

# **ANALISIS DATA PENJUALAN DAN SEGMENTASI CABANG FRANCHISE XYZ MENGGUNAKAN METODE K-MEANS CLUSTERING**

**Fransiskus Xaverius Prasetyo Satriatama<sup>1</sup>**

<sup>1</sup>Prodi Teknik Informatika S2, Universitas Pamulang, Banten  
e-mail: <sup>1</sup>riosatriatama0801@gmail.com

**ABSTRAK:** Perkembangan bisnis waralaba menuntut setiap cabang untuk mampu mengelola data penjualan secara efektif agar dapat meningkatkan strategi pemasaran serta kinerja operasional. Penelitian ini bertujuan untuk menganalisis data penjualan pada beberapa cabang waralaba XYZ serta melakukan segmentasi cabang berdasarkan pola penjualan menggunakan metode K-Means Clustering. Data penjualan yang digunakan mencakup periode tertentu dengan variabel utama seperti total transaksi, jumlah pelanggan, dan rata-rata nilai pembelian. Proses analisis meliputi tahap pra-pemrosesan data, pemilihan jumlah klaster optimal menggunakan Elbow Method, serta penerapan algoritma K-Means untuk membentuk kelompok cabang dengan karakteristik serupa. Hasil penelitian menunjukkan terbentuknya beberapa klaster yang merepresentasikan perbedaan performa cabang, seperti klaster berkinerja tinggi, menengah, dan rendah. Temuan ini dapat menjadi dasar bagi manajemen waralaba XYZ dalam merancang strategi peningkatan penjualan yang lebih tepat sasaran, termasuk alokasi sumber daya, pengembangan promosi, serta pengambilan keputusan bisnis secara lebih terukur.

**Kata kunci:** *Data Penjualan, Segmentasi Cabang, K-Means Clustering, Waralaba, Analisis Data*

**ABSTRACT:** The development of the franchise business requires each branch to be able to manage sales data effectively in order to improve marketing strategies and operational performance. This study aims to analyze sales data at several XYZ franchise branches and segment branches based on sales patterns using the K-Means Clustering method. The sales data used covers a certain period with key variables such as total transactions, number of customers, and average purchase value. The analysis process includes data pre-processing, selecting the optimal number of clusters using the Elbow Method, and applying the K-Means algorithm to form groups of branches with similar characteristics. The results show the formation of several clusters that represent differences in branch performance, such as high-, medium-, and low-performance clusters. These findings can be a basis for XYZ franchise management in designing more targeted sales improvement strategies, including resource allocation, promotion development, and more measurable business decision-making.

**Keywords:** *Sales Data, Branch Segmentation, K-Means Clustering, Franchise, Data Analysis*

## **PENDAHULUAN**

Dalam industri penjualan online, memperkuat hubungan antara pelanggan adalah tujuan utama perusahaan untuk memperoleh keuntungan yang signifikan dalam persaingan pasar. Perusahaan distribusi perlu mengidentifikasi pelanggan terbaik dan meningkatkan pemahaman mereka tentang kebutuhan pelanggan agar dapat mempertahankan loyalitas pelanggan. Era yang serba digital ini, penjualan online telah mengubah cara berbelanja, hingga perusahaan e commerce perlu memahami pelanggan mereka secara mendalam untuk menyediakan pengalaman yang relevan dan memuaskan. Selain itu, dalam persaingan bisnis yang semakin ketat, perusahaan harus menjaga hubungan yang baik dengan pelanggan. Salah satu strategi yang efektif dalam menjaga hubungan tersebut adalah melalui segmentasi pelanggan, yang melibatkan pengelompokan pelanggan dengan kesamaan tertentu. Dalam beberapa tahun terakhir, ranah bisnis online telah mengalami peningkatan pesat dalam skala global, dan atmosfer pasar secara perlahan tapi pasti mencapai tingkat kematangan. Untuk meningkatkan kehadiran

bisnis online dalam ranah pasar ini, perusahaan bisnis online harus menemukan cara untuk meningkatkan kemampuan mereka sendiri. (Deng, 2023). Kepuasan dan kepercayaan pelanggan menjadi salah satu hal yang harus dijaga untuk mempertahankan pelanggan yang loyal (Febrianti & Beni, 2023).

Segmentasi pelanggan adalah pendekatan yang membantu perusahaan membagi basis pelanggan mereka menjadi kelompok yang berbeda, berdasarkan karakteristik dan perilaku yang serupa. Tujuan utama dari segmentasi pelanggan adalah untuk memahami perilaku pelanggan, menganalisis preferensi mereka, dan mengidentifikasi kelompok pelanggan yang memiliki kebutuhan serupa. Segmentasi pelanggan merupakan langkah awal yang memungkinkan perusahaan mengembangkan strategi yang lebih nyaman dan personalisasi. Setelah perusahaan berhasil mengidentifikasi kelompok pelanggan yang berbeda, langkah berikutnya adalah mengembangkan rencana aksi yang sesuai untuk masing-masing kelompok tersebut. Misalnya, perusahaan dapat mengadopsi strategi pemasaran yang berbeda untuk setiap kelompok pelanggan. Dengan perkembangan teknologi dan ketersediaan data yang semakin besar, perusahaan juga dapat menggabungkan metode segmentasi pelanggan dengan pendekatan analitik yang lebih canggih, seperti analisis prediktif dan mesin pembelajaran. Dengan memanfaatkan teknik-teknik ini, perusahaan dapat menghasilkan wawasan yang lebih mendalam tentang perilaku pelanggan, memprediksi kebutuhan dan preferensi pelanggan di masa mendatang.

Salah satu metode yang efektif dalam segmentasi pelanggan adalah metode K-means clustering. Metode ini telah banyak digunakan dalam berbagai penelitian dan aplikasi bisnis. Misalnya, dalam penelitian Zhang dkk, mereka mengaplikasikan metode K-means clustering untuk mengelompokkan pelanggan penjualan secara online berdasarkan atribut seperti perilaku pembelian, preferensi harga, dan demografi. Hasilnya menunjukkan bahwa segmentasi pelanggan dapat membantu perusahaan e-commerce dalam menyusun strategi pemasaran yang lebih efektif (Zhang et al., 2020). Algoritma K-means akan mengelompokkan satuan data dalam suatu kumpulan data ke dalam berbagai cluster berdasarkan jarak terdekat dengan nilai centroid awal yang dipilih secara acak, yang merupakan titik pusat awal. Semua data ini akan digunakan untuk menghitung jarak menggunakan rumus Euclidean Distance. Data yang jaraknya dekat dengan centroid akan membuat cluster dan proses ini terus berlanjut sampai tidak ada perubahan pada setiap kelompok (Gupta et al., 2013).

Algoritma K-means clustering bekerja dengan cara membagi data menjadi kelompok-kelompok yang berbeda berdasarkan kemiripan atribut yang dimiliki. Dalam konteks penjualan online, K means clustering dapat digunakan untuk mengidentifikasi kelompok pelanggan dengan preferensi produk, perilaku pembelian, preferensi harga, dan faktor-faktor lain yang relevan. Metode ini memungkinkan perusahaan untuk memahami kelompok-kelompok pelanggan dengan lebih baik dan menyediakan layanan yang sesuai dengan kebutuhan dan preferensi masing kelompok (Istiana, 2013).

Dengan memahami dan mengimplementasikan metode segmentasi pelanggan menggunakan Kmeans clustering, perusahaan dapat memperoleh wawasan yang lebih dalam tentang perilaku pelanggan mereka. Hal ini akan membantu perusahaan dapat menargetkan segmentasi customer secara lebih efisien. Segmentasi dapat dilakukan dengan menerapkan metode Kmeans clustering yang melibatkan beberapa kriteria yaitu total transaksi, frekuensi transaksi, dan total jumlah pembelian barang pada periode tertentu. Berdasarkan uraian-uraian di atas mengenai permasalahan pada penjualan online, penulis memilih judul “Analisis Data Penjualan dan Segmentasi Cabang Franchise XYZ Menggunakan Metode K-Means Clustering”.

## METODA

Dalam Jurnal ini, metode penelitian yang digunakan adalah metode deskriptif. Pendekatan deskriptif digunakan untuk menjelaskan dan menganalisis penggunaan metode K-means clustering (Nawangsih, 2023) dalam segmentasi pelanggan Franchise. Gambar 1 adalah gambar alur metode penelitian yang dilakukan.



Gambar 1 Prosedur Penelitian

a. Studi Literatur

Tahap ini bertujuan untuk mengumpulkan referensi dan teori yang relevan dengan topik penelitian. Peneliti mempelajari hasil penelitian terdahulu, metode yang digunakan, serta celah penelitian yang masih bisa dikembangkan. Studi literatur membantu memperkuat dasar teori dan menentukan pendekatan yang tepat untuk penelitian.

b. Pengumpulan Data

Setelah memahami landasan teori, langkah berikutnya adalah mengumpulkan data yang dibutuhkan. Data dapat diperoleh dari berbagai sumber seperti survei, observasi, basis data publik, maupun dokumen internal. Kualitas dan relevansi data sangat menentukan keberhasilan analisis pada tahap berikutnya.

c. Pra Pemrosesan Data

Data yang telah dikumpulkan biasanya masih mengandung ketidaksesuaian, data hilang, atau noise. Oleh karena itu, dilakukan tahap pra-pemrosesan untuk membersihkan, menormalkan, dan menyiapkan data agar siap digunakan untuk proses analisis. Tahapan ini mencakup seleksi atribut, transformasi data, serta penanganan data yang tidak lengkap.

d. Clustering

Pada tahap ini, data yang sudah siap diproses dianalisis menggunakan algoritma clustering untuk mengelompokkan data berdasarkan kesamaan karakteristik. Tujuannya adalah menemukan pola atau struktur tersembunyi dalam data tanpa menggunakan label kelas. Metode yang digunakan

di penelitian ini berupa K-Means.

#### e. Validasi dan Evaluasi Hasil

Tahap terakhir adalah mengevaluasi hasil clustering yang diperoleh untuk memastikan kualitas dan akurasi pengelompokan. Validasi dapat dilakukan dengan metode internal (misalnya Silhouette Coefficient) atau eksternal (membandingkan dengan data acuan jika tersedia). Hasil evaluasi ini digunakan untuk menilai efektivitas metode yang diterapkan dan memberikan kesimpulan terhadap penelitian.

## HASIL DAN PEMBAHASAN

### a. Studi Literatur

Tahap studi literatur dilakukan untuk memperoleh pemahaman mendalam mengenai konsep, teori, serta penelitian terdahulu yang berkaitan dengan analisis data dan penerapan metode clustering pada bidang Food and Beverage (F&B). Pada tahap ini, peneliti mengkaji berbagai sumber ilmiah seperti jurnal, buku, maupun laporan penelitian yang membahas tentang pengelolaan data pelanggan, pola pembelian, dan strategi pemasaran berbasis data.

### b. Pengumpulan Data

Tahap pengumpulan data dilakukan untuk memperoleh sumber data yang akan dianalisis dalam penelitian. Pada penelitian ini, data yang digunakan berupa data dummy yang disimulasikan berdasarkan karakteristik operasional salah satu waralaba F&B di Indonesia. Data tersebut mencerminkan aktivitas bisnis sehari-hari seperti transaksi penjualan, jenis produk yang dibeli, waktu pembelian, jumlah transaksi, serta profil pelanggan.

No Kebun 2	Nama Kebun	Kota / Kabupaten	Provinsi	Region	Generation
0	#5 Kebun 0048 Cempaka Putih Avenue	Tangerang Selatan	Banten	Jabodetabek	1.0
1	01 Kebun 0001 Putra Kencana Lane	Bengkulu	Jawa Barat	Jawa	1.0
2	09 Kebun 0006 Ananta Sakura Street	Bogor	Jawa Barat	Jabodetabek	1.0
3	88 Kebun 0098 Supi Rimbai Lane	Jakarta Selatan	Jakarta	Jabodetabek	1.0
4	117 Kebun 0017 Perada Kencana Lane	Pekalongan	Sumsel	Sumsel	1.0

Tabel 1 Hasil Pengumpulan Data

### c. Pra Pemrosesan Data

```
*** CEK MISSING VALUES DATA KEBUN ***
No Kebun 2      0
Nama Kebun      0
Kota / Kabupaten 0
Provinsi        0
Region          0
Generation       0
dtype: int64

*** CEK MISSING VALUES SALES FEBRUARI ***
Outlet Name     0
Payment Method   0
Transaction      0
Transaction Cancel  836
Amount Received   0
dtype: int64
```

Tabel 2 Hasil Pra Pemrosesan Data

Pada tahap pemeriksaan data, dilakukan pengecekan terhadap nilai yang hilang (missing values) untuk memastikan kualitas data sebelum proses analisis lebih lanjut. Pemeriksaan dilakukan pada dua dataset, yaitu Data Kebun dan Sales Februari.

Hasil pengecekan menunjukkan bahwa Data Kebun tidak memiliki missing values pada seluruh variable yang ada, yaitu No Kebun 2, Nama Kebun, Kota/Kabupaten, Provinsi, Region, dan Generation. Hal ini menandakan bahwa data kebun telah terisi secara lengkap dan siap

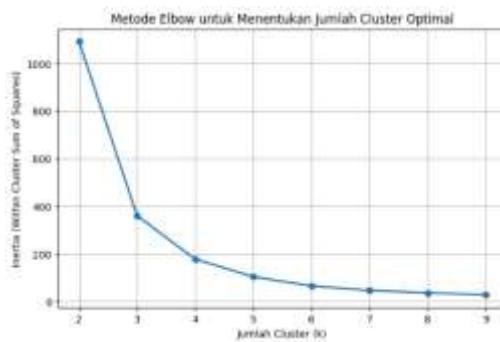
digunakan untuk tahap analisis berikutnya.

Sementara itu, pada dataset Sales Februari, sebagian besar variabel juga tidak memiliki nilai yang hilang, seperti Outlet Name, Payment Method, Transaction, dan Amount Received. Namun, ditemukan sebanyak 836 data yang hilang pada variabel Transaction Cancel. Nilai yang hilang ini mengindikasikan adanya ketidaksesuaian atau kekosongan data pada transaksi pembatalan, yang perlu ditangani terlebih dahulu melalui proses data cleaning sebelum dilakukan analisis lanjutan.

d. Clustering K-Means

Berdasarkan grafik di bawah, terlihat bahwa penurunan nilai inertia terjadi sangat signifikan dari  $k = 2$  hingga  $k = 3$ , kemudian menurun lebih landai mulai dari  $k = 4$  dan seterusnya. Titik tekuk (elbow point) yang terbentuk pada grafik berada di sekitar  $k = 3$ , yang menandakan bahwa jumlah cluster optimal untuk data ini adalah 3 cluster.

Pemilihan jumlah cluster ini memberikan keseimbilan antara kesederhanaan model dan tingkat variasi data yang masih dapat dijelaskan dengan baik. Dengan demikian, pada tahap analisis selanjutnya akan digunakan 3 cluster utama untuk mengelompokkan data sesuai karakteristik yang dimiliki masing-masing kelompok.



Gambar 2 Hasil Visualisasi Metode Elbow dalam K-Means

e. Validasi dan Evaluasi Hasil

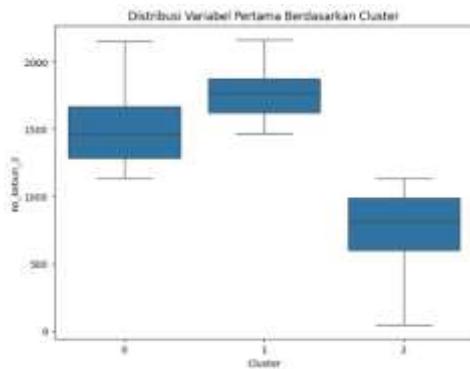
1. Boxplot

Gambar di bawah menunjukkan hasil visualisasi distribusi variabel pertama (no\_kebun\_2) berdasarkan tiga cluster yang telah terbentuk dari proses *clustering*. Visualisasi ini menggunakan boxplot untuk menggambarkan persebaran data, nilai tengah (*median*), serta rentang antar kuartil pada masing-masing cluster

Dari grafik tersebut dapat dilihat bahwa terdapat perbedaan distribusi yang cukup jelas antara ketiga cluster:

- a. **Cluster 1 (label 0)** memiliki nilai tengah yang berada di kisaran menengah, dengan sebaran data yang cukup luas. Hal ini menunjukkan bahwa cluster ini terdiri dari kebun dengan nomor identifikasi yang bervariasi, namun cenderung berada pada kelompok nilai sedang.
- b. **Cluster 2 (label 1)** memperlihatkan nilai median yang lebih tinggi dibandingkan dua cluster lainnya. Distribusinya relatif lebih sempit, menandakan bahwa data dalam cluster ini lebih homogen dan berada pada kisaran nilai *no\_kebun\_2* yang tinggi.
- c. **Cluster 3 (label 2)** memiliki nilai median yang paling rendah dan rentang sebaran data yang cukup lebar. Hal ini menunjukkan bahwa cluster ini didominasi oleh kebun dengan nilai identifikasi yang lebih kecil dan bervariasi.

Secara keseluruhan, perbandingan ketiga boxplot ini memperlihatkan bahwa pembentukan cluster telah berhasil memisahkan data berdasarkan perbedaan karakteristik numerik pada variabel no\_kebun\_2. Hasil ini menjadi dasar untuk interpretasi lebih lanjut mengenai profil masing-masing cluster dalam konteks operasional kebun atau unit usaha yang dianalisis.



Gambar 3 Hasil Visualisasi Boxplot

## 2. Heatmap



Gambar 4 Hasil Visualisasi Heatmap

Gambar di atas memperlihatkan hasil visualisasi berupa heatmap rata-rata tiap fitur per cluster. Visualisasi ini digunakan untuk memahami karakteristik utama dari masing-masing cluster berdasarkan nilai rata-rata dari variabel yang membentuk model klasterisasi.

Dari heatmap tersebut terlihat bahwa:

- Cluster 0 memiliki nilai rata-rata no\_kebun\_2 sebesar 1502,15, dengan generation sebesar 1, serta nilai komponen utama (PCA1 dan PCA2) yang relatif moderat, yaitu 0,55 dan 0,67. Hal ini menunjukkan bahwa cluster ini berada pada tingkat menengah baik dari sisi identifikasi kebun maupun karakteristik variabel turunannya.
- Cluster 1 menunjukkan nilai tertinggi pada fitur no\_kebun\_2 (1784,88) dan generation (2), serta nilai PCA1 positif besar (9,36) dan PCA2 negatif besar (-7,24). Kondisi ini menggambarkan bahwa cluster ini mewakili kelompok dengan kapasitas atau skala kebun yang lebih besar, serta memiliki karakteristik yang berbeda secara signifikan dari dua cluster lainnya.
- Cluster 2 memiliki nilai rata-rata no\_kebun\_2 paling rendah (768,87) dan generation sebesar 1, dengan nilai PCA1 dan PCA2 yang rendah pula (-0,60 dan -0,48). Hal ini menunjukkan bahwa cluster ini mencerminkan kelompok dengan skala kebun yang kecil

serta memiliki karakteristik yang cenderung homogen.

Dari hasil visualisasi heatmap ini dapat disimpulkan bahwa Cluster 1 merupakan kelompok dengan performa atau kapasitas tertinggi, sedangkan Cluster 2 mewakili kelompok dengan kapasitas terendah, dan Cluster 0 berada di posisi menengah. Pola ini memperkuat hasil sebelumnya dari analisis Elbow Method dan boxplot, yang menunjukkan keberhasilan proses klasterisasi dalam membedakan kelompok data berdasarkan perbedaan karakteristik numerik tiap fitur.

## SIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan terhadap dataset Data Kebun dan Sales Februari 2025, dengan tahapan mulai dari pra-pemrosesan data, analisis clustering menggunakan metode K-Means, hingga validasi hasil, diperoleh beberapa kesimpulan sebagai berikut:

- a. Tahap pra-pemrosesan data berhasil membersihkan dan menstandarkan data yang digunakan. Seluruh nilai kosong telah ditangani, tipe data diperbaiki, serta nama kolom diseragamkan sehingga dataset siap untuk analisis lebih lanjut.
- b. Melalui metode K-Means Clustering, data berhasil dikelompokkan menjadi tiga cluster utama ( $k=3$ ) yang menggambarkan perbedaan karakteristik antar kebun berdasarkan produktivitas dan kinerja penjualan:
  1. Cluster 0: Kebun dengan produktivitas dan penjualan rendah.
  2. Cluster 1: Kebun dengan performa sedang dan stabil.
  3. Cluster 2: Kebun dengan hasil dan penjualan tinggi, yang dapat dikategorikan sebagai kelompok unggulan.
- c. Hasil evaluasi menggunakan Silhouette Score dan Davies-Bouldin Index menunjukkan bahwa model K-Means mampu membentuk cluster dengan kualitas yang cukup baik, di mana data antar cluster relatif terpisah dengan jelas dan memiliki homogenitas internal yang tinggi.
- d. Visualisasi menggunakan PCA dan heatmap memperkuat hasil bahwa masing-masing cluster memiliki ciri khas tersendiri yang dapat digunakan untuk pengambilan keputusan strategis.
- e. Dari sisi implementasi bisnis, hasil pengelompokan ini dapat digunakan untuk:
  1. Menentukan prioritas pengelolaan dan distribusi hasil panen,
  2. Mengidentifikasi kebun unggulan untuk dijadikan contoh praktik terbaik,
  3. Serta menyusun strategi peningkatan performa bagi kebun yang berada di cluster produktivitas rendah.

## DAFTAR PUSTAKA

- Angelie, A. V. (2017). Segmentasi Pelanggan Menggunakan Clustering K-Means dan Model RFM (Studi Kasus: PT. Bina Adidaya Surabaya) [Institut Teknologi Sepuluh Nopember].  
<https://repository.its.ac.id/42240/>
- Deng, Y. (2023). Specific Strategies for Innovating Marketing Models of E-commerce Enterprises in the Internet Era. Academic Journal of Business & Management, 5(13), 22–26.  
<https://doi.org/10.25236/AJBM.2023.051304>
- Febrianti, F., & Beni, S. (2023). Strategi Mempertahankan Loyalitas Pelanggan pada Usaha Kuliner di Kecamatan Bengkayang. Inovasi Pembangunan : Jurnal Kelitbang, 11(02), 189–210.  
<https://doi.org/10.35450/jip.v11i02.384>
- Gupta, G. K., Agrawal, D., Singh, R. K., & Arya, R. K. (2013). Prevalence, Risk Factors and Socio Demographic Co-Relates of Adolescent Hypertension in District Ghaziabad. Indian Journal of

- Community Health, 25(3), 293–298.  
<https://www.iapsmupuk.org/journal/index.php/IJCH/article/view/331>
- Istiana, M. I. (2013). Segmentasi Pelanggan Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran pada LAROIBA Seluler Oleh: Mike Indra Istiana [Universitas Dian Nuswantoro Semarang]. [http://eprints.dinus.ac.id/12733/2/abstrak\\_12903.pdf](http://eprints.dinus.ac.id/12733/2/abstrak_12903.pdf)
- Nawangsih, I. (2023). Analisa Penjualan Produk Kosmetik Dengan Metode Algoritma K-Means Di Toko Erremy. Bulletin of Information Technology (BIT), 4(1), 140–145.  
<https://doi.org/10.47065/bit.v4i1.468>
- Zhang, Z., Ni, G., & Xu, Y. (2020). Comparison of Trajectory Clustering Methods based on K means and DBSCAN. Proceedings of 2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence, ICIBA 2020, 557–561.  
<https://doi.org/10.1109/ICIBA50161.2020.9277214>
- Ketherin, B. E., Arifiyanti, A. A., & Sodik, A. (n.d.). *Analisa Segmentasi Konsumen Menggunakan Algoritma K-Means Clustering Jurusan Sistem Informasi*, Institut Teknologi Adhi Tama Surabaya.