

## Clustering Penjualan Terbaik Dengan Sum of Squares Error Dan Menentukan Nilai K Menggunakan Algoritma K-Means

Eric Kamti<sup>1</sup>, Ade Putra Sanjaya<sup>2</sup>, Marlince Nk Nababan<sup>3</sup>

<sup>1,2,3</sup> Sarjana Sistem Informasi, Universitas Prima Indonesia, Medan, Indonesia

Email: <sup>1</sup>Erickamti16@gmail.com, <sup>2</sup>Adeputrasanjaya2001@gmail.com, <sup>3</sup>marlince@unprimdn.ac.id

Email Penulis Korespondensi: [marlince@unprimdn.ac.id](mailto:marlince@unprimdn.ac.id)

### Article History:

Received May 13<sup>th</sup>, 2024

Revised Jun 27<sup>th</sup>, 2024

Accepted Aug 09<sup>th</sup>, 2024

### Abstrak

Pada umumnya Toko Swalayan tidak tau untuk item yang laku , yang tidak laku dan menengah jika item yang ada juga terlalu banyak . Oleh karena itu biasanya banyak barang barang yang sudah lama tidak terjual dan akhirnya expired . Namun sebenarnya hal tersebut sangat perlu diperhatikan oleh pihak Toko , agar profit yang dihasilkan tidak untuk menutup barang barang yang tidak laku itu . Sebab , jika barang yang tidak laku memiliki stok yang banyak , maka lama kelamaan barang tersebut akan mencapai tanggal expirednya . Setelah itu makanan tersebut tidak bisa dikonsumsi lagi . Untuk mengatasi permasalahan diatas dapat digunakan metode analisa data yang bernama K-Means Clustering dan juga Sum Of Square Error. K-means Digunakan untuk menganalisa data penjualan toko dan di kelompokkan menjadi 3 cluster . Cluster pertama adalah cluster dengan tingkat penjualan yang rendah , Cluster yang kedua ada cluster dengan tingkat penjualan yang tinggi , dan Cluster yang ketiga adalah cluster dengan tingkat penjualan yang rendah . Sum Of Square Error digunakan untuk mencari cluster terbaik. Penelitian ini diharapkan dapat membantu Toko agar bisa lebih baik dalam pengolahan stok . Untuk penjualan yang rendah bisa dikurangi untuk stok barangnya dan diperbanyak stok untuk penjualan yang menengah dan penjualan yang tinggi.

**Kata Kunci** :K-Means, Sum Of Squares Error , Clustering

### Abstract

*In general, supermarkets don't know which items are selling, which aren't selling and whether there are too many items. Therefore, usually many items have not been sold for a long time and eventually expire. However, the shop really needs to pay attention to this matter, so that the profit generated does not cover the goods that do not sell. Because, if an unsold item has a lot of stock, then over time the item will reach its expiry date. After that the food cannot be consumed again. To overcome the above problems, a data analysis method called K-Means Clustering and also Sum Of Square Error can be used. K-means is used to analyze store sales data and grouped into 3 clusters. The first cluster is a cluster with a low level of sales, the second cluster is a cluster with a high level of sales, and the third cluster is a cluster with a medium level of sales. Sum Of Square Error is used to find the best cluster. It is hoped that this research can help shops to be better at processing stock. For low sales, stock can be reduced and stock increased for medium sales and high sales.*

**Keyword** :K-Means, Sum Of Squares Error, Clustering

## 1. PENDAHULUAN

Toko Swalayan adalah yang sering dihadapi adalah kurangnya pengetahuan dalam menentukan produk yang paling diminati konsumen. Hasil penelitian menunjukkan bahwa penggunaan Data Mining K-Means Clustering memberikan informasi yang berguna terkait pengelompokan penjualan produk produk di Toko swalayan.[1]

Ada beberapa cara yang dapat digunakan untuk memprediksi kebutuhan stok berdasarkan jenis cluster dari sejumlah besar data, dan salah satu di antaranya adalah melalui metode clustering. Metode clustering dapat diimplementasikan dengan menggunakan Algoritma K-Means, yang dikenal karena kemudahannya implementasinya.[2].

Secara umum, cluster adalah kelompok objek yang lebih serupa satu sama lain daripada dengan anggota kelompok lainnya. Algoritma K-Means Clustering adalah salah satu teknik pengelompokan yang paling umum digunakan untuk mengkategorikan data. Dalam konteks analisis penjualan, K-Means dapat diterapkan dengan dua kriteria utama, yaitu penjualan tinggi dan rendah. Metode K-Means efisien dan cepat dalam mengelompokkan data dalam jumlah besar.[3]

Penelitian ini bertujuan untuk Membuat sistem Cluster pengelompokkan transaksi barang menggunakan algoritma K-means pada Toko Swalayan dalam menentukan strategi penentuan stok barang, sehingga kita dapat menyesuaikan Produk yang terlaris agar dibuat lebih banyak persediaannya, dan Produk yang terendah agar dibuat lebih sedikit untuk ketersediaan stoknya, agar produk yang tidak begitu laris tidak perlu menumpuk modal yang besar, sehingga modalnya bisa digunakan untuk menyetok produk yang laris).

Batasan masalah dari penelitian ini adalah mengurutkan jumlah permintaan sesuai jumlah cluster, analisis yang digunakan merupakan algoritma k-means dan Sum Of Squares Error. Manfaat dari penelitian ini adalah untuk membantu toko swalayan dalam menentukan stok barang. Menstabilkan dan meningkatkan perekonomian, karena stok barang tidak sampai kosong dan juga tidak mengalami kelebihan stok.

Definisi clustering menurut Tan (2006) adalah proses pengelompokan data ke dalam beberapa cluster atau kelompok, di mana data dalam satu cluster memiliki tingkat kemiripan yang tinggi, sementara data antar cluster memiliki tingkat kemiripan yang rendah. Proses clustering melibatkan partisi satu set objek data ke dalam himpunan bagian yang disebut cluster[4]. Objek yang berada dalam satu cluster memiliki karakteristik yang mirip satu sama lain dan berbeda dari cluster lainnya. Partisi ini tidak dilakukan secara manual, melainkan menggunakan suatu algoritma clustering. Oleh karena itu, clustering memiliki kegunaan yang signifikan dalam mengidentifikasi kelompok atau grup yang mungkin tidak dikenali dalam data. Clustering banyak digunakan dalam berbagai aplikasi seperti business intelligence, pengenalan pola citra, pencarian web, ilmu biologi, dan keamanan.

Dalam konteks business intelligence, clustering dapat membantu mengelompokkan banyak pelanggan ke dalam kelompok-kelompok yang memiliki karakteristik serupa. Sebagai contoh, dapat digunakan untuk mengelompokkan pelanggan ke dalam beberapa cluster berdasarkan kesamaan karakteristik yang kuat. [5]Clustering juga dikenal sebagai data segmentation karena mempartisi banyak dataset ke dalam kelompok berdasarkan kesamaan mereka. Selain itu, clustering juga dapat digunakan sebagai pendeteksian outlier. Clustering sering menggunakan metrik atau jarak untuk mengukur kemiripan atau perbedaan antara objek-objek. Fungsi objektif digunakan untuk mengukur kualitas pengelompokan, seperti minimisasi varian dalam kelompok atau maksimisasi jarak antara kelompok-kelompok. Clustering memiliki aplikasi luas dalam berbagai bidang, seperti pengelompokan konsumen berdasarkan perilaku pembelian, analisis data biomedis, segmentasi pasar, pengelompokan teks, pengelompokan gambar, dan banyak lagi.

## **2. METODOLOGI PENELITIAN**

### **2.1 Tahapan Penelitian**

Penjelas terkait masalah yang diteliti itu adalah pada umumnya toko swalayan tidak bisa menganalisa item yang tidak laku, dikarenakan banyaknya jenis barang. Oleh karena itu penulis akan menganalisa dan mengelompokkan kedalam 3 jenis cluster, yaitu cluster dengan penjualan rendah, Cluster dengan penjualan tinggi dan Cluster dengan penjualan menengah. Studi literatur untuk mengecek stok barang dan mengelompokkannya menggunakan algoritma k-means clustering dan menggunakan Sum of Squares Error untuk mencari cluster terbaik[6]. Input dan output juga merupakan bagian yang sangat penting yang harus dijelaskan secara rinci dan lengkap. Input yang diperlukan adalah data penjualan toko swalayan dan output merupakan keluaran dari input yang sudah diproses.

Metode Penelitian ini adalah langkah yang dilakukan peneliti untuk mengumpulkan informasi atau data serta melakukan investigasi pada data yang telah didapatkan tersebut. Metode Pengumpulan Data dalam melakukan pengumpulan data, penulis menggunakan metode dengan cara mengumpulkan data dari jurnal atau karya tulis yang relevan, hal tersebut dapat membantu penulis untuk menambah referensi sesuai dengan permasalahan yang ada. Tantangan utama dalam Analisa cluster adalah penentuan jumlah cluster atau parameter model, yang sering kali tidak di ketahui sebelumnya dan perlu ditetapkan sebelum proses pengelompokkan. Terdapat beberapa algoritma pengelompokkan yang telah diajukan, salah satunya adalah metode K-Means.[7]

Data mining merupakan proses analisis data untuk mengidentifikasi hubungan yang terdefinisi dengan jelas dan menghasilkan informasi baru yang sebelumnya tidak diketahui. Pendekatan ini memberikan manfaat yang signifikan bagi pemilik data dengan menggunakan metode terkini. Salah satu teknik yang umum digunakan, seperti metode k-means, membagi data menjadi kelompok-kelompok berdasarkan karakteristik serupa, memungkinkan data dengan sifat yang mirip dikelompokkan bersama, sementara data dengan sifat yang berbeda ditempatkan dalam kelompok yang berbeda.[8]

Benri Melpa Metisen (2015) dengan judul analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk di Supermarket Fadhillah menjelaskan bahwa data diambil dari Supermarket Fadhillah sehingga menghasilkan dua jenis kelompok data. Data penjualan rendah dan data penjualan tinggi. Nah, dengan mengelompokkan data tersebut bisa diketahui jenis barang mana yang laris manis dan barang mana yang ada di gudang tidak menumpuk.[9]

## 2.2 Metode Penyelesaian Masalah

Penggunaan data mining dirasa tepat untuk pengolahan data penjualan di supermarket, karena dengan menggunakan data mining data dalam jumlah besar dapat dikelola secara efisien. Data penjualan dari cabang perusahaan di berbagai wilayah bisa sangat besar dan rumit untuk dikelola secara manual. Penambangan data membantu dalam mengelompokkan dan mengatur data ke dalam kelompok yang bermakna, membuat analisis dan pengambilan keputusan menjadi lebih mudah. [10]

Definisi data mining secara formal adalah proses mengekstrak informasi yang valid, bermanfaat, tak dikenal, dan dapat dipahami dari data dan menggunakannya untuk membuat keputusan bisnis.[11]

Pengertian data mining sendiri adalah media yang digunakan untuk mengolah dan mengelompokkan data yang tergabung kedalam suatu basedata.[12] Salah satu metode yang digunakan adalah clustering. Clustering adalah salah satu metode yang digunakan untuk menggolongkan data menjadi beberapa bagian yang telah ditentukan sebelumnya [13]

Data mining yaitu suatu proses penggalian data dari sebuah informasi yang sangat penting. Data Mining juga merupakan suatu proses untuk menggali pola-pola dari data. Pola-pola itu didapatkan dari berbagai jenis basis data seperti basis data relasional, data warehouse, data transaksi, dan data berorientasi objek. Penggunaan data mining dapat membantu para pebisnis dalam pengambilan keputusan secara cepat dan tepat [14]

Algoritma K-means clustering adalah suatu metode yang dapat digunakan untuk proses pemodelan data tanpa supervise(Unsupervised).[15] Artinya, dalam clustering, tidak ada label atau kelas yang sudah ditentukan sebelumnya untuk mengarahkan pembagian data ke dalam kelompok. Algoritma clustering mencoba untuk menemukan pola atau struktur dalam data tanpa bantuan label. Clustering didasarkan pada konsep bahwa objek-objek dalam satu kelompok (cluster) memiliki kemiripan atau kesamaan tertentu. K-means ini juga merupakan salah satu algoritma yang digunakan dalam proses data mining.[16] engelompokan dilakukan dengan meminimalkan perbedaan antara objek dalam satu kelompok dan memaksimalkan perbedaan antara kelompok-kelompok yang berbeda.

Clustering adalah suatu metode dalam analisis data yang digunakan untuk mengelompokkan data atau objek ke dalam kelompok-kelompok yang serupa berdasarkan karakteristik atau kesamaan tertentu. Clustering juga digunakan untuk mengetahui komposisi didalam data. Metode untuk melakukan cluster dalam penelitian ini adalah K-Means.[17]

Tujuan utama dari clustering adalah mengidentifikasi struktur dalam data. Ini bisa berupa kelompok-kelompok yang tidak diketahui sebelumnya, pola-pola tersembunyi, atau hubungan antara objek-objek dalam dataset. Terdapat dua jenis utama dari metode clustering, yaitu non-hierarchical (partitional) dan hierarchical. Metode partitional, seperti K-Means, membagi data menjadi beberapa kelompok yang berbeda secara bersamaan. Sedangkan metode hierarchical membangun hirarki dari kelompok-kelompok yang lebih kecil hingga kelompok yang lebih besar.

Data Scalling termasuk kedalam data preprocessing . pada langkah ini penulis menggunakan Min Max Scaller yang merupakan bagian dari data scalling. Scaller ini membuat data berada pada rentang 0-1. Min Max Scaller juga merupakan varian untuk menormalisasi data pelatihan model machine learning , yaitu membawa nilai numerik ke skala yang seragam. Hal ini menyebabkan model belajar dan melakukan konvergensi lebih cepat karena gradien berubah secara seragam dan tidak membuat lompatan besar karena skala yang berbeda. Cara kerja dari Min Max Scaller sendiri merupakan bentuk normalisasi nilai antara 0 dan 1 . Namanya didapatkan karena nilai maksimum dan minimum.

Sum of Squares berarti Sum of Squares dari bilangan-bilangan yang diberikan. Dalam statistik, ini adalah jumlah kuadrat variasi suatu kumpulan data. Untuk ini, kita perlu mencari mean data dan mencari variasi setiap titik data dari mean, mengkuadratkannya, dan menjumlahkannya. Dalam aljabar, jumlah kuadrat dua bilangan ditentukan dengan menggunakan identitas  $(a + b)^2$ . Rumusnya juga dapat mencari jumlah kuadrat dari  $n$  bilangan asli pertama dengan menggunakan rumus. Rumusnya dapat diturunkan dengan menggunakan prinsip induksi matematika Untuk melakukan operasi aritmatika dasar yang diperlukan dalam statistik dan aljabar. Ada beberapa teknik berbeda untuk mencari jumlah kuadrat suatu bilangan.. Untuk menghitung jumlah dua atau lebih kuadrat dalam suatu ekspresi, digunakan rumus jumlah kuadrat. Selain itu, rumus jumlah kuadrat digunakan untuk menggambarkan seberapa baik data yang dimodelkan diwakili oleh suatu model. SSE adalah jumlah selisih kuadrat antara setiap observasi dan mean kelompoknya. Ini dapat digunakan sebagai ukuran variasi dalam suatu cluster.

## 3. HASIL DAN PEMBAHASAN

Berikut dibawah ini merupakan Hasil Dan Pembahasan dari penelitian yang dibuat oleh penulis. Dibawah ini akan dibahas secara lengkap untuk hasil dan juga pembahasannya.

### 3.1 Data Uji (Training)

Data yang digunakan merupakan data transaksi penjualan yang berjumlah 7403 data , yang akan di kelompokkan menjadi 3 jenis cluster .

### 3.2 Hasil Pengolahan Data

Berikut dibawah ini merupakan hasil data yang telah diolah menggunakan Algoritma K-Means dan Sum Of Squares Error.

	kode_barang	nama_barang	jumlah_transaksi	total_penjualan	rata_rata
0	2 TANG BLACK TEA 1 RENCENG (ISI 10)	2 TANG BLACK TEA 1 RENCENG (ISI 10)	1	1	10.000
1	2 TANG MELATI 1 RENCENG (ISI 10)	2 TANG MELATI 1 RENCENG (ISI 10)	1	1	10.000
2	AQUA 1500 ML 1 DUS	AQUA 1500 ML 1 DUS	10	11	11.000
3	RIBUT KILOAN	RIBUT KILOAN	77	83	10.779
4	7916248823	MINYAK TAWON FF	8	9	11.250

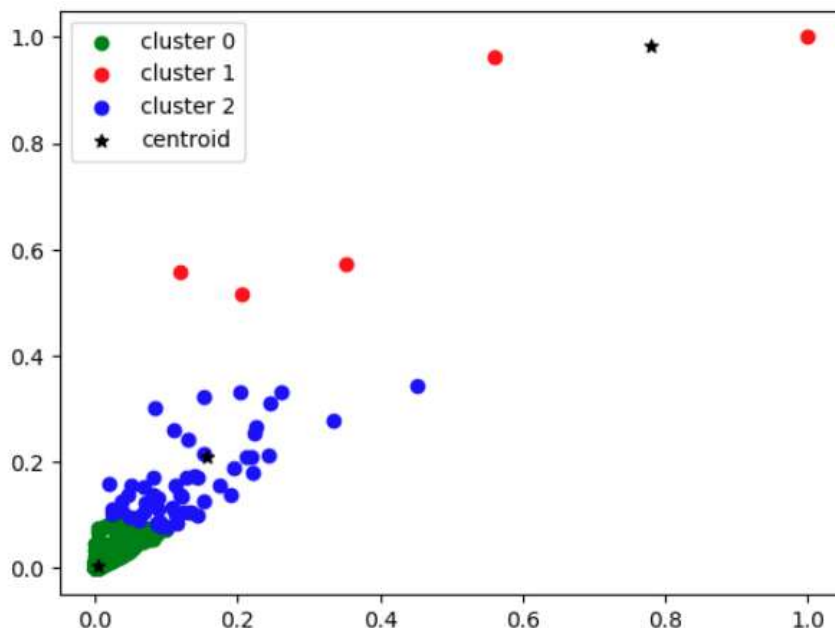
Gambar 1. data bagian atas

Gambar ini merupakan gambar yang berasal dari sample data penjualan supermarket yang belum diproses, namun data yang penulis tampilkan merupakan hanya data bagian atas.

	kode_barang	nama_barang	jumlah_transaksi	total_penjualan	rata_rata	cluster
0	2 TANG BLACK TEA 1 RENCENG (ISI 10)	2 TANG BLACK TEA 1 RENCENG (ISI 10)	1	1	10.000	0
1	2 TANG MELATI 1 RENCENG (ISI 10)	2 TANG MELATI 1 RENCENG (ISI 10)	1	1	10.000	0
2	AQUA 1500 ML 1 DUS	AQUA 1500 ML 1 DUS	10	11	11.000	0
3	RIBUT KILOAN	RIBUT KILOAN	77	83	10.779	0
4	7916248823	MINYAK TAWON FF	8	9	11.250	0

Gambar 2. menambahkan kolom cluster

Gambar ini merupakan gambar yang berasal dari sample data penjualan supermarket yang telah diproses , namun data yang diproses disini hanya menambahkan bagian kolom cluster , sehingga pada bagian itulah nanti penulis dapat memberikan gambar tentang penjualan yang terbaik , terendah , dan sedang.



Gambar 3. Hasil Penyebaran Data dan Centroid

Pada gambar ini dapat dilihat merupakan gambar hasil penyebaran data menggunakan kmeans. Pada gambar ini penulis menggunakan 3 jenis data frame dan 3 jenis cluster , data frame 1 = cluster 0 , data frame 2 = cluster 1 , data frame 3 = cluster 2 . Cluster 0 berwarna hijau , Cluster 1 berwarna merah dan Cluster 2 berwarna biru . Dapat dilihat bahwa cluster 0 berwarna hijau dan merupakan cluster dengan data penjualan yang paling rendah , Cluster 1 berwarna merah dengan tingkat penjualan yang paling tinggi , dan Cluster 2 berwarna biru dengan tingkat penjualan yang menengah. Analisa data yang ada dibawah ini sangat sulit jika dihitung manual menggunakan excel , dikarena untuk pencarian centroid pertama kedua dan ketiga . Tidak ada aturan khusus untuk mencari centroid , nilai centroid itu nilainya bebas , jadi jika ingin mencocokkan hasil yang ada disistem dengan yang ada di excel itu tidaklah mungkin.

```
sse
[8.233054446326904,
3.815409602402714,
2.1739959690437543,
1.2736086834800198,
0.9349618202773462,
0.7089127401197525,
0.5730184864884318,
0.47586919982397,
0.39876162974000307]
```

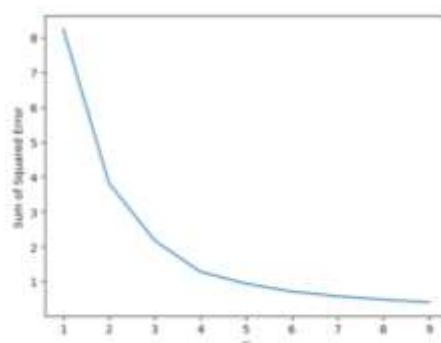
Gambar 4. Hasil Sum of Squares Error

Kemudian dapat dilihat pada gambar c bahwa nilai tertinggi ada pada K=2 , dimana k=2 yang bernilai 8.23... dan nilai minimum pada k=10 yang bernilai 0.39.... . Disini penulis akan menjelaskan , kenapa tidak ada K=1 . Itu disebabkan karena jika kita ingin mengelompokkan data dengan hanya menjadi 1 cluster , maka k=1 adalah data asli yang jika di proses maka akan sama saja dengan data aslinya . oleh karena itu di k-means menggunakan Sum Of Squares Error tidak ada K=1 , Minimal adalah K=2 . jika K=2 maka yang diteliti adalah data penjualan terbanyak dan data penjualan terkecil. Selanjutnya proses yang akan dilakukan adalah menghitung nilai selisih dari Sum Of Squares Error . Penulis akan mencari nilai selisih yang mendapatkan penurunan paling signifikan.Perhitungan selisih pada nilai Sum Of Squares Error dapat di lihat pada gambar 5

Cluster(k)	Hasil sse	Selisih
2	8,23305444632690	0
3	3,81540960240271	4,41764484392419
4	2,17399596904375	1,6414136335896
5	1,27360868348001	0,90038728556374
6	0,93543122897198	0,33817745450803
7	0,70891331659378	0,22651791237820
8	0,57313211650903	0,13578120008475
9	0,47585601322027	0,09727610328876
10	0,39749846948489	0,07835754373538

Gambar 5. Hasil perhitungan selisih SSE

Cara Menghitung selisih pada gambar table tersebut adalah dengan cara mengurangkan k=2 dengan k=3 . Maka hasil selisih pada k=3 adalah 8.23-3.81 maka di dapatkan hasil selisih dari k3 = 4.42. Untuk mencari nilai selisih k4 , k5 dan seterusnya menggunakan cara yang sama. Perhitungan selisih nilai Sum Of Squares Error pada setiap K cluster pada akhirnya menghasilkan nilai k yang optimal, dapat dilihat pada gambar d . Nilai K yang optimal ada pada k=3 yang merupakan selisih Sum Of Squares Error terbesar.



Gambar 6. Grafik SSE

Pada Grafik nilai Sum Of Squares Error terlihat adanya pola elbow yang mana menyatakan bahwa suatu cluster mempunyai nilai selisih terbesar antara nilai Sum Of Squares Error pada setiap data merupakan klaster yang Optimal. Maka dari itu ditemukan selisih terbesar ada pada k=3 dan merupakan klaster yang Optimal.Kemudian untuk gambar yang ke e bisa dilihat bahwa nilai penurunan Sum Of Square Error yang cukup konstan . dengan kesalahan atau penurunan yang paling signifikan ada pada 3 cluster. Dengan penguran jarak cluster ke 2 dan jarak cluster ke 3. Sehingga cluster yang paling optimal adalah cluster 3.

	kode_barang	nama_barang	jumlah_transaksi	total_penjualan	rata_rata	cluster
0	2 TANG BLACK TEA 1.RENCENG (ISI 10)	2 TANG BLACK TEA 1.RENCENG (ISI 10)	0.000000	0.000000	10.000	0
4919	8.99887E+12	EMERON HJB ANTI DANDRUFF RENCENG	0.001394	0.000869	10.000	0
4918	8.99887E+12	EMERON HIJAB ANTI DANDRUFF 170ML	0.000774	0.000483	10.000	0
4917	8.99887E+12	LOVELY COMPACT POWDER NO.03	0.000620	0.000388	10.000	0
4916	8.99887E+12	LOVELY COMPACT POWDER NO.02	0.000310	0.000193	10.000	0
4915	8.99887E+12	LOVELY COMPACT POWDER NO.01	0.000310	0.000193	10.000	0
4914	8.99887E+12	LOVELY LOOSE POWDER NO.02	0.000000	0.000000	10.000	0
4913	8.99887E+12	LOVELY LOOSE POWDER NO.01	0.000000	0.000000	10.000	0
4912	8.99887E+12	POSH HIJAB WINTER MAGIC 150ML	0.002169	0.001351	10.000	0
4911	8.99887E+12	POSH HIJAB PURPLE WISH 150ML	0.002788	0.001737	10.000	0
4910	8.99887E+12	POSH HIJAB GREEN BLOSSOM 150ML	0.003872	0.002509	10.385	0
4909	8.99887E+12	MAMA LIME CHARCOAL 120ML	0.028487	0.027311	16.512	0
4920	8.99887E+12	MAMA LIME 220ML	0.020446	0.014476	11.353	0
4908	8.99887E+12	EMERON CLEAN&FRESH RENCENG	0.003717	0.002316	10.000	0
4906	8.99887E+12	POSH CHARCOAL 50ML	0.001239	0.000772	10.000	0
4905	8.99887E+12	ROLL ON POSH ANTI STAIN 50ML	0.001704	0.001082	10.000	0
4904	8.99887E+12	ROLL ON POSH WHITENING 50ML	0.002323	0.001448	10.000	0
4903	8.99887E+12	MAMA CHARCOAL 800ML	0.002189	0.001351	10.000	0

Gambar 7. Hasil Data Yang telah diolah Cluster 0

	kode_barang	nama_barang	jumlah_transaksi	total_penjualan	rata_rata	cluster
4941	SEDAP AYAM BAWANG	SEDAP AYAM BAWANG	0.118269	0.558483	75.071	1
4940	BIMOLI 2 LITER	BIMOLI 2 LITER	0.205235	0.517661	40.460	1
6477	SEDAP GORENG	SEDAP GORENG	1.000000	1.000000	16.049	1
6552	GULA PASIR 1 KG	GULA PASIR 1 KG	0.353005	0.572662	26.039	1
6553	GULA PASIR 1/2 KG	GULA PASIR 1/2 KG	0.560409	0.864196	27.610	1
1395	8.9927E+12	BEAR BRAND 189ML	0.108116	0.115036	17.067	2
6942	PERMEN COIN	PERMEN COIN	0.088154	0.153445	36.077	2
32	89886010343	INDOME SOTO ME	0.082559	0.138101	26.816	2
36	89886014280	SUPERMI AYAM BAWANG	0.037949	0.127196	53.618	2
35	INTERME KALDU AYAM	INTERME KALDU AYAM	0.129337	0.243968	30.251	2
1854	8.99275E+12	FF SKM COKLAT RENCENG	0.152107	0.127003	13.398	2
1845	8.99275E+12	FF SKM PUTIH RENCENG	0.243649	0.213279	14.047	2
31	89886010046	INDOME AYAM SPESIAL	0.070632	0.123142	27.943	2
39	89886023770	SAKURA GORENG	0.018742	0.157885	134.180	2
456	7.11844E+11	ABC KECAP MANIS REF 60ML	0.038245	0.107605	47.489	2
4942	BIMOLI 1 LITER	BIMOLI 1 LITER	0.084108	0.303223	57.776	2
5481	8.99899E+12	GG DJAYA	0.152571	0.214244	22.525	2
2178	8.99278E+12	2TANG PREMIUM 1 PACK	0.090613	0.088110	15.597	2
2179	8.99278E+12	2TANG SUPER 1 PACK	0.118959	0.137136	18.492	2
2180	8.99278E+12	2TANG HIJAU 1 PACK	0.203841	0.332272	26.150	2
6550	GULA MERAH PURNAMA 1/4 KG	GULA MERAH PURNAMA 1/4 KG	0.093556	0.077591	13.306	2

Gambar 8. Hasil Data Yang telah diolah menghasilkan Cluster 1 dan Cluster 2

Setelah penulis menunjukkan hasil data yang telah diolah menggunakan K-Means , Selanjutnya penulis akan menjelaskan arti gambar gambar diatas. Penulis akan menjelaskan gambar yang telah ada , pertama akan membahas gambar 7 dan 8 , pada gambar tersebut dapat dilihat merupakan gambar yang telah diolah menggunakan analisa k means clustering , yang menghasilkan 3 jenis cluster . Cluster Pertama merupakan c0 yang merupakan cluster dengan tingkat penjualan yang rendah , kemudian Cluster Kedua merupakan c1 yang merupakan cluster dengan tingkat penjualan yang tinggi , dan yang terakhir merupakan cluster ketiga merupakan c2 yang merupakan cluster dengan tingkat penjualan yang rendah. Pada table yang di lampirkan hanya sebagian dari data yang telah ada , karena data yang ada sebanyak 7.403 data , dan data yang ditampilkan hanyalah sebagian.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan oleh penulis menarik kesimpulan bahwa, Penerapan algoritma K-Means Clustering pada data penjualan swalayan menghasilkan informasi mengenai data pengelompokkan berdasarkan minat konsumen rendah , tinggi dan juga menengah. Dari 7403 data yang telah diteliti terdapat 7345 produk yang merupakan anggota cluster 0 atau dapat diartikan sebagai produk dengan minat konsumen yang rendah . 5 produk yang termasuk kedalam cluster 1 yang dapat diartikan sebagai minat konsumen tinggi dan 53 produk yang masuk kedalam cluster 2 yang dapat diartikan sebagai produk yang memiliki minat konsumen Menengah . Penelitian ini diharapkan penulis dapat membantu toko supermarket dalam menentukan stok barang . agar stok barang yang tidak laku tidak

mengalami kelebihan stok, dan untuk barang yang laku agar tidak mengalami kekurangan stok, agar dapat meningkatkan profit. Pengelolaan stok dengan cara menganalisa data penjualan merupakan cara yang cukup efektif agar modal yang ada tidak tertumpuk kedalam barang yang tidak laku, sehingga penjualan bisa dilakukan dengan lebih optimal.

### UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada semua pihak-pihak yang telah mendukung terlaksananya penelitian ini. Terima kasih diucapkan untuk Dosen Pembimbing yang telah membimbing penulis hingga penelitian ini dapat berjalan dengan baik dan berjalan sesuai rencana.

### DAFTAR PUSTAKA

- [1] H. Agustian Siregar, N. Yanti Lumban Gaol, S. Informasi, and S. Triguna Dharma, "Penerapan Data Mining Pada Penjualan Rumah Makan Kasih Ibu Menggunakan Metode K-Means Clustering", [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsi>
- [2] R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," 2020. [Online]. Available: <https://journal-computing.org/index.php/journal-sea/index>
- [3] D. Ariyanto, "Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut," *J. Sistim Inf. dan Teknol.*, pp. 13–18, Feb. 2022, doi: 10.37034/jsisfotek.v4i1.117.
- [4] S. Baehera, Utami Dyah Syafitri, and Agus Mohamad Soleh, "Evaluasi Perbandingan Kinerja Algoritma Cheng and Church Biclustering Terhadap Algoritma Clustering Klasik K-Means untuk Mengidentifikasi Pola Distribusi Barang Ekspor Indonesia," *J. Stat. dan Apl.*, vol. 7, no. 2, pp. 149–161, 2023, doi: 10.21009/jsa.07204.
- [5] E. F. L. Awalina and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023, doi: 10.34010/jati.v13i2.10090.
- [6] J. Jabbar, "Sistem Informasi Stok Barang Menggunakan Metode Clustering Kmeans (Studi Kasus Rmd Store)," *INFOTECH J.*, vol. 8, no. 1, pp. 70–75, 2022, doi: 10.31949/infotech.v8i1.2280.
- [7] S. Handoko, F. Fauziah, and E. T. E. Handayani, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [8] N. Irsa Syahputri and S. Sundari, "Implementasi Data Mining Pada Hasil Penjualan Makanan Beku Menggunakan K Means Clustering Implementation of Data Mining on Froozen Food Sales Results Using K Means Clustering," 2023, [Online]. Available: <https://jurnal.unity-academy.sch.id/index.php/jirsi/index>
- [9] S. Marliska Hutabarat, A. Sindar, and S. Pelita Nusantara Jl Iskandar Muda No, "Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 2, no. 2, 2019.
- [10] E. Mayoana Fitri, R. Randy Suryono, and A. Wantoro, "KLAUSTERISASI DATA PENJUALAN BERDASARKAN WILAYAH MENGGUNAKAN METODE K-MEANS PADA PT XYZ."
- [11] A. Muhidin and S. Kevin Alfandara, "ANALISIS DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK CLUSTERING PENJUALAN STUDI KASUS DAPUR BU IPUNG," vol. 13, 2022.
- [12] D. F. Pasaribu, I. S. Damanik, E. Irawan, Suhada, and H. S. Tambunan, "Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, pp. 11–20, Mar. 2021, doi: 10.37148/bios.v2i1.17.
- [13] O. : Noviati, D. A. Kurnia, and A. R. Rinaldi, "Clustering Data Penjualan Produk Makanan pada Toko Toserba Yogya Siliwangi dengan Menggunakan Metode K-Means," vol. 7, no. 1, [Online]. Available: [http://ejournal.ust.ac.id/index.php/Jurnal\\_Means/](http://ejournal.ust.ac.id/index.php/Jurnal_Means/)
- [14] S. Nurajizah and A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten," *J. Tek. Komput. AMIK BSI*, vol. 7, no. 2, 2021, doi: 10.31294/jtk.v4i2.
- [15] S. Parsaoran Tamba and F. Toknady Kesuma, "PENERAPAN DATA MINING UNTUK MENENTUKAN PENJUALAN SPAREPART TOYOTA DENGAN METODE K-MEANS CLUSTERING," *J. Sist. Inf. Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 2, no. 2, 2019.
- [16] G. Triyandana, L. A. Putri, and Y. Umaidah, "Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means," 2022. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [17] I. Virgo, S. Defit, and Y. Yuhandri, "Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering," *J. Sistim Inf. dan Teknol.*, pp. 23–28, Mar. 2020, doi: 10.37034/jsisfotek.v2i1.17.