

Analisis Keadaan Banyaknya Jenis Produk Terhadap Total Penjualan pada Supermarket

Nama Anggota:

- Andre Budiman / 2440050344
- Fabian / 2401961942
- Ricky Aryatama Yaputera / 2401955126
- Kenny / 2401959111
- William Valentino Ricard / 2440041674

BAB 1. PENDAHULUAN

1.1. Latar Belakang

Pasar merupakan tempat bertemunya antara penjual dan pembeli untuk melakukan suatu transaksi. Di pasar ini, penjual akan menjajakan barang dagangannya untuk dapat dilihat dan dibeli oleh pelanggan. Pasar di sini memiliki beberapa fungsi, yaitu sebagai sarana untuk hasil distribusi dari produsen, sehingga pasar bekerja sebagai pihak ketiga antara produsen dengan konsumen. Lalu, fungsi yang kedua adalah sebagai tempat untuk menciptakan *trend* untuk barangnya atau tempat menetapkan harga. Lalu, fungsi yang terakhir adalah sebagai tempat untuk menjalin komunikasi antar berbagai pihak, baik antar pegawai dengan pembeli atau pegawai dengan pegawai ataupun antar pembeli dengan pembeli.

Pasar sendiri dibagi menjadi beberapa jenis, ada yang pasar yang dibedakan sesuai transaksi, pasar yang dibedakan sesuai bentuk kegiatannya, dan pasar yang dibedakan sesuai produk yang dijual. Jika kita melihat ke salah satu jenis pasar, yaitu pasar yang dibedakan sesuai jenis transaksinya, maka kita bisa mendapat 2 jenis pasar lagi, yaitu pasar tradisional dan pasar modern. Menurut Wicaksono dkk. (2011), pasar tradisional merupakan tempat bertemunya penjual dan pembeli serta ditandai dengan adanya transaksi penjual pembeli secara langsung. Sedangkan menurut Sinaga (2006), pasar modern adalah pasar yang dikelola dengan manajemen modern, umumnya terdapat di kawasan perkotaan, sebagai penyedia barang dan jasa dengan mutu dan pelayanan yang baik kepada konsumen. Menurut Pengantar Bisnis: Merajut Bisnis Lokal Menuju Bisnis Global oleh I Putu Artaya dan Sulistyani Eka Lestari (2021:63), pasar modern terdiri dari beberapa jenis, yaitu supermarket, minimarket, hypermarket, grosir, dan toko serba ada.

Supermarket atau pasar swalayan adalah suatu pasar yang menjual secara eceran produk-produk kebutuhan sehari-hari seperti makanan, minuman, sabun, shampo, perabot rumah tangga, dan kebutuhan lainnya. Supermarket juga kerap disebut sebagai toserba atau toko serba ada karena supermarket sendiri menjual berbagai produk-produk yang sering dipakai oleh masyarakat. Di supermarket sendiri, jenis produk yang dijual sangatnya beragam serta memiliki kualitas yang baik dan dalam supermarket juga tidak terjadi aktivitas tawar-menawar seperti yang ada di pasar tradisional karena setiap produk yang dijual sudah memiliki harga tetap dan terdapat barcode di setiap produknya.

Tentunya supermarket sendiri harus memiliki rencana yang matang dan sempurna, dikarenakan biaya operasional yang ditanggung oleh perencana bisnis tentunya akan sangat besar jika dibandingkan oleh minimarket. Oleh karena itu, perlu ada rencana dan analisis terkait dengan transaksi yang terjadi di supermarket untuk tetap menjaga kesehatan dari ekonomi supermarket tersebut. Oleh karena itu, kami ingin mencoba untuk menganalisis keadaan yang terjadi di supermarket, lebih tepatnya apakah banyaknya jenis produk dengan tingkat penjualan sudah sebanding, dengan menggunakan teknik clustering. Dengan kita menganalisis apakah banyak jenis produk berpengaruh terhadap suatu tingkat penjualan, maka kita mampu mengatasi kemungkinan terburuk, yaitu produk yang menumpuk di toko ataupun gudang.

1.2. Tujuan

- a) Untuk mengetahui apakah banyaknya jenis produk sudah sebanding dengan penjualan pada Supermarket.
- b) Untuk membuat algoritma cluster analysis yang dapat mengolah dan menganalisis data dari dataset Supermarket.
- c) Untuk mengetahui berbagai data yang memiliki karakteristik atau nilai yang mirip lalu dibagi menjadi cluster-cluster.

1.3. Manfaat

Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

- a) Bagi Penulis
Penelitian ini diharapkan dapat digunakan oleh penulis sebagai media untuk mengaplikasikan ilmu yang telah diperoleh di perkuliahan.
- b) Bagi Peneliti
Penelitian ini diharapkan dapat digunakan oleh peneliti lainnya sebagai referensi untuk penelitian selanjutnya.
- c) Bagi Supermarket
Penelitian ini diharapkan dapat dijadikan sebagai acuan bagi supermarket untuk meningkatkan penjualan dan menyesuaikan jenis produk yang akan dijual.

BAB 2. METODOLOGI

2.1. Dataset

Dataset yang digunakan diambil dari Kaggle dengan nama dataset Supermarket store branches sales analysis yang mempunyai 896 data. Adapun alasan dari kami untuk mengambil data set tersebut dikarenakan dataset tersebut merupakan data yang sesuai dengan topik kami, yaitu memiliki data penjualan serta faktor dari penjualan yang ingin kami analisis, yaitu banyaknya jenis produk. Contoh data dari dataset dapat dilihat pada gambar di bawah.

	Store ID	Store_Area	Items_Available	Daily_Customer_Count	Store_Sales
0	1	1659	1961	530	66490
1	2	1461	1752	210	39820
2	3	1340	1609	720	54010
3	4	1451	1748	620	53730
4	5	1770	2111	450	46620

Tautan dari dataset yang diambil dari Kaggle.

(<https://www.kaggle.com/datasets/surajjha101/stores-area-and-sales-data>)

2.2. Deskripsi Dataset

Store_ID: (Indeks) ID dari toko tertentu.

Store_Area: Luas toko dalam satuan yard kuadrat.

Items_Available: Jumlah item berbeda yang tersedia di toko yang sesuai.

DailyCustomerCount: Jumlah rata-rata pelanggan yang berkunjung ke toko selama sebulan.

Store_Sales: Penjualan dalam (US \$) yang dilakukan toko.

Tipe data:

Store_ID char (3)

Store_Area int

Items_Available int

DailyCustomerCount int

Store_Sales int

Dari dataset di atas, kami akan menggunakan data dari 2 field yaitu Items_Available dan Store_Sales.

2.3. Proses pengolahan

Melalui dataset yang digunakan kami menentukan hyperparameter yang akan digunakan yaitu Items_Available dan Store_Sales yang dimana Items_Available bisa mempengaruhi Store_Sales. Untuk mendukung analisis yang akan dilakukan kami menggunakan bahasa pemrograman Python dengan menggunakan library-library seperti sklearn, pandas,

matplotlib. sklearn berfungsi untuk membantu melakukan processing data atau training data dalam machine learning, pandas menyediakan alat analisis dan pengolahan data dengan performa tinggi serta data yang terstruktur dengan baik, dan matplotlib untuk membantu dalam visualisasi data. Melalui library sklearn melakukan import KMeans yang menjadi metode untuk clustering, MinMaxScaler untuk melakukan penskalaan data dalam skala 0-1, LabelEncoder untuk melakukan labeling data. Melalui library matplotlib melakukan import pyplot yang membantu membuat beberapa perubahan pada gambar seperti membuat gambar, membuat area plot dalam gambar, menambah label di plot dan lainnya.

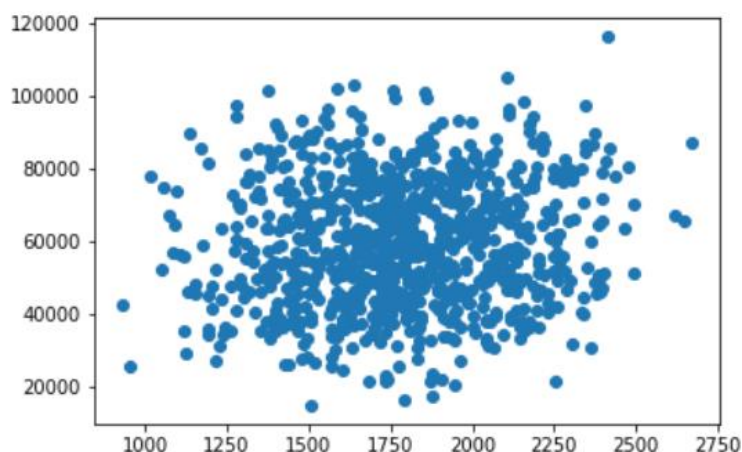
```
from sklearn.cluster import KMeans
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from matplotlib import pyplot as plt
from sklearn.preprocessing import LabelEncoder
%matplotlib inline
```

Pertama, mengambil dataset dengan melakukan read csv

```
df = pd.read_csv('../input/stores-area-and-sales-data/Stores.csv')
```

Kedua, melakukan visualisasi data dengan melihat persebaran data menggunakan scatter plot, menghasilkan tampilan seperti di bawah ini.

```
plt.scatter(df['Items_Available'], df['Store_Sales'])
```



Ketiga, melakukan penskalaan data dengan menggunakan MinMaxScaler sehingga nilai dari data Items_Available dan Store_Sales berada dalam skala 0-1 sehingga data lebih ternormalisasi.

```
scaler = MinMaxScaler()
scaler.fit(df[['Store_Sales']])
df['Store_Sales'] = scaler.transform(df[['Store_Sales']])

scaler.fit(df[['Items_Available']])
df['Items_Available'] = scaler.transform(df[['Items_Available']])
```

Tampilan data yang telah dilakukan penskalaan.

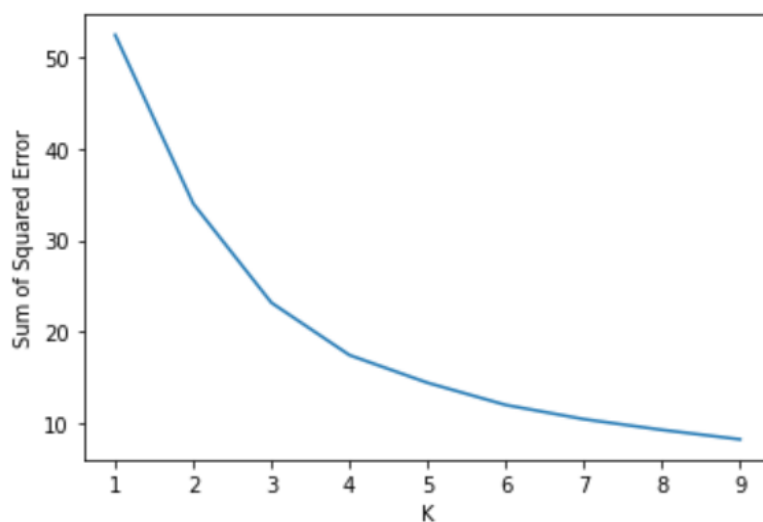
	Store ID	Store_Area	Items_Available	Daily_Customer_Count	Store_Sales
0	1	1659	0.593084	530	0.508580
1	2	1461	0.472622	210	0.245562
2	3	1340	0.390202	720	0.385503
3	4	1451	0.470317	620	0.382742
4	5	1770	0.679539	450	0.312623

Keempat, menentukan banyaknya cluster dengan menggunakan elbow method. Elbow method membutuhkan menggambar plot garis antara SSE (Dalam-cluster Sum of Squared error) dengan jumlah cluster.

```
k_range = range(1,10)
sse = []

for k in k_range:
    km = KMeans(n_clusters=k)
    km.fit(df[['Items_Available', 'Store_Sales']])
    sse.append(km.inertia_)

plt.xlabel('K')
plt.ylabel('Sum of Squared Error')
plt.plot(k_range, sse)
```



Melalui hasil plot garis dari elbow method dapat dilihat titik siku terbaik terdapat pada $k = 3$, sehingga data akan di clustering ke dalam 3 cluster.

Kelima, melakukan prediksi clustering melalui dua variable yang telah ditentukan yaitu Items_Available dan Store_Sales.

```
km = KMeans(n_clusters=3)
y_predicted = km.fit_predict(df[['Items_Available', 'Store_Sales']])
```

Menambahkan kolom cluster dalam table agar dapat melihat hasil prediksi clustering dalam bentuk kolom.

```
df['cluster'] = y_predicted
df.head()
```

Hasil prediksi akan tertampil dalam kolom seperti di bawah ini.

	Store ID	Store_Area	Items_Available	Daily_Customer_Count	Store_Sales	cluster
0	1	1659	0.593084	530	0.508580	2
1	2	1461	0.472622	210	0.245562	1
2	3	1340	0.390202	720	0.385503	1
3	4	1451	0.470317	620	0.382742	1
4	5	1770	0.679539	450	0.312623	2

Keenam, melakukan visualisasi dari hasil prediksi clustering.

```
km.cluster_centers_
```

cluster_center berfungsi untuk melihat koordinat centroid dari setiap cluster dan menghasilkan tampilan seperti di bawah ini.

```
array([[0.42789974, 0.59962538],
       [0.38450228, 0.29378607],
       [0.68099135, 0.43282446]])
```

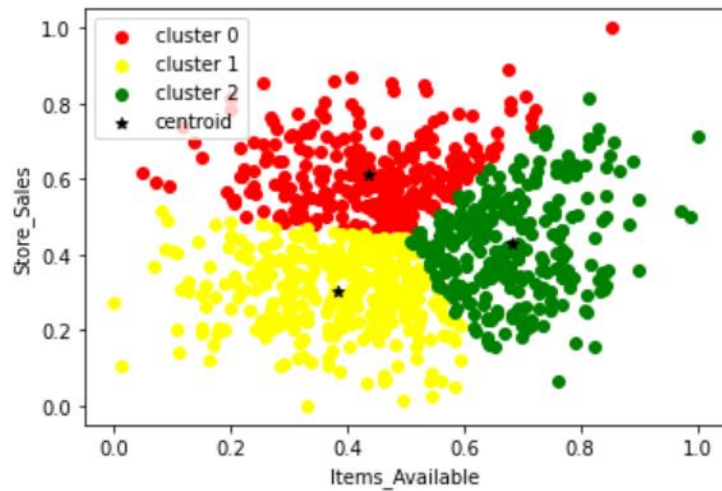
```
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]

plt.scatter(df1['Items_Available'], df1['Store_Sales'], color='red', label='cluster 0')
plt.scatter(df2['Items_Available'], df2['Store_Sales'], color='yellow', label='cluster 1')
plt.scatter(df3['Items_Available'], df3['Store_Sales'], color='green', label='cluster 2')

plt.scatter(km.cluster_centers_[0], km.cluster_centers_[1], color='black', marker='*', label='centroid')

plt.xlabel('Items_Available')
plt.ylabel('Store_Sales')
plt.legend()
```

Pemberian warna pada masing-masing cluster dan penentuan bentuk serta warna centroid. Lalu menghasilkan visualisasi cluster seperti di bawah ini.



Ketujuh, untuk melihat hasil clustering secara lebih mudah maka dilakukan pengelompokan data hasil clustering dengan rata-rata dari masing-masing tiap cluster.

```
selected_cols = ['Items_Available', 'Store_Sales']
cluster_data = df.loc[:, selected_cols]

kmeans_sel = KMeans(init='k-means++', n_clusters=3, n_init=100, random_state=2).fit(cluster_data)
labels = pd.DataFrame(kmeans_sel.labels_)
clustered_data = cluster_data.assign(Cluster=labels)

grouped_km = clustered_data.groupby(['Cluster']).mean().round(1)
grouped_km
```

	Items_Available	Store_Sales
Cluster		
0	0.4	0.6
1	0.4	0.3
2	0.7	0.4

BAB 3. PENUTUP

3.1. Hasil dan Kesimpulan

Hasil yang didapatkan terbagi menjadi 3 cluster, yaitu cluster 0, cluster 1, dan cluster 2. Pada cluster 0, nilai rata-rata yang didapatkan adalah 0.4 untuk Items_Available dan 0.6 untuk Store_sales. Pada cluster 1, nilai rata-rata yang didapatkan adalah 0.4 untuk Items_Available dan 0.3 untuk Store_Sales. Pada cluster 2, nilai rata-rata yang didapatkan adalah 0.7 untuk Items_Available dan 0.4 untuk Store_Sales.

Kesimpulan yang bisa didapat dari descriptive analysis dengan menggunakan clustering adalah cluster 0 merupakan cluster dengan penjualan (Store_Sales) terbesar dengan nilai rata-rata 0.7 dengan nilai rata-rata banyak jenis produk (Items_Available) 0.4 yang merupakan nilai rata-rata paling kecil untuk banyaknya jenis produk. Sedangkan, cluster 2 memiliki penjualan (Store_Sales) 0.4 dengan nilai rata-rata banyak jenis produk (Items_Available) 0.7 yang merupakan nilai rata-rata paling besar untuk banyaknya jenis produk. Kedua cluster ini memiliki hasil yang sangat berlawanan, di mana jika kita kaitkan dengan kenyataannya, cluster 0 memiliki penjualan yang lebih banyak dibandingkan cluster 2, padahal cluster 2 memiliki jenis produk yang lebih banyak. Dengan demikian, kita bisa memindahkan beberapa jenis produk ke cluster 0 dari cluster 2 dikarenakan cluster 0 memiliki penjualan yang lebih besar dibandingkan cluster 2.

3.2. Hal yang Perlu Ditingkatkan

Hasil clustering menunjukkan adanya perbedaan yang cukup signifikan dari Items_Available dengan Store_Sales, yang seharusnya semakin banyak jenis produk maka penjualan seharusnya juga menjadi semakin besar. Maka dari itu diperlukan penelitian lanjutan untuk mengetahui faktor yang mempengaruhi rendahnya Store_Sales pada Supermarket di cluster 2 meskipun Items_Available tergolong banyak.