

SEGMENTASI PELANGGAN ONLINE DENGAN METODE K-MEANS CLUSTERING

Herian Daulat Silalahi¹, Doli Hasibuan², Marzuki Sinambela³
^{1,2,3}Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Histori Artikel:

Received, Januari, 2023
Revised, Januari, 2023
Accepted, Februari, 2023

Keywords:

Segmentasi,
Pelanggan Online,
K-means Clustering,

ABSTRAK

Hasil penelitian menunjukkan bahwa pasar konsumen Online terdiri dari tiga segmen yang umumnya digunakan yaitu: Pola segmentasi Dasar Komunikasi, pada segmen konsumen inilah yang menggunakan internet terutama untuk berkomunikasi via e-mail, maka segmen Commerce atau belanja di segmen ini. konsumen yang menggunakan internet dan melihat-lihat toko online dan segmen ketiga hubungan sosial dan kesenangan, di segmen konsumen inilah yang memanfaatkan internet menggunakan fitur interaktif untuk berinteraksi dengan cara chatting, video streaming, dan mendownload Keterbatasan meliputi data yang dikumpulkan melalui survei yang informasinya berasal dari internet, sehingga dibutuhkan pandangan pihak lain yang bukan pengguna tapi pengamat bisnis online yang memiliki pendapat lain mengenai segmentasi ini. Nantinya segmentasi ini sebenarnya masih bisa dipecah menjadi beberapa sub segmen, selaku usaha penelitian lanjutan.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Koresponden:

Herian Daulat Silalahi,
Fakultas Ilmu Komputer,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: heriandaulatsilalahirian@gmail.com

1. PENDAHULUAN

Pada masa kegiatan Masyarakat Ekonomi ASEAN (MEA), Indonesia dipilih oleh beberapa perusahaan dari negara lain untuk menjual produk karena populasi diantaranya perusahaan cat avitex luar negeri. Peningkatan jumlah perusahaan yang ada saat ini, membuat persaingan global yang terus meningkat termasuk persaingan pasar dalam industri cat avitex menjadi sangat sengit. Peningkatan jumlah perusahaan juga menjadikan pemahaman mengenai pelanggan dibutuhkan dalam membangun model keputusan bisnis dan analisis data transaksi pelanggan untuk membantu perencanaan komersial. Agar tidak kalah bersaing, perusahaan dalam negeri mengembangkan pasar dan menyesuaikan perkembangan yang terjadi saat ini dengan menerapkan strategi bisnis yang didapatkan dari menganalisis kebutuhan pelanggannya.

Salah satu perusahaan cat avitex dalam negeri telah mengembangkan sistem baru yang diperlukan pelanggan dalam proses transaksi. Oleh karena itu, perusahaan juga harus memperhatikan kebutuhan pelanggannya selain berfokus dengan produk yang dijual. Pemahaman mengenai pelanggan terdapat dalam pengelolaan hubungan pelanggan yang menggambarkan strategi komprehensif dalam proses memperoleh, mempertahankan dan bermitra dengan pelanggan tertentu untuk menciptakan nilai bagi perusahaan dan pelanggan. Tugas terpenting perusahaan merupakan mencari pelanggan, namun mempertahankan pelanggan lebih penting karena kehilangan pelanggan berarti kehilangan seluruh aliran bisnis dengan 2 pelanggan. Dalam lingkungan yang

dinamis perusahaan yang kompetitif mencoba untuk memperoleh wawasan mengenai kebutuhan, sikap, dan perilaku pelanggan yang akan dipertahankan. Namun pada kenyataannya, perusahaan pasti memiliki pelanggan yang perilakunya berbeda-beda dalam melakukan transaksi. Idealnya setiap pelanggan harus diperlakukan sebagai individu dan perusahaan harus memperhatikan pelanggan satu persatu, akan tetapi pendekatan ini jelas tidak mungkin. Oleh karena itu, alternatif yang efisien adalah melakukan segmentasi pelanggan. Segmentasi pelanggan merupakan sebuah model yang dibangun dalam mengelompokkan pelanggan sesuai dengan standar tertentu untuk digunakan menjadi variabel pengelompokan.

2. METODE PENELITIAN

Tahapan Ini menjelaskan kronologis penelitian, menentukan jumlah cluster untuk diterapkan pada Algoritma K-Means, perlu untuk menentukan jumlah cluster yang akan digunakan. Sebaiknya menggunakan jumlah cluster yang optimal. Ada dua metode sangat populer dalam menentukan cluster optimal, yakni:

a. Metode Siku (Elbow method)

Tujuan utama di balik metode partisi cluster seperti k-means adalah untuk mendefinisikan cluster sehingga metode Elbow menggunakan variasi intra-cluster tetap minimum.

$$\text{mimize}(\text{Sum } W(C_k)).k = 1 \dots k$$

dimana (C_k) mewakili cluster K dan $W(C_k)$

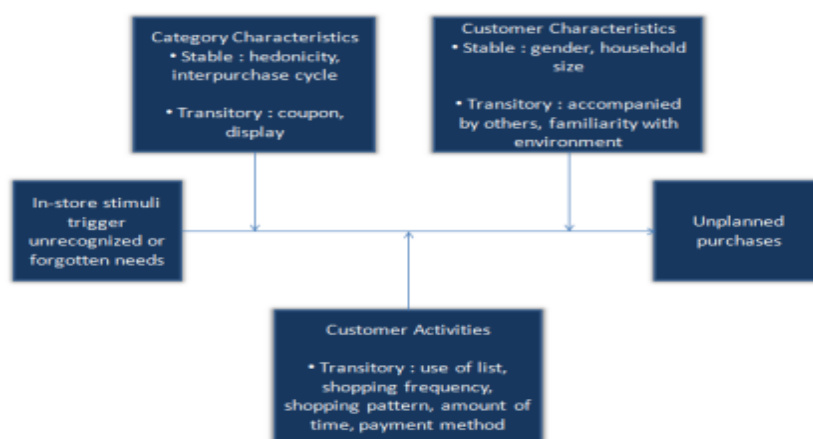
b. Metode Siluet (Silhouette Method)

Dengan bantuan metode siluet rata-rata, dapat diukur kualitas operasi pengelompokan yang dilakukan. Dengan metode ini, dapat ditentukan seberapa baik cluster pada setiap variabel dalam dataset. Jika diperoleh luas rata-rata siluet yang tinggi, itu berarti kita memiliki pengelompokan yang baik. Metode siluet rata-rata dilakukan dengan menghitung rata-rata pengamatan siluet untuk setiap nilai K_k yang berbeda. Dengan jumlah optimal K_k cluster, dapat dimaksimalkan siluet rata-rata di atas nilai signifikan untuk K_k cluster.

Ada empat poin keberhasilan untuk melakukan segmentasi pelanggan antara lain:

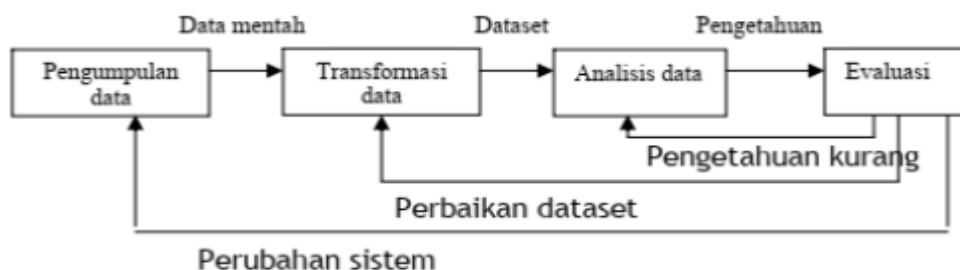
1. Memahami biaya bisnis: mengetahui bukan hanya tentang biaya bisnis secara keseluruhan, tetapi juga biaya yang setiap detail transaksi pelanggan. Hal ini diharapkan membantu meningkatkan customers profit potential dan customer lifetime values (CLV) [2]
2. Identifikasi pelanggan terbaik: mengetahui pelanggan yang potensial dan memberikan pelayanan serta rekomendasi terbaik yang diharapkan dapat memberi nilai kepuasan pada pelanggan
3. Mengumpulkan data dari sumber eksternal: mengumpulkan setiap informasi terkait transaksi pelanggan untuk menciptakan pengalaman berbelanja yang relevan dan tepat sasaran
4. Membuat profil pelanggan: ini ditujukan untuk memahami kebutuhan, tantangan dan pola pikir terhadap pelanggan [1]

Karakteristik Pelanggan adalah pola sifat menguntungkan dari pelanggan yang berupa informasi yang dapat digunakan untuk membantu pengembangan produk atau layanan baru dengan mengacu pola yang didapat dari pelanggan [2]. Karakteristik dapat dicontohkan seperti status, kualifikasi, spesialisasi, sifat pekerjaan, kegiatan dan prestasi profesional dari pelanggan [3]. Karakteristik pelanggan dapat digunakan untuk menentukan target pasar lebih spesifik dan dapat membuat strategi pemasaran yang cocok untuk pelanggan mereka [4] dapat dilihat pola informasi pelanggan yang dapat dijadikan rujukan untuk membangun karakteristik yang dibutuhkan sebagai peningkatan pelayanan terhadap pelanggan.



Gambar 1 Data Mini

Data Mining Data mining adalah bagian dari knowledge discovery in databases (KDD) yang terdiri dari beberapa proses seperti data cleaning, data selection, data integration, data transformation, data mining, pattern evaluation) dan knowledge presentation [5]. Data mining dapat digunakan sebagai salah satu cara praktis untuk automasi data dari berbagai source seperti database, data warehouse dan penyimpanan data lainnya [6]. Adapun langkah-langkah proses pada data mining.



Gambar 2 Gambar Mini

Pada data mining, pola atau informasi yang dibutuhkan untuk proses pengembangan pelayanan dapat dipilih dengan menggunakan teknik atau metode tertentu (algoritma). Analisa data informasi yang dikumpulkan diharapkan dapat mencapai tujuan bisnis untuk meningkatkan kepuasan pelanggan sesuai dengan target dan tepat sasaran.

3. HASIL DAN PEMBAHASAN

Ditahap ini dilakukan analisa kebutuhan data. Data-data yang akan diinputkan ke sistem saling berelasi antara satu sama lainnya. Relasi data yang akan menjadi satu kesatuan basis data utuh. Data-data yang dibutuhkan sistem adalah sebagai berikut:

1. Data konsumen Data-data konsumen yang melakukan transaksi.
2. Data transaksi Data-data transaksi yang telah dilakukan oleh konsumen. Kriteria yang digunakan untuk proses pengelompokan.

3.1. Data Input

Pada sub tahapan analisa proses data inputan ini akan ditentukan data yang menjadi data input yang akan dimasukkan ke dalam sistem yang akan dirancang nantinya. Data input ini berupa kurang lebih 200 data pelanggan online yang diambil dari *website kaggle.net*.

	A	B	C	D	E
1	Customer	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
2	1	Male	19	15	39
3	2	Male	21	15	81
4	3	Female	20	16	6
5	4	Female	23	16	77
6	5	Female	31	17	40
7	6	Female	22	17	76
8	7	Female	35	18	6
9	8	Female	23	18	94
10	9	Male	64	19	3
11	10	Female	30	19	72
12	11	Male	67	19	14
13	12	Female	35	19	99
14	13	Female	58	20	15
15	14	Female	24	20	77
16	15	Male	37	20	13
17	16	Male	22	20	79
18	17	Female	35	21	35
19	18	Male	20	21	66
20	19	Male	52	23	29
21	20	Female	35	23	98
22	21	Male	35	24	35
23	22	Male	25	24	73

Gambar 3 Sampel data Pelanggan online Mall

3.2. Menentukan Pengelompokan Cluster

Langkah selanjutnya adalah langkah untuk menentukan data yang masuk kedalam suatu cluster. Suatu data yang akan menjadi anggota dari suatu cluster adalah data yang memiliki jarak terkecil dari pusat clusternya.

Berikut ini merupakan langkah-langkah pengelompokan pelanggan menggunakan K-Means:

1. Memasukkan data pelanggan yang akan di cluster
Data pelanggan akan di cluster. Dalam penelitian ini data yang digunakan adalah data pelanggan sebanyak 90 data.
2. Menentukan jumlah data cluster
Langkah kedua merupakan langkah yang digunakan untuk menentukan jumlah cluster yang akan dibentuk. Dalam penelitian ini jumlah cluster yang akan dibentuk adalah menjadi 3 cluster.
3. Menentukan data yang akan diproses
Langkah selanjutnya adalah menentukan data yang akan diproses atau data yang akan di-cluster. Data yang akan diproses atau di-cluster adalah Data pelanggan yang digunakan adalah data sampel, yaitu sekitar 10% dari seluruh pelanggan, dengan jumlah pelanggan 90 yang merujuk dari tabel pelanggan.

3.3 Pembagian Data

Pada sub tahapan analisa model pembagian data ini dilakukan untuk proses pembagian data dari seluruh data yang telah didapatkan. Pembagian data ini dilakukan untuk mempermudah dalam proses pelatihan dan pengujian dengan menggunakan metode Logistic Regression nantinya. Dari data yang didapatkan, kurang lebih 200 data kasus Planggan Mall akan dibagi menjadi pengujian data latih sebanyak 56% dan data uji sebanyak 44%.

Tabel 1 Kolom Id Pelanggan

	Cutomer ID	Gender	Age	Annual income	Spending score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Famale	20	16	6
3	4	Famale	23	16	77
4	5	Famale	31	17	40

3.4 Pengujian

Untuk mengetahui keberhasilan dari implementasi peramalan yang telah dilakukan, selanjutnya melakukan tahap pengujian terhadap sistem. Pengujian pada penelitian ini dilakukan untuk mengukur akurasi hasil prediksi menggunakan indeks evaluasi yaitu dengan menggunakan metode K-Means, lalu hasil pengelompokan yang telah dilakukan lalu melalui proses segmentasi yang bertujuan memberi grade-grade dari kelompok konsumen untuk membagi konsumen potensial dan kurang potensial dengan menggunakan sistem. Setiap pengujian dilakukan dengan komposisi data yang didapatkan, dari k-means clustering kurang lebih 100 data kasus Planggan Mall akan dibagi menjadi pengujian data latih sebanyak 56% dan data uji sebanyak 44% untuk setiap masing-masing variabel peramalan seperti yang sudah dijelaskan pada bab sebelumnya. Berikut hasil indeks evaluasi untuk setiap variabel prediksi.

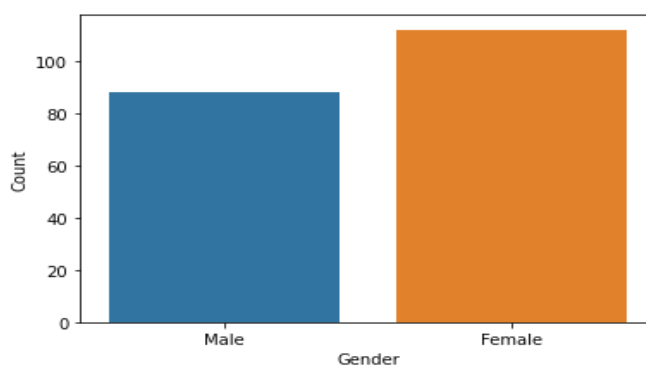
	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
1					
2	1	Male	19	15	39
3	2	Male	21	15	81
4	3	Female	20	16	6
5	4	Female	23	16	77
6	5	Female	31	17	40
7	6	Female	22	17	76
8	7	Female	35	18	6
9	8	Female	23	18	94
10	9	Male	64	19	3
11	10	Female	30	19	72
12	11	Male	67	19	14
13	12	Female	35	19	99
14	13	Female	58	20	15
15	14	Female	24	20	77
16	15	Male	37	20	13
17	16	Male	22	20	79
18	17	Female	35	21	35
19	18	Male	20	21	66
20	19	Male	52	23	29
21	20	Female	35	23	98
22	21	Male	35	24	35
23	22	Male	25	24	73

Gambar 4 Nilai K-means untuk variabel Customer ID,Gender,Age,Annual

179	178	Male	27	88	69
180	179	Male	59	93	14
181	180	Male	35	93	90
182	181	Female	37	97	32
183	182	Female	32	97	86
184	183	Male	46	98	15
185	184	Female	29	98	88
186	185	Female	41	99	39
187	186	Male	30	99	97
188	187	Female	54	101	24
189	188	Male	28	101	68
190	189	Female	41	103	17
191	190	Female	36	103	85
192	191	Female	34	103	23
193	192	Female	32	103	69
194	193	Male	33	113	8
195	194	Female	38	113	91
196	195	Female	47	120	16
197	196	Female	35	120	79
198	197	Female	45	126	28
199	198	Male	32	126	74
200	199	Male	32	137	18

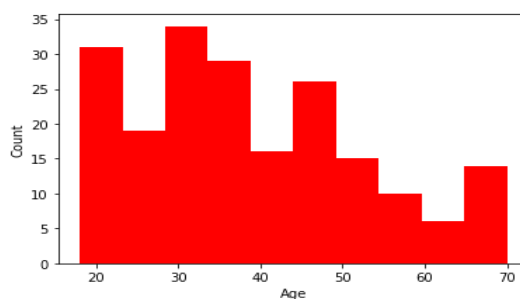
Gambar 5 Nilai K-means untuk variabel CustomerID,Gender,Age,Annual

Income (kRp),Spending Score (1-100)



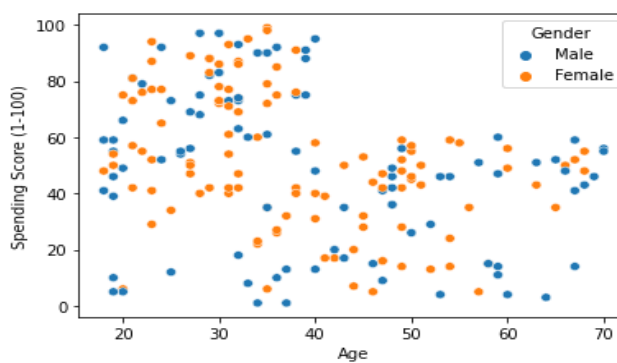
Gambar 6 Grafik Hasil Prediksi Plot gender - Pengamatan

Dari Count plot terlihat bahwa jumlah pelanggan Wanita lebih banyak dari jumlah total pelanggan Pria

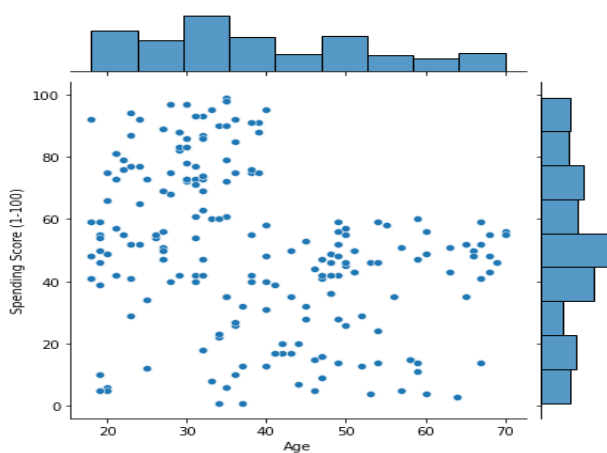


Gambar 7 Plot usia – Observasi

Dari Histogram terlihat ada 3 kelompok umur yang lebih sering berbelanja di mall, yaitu: 21-29 tahun, 30-40 tahun dan 45-50 tahun.



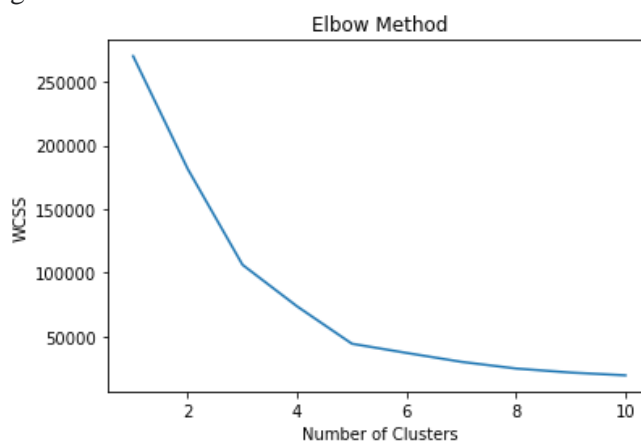
Gambar 8 Spending Score Gender



Gambar 9 Skor Usia Pengeluaran dan Pengamatan

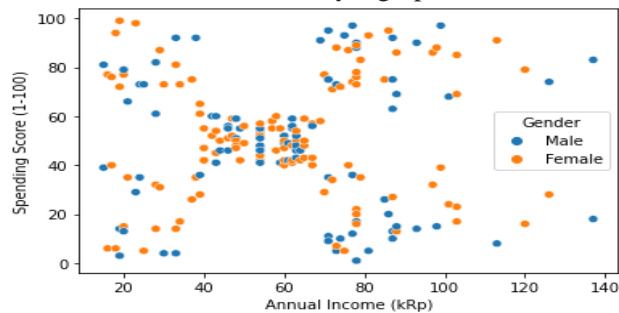
Dari plot Skor pengeluaran dan pengamatan bahwa ada 2 cluster dan dapat dikategorikan sebagai berikut :

- 1) Dari plot Age Vs Spending Score kami mengamati bahwa pelanggan yang memiliki nilai pengeluaran lebih dari 65 memiliki Usia mereka dalam kisaran 21-38 tahun. Juga dari plot Scatter diamati bahwa pelanggan yang skor pengeluarannya lebih dari 65 terdiri dari lebih banyak Wanita daripada Pria.
- 2) Pelanggan yang memiliki rata-rata nilai pembelanjaan yaitu: pada rentang 40-60 terdiri dari kelompok usia rentang 15-35 tahun dan jumlah Pria dan Wanita pada kelompok usia ini juga kurang lebih sama.

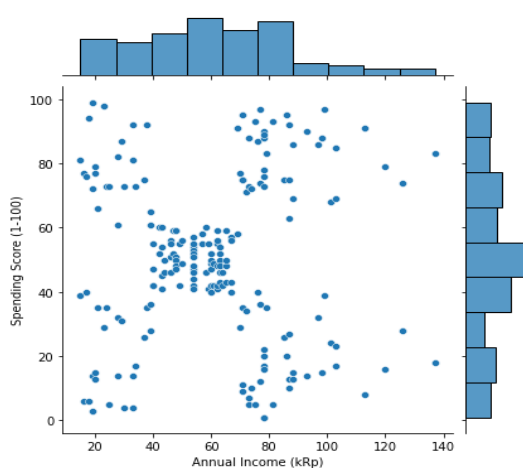


Gambar 10 Gambar Elbow Method

Dari plot di atas diamati bahwa 5 cluster yang optimal untuk dataset yang diberikan.



Gambar 11 Spending Score



Gambar 12 Skor Pendapatan Vs Pengeluaran Tahunan

Dari plot Skor Pendapatan Vs Pengeluaran Tahunan bahwa ada 5 cluster dan dapat dikategorikan sebagai:

- Pendapatan Tinggi, Skor Pengeluaran Tinggi (Cluster Kanan Atas)
- Pendapatan Tinggi, Skor Pengeluaran Rendah (Cluster Kanan Bawah)
- Pendapatan Rata-Rata, Skor Pengeluaran Rata-Rata (Cluster Pusat)
- Pendapatan Rendah, Skor Pengeluaran Tinggi (Cluster Kiri Atas)
- Pendapatan Rendah, Skor Pengeluaran Rendah (Cluster Kiri Bawah)



Gambar 13 Pengelompokan Observasi

Dari pengelompokan observasi bahwa ada 5 cluster dan dapat dikategorikan sebagai:

- A. Pendapatan Tinggi, Skor Pengeluaran Tinggi (Cluster 5) - Targetkan pelanggan ini dengan mengirimkan peringatan produk baru yang akan meningkatkan pendapatan yang dikumpulkan oleh mal karena mereka adalah pelanggan setia.
- B. Pendapatan Tinggi, Skor Pengeluaran Rendah (Cluster 3) - Targetkan pelanggan ini dengan meminta umpan balik dan mengiklankan produk dengan cara yang lebih baik untuk mengubahnya menjadi pelanggan Cluster 5.
- C. Pendapatan Rata-Rata, Skor Pengeluaran Rata-Rata (Cluster 2) - Dapat menargetkan kumpulan pelanggan ini dengan memberi mereka EMI berbiaya rendah, dll.
- D. Pendapatan Rendah, Skor Pengeluaran Tinggi (Cluster 1) - Mungkin atau mungkin tidak menargetkan kelompok pelanggan ini berdasarkan kebijakan mal.
- E. Pendapatan Rendah, Skor Pengeluaran Rendah (Cluster 4) - Jangan targetkan pelanggan ini karena mereka memiliki pendapatan lebih sedikit dan perlu menghemat uang.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan yang sudah dijabarkan diatas, kesimpulan yang didapat dalam menjalani penelitian ini serta saran yang diharapkan dapat bermanfaat dalam usaha untuk melakukan perbaikan dan pengembangan penelitian ini.

REFERENSI

- [1] Harani, N. H., Prianto, C., & Nugraha, F. A. (2020). Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python. *Jurnal Manajemen Informatika (JAMIKA)*, 10(2), 133–146. <https://doi.org/10.34010/jamika.v10i2.2683>
- [2] Puspitasari, N., Widians, J. A., & Setiawan, N. B. (2020). Customer segmentation using bisecting k-means algorithm based on recency, frequency, and monetary (RFM) model. *Jurnal Teknologi Dan Sistem Komputer*, 8(2), 78–83. <https://doi.org/10.14710/jtsiskom.8.2.2020.78-83>
- [3] Akhir, T., Sofiah, M., Manajemen, P. S., Ekonomi, F., Ilmu, D. A. N., & Bakrie, U. (2021). Analisis Segmentasi Pelanggan Pada Top 5 Marketplace Di Indonesia Menggunakan Metode K- Means Clustering Analisis Segmentasi Pelanggan Pada Top 5 Marketplace Di Indonesia Menggunakan Metode K-.
- [4] Bassy, N. T., Studi, P., Fakultas, I., Informasi, T., Kristen, U., & Wacana, D. (2018). Implementasi k-means untuk segmentasi data pelanggan dengan model lrjm.
- [5] Sani, A., Studi, P., & Informatika, T. (n.d.). *PENERAPAN METODE K- MEANS CLUSTERING PADA*.
- [6] Febriani, A., & Putri, S. A. (2020). Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means. *Jiems (Journal of Industrial Engineering and Management Systems)*, 13(2), 52–57. <https://doi.org/10.30813/jiems.v13i2.2274>
- [7] Murpratiwi, S. I., Agung Indrawan, I. G., & Aranta, A. (2021). Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 18(2), 152. <https://doi.org/10.23887/jptk-undiksha.v18i2.37426>
- [8] Prasetyo, S. S., Mustafid, M., & Hakim, A. R. (2020). Penerapan Fuzzy C-Means Kluster Untuk Segmentasi Pelanggan E-Commerce Dengan Metode Recency Frequency Monetary (Rfm). *Jurnal Gaussian*, 9(4), 421–433. <https://doi.org/10.14710/j.gauss.v9i4.29445>
- [9] Angelie, A. V. (2017). *Segmentasi Pelanggan Menggunakan Clustering K-Means Dan Model Rfm (Studi Kasus : Pt . Bina Adidaya Surabaya) Customer Segmentation Using K-Means Clustering and Rfm Model (Case Study : Pt . Bina Adidaya Surabaya Segmentasi Pelanggan Menggunakan Clusteri*. 163.
- [10] Studi, P., Informasi, S., & Sarjana, J. (2019). Oleh Nadhira Amalia Diandra.