

# Analisis *Cluster K-Means* dengan Metode *Elbow* untuk Menentukan Pola Penjualan Produk Traffic Room Summarecon Mal Bekasi

Ajie Prasetya<sup>1</sup>, Ratna Salkiawati<sup>1,\*</sup>, Allan D Alexander<sup>1</sup>

\* Korespondensi: e-mail: [ratna\\_tind@dsn.ubharajaya.ac.id](mailto:ratna_tind@dsn.ubharajaya.ac.id)

<sup>1</sup> Informatika; Universitas Bhayangkara Jakarta Raya; Jalan Raya Perjuangan, Marga Mulya, Bekasi Utara, Jawa Barat 17121. Telp: 021-88955882; e-mail: [ajie.prasetya19@mhs.ubharajaya.ac.id](mailto:ajie.prasetya19@mhs.ubharajaya.ac.id), [ratna\\_tind@dsn.ubharajaya.ac.id](mailto:ratna_tind@dsn.ubharajaya.ac.id), [allan@ubharajaya.ac.id](mailto:allan@ubharajaya.ac.id)

Submitted : 12 April 2023  
Revised : 25 April 2023  
Accepted : 19 Mei 2023  
Published : 30 Mei 2023

## Abstract

An effective sales strategy in the fashion retail business is essential to determine the success of the company or store. Like the Traffic Room store, which is a vintage fashion retail store that sells a variety of products. Although there are many products on sale, this store has not utilized sales data to determine product sales patterns, causing negative impacts such as there are still many products that are in short supply and products are not sold with predetermined targets. So the purpose of this study is to determine product sales patterns in order to improve product inventory. To solve this problem, the analysis used is the K-Means algorithm to find product sales patterns assisted by the elbow method in determining the optimal cluster. As well as the flow in this research process is the CRISP-DM method with steps namely business understanding, data understanding, data preparation, modeling, evaluation and deployment. The results of this study obtained 4 clusters, namely cluster 2 or very in demand there are 2 products, cluster 3 or in demand there are 5 products, cluster 1 or quite in demand there are 5 products and cluster 4 or less in demand there are 3 products. The evaluation results get the optimal Sum of Square Error (SSE) value of 594,366.733 or 65.5%. From the evaluation results, it means that the performance of the K-Means algorithm used is good.

**Keywords:** CRISP-DM, Elbow Method, K-Means Algorithm, Product Sales Pattern, Sum of Square Error (SSE)

## Abstrak

Strategi penjualan yang efektif dalam bisnis ritel fashion sangatlah penting untuk menentukan keberhasilan perusahaan atau toko. Seperti toko Traffic Room yaitu toko ritel fashion vintage yang menjual berbagai macam produk. Walaupun banyaknya produk yang di jual, toko ini belum memanfaatkan data penjualan untuk menentukan pola penjualan produk sehingga menimbulkan dampak negatif seperti masih banyak produk yang kekurangan persediaan dan produk tidak terjual dengan target yang sudah ditentukan. Maka tujuan dari penelitian ini untuk mengetahui pola penjualan produk agar bisa memperbaiki persediaan produk. Untuk mengatasi permasalahan ini, analisis yang digunakan yaitu algoritma K-Means untuk mencari pola penjualan produk dibantu dengan metode elbow dalam menentukan cluster yang optimal. Serta yang menjadi alur dalam proses penelitian ini yaitu metode CRISP-DM dengan langkah-langkahnya yakni business understanding, data understanding, data preparation, modeling, evaluation dan deployment. Hasil dari penelitian ini mendapatkan 4 cluster yaitu cluster 2 atau sangat laris ada 2 produk, cluster 3 atau laris ada 5 produk, cluster 1 atau cukup laris

ada 5 produk dan *cluster* 4 atau kurang laris ada 3 produk. Hasil evaluasi mendapatkan nilai *Sum of Square Error* (SSE) optimal yaitu 594.366,733 atau 65,5%. Dari hasil evaluasi artinya kinerja algoritma *K-Means* yang digunakan sudah baik.

**Kata kunci:** Algoritma *K-Means*, CRISP-DM, Metode *Elbow*, Pola Penjualan Produk, *Sum of Square Error* (SSE)

---

## 1. Pendahuluan

Retail *fashion* atau bisnis ritel pakaian adalah bisnis yang menjual produk-produk pakaian dan aksesoris kepada konsumen. Bisnis ritel *fashion* sangat dipengaruhi oleh tren dan gaya hidup yang berubah-ubah, sehingga pelaku bisnis harus selalu memantau tren terbaru dan menghadirkan produk-produk yang inovatif dan menarik bagi konsumen (Ferdian et al., 2021). Di Indonesia bisnis ritel *fashion* terus berkembang dan memiliki potensi yang besar di masa depan, seiring dengan pertumbuhan ekonomi dan meningkatnya daya beli masyarakat. Namun, persaingan di bisnis ritel *fashion* juga sangat ketat, sehingga pelaku bisnis harus bisa membuat strategi yang tepat dan membedakan diri dari pesaing agar bisa bersaing dan berhasil di pasar (Sono et al., 2023). Pertumbuhan penjualan ritel di Indonesia naik 1,7 % dari tahun 2011 – 2023, yang sebelumnya turun 0,7 % di tahun 2022 (CEIC DATA, 2023). Bahwa banyaknya toko ritel yang menjual produk *fashion* di Indonesia. salah satunya adalah Traffic Room.

Traffic Room merupakan salah satu toko pakaian *brand* lokal dengan konsep *vintage*. *Vintage* adalah *tagline* Traffic Room karena konsepnya menawarkan gaya klasik, nyaman dan tahan lama yang dapat dilihat dari desain *vintage* seperti otomotif klasik, band legendaris dan lain-lain. Produk yang tersedia antara lain adalah *T-Shirt*, *crewneck*, *long sleeve*, jaket, celana, tas dan lain-lain.

Banyaknya produk yang di jual menjadikan Toko Traffic Room harus menjualkan produk sesuai target yang sudah ditetapkan. Tetapi masalah yang dialami yaitu kurangnya persediaan produk dalam penjualan dan berlebihnya produk yang kurang diminati sehingga berdampak kerugian bagi Traffic Room. Bila dilihat dari konsep jual beli Traffic Room sendiri memang sudah menggunakan teknologi komputer sebagai alat penginputan data serta berfungsi untuk pencetakan/*print out*. Namun dari data penjualan Traffic Room selama ini belum tersusun dengan baik dan belum dimanfaatkan untuk pengembangan pola penjualan produk, sehingga mengakibatkan dalam membuat promosi-promosi produk masih memakai cara melihat dari persediaan produk yang ada di toko.

Dari beberapa permasalahan, Traffic Room berharap adanya cara untuk mengetahui pola penjualan produk agar dapat diketahui produk yang harus ditingkatkan atau dikurangkan dalam produksi. Untuk memenuhi keinginan Toko Traffic Room yaitu ada beberapa cara salah satunya melakukan pemanfaatan pengelompokan produk penjualan memakai Algoritma *K-Means* dan metode *elbow* untuk proses evaluasi dan juga berfungsi untuk menentukan nilai *k* optimal dari perulangan proses uji nilai *k*.

Maka pengelompokan penjualan produk memakai *K-Means* bisa menjadi jalan keluar dari beberapa masalah yang ada di Traffic Room. Harapan pengelompokan ini untuk mengetahui pola penjualan produk Traffic Room yang bertujuan untuk memudahkan pengelola toko dalam melakukan persediaan produk yang akan di jual.

### **State of The Art**

Penelitian berjudul penerapan *K-Means clustering* dan *cross-industry standard process for data mining* (CRISP-DM) untuk Mengelompokan Penjualan Kue. Penggunaan algoritma *K-Means* ditujukan untuk membantu produsen agar dapat memprediksi jenis kue yang laris atau habis pada waktu tertentu. Hasil dari penelitian memakai tiga *cluster* yaitu penjualan rendah, menengah dan tinggi. Dari memprediksi jenis kue ini harapannya membantu produsen untuk mempersiapkan produknya dengan lebih efektif dan efisien sehingga mengurangi hasil produksi yang terbuang percuma (Muttaqin et al., 2022).

Penelitian selanjutnya berjudul optimasi performa *cluster K-Means* menggunakan *Sum of Square Error (SSE)*. Penggunaan pada algoritma ini bertujuan untuk mengelompokan berdasarkan wilayah, pekerjaan dan umur pasien dari hasil perhitungan *Sum of Square Error (SSE)*. Hasil penelitian, penggunaan algoritma *K-Means* mendapatkan tiga *cluster* yang nantinya akan pakai untuk pihak rumah sakit sehingga dapat mengetahui informasi lebih lanjut tentang pasien (Nainggolan & Lumbantoruan, 2018).

Penelitian berjudul penerapan algoritma *K-Means* untuk mengetahui pola persediaan barang pada Toko Raja Bekasi. Penggunaan algoritma *K-Means* bertujuan mengelompokan produk untuk penentuan produksi persediaan barang, dan mengelompokkan barang yang ada di Toko Raja. Hasil penelitian, penggunaan algoritma *K-Means* mendapatkan tiga *cluster*. Sehingga bisa mengetahui seberapa besar hasil dari pengelompokkan barang berpengaruh terhadap kebutuhan dari konsumen sehingga tidak adanya lagi barang yang kelebihan persediaan atau kurang persediaan (Safira et al., 2022).

## **2. Metode Penelitian**

### **2.1. Data mining**

Merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. *Data mining* mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data. *Data mining* bertujuan untuk menemukan pola yang sebelumnya tidak diketahui. Jika pola-pola tersebut telah diperoleh maka dapat digunakan untuk menyelesaikan berbagai macam permasalahan (Arhami & Nasir, 2020).

### **2.2. K-Means**

Salah satu metode *clustering* yang menentukan jumlah k untuk mengelompokan data ke dalam k kelompok. Tujuan *K-Means* yaitu untuk mengoptimalkan data dan *centroid* kelompoknya, sehingga dapat meminimalisir berbedanya antara satu kelompok dan memaksimalkan perbedaan data dengan kelompok lain (Werdiningsih et al., 2022). Untuk menghitung jarak antara objek dengan *centroid* dapat menggunakan *Euclidean Distance* pada rumus 1.

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, i = 1, 2, 3, \dots, n \quad (1)$$

Dimana:

$x_i$  = objek x ke- $i$ ,

$y_i$  = objek y ke- $i$  dan

$n$  = banyaknya objek.

### 2.3. CRISP-DM

Singkatan dari *Cross Industry Standard for Data mining*, yaitu sebuah metode yang umum digunakan dengan menggunakan proses pemodelan data di dalamnya. Tujuan dari metode CRISP-DM yaitu membantu menemukan pola yang menarik untuk mengembangkan model pada data yang digunakan (Zuhri & Harani, 2023). Ada beberapa proses CRISP-DM yaitu *business understanding, data understanding, data preparation, modeling, evaluation* dan *deployment*. metode ini membantu memastikan bahwa proses pengolahan data dilakukan secara terstruktur dan sistematis, sehingga dapat menghasilkan hasil analisis yang lebih akurat dan dapat diandalkan (Purnomo et al., 2023).

### 2.4. Metode *Elbow*

Metode *elbow* adalah metode yang digunakan dalam *clustering* untuk menentukan jumlah *cluster* optimal atau terbaik. Metode ini dinamakan "*Elbow*" karena grafik yang dihasilkan biasanya berbentuk siku pada titik yang dianggap sebagai jumlah optimal dari *cluster* (Hartini et al., 2022). Hasil persentase yang berbeda dari setiap nilai *cluster* dapat ditunjukkan dengan menggunakan grafik sebagai sumbernya informasi. Jika nilai *cluster* turun membentuk sudut siku dan setelah *cluster* itu menurunnya nilai tidak terlalu drastis maka nilai *cluster* tersebut adalah yang terbaik (Muningsih, 2017).

### 2.5. Evaluasi *Cluster Sum Square Error (SSE)*

Merupakan salah satu matriks evaluasi internal yang digunakan untuk mengevaluasi kualitas dari pengelompokan data yang dihasilkan oleh algoritma *K-Means* (Patgiri et al., 2021). Evaluasi internal adalah metode yang digunakan untuk mengevaluasi kualitas pengelompokan hanya dengan menggunakan data pengelompokan itu sendiri. Evaluasi internal biasanya dilakukan untuk mengukur seberapa baik algoritma *clustering* dalam mengelompokkan data atau untuk menentukan jumlah *cluster* yang optimal (Thamrin & Wijayanto, 2021). SSE dapat dihitung dengan rumus 2.

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} d(m_i, x)^2 \quad (2)$$

Dimana:

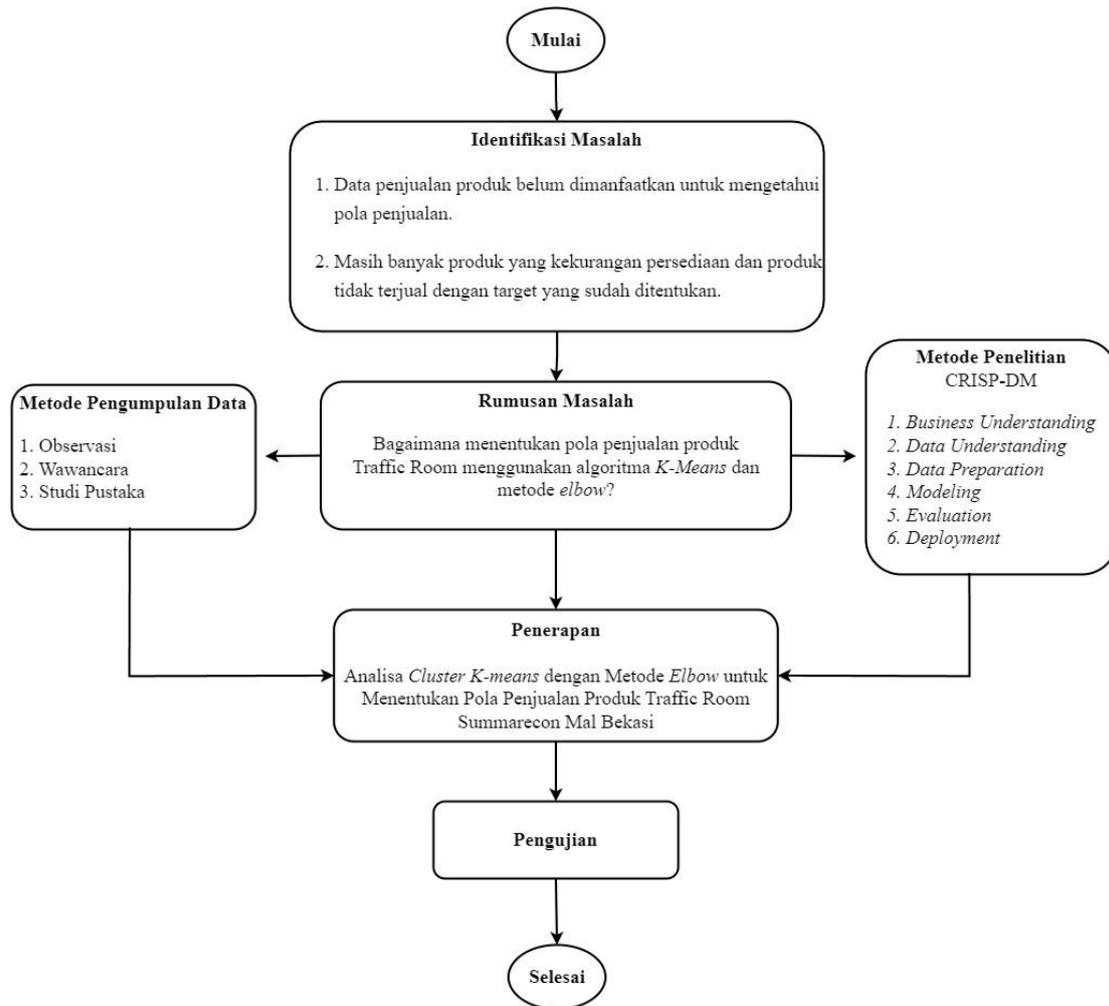
$K$  = banyaknya *cluster*,

$C_i$ ,  $d$  = *cluster* ke- $i$  dan jarak serta

$m_i$ ,  $x$  = centroid *cluster* ke- $i$  dan data masing-masing *cluster*.

### 2.6. Kerangka Penelitian

Peneliti membuat kerangka untuk menggambarkan proses penelitian yang ditempuh secara keseluruhan.



Sumber: Hasil Penelitian (2023)

Gambar 1. Kerangka Penelitian

## 2.7. Metode Pengumpulan Data

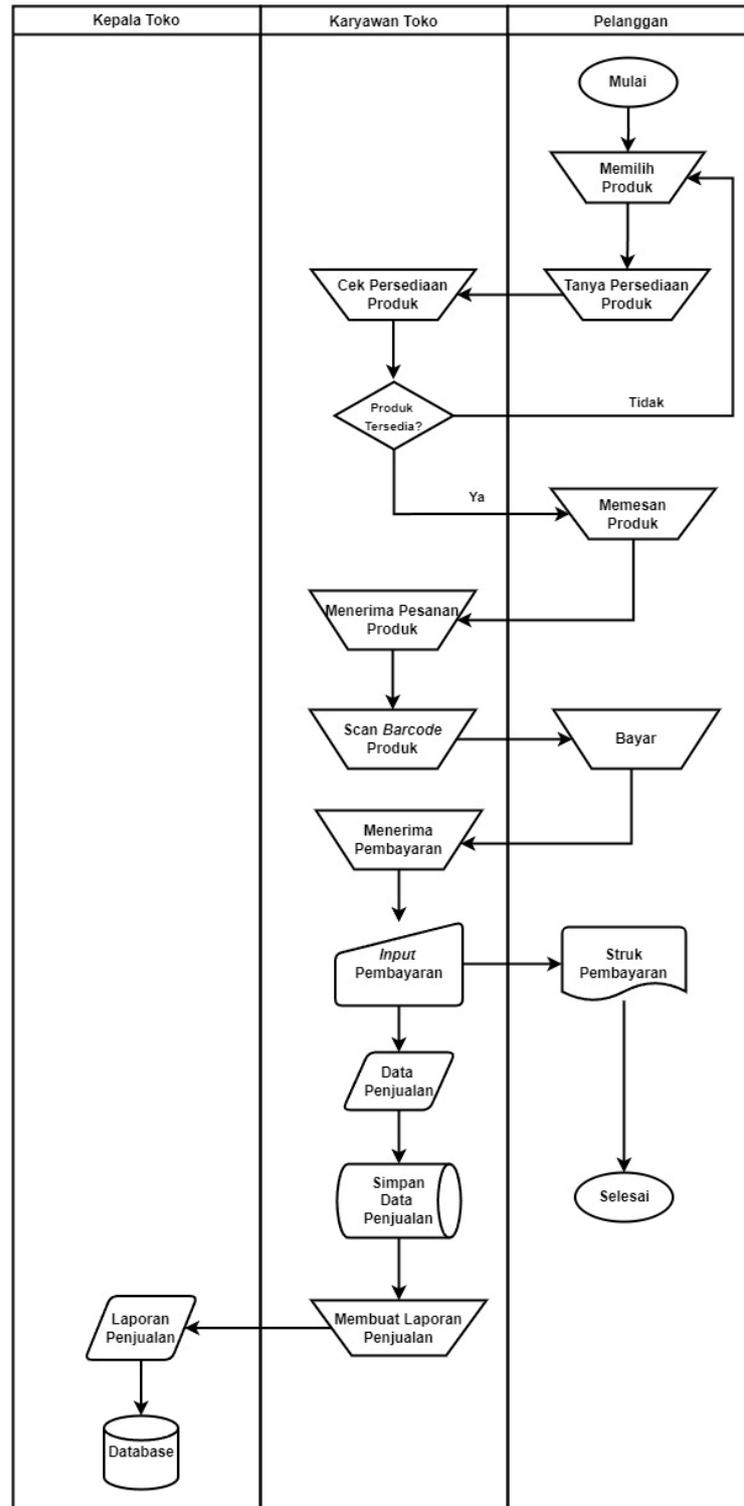
Observasi, melakukan pengumpulan data dengan pengamatan langsung ke Traffic Room mengenai nama-nama produk dan data penjualan tahunan. Data yang dikumpulkan yaitu pada tahun 2018-2022. Wawancara, melakukan tanya jawab dengan narasumber untuk mengumpulkan data dengan mempertanyakan pokok tentang permasalahan skripsi yang diteliti. Studi pustaka melakukan pencarian referensi melalui buku atau jurnal yang berhubungan dengan penelitian yang sedang dilaksanakan.

## 2.8. Metode Analisis

Pada metode analisis, beberapa analisis yang dibahas yaitu analisis sistem berjalan, analisis permasalahan dan analisis yang diusulkan.

### a. Analisis Sistem Berjalan

Peneliti melakukan analisis terhadap proses penjualan yang terjadi pada Traffic Room dimulai dari pelanggan memilih produk, selanjutnya menanyakan stok kepada karyawan toko, apabila stok tersedia karyawan toko melakukan *input* data transaksi ke dalam komputer. Karyawan toko membuat laporan penjualan untuk diserahkan ke Kepala Toko.



Sumber : Hasil Penelitian (2023)

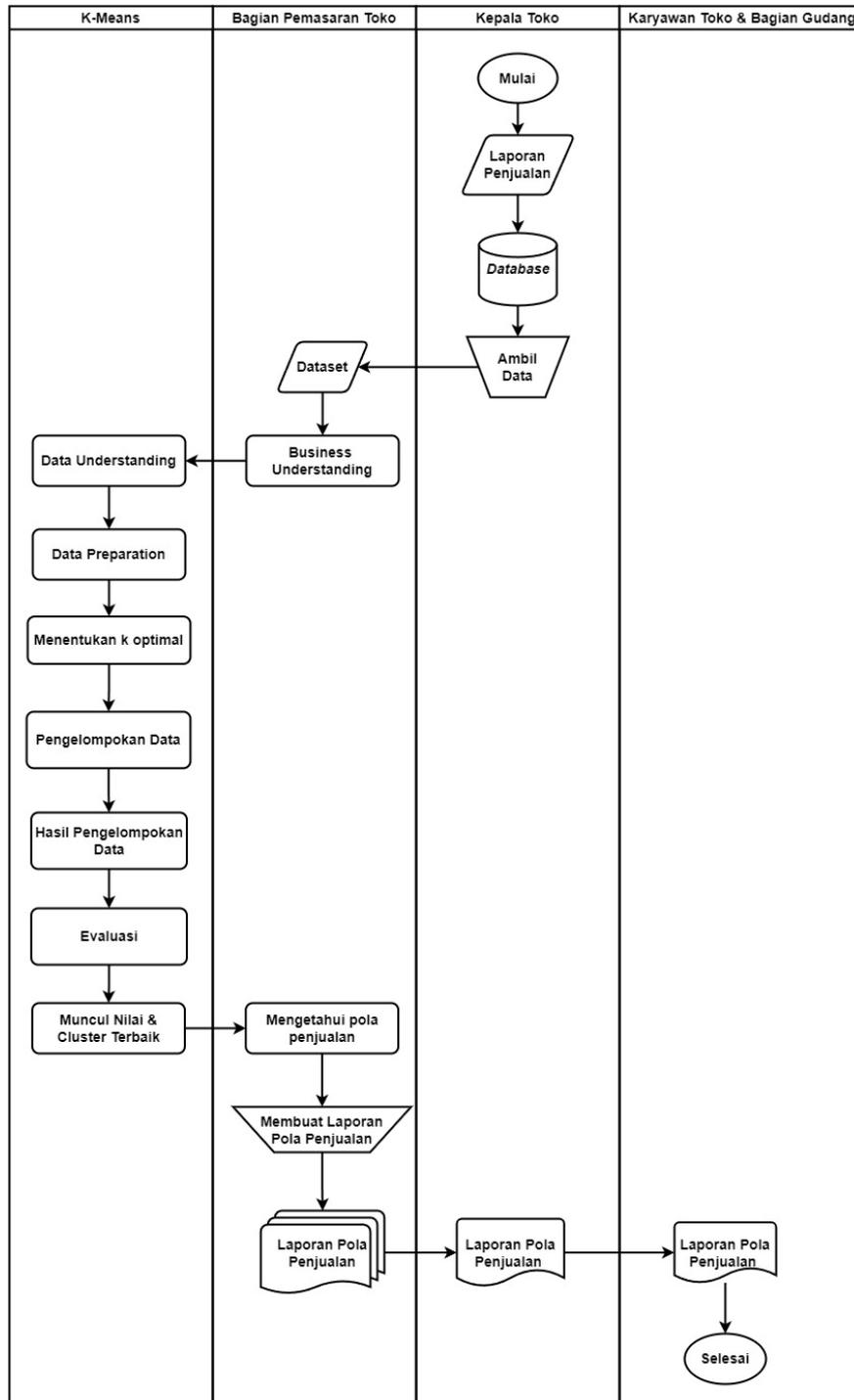
Gambar 2. Analisis Sistem Berjalan

b. Analisis Permasalahan

Berdasarkan sistem berjalan penjualan pada Gambar 2, maka terdapat kekuarangan di antaranya tidak ada tindak lanjut pemanfaatan data penjualan untuk mengetahui pola penjualan produk dan untuk karyawan toko harus lebih memperhatikan persediaan produk.

c. Analisis Sistem yang Diusulkan

Laporan penjualan yang ada pad kepala toko diserahkan ke bagian pemasaran untuk diolah sebagai dataset, dengan menggunakan algoritma *K-Means* diperoleh evaluasi hasil berupa pola penjualan sehingga hal ini dapat menjadi solusi untuk membantu mengetahui produk yang diminati pembeli.



Sumber : Hasil Penelitian (2023)

Gambar 3. Analisis Sistem yang Diusulkan

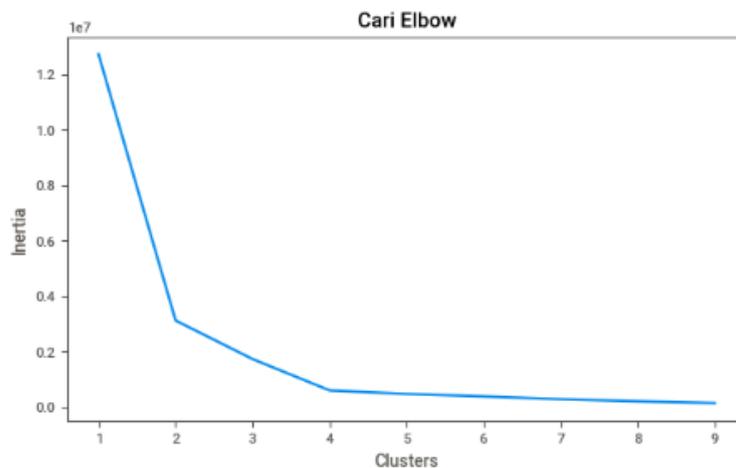


### 3.4. Penerapan (*Modeling*)

Pada modeling dilakukan proses menentukan jumlah *cluster* optimal dengan Grafik metode Elbow, kemudian penerapan algoritma *K-Means*, dan perhitungan menggunakan *Microsoft Excel*.

#### a. Menentukan Jumlah *Cluster* Optimal

Dapat dilihat hasil grafik pada Gambar 6 setelah melakukan teknik metode *Elbow* grafik menghasilkan dua sudut/siku di  $k=2$  dan  $k=4$ . Diantara  $k=2$  dan  $k=4$  menunjukkan bahwa kondisi *Elbow* atau  $k$  optimal ditemukan pada nilai  $k=4$ . Hal tersebut menunjukkan bahwa kondisi optimal dari proses klasterisasi ditemukan pada nilai  $k=4$  atau data dapat dikelompokkan pada 4 *cluster*.

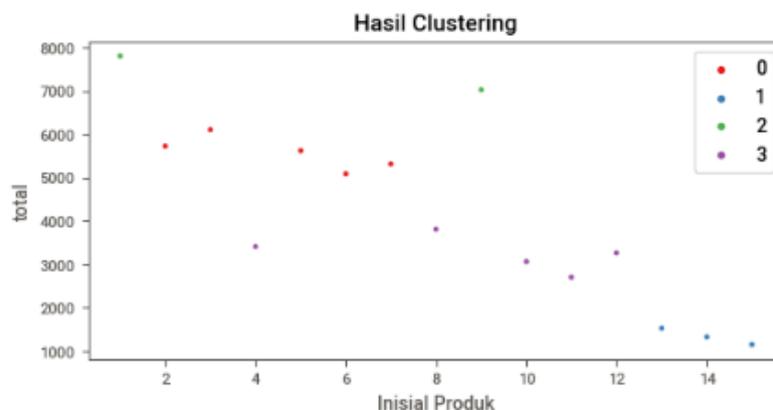


Sumber : Hasil Penelitian (2023)

Gambar 6. Grafik Metode *Elbow*

#### b. Penerapan Algoritma *K-Means*

Terlihat bahwa grafik *scatter plot* pada Gambar 7 mengelompokkan masing-masing produk. Di dalam grafik terdapat sumbu x yaitu inisial produk dan sumbu y adalah jumlah total produk pada tahun 2018-2022 masing-masing produk.



Sumber : Hasil Penelitian (2023)

Gambar 7. Grafik Hasil *Clustering*

Selanjutnya pada Gambar 8 terlihat hasil penjualan untuk setiap produk beserta pengelompokannya produk sangat laris, laris, cukup laris dan kurang laris.

produk dan tahun	Keterangan
t-shirt kids	Cukup Laris
t-shirt	Sangat Laris
sandal	Cukup Laris
syal	Kurang Laris
polo shirt	Cukup Laris
longsleeve	Laris
blazzer	Cukup Laris
jaket	Sangat Laris
hoodie	Laris
crewneck	Laris
celana pendek	Cukup Laris
tas	Kurang Laris
celana panjang denim	Laris
celana panjang cargo	Laris
bucket hat	Kurang Laris

Sumber : Hasil Penelitian (2023)

Gambar 8. Hasil Pola Penjualan Produk

c. Penerapan Perhitungan *Microsoft Excel*

Dalam penelitian ini, iterasi yang di dapatkan sampai iterasi ke-9 dimana *cluster* tidak adanya *cluster* berubah dan tidak ada lagi data yang berpindah dari *cluster* satu ke *cluster* lainnya.

Tabel 1 merupakan hasil iterasi ke-1 dan Tabel 2 merupakan hasil iterasi ke-9

Tabel 1. Perhitungan Data Iterasi 1

Data ke-i	C1 t-shirt kids	C2 t-shirt	C3 sandal	C4 syal	Min Jarak	Anggota Cluster
1	0	2314,37227	315,4283	703,6185	0	C1
2	2314,372269	0	2088,129	3006,637	0	C2
3	315,4282803	2088,12931	0	949,1554	0	C3
4	703,6185046	3006,63666	949,1554	0	0	C4
5	378,6674002	2029,04362	205,3144	1022,685	205,314393	C3
6	1519,664108	877,752813	1287,693	2217,585	877,752813	C2
7	386,4970893	2240,25155	310,7137	911,5728	310,713695	C3
8	1944,17206	421,808013	1723,528	2642,274	421,808013	C2
9	1324,690907	1053,7661	1077,14	2008,681	1053,7661	C2
10	1354,649032	1060,79074	1125,258	2051,275	1060,79074	C2
11	524,3929824	1858,66269	311,2732	1202,929	311,273192	C3
12	635,0677129	2945,35872	876,9595	138,7227	138,722745	C4
13	1070,76935	1289,06167	828,1353	1760,21	828,135255	C3
14	1179,787693	1206,72366	950,2579	1869,512	950,25786	C3
15	532,5410782	2836,22302	785,545	182,3595	182,359535	C4

Sumber : Hasil Penelitian (2023)

Tabel 2. Perhitungan Data Iterasi 9

Data ke-i	C1 t-shirt kids	C2 t-shirt	C3 sandal	C4 syal	Min Jarak	Anggota Cluster
1	290,793535	2126,87129	1280,89	620,5734	290,793535	C1
2	2100,596458	210,904007	1085,33	2928,736	210,904007	C2
3	134,6368449	1902,87585	1042,984	868,021	134,636845	C1
4	946,0880931	2822,45576	1975,82	92,83181	92,8318076	C4
5	141,4845575	1842,81375	978,4479	942,539	141,484557	C1
6	1288,490388	674,955924	258,5657	2134,15	258,565736	C3
7	226,3123505	2049,82914	1165,643	824,3579	226,312351	C1
8	1729,775789	210,904007	702,3249	2562,083	210,904007	C2
9	1080,128918	877,995729	210,6163	1928,964	210,616334	C3
10	1121,588909	857,668642	160,6725	1966,578	160,672462	C3
11	265,5644555	1670,73023	802,4156	1117,118	265,564455	C1
12	863,3712295	2757,95241	1899,585	81,73337	81,7333745	C4
13	845,007858	1094,57503	252,7897	1678,666	252,789715	C3
14	937,7768818	1014,77953	182,6938	1788,14	182,693842	C3
15	782,4168199	2652,0412	1807,974	106,5328	106,532833	C4

Sumber : Hasil Penelitian (2023)

Dengan iterasi yang dilakukan maka di dapatkan pola penjualan produk Traffic Room yaitu:

Sangat Laris (C2) : *T-Shirt* dan *Jaket*

Laris (C3) : *Longsleeve, Hoodie, Crewneck, Celana Panjang Denim dan Celana Panjang Kargo*

Cukup Laris (C1) : *T-Shirt Kids, Sandal, Polo Shirt, Blazzer, dan Celana Pendek.*

Kurang Laris (C4) : *Syal, Tas dan Bucket Hat*

### 3.5. Pengujian

Evaluasi *Sum of Square Error* (SSE), dari perhitungan *google colab*, mendapatkan *cluster 4* yang optimal karena membentuk siku dan dilanjutkan penurunan nilai *cluster* tidak terlalu signifikan. Nilai dari *cluster 4* yaitu 594.366,733 atau 65,5%.

```

1 12712401.2
2 3115814.3214285714
3 1724543.504761905
4 594366.7333333333
5 468319.4333333333
6 362618.3333333334
7 273657.3333333334
8 202854.8333333337
9 138723.3333333333
    
```

Sumber : Hasil Penelitian (2023)

Gambar 9. Nilai Cluster

Perhitungan pada Gambar 9, juga terbukti dengan perhitungan manual memakai *Microsoft Excel* yaitu pada Tabel 3.

Tabel 3 Hasil Nilai SSE

Min Jarak	Anggota <i>Cluster</i>	Kuadrat Min Jarak
290,793535	C1	84.560,88
210,904007	C2	44.480,5
134,636845	C1	18127,08
92,8318076	C4	8617,7445
141,484557	C1	20017,88
258,565736	C3	66856,24
226,312351	C1	51217,28
210,904007	C2	44480,5
210,616334	C3	44359,24
160,672462	C3	25815,64
265,564455	C1	70524,48
81,7333745	C4	6680,3445
252,789715	C3	63902,64
182,693842	C3	33377,04
106,532833	C4	11349,2445
		594.366,7335

Sumber : Hasil Penelitian (2023)

Untuk melihat nilai SSE dalam bentuk persentase yaitu pada Tabel 4.

Tabel 4. Nilai SSE dan Selisih SSE

<i>Cluster</i>	Nilai SSE	Selisih SSE
1	12.712.401,2	0
2	3.115.814,321	9.596.586,88
3	1.724.543,504	1.391.270,82
4	594.366,733	1.130.176,77
5	468.319,433	126.047,3
6	362.618,333	105.701,1
7	273.657,333	88.961
8	202.854,833	70.802,5
9	138.723,333	64.131,5

Sumber : Hasil Penelitian (2023)

$$persentase = \left( \frac{\text{selisih SSE}}{(\text{nilai SSE} + \text{selisih SSE})} \right) \times 100$$

$$persentase = \left( \frac{1.130.176,77}{(594.366,733 + 1.130.176,77)} \right) \times 100$$

$$persentase = 65,5\%$$

Maka dari nilai SSE 594.366,733 dengan persentase yang diperoleh yaitu 65,5%.

### 3.6. Deployment

Tahap terakhir yaitu tahap penyebaran hasil dari penelitian berdasarkan pemodelan dan evaluasi pada proses *data mining* yang sudah dilakukan. Hasil akan disampaikan berupa laporan kepada pihak toko Traffic Room sebagai rekomendasi untuk menentukan pola penjualan pada tahun-tahun yang akan datang.

#### **4. Kesimpulan**

Hasil pengelompokan algoritma *K-Means* yang dibantu dengan metode *Elbow*. Mendapatkan empat *cluster* dari grafik metode *elbow* maka pola penjualan produk Traffic Room yaitu Sangat Laris (C2) untuk *T-Shirt* dan Jaket; Laris (C3) untuk *Longsleeve*, *Hoodie*, *Crewneck*, Celana Panjang Denim dan Celana Panjang Kargo; Cukup Laris (C1) untuk *T-Shirt Kids*, Sandal, Polo Shirt, Blazzer, dan Celana Pendek; dan Kurang Laris (C4) Syal, Tas dan *Bucket Hat*. Dari pola penjualan yang sudah dapat diatas, sehingga dapat memperbaiki persediaan produk pada toko Traffic Room yaitu untuk kategori C2 dan C3 (penjualan tinggi) sehingga persediaan produk yang termasuk kategori C2 dan C3 diperbanyak dalam produksi, agar memenuhi penjualan toko; untuk kategori C1 (penjualan sedang) sehingga persediaan produk yang termasuk kategori C1 dikurangi setengah dalam produksi; dan untuk kategori C4 (penjualan rendah) sehingga persediaan produk yang termasuk kategori C4 tidak diperbanyak dalam produksi, agar tidak ada produk yang menumpuk di toko. Pada tahap evaluasi mendapatkan *cluster* optimal menggunakan perhitungan *Sum of Square Error* (SSE) yaitu *cluster* 4 dikarenakan pada grafik membentuk siku lalu di ikuti penurunan nilai *cluster* yang tidak terlalu signifikan. Serta menghasilkan nilai *Sum of Square Error* (SSE) optimal yaitu 594.366,733 dengan persentase 65,5%. Maka arti dari tahap evaluasi yaitu kinerja algoritma *K-Means* yang digunakan sudah baik.

#### **Daftar Pustaka**

- Arhami, M., & Nasir, M. (2020). *Data Mining - Algoritma dan Implementasi*. Penerbit Andi.  
<https://books.google.co.id/books?id=AtcCEAAAQBAJ>
- CEIC DATA. (2023). *Indonesia Pertumbuhan Penjualan Ritel*. <https://www.ceicdata.com/>  
<https://www.ceicdata.com/id/indicator/indonesia/retail-sales-growth>
- Ferdian, M., Rumanti, A. A., & Rizana, A. F. (2021). Perancangan Sistem Informasi Manajemen Pengelolaan Persediaan Barang dan Keuangan pada Penjualan Menggunakan Metode Scrum pada Fashion Retail Magna. *E-Proceeding of Engineering*, 8(5), 8128–8142.
- Hartini, Y. S., Lefanska, A. B. P., Ursia, A. A., Prasetyo, D. A. B., & Sugiharto, B. (2022). Pengembangan, Penerapan Dan Pendidikan 'Sains Dan Teknologi' Pasca Pandemi. *Prosiding Seminar Nasional Sanata Dharma Berbagi*.  
<https://books.google.co.id/books?id=qCSjEAAAQBAJ>
- Muningsih, E. (2017). Optimasi Jumlah Cluster K- Means Dengan Metode Elbow Untuk Pemetaan Pelanggan. *Prosiding Seminar Nasional Elinvo*, 105–114.
- Muttaqin, M. R., Hermanto, T. I., & Sunandar, M. A. (2022). Penerapan *K-Means* Clustering dan Cross-Industry Standard Process For Data Mining (CRISP-DM) untuk Mengelompokan Penjualan Kue. *Journal.Unpak.Ac.Id*, 19(1), 38–53.  
<http://journal.unpak.ac.id/index.php/komputasi/article/view/3976>
- Nainggolan, R., & Lumbantoruan, G. (2018). Optimasi Performa Cluster *K-Means* Menggunakan Sum of Squared Error (SSE). *METHOMIKA: Jurnal Manajemen Informatika &*

- Komputerisasi Akuntansi*, 2(2), 103–108.  
<https://doi.org/https://doi.org/10.46880/jmika.Vol2No2.pp103-108>
- Patgiri, R., Bandyopadhyay, S., & Balas, V. E. (2021). *Proceedings of International Conference on Big Data, Machine Learning and Applications* (Ripon Patgiri, S. Bandyopadhyay, & V. E. Balas (eds.); Vol. 180). Springer Singapore. <https://doi.org/10.1007/978-981-33-4788-5>
- Purnomo, N. H., Pamungkas, B., & Juliane, C. (2023). Penerapan Algoritma C4 . 5 Untuk Klasifikasi Tren Pelanggaran Kendaraan Angkutan Barang dengan Metode CRISP-DM. *Jurnal Media Informatika Budidarma*, 7(1), 30–40. <https://doi.org/10.30865/mib.v7i1.5247>
- Safira, I., Salkiawati, R., & Priatna, W. (2022). Penerapan Algoritma *K-Means* untuk Mengetahui Pola Persediaan Barang pada Toko Raja Bekasi. *Journal of Informatic and Information Security*, 3(1), 99–110. <https://doi.org/10.31599/jiforty.v3i1.1253>
- Sono, M. G., Assayuti, A. A., & Rukmana, A. Y. (2023). Hubungan Antara Perencanaan Strategis, Ekspansi Pasar, Keunggulan Kompetitif Terhadap Pertumbuhan Perusahaan Fashion di Jawa Barat. *Jurnal Bisnis Dan Manajemen West Science*, 2(02), 81–91. <https://doi.org/10.58812/jbmws.v2i02.340>
- Thamrin, N., & Wijayanto, A. W. (2021). Comparison of Soft and Hard Clustering: A Case Study on Welfare Level in Cities on Java Island. *Indonesian Journal of Statistics and Its Applications*, 5(1), 141–160. <https://doi.org/10.29244/ijsa.v5i1p141-160>
- Werdiningsih, I., Novitasari, D. C. R., & Haq, D. Z. (2022). *Pengelolaan Data Mining dengan Pemrograman Matlab*. Airlangga University Press. <https://books.google.co.id/books?id=CgOdEAAAQBAJ>
- Zuhri, B., & Harani, N. H. (2023). *Aplikasi Rekrutmen Karyawan Dengan Algoritma Artificial Neural Network (ANN) dan Framework Flask*. Penerbit Buku Pedia. <https://books.google.co.id/books?id=yeCuEAAAQBAJ>