any(axis=1).sum()}")
    st.write(f"Produk yang dibeli: {(processed_data.iloc[:, 1:] > 0).any(axis=0).sum()}")

# Generate association rules only if we have frequent itemsets
if not frequent_itemsets.empty:
    with st.spinner("Menghitung association rules..."):
        rules = generate_association_rules(frequent_itemsets, min_confidence)
```

Gambar 8 Kode Program Pemodelan Apriori

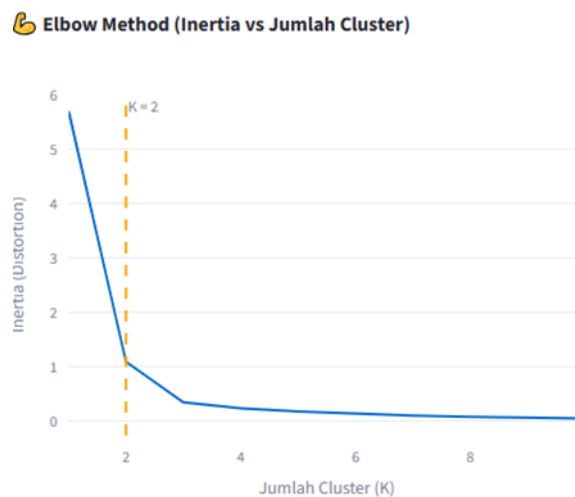
Gambar 8 menunjukkan kode program proses pemodelan yang diawali dengan menghitung frequent itemsets dengan fungsi `generate_frequent_itemsets` untuk mencari kombinasi produk berdasarkan nilai minimum support yang telah ditentukan sebelumnya. Setelah itu jika kombinasi produk yang memenuhi syarat ditemukan maka sistem akan membentuk association rules melalui fungsi `generate_association_rules` dengan evaluasi menggunakan nilai minimum Confidence yang telah ditentukan pengguna sebelumnya juga. Hasil dari proses ini menampilkan aturan seperti “jika membeli produk A maka kemungkinan akan membeli produk B” yang dapat dipergunakan untuk mendukung strategi rekomendasi produk promosi pada Cindrelia.id.

Evaluasi

Evaluasi K-Means

Pengujian metode K-means dilakukan untuk mengetahui hasil pengelompokan data transaksi pembelian produk fashion di Cindrelia.id. Tujuan utama pengujian ini adalah untuk menemukan pola perilaku wilayah pelanggan serta mengidentifikasi kelompok wilayah yang memiliki karakteristik pembelian serupa. Pada proses pemodelan K-means penelitian ini menggunakan 2 pilihan metode evaluasi yaitu metode elbow dan metode Silhouette Score. Proses evaluasi ini menggunakan bahasa pemrograman python.

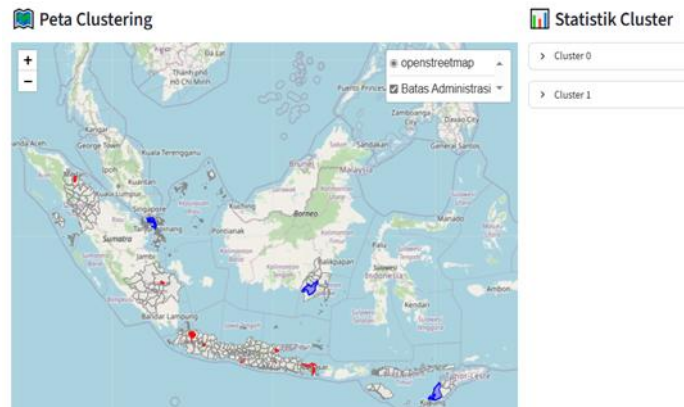
Dalam perancangan web pemetaan ini nilai K dapat ditentukan menggunakan metode Elbow Method. Cara kerja dari metode elbow adalah dengan melihat titik “siku” pada grafik Within-Cluster Sum of Squares yang biasanya membentuk grafik seperti siku. Titik siku tersebut dapat disebut titik Elbow dan menunjukkan jumlah cluster terbaik. Hasil penerapan pemilihan nilai k menggunakan metode elbow dapat dilihat pada Gambar 9.



Gambar 9 Hasil Pemilihan Nilai K Dengan Metode Elbow

Gambar 9 menunjukkan hasil penggunaan metode Elbow untuk menentukan jumlah cluster (K) yang optimal dalam web pemetaan. Pada grafik diatas memberikan hasil bahwa titik "siku" terlihat pada K=2. Maka dari hal tersebut web pemetaan menggunakan nilai K=2 untuk melakukan pemodelan K-means. Hasil dari pemetaan menggunakan

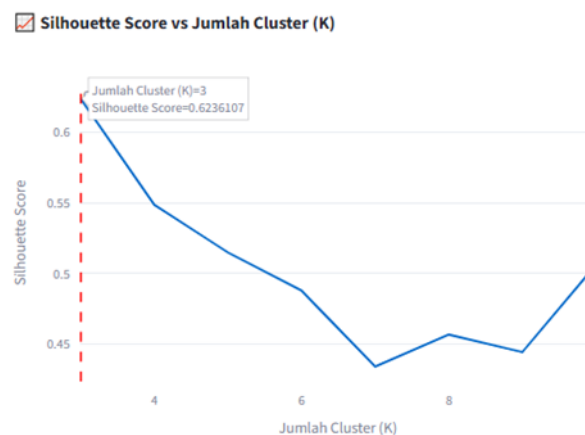
metode elbow untuk pemilihan nilai K dapat dilihat pada Gambar 10:



Gambar 10 Visualisasi Hasil Pemilihan Nilai K dengan Elbow

Gambar 10 menunjukkan hasil visualisasi pengelompokan menggunakan nilai $K=2$ berdasarkan metode Elbow. Hasil pengelompokan menunjukkan 2 kelompok yang terbentuk yaitu cluster 0 dan cluster 1. Cluster 0 memiliki 24 wilayah sedangkan cluster 1 memiliki 6 wilayah.

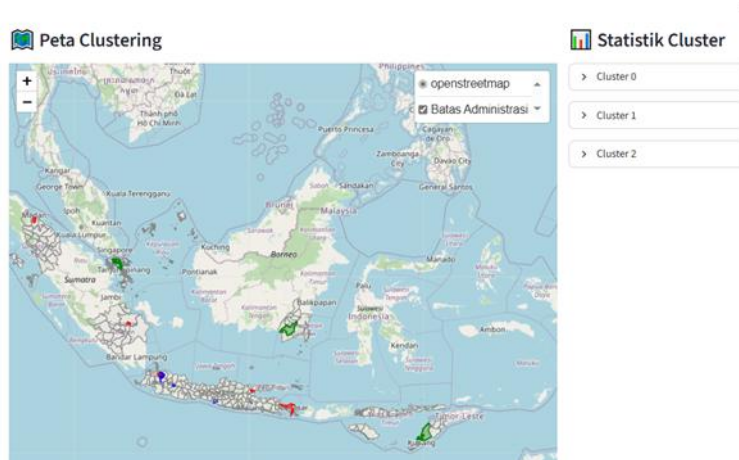
Pada tahap Evaluasi Silhouette Score pada web pemetaan ini maka penentuan nilai K optimal dilakukan dengan mengukur tingkat kelayakan data dikelompokkan pada setiap percobaan nilai K . Silhouette Score menunjukkan tingkat pengelompokan antarcluster dengan nilai antara -1 sampai 1. Apabila nilai Silhouette Score semakin mendekati 1 maka semakin baik hasil pengelompokan karena data dalam satu cluster lebih mirip dan berbeda jelas dengan cluster lain. Hasil dari pemilihan nilai k optimal menggunakan evaluasi Silhouette Score dapat dilihat pada Gambar 11:



Gambar 11 Hasil Pemilihan Nilai K Dengan Metode Silhouette Score

Gambar 11 menunjukkan hasil pengelompokan menggunakan pemilihan K dengan metode Silhouette Score yang memberikan hasil bahwa $K=3$ sebagai jumlah cluster optimal dengan nilai Silhouette Score sebesar 0.623. Hasil pemetaan wilayah

menggunakan nilai K berdasarkan metode Silhouette Score dapat dilihat pada Gambar 12



Gambar 12 Hasil Pemilihan Nilai K Dengan Metode Silhouette Score

Gambar 12 menunjukkan visualisasi hasil pemetaan menggunakan nilai K=3 dari metode Silhouette Score. Peta diatas menunjukkan 3 kelompok besar yang terbentuk yaitu cluster 0, 1 dan 2. Cluster 0 memiliki 12 kota, Cluster 1 memiliki 12 kota dan Cluster 2 memiliki 6 kota. Hasil evaluasi Silhouette Score berdasarkan nilai k yaitu:

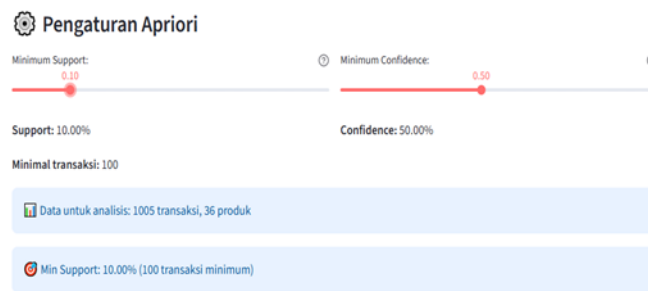
Tabel 1 Hasil Evaluasi Silhouette Score

Nilai k	<i>Silhouette Score</i>
3	0,623
4	0,5485
5	0,5148
6	0,4879
7	0,4341
8	0,4567
9	0,4443
10	0,5063

Berdasarkan Tabel 1 hasil evaluasi Silhouette Score menunjukkan bahwa nilai k = 3 memiliki skor tertinggi yaitu sebesar 0,623 dibandingkan dengan nilai k lainnya. Hal ini menunjukkan bahwa pengelompokan dengan tiga cluster menghasilkan tingkat keseragaman dan pemisahan antar kelompok yang paling baik.

Evaluasi Apriori

Setelah dilakukan pengelompokan wilayah menggunakan K-means, tahap selanjutnya adalah pengujian metode Apriori. Pengujian ini dilakukan untuk menemukan asosiasi antar produk yang sering dibeli secara bersamaan sebagai dasar rekomendasi promosi di Cindrelia.id.



Gambar 13 Pengaturan Apriori

Gambar 13 menunjukkan pengaturan algoritma Apriori menggunakan minimum support 10% dan minimum Confidence 50%. Pengaturan tersebut digunakan untuk menyaring aturan asosiasi sehingga hanya transaksi dengan support minimal 100 kali pembelian produk dan setiap aturan yang dihasilkan memiliki tingkat Confidence minimal 50%.

Association Rules

	Antecedent	Consequent	Support	↑ Confidence	Lift
0	BIEL TOP	ALICIE DRESS	0.262	0.886	2.023
6	ALICIE DRESS, TAVI SKIRT	ROSIE TOP	0.110	0.881	2.767
4	TAVI SKIRT	ROSIE TOP	0.255	0.856	2.689
2	CHEL DRESS	ALICIE DRESS	0.215	0.831	1.898
3	ROSIE TOP	TAVI SKIRT	0.255	0.800	2.689
5	ROSIE TOP, ALICIE DRESS	TAVI SKIRT	0.110	0.799	2.684
1	ALICIE DRESS	BIEL TOP	0.262	0.598	2.023

Gambar 14 Hasil Rekomendasi Produk dan Evaluasi

Gambar 14 menunjukkan hasil aturan asosiasi dengan kolom Antecedent yaitu produk yang dibeli sedangkan Consequent adalah produk penyerta yang dibeli bersamaan dengan produk yang dibeli pada kolom Antecedent. Confidence menunjukkan tingkat interval waktu produk dibeli secara bersamaan misalnya Confidence 0.886 pada "BIEL TOP → ALICIE DRESS" yang menunjukkan bahwa 88.6% pelanggan yang membeli BIEL TOP juga membeli ALICIE DRESS. Sementara Lift menunjukkan kekuatan hubungan misalnya nilai Lift 2.023 pada "BIEL TOP → ALICIE DRESS" menunjukkan bahwa pelanggan yang membeli BIEL TOP memberikan kemungkinan 2 kali lebih besar juga akan membeli ALICIE DRESS. Pada penelitian ini terdapat 7 aturan yang terbentuk dengan nilai Confidence tertinggi sebesar 88,5% pada aturan.

KESIMPULAN

Berdasarkan hasil perancangan, implementasi dan pengujian sistem pemetaan pengelompokan pembelian produk fashion Cindrelia.id menggunakan metode K-means dan Apriori dapat disimpulkan bahwa sistem yang dibangun telah mampu membantu pihak Cindrelia.id dalam menganalisis data penjualan secara lebih cepat dan mudah.

Penerapan algoritma K-means pada data transaksi berhasil memetakan persebaran wilayah pembeli ke dalam beberapa kelompok dengan tingkat pembelian yang berbeda yaitu tinggi, sedang, dan rendah. Berdasarkan hasil evaluasi dengan metode Silhouette Score nilai K yang paling optimal adalah tiga cluster dengan skor sebesar 0,623, yang menunjukkan kualitas pengelompokan cukup baik dan mampu menggambarkan pola pembelian berdasarkan wilayah secara jelas. Selain itu penerapan algoritma Apriori juga berhasil menemukan hubungan antar produk yang sering dibeli secara bersamaan. Dengan menggunakan nilai minimum support sebesar 10% dan Confidence sebesar 50%, sistem berhasil menghasilkan tujuh aturan asosiasi yang dapat dijadikan dasar dalam menentukan rekomendasi produk promosi atau bundling.

DAFTAR PUSTAKA

- Akbar, M., Wuriyanto, T., & Lusiani, T. (2023). *Penerapan Algoritma Apriori untuk Rekomendasi Bundling Produk Pada Toko Remaja*.
- Amira, B., & Nasution, M. I. P. (2024). Peluang pemanfaatan teknologi big data pada UMKM di Indonesia untuk meningkatkan kinerja perdagangan di era industri 4.0. *Jurnal Ilmiah Nusantara*, 1(4), 695–704.
- Barkah, N., Sutinah, E., & Agustina, N. (2020). Metode Asosiasi Data Mining Untuk Analisa Persediaan Fiber Optik Menggunakan Algoritma Apriori. *Jurnal Kajian Ilmiah*, 20(3), 237–248.
- Butsianto, S., Naya, C., & Rifa'i, A. M. (2025). Implementasi Algoritma Apriori dalam Menemukan Pola Asosiasi pada Data Penjualan Produk Retail. *Bulletin of Computer Science Research*, 5(5), 938–947.
- Gunawan, C., Susanti, W., & Duha, Y. (2024). Sistem rekomendasi produk e-commerce menggunakan algoritma apriori. *Jurnal Mahasiswa Aplikasi Teknologi Komputer Dan Informasi (JMApTeKsi)*, 6(2), 84–90.
- Handayani, E., & Hendrastuty, N. (2025). Klasterisasi Wilayah Penjualan Pada Toko Roni Menggunakan K-Means. *Informasi (Jurnal Informatika Dan Sistem Informasi)*, 17(1), 163–180.
- Handayani, R. S., Diva, G. A., Hermansyah, E. N., & Siswati, A. (2025). Transformasi Digital E-Commerce di Era Post-Pandemic: Analisis Tren dan Dampak pada Perilaku Konsumen Indonesia. *Jurnal Informatika Dan Kesehatan*, 2(2), 108–117.
- Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering. *J. Ilm. Teknol. Dan Rekayasa*, 25(1), 76–88.
- Kurniawan, L. (2023). Penerapan Pengambilan Keputusan Strategis Pada Pesantren Udara 14.2690 MHz. *Tanzhim: Jurnal Dakwah Terprogram*, 1(1), 81–100.
- Mustika, W., Kurniawati, M., & Sari, M. P. (2023). Pengaruh motivasi belanja hedonis terhadap perilaku pembelian impulsif kategori produk fashion online pada marketplace. *Primanomics: Jurnal Ekonomi & Bisnis*, 21(1), 141–151.
- Nanda, S., Shafira, H., & Zahira, T. (2025). Implementasi Data Mining di Sektor E-Commerce. *Jurnal Ilmu Komputer (JILKOMP)*, 1(1), 13–16.
- Sari, P. (2024). Algoritma K-Means Clustering: Sebuah Studi Literatur. *Jurnal Informatika*, 1(01), 1–7.
- Sari, S., & Utamajaya, J. N. (2022). Sistem Pendukung Keputusan Penerima Bantuan

- Langsung Tunai Dana Desa Menggunakan Metode Algoritma K-Means Clustering. *JUPITER: Jurnal Penelitian Ilmu Dan Teknologi Komputer*, 14(1), 150–160.
- Septiawan, B., & Sunandes, A. (2024). Implementasi Intuisi dalam Manajemen Bisnis (Sebuah Studi Literatur). *Jurnal Penelitian Manajemen Terapan (PENATARAN)*, 9(2), 199–209.
- Sianturi, R. D. (2020). Manajemen pemasaran menggunakan analisis SWOT pada UMKM guna meningkatkan daya saing UMKM. *Journal of Business and Economics Research (JBE)*, 1(1), 45–50.
- Silaen, K., Bachtar, L. E., Silaen, M., & Parhusip, A. (2024). Strategi Pemasaran Digital Untuk Meningkatkan Penjualan Produk Lokal Diera Online. *Jurnal Pengabdian Kolaborasi Dan Inovasi IPTEKS*, 2(3), 1072–1079.
- Suarna, I. F., Himawan, R. A., & Rezqia, A. P. (2025). Peranan Penting Sosial Media Pada Penjualan Umkm Fashion Brand 3second Di Kota Bandung. *Jurnal Integrasi Pengetahuan Disiplin*, 6(1).
- Syalwa, R. V., Nurcahyawati, V., & Wuriyanto, T. (2023). Penerapan Algoritma K-Means Untuk Penentuan Wilayah Penjualan Potensial Pada Perusahaan Jasa Cleaning Service. *TEMATIK*, 10(2), 204–209.
- Warianti, D. T., Astagina, P., Julianto, R., & Arini, F. Y. (2024). Optimasi K-Means Menggunakan Algoritma Firefly Untuk Segmentasi Pelanggan pada E-commerce. *JURNAL FASILKOM*, 14(3), 775–785.
- Yunitasari, Y., Nenobais, H., & Marwan, J. (2025). Peningkatan Daya Saing UMKM Berbasis Media dan Teknologi untuk Pasar Internasional dan Ekonomi Indonesia. *ABDI MOESTOPO: Jurnal Pengabdian Pada Masyarakat*, 8(1), 174–182.



© 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>)