

ANALISIS SEGMENTASI KLIEN MENGGUNAKAN K-MEANS CLUSTERING PADA AGENSI KREATIF ANOORAPRO

Andreas Bayu Krisna, Devita Maulina Putri
Sistem Informasi, Universitas Merdeka Malang
Jalan Terusan Dieng, Malang, Indonesia
devita.maulina@unmer.ac.id

ABSTRAK

Pemanfaatan media sosial dalam pemasaran digital telah menjadi salah satu strategi utama bagi banyak perusahaan. Semakin banyak pengguna media sosial untuk berbisnis, agensi kreatif juga mulai menawarkan jasa mereka. Anoorapro merupakan sebuah agensi kreatif yang bergerak di bidang visual dan komunikasi kreatif. Karena klien Anoorapro semakin bertambah, pihak mereka ingin mendalami kebutuhan dan berencana untuk melakukan marketing kepada klien. Oleh karena itu, segmentasi klien perlu dilakukan untuk memahami karakteristik dan kebutuhan para klien. Dengan pemahaman tersebut, pihak Anoorapro dapat terbantu untuk membuat strategi pemasaran yang lebih sesuai dan efektif. Dalam penelitian ini, *clustering* akan menggunakan algoritma K-Means. Klien-klien tersebut akan dikelompokkan berdasarkan total anggaran, frekuensi dan recency. Untuk alur proses akan menggunakan metode SEMMA. Hasil dari penerapan *Elbow method* menunjukkan nilai $k = 4$, tetapi terdapat 3 *cluster outlier* yang memiliki nilai bisnis tinggi sehingga tetap dipertahankan dan menghasilkan total 7 *cluster client*. Ketujuh *cluster* tersebut memberikan pemahaman bahwa diperlukan strategi pendekatan yang berbeda agar dapat meningkatkan kepuasan client dan efektivitas pemasaran.

Kata kunci: *Clustering, K-Means, Segmentasi klien*

1. PENDAHULUAN

Dengan adanya internet banyak sekali kemudahan yang yang kita dapatkan pada jaman sekarang. Teknologi internet memberikan dampak yang sangat signifikan terutama pada sektor bisnis. Kehadiran teknologi internet telah merubah wajah bisnis yang semula di dominasi transaksi secara offline secara perlahan bergerak ke arah pemasaran online [1]. Pemanfaatan media sosial dalam pemasaran digital telah menjadi salah satu strategi utama bagi banyak perusahaan dalam beberapa tahun terakhir [2]. Bertambahnya pengguna media sosial untuk bisnis dan bertambahnya kebutuhan informasi tentang suatu produk, agensi kreatif juga mulai menawarkan jasa mereka. Salah satunya adalah Anoorapro.

Anoorapro merupakan sebuah agensi kreatif yang berlokasi di Perumahan De Prima Blok AA-12, Tunggulwulung, Lowokwaru, Malang. Anoorapro sendiri bergerak di bidang visual dan komunikasi kreatif. Sudah banyak brand-brand yang telah bekerja sama dengan Anoorapro, mulai dari brand nasional hingga brand internasional. Pihak Anoorapro ingin mendalami kebutuhan klien mereka agar bisa melayani setiap klien dengan lebih efektif dan sesuai. Selain ingin mendalami kebutuhan klien mereka, tim bisnis Anoorapro berencana untuk melakukan marketing kepada klien mereka. Oleh karena itu, diperlukan segmentasi klien agar dapat mencapai klien yang sesuai dengan keperluan.

Analisis segmentasi klien sudah sering dilakukan oleh organisasi ataupun perusahaan. Data transaksi pelanggan dapat digunakan sebagai sumber data yang dapat diproses untuk memproduksi informasi mengenai klasifikasi berdasarkan level transaksi yang

telah dilakukan, baik dari nominal maupun jumlah transaksi dapat membantu perusahaan meningkatkan strategi yang berfokus pada kebutuhan pelanggan [3]. K-Means adalah salah satu algoritma klasterisasi yang populer dan efisien untuk mengelompokkan data ke dalam beberapa klaster berdasarkan kesamaan karakteristik [4]. Beberapa penelitian terdahulu yang menggunakan metode *clustering*, diantaranya adalah penelitian yang dilakukan oleh Auliasari pada tahun 2019. Pada penelitian tersebut menggunakan K-Means *Clustering* untuk menentukan jumlah *cluster* pada konsumen PT Super Sukses Niaga dan menghasilkan 3 kluster. Penelitian lain yang ditulis oleh Nazelda Raina Noperahila dan Feri Sulianta pada tahun 2024. Penelitian ini menggunakan metode tersebut untuk menganalisis pola perilaku konsumen. Penggunaan K-Means *Clustering* menghasilkan $k = 4$ sehingga menggunakan 4 kluster sebagai nilai optimal. Penelitian terdahulu yang ditulis oleh Febryanti E dan Woro Rahayu pada tahun 2023. Penelitian tersebut menggunakan K-Means *clustering* pada transaksi online retail untuk mengoptimisasi strategi pemasaran.

Dengan demikian analisis segmentasi klien ini memang perlu dilakukan, karena memberikan wawasan mendalam tentang karakteristik klien Anoorapro dan juga memungkinkan Anoorapro untuk menargetkan pemasaran kepada kelompok yang sesuai atau relevan. Dengan pemahaman ini, tim marketing Anoorapro dapat membuat strategi pemasaran yang lebih sesuai sehingga dapat meningkatkan penjualan dan meningkatkan layanan yang sesuai dengan kebutuhan kelompok klien.

2. TINJAUAN PUSTAKA

2.1. Clustering

Clustering adalah sebuah metode dengan pendekatan analisis data yang digunakan untuk mengelompokkan objek atau data ke dalam kelompok atau kluster yang memiliki kedekatan atau kemiripan berdasarkan karakteristik tertentu [5].

Tujuan dari segmentasi pelanggan ini adalah untuk lebih mengetahui dan memahami target pasar serta promosi apa yang paling cocok untuk diberikan kepada setiap segmen pelanggan [6].

2.2. K-Means

K-Means adalah bentuk mekanisme data mining yang berguna untuk mengelompokkan data tanpa struktur dan hanya mempunyai tingkatan atau cenderung partisi yang terbagi kedalam kelompok-kelompok. Metode ini mampu memisahkan data ke dalam satu kelompok yang memiliki kesamaan dan memisahkan dengan kelompok lain. Teknik itu berlandaskan jarak antar data yang terdapat dalam cluster yang telah terbentuk, di mana semakin kecil jaraknya menunjukkan tingkat kemiripan antara data satu dengan data lainnya [7]. Tahapan algoritma K-Means mencakup beberapa hal:

- a. Menetapkan jumlah cluster (k).
- b. Menetapkan pusat cluster atau biasa disebut centroid.
- c. Mengkalkulasi jarak pada setiap data terhadap masing masing pusat.

Untuk menghitung jarak rumus yang diterapkan adalah rumus Euclidean (Euclidean Distance) dengan persamaan yaitu:

$$d(a, b) = \sqrt{\sum_{k=1}^n (x_{ak} - x_{bk})^2} \quad (1)$$

Keterangan:

- d (a, b) : Distance (jarak) antara objek a dan b
- n : Jumlah atribut
- x_{ak} : Nilai pusat dari objek a pada dimensi k
- x_{bk} : Nilai pusat dari objek b pada dimensi k
- d. Mengelompokkan data yang didasarkan pada jarak terdekat ke titik pusat.
- e. Menetapkan nilai pusat yang baru dengan mengkalkulasi rata-rata dari cluster.

2.3. SEMMA

SEMMA merupakan singkatan dari Sample, Explore, Modify, Model, dan Assess. Metode ini. SEMMA mengacu pada standar proses data mining yang merupakan tahapan strategi dalam pemecahan masalah pada bisnis atau unit penelitian [8]. Metode ini terdiri dari 5 tahapan yaitu:

- a. Pengumpulan data (sample)
Sample proses mengumpulkan data yang akan digunakan.
- b. Eksplorasi data (explore)
Explore memilah data yang akan digunakan dalam proses analisis atau disebut preprocessing. Mulai dari pembersihan data, memberikan label, dan normalisasi.

- c. Transformasi data (modify)
Modify mengubah atau memodifikasi data yang sudah didapatkan untuk digunakan untuk proses selanjutnya. Seperti merubah data tipe kategorikal menjadi numerik atau sebaliknya sesuai kebutuhan.
- d. Pemodelan data (model)
Model adalah proses memodelkan data agar sesuai dengan hasil yang diinginkan. Contohnya menggunakan elbow method untuk menentukan nilai k pada clustering dan sebagainya.
- e. Evaluasi (assess)
Assess adalah penilaian data untuk evaluasi. Metode ini menjadi kerangka atau alur saat melakukan data mining.

2.4. Python

Python sudah digunakan secara luas dalam analisis data karena Python memiliki libraries yang cukup lengkap. Librarians ini sudah menjadi fondasi dalam data science, dapat membuat para analyst untuk memproses dataset yang besar dengan efisien, membuat model yang kompleks, dan membuat visual yang representatif tentang sebuah data [9]. Contohnya seperti NumPy untuk komputasi numerik. Pandas untuk data preprocessing, mengatasi missing value dan sebagainya. Matplotlib dan Seaborn untuk data visualisasi.

2.5. Colab

Colaboratory atau singkatnya “Colab”, adalah sebuah product dari Google Research. Colab memungkinkan siapa saja untuk menulis dan mengeksekusi code python melalui browser dan sangat cocok untuk machine learning, data analisis, juga edukasi [10]. Melalui platform ini, pengguna dapat menulis dan menjalankan Python tanpa melakukan instalasi apapun. Keuntungan menggunakan Google Colab yang lain adalah berbasis cloud, sehingga bisa diakses dari perangkat apapun selama terhubung dengan internet, termasuk ponsel.

2.6. Tableau

Tableau adalah perangkat lunak Business Intelligence yang relatif sederhana untuk digunakan dalam merancang visualisasi data, melakukan analisis data, dan menyusun laporan dengan berbagai bentuk, seperti Pie Chart, Scatter Plot, Grafik Map, dan Dasbor [11]. Fitur utama Tableau mendukung visualisasi data, seperti:

- a. Dashboard Interaktif: Menyediakan tampilan keseluruhan dari informasi penting dalam format visual.
- b. Kemampuan Blending Data: Menggabungkan data dari berbagai sumber untuk analisis yang lebih komprehensif

Aplikasi ini mampu menganalisis sejumlah besar data dari berbagai sumber, dan dengan menggunakan

sistem dashboard, Tableau dapat memudahkan proses pengambilan keputusan yang cerdas [11].

3. METODE PENELITIAN

Metode yang akan digunakan dalam menganalisis klien Anoorapro adalah metode K-means Clustering. Metode ini dipilih karena data yang akan diteliti berupa data numerik sehingga metode K-Means ini cocok untuk digunakan. Penelitian ini akan mengelompokkan setiap client berdasarkan kemiripan karakteristik mereka. Hasil clustering menggunakan K-means ini nantinya akan digunakan untuk membuat

rekomendasi strategi yang sesuai dengan karakteristik klien.

Untuk keperluan penelitian, penulis menggunakan data pribadi milik agensi kreatif Anoorapro. Dataset yang akan digunakan adalah data transaksi atau data klien yang telah bekerja sama dengan Anoorapro. Agensi kreatif ini berlokasi di Perumahan De Prima Blok AA-12, Tunggulwulung, Lowokwaru, Malang, Jawa Timur.

Tabel 1. Data Klien Anoorapro

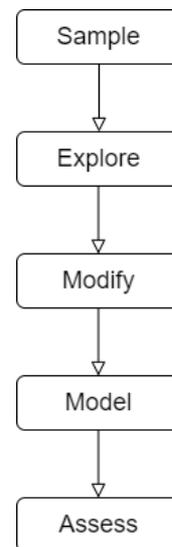
invoice_id	invoice_date	brand_id	produk	harga	qty	total_produk	penjualan_tambahan	harga_tambahan	is_discount	discount_amount	total_tagihan
P/I/622	25/01/2024	July's Scent	Modern n' Clean	237.500	8	1.900.000	none	0	0	0	1.900.000
P/I/620	26/01/2024	Facetology	Catalogue	141.667	24	3.400.000	none	0	0	0	3.400.000
P/I/618	28/01/2024	Halve Skin	Side with Nature	240.000	5	1.200.000	props	500.000	0	0	1.700.000
P/I/621	29/01/2024	Skinberries	Side with Nature	225.000	20	4.500.000	none	0	0	0	4.500.000
P/II/616	05/02/2024	Sealeia	Catalogue	300.000	2	600.000	none	0	0	0	600.000
...
...
P/X/715	12/10/2024	Gummy Nails	Catalogue	194.813	5	974.063			0	0	974.063
P/X/716	14/10/2024	Santai Candles	Side with Nature	243.126	12	2.917.511	Additional Properties	1502293	0	0	4.419.804
C/X/717	17/10/2024	Revered Artistry	Design Graphic	829.010	1	829.010			0	0	829.010
P/X/718	22/10/2024	Eyesight Lenses	Catalogue	150.000	15	2.250.000			0	0	2.250.000

Tools yang akan digunakan adalah python, colab, dan tableau. Library python yang digunakan contohnya seperti pandas, seaborn, matplotlib dan lain-lain. Tableau digunakan sebagai alat bantu untuk memvisualisasikan. Untuk alur proses menggunakan proses model SEMMA (Sample, Explore, Modify, Model, Assess). SEMMA adalah sebuah metode data mining yang menyediakan pendekatan terstruktur untuk memecahkan masalah dalam sebuah bisnis [8].

Langkah-langkah yang dilakukan untuk mengelola data tersebut sebagai berikut:

- a. Mengumpulkan dan memasukkan data
Data-data yang telah didapatkan yaitu data transaksi Anoorapro, diimport ke dalam Google Colab. Bersama dengan import python libraries.
- b. Eksplorasi data
Eksplorasi data adalah proses untuk menganalisis data menggunakan berbagai teknik khususnya secara grafis [12]. Seperti menggunakan heatmap untuk melihat hubungan antar variabel dan memahami setiap variabel di dalam data. Selain itu eksplorasi juga harus dilakukan untuk mengidentifikasi data yang kosong, outlier. Jika

terdapat data kosong, data tersebut dapat dihapus atau diisi dengan nilai lain tergantung konteks data tersebut.



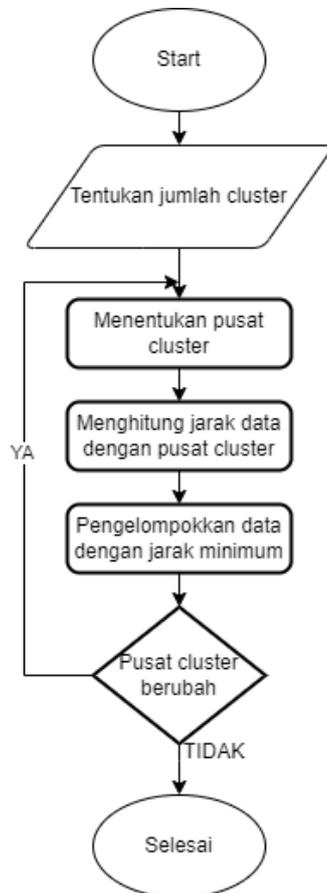
Gambar 1. Model SEMMA

c. Transformasi / Modify data

Mengubah format data agar lebih mudah untuk diproses. Format data diubah sesuai dengan kategorinya sehingga tidak memakan waktu dalam memprosesnya. Membuat fitur atau variabel baru (agregasi) jika diperlukan, menyesuaikan kebutuhan untuk analisis data. Contohnya seperti menghitung berapa lama pembelian terakhir terjadi (Recency). Bisa juga mengubah data kategorikal menjadi numerikal dan juga normalisasi atau standarisasi data.

d. Pemodelan data

Setelah melakukan proses eksplorasi data dan membersihkan data, selanjutnya adalah proses pemodelan. Karena pada penelitian ini adalah *clustering* menggunakan K-Means, sehingga dalam menentukan jumlah *cluster* akan digunakan elbow method. Berikut adalah proses *clustering* algoritma K-Means:



Gambar 2. Flowchart K-Means

e. Evaluasi / Assess data

Pada tahap ini *clustering* atau segmentasi sudah dibuat. Selanjutnya adalah mengevaluasi *clustering* tersebut. Apakah data *clustering* sesuai dengan data aktual. Karena tujuannya adalah memberikan rekomendasi yang dapat diimplementasikan untuk meningkatkan pemasaran.

4. HASIL DAN PEMBAHASAN

4.1. Import Data

Data yang didapatkan dari Anoorapro diimport ke dalam Google Colab. Data ini dapat memberikan gambaran tentang karakteristik klien nantinya.

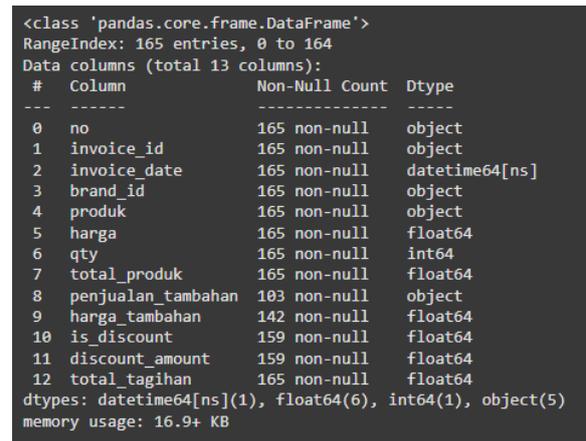


Gambar 3. Import data

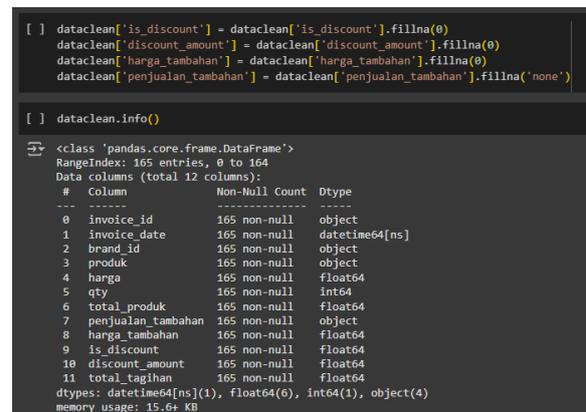
Setelah data berhasil di import ke dalam Google Colab. Selanjutnya adalah eksplorasi data

4.2. Pengecekan Missing Value

Sebelum data-data tersebut *di clustering* atau dikelompokkan, data tersebut harus dieksplor terlebih dahulu. Saat eksplorasi data, hal terpenting adalah memahami variabel data tersebut. Setelah itu adalah melakukan data cleaning, yaitu pengecekan missing value dan duplikat.



Gambar 4. Pengecekan missing value



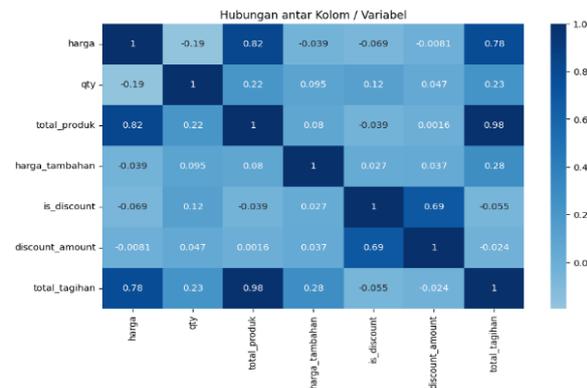
Gambar 5. Mengatasi Missing value

Pada gambar 4 dapat dilihat bahwa terdapat value atau nilai yang kosong pada beberapa kolom. Hal itu dapat diatasi dengan cara menghapus atau mengisi nilai kosong tersebut. Dengan melihat kolom-kolom yang memiliki nilai kosong seperti kolom

penjualan_tambahan, nilai kosong berarti klien tidak membeli produk tambahan. Sehingga nilai kosong tersebut bisa diganti dengan nilai 0.

4.3. Hubungan antar Variabel Menggunakan Heatmap

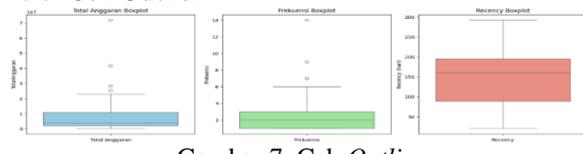
Setelah pengecekan, selanjutnya adalah melihat hubungan antar variabel. Dengan melihat hubungan antar variabel atau kolom, bisa diketahui apakah kolom-kolom tersebut saling berkaitan atau tidak.



Gambar 6. Hubungan antar variabel

Jika nilai semakin mendekati 1, maka terdapat korelasi atau saling berhubungan. Berdasarkan gambar 6 dapat diketahui bahwa terdapat korelasi positif antara quantity dengan adanya discount meskipun sangat kecil.

4.4. Cek Outlier



Gambar 7. Cek Outlier

Berdasarkan gambar di atas, terdapat beberapa client yang termasuk outlier. Outlier adalah sebuah data yang memiliki nilai jauh berbeda daripada nilai data yang lainnya. Terdapat 4 brand yang termasuk dalam outlier Total Anggaran dan 4 brand pada outlier Frekuensi. Di sini berarti klien outlier tersebut mengeluarkan total anggaran yang banyak dan memiliki frekuensi beli yang lebih daripada klien yang lainnya. Karena tidak mungkin untuk menghapus data klien yang berpotensi tersebut, maka akan dipisahkan dari data lainnya sementara.

4.5. Agregasi Data

Agregasi data dibuat untuk menyederhanakan dan lebih menonjolkan pola yang relevan sehingga akan membantu proses analisis. Variabel-variabel tersebut akan dicocokkan berdasarkan nama brandnya.

	brand_id	TotalAnggaran	Frekuensi	PembelianTerakhir	Recency (hari)
0	Acha Skincare	2750000.00	1	2024-07-10	125
1	Amura Beauty Care	11350000.00	3	2024-07-23	112
2	Anessa	7531750.00	1	2024-05-06	190
3	BG Skin	41920000.00	14	2024-09-30	43
4	Basic Skincare	1550000.00	1	2024-06-08	157

Gambar 8. Agregasi data

Variabel Total Anggaran didapatkan dari jumlah total_tagihan. Frekuensi dihitung dari berapa banyak brand itu melakukan pembelian. Sedangkan Recency dihitung dari tanggal terakhir klien melakukan pembelian dikurangi dengan tanggal waktu analisis ini dilakukan yaitu 12 November 2024.

4.6. Clustering

Sebelum menerapkan *elbow method*, data yang merupakan outlier dipisahkan sementara.

```
data_noutlier = data_agg[~data_agg.index.isin(dataA_outlier.index) & ~data_agg.index.isin(dataF_outlier.index)]
data_noutlier.describe()
```

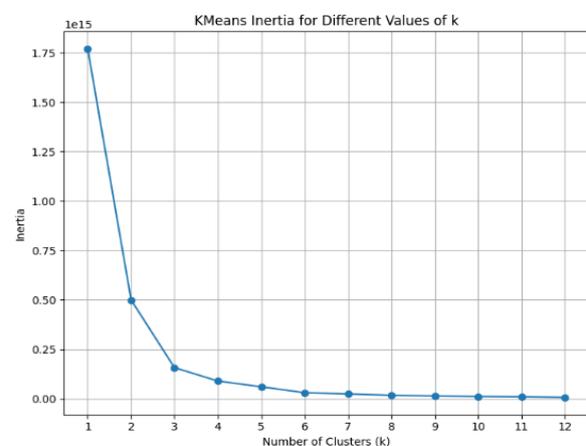
	TotalAnggaran	Frekuensi	PembelianTerakhir	Recency (hari)
count	53.00	53.00	53	53.00
mean	6264017.88	2.32	2024-06-07 18:33:57.735848960	157.23
min	2500000.00	1.00	2024-01-25 00:00:00.000000	21.00
25%	19000000.00	1.00	2024-04-29 00:00:00.000000	110.00
50%	35000000.00	2.00	2024-05-30 00:00:00.000000	166.00
75%	95000000.00	3.00	2024-07-25 00:00:00.000000	197.00
max	230000000.00	6.00	2024-10-22 00:00:00.000000	292.00
std	5832168.71	1.31	NaN	75.97

Gambar 9. Pemisahan data outlier

Setelah dipisah, barulah diterapkan *elbow method* untuk mendapatkan nilai k. Nilai k ini menunjukkan jumlah cluster yang akan terbentuk dan juga akan mempengaruhi baik atau buruknya cluster.

```
max_k = 12
inertia = []
k_values = range(1, max_k + 1)
for k in k_values:
    kmeans = KMeans(n_clusters=k, random_state=42, max_iter=1000)
    cluster_labels = kmeans.fit_predict(data_noutlier[['TotalAnggaran', 'Frekuensi', 'Recency (hari)']])
    inertia.append(kmeans.inertia_)
plt.figure(figsize=(14, 6))
plt.subplot(1, 2, 1)
plt.plot(k_values, inertia, marker='o')
plt.title('KMeans Inertia for Different Values of k')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Inertia')
plt.xticks(k_values)
plt.grid(True)
plt.tight_layout()
plt.show()
```

Gambar 10. Elbow method



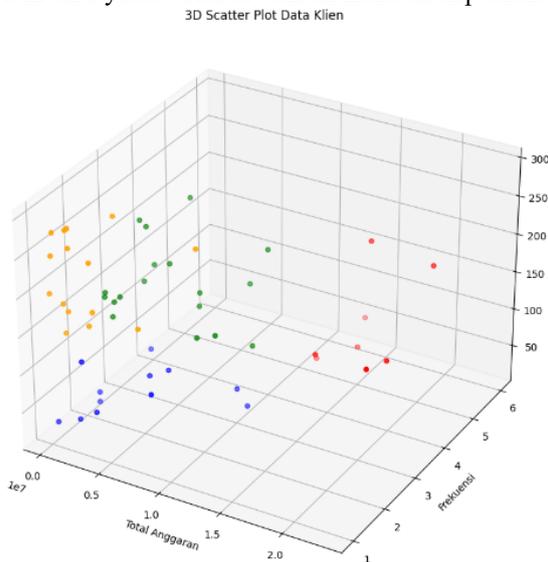
Gambar 11. Visualisasi Elbow method

Hasil dari elbow method menunjukkan bahwa k terbaik ada di k = 4. Karena titik di mana inersia mulai mendatar itulah yang disebut elbow. Sehingga hasil clustering nanti akan menunjukkan 4 kelompok tanpa client outlier. Selanjutnya mengimplementasikan K-Means dan membuat visualisasi clustering pada variabel Total Anggaran, Frekuensi, dan Recency.

```
kmeans = KMeans(n_clusters=4, random_state=42, max_iter=1000)
cluster_labels = kmeans.fit_predict(fix_scale)
cluster_labels
array([0, 2, 1, 1, 1, 2, 2, 2, 0, 1, 2, 1, 1, 2, 2, 3, 3, 2, 1, 3, 0, 1, 1,
       3, 2, 2, 0, 1, 2, 3, 1, 2, 0, 2, 0, 0, 2, 0, 0, 0, 1, 3, 1, 3, 3,
       1, 1, 2, 0, 1, 0, 2, 1, 2], dtype=int32)
data_nooutlier["Cluster"] = cluster_labels
data_nooutlier
```

Gambar 12. Implementasi K-Means

Setelah implementasi K-Means, akan dibuat kolom baru yaitu "Cluster" untuk melabeli tiap client.



Gambar 13. 3D Scatter plot

```
[ ] gabungan = dataTA_outlier.index.intersection(dataF_outlier.index)
anggaranoutlier = dataTA_outlier.drop(gabungan)
frekuensioutlier = dataF_outlier.drop(gabungan)
anggaranfrekuensi_outlier = dataTA_outlier.loc[gabungan]
anggaranoutlier["Cluster"] = -1
frekuensioutlier["Cluster"] = -2
anggaranfrekuensi_outlier["Cluster"] = -3
cluster_outlier = pd.concat([anggaranoutlier, frekuensioutlier, anggaranfrekuensi_outlier])
cluster_outlier
```

brand_id	TotalAnggaran	Frekuensi	PembelianTerakhir	Recency (hari)	Cluster
36	Pucelle 7200000.00	2	2024-06-14	151	-1
38	Quixx Creatively 28488100.00	3	2024-07-09	126	-1
5	Biotalk 14050000.00	7	2024-09-25	48	-2
38	Pvirebeauty 4067250.00	7	2024-04-05	221	-2
3	BG Skin 41920000.00	14	2024-09-30	43	-3
22	Gummy Nails 25541263.39	9	2024-10-12	31	-3

Gambar 14. Client outlier

Melihat visualisasi di atas, bisa dilihat setiap client tersebar berdasarkan variabel Total Anggaran, Frekuensi, dan Recency. Warna biru adalah cluster 0, warna oranye / kuning adalah cluster 1, warna hijau

adalah cluster 2, dan warna merah adalah cluster 3. Ini adalah empat cluster tanpa data outlier. Sedangkan untuk data klien outlier, akan dibuatkan cluster sendiri karena klien tersebut juga memiliki karakteristik sendiri.

Meskipun merupakan outlier, client ini memiliki nilai yang tinggi di mana hal tersebut membawa keuntungan bagi agensi. Untuk klien atau brand yang terdapat pada outlier anggaran akan masuk ke cluster -1, untuk klien yang masuk ke cluster -2 adalah klien yang terdapat pada outlier frekuensi, sedangkan cluster -3 adalah klien yang terdapat pada outlier frekuensi dan anggaran. Setelah semua client memiliki label, kedua data tersebut digabungkan kembali.

```
datafull = pd.concat([data_nooutlier, cluster_outlier])
datafull
```

brand_id	TotalAnggaran	Frekuensi	PembelianTerakhir	Recency (hari)	Cluster
0	Acha Skincare 2750000.00	1	2024-07-10	125	0
1	Amura Beauty Care 11350000.00	3	2024-07-23	112	2
2	Anessa 7531750.00	1	2024-05-06	190	1
4	Basic Skincare 1550000.00	1	2024-06-08	157	1
6	Blishfull 2700000.00	2	2024-02-09	277	1
7	Castury 4500000.00	3	2024-05-07	189	2
8	Ciella and Co 2560000.00	3	2024-03-28	229	2
9	Curble 1900000.00	2	2024-05-21	175	2
10	Dearly Herb 8500000.00	1	2024-07-25	110	0
11	Deo Pow 3500000.00	1	2024-05-21	175	1
12	Desir Fragrance 2300000.00	3	2024-06-09	156	2

Gambar 15. Penggabungan data

Dari hasil clustering tersebut, dapat diketahui karakteristik dari tiap cluster sebagai berikut:

Cluster 0: Pelanggan dalam cluster ini memiliki anggaran menengah, melakukan pembelian sebanyak 1-3 kali, dan baru-baru ini melakukan transaksi. Hal ini menunjukkan bahwa mereka memiliki potensi untuk menjadi pelanggan loyal apabila dikelola dengan strategi pemasaran yang tepat, seperti penawaran eksklusif untuk pembelian berikutnya.

- Cluster 1: Pelanggan ini memiliki anggaran terkecil dibandingkan cluster lainnya, dengan frekuensi pembelian 1-2 kali, dan sudah lama tidak bertransaksi. Kelompok ini cenderung menjadi pelanggan yang beresiko menghilang, sehingga memerlukan strategi re-engagement, seperti pemberian diskon atau pengingat melalui email.
- Cluster 2: Cluster ini berisi pelanggan dengan anggaran menengah, frekuensi pembelian 2-4 kali, tetapi sudah lama tidak melakukan transaksi. Mereka pernah aktif sebelumnya, sehingga terdapat peluang untuk mengaktifkan kembali mereka melalui strategi promosi yang menarik atau komunikasi personal.
- Cluster 3: Pelanggan ini memiliki anggaran tinggi dan frekuensi pembelian 2-6 kali, tetapi sudah cukup lama tidak bertransaksi (meskipun tidak selama Cluster 1 dan 2). Potensi pendapatan dari cluster ini cukup besar, sehingga penting untuk menyusun yang terfokus untuk membuat klien di cluster ini menjadi loyal, seperti pemberian penawaran eksklusif.

- d. Cluster -1: Cluster ini terdiri dari pelanggan dengan anggaran tertinggi tetapi frekuensi pembelian yang rendah. Mereka juga telah lama tidak bertransaksi. Mengingat nilai transaksinya yang besar, strategi retensi yang lebih personal, seperti layanan khusus atau konsultasi, dapat digunakan untuk menarik mereka kembali.
- e. Cluster -2: Pelanggan ini memiliki anggaran besar dan sering berbelanja hingga 7 kali, tetapi tidak melakukan transaksi dalam kurun waktu yang agak lama. Cluster ini menunjukkan loyalitas di masa lalu dan berpeluang besar untuk diaktifkan kembali dengan memberikan insentif seperti diskon loyalitas atau program penghargaan.
- f. Cluster -3: Cluster ini berisi pelanggan dengan anggaran tinggi, frekuensi pembelian yang tinggi, dan baru-baru ini melakukan transaksi. Mereka adalah pelanggan terbaik (VIP) yang perlu diprioritaskan dengan layanan premium, program loyalitas eksklusif, atau akses khusus ke produk/layanan baru.

4.7. Label Cluster

Hasil *clustering* menghasilkan total 7 *cluster* Ketujuh *cluster* tersebut akan diberi label atau nama sesuai dengan karakteristik mereka agar lebih mudah untuk dikenali. Juga agar memudahkan dalam pemberian rekomendasi strategi untuk tim pemasaran.

Cluster 0 : Client Baru memiliki total pembelian yang sedikit dan baru-baru ini melakukan transaksi, mereka memiliki potensi jika dikelola dengan tepat. Rekomendasi yang dapat dilakukan adalah memberikan diskon untuk pembelian selanjutnya (repeat).

Cluster 1: One Time Buyer memiliki total pembelian paling sedikit dan sudah lama tidak melakukan transaksi. Rekomendasi yang bisa dilakukan adalah re-engagement dengan memberikan penawaran khusus.

Cluster 2: Client Pasif ini sempat menjadi client aktif dengan anggaran yang menengah, tetapi lama tidak melakukan transaksi lagi. Rekomendasi yang bisa dilakukan adalah komunikasi personal seperti mengirim pesan dan beri promosi yang menarik agar mereka dapat aktif kembali.

Cluster 3: Client Hidden Gem memiliki anggaran yang besar sehingga penting untuk menjadikan client pada *cluster* ini loyal. Rekomendasi yang bisa dilakukan adalah dengan memberikan penawaran yang eksklusif dan fokus untuk menciptakan hubungan jangka panjang dengan mereka.

Cluster -1: Client Spender Besar memiliki nilai transaksi yang sangat tinggi sehingga dibutuhkan strategi retensi. Rekomendasi yang bisa dilakukan adalah menyediakan layanan personal seperti konsultasi eksklusif.

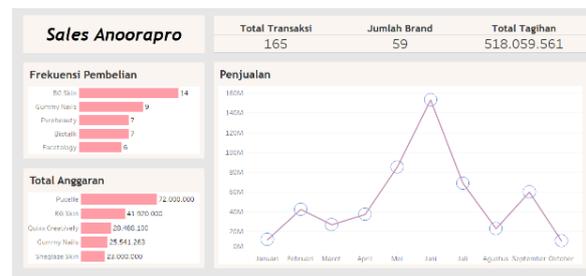
Cluster -2: Client Loyal memiliki anggaran yang cukup besar dan total transaksi yang cukup banyak. Rekomendasi yang bisa dilakukan adalah memberikan

penawaran atau diskon loyalitas, sehingga kerja sama tetap dapat berjalan dengan baik.

Cluster -3: Client Terbaik memiliki anggaran yang besar dan frekuensi pembelian yang tinggi, sehingga bisa dibilang mereka adalah client prioritas. Rekomendasi yang bisa dilakukan adalah prioritaskan dengan layanan premium dan buat program loyalitas seperti produk ekstra setelah transaksi.

4.8. Visualisasi Overview

Setelah melakukan analisis data klien Anoorapro, hasilnya bisa divisualisasikan agar lebih mudah untuk dipahami oleh tim pemasaran Anoorapro untuk mendapatkan insight.



Gambar 16. Visualisasi overview

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis *clustering* mengenai klien Anoorapro, dapat disimpulkan beberapa poin penting yaitu hubungan antara quantity dengan adanya diskon menunjukkan adanya korelasi / hubungan yang positif, tetapi sangat lemah. Artinya dengan adanya diskon, maka quantity juga meningkat meskipun sangat sedikit. Dari penerapan *Elbow method* menghasilkan nilai k terbaik ada di k = 4. Dalam penelitian ini, *clustering* menggunakan 3 variabel yaitu total anggaran, frekuensi, dan recency. Hasil dari *clustering* menunjukkan bahwa client Anoorapro terbagi menjadi 7 kelompok atau cluster. Cluster 0 (Client baru) terdiri dari client yang memiliki anggaran menengah, melakukan pembelian sebanyak 1-3 kali, dan baru-baru ini melakukan transaksi. Cluster 1 (One Time buyer) terdiri dari client yang memiliki anggaran terkecil dibandingkan cluster lainnya, dengan frekuensi pembelian 1-2 kali, dan sudah lama tidak bertransaksi. Cluster 2 (Client Pasif) terdiri dari client dengan anggaran menengah, frekuensi pembelian 2-4 kali, tetapi sudah lama tidak melakukan transaksi. Cluster 3 (Hidden Gem) terdiri dari client yang memiliki anggaran tinggi dan frekuensi pembelian 2-6 kali, tetapi sudah cukup lama tidak bertransaksi (meskipun tidak selama Cluster 1 dan 2). Cluster -1 (Spender Besar) ini terdiri dari client anggaran tertinggi tetapi frekuensi pembelian yang rendah. Mereka juga telah lama tidak bertransaksi. Cluster -2 (Client Loyal) terdiri dari client yang memiliki anggaran besar dan sering berbelanja hingga 7 kali, tetapi tidak melakukan transaksi dalam kurun waktu yang agak lama. Cluster -3 (Client Terbaik) ini berisi pelanggan dengan anggaran tinggi, frekuensi

pembelian yang tinggi, dan baru-baru ini melakukan transaksi. Dari penjelasan di atas, dapat diketahui bahwa setiap *cluster* memiliki karakteristik yang berbeda sehingga perlu strategi pendekatan yang sesuai. Saran yang dapat diberikan untuk pengembangan penelitian selanjutnya adalah melakukan perbandingan dengan algoritma lain seperti Hierarchical atau DBSCAN untuk mendapatkan nilai k terbaik.

DAFTAR PUSTAKA

- [1] Kussudyarsana and L. Rejeki, "PENGARUH MEDIA SOSIAL ONLINE DAN MEDIA PROMOSI OFFLINE TERHADAP PEMILIHAN MEREK PRODUK SKINCARE DAN KLINIK KECANTIKAN," 2020.
- [2] R. Y. Raffi, N. Ketut, P. Sarjani, A. Ngurah, and A. Putraka, "PERANCANGAN STRATEGI PEMASARAN VISUAL MELALUI MEDIA SOSIAL UNTUK MEREK DAGANG 'ZOUSZ GROOMING' PADA AGENSI KREATIF LAPOMPS DI BALI," 2024.
- [3] C. Devi, O. Soleman, N. Pramaita, and M. Sudarma, "Classification Of Loyalty Customer Using K-Means Clustering, Studi Case : PT. Sucofindo (Persero) Denpasar Branch," *International Journal of Engineering and Emerging Technology*, vol. 5, no. 2, 2020.
- [4] E. Febrianty, L. Awalina, and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means *Clustering* pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means *Clustering* on Online Retail Transactions," *Jurnal Teknologi dan Informasi (JATI)*, vol. 13, 2023, doi: 10.34010/jati.v13i2.
- [5] Dhea Agustina Akmal, Relita Buaton, and Anton Sihombing, "Klasifikasi Tingkat Minat Belanja Online Melalui Media Sosial pada Masyarakat di Kota Binjai Menggunakan Algoritma K-Means," *Bridge : Jurnal publikasi Sistem Informasi dan Telekomunikasi*, vol. 2, no. 3, pp. 214–228, Aug. 2024, doi: 10.62951/bridge.v2i3.169.
- [6] E. Febrianty, L. Awalina, and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions," *Jurnal Teknologi dan Informasi (JATI)*, vol. 13, 2023, doi: 10.34010/jati.v13i2.
- [7] N. Sukarno Wijaya, M. Jajuli, and B. A. Dermawan, "PENERAPAN ALGORITMA K-MEANS *CLUSTERING* DALAM MENENTUKAN DAERAH PRIORITAS PENANGANAN KEMISKINAN DI WILAYAH JAWA TIMUR," 2024.
- [8] N. R. Noperahila and F. Sulianta, "Analisis Pola Perilaku Konsumen dan Kebiasaan Belanja menggunakan Metode K-Means *Clustering*," 2024. [Online]. Available: <https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and->
- [9] M. A. Kabir, F. Ahmed, M. M. Islam, and Md. R. Ahmed, "Python For Data Analytics: A Systematic Literature Review Of Tools, Techniques, And Applications," *ACADEMIC JOURNAL ON SCIENCE, TECHNOLOGY, ENGINEERING & MATHEMATICS EDUCATION*, vol. 4, no. 04, pp. 134–154, Nov. 2024, doi: 10.69593/ajsteme.v4i04.146.
- [10] P. Naik, "Conceptualizing Python in Google COLAB," 2022. [Online]. Available: <https://www.researchgate.net/publication/357929808>
- [11] D. Chika Fransisca and A. Raharja, "SOROT (Jurnal Pengabdian Kepada Masyarakat) Pelatihan Tableau Untuk Visualisasi Data Secara Cepat dan Mudah," vol. 3, no. 1, pp. 17–21, 2024, doi: 10.32699.
- [12] A. Sujjada, G. Purnama Insany, and S. Noer, "Jurnal Teknologi dan Manajemen Informatika Analisis *Clustering* Data Penyandang Disabilitas Menggunakan Metode Agglomerative Hierarchical *Clustering* dan K-means," vol. 10, pp. 1–12, 2024, [Online]. Available: <http://http://jurnal.unmer.ac.id/index.php/jtmi>