

Klasterisasi Customer Lifetime Value Pelanggan Menggunakan K-Means pada Online Travel Agency

I Putu Ananta Wijaya^{a1}, I Ketut Gede Suhartana^{a2}

^aInformatika, Universitas Udayana
Kampus Bukit Jimbaran, Indonesia
¹anantaw81@gmail.com
²suhartana@gmail.com

Abstraksi

Penjualan produk didapati pada perusahaan yang bergerak di bidang *e-commerce*, salah satunya adalah *online travel agency* (OTA). Setelah *covid-19*, perusahaan sedang bangkit dengan persaingan yang semakin banyak muncul dan pelanggan yang beralih ke *e-commerce* pesaing. Menjaga hubungan dari satu pelanggan ke pelanggan lainnya diperlukan. Data historis transaksi pelanggan dapat digunakan untuk melakukan klasterisasi. Sebelum itu, dilakukan pembentukan model RFM yaitu *recency*, *frequency*, dan *monetary* supaya bisa dilakukan perhitungan *customer lifetime value*. Klasterisasi yang digunakan adalah K-Means karena sederhana dan efisien. Penentuan jumlah klaster optimal menggunakan metode *Gap Statistic*. Jumlah klaster akan divalidasi separasinya oleh metode *davies bouldin index*. Setiap data klaster kemudian dicari rata-rata untuk nilai setiap fitur model RFM. Hasilnya akan dikali dengan bobot yang didapat dari AHP. Klaster dengan loyalitas tertinggi didapatkan pada klaster 2 yaitu 0.0297. Nilai fitur *frequency* pada klaster 2 juga merupakan nilai tertinggi. Hal ini menandakan klaster 2 merupakan pelanggan dengan nilai loyalitas yang tertinggi. Perusahaan dapat merancang strategi untuk pelanggan dengan program seperti pemberian diskon atau poin. Sedangkan, pada klaster dengan nilai loyalitas terendah bisa dibuatkan program *personalized product*.

Kata kunci: *Image Processing, Klasterisasi, Customer Lifetime Value, RFM, K-Means, Gap Statistic*

1. Pendahuluan.

Pelanggan merupakan kunci dari penjualan produk. Penjualan produk didapati pada perusahaan yang bergerak di bidang *e-commerce*, salah satunya adalah *online travel agency* (OTA). Pada saat *covid-19*, perusahaan OTA mengalami penurunan pelanggan dikarenakan *lockdown* [10]. Pada saat ini, perusahaan sedang bangkit dengan persaingan yang semakin banyak muncul dan pelanggan yang beralih ke *e-commerce* pesaing. Oleh karena itu, salah satu tujuan dari perusahaan *e-commerce* adalah menjaga hubungan dari satu pelanggan ke pelanggan lainnya. Hal ini menunjukkan perusahaan harus mampu mengidentifikasi pelanggan karena setiap pelanggan itu unik yaitu punya preferensi dan perilaku berbeda.

Dalam *data mining*, penggunaan data historis transaksi pelanggan dapat digunakan untuk melakukan klasterisasi. Metode klasterisasi yang digunakan bertipe non hierarki karena tidak mempunyai tingkatan. Metode yang dipilih adalah *k-means*. Menurut Bao Chong [3], *k-means* mempunyai keuntungan yaitu sederhana, sangat efisien, dan gampang diimplementasikan sehingga dipakai di banyak bidang. Algoritma ini juga lebih cepat dibandingkan algoritma klasterisasi yang lain dari segi waktu eksekusi. Algoritma *k-means* sendiri tidak bisa menentukan jumlah klaster yang akan dibuat. Oleh karena itu, digunakan *gap statistic* yang mengambil nilai *gap* terbesar dari nilai perhitungan data asli dan data referensi untuk digunakan sebagai jumlah klaster optimal [7]. Setelah itu, jumlah klaster akan divalidasi dengan *davies bouldin index* (DBI) untuk menentukan apakah data sudah terseparasi dengan baik. Nilai DBI akan optimal jika semakin kecil [9].

Untuk mendapatkan klaster dengan nilai loyalitas terbaik, maka digunakan perhitungan *customer lifetime value* (CLV). Sejumlah nilai potensial yang dimiliki oleh setiap pelanggan dalam bentuk mata uang, yang dapat diperoleh oleh perusahaan dari pembelian produk atau layanan, berdasarkan tingkat kesetiaan pelanggan tersebut di masa depan. Nilai inilah yang diamati karena manfaat yang diberikan

saat ini hingga di masa depan bagi perusahaan [5]. Perhitungan CLV sendiri memerlukan pembuatan model sebelum klasterisasi, yaitu model *recency*, *frequency*, dan *monetary* (RFM). Setelah klasterisasi, setiap klaster akan memiliki rata-rata setiap fitur model. Rata-rata ini akan dikalikan dengan bobot yang didapatkan dari proses *analytical hierarchy process* (AHP). AHP adalah pendukung keputusan yang akan menguraikan masalah multi faktor atau multi kriteria yang kompleks menjadi suatu hirarki. Nilai yang terbesar didapatkan oleh CLV akan menjadi klaster dengan loyalitas terbaik. Klaster ini dapat diberikan perlakuan khusus seperti memberikan diskon dan poin.

2. Metode Penelitian

Penelitian menggunakan data historis transaksi pelanggan dari sebuah perusahaan OTA dengan rentang waktu Juni 2022 hingga Juni 2023. Kemudian data akan dilakukan *preprocessing* yaitu transformasi ke model RFM dan normalisasi dengan *min-max*. Setelah itu, dilakukan klasterisasi dengan *k-means* dan bantuan *gap statistic* untuk menentukan jumlah klaster optimal. Hasil klaster kemudian akan divalidasi dengan DBI untuk menentukan data per klaster sudah terseparasi dengan baik. Setelah itu dilakukan pencarian nilai CLV dengan bantuan bobot dari AHP.

a. K-Means

Metode ini merupakan salah satu yang tertua dan sangat populer di kalangan praktisi karena kemudahan implementasi dan kecepatan prosesnya [6]. *K-means* berbasis titik pusat (*centroid*) dengan metode partisi (*partitioning method*). *clustering* data menggunakan dengan metode *k-means* bertujuan untuk mempartisi atau membagi n objek data ke dalam k kelompok data (*klaster*). Kemiripan data pada suatu *klaster* diukur sesuai dengan kedekatan obyek terhadap nilai rata-rata pada *klaster* atau disebut sebagai *centroid*. Berikut ini merupakan langkah-langkah dari metode *clustering* menggunakan algoritma *k-means*:

1. Algoritma *k-means* menerima inputan dari pengguna berupa nilai jumlah klaster sebanyak k klaster, dan suatu dataset yang ingin dikelompokkan.
2. Sebanyak k data dipilih secara *random* dari suatu dataset yang kemudian ditentukan sebagai *initial centroid*.
3. Hitung jarak masing-masing titik objek data ke masing-masing *initial centroid*. Perhitungan jarak dapat dilakukan menggunakan persamaan ukuran jarak *euclidian distance* ditampilkan pada (1).

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

4. Mengalokasikan masing-masing titik data ke dalam *klaster* dengan jarak paling minimum.
5. Menghitung rata-rata dari semua data yang terdapat di dalam *klaster* tersebut sebagai pusat klaster baru.
6. Ulangi langkah 4, 5, dan 6 hingga tidak ada lagi objek yang berubah di dalam suatu klaster.

b. Gap Statistic

Dalam menentukan k klaster optimal untuk klasterisasi, digunakan metode *gap statistic*. *Gap statistic* dipublikasikan oleh Tibshirani [7] *Gap statistic* adalah metode yang digunakan untuk mengukur "gap" antara kualitas klaster dari data asli dan kualitas dari data referensi. *Gap statistic* memerlukan dua data matriks yaitu data asli dan data referensi. Berikut langkah-langkah algoritma *gap statistic*

1. Buat data referensi (*references*) menggunakan *sampling* secara *uniformly* (distribusi secara rata) dari data asli. Batas data referensi yang dibuat mengikuti jarak terjauh titik data pada data asli.
2. Lakukan klasterisasi pada data referensi dengan batas sampai $n_references$ (nilai dapat disesuaikan). Hitung *euclidean distance* untuk setiap subset hasil klaster dan rata-ratakan. Ulangi sampai $n_references$ dan jumlahkan kemudian dirata-ratakan. Nilai ini dimasukkan ke *array* baris ke 1. Lakukan sampai batas k_max (batas jumlah klaster yang ingin dicari nilai *gap statistic*-nya) sehingga hasilnya berupa *array* dengan panjang k_max .
3. Lakukan klasterisasi pada data asli. Hitung *euclidean distance* untuk setiap subset hasil klaster. Setelah itu hasilnya langsung masuk ke *array* baris ke 1, dilakukan sampai batas k_max sehingga tercipta juga *array* dengan panjang k_max .
4. Hitung nilai logaritma setiap *item* dari dua *array* (data referensi dan data asli).
5. Kurangi setiap *item* pada *array* hasil data referensi dengan *array* hasil data asli. Didapatkan *array* baru berisi nilai *gap* dan ambil yang paling besar.

c. Davies Bouldin Index

Davies bouldin index (DBI) adalah metode evaluasi untuk mengukur validitas kluster. Pengukuran DBI dilakukan dengan mengukur nilai maksimal jarak *inter-cluster* dan nilai minimal jarak antar kluster. Apabila jarak *inter-cluster* maksimal, maka karakteristik masing-masing kluster akan terlihat jelas. Menurut Bates dan Kalita [4], semakin kecil nilai DBI yang diperoleh atau semakin mendekati nol, maka semakin baik kluster tersebut dalam kepadatan dan separasi. Rumus DBI dapat dilihat pada persamaan (2).

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (2)$$

d. Model RFM

Model RFM digunakan untuk analisis perilaku pelanggan [8]. Penjelasan setiap fitur dari model tersebut sebagai berikut.

1. *Recency*
Tanggal pembelian terakhir pelanggan dan tanggal referensi pada periode penelitian. Pelanggan yang baik, yang berulang kali mengunjungi perusahaan akan memiliki nilai yang kecil.
2. *Frequency*
Mewakili berapa kali pembelian yang dilakukan oleh pelanggan. Pelanggan dianggap setia asalkan nilai frekuensinya besar.
3. *Monetary*
Mewakili jumlah total yang dibelanjakan pelanggan selama periode penelitian. Semakin besar, semakin banyak uang yang dihasilkan perusahaan.

e. *Analytical Hierarchy Process* dan *Customer Lifetime Value*

Analytical hierarchy process (AHP) adalah sistem pendukung keputusan dengan sebuah hierarki fungsional dengan input utamanya persepsi manusia [2]. Dengan hierarki, suatu masalah kompleks dan tidak terstruktur dipecahkan ke dalam kelompok-kelompok tersebut diatur menjadi suatu bentuk hierarki. Persepsi manusia yang di input adalah setujuannya suatu fitur RFM dalam menentukan loyalitas suatu pelanggan. Hasilnya adalah pembobotan setiap fitur yang ada. *Customer Lifetime Value* (CLV) digunakan untuk mengukur profitabilitas dari setiap pelanggan. Penggunaan CLV dapat membantu perusahaan untuk menentukan strategi terbaik untuk mencetak laba (Kurniawan, dkk 2021). Perhitungan CLV dilakukan setelah menyelesaikan proses klusterisasi. Setelah itu untuk setiap kluster akan diranking. Rumus CLV untuk suatu kluster k ditampilkan pada persamaan (3).

$$CLV_k = W_r C_r^k + W_f C_f^k + W_m C_m^k \quad (3)$$

Keterangan

CLV_k : Nilai CLV untuk kluster k

C_r^k, C_f^k, C_m^k : Nilai rata-rata R, F, M untuk kluster k

W_r, W_f, W_m : Bobot yang dihasilkan dari AHP

3. Hasil dan Pembahasan

3.1. Data dan Model RFM

Data yang didapatkan adalah data primer historis transaksi pelanggan dari salah satu perusahaan OTA berlokasi di Bali dari Juni 2022 hingga Juni 2023. Data tersebut saat digabungkan berjumlah 10946 baris. Cuplikan singkat datanya berupa nama kolom dan isi data dapat dilihat pada tabel 1.

Tabel 1. Cuplikan Data Historis Transaksi

<i>id</i>	<i>user_id</i>	<i>price</i>	<i>payment_date</i>	<i>currency</i>	<i>type</i>
50662	16923	169	01/06/2022 03:19	AUD	flight
50686	26945	3018	01/06/2022 07:27	AUD	flight
50724	42968	1711	01/06/2022 10:40	AUD	flight

50821	37038	141	02/06/2022 10:22	AUD	flight
50844	43410	4196	02/06/2022 13:15	AUD	flight
50858	43424	170	02/06/2022 21:06	AUD	flight
...

Atribut yang diproitasakan untuk transformasi data adalah kolom *price*, *payment_date*, dan *user_id*. Model RFM didapatkan dari selisih tanggal sekarang dengan *payment_date*, jumlah transaksi yang pernah dilakukan *user_id*, total *price* setiap *user_id*. Kode *sql* yang digunakan untuk transformasi adalah pada tabel 2. Kode ini melakukan pembuatan tabel dengan pengelompokkan berdasarkan *user_id* setelah itu diambil dari tabel *combined_transactions* berupa *user_id*, *recency* dari perbedaan tanggal dari tanggal sekarang dengan tanggal maksimal yang ada pada *user_id* tersebut, *frequency* dari jumlah id yang ada pada *user_id* tersebut, *monetary* dari jumlah *price* yang ada pada *user_id* tersebut.

Tabel 2. Kode Transformasi RFM

```
CREATE TABLE rfm_transactions AS SELECT user_id, DATEDIFF(CURDATE(),
MAX(payment_date)) AS recency, COUNT(id) AS frequency, SUM(price) AS monetary
FROM combined_transactions GROUP BY user_id
```

Dari hasil tabel baru tersebut, didapatkan statistik datanya untuk setiap fitur RFM (*recency*, *frequency*, *monetary*) berupa jumlah, rata-rata, standar deviasi, nilai terkecil, kuartil, nilai terbesar. Statistik tersebut ditunjukkan oleh tabel 3. Cuplikan data pada tabel baru bernama *rfm_transactions* ditunjukkan oleh tabel 4 yang berisi tipe data numerik untuk fitur RFM-nya.

Tabel 3. Statistik Deskriptif

	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
jumlah	7808	7808	7808
rata-rata	402.626537	2.292008	1430.975282
standar deviasi	114.431750	2.261620	2155.253033
nilai terkecil	162.000000	1.000000	5.000000
kuartil 1 (25%)	304.000000	1.000000	474.000000
kuartil 2 (50%)	395.000000	2.000000	840.500000
kuartil 3 (75%)	501.000000	2.000000	1612.500000
nilai terbesar	612.000000	54.000000	69994.000000

Tabel 4. Cuplikan Data RFM

<i>user_id</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
97	272	20	23135
443	326	3	1749
487	265	2	5540
619	274	4	3224
633	410	1	996
639	566	1	222
...

Setelah melalui proses transformasi, selanjutnya adalah normalisasi data. Normalisasi data dapat membantu proses klasterisasi agar lebih cepat. Metode *min-max* digunakan untuk proses nilai minimum

dan maksimum dari setiap variabel sehingga hasilnya berada di rentang 0 sampai 1. Ilustrasi proses diberikan untuk *user_id* 97 (1) dan *user_id* 443 (2) untuk kolom *recency*.

$$x'_1 = \frac{X_1 - X_{min}}{X_{max} - X_{min}} = \frac{612 - 162}{272 - 162} = 0,24444$$

$$x'_2 = \frac{X_1 - X_{min}}{X_{max} - X_{min}} = \frac{326 - 162}{612 - 162} = 0,36444$$

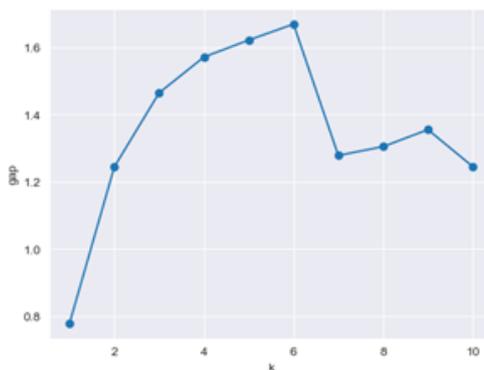
Didapatkan hasil normalisasi masing-masing 0,24444 dan 0,36444 untuk *user_id* 97 dan 443 untuk kolom *recency*. Hasil cuplikan beberapa data setelah normalisasi *min-max* yang datanya terlihat memiliki rentang di antara 0 hingga 1 ditunjukkan oleh tabel 5.

Tabel 5. Cuplikan Tabel Normalisasi RFM

	<i>user_id</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
0	97	0.244444	0.358491	0.330481
1	443	0.364444	0.037736	0.024918
2	487	0.228889	0.018868	0.079084
3	619	0.248889	0.056604	0.045993
4	633	0.551111	0.000000	0.014159
5	639	0.897778	0.000000	0.003100
6	650	0.175556	0.018868	0.043435
...

3.2. Klusterisasi dengan *K-Means* dan *Gap Statistic*

Klusterisasi pelanggan menggunakan salah satu metode klusterisasi non hirarki *k-means* yang akan membagi data menjadi beberapa kluster tertentu. Penentuan jumlah kluster yang optimal akan dibantu dengan metode *gap statistic* yang mengambil selisih terbesar dari nilai total jarak *euclidean* seluruh kluster untuk data referensi distribusi dengan data asli. Hasilnya jumlah kluster (*k*) optimal yang didapatkan adalah 6 seperti pada gambar 1. Hasil optimal dapat terlihat dari nilai *gap* tertinggi yang didapatkan.



Gambar 1. Grafik Hasil Penentuan *k* Optimal *Gap Statistic*

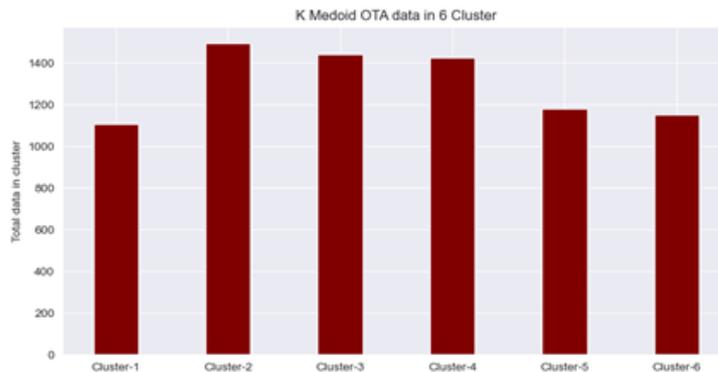
Klusterisasi akan dilakukan menggunakan *method KMeans* pada *library* *sklearn* di *python*. Kode yang digunakan ditunjukkan oleh tabel 6. Kode ini melakukan pembuatan objek *kmeans* yang selanjutnya inialisasi parameter jumlah kluster sebanyak 6 di *n_clusters*. Setelah itu, data yang telah dinormalisasi diubah terlebih dahulu ke tipe *numpy* yang sebelumnya bertipe *dataframe*. Setelah itu, dipanggil metode *fit_predict* untuk melakukan klusterisasi *k-means* tersebut.

Tabel 6. Kode Klasterisasi dengan *Sklearn*

```

from sklearn.cluster import KMeans
kmeans = KMeans()
kmeans.n_clusters = 6
data_target_array = df_normalized[['recency', 'frequency',
'monetary']].to_numpy()
cluster_no_every_data = kmeans.fit_predict(data_target_array)
    
```

Hasil klasterisasi menciptakan 6 klaster dengan pembagian jumlah tiap klaster yang merata ditampilkan pada gambar 2. Hasil cuplikan data setelah melakukan klasterisasi ditampilkan pada gambar 3. Terlihat data berupa kolom *user_id*, kolom RFM, serta nomor klaster yang didapatkan.



Gambar 2. Jumlah Data Setiap Klaster

Tabel 7. Cuplikan Hasil Klasterisasi

	<i>user_id</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>	<i>cluster</i>
0	97	0.244444	0.358491	0.330481	6
1	443	0.364444	0.037736	0.024918	2
2	487	0.228889	0.018868	0.079084	6
3	619	0.248889	0.056604	0.045993	6
4	633	0.551111	0.000000	0.014159	3
5	639	0.897778	0.000000	0.003100	4
6	650	0.175556	0.018868	0.043435	6
...

Setelah mendapatkan nomor klaster untuk setiap data, selanjutnya dilakukan validasi jumlah klaster dengan *davies bouldin index* untuk menentukan hasil klaster memiliki kepadatan yang baik dan terseparasi. Semakin mendekati nol, maka semakin baik hasil klasterisasi. Penentuan nilai menggunakan metode *davies_bouldin_index* dari *library perimetrics* pada bahasa pemrograman *python*. Hasilnya, jumlah klaster 6 mendapatkan nilai 0,45419 yang mendekati nol sehingga masih bisa dibilang klaster memiliki kepadatan baik dan terseparasi dengan baik.

3.3. Pembobotan AHP dan Perangkingan CLV

AHP mengambil keputusan secara hirarki yang dipilih dari berbagai kriteria dan alternatif. Dalam permasalahan klasterisasi, hierarki yang dievaluasi adalah model RFM. Dalam hal ini, peneliti menggunakan bobot yang sama dengan penelitian [11] dengan nilai W_r , W_f , W_m yaitu 0.058, 0.546, dan 0.395. Penentuan nilai CLV tertinggi akan dilakukan dengan penjumlahan hasil perkalian setiap

rata-rata fitur model dengan setiap bobot. Rata-rata setiap fitur kluster ditampilkan pada tabel 8. Hasil perkalian rata-rata fitur dengan bobot fitur ditampilkal pada tabel 9.

Tabel 8. Rata-rata Nilai Fitur per Kluster

<i>cluster</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
1	0.929360	0.018136	0.015502
2	0.189759	0.035291	0.023710
3	0.492069	0.021958	0.018664
4	0.343677	0.028830	0.024190
5	0.784051	0.019539	0.017239
6	0.637803	0.018672	0.021350
Rata-rata	0.56278638	0.02373775	0.02010918

Tabel 9. Perangkingan Customer Lifetime Value (CLV)

No cluster	$W_r C_r$	$W_f C_f$	$W_m C_m$	CLV	Peringkat
1	0.0053820800	0.009901456	0.00612439	0.0214	6
2	0.0011005820	0.019283686	0.00934895	0.0297	1
3	0.0028510022	0.011978188	0.00736328	0.0222	4
4	0.0019940986	0.015743580	0.00952755	0.0273	2
5	0.0045419458	0.010659294	0.00680005	0.0220	5
6	0.0036960474	0.010178112	0.00841675	0.0223	3
Rata-rata	0.00292	0.0139	0.00712	0.0249	

Didapatkan hasil kluster dengan nilai CLV dan peringkatnya. Pemeringkatan dilakukan untuk mengetahui kluster yang memiliki nilai hidup pelanggan tertinggi. Hasilnya, kluster 2 memiliki nilai CLV tertinggi sebesar 0,0297 yang berarti kluster 2 merupakan pelanggan dengan nilai loyalitas tertinggi.

4. Kesimpulan

Klasterisasi data pelanggan OTA untuk Juni 2022 hingga Juni 2023 menggunakan *k-means* dan *gap statistic* diperoleh jumlah kluster optimal sebanyak 6, dari kluster kecil masing-masing jumlahnya 1109, 1497, 1441, 1428, 1180, 1153. Hasil validasi dari *davies bouldin index* juga nilai 0,45419 yang artinya mendekati 0 sehingga dapat dikatakan kluster terseparasi dengan baik dan memiliki kepadatan yang baik.

Nilai *customer lifetime value* (CLV) yang tertinggi didapatkan pada kluster 2 yaitu 0.0297. Nilai fitur *frequency* pada kluster 2 juga merupakan nilai tertinggi. Hal ini menandakan kluster 2 merupakan pelanggan dengan nilai loyalitas yang tertinggi. Perusahaan dapat merancang strategi untuk mempertahankan pelanggan dengan program-program seperti pemberian diskon atau poin sehingga pelanggan sehingga keuntungan meningkat. Sedangkan, untuk nilai CLV terendah ada pada kluster 1, yaitu 0.0214. Perusahaan dapat membuat *personalized product* yaitu menampilkan produk-produk terkait untuk diberikan kepada pelanggan kluster 1 sehingga mereka akan lebih menjelajahi produk tersebut.

Referensi

- [1] A. R. Kurniawan, R. Mayasari and J. Febriantoko, "CUSTOMER LIFETIME VALUE METRICS AS CUSTOMER RELATIONSHIP MANAGEMENT INDICATORS" *International Journal of Economics, Business and Accounting Research (IJEBAR)*, vol 5, no. 3, p. 956-966, 2021.

- [2] A. Natalis, Y. Natalian, "Pemanfaatan k-Means Clustering dan Analytic Hierarchy Process terhadap Penilaian Prestasi Kerja Pegawai" *Jurnal Teknik Informatika dan Sistem Informasi*, vol 8, no. -, p. 88 – 99, 2022.
- [3] B. Chong, "K-means clustering algorithm: a brief review" *Academic Journal of Computing & Information Science*, vol. 4, no. 5, p. 37-40, 2021.
- [4] Bates, Andrew and K. Jugal, "Counting clusters in twitter posts" *Proceedings of the 2nd International Conference on Information Technology for Competitive Strategies*, vol. -, no. -, p. 85, 2016
- [5] E. A. Pratomo, M. Najib and H. Mulyati, "CUSTOMER SEGMENTATION ANALYSIS BASED ON THE CUSTOMER LIFETIME VALUE METHOD" *Journal of Applied Management*. vol. 17, no. 3, p. 408 – 415, 2019.
- [6] Suyanto, MACHINE LEARNING TINGKAT DASAR DAN LANJUT, 1 ed., Bandung: Informatika, 2018, pp. –
- [7] R. Tibshirani, G. Walther and Hastie, "Estimating the number of clusters in a data set via the gap statistic" *J. R. Statist. Soc. B*, vol. 63, no. 2, p. 411–423, 2001.
- [8] W. A. Taqwim, N. Y. Setiawan and F. A. Bachtiar, "Analisis Segmentasi Pelanggan Dengan RFM Model Pada Pt. Arthamas Citra Mandiri Menggunakan Metode Fuzzy C-Means Clustering" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 2, p. 1986 -1993, 2019.
- [9] Y. Asriningtias, J. Aryanto, "K-Means Algorithm with Davies Bouldin Criteria for Clustering Provinces in Indonesia Based on Number of Events and Impacts of Natural Disasters" *International Journal of Engineering, Technology and Natural Sciences*, vol. 4, no. 1, p. 75 – 80, 2022.
- [10] Z. Liu, J. Zeng and J. Zhou, "Impact of COVID-19 on OTA Platforms and Development: A Case Study of Ctrip" *International Conference on Economic Management and Green Development*, vol. -, no. -, p. 55-60, 2023.
- [11] S. Monalisa, P. Nadya and R. Novita, "Analysis for Customer Lifetime Value Categorization with RFM Model" *The Fifth Information Systems International Conference*, vol. 161, no. -, p. 834 – 840, 2019.