

STRATEGI PENGOLAHAN STOK BERBASIS ANALISIS CLUSTER KEPOPULERAN PRODUK

Anjar Fahrian¹, Muhammad Naufal Fauzi², Shabrina Putri Ramadhani³, Iman Paryudi⁴

Program Studi Teknik Informatika, Universitas Pancasila, Jakarta^{1,2,3,4}

4520210016@univpancasila.ac.id¹, 4520210005@univpancasila.ac.id², 4520210024@univpancasila.ac.id³, iman.paryudi@univpancasila.ac.id⁴

Abstract— Peningkatan layanan pelanggan dalam hal ketersediaan stok tetap menjadi tantangan. Seringkali, industri supermarket menghadapi masalah seperti kekurangan stok dan kelebihan persediaan, yang mengakibatkan ketidakpuasan pelanggan. Menanggapi masalah ini, dilakukan sebuah penelitian terkait manajemen stok berdasarkan analisis kluster terhadap popularitas produk, yang diharapkan dapat mengoptimalkan strategi pemasaran terkait ketersediaan produk. Penelitian ini menggunakan algoritma klustering *K-means* dengan bantuan *tools* Google Colaboratory. Hasil dari klustering menunjukkan bahwa kluster 1 mewakili kelompok produk populer, dilihat dari rata-rata jumlah produk yang terjual dan rata-rata tingkat kepuasan pelanggan yang relatif lebih tinggi dibandingkan dengan kluster 0 dan 2. Kluster 0 mewakili kelompok produk yang cukup populer, sedangkan kluster 2 mewakili kelompok produk yang kurang populer. Berdasarkan visualisasi yang dilakukan, dapat diamati bahwa popularitas produk di setiap kota menunjukkan bahwa produk elektronik paling populer di Mandalay dan Naypyidaw, sedangkan produk kesehatan dan kecantikan kurang populer atau termasuk dalam kluster 2 di Naypyidaw. Namun, produk kesehatan dan kecantikan menjadi produk populer dengan penjualan tertinggi di Yangon.

Kata Kunci — Pelanggan, Layanan, Ketersediaan, Kluster, Supermarket

I. PENDAHULUAN

Sandang, pangan, dan papan adalah kebutuhan dasar manusia, dimana zaman berkembang makin modern sehingga keperluan manusia juga akan bermacam – macam [1]. Perbedaan kebutuhan sandang, pangan, dan papan pada berbagai wilayah dapat dipengaruhi oleh faktor-faktor seperti ketersediaan sumber daya, iklim, tingkat ekonomi, dan budaya setempat [2]. Dalam menghadapi keanekaragaman ini, industri ritel harus mengadopsi strategi yang dapat menyesuaikan penawaran produk mereka sesuai dengan preferensi lokal. Banyak negara saat ini berupaya mendorong pertumbuhan ekonominya dengan meningkatkan output secara konsisten melalui penyediaan barang modal, teknologi, dan sumber daya manusia [3]. Tidak semua negara berhasil mencapai pertumbuhan ekonomi yang diharapkan. Hal ini disebabkan oleh beberapa faktor, salah satunya adalah ketidakmampuan negara tersebut untuk memenuhi kebutuhannya secara mandiri

di dalam negeri. Jika suatu negara gagal meningkatkan pertumbuhan ekonominya, hal ini dapat memicu berbagai masalah baru, baik ekonomi maupun sosial, seperti meningkatnya tingkat korupsi dan inflasi [4].

Strategi manajemen stok yang efektif dan fleksibel menjadi penting bagi retailer, terutama supermarket, untuk memastikan bahwa produk yang relevan selalu tersedia sesuai dengan permintaan setempat. Data dalam laporan "State of the Retail Supply Chain Report 2020" menunjukkan beberapa tren penting dalam strategi rantai pasokan untuk industri ritel, terutama yang terkait dengan kebutuhan akan digitalisasi dan efisiensi operasional. Strategi supply chain untuk industri ritel kini semakin mengarah pada otomatisasi pemenuhan, investasi dalam digitalisasi rantai pasokan, dan kolaborasi erat dengan pemasok untuk mengelola risiko. Langkah-langkah seperti diversifikasi sumber pemasok dan kolaborasi finansial dengan pemasok menjadi lebih penting untuk mempertahankan stabilitas rantai pasokan, terutama dalam situasi krisis[5].

Dengan memahami bahwa kebutuhan sandang, pangan, dan papan di berbagai wilayah dipengaruhi oleh faktor seperti ekonomi dan sumber daya lokal, ritel perlu menerapkan strategi pengelolaan stok yang fleksibel dan adaptif. Dalam era globalisasi berbasis teknologi, digitalisasi telah mendorong transformasi signifikan dalam industri ritel. Teknologi digital membantu memahami preferensi pelanggan dengan lebih baik, memungkinkan penyesuaian manajemen stok dan distribusi berdasarkan faktor lokal seperti iklim, budaya, dan ekonomi.

II. TINJAUAN PUSTAKA

2.1 Supermarket

Supermarket adalah pasar swalayan atau toko besar yang menjual sebagian besar jenis makanan dan barang-barang lain yang dibutuhkan untuk rumah tangga [5]. Menurut Kotler dan Armstrong (1996), supermarket adalah organisasi ritel yang menawarkan berbagai lini produk, seperti pakaian, perabot rumah tangga, dan barang kebutuhan rumah tangga, di mana setiap lini dikelola sebagai departemen terpisah oleh pembeli atau pengelola khusus.

2.2 Clustering

Clustering merupakan pengelompokan data yang dilakukan tidak berdasarkan kelas data tersebut kedalam kelas objek yang sama sesuai dengan topik yang diambil dalam suatu

penelitian[6]. Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan (kemiripan). Sebuah cluster tidak harus sama akan tetapi pengelompokannya berdasarkan pada kedekatan dari suatu karakteristik sampel yang ada salah satunya dengan menggunakan rumus jarak Eclidean[7].

2.3 Algoritma K-Means

Algoritma k-means adalah metode yang sederhana untuk mengklasifikasikan atau mengelompokkan sejumlah besar objek berdasarkan atribut tertentu ke dalam beberapa kelompok (cluster) sebanyak K. Dalam algoritma ini, jumlah cluster K ditentukan sebelumnya. K-means termasuk dalam metode clustering non-hirarki yang bertujuan membagi data ke dalam cluster sedemikian rupa sehingga data dengan karakteristik serupa dikelompokkan dalam cluster yang sama, sementara data dengan karakteristik yang berbeda dimasukkan ke dalam cluster yang berbeda[7].

Dalam setiap cluster memiliki karakter yang sama namun antar cluster diciptakan tingkat variasi yang kecil, sehingga memberikan karakteristik yang berbeda. Proses clustering mempergunakan algoritma K-Means secara umum mempunyai tahapan, yakni:[8]

- Memasukkan data
- Menentukan jumlah *cluster*
- Mengambil data secara acak berdasarkan jumlah cluster yang akan digunakan untuk *centroid*.
- Menghitung jarak antara data dengan *centroid* menggunakan persamaan *euclidean distance* berikut:

$$D_{ij} = \sqrt{(x_{1j} - v_{1j})^2 + \dots + (x_{ij} - v_{kj})^2}$$

Dimana:

Dij = jarak data ke i ke *centroid* ke k

Xij = data ke i, pada atribut ke j

Vkj = titik pusat ke k, pada atribut ke j

- Menghitung centroid dengan anggota *cluster* yang baru.

- Jika *centroid* tidak berubah maka proses cluster sudah selesai, sedangkan jika *centroid* berubah maka ulangi langkah menghitung jarak pada poin d sampai *centroid* tidak berubah lagi.

III. METODOLOGI

Metode penelitian adalah serangkaian langkah dan metode yang dilakukan selama penelitian mulai dari tahap awal pengumpulan data hingga hasil penelitian. Berikut beberapa tahapan dalam penelitian yang akan dilalui.

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data yang diambil dari Kaggle berupa data supermarket dari 3 kota yang berbeda di Myanmar.

3.2 Pengecekan Missing Value

Pada tahap ini dilakukan pengecekan terhadap data yang sekiranya memiliki missing value. Hasil menunjukan bahwa data bersih dari missing value.

3.3 Seleksi Data

Sebelum melakukan pemodelan, terlebih dahulu menyeleksi variabel-variabel dari data yang akan digunakan. Variabel yang

digunakan dalam proses clustering meliputi data-data numerik seperti unit price, tax, rating, dan lain-lain.

3.4 Normalisasi Data

Variabel yang digunakan dalam proses clustering memiliki skala yang beragam sehingga akan lebih baik dilakukan normalisasi demi hasil yang maksimal. Normalisasi yang digunakan adalah metode normalisasi Min-max.

3.5 Implementasi Algoritma

Setelah melakukan normalisasi pada data yang akan digunakan, tahap selanjutnya adalah implementasi dari algoritma *K-means*. Tahapan dari implementasi algoritma sebagai berikut.

- Meng-import library *scikit-learn*.
- Menggunakan data *frame* baru yang berisi variabel-variabel yang telah diseleksi dan dinormalisasi.
- Menentukan n *cluster*. Pada penelitian ini, n *cluster* ditentukan menggunakan metode *elbow*.
- Melakukan *cluster*.
- Hasil cluster kemudian dimasukkan ke dalam dataset dengan kolom baru yang dinamakan “cluster”.

IV. HASIL DAN PEMBAHASAN

Langkah pertama yang dilakukan sebelum memulai penelitian adalah dengan melakukan pengecekan terhadap missing value seperti pada gambar berikut.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 17 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Invoice ID             1000 non-null   object
 1   Branch                1000 non-null   object
 2   City                  1000 non-null   object
 3   Customer type         1000 non-null   object
 4   Gender                1000 non-null   object
 5   Product line          1000 non-null   object
 6   Unit price            1000 non-null   float64
 7   Quantity              1000 non-null   int64
 8   Tax 5%                1000 non-null   float64
 9   Total                 1000 non-null   float64
10   Date                  1000 non-null   object
11   Time                  1000 non-null   object
12   Payment               1000 non-null   object
13   cogs                  1000 non-null   float64
14   gross margin percentage 1000 non-null   int64
15   gross income          1000 non-null   float64
16   Rating                1000 non-null   float64
dtypes: float64(6), int64(2), object(9)
memory usage: 132.9+ KB
```

Gambar 4. 1 Pengecekan Missing Value

Hasil menunjukan bahwa tidak ditemukan adanya *missing value* pada data sehingga bisa dilanjutkan ke tahap selanjutnya yaitu seleksi dan normalisasi data.

Normalisasi data pada penelitian ini menggunakan metode *Min-max* yang dilakukan lewat *google colab*. Rumus dari metode normalisasi *Min-max* adalah sebagai berikut.

$$X_{norm} = \frac{x' - \min(x)}{\max(x) - \min(x)} (\text{new}_{\max}(x) - \text{new}_{\min}(x)) + \text{new}_{\min}(x)$$

Setelah normalisasi, didapat data yang konsisten dan siap untuk diklasterisasi.

Langkah berikutnya adalah menentukan jumlah *cluster* dengan menggunakan metode *elbow*. Implementasi metode *elbow* diperlukan guna memperoleh jumlah *cluster* yang optimal. Implementasi metode *elbow* pada *python* dapat dilihat pada kode berikut.

```
cluster_range=range(1,10)
cluster_wss=[]
```

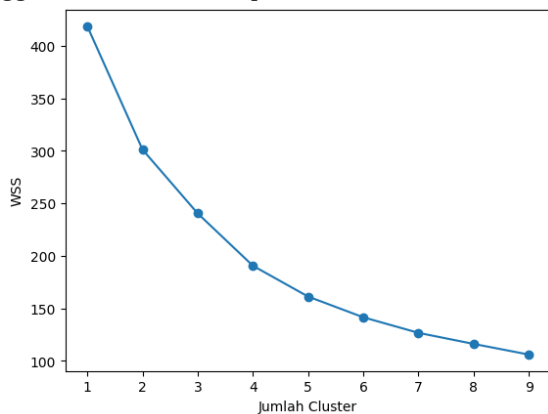
```
for num_cluster in cluster_range:
```

```
clusters=KMeans(num_cluster)
clusters.fit(x)
cluster_wss.append(clusters.inertia_)
```

```
plt.xlabel("Jumlah Cluster")
plt.ylabel("WSS")
plt.plot(cluster_range,cluster_wss, marker="o")
```

```
plt.show()
```

Variabel “x” pada kode di atas merupakan *data frame* yang berisikan data yang telah melalui proses seleksi dan normalisasi sehingga telah bersih dan siap untuk diklasterisasi.



Gambar 4. 2 Elbow Method

Dari visualisasi di atas, kemudian diambil kesimpulan bahwa jumlah n cluster yang maksimal adalah 3 cluster.

Tahap akhir pada proses klasterisasi adalah dengan mengimplementasi data ke dalam algoritma. Proses tersebut dapat dilihat pada kode berikut.

```
cls=KMeans(n_clusters=3, random_state=0)
model_cls=cls.fit(x)
```

Dengan jumlah cluster sebanyak 3, diperoleh hasil cluster sebagai berikut

```
[ ] df_1['cluster'].value_counts()

2    527
1    383
0     90
Name: cluster, dtype: int64
```

Gambar 4. 3 Jumlah Anggota per Cluster

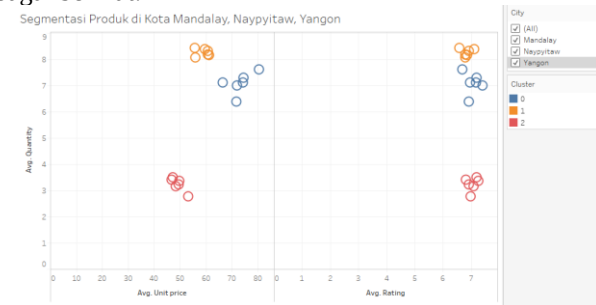
Diperoleh hasil cluster 0 sebanyak 90, cluster 1 sebanyak 383, dan cluster 2 sebanyak 527. Jika ditampilkan kedalam bentuk tabel maka akan terlihat hasil *clustering* sebagai berikut.

Tabel 4. 1 Hasil Cluster

No.	Unit price	Quantit y	Tax 5%	Total	cogs	gross income	Rating	cluster
0	0,72	0,67	0,58	0,58	0,52	0,58	0,85	0

1	0,06	0,44	0,00	0,00	0,07	0,00	0,93	2
2	0,40	0,67	0,36	0,36	0,32	0,36	0,57	0
3	0,54	0,78	0,05	0,05	0,46	0,05	0,73	1
4	0,85	0,67	0,67	0,67	0,60	0,67	0,22	0
5	0,84	0,67	0,67	0,67	0,60	0,67	0,02	0

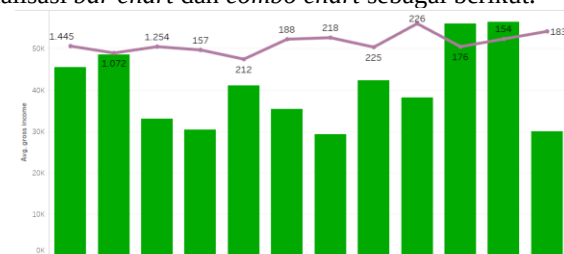
Kemudian untuk menentukan interpretasi dari kelompok cluster 0, 1, dan 2 dilakukan visualisasi dengan *scatter plot* sebagai berikut.



Gambar 4. 4 Segmentasi Produk

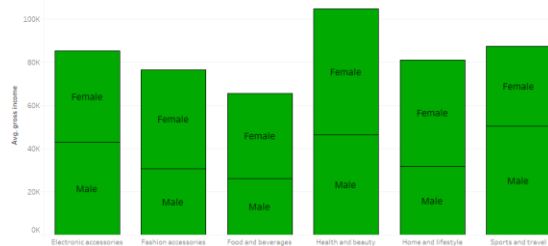
Berdasarkan hasil visualisasi kemudian dengan melihat dari kuantitas pembelian produk, harga per unit, dan rating, dapat diambil kesimpulan bahwa kelompok cluster 1 menyatakan kelompok produk yang populer karena rata-rata kuantitas pembelian dan rating yang lebih tinggi dibanding kelompok cluster lainnya. Kemudian kelompok cluster 0 menyatakan kelompok produk yang cukup populer karena letaknya yang berada di tengah-tengah. Kemudian cluster 2 menyatakan kelompok produk yang kurang populer karena rata-rata kuantitas pembeliannya yang berada jauh di bawah kelompok cluster lain.

Untuk mempermudah dalam mengambil keputusan terkait segmentasi produk, diperlukan pemahaman terhadap data. Oleh karena itu dilakukan beberapa visualisasi data untuk melihat pola dalam data. Visualisasi data yang dilakukan meliputi jenis visualisasi *bar chart* dan *combo chart* sebagai berikut.



Gambar 4. 5 Perkembangan Penjualan

Grafik di atas menunjukkan pola perkembangan penjualan selama 1 tahun dilihat dari jumlah pendapatan kotor (*gross income*) dan kuantitas produk yang terjual. Selain ini, dilakukan visualisasi lain terkait pendapatan kotor berdasarkan gender dari konsumen sebagai berikut.



Gambar 4. 6 Grafik Pendapatan berdasarkan gender

Dari visualisasi tersebut memperlihatkan jumlah *income* yang didapatkan dari aktivitas pembelian konsumen berdasarkan jenis kelamin. Konsumen berjenis kelamin perempuan mendominasi jumlah pendapatan kotor pada setiap kategori produk, kecuali pada kategori “Sports and travel” yang jumlah *income*-nya lebih tinggi diperoleh dari konsumen berjenis kelamin laki-laki.

Selain itu, dilakukan juga visualisasi terhadap kategori produk dengan rata-rata rating tertinggi berdasarkan kelompok *cluster*-nya sebagai berikut.

Tabel 4. 2 Rating Tiap Kategori Produk

Cluster	Electronic accessories	Fashion accessories	Food and beverages	Health and beauty	Home and lifestyle	Sports and travel
0	6.6846	6.9563	6.9231	7.3947	7.1846	7.1625
1	6.5800	7.1183	6.9288	6.7902	6.7924	6.8462
2	7.2022	6.9882	7.2674	7.0806	6.8185	6.9235

Dapat dilihat bahwa rata-rata rating terendah berada pada kelompok *cluster* 1 dengan kategori produk elektronik, sedangkan rata-rata rating tertinggi ada pada kelompok *cluster* 0 dengan kategori produk kesehatan dan kecantikan.

V. KESIMPULAN

Penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam meningkatkan pemahaman tentang perilaku konsumen, mengoptimalkan strategi pengelolaan stok, dan memberikan dasar yang kuat untuk pengambilan keputusan yang lebih baik dalam konteks industri supermarket di Myanmar.

Dari hasil pembahasan yang telah dijabarkan sebelumnya, maka optimasi yang dapat dilakukan terkait pengelolaan stok produk adalah sebagai berikut:

1. Pada kota Mandalay menunjukkan kelompok *cluster* 0 atau kelompok produk cukup populer memiliki nilai kuantitas penjualan produk yang lebih tinggi dibanding kota lain. Rating dari *cluster* 0 di kota Mandalay pun relatif tinggi sehingga kelompok *cluster* 0 merupakan kelompok produk yang populer hanya pada kota tersebut. Kategori produk yang paling banyak dibeli merupakan produk elektronik sehingga penyediaan stok produk tersebut di kota Mandalay perlu ditingkatkan.

2. Segmentasi produk di kota Naypyaw menunjukkan produk elektronik pada kelompok *cluster* 1 memiliki nilai kuantitas penjualan tertinggi sehingga penyediaan stok produk elektronik pada kota tersebut perlu ditingkatkan. Sedangkan untuk produk health and beauty menunjukkan kuantitas penjualan yang sangat rendah dan termasuk kelompok *cluster* produk yang kurang populer sehingga penyediaan stok produk

ini pada kota Naypyaw perlu lebih dikelola agar tidak terjadi kelebihan stok.

3. Pada Kota Yangon, menunjukan kelompok *cluster* 1 Atau kelompok produk populer memiliki nilai kuantitas penjualan produk health and beauty sehingga penyediaan stok ke depannya dapat di sesuaikan dengan permintaan konsumen pada riwayat transaksi sebelumnya. Sedangkan untuk produk Fashion and accessories menunjukan riwayat pembelian yang sangat rendah dan termasuk *cluster* 2 atau kelompok kurang populer sehingga dapat menyesuaikan permintaan stok pada segment fashion and accessories di Kota Yangon.

REFERENCES

- [1] R. A. Chintia And R. Destiningsih, “Pengaruh Harga Komoditas Pangan Terhadap Inflasi Di Kota Semarang,” *Jurnal Ilmiah Ekonomi Bisnis*, vol. 27, no. 2, pp. 244–258, 2022, doi:10.35760/eb.2022.v27i2.4948.
- [2] A. A. K. Sumunar and S. Budiman, “Proyeksi Ketahanan Pangan Provinsi Nusa Tenggara Timur Tahun 2015-2045 Ditinjau Dari Ketersediaan, Kebutuhan, Dan Persediaan Beras,” *SEPA: Jurnal Sosial Ekonomi Pertanian dan Agribisnis*, vol. 18, no. 1, p. 80, Sep. 2021, doi: 10.20961/sepa.v18i1.47131.
- [3] A. R. Suleman *et al.*, *Perekonomian Indonesia*, 1st ed., vol. 1. 2021.
- [4] Darwin Damanik and M. Saragih, “Korupsi, Inflasi dan Pertumbuhan Ekonomi di ASEAN,” *Jurnal Ekuilnomi*, vol. 5, no. 1, May 2023, doi: 10.36985/ekuilnomi.v5i1.494.
- [5] E. L. Tandipayuk, “Analisis Yuridis Terhadap Pembangunan Minimarket Yang Jaraknya Berdekatan Dengan Usaha Toko Tradisional Ditinjau Dari Peraturan Presiden No. 112 Tahun 2007 1.” [Online]. Available: [https://kumparan.com/kabar-harian/pasar-](https://kumparan.com/kabar-harian/pasar-S. R. Hani,)
- [6] S. R. Hani, “Clustering Data Pencari Kerja Menurut Tingkat Pendidikan Menggunakan Algoritma K-Means,” *Jurnal Minfo Polgan*, vol. 12, no. 1, pp. 1–14, Mar. 2023, doi: 10.33395/jmp.v12i1.12217.
- [7] Rindi Asti Ananda, Yani Maulita, and Husnul Khair, “Clustering Menggunakan Algoritma K-Means untuk Mengelompokan Data Perjudian Berdasarkan Wilayah di Kota Binjai (Studi Kasus : Pengadilan Negeri Binjai),” *Switch : Jurnal Sains dan Teknologi Informasi*, vol. 2, no. 4, pp. 167–180, Sep. 2024, doi: 10.62951/switch.v2i4.226.
- [8] M. Maesaroh, T. Nur Padilah, And J. Haerul Jaman, “Penerapan Algoritma K-Means Clustering Pada Pengelompokan Daerah Penyebaran Diare Di Provinsi Jawa Barat,” 2023.