



**MARKET-DRIVEN STRATEGY DENGAN PEMANFAATAN BIG DATA ANALYTIC
TOOLS: KLASTERISASI UNTUK OPTIMALISASI STRATEGI PEMASARAN DI SOULJA
COFFEE**

Melani Defina Maharani¹, Adam Hermawan², Muhammad Rizki Nugraha³

^{1,2,3}Program Studi Bisnis Digital, Universitas Pendidikan Indonesia, Indonesia

Correspondence: E-mail: ¹melanimaha123@upi.edu, ²adamhermawan@upi.edu, ³murinu@upi.edu

ABSTRACT

Penelitian ini berfokus pada peningkatan efektivitas strategi pemasaran Soulja Coffee melalui pendekatan berbasis data dengan prinsip *market-driven strategy*. Data pelanggan yang dikumpulkan melalui sistem CRM dianalisis menggunakan algoritma *K-Means Clustering* untuk membagi konsumen ke dalam beberapa segmen, sedangkan metode *Decision Tree* digunakan untuk mengenali pola perilaku khas dari setiap kelompok tersebut. Hasil dari analisis ini mengungkapkan adanya empat kelompok pelanggan dengan preferensi dan perilaku belanja yang bervariasi. Segmentasi ini menjadi dasar dalam merancang strategi pemasaran yang lebih spesifik dan efisien sesuai dengan karakter masing-masing segmen. Model klasifikasi yang dibangun menunjukkan akurasi yang baik serta memberikan hasil yang mudah dipahami. Pendekatan ini menawarkan Soulja Coffee kemampuan untuk membuat keputusan pemasaran yang lebih terarah, objektif, dan berkelanjutan. Secara keseluruhan, studi ini menegaskan bahwa pemanfaatan *big data* dan teknologi *machine learning* dapat menjadi solusi strategis untuk meningkatkan efektivitas pemasaran, terutama bagi UMKM di tengah transformasi digital.

ARTICLE INFO

Article History:

Submitted/Received 12 Juni 2025

First Revised 15 Juli 2025

Accepted 20 Juli 2025

First Available online 31 Juli 2025

Publication Date 31 Juli 2025

Keyword:

Market-Driven Strategy, K-Means Clustering, Decision Tree, Data-Driven Marketing, Soulja Coffee, Segmentasi Konsumen.

1. INTRODUCTION

Industri F&B, khususnya sektor minuman kopi, mengalami pertumbuhan signifikan di Indonesia seiring meningkatnya konsumsi dari berbagai kalangan, seperti Gen Z, Milenial, dan Gen X. Tren gaya hidup dan sosial turut mendorong lonjakan minat terhadap produk kopi, dari kopi siap saji hingga varian dengan cita rasa unik. Menurut Snapcart (2023), mayoritas konsumen dari berbagai generasi tersebut mengonsumsi kopi setiap hari, dengan preferensi yang terus berkembang.

Pertumbuhan ini juga mendorong persaingan yang semakin ketat. Di Kota Tasikmalaya, terdapat lebih dari 310 kedai kopi pada 2023 (Open Data Kota Tasikmalaya, 2025), menuntut pelaku usaha seperti Soulja Coffee untuk menerapkan strategi pemasaran yang adaptif dan berbasis pemahaman perilaku konsumen yang dinamis.

Namun perilaku konsumen terus mengalami perubahan waktu ke waktu karena pengaruh sosial, psikologis, dan personal (Eka et al., 2022; Hermawan et al., 2020), bahkan dalam memilih kopi, faktor seperti *brand attachment* tak kalah penting dari harga dan rasa (Pangestu & Aribowo, 2023). Oleh sebab itu, penting menerapkan *market-driven strategy* yaitu pendekatan yang menitikberatkan pada analisis pasar berbasis data untuk menyusun strategi yang kompetitif (Fauzi et al., 2022).

Kunci keberhasilan pendekatan ini terletak pada segmentasi pasar yang akurat. Melalui alat analitik seperti *K-Means Clustering* dan *Decision Tree*, pelaku usaha dapat memahami karakteristik kelompok konsumen bernilai tinggi (Galahartlambang et al., 2022). Maka pendekatan *data-driven marketing* memungkinkan strategi yang lebih personal, efisien, dan sesuai kebiasaan konsumen (Mahadewi & Iswara, 2025).

Bagi UMKM, keterbatasan akses data demografis dapat diatasi dengan memanfaatkan data historis dari sistem POS seperti frekuensi kunjungan, nominal transaksi, dan produk yang dibeli (Astikarani & I Komang, 2024). Informasi ini mendukung segmentasi berbasis perilaku untuk strategi yang lebih tajam.

Sebagai pelaku usaha lokal, Soulja Coffee menghadapi tantangan dalam efisiensi promosi dan menjaga loyalitas pelanggan. Hal ini didukung melalui data biaya operasional, terutama pada biaya pemasaran antara September 2024 hingga Maret 2025, melonjak dari 5,34% dan paling tinggi di angka 19,68% (Laporan Keuangan Soulja Coffee (Data Internal), 2025). Namun hal itu tidak diikuti kenaikan profit sepadan. Rasio BOPO yang melebihi 85% menunjukkan inefisiensi, jauh di atas standar Bank Indonesia sebesar 50–70% (Kahar et al., 2024).

Situasi ini menegaskan pentingnya strategi pemasaran yang lebih *data-driven* dan terukur. Sejalan dengan Armstrong & Kotler (2021), keberhasilan pemasaran menuntut pemahaman mendalam terhadap konsumen. Oleh karena itu, penelitian ini mengkaji penerapan *data-driven marketing* menggunakan *K-Means Clustering* dan *Decision Tree* untuk menyusun strategi pemasaran Soulja Coffee yang lebih tepat sasaran, efisien, dan mendukung pertumbuhan berkelanjutan.

2. TINJAUAN PUSTAKA

2.1 Pemasaran

Pemasaran adalah proses strategis yang bertujuan memenuhi kebutuhan konsumen secara menguntungkan dengan menciptakan dan menukar nilai secara berkelanjutan. Menurut (Kotler & Keller, 2016), pemasaran mencakup interaksi sosial dan manajerial yang memungkinkan individu atau kelompok memperoleh apa yang mereka butuhkan melalui pertukaran nilai. Definisi dari American Marketing Association (AMA) memperluas konsep ini

dengan menekankan bahwa pemasaran melibatkan aktivitas, lembaga, dan proses untuk menciptakan, menyampaikan, serta menukar penawaran bernilai kepada pelanggan, mitra, dan masyarakat. Dalam praktiknya, pemasaran modern menekankan pada *value creation and sharing*, yaitu penciptaan nilai bersama konsumen yang diposisikan sebagai mitra.

2.2 Perilaku Konsumen

Perilaku konsumen mencerminkan bagaimana individu atau kelompok mengambil keputusan terkait pembelian, penggunaan, dan evaluasi produk atau jasa, yang dipengaruhi oleh interaksi antara emosi, pemikiran, tindakan, serta lingkungan sekitar (Setiadi, 2023). Kotler & Keller (2016) mengelompokkan pengaruh ini ke dalam empat faktor utama: budaya (termasuk nilai, subkultur, dan kelas sosial), sosial (kelompok referensi, peran, dan status), pribadi (usia, pekerjaan, pendapatan, gaya hidup), serta psikologis (motivasi, persepsi, pembelajaran, dan sikap). Sari (2023) menyoroti peran dorongan bawah sadar dalam keputusan pembelian, sementara Calicchio (2023) menyatakan bahwa kebutuhan manusia berkembang dari dasar hingga aktualisasi diri. Pengalaman berulang juga membentuk persepsi dan loyalitas terhadap merek, sehingga konsumen cenderung bertindak berdasarkan keyakinan dan pengalaman sebelumnya. Dengan memahami dinamika ini secara menyeluruh, perusahaan dapat mengembangkan strategi pemasaran yang lebih relevan dan responsif terhadap perubahan perilaku dan ekspektasi konsumen, terutama di tengah percepatan transformasi digital.

2.3 Segmentasi Pasar

Segmentasi pasar merupakan strategi inti dalam pemasaran modern yang memungkinkan perusahaan membagi pasar menjadi kelompok konsumen homogen berdasarkan karakteristik tertentu (Setiadi, 2023). Pendekatan ini penting karena tidak semua konsumen dapat dijangkau dengan strategi tunggal (Kotler & Keller, 2016). Cravens & Piercy (2013) menekankan bahwa segmentasi adalah proses strategis yang meliputi identifikasi pasar, pemilihan variabel, dan perancangan strategi tiap segmen. Segmentasi dapat berbasis demografis, geografis, psikografis, dan perilaku (Schiffman, 2015), serta didukung teknologi seperti CRM dan *data mining* untuk analisis konsumen yang lebih akurat.

Dalam praktiknya, segmentasi berkembang ke arah mikrosegmentasi dan personalisasi massal dengan memanfaatkan data riwayat transaksi, preferensi, dan waktu pembelian. Teknik seperti *K-Means Clustering* dan *Decision Tree* digunakan untuk mengungkap pola perilaku dan menghasilkan model yang mudah diterapkan (Abdul-Rahman et al., 2021; Guo et al., 2020). Strategi ini memungkinkan perusahaan menawarkan produk yang lebih relevan dan mempertimbangkan potensi pasar serta daya saing tiap segmen (Octiva et al., 2024). Penelitian terbaru menunjukkan bahwa segmentasi yang tepat meningkatkan efisiensi sumber daya, efektivitas pemasaran, dan loyalitas pelanggan (Wulandari et al., 2024).

2.4 Market Driven Strategy

Market-driven strategy adalah pendekatan pemasaran yang menitikberatkan pada pemahaman mendalam terhadap kebutuhan pelanggan, tren pasar, dan persaingan guna menciptakan keunggulan kompetitif yang berkelanjutan (Cravens & Piercy, 2013). Strategi ini menuntut perusahaan untuk bersikap adaptif dan proaktif dalam mengenali peluang pasar, dengan dukungan evaluasi kapabilitas internal, inovasi teknologi informasi, dan perbaikan strategi yang terus menerus. *Market-driven strategy* identic dengan segmentasi pasar yang menjadi fondasi utama karena memungkinkan penyusunan strategi yang sesuai dengan karakteristik tiap segmen. Cravens & Piercy (2013) menyebutkan enam tahapan inti

strategi ini: identifikasi nilai pelanggan, penyesuaian kapabilitas, pemilihan target, pengembangan *positioning*, implementasi, dan evaluasi berkelanjutan.

Positioning, menurut Kotler & Keller (2016), adalah proses merancang citra dan penawaran agar menempati posisi yang unik dan bernilai di benak konsumen, yang mencakup aspek diferensiasi produk, harga, distribusi, dan promosi. Dalam konteks penelitian ini, *positioning* digunakan untuk mengoptimalkan strategi berbasis hasil segmentasi pelanggan, sehingga menghasilkan strategi pemasaran yang lebih akurat, personal, dan kompetitif. Selaras dengan itu, Hendarwan, (2023) menyatakan bahwa penerapan strategi *market-driven* mampu meningkatkan kapabilitas internal dan kinerja bisnis secara keseluruhan. Strategi ini menekankan pentingnya koordinasi informasi pasar dalam pengambilan keputusan strategis dan terbukti mendukung penciptaan nilai pelanggan yang unggul serta respons yang adaptif terhadap perubahan pasar.

2.5 Customer Relationship Management (CRM)

Customer Relationship Management (CRM) adalah pendekatan bisnis terintegrasi yang menggabungkan strategi, operasional, dan teknologi untuk membangun serta mempertahankan hubungan jangka panjang dengan konsumen (Prior et al., 2024). CRM terbagi menjadi tiga jenis: *strategic CRM* yang berfokus pada orientasi konsumen, *operational CRM* yang mengotomatisasi proses bisnis, dan *analytical CRM* yang menganalisis data pelanggan melalui *data mining* dan *big data*.

Penerapannya, seperti melalui aplikasi POS Majoo, memungkinkan pencatatan transaksi, pengelolaan relasi, dan pemantauan penjualan. Dengan menyediakan layanan yang lebih personal, CRM menjadi kunci dalam meningkatkan loyalitas pelanggan dan mendukung pertumbuhan bisnis berkelanjutan.

2.6 Data Driven Marketing

Pemasaran berbasis data adalah strategi yang memanfaatkan informasi pelanggan secara sistematis untuk merancang dan mengevaluasi aktivitas pemasaran secara terukur. Jeffery (2010) menyatakan bahwa pendekatan ini tidak hanya memberikan keunggulan kompetitif, tetapi juga berdampak positif pada kinerja keuangan. Strategi ini dimulai dari penentuan tujuan yang jelas dan pengumpulan data yang relevan, dengan menerapkan prinsip 80/20 untuk memaksimalkan nilai dari sebagian kecil data yang paling berpengaruh. Teknik analitik seperti *propensity modeling*, *market basket analysis*, dan *decision tree* digunakan untuk memahami pola perilaku konsumen dan menyusun strategi penargetan yang lebih efektif.

Sistem pendukung seperti *Customer Relationship Management* (CRM) dan *Enterprise Data Warehouse* (EDW) memungkinkan pengelolaan data yang terintegrasi, sehingga perusahaan dapat merespons kebutuhan pasar secara cepat dan berbasis bukti. Di samping itu, transparansi dan perlindungan data menjadi aspek penting dalam menjaga kepercayaan pelanggan dan mematuhi regulasi. Kotler & Keller, (2012) menekankan pentingnya pemanfaatan data pelanggan, seperti riwayat pembelian, karakteristik demografis, dan perilaku digital untuk personalisasi penawaran dan peningkatan strategi retensi. Penerapan pemasaran berbasis data mendorong efisiensi biaya, memperkuat loyalitas konsumen, dan mendukung pertumbuhan bisnis yang berkelanjutan.

2.7 Data Mining

Data mining adalah proses analitis yang digunakan untuk menggali pola tersembunyi, tren, dan segmentasi konsumen dari kumpulan data berukuran besar. Han et al., (2023)

menjelaskan bahwa teknik ini memungkinkan perusahaan menemukan informasi penting yang mendukung keputusan strategis pemasaran. Dengan memanfaatkan basis data pelanggan, bisnis dapat mengidentifikasi prospek, menyesuaikan penawaran, meningkatkan retensi, mengaktifkan kembali pelanggan tidak aktif, serta mengurangi kesalahan layanan.

Salah satu metode utama dalam *data mining* adalah *clustering*, di mana teknik *K-Means* digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik secara efisien. Metode ini menghasilkan kluster yang konsisten dan terpisah jelas, meskipun sensitif terhadap *noise*. Sementara itu, *decision tree* dimanfaatkan untuk klasifikasi, bekerja dengan pendekatan *top-down* untuk membentuk struktur pembagian data yang optimal. Kombinasi antara *K-Means* dan *decision tree* memperkuat analisis perilaku pelanggan secara tersegmentasi dan prediktif, sehingga mendukung pengembangan strategi pemasaran yang lebih tepat sasaran dan kompetitif,

3. METODE

Dalam penelitian ini, metode yang digunakan disusun secara sistematis untuk mencapai tujuan penelitian yang telah ditetapkan. Penelitian dilakukan dengan pendekatan kuantitatif berbasis analisis data sekunder yang diperoleh dari sistem *Point of Sale (POS)* Soulja Coffee. Fokus utama dari metode ini adalah penerapan teknik *data mining* untuk mengidentifikasi segmentasi pelanggan secara lebih akurat melalui pendekatan *K-Means Clustering*, serta mengevaluasi performa klasifikasi hasil segmentasi dengan algoritma *Decision Tree*. Dengan memanfaatkan *big data analytic tools*, penelitian ini bertujuan untuk merancang strategi pemasaran berbasis data (*data-driven marketing*) yang mampu meningkatkan efektivitas serta efisiensi dalam menjangkau target pasar secara lebih relevan dan terarah. Tahapan-tahapan analisis dalam penelitian ini mengacu pada model segmentasi konsumen yang dikembangkan oleh Cravens dan Piercy (2013), yang terdiri dari lima tahapan strategis, yaitu:

1. Menentukan Konsumen yang Akan Disegmentasi

Objek Penelitian ini bertujuan menerapkan *data-driven marketing* untuk menganalisis strategi pemasaran berbasis *market-driven strategy* menggunakan algoritma *K-Means Clustering* dan *Decision Tree* pada data transaksi pelanggan Soulja Coffee. Fokus utama penelitian adalah mengoptimalkan strategi pemasaran melalui pemahaman pola konsumsi aktual. Subjek penelitian mencakup konsumen Soulja Coffee di Kota Tasikmalaya selama periode Desember 2024 hingga Maret 2025.

Dengan pendekatan kuantitatif, penelitian ini menggabungkan algoritma *unsupervised (K-Means)* dan *supervised (Decision Tree)* dalam analisis numerik (Bougie & Sekaran, 2019). Data sekunder diambil dari sistem *Point of Sale (POS)* yang terintegrasi dengan CRM dan platform Majoo. Proses pengumpulan data dilakukan melalui ekspor transaksi ke Excel, mencakup variabel seperti lokasi, produk, metode pembayaran, dan waktu pembelian untuk analisis segmentasi pelanggan.

2. Mengidentifikasi Segmentasi Konsumen

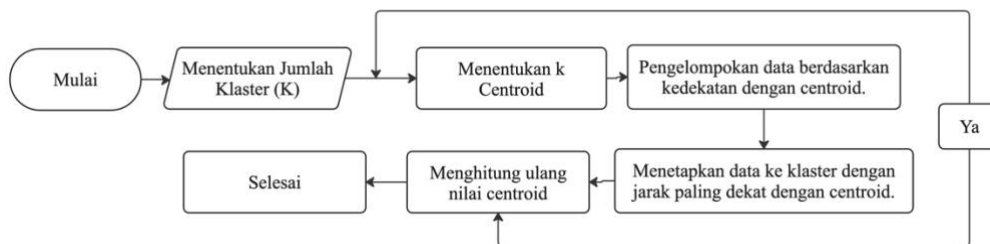
Tahapan ini mengandalkan teknik *Exploratory Data Analysis (EDA)* sebagai bagian dari instrumen penelitian. EDA berfungsi untuk memahami struktur dan distribusi data, mengidentifikasi *outlier*, serta memastikan kualitas data sebelum diterapkan ke algoritma analitik. Menurut Angela et al., (2023), Langkah-langkah EDA meliputi:

- a. *Data Understanding*: Penelaahan variabel numerik dan kategorikal, distribusi data, dan keterkaitan antar variabel.
- b. *Data Preprocessing*: Proses pembersihan dan validasi data untuk menghilangkan duplikasi, nilai kosong, dan format yang tidak konsisten.

- c. *Formatting & Encoding*: Koreksi tipe data (misalnya konversi kolom tanggal menjadi datetime) dan penyediaan data kategorikal.

3. Membentuk Segmentasi Konsumen

Segmentasi dilakukan menggunakan algoritma *K-Means Clustering*, yang mengelompokkan konsumen ke dalam kluster berdasarkan kemiripan pola transaksi. Menurut Sarker et al., (2018) Tahapan pelaksanaan algoritma *K-Means* dijelaskan pada gambar 3.1 sebagai berikut:



Gambar 3.1 Flowchart K-Means Clustering

1. Menentukan jumlah kluster optimal menggunakan metode Elbow
2. Memilih K centroid awal secara acak dan menghitung *sum of squared distances (SSD)*
3. Mengelompokkan data berdasarkan jarak terdekat ke *centroid* menggunakan *Euclidean Distance*
4. Menetapkan data ke kluster dengan *centroid* terdekat
5. Menghitung ulang centroid berdasarkan rata-rata anggota kluster
6. Mengulangi proses hingga posisi *centroid* stabil atau mencapai batas iterasi

Algoritma *K-Means* bertujuan meminimalkan *intra-cluster distance* dan memaksimalkan *inter-cluster variation*, sehingga kluster yang terbentuk bersifat homogen secara internal namun berbeda signifikan antar-kluster.

4. Strategi Segmentasi yang Lebih Mendalam

Setelah klusterisasi dilakukan, algoritma *Decision Tree* digunakan untuk mengevaluasi akurasi model dan mengidentifikasi karakteristik khas masing-masing kluster. Struktur pohon keputusan dibentuk melalui pengukuran *Information Gain*. Untuk mengevaluasi sejauh mana model dapat memprediksi kluster pelanggan secara akurat, digunakan pendekatan *confusion matrix* seperti pada tabel 3.2 dan metrik evaluasi klasifikasi.

Tabel 3. 1 *Confusion Matrix*

Actual \ Predicted	Cluster 0	Cluster 1	Cluster 2	...	Cluster n	Total
Cluster 0	TP (C0)	FP (C1 on C0)	FP (C2 on C0)	...	FP (Cn on C0)	P ₀
Cluster 1	FN (C0 from C1)	TP (C1)	FP (C2 on C1)	...	FP (Cn on C1)	P ₁
Cluster 2	FN (C0 from C2)	FN (C1 from C2)	TP (C2)	...	FP (Cn on C2)	P ₂
...
Cluster n	FN (C0 from Cn)	FN (C1 from Cn)	FN (C2 from Cn)	...	TP (Cn)	P _n

Tabel 3.2 merupakan tabel yang menunjukkan kinerja model klasifikasi dengan membandingkan kelas sebenarnya (*actual labels*) dan kelas yang diprediksi (*predicted*)

labels) untuk lebih dari dua kelas. Han et al., (2011) juga menyebutkan beberapa metrik evaluasi kinerja model klasifikasi, yaitu seperti: *Accuracy*, *Precision*, *Recall (Sensitivity)*, *F1-Score*, *Macro Average*, dan *Weighted Average*. Metrik ini memberikan gambaran menyeluruh terhadap performa model, terutama dalam konteks *multi-class classification*.

5. Memilih Strategi Segmentasi yang Optimal

Pendekatan yang digunakan dalam penelitian ini berfokus pada *market-driven strategy* dengan mengimplementasikan dengan pendekatan yang efektif yaitu menerapkan strategi *positioning*. Menurut Kotler & Keller (2016), Penerapan strategi *positioning* dalam penelitian ini mengacu pada kerangka *marketing mix 4P*, yaitu:

- a. *Product*: Menyesuaikan produk unggulan sesuai preferensi dominan klaster.
- b. *Price*: Menyusun harga sesuai sensitivitas konsumen tiap segmen.
- c. *Place*: Menentukan saluran distribusi dan pemesanan berdasarkan kebiasaan klaster.
- d. *Promotion*: Menyesuaikan waktu dan bentuk promosi dengan karakteristik perilaku transaksi klaster.

Pendekatan ini menjadikan segmentasi tidak hanya sebagai alat analisis, tetapi juga sebagai dasar pengambilan keputusan strategis dalam pemasaran.

4. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan mengelompokkan konsumen Soulja Coffee berdasarkan perilaku transaksional mereka dengan menerapkan metode *K-Means Clustering* dan *Decision Tree*. Analisis dilakukan menggunakan data transaksi periode Desember 2024–Maret 2025 untuk memahami pola konsumsi aktual pelanggan, dengan pendekatan kuantitatif dan pengolahan data secara eksploratif.

1. Menentukan Konsumen yang Akan Disegmentasi

Objek dalam penelitian ini adalah seluruh konsumen yang pernah bertransaksi di Soulja Coffee selama periode Desember 2024 hingga Maret 2025, tanpa batasan frekuensi, waktu, atau jenis produk. Variabel yang digunakan dalam data transaksi mencakup waktu order, metode pembayaran, jenis pesanan, produk, dan nominal transaksi. Tabel 1 berikut merangkum atribut dari dataset:

Tabel 4. 1 Atribut Dataset Transaksi Konsumen

No.	Variable	Deskripsi	Tipe Atribut
1.	No Transaksi	Kode unik untuk setiap transaksi	Nominal
2.	Waktu Order	Tanggal dan jam pelanggan melakukan order	Nominal
3.	Waktu Bayar	Tanggal dan jam pelanggan melakukan pembayaran	Nominal
4.	Outlet	Lokasi tempat transaksi dilakukan.	Nominal
5.	Produk	Nama produk yang dibeli dalam transaksi tersebut.	Numerik
6.	Jenis Order	Kategori jenis pesanan seperti <i>outlet</i> (ditempat) atau melalui layanan <i>ojek online</i> (GoFood, GrabFood, ShopeeFood)	Numerik
7.	Sisa Tagihan	Jumlah sisa tagihan yang belum dibayar untuk transaksi tersebut.	Nominal
8.	Total Penjualan	Total nilai moneter dari transaksi (penjualan) yang dilakukan.	Numerik

9.	Metode Pembayaran	Jenis pembayaran (Cash, QRIS, dll)	Numerik
10.	Bayar	Channel yang dipilih ketika pelanggan melakukan pembayaran	Numerik
11.	Order	Channel yang dipilih ketika pelanggan melakukan Order	Numerik

Tabel 4.1 memuat 11 atribut dari data transaksi pelanggan Soulja Coffee untuk analisis segmentasi, mencakup informasi transaksi, waktu, lokasi, produk, metode pemesanan, dan pembayaran. Sebagian besar atribut bertipe numerik untuk mendukung pemrosesan algoritma, sementara sisanya berupa data nominal sebagai kategori. Kombinasi atribut ini merepresentasikan pola konsumsi pelanggan secara menyeluruh.

2. Mengidentifikasi Segmentasi Konsumen

Segmentasi konsumen diawali dengan eksplorasi dataset transaksi Soulja Coffee (Desember 2024–Maret 2025) menggunakan pendekatan *Exploratory Data Analysis* (EDA) untuk memastikan kelayakan data sebelum dilakukan klusterisasi.

a. Data Understanding

Analisis dimulai dengan menelaah struktur 11 atribut transaksi pelanggan. Beberapa atribut waktu seperti “Waktu Bayar” diolah menjadi “Jam Bayar” dalam format desimal dan “Hari Bayar” sebagai angka 0–6. Variabel yang relevan untuk segmentasi ditetapkan berdasarkan kekonsistenan nilai, distribusi data, dan keterkaitannya terhadap perilaku pembelian. Hasilnya, lima fitur utama dipilih: jam bayar, hari bayar, produk, jenis order, dan metode pembayaran.

b. Data Preprocessing

Data dibersihkan dari kolom tidak relevan (ID transaksi, outlet, dll.), data kosong, serta duplikasi atau *noise* sistem. Produk *bundle* disederhanakan, dan transaksi tidak sah dihapus. Penyamaan label dilakukan agar tidak terjadi redundansi, memastikan hanya data valid yang dipertahankan untuk proses analisis lanjutan.

c. Formatting dan Tipe Data

Seluruh variabel dikonversi ke bentuk numerik melalui *Label Encoding* agar sesuai dengan algoritma *machine learning*. Fitur waktu diformat ulang, produk dieksplorasi menjadi satuan, dan seluruh kategori diseragamkan (seperti: “Ojek Online – Grab” menjadi GRABFOOD). Dataset akhir siap digunakan untuk tahap segmentasi berbasis *clustering*.

3. Membentuk Segmentasi Konsumen

Segmentasi pelanggan dilakukan menggunakan pendekatan *unsupervised learning* melalui algoritma *K-Means Clustering*, karena kemampuannya mengelompokkan data tanpa label berdasarkan kemiripan pola transaksi pelanggan.

a. Pemilihan Fitur Segmentasi

Lima fitur utama digunakan sebagai dasar segmentasi: Jam Bayar, Produk_encoded, Jenis Order_encoded, Metode Pembayaran_encoded, dan Hari Bayar_encoded. Fitur-fitur ini merepresentasikan dimensi perilaku transaksi pelanggan dari sisi waktu, produk, jenis pemesanan, dan metode pembayaran.

b. Penentuan Jumlah Kluster

Metode Elbow digunakan untuk menentukan jumlah kluster optimal. Nilai *Sum of Squared Errors* (SSE) diplot dari k=1 hingga k=10 dan menghasilkan titik elbow pada k=4, yang menunjukkan bahwa penambahan kluster di atas nilai ini tidak

meningkatkan hasil segmentasi secara signifikan. Oleh karena itu, nilai $k=4$ digunakan dalam model.

c. Proses dan Hasil Klasterisasi

Model *K-Means* dilatih dengan $k=4$, lalu setiap entri data diklasifikasikan ke salah satu dari empat klaster. Visualisasi klaster menggunakan *scatterplot* (Jam Bayar vs Produk_encoded) menunjukkan penyebaran yang cukup terpisah antar segmen. Ringkasan karakteristik tiap klaster dirangkum sebagai berikut:

Tabel 4.2. Hasil Segmentasi Konsumen

Cluster	Jumlah_Data	Jam_Bayar_Rata2	Produk_Dominan	Jenis_Order_Dominan	Metode_Pembayaran_Dominan	Hari_Dominan
0	746	18:02	Long Black Iced	OUTLET	Cash	Sabtu
1	1215	17:17	Butterscotch	OUTLET	Cash	Sabtu
2	2533	17:13	Soulful	OUTLET	GRAB	Sabtu
3	886	17:23	Coffee Lemon	OUTLET	Cash	Sabtu

Hasil pada tabel 4.2 menunjukkan bahwa sebagian besar pelanggan melakukan pembelian pada sore hari, dengan klaster 0 sebagai kelompok dominan dalam jumlah dan preferensi produk. Variasi metode pembayaran dan produk antar klaster memberikan dasar diferensiasi untuk strategi pemasaran berbasis data.

4. Strategi Segmentasi yang Lebih Mendalam

Setelah membentuk klaster pelanggan menggunakan *K-Means*, langkah selanjutnya adalah memahami pola segmentasi secara lebih dalam melalui model prediktif. Untuk itu digunakan *Decision Tree Classifier*, yang mampu menjelaskan klasifikasi secara interpretatif berdasarkan pola perilaku transaksi pelanggan.

a. Pelatihan dan Validasi Model

Model *Decision Tree* dilatih menggunakan hasil klasterisasi sebagai label target (Cluster 0–3), dengan fitur Jam Bayar, Produk_encoded, Jenis Order_encoded, Metode Pembayaran_encoded, dan Hari Bayar_encoded. Dataset dibagi 70% data pelatihan dan 30% data pengujian, menggunakan parameter regularisasi untuk menghindari *overfitting*. Hasilnya, model mencatat akurasi tinggi dengan: Akurasi pelatihan: 98.8%; Akurasi pengujian: 98.5%; Cross-validation: $98.9\% \pm 0.9\%$; Gap overfitting: 0.3%. Hal ini menunjukkan generalisasi model yang sangat baik.

b. Evaluasi *Confusion Matrix* dan Metrik Klasifikasi

Confusion matrix memperlihatkan bahwa klasifikasi Cluster 1, 2, dan 3 berjalan sempurna. Namun, Cluster 0 mengalami sedikit kesalahan klasifikasi ke Cluster 3, akibat kemiripan fitur. Metrik evaluasi menunjukkan:

Tabel 4.3 Metrik Evaluasi Kinerja Model Klasifikasi

	precision	recall	f1-score	support
Cluster 0	1.00	0.89	0.94	224.00
Cluster 1	1.00	1.00	1.00	364.00
Cluster 2	1.00	1.00	1.00	760.00
Cluster 3	0.91	1.00	0.96	266.00
accuracy	0.98	0.98	0.98	0.98
macro avg	0.98	0.97	0.97	1614.00

weighted avg	0.99	0.98	0.98	1614.00
---------------------	------	------	------	---------

Tabel 4.4 menunjukkan bahwa Cluster 1 dan 2 mencapai skor evaluasi sempurna (precision, recall, dan f1-score sebesar 1.00), menandakan performa model yang sangat konsisten dalam mengenali kedua klaster. Cluster 3 juga tampil baik dengan recall 1.00 dan precision 0.91, menunjukkan semua data aktual teridentifikasi meski ada sedikit kesalahan prediksi. Cluster 0 memiliki precision tinggi (1.00), namun recall 0.89, mengindikasikan sebagian data tidak terklasifikasi dengan tepat, sesuai hasil *confusion matrix* sebelumnya. Secara keseluruhan, model mencapai akurasi 98%, dengan macro f1-score 0.97 dan weighted f1-score 0.98, mencerminkan performa yang andal dan stabil dalam membedakan pola perilaku antar segmen pelanggan, meskipun ada ketidakseimbangan jumlah data antar klaster.

c. Interpretasi Decision Tree

Visualisasi *decision tree* menunjukkan bahwa variabel Produk_encoded menjadi pemisah utama di simpul akar, diikuti oleh Jam Bayar dan Metode Pembayaran. Struktur pohon menggambarkan aturan klasifikasi seperti pada tabel 4.4:

Tabel 4.4 Interpretasi Cluster pada *Decision Tree*

Cluster	Aturan Kunci (Kondisi)	Jumlah Sampel (Kemurnian Gini)
Cluster 0	Produk_encoded <= 55.5 AND Produk_encoded > 19.5 AND Jam Bayar <= 0.5 OR Produk_encoded > 55.5	1993 (gini = 0.667) & 1773 (gini = 0.0)
Cluster 1	Produk_encoded <= 19.5 AND Produk_encoded <= 18.5	851 (gini = 0.0)
Cluster 2	Produk_encoded > 55.5	1773 (gini = 0.0)
Cluster 3	Produk_encoded <= 19.5 AND Produk_encoded > 18.5 AND Metode Pembayaran_encoded <= 0.5 OR Produk_encoded <= 19.5 AND Produk_encoded > 18.5 AND Metode Pembayaran_encoded > 0.5 AND Produk_encoded <= 38.5 OR Produk_encoded <= 19.5 AND Produk_encoded > 18.5 AND Metode Pembayaran_encoded > 0.5 AND Produk_encoded > 38.5	254 (gini = 0.489) & 198 (gini = 0.465) & 146 (gini = 0.448)

Tabel 4. menunjukkan bahwa Cluster 1 dan 2 bersifat sangat homogen (gini = 0.0), masing-masing terdefinisi oleh preferensi produk bernilai rendah (≤ 18.5) dan tinggi (> 55.5). Sebaliknya, Cluster 0 dan 3 lebih heterogen (gini > 0), dengan Cluster 0 sebagai klaster dominan namun beragam, dan Cluster 3 mencerminkan kombinasi preferensi produk menengah serta metode pembayaran yang bervariasi. Secara keseluruhan, *Decision Tree* memvalidasi segmentasi *K-Means* dengan menghasilkan aturan *if-then* yang jelas, menunjukkan bahwa fitur produk, waktu pembayaran, dan metode pembayaran merupakan pemisah utama antar klaster yang karakteristiknya konsisten dengan temuan dominan dari analisis *K-Means*.

5. Memilih Strategi Segmentasi yang Optimal

Penelitian ini mengadopsi pendekatan *market-driven strategy* yang menekankan eksplorasi kebutuhan pelanggan sebagai dasar penyusunan strategi pemasaran. Segmentasi berbasis *K-Means* yang dibantu *Decision Tree* menghasilkan empat klaster pelanggan Soulja Coffee yang memiliki karakteristik unik dalam hal waktu transaksi, preferensi produk, metode pembayaran, dan gaya pemesanan.

Analisis struktur *decision tree* menunjukkan bahwa atribut produk menjadi pemisah paling signifikan antar segmen. Sementara itu, pola waktu pembelian dan metode pembayaran turut berperan penting dalam membedakan klaster-klaster pelanggan. Pemahaman ini memberikan dasar kuat dalam menyusun strategi pemasaran yang relevan untuk tiap segmen. Berikut pada tabel 4.5 adalah Deskripsi analisis pada setiap cluster.

Tabel 4.5 Analisis Cluster

Cluster	Deskripsi Analisis
Cluster 0	Klaster ini mencakup pelanggan yang cenderung membeli <i>Long Black Iced</i> secara langsung di outlet dengan pembayaran tunai. Transaksi rata-rata terjadi pukul 18:02 pada hari Sabtu, mencerminkan kebiasaan menikmati kopi klasik di malam akhir pekan.
Cluster 1	Kelompok ini didominasi oleh pembeli <i>Butterscotch</i> yang melakukan transaksi tunai di outlet, dengan waktu rata-rata sekitar 17:17 pada hari Sabtu. Ini menunjukkan kecenderungan terhadap minuman manis yang dinikmati di sore hari menjelang akhir pekan.
Cluster 2	Sebagai klaster dengan jumlah pelanggan terbanyak, segmen ini mengutamakan <i>Soulful</i> , dibeli di outlet namun dibayar melalui layanan <i>Grab</i> . Transaksi umumnya terjadi sekitar pukul 17:13 di hari Sabtu, menandakan preferensi terhadap pembelian praktis berbasis digital.
Cluster 3	Klaster ini terdiri dari pelanggan yang memilih <i>Coffee Lemon</i> , dengan pembelian langsung di outlet dan pembayaran tunai. Waktu transaksi rata-rata adalah pukul 17:23 pada hari Sabtu, menandakan ketertarikan pada minuman dengan cita rasa unik dan segar di akhir pekan.

Tabel 4.5 merangkum karakteristik tiap klaster berdasarkan produk utama, metode pembayaran, waktu, dan hari transaksi. Meskipun seluruh klaster cenderung bertransaksi pada hari Sabtu, terdapat perbedaan dalam jenis produk dan waktu pembelian. Temuan ini dapat digunakan untuk merancang promosi dan penawaran produk yang lebih sesuai dengan profil masing-masing segmen.

Berdasarkan hasil segmentasi tersebut, penulis merumuskan strategi untuk mengoptimalkan pemasaran dan operasional Soulja Coffee dengan mempertimbangkan efisiensi biaya, pengembangan produk, penawaran harga, serta program loyalitas yang sesuai karakteristik tiap klaster. Meskipun preferensi produk berbeda antar segmen, pola perilaku pelanggan, seperti waktu transaksi, metode pembayaran, dan Lokasi yang menunjukkan kesamaan yang cukup kuat. Oleh karena itu, pendekatan strategi terpadu berbasis kerangka 4P (*Product, Price, Place, Promotion*) dinilai paling efektif untuk menjangkau seluruh segmen melalui satu strategi gabungan yang tetap mempertimbangkan kebutuhan masing-masing kelompok. Berikut pada tabel 4.6 adalah rekomendasi strategi pemasaran 4P dari seluruh segmen.

Tabel 4.6 Rekomendasi Strategi Pemasaran 4P

4P	Strategi Gabungan
Product (Produk)	1. Tawarkan 3 kategori utama: <ol style="list-style-type: none"> Kopi klasik (<i>Long Black Iced</i>) untuk penikmat kopi murni. Minuman manis dan creamy (<i>Butterscotch, Soulful</i>) untuk segmen muda atau casual drinkers.

	<ul style="list-style-type: none"> c. Minuman eksperimental (Coffee Lemon) untuk yang mencari pengalaman rasa unik. 2. Sediakan <i>seasonal drinks</i> untuk memancing ekspektasi baru dan menjaga loyalitas.
Price (Harga)	<p>Terapkan strategi harga diferensiasi ringan:</p> <ul style="list-style-type: none"> a. Harga premium untuk kopi klasik. b. Diskon <i>bundling</i> untuk pembelian minuman manis/eksperimental. c. <i>Happy Hour</i> (16.30–18.30) di akhir pekan untuk semua kategori minuman.
Place (Tempat)	<ul style="list-style-type: none"> 1. Fokus pada outlet fisik karena semua klaster dominan di sini. 2. Tingkatkan kemitraan dengan platform seperti Grab untuk penjualan dan promosi digital. 3. Evaluasi dan optimalkan jam operasional: pastikan siap di puncak transaksi Sabtu sore menjelang malam.
Promotion (Promosi)	<p>Gunakan kombinasi digital dan offline:</p> <ul style="list-style-type: none"> a. Digital: Promosi di GrabFood, Instagram Reels, kampanye UGC (User-Generated Content). b. Offline: Banner khusus <i>weekend drinks</i>, diskon <i>cash-only</i>, sampling minuman baru. c. Bangun tema promosi akhir pekan, seperti: “Temukan Rasa Favoritmu Setiap Sabtu!” atau “Waktunya Menyegarkan Soremu”.

Strategi pemasaran 4P pada Tabel 4.7 disusun sebagai respons terhadap tantangan bisnis Soulja Coffee, khususnya dalam menekan biaya promosi dan meningkatkan efisiensi operasional. Dengan pendekatan berbasis data, fokus produk diarahkan pada minuman yang paling diminati, strategi harga dibuat kompetitif melalui *bundling* dan promo waktu tertentu, serta distribusi dipusatkan pada outlet fisik dengan penyesuaian jam operasional. Promosi dilakukan secara terarah melalui kombinasi kanal digital dan *offline* untuk hasil maksimal dengan anggaran terbatas. Secara keseluruhan, strategi ini bertujuan meningkatkan profit, menekan pemborosan, dan menjaga efisiensi bisnis secara berkelanjutan di tengah persaingan pasar kopi.

5. CONCLUSION

Berdasarkan hasil analisis segmentasi pelanggan Soulja Coffee menggunakan algoritma *K-Means Clustering* dan *Decision Tree*, ditemukan empat klaster dengan perilaku konsumsi yang berbeda. Segmentasi dilakukan berdasarkan variabel jam pembayaran, jenis produk, metode pemesanan, metode pembayaran, dan hari transaksi. Hasil klasterisasi menunjukkan pola konsisten di setiap segmen, seperti preferensi produk tertentu (Long Black Iced, Butterscotch, Soulful, dan Coffee Lemon), dominasi transaksi pada hari Sabtu, serta kecenderungan metode pembayaran, baik tunai maupun melalui Grab. Model *Decision Tree* memvalidasi segmentasi dengan akurasi tinggi, yaitu 98 persen secara keseluruhan, dengan nilai *f1-score* rata-rata 0.98 dan *macro average* 0.97. Produk terbukti menjadi variabel paling membedakan antar klaster, diikuti oleh jam transaksi dan metode pembayaran.

Strategi pemasaran disusun dengan pendekatan *market-driven strategy* dan diperkuat melalui kerangka *marketing mix 4P*. Meskipun terdapat variasi preferensi produk antar klaster, kemiripan dalam perilaku inti seperti waktu, hari, dan lokasi transaksi memungkinkan diterapkannya strategi terpadu. Elemen produk, harga, tempat, dan promosi dirancang untuk menjangkau seluruh segmen secara efisien, sambil tetap menyesuaikan dengan karakteristik

masing-masing klaster. Pendekatan ini mendukung prinsip *market-driven* yang berfokus pada pemahaman konsumen sebagai landasan penyusunan strategi bisnis yang relevan, terukur, dan berkelanjutan.

6. REFERENCES

- Abdul-Rahman, S., Arifin, N. F. K., Hanafiah, M., & Mutalib, S. (2021). Customer Segmentation and Profiling for Life Insurance using K-Modes Clustering and Decision Tree Classifier. *International Journal of Advanced Computer Science and Applications*, 12(9). <https://doi.org/10.14569/IJACSA.2021.0120950>
- Armstrong, G., & Kotler, P. (2021). *Dasar-Dasar Pemasaran (Alih Bahasa Alexander Sindoro dan Benyamin Molan)(I)*.
- Angela, J. B., Islamiyah, & Irsyad, A. (2023). Implementasi Visualisasi Data Berbasis Web Pada Exploratory Data Analysis Profil Kesehatan Kota Samarinda. *Kreatif Teknologi Dan Sistem Informasi (KRETISI)*, 1(1), 9–16. <https://doi.org/10.30872/kretisi.v1i1.447>
- Astikarani, N. M. D., & I Komang, Y. W. P. (2024). Pemanfaatan Digital Marketing Sebagai Strategi Komunikasi Pemasaran Efektif dan Inovatif untuk Menarik Konsumen UMKM. *Waisya : Jurnal Ekonomi Hindu*, 3(2), 133–147. <https://doi.org/10.53977/jw.v3i2.2006>
- Bougie, R., & Sekaran, U. (2019). *Research Methods For Business: A Skill Building Approach* (8th ed.).
- Calicchio, S. (2023). *Abraham Maslow, dari hierarki kebutuhan hingga pemenuhan diri: Sebuah perjalanan dalam psikologi humanistik melalui hierarki kebutuhan, motivasi, dan pencapaian potensi manusia sepenuhnya*.
- Cravens, D. W., & Piercy, N. (2013). *Strategic Marketing*.
- Eka, E. P. Y., Deviawati, D., Maulani, N. F., & Shidiq, M. J. (2022). Perubahan Perilaku Konsumen Pasar Tradisional & Pasar Modern Dimasa Pandemic dan Masa “New Normal” di Kabupaten Garut. *Jurnal Ekobistek*, 258–264. <https://doi.org/10.35134/ekobistek.v11i4.362>
- Fauzi, C. B., Mandey, S. L., & Rotinsulu, J. J. (2022). Analisis Strategi Pemasaran Untuk Meningkatkan Daya Saing di Coffee Shop Pulang Manado. *Jurnal EMBA : Jurnal Riset Ekonomi, Manajemen, Bisnis Dan Akuntansi*, 10(2), 51. <https://doi.org/10.35794/emba.v10i2.39392>
- Galahartlambang, Y., Khotiah, T., & Jumain, J. (2022). Pemetaan Konsumen Berdasarkan Perilaku Transaksi Produk Ritel Pada Distributor XYZ Semarang. *Journal Of Electrical Engineering And Technology*, 3(2).
- Guo, M., Liao, X., Liu, J., & Zhang, Q. (2020). Consumer preference analysis: A data-driven multiple criteria approach integrating online information. *Omega*, 96, 102074. <https://doi.org/10.1016/j.omega.2019.05.010>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.
- Han, J., Pei, J., & Tong, H. (2023). *Data Mining: Concepts and Techniques*.
- Hendarwan, D. (2023). Analysis of Market Driven Strategies to Increase Capabilities and Performances Advantages in Business. *Journal of Economic Empowerment Strategy (JEES)*, 6(1), 1–14. <https://doi.org/10.23969/jees.v6i1.5414>
- Hermawan, A., Maesaroh, S. S., & Purwaamijaya, B. M. (2020). Digital marketing strategy based on user experience to increase user growth and engagement start-up in Tasikmalaya. In *Advances in Business, Management and Entrepreneurship* (1st ed., pp. 200–203). CRC Press.

- Jeffery, M. (2010). *Data-Driven Marketing The 15 Metrics Everyone in Marketing Should Know Includes Downloadable ROI Spreadsheets Kellogg School of Management*.
- Kahar, R. M., Palisuri, P., & Abubakar, H. (2024). Analisis Biaya Operasional Terhadap Profitabilitas Pada PT. Sejahtera Buana Raya. *Journal of Economy Business Development*, 2(2), 171–176. <https://doi.org/10.56326/jebd.v2i2.2739>
- Kotler, P., & Keller, K. L. (2016). *Marketing Management (15th ed.)*. Pearson Education.
- Kotler, Philip., & Keller, K. Lane. (2012). *Marketing management*. Prentice Hall.
- Laporan Keuangan Soulja Coffee (Data Internal). (2025). *Laporan Keuangan dan Data Penjualan Periode Desember 2024 – Maret 2025*. Data Internal.
- Mahadewi, E. P., & Iswara, A. (2025). Strategi Pemasaran Berbasis Data (Data-Driven Marketing) untuk Meningkatkan Kinerja Penjualan: Studi Empiris di Perusahaan Startup. *Jurnal Ekonomi Utama*, 4(1), 98–107. <https://doi.org/10.55903/juria.v4i1.245>
- Octiva, C. S., Fajri, T. I., Sulistiarini, E. B., Suharjo, S., & Nuryanto, U. W. (2024). Penggunaan Teknik Data Mining untuk Analisis Perilaku Pengguna pada Media Sosial. *Jurnal Minfo Polgan*, 13(1), 1074–1078. <https://doi.org/10.33395/jmp.v13i1.13936>
- Open Data Kota Tasikmalaya. (2025). *Jumlah Coffee Shop dan Cafe di Kota Tasikmalaya*. <https://opendata.tasikmalayakota.go.id/Dataset/Jumlah-Coffee-Shop-Dan-Cafe-Di-Kota-Tasikmalaya>.
- Pangestu, A. D., & Aribowo, A. (2023). Pengaruh Komponen Preferensi Terhadap Keputusan Pembelian Konsumen Coffee Shop di Kota Bandung. *Jurnal Ilmiah Manajemen, Ekonomi, & Akuntansi (MEA)*, 7(3), 813–827. <https://doi.org/10.31955/mea.v7i3.3405>
- Prior, D. D., Buttle, F., & Maklan, S. (2024). *Customer Relationship Management: Concepts, Applications and Technologies*. Taylor & Francis, 2024.
- Sari, M. D. (2023). *Perilaku Konsumen*. Uwais Inspirasi.
- Sarker, A., Shamim, S. M., Zaman, Dr. Md. S., & Rahman, Md. M. (2018). *Employee's Performance Analysis and Prediction Using K-means Clustering & Decision Tree Algorithm*. <https://www.researchgate.net/publication/326258764>
- Schiffman, L. G. (2015). *Consumer Behavior, 11th edition*. Harlow : Pearson Education., 2015.
- Setiadi, N. J. (2023). *Perilaku Konsumen : Persektif Komtemporer pada motif, tujuan, dan keinginan konsumen (Revisi)*. Kencana Prenada Media Group.
- Snapcart. (2023). *Indonesia's Coffee Consumption Trends in 2023*. Snapchart Global.
- Wulandari, R. A., Indah, P. N., & Laily, D. W. (2024). Analisis Pengaruh Segmentasi Pasar Terhadap Peningkatan Volume Penjualan Pada Swakarja Coffe Space di Gresik. *Berkala Ilmiah AGRIDEVINA*, 13(2). <https://doi.org/10.33005/agridevina.v13i2.4532>