Jurnal Informatika: Jurnal pengembangan IT

Vol. 10, No. 3, 2025
ISSN : 2477-5126, DOI:10.30591/jpit.v10i3.8735

Segmentasi Pelanggan Berdasarkan Model LRFM Menggunakan Algoritma K-Means dan Optimasi Klaster Dinamis

Riyan wahyudi [1], Achmad solihin [2]
[1,2] Magister Ilmu Kompputer Fakultas Teknologi, Universitas Budi Luhur
Jl. Ciledug Raya, RT.10/RW.2, Petukangan Utara, Kec. Pesanggrahan,
Kota Jakarta Selatan, DKI Jakarta 12260, Indonesia

Info Artikel

Riwayat Artikel:

Received 2025-05-14
Revised 2025-06-23
Accepted 2025-06-28

Corresponding Author:

Riyan wahyudi

Email: riaanwhy@gmail.com

This is an open access article under the CC BY 4.0 license.

Abstract *–The number of tax training participants often does not meet the minimum quota, resulting in the cancellation of many training classes. Throughout 2022, there were 27 training classes that failed to take place due to a lack of participants. One of the reasons is that promotions have not utilised historical customer data to set marketing targets more precisely. By utilising historical customer data, companies can design more targeted promotional strategies and increase the number of training participants. Therefore, this research aims to segment customers using the dynamic K-Means algorithm based on the Length, Recency, Frequency, and Monetary (LRFM) model, so that customer behaviour patterns when registering for training can be identified. The clustering results are then visualised to facilitate analysis and decision-making. This research resulted in three customer segments, namely Loyal customers (Gold, 17%), Lost customers (Diamond, 64%), and New customers (Silver, 17%). With this segmentation, it is expected that the company can conduct more effective promotions and increase the number of trainees in the future.*

Keywords: *Data Mining; Customer Segmentation; K-Means; Model LRFM.*

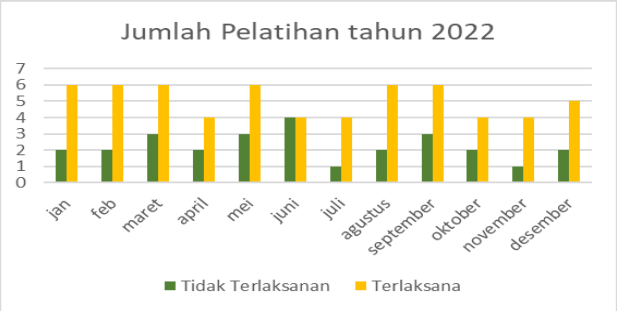
Abstrak *– Jumlah peserta pelatihan pajak sering kali tidak memenuhi kuota minimum, sehingga banyak kelas pelatihan batal dilaksanakan. Sepanjang tahun 2022, terdapat 27 kelas pelatihan yang gagal terlaksana akibat kurangnya jumlah peserta. Salah satu penyebabnya adalah promosi yang belum memanfaatkan riwayat data pelanggan untuk menetapkan target pemasaran secara lebih tepat. Dengan memanfaatkan riwayat data pelanggan, perusahaan dapat merancang strategi promosi yang lebih terarah dan meningkatkan jumlah peserta pelatihan. Oleh karena itu, penelitian ini bertujuan melakukan segmentasi pelanggan menggunakan algoritma K-Means dinamis berdasarkan model Length, Recency, Frequency, dan Monetary (LRFM), agar pola perilaku pelanggan saat mendaftar pelatihan dapat diidentifikasi. Hasil klasterisasi kemudian divisualisasikan untuk mempermudah analisis dan pengambilan keputusan. Penelitian ini menghasilkan tiga segmen pelanggan, yaitu Loyal customer (Gold, 17%), Lost customer (Diamond, 64%), dan New customer (Silver, 17%). Dengan adanya segmentasi ini, diharapkan perusahaan dapat melakukan promosi yang lebih efektif dan meningkatkan jumlah peserta pelatihan di masa mendatang*

Kata Kunci*: Data Mining, Segmentasi Pelanggan, K-Means, Model LRFM.*

I. PENDAHULUAN

Pelatihan perpajakan merupakan salah satu program unggulan yang diadakan secara rutin setiap minggu untuk mendukung pengembangan kompetensi praktis dan teoritis para pelaku bisnis maupun individu dalam bidang perpajakan. Tema yang diangkat dalam pelatihan ini bervariasi dan disesuaikan dengan kebutuhan terkini, Peserta pelatihan berasal dari beragam latar belakang, meliputi staf keuangan, akuntan, konsultan pajak, pelaku UMKM, hingga pejabat perusahaan yang menangani aspek perpajakan secara langsung. Dengan adanya variasi tema ini, setiap peserta diharapkan mampu memahami ketentuan perpajakan secara komprehensif dan mengimplementasikannya dalam operasional bisnis sehari-hari.

Namun, dalam kenyataannya, tidak semua kelas berhasil mencapai jumlah minimal peserta. Berdasarkan catatan data internal, sepanjang tahun 2022 terdapat 27 dari 87 kelas pelatihan pajak yang harus dibatalkan. Kondisi ini disebabkan kurang efektifnya promosi dan penentuan target pasar, sehingga kelas tidak memenuhi kuota minimal.



Gambar 1. Pelatihan Tahun 2022

Gambar 1 menunjukkan bahwa 30% dari total 87 pelatihan yang dijadwalkan sepanjang tahun 2022 tidak terlaksana. Berdasarkan tren pembatalan, terdapat 16 pelatihan yang dibatalkan pada semester pertama (Januari hingga Juni), dan 11 pelatihan lainnya dibatalkan pada semester kedua (Juli hingga Desember). Penyebab utama pembatalan pelatihan adalah kurangnya jumlah peserta, yang salah satunya dipengaruhi oleh kegiatan promosi yang belum memanfaatkan riwayat data pelanggan yang pernah mengikuti pelatihan sebelumnya. Dengan memanfaatkan riwayat data pelanggan, perusahaan dapat menetapkan target pemasaran secara lebih tepat dan efektif. Salah satu pendekatan yang dapat diterapkan adalah segmentasi data pelanggan, sehingga promosi dapat disesuaikan dengan karakteristik masing-masing segmen pelanggan. Dengan adanya segmentasi ini, diharapkan jumlah peserta pelatihan meningkat dan pelaksanaan pelatihan menjadi lebih optimal [1].

Pada dasarnya, terdapat berbagai metode yang dapat digunakan untuk melakukan segmentasi pelanggan. Salah satu pendekatan yang umum digunakan adalah mengekstraksi data riwayat transaksi pelanggan dalam periode waktu tertentu. Berdasarkan data tersebut, sejumlah variabel pembentuk segmentasi pelanggan dapat diidentifikasi sebagai kriteria utama dalam analisis. Model *Recency*, *Frequency*, dan *Monetary* (RFM) merupakan salah satu metode analisis nilai pelanggan yang sering diterapkan dalam segmentasi. Model ini banyak digunakan untuk mengevaluasi pola perilaku pelanggan dan menentukan prioritas strategi pemasaran agar lebih efektif dan tepat sasaran.[2], [3].

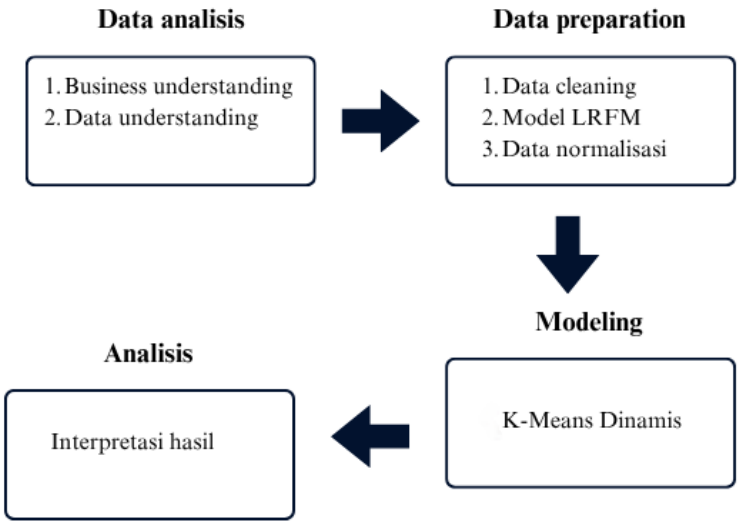
Model RFM kemudian dikembangkan lebih lanjut dengan menambahkan satu variabel baru, yaitu *Length* (L), sehingga menjadi model LRFM (*Length*, *Recency*, *Frequency*, *Monetary*). Variabel *Length* mengukur durasi waktu sejak transaksi pertama hingga transaksi terakhir pelanggan dalam periode tertentu. Penggunaan variabel ini penting untuk memperkuat analisis *Recency*, karena memberikan informasi tambahan mengenai lama hubungan pelanggan dengan perusahaan selain jarak waktu sejak transaksi terakhir. Dengan demikian, model LRFM lebih sesuai untuk digunakan dalam analisis *Customer Lifetime Value* (CLV).

Kelemahan utama model RFM adalah kurangnya pertimbangan terhadap faktor loyalitas pelanggan. Kekurangan ini diatasi dalam model LRFM, di mana variabel *Length* mampu memberikan indikasi loyalitas pelanggan dari sisi durasi hubungan. Selain itu, beberapa penelitian juga mencoba memodifikasi model RFM lebih jauh, misalnya dengan menambahkan parameter perilaku pengguna berbasis *event tracking* [4]. Penelitian sebelumnya yang bertujuan untuk mengkelompokan data pelanggan dengan teknik clustering diantaranya [5], [6], [7]. K-Means merupakan algoritma klastering yang sering digunakan. Namun demikian, performa K-Means dipengaruhi oleh penentuan titik awal pusat klaster [8], [9], [10], [11], [12]

Dynamic K-means *clustering* merupakan evolusi dari algoritma K-means yang memeriksa kembali kualitas klaster pada setiap iterasi dan memungkinkan jumlah klaster diubah untuk memenuhi kecukupan kualitas klaster. Algoritma klaster dinamis ini bertujuan untuk meningkatkan kualitas klaster sekaligus menentukan jumlah klaster (k) secara optimal agar hasil segmentasi lebih sesuai kebutuhan [13]. Dalam penelitian ini, data pelanggan yang mengikuti pelatihan pajak disegmentasi menggunakan model *Length*, *Recency*, *Frequency*, dan *Monetary* (LRFM), dan performa K-Means ditingkatkan melalui optimasi inisialisasi titik awal agar mampu memetakan pelanggan potensial secara lebih baik dan akurat [14]

II. METODE

Langkah-langkah penelitian untuk melakukan segmentasi pelanggan mencakup tahapan *business understanding, data understanding*, *data preparation*, modeling dan analisis. Pada tahap evaluation, hasil klasterisasi dianalisis untuk menilai kualitas dan validitas model segmentasi.tahap terakhir penerapan model segmentasi yang sesuai untuk mendukung proses pengambilan keputusan dan perancangan strategi pemasaran yang lebih tepat sasaran.[15]



Gambar 2. Alur penelitian

A. Data analisis

Pada tahap data analisis, terdapat dua proses utama yang dijalankan, yaitu business understanding dan data understanding. Proses-proses ini bertujuan untuk memastikan permasalahan bisnis teridentifikasi secara jelas serta kebutuhan data untuk mendukung analisis terpenuhi.

1) *Business understanding:* Pada tahap ini dilakukan studi lapangan terhadap pihak terkait, yaitu bagian divisi pemasaran di Esindo Multi Tata. Tujuan kegiatan ini adalah untuk mengidentifikasi permasalahan yang dihadapi dan menentukan solusi yang sesuai. Berdasarkan hasil studi lapangan, diperoleh beberapa informasi penting yang akan digunakan dalam penelitian, di antaranya: (1) permasalahan bisnis yang dihadapi, (2) proses bisnis perusahaan, dan (3) alternatif solusi untuk permasalahan tersebut
2) *Data understanding:* Pada tahap ini dikumpulkan informasi dan data yang relevan untuk mendukung analisis, khususnya dalam proses klastering. Dataset yang digunakan berasal dari basis data perusahaan dan mencakup: (1) Data pelanggan yang pernah mengikuti pelatihan perpajakan. (2) Data pelatihan pada periode 2017–2022. Dataset tersebut memiliki atribut-atribut penting untuk kebutuhan klastering, seperti tanggal pelatihan, nama perusahaan, jenis produk, jumlah total peserta, dan total biaya. Gambaran awal data transaksi pelanggan yang diperoleh dari *database* perusahaan ditunjukkan pada Tabel 1.

TABEL 1
DATASET TRANSAKSI PELATIHAN

PelangganId	Tgl	Ktgri	jmlHari	nopo	mktId	total	kodePel
1416	2017-01-03	2	2	JKT-2017-1905	13	6000000	3
2379	2017-01-03	1	2	JKT-2017-1907	11	2500000	3
1741	2017-01-05	1	2	JKT-2017-1915	11	8550000	3
2382	2017-01-06	1	2	JKT-2017-1918	18	3000000	3
368	2017-01-10	1	2	JKT-2017-1927	11	3000000	3

Dataset yang diperoleh kemudian melalui tahap pengolahan lebih lanjut berdasarkan kriteria dan metode yang telah ditetapkan. Gambaran umum dari dataset tersebut disajikan dalam bentuk statistik deskriptif untuk memberikan representasi awal sebelum dilakukan analisis lebih mendetail, seperti terlihat pada Tabel 2.

TABEL 2
STATISTIK DATA

Statistik	Total_bayar	Tgl	Jml_hari	Jml_transaksi
min	250000	03/01/2017	2	1
max	105550000	29/12/2022	32	16
mean	6575470	21/07/2019	3	2
median	3750000	01/03/2019	2	1

Berdasarkan Tabel 2, dapat dilihat bahwa terdapat variasi cukup signifikan dalam nilai transaksi maupun jumlah hari dan jumlah transaksi per pelanggan. Nilai pembayaran terendah adalah Rp 250.000, sedangkan nilai tertinggi mencapai Rp 105.550.000, menunjukkan adanya perbedaan daya beli dan kebutuhan pelanggan dalam mengikuti pelatihan. Selain itu, secara rata-rata pelanggan melakukan transaksi sebesar Rp 6.575.470 dan mayoritas hanya melakukan pembelian sekali hingga dua kali dalam periode pengamatan.

Dari sisi waktu, distribusi transaksi pelanggan berlangsung sepanjang periode 2017 hingga 2022, di mana nilai median transaksi terjadi pada 01 Maret 2019. Sementara itu, jumlah mengikuti pelatihan bervariasi mulai dari 2 hingga 32 kali mengikuti pelatihan, dengan nilai median 2 hari. Informasi ini menunjukkan bahwa mayoritas pelanggan cenderung mengikuti pelatihan dalam durasi singkat dan melakukan transaksi dalam jumlah terbatas.

Statistik deskriptif ini memberikan gambaran awal mengenai karakteristik pelanggan dan menjadi dasar untuk tahap selanjutnya, yaitu pembentukan model LRFM dan segmentasi pelanggan. Dengan memahami distribusi data secara keseluruhan, proses klastering diharapkan lebih sesuai dan mampu menangkap perbedaan pola transaksi antar pelanggan secara lebih optimal.

B. Data preparation

Pada tahap ini, dilakukan prapemrosesan data untuk memastikan bahwa data set yang terkumpul sesuai dengan kebutuhan analisis. Data diperoleh dari basis data transaksi penjualan yang dikumpulkan pada fase sebelumnya. Tahap prapemrosesan ini mencakup tiga proses utama, yaitu: (1) pembersihan data untuk menghilangkan data duplikat dan nilai kosong, (2) pembentukan model *Length, Recency, Frequency, Monetary* (LRFM) untuk menggambarkan karakteristik pelanggan, dan (3) normalisasi data agar setiap atribut memiliki skala yang seragam dan siap digunakan dalam proses klastering.

1) *Data cleaning:* Pada tahap Pembersihan data adalah subproses untuk membersihkan data set yang sudah dimiliki. Data yang perlu dibersihkan adalah data kosong, negatif, atau di bawah standar. Tujuan pembersihan data adalah untuk menghasilkan data yang bersih dan siap untuk digunakan.
2) *Pemodelan LRFM:* Pada tahap pemodelan LRFM membentuk variabel *Length, Recency, Frequency* dan *Monetary* data set hasil cleaning. Pemodelan variabel tersebut meliputi:
   a. *Length*: ditentukan dari selisih waktu antara transaksi pertama dan terakhir pelanggan. Atribut tanggal digunakan untuk menghitung nilai variabel ini.
   b. *Recency*: Variabel ini ditentukan dari selisih waktu antara keikutsertaan terakhir yang dilakukan oleh masing-masing pelanggan dengan tanggal 31 Desember 2022. Atribut tanggal transaksi digunakan untuk menghitung nilai variabel.
   c. *Frequency*: variabel ini dari jumlah transaksi keikutsertaan pelatihan Nilai diperoleh dari jumlah keikutsertaan di kali jumlah hari pelatihan yang dilakukan pada periode tahun 2017 sampai tahun 2022.
   d. *Monetary*: Variabel ini didapatkan dari total jumlah biaya pelatihan yang dikeluarkan pelanggan dalam setiap keikutsertaan pelatihan. Atribut Total harga digunakan untuk menghitung nilai variable.

TABEL 3
DATA LRFM

Detail	Model
Jumlah bulan antara pertama kali ikut pelatihan sampai tanggal terakhir kali ikut pelatihan	*Length*
Tanggal terakhir ikut pelatihan sampai tanggal acuan tanggal terakhir pengambilan data set yaitu bulan desember 2022	*Recency*
Jumlah hari keikutsertan pelanggan pada pelatihan yang diadakan diambil dari variable jumlah hari dikali jumlah keikutsertaan	*Frequency*
Total uang yang dibayarkan selama mengikuti pelatihan	*Monetary*

Setelah data diolah dan diterapkan ke dalam model LRFM, diperoleh data berjumlah 958 data, masing-masing merepresentasikan kode pelanggan beserta nilai atribute *Length*, *Recency*, *Frequency*, dan *Monetary* untuk setiap pelanggan. Data hasil pemodelan LRFM tersebut dapat dilihat pada Tabel 4.

TABEL 4
HASIL PEMODELAN LRFM

No	Kode pelanggan	*Length*	*Recency*	*Monetary*	*Frequency*
0	11	1	68	3000000	2
1	23	39	24	7500000	6
2	27	65	3	16250000	10
3	40	1	52	550000	2
4	41	1	3	14400000	2
----	-------	-------	-------	-----	-----
953	3462	1	1	3750000	2
954	3463	1	1	3750000	2
955	3467	1	1	3750000	2
956	3468	1	1	3600000	2
957	3470	1	1	7000000	2

Pada tahap pemodelan LRFM, setiap pelanggan direpresentasikan sebagai satu baris data, di mana setiap atribut menggambarkan aspek perilaku transaksional pelanggan.

Sebagai contoh, pelanggan dengan kode perusahaan 11 memiliki nilai Length sebesar 1, Recency 68 hari, Monetary Rp3.000.000, dan Frequency 2 transaksi. Ini menunjukkan bahwa pelanggan tersebut baru bergabung (length = 1), sudah cukup lama sejak transaksi terakhirnya (recency = 68), dan telah melakukan pembelian sebanyak 2 kali.

Sebaliknya, pelanggan berkode 27 memiliki nilai Length 65 hari, Recency 3 hari, Monetary Rp16.250.000, dan Frequency 10 kali transaksi. Ini mencerminkan bahwa pelanggan tersebut merupakan pelanggan lama dan cukup aktif bertransaksi dengan nominal pembelian yang besar. Secara keseluruhan, tabel hasil pemodelan LRFM ini menjadi dasar untuk tahapan segmentasi, di mana nilai-nilai tersebut akan digunakan sebagai input ke dalam algoritma K-means dinamis untuk mengelompokkan pelanggan sesuai pola transaksionalnya. Setelah pembentukan model LRFM, tahap berikutnya adalah normalisasi data agar seluruh atribut memiliki skala nilai yang seragam.

3) *Normalisasi:* Tahap pra-pemrosesan yang bertujuan untuk menyamakan skala nilai dari setiap atribut agar berada dalam rentang yang sama. Proses ini sangat penting dalam analisis klastering, terutama ketika menggunakan algoritma seperti K-Means, karena perbedaan skala antaratribut dapat menyebabkan bias pada hasil perhitungan jarak antar-observasi. Dengan melakukan normalisasi, setiap variabel akan berkontribusi secara seimbang dalam pembentukan klaster.tahap normalisasi Diperlukan untuk menormalkan nilai data untuk menemukan keseimbangan antara satu variabel dengan variabel lainnya.

normalisasi perlu dilakukan pada Semua variabel. Algoritma min-max normalisasi dengunakan untuk menormalisasi data dengan rentang 0 sampai 1. Hasil normalisasi dapat dilihat pada Tabel 5

TABEL 5
DATA NORMALISASI

Kode_pelanggan	*Length*	*Recency*	*Frequency*	*Monetary*
11	*0*	*0,9571*	*0*	*0,0261*
23	*0,5571*	*0,329*	*0,1875*	*0,0711*
27	*0,9286*	*0,0378*	*0,3125*	*0,154*
40	*0,0143*	*0,7323*	*0,0625*	*0,0052*
41	*0,0143*	*0,0417*	*0,0625*	*0,1364*
52	*0,0143*	*0,5631*	*0,0625*	*0,0284*
70	*0,8857*	*0,0699*	*0,25*	*0,1303*

C. Modeling

Pada tahap ini dilakukan implementasi klaster dinamis pada k-means untuk melakukan proses clustering. Data yang akan di proses data yang telah di normalisasi pada tahap sebelumnya. *Clustering* digunakan untuk mencari pola atau struktur yang bermakna dalam pengelompokan sejumlah data di dalam dataset [16]. *Clustering* memungkinkan untuk mengetahui karakteristik dan perilaku kelompok pelanggan sasaran dengan mengelompokkan pelanggan dengan kesamaan yang tinggi. Perilaku yang dijelaskan di sini berhubungan dengan empat variabel yang digunakan sebagai kriteria: length, recency, frequency, dan monetary Pada tahap ini data yang telah melalui proses normalisasi diolah untuk membentuk segmen pelanggan menggunakan algoritma K-Means dan variannya, yaitu K-Means dinamis. Kedua metode ini diaplikasikan secara berurutan untuk memaksimalkan kualitas klaster dan memastikan pembentukan segmen yang optimal.

1) *Algoritma K-Means Konvensional:* Algoritma K-Means merupakan salah satu teknik *partition-based clustering* yang banyak digunakan dalam segmentasi pelanggan. Proses klasterisasi dimulai dengan menentukan jumlah klaster k secara apriori dan menginisialisasi pusat klaster (*centroid*) secara acak. Setiap data x_i kemudian dialokasikan ke klaster terdekat berdasarkan jarak *Euclidean* sebagai berikut:

$$\left(D(X_2, X_1) = ||X_2 - X_1||^2 = \sqrt{\sum_{j=i}^{p} |X_{2j-X_{1j}}|^2}\right. \quad (1)$$

di mana c_j adalah pusat klaster ke-j, dan M merupakan jumlah atribut. Setelah semua data dialokasikan, pusat klaster dihitung ulang sebagai rata-rata dari anggota klaster:

$$c_j(new) = \left(\frac{1}{|S_j|}\right) \sum_{\{x_i \in S_j\}} x_i \quad (2)$$

Proses ini diulang hingga nilai pusat klaster konvergen atau hingga iterasi maksimum tercapai.

2) *Algoritma K-Means Dinamis:* Algoritma K-Means Dinamis merupakan pengembangan dari algoritma K-Means konvensional. Algoritma ini dirancang untuk menentukan jumlah klaster secara adaptif berdasarkan kualitas klaster. Tidak seperti K-Means biasa, yang memerlukan jumlah klaster k ditentukan di awal, algoritma K-Means dinamis mengevaluasi kualitas klaster di setiap iterasi dan menyesuaikan jumlah klaster sesuai kebutuhan hingga mencapai kualitas klaster optimal. Pada setiap iterasi, algoritma menghitung dua jenis jarak sebagai tolok ukur kualitas klaster: jarak antar klaster (inter-cluster) dan jarak dalam klaster (intra-cluster). Jarak antar klaster (inter) didefinisikan sebagai jarak minimum antar pusat klaster dan bertujuan untuk mengukur pemisahan antar klaster. Persamaan jarak antar klaster dapat dirumuskan sebagai berikut:

$$inter = min\{||m\_k - m\_(k+1)||\}, untuk\ k = 1, 2, \dots, K - 1 \quad (3)$$

Keterangan:

m_k = pusat klaster ke-k, dan K adalah jumlah klaster saat ini.

Jarak dalam klaster (intra) digunakan untuk mengukur kekompakan setiap klaster dan dihitung menggunakan simpangan baku (standard deviation) dari jarak setiap data ke pusat klaster.

$$\sqrt{\frac{1}{n-1}\Sigma_{i=}^{n}(x_i - x_M)^2 + b^2} \quad (4)$$

Berdasarkan pengukuran jarak intra dan inter klaster tersebut, algoritma menentukan apakah perlu menambah jumlah klaster. Jika jarak intra baru lebih kecil dari jarak intra lama dan jarak inter baru lebih besar dari jarak inter lama, maka jumlah klaster k akan ditingkatkan sebesar 1, dan proses klasterisasi diulang. Jika kondisi tersebut tidak terpenuhi, maka iterasi dihentikan. Secara lebih rinci**,** tahapan dalam algoritma K-Means dinamis adalah sebagai berikut:

1. Bentuk partisi awal sejumlah k klaster.
2. Tentukan pusat klaster awal (*centroid*) secara acak.
3. Hitung jarak semua data ke setiap pusat klaster.
4. Alokasikan setiap data ke klaster terdekat.
5. Perbarui pusat klaster sebagai rata-rata data dalam klaster tersebut.
6. Ulangi Langkah 3–5 hingga konvergensi (tidak terjadi perubahan keanggotaan klaster).
7. Hitung jarak antar klaster (inter) menggunakan Persamaan (3).
8. Hitung jarak dalam klaster (intra) menggunakan Persamaan (4).
9. Jika intra baru < intra lama dan inter baru > inter lama, maka tingkatkan jumlah klaster (k = k + 1) dan ulangi proses dari Langkah 1. Jika tidak, lanjutkan ke Langkah 10.
10. Hentikan proses (STOP) dan gunakan jumlah klaster terakhir sebagai jumlah klaster optimal

D. Evaluasi dan Analisis akhir

Pada tahap ini, peneliti mengevaluasi kualitas cluster yang terbentuk dengan mengukur kedekatan antar data dalam cluster serta jarak antar cluster. Analisis dilakukan menggunakan *Sum Square Error* (SSE) dan *Silhouette Coefficient Score* (SCS) untuk menentukan efektivitas segmentasi.

Dalam metode SSE, setiap titik data dalam cluster dihitung berdasarkan jaraknya terhadap centroid cluster. Nilai ini kemudian dikuadratkan dan dijumlahkan untuk mendapatkan total SSE. Proses ini dilakukan untuk setiap cluster dan digabungkan sehingga setiap nilai K memiliki SSE yang terpisah. Nilai K yang optimal adalah yang memiliki nilai SSE terendah, menandakan bahwa data dalam cluster lebih terpusat.

Selain itu, perlu dicari titik keseimbangan, agar jumlah cluster yang terbentuk tidak terlalu banyak atau terlalu sedikit, sehingga tetap representatif terhadap pola data

Tahap selanjutnya Analisa hasil klaster yang terbentuk agar dapat memudahkan Pengambilan keputusan pada pihak perusahaan. Klaster yang telah dihasilkan tersebut akan dikalikan dengan bobot nilai LRFM dengan menggunakan metode Pembobotan. Penelitian ini menggunakan bobot dengan nilai WL, WR, WF, WM yaitu 0,238, 0,088, 0,326, dan 0,348. Dari data yang telah menjadi hasil segmentasi kemudian akan dihitung nilai rata-rata dari setiap kolom atau features yang ada.[17].

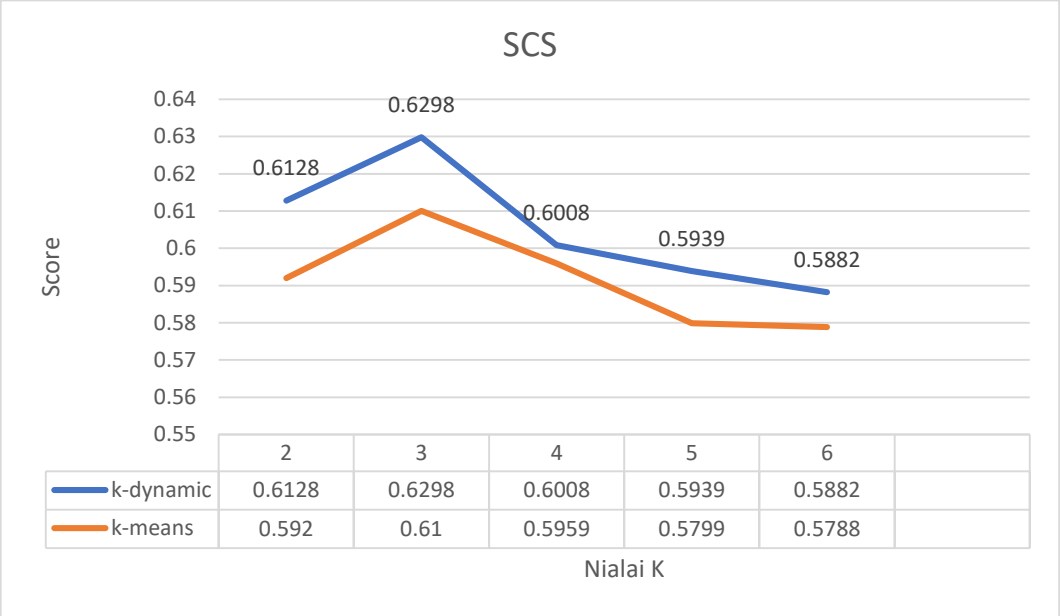
III. HASIL DAN PEMBAHASAN

A. Hasil

Dalam penelitian ini, dilakukan dua pendekatan algoritma untuk segmentasi pelanggan berdasarkan model LRFM, yaitu LRFM + K-Means dan LRFM + K-Means Dinamis. Kedua algoritma tersebut diuji untuk mengevaluasi kinerja segmentasi pelanggan berdasarkan variabel *Length*, *Recency*, *Frequency*, dan *Monetary*, serta untuk melihat pengaruh teknik inisialisasi dan pembaruan pusat klaster terhadap kualitas pembentukan segmen.

Pengujian kinerja masing-masing algoritma dievaluasi menggunakan *Silhouette Coefficient* sebagai metrik validasi internal. *Silhouette Coefficient* menilai sejauh mana setiap data sesuai dalam klaster dan seberapa baik data tersebut terpisah dari klaster lainnya. Semakin mendekati 1, semakin baik kualitas klasterisasi.

Hasil pengujian menunjukkan bahwa pendekatan K-Means Dinamis memberikan nilai Silhouette Coefficient yang lebih tinggi dibandingkan K-Means tradisional. Dengan pembaruan jumlah klaster secara iteratif dan penggunaan jarak intra-klaster maupun inter-klaster sebagai kriteria pembentukan klaster baru, K-Means Dinamis lebih adaptif dalam menemukan struktur klaster yang optimal, terutama untuk data pelanggan yang memiliki karakteristik transaksi beragam. Selain itu, penghitungan jarak intra- dan inter-klaster secara lebih terkendali membuat pembentukan klaster menjadi lebih stabil dan lebih sesuai untuk segmentasi pelanggan.



Gambar 3.Visualisasi Silhouette Coefficient Score

Gambar 3 memperlihatkan perbandingan nilai *Silhouette Coefficient Score* (SCS) untuk algoritma K-Means dan K-Means Dinamis pada variasi jumlah klaster (K). Terlihat bahwa K-Means Dinamis secara konsisten menunjukkan nilai SCS yang lebih tinggi dibandingkan K-Means tradisional. Peningkatan tertinggi terjadi pada jumlah klaster sama dengan 3, di mana K-Means Dinamis mencapai skor SCS sebesar 0,6298, lebih baik dibandingkan skor SCS K-Means tradisional sebesar 0,61. Selain itu, nilai SCS untuk algoritma K-Means Dinamis menunjukkan tren yang lebih stabil dan tetap unggul hingga jumlah klaster meningkat. Dengan demikian, dapat disimpulkan bahwa K-Means Dinamis mampu memberikan kualitas klasterisasi yang lebih baik dan lebih sesuai untuk segmentasi pelanggan berdasarkan metrik LRFM.

B. Analisa hasil

Setelah ditetapkan bahwa algoritma K-Means Dinamis lebih baik berdasarkan evaluasi *Silhouette Coefficient*, tahap selanjutnya adalah melakukan klastering dengan K-Means Dinamis menggunakan K=3. Setelah dilakukan segmentasi pelanggan menggunakan model LRFM dengan algoritma K-Means Dinamis, pelanggan berhasil dikelompokkan ke dalam tiga klaster berbeda. Evaluasi kualitas klaster dengan metode *Silhouette Coefficient* menunjukkan bahwa algoritma ini mampu menghasilkan segmentasi yang baik dan terpisah dengan jelas. Karakteristik masing-masing segmen kemudian dianalisis berdasarkan nilai rata-rata variabel *Length* (L), *Recency* (R), *Frequency* (F), dan *Monetary* (M) untuk setiap klaster, seperti disajikan pada Tabel 6.

TABEL 6
HASIL RATA-RATA KLASTER

Klaster	Populasi	Length (L)	Recency (R)	Frequency (F)	Monetary (M)
0	620	0,0459	0,9444	0,1472	0,0848
1	169	0,8647	0,3789	0,4854	0,2951
2	169	0,2027	0,3495	0,1717	0,0943

Dilakukan proses profil pelanggan dari klaster yang terbentuk dengan penjelasan sebagai berikut:

Klaster 0 merupakan segmen pelanggan terbesar, dengan populasi sebanyak 620 individu. Pelanggan pada klaster ini memiliki nilai *Length* yang sangat rendah, yaitu 0,0459, yang menunjukkan bahwa masa hubungan mereka dengan perusahaan relatif singkat atau aktivitas bertransaksi sangat minim. Nilai *Recency* yang tinggi (0,9444) mengindikasikan bahwa pelanggan dalam klaster ini sudah lama tidak melakukan transaksi terakhir. Selain itu, *Frequensi* pembelian (0,1472) dan nilai *Monetary* (0,0848) juga tergolong rendah, sehingga menunjukkan keterlibatan dan kontribusi transaksi yang minimal. Berdasarkan karakteristik tersebut, klaster 0 dapat dikategorikan sebagai pelanggan pasif atau berisiko churn, sehingga memerlukan strategi reaktivasi untuk meningkatkan loyalitas dan keterlibatan mereka.

Klaster 1 terdiri dari 169 pelanggan dengan karakteristik *Length* yang tinggi, yaitu 0,8647, yang mencerminkan hubungan jangka panjang dengan perusahaan. Nilai *Recency* yang relatif rendah (0,3789) menunjukkan bahwa transaksi terakhir pelanggan terjadi baru-baru ini. Selain itu, *Frequency* pembelian (0,4854) dan nilai *Monetary* (0,2951) merupakan yang tertinggi dibandingkan klaster lainnya, sehingga mengindikasikan

tingkat aktivitas dan kontribusi pendapatan yang signifikan. Dengan demikian, klaster 1 dapat diklasifikasikan sebagai pelanggan loyal dan bernilai tinggi, yang sebaiknya menjadi fokus utama dalam strategi retensi dan pengembangan program loyalitas perusahaan

Klaster 2 memiliki jumlah pelanggan yang sama dengan klaster 1, yaitu 169 individu. Nilai *Length* (0,2027) dan *Recency* (0,3495) berada pada level sedang, sedangkan *Frequency* pembelian (0,1717) dan nilai *Monetary* (0,0943) tergolong rendah. Klaster ini merepresentasikan pelanggan baru atau kurang aktif, tetapi memiliki potensi untuk dikembangkan lebih lanjut. Oleh karena itu, perusahaan perlu merancang strategi pemasaran yang lebih personal dan tepat sasaran untuk meningkatkan tingkat pembelian, keterlibatan, serta nilai transaksi dari segmen ini.

TABEL 7
PROFIL KLASTER

Klaster	Populasi	Length (L)	Recency (R)	Frequency (F)	Monetary (M)	Karaktersitik
0	620	0,0459	0,9444	0,1472	0,0848	Lost customer
1	169	0,8647	0,3789	0,4854	0,2951	Loyal customer
2	169	0,2027	0,3495	0,1717	0,0943	New customer

Secara keseluruhan, ketiga klaster tersebut mencerminkan segmen pelanggan dengan perilaku dan nilai yang berbeda, yang dapat menjadi dasar bagi perusahaan dalam merancang strategi pemasaran yang spesifik dan efektif sesuai karakteristik masing-masing segmen. Berdasarkan karakteristik masing-masing segmen, ketiga klaster yang terbentuk diberi label atau alias sebagai berikut:

TABEL 8
ALIAS KLASTER

Klaster	Populasi	Karaktersitik	Label / alias
0	620	Lost customer	Diamond
1	169	Loyal customer	Gold
2	169	New customer	Silver

1. Pada Kategori *Loyal Customer* (*Gold*)

Klaster ini terdiri dari pelanggan paling loyal dan bernilai tinggi, Mereka memberikan keuntungan besar bagi perusahaan melalui pembelian massal yang berulang. Karena pelanggan ini telah menjalin hubungan jangka panjang dengan perusahaan, mereka dikategorikan sebagai pelanggan setia. Oleh karena itu, strategi pemeliharaan pelanggan harus diterapkan guna meningkatkan loyalitas mereka. Pola pendekatan terbaik untuk pelanggan setia meliputi:

1) Mengakui dan menghargai mereka sebagai pelanggan paling berharga bagi perusahaan.
2) Menyelenggarakan seminar atau acara khusus untuk memperkuat hubungan dan meningkatkan keterlibatan mereka.
3) Menyediakan layanan kelas satu, seperti dukungan prioritas dan pengalaman eksklusif.
4) Menyesuaikan dan menawarkan produk berkualitas tinggi yang dirancang khusus untuk memenuhi kebutuhan mereka.

Karena pelanggan setia sangat penting bagi keberlanjutan bisnis, mereka berhak mendapatkan berbagai keuntungan, termasuk:

1) Diskon khusus untuk pembelian produk.
2) Layanan cepat dan berkualitas tinggi, baik sebelum maupun sesudah penjualan.
3) Pemberitahuan produk secara tepat waktu, agar mereka selalu mendapatkan informasi terbaru.
4) Akses mudah dalam berkomunikasi dengan perusahaan, sehingga meningkatkan kenyamanan dan kepuasan mereka.

Dengan strategi yang tepat, perusahaan dapat memastikan bahwa pelanggan setia tetap merasa dihargai, sehingga memperkuat retensi dan hubungan jangka panjang.

2. Pada Kategori *Lost Customer* (*Diamond*)

Kategori pelanggan yang hilang (lost customer) mencakup pelanggan yang belum memiliki hubungan jangka panjang dengan perusahaan serta belum melakukan pembelian produk dalam periode terbaru. Selain itu, frekuensi transaksi dan nilai moneter yang dihasilkan pelanggan dalam klaster ini juga rendah, sehingga mereka dianggap kurang bernilai bagi perusahaan.

Langkah pertama dalam menangani pelanggan ini adalah mengidentifikasi alasan yang menyebabkan mereka memutuskan untuk berhenti berinteraksi dengan perusahaan. Dalam hal ini, sistem CRM yang kuat dapat membantu dengan mengungkap faktor-faktor yang menyebabkan friksi pelanggan serta alasan di balik ketidakpuasan mereka. Untuk mengurangi kemungkinan kehilangan pelanggan lebih lanjut, perlu diterapkan strategi retensi, seperti:

1) Memberikan bonus khusus dan penawaran menarik untuk meningkatkan ketertarikan mereka kembali.
2) Memastikan pengalaman pelanggan tetap positif guna mempertahankan interaksi dengan perusahaan.

Namun, mengingat rendahnya nilai dari klaster ini serta fakta bahwa mereka tidak memberikan kontribusi signifikan terhadap profitabilitas perusahaan, perlu dilakukan evaluasi ekonomi untuk menentukan apakah mengembalikan pelanggan ini masih layak secara finansial atau tidak.

3. Pada Kategori *New Customer* (*Silver*)

Klaster 3 memiliki jumlah pelanggan paling sedikit dan bersifat tidak pasti karena mereka baru mulai bertransaksi dan belum memiliki hubungan jangka panjang dengan perusahaan. Oleh karena itu, mereka dikategorikan sebagai pelanggan baru.

Strategi yang tepat untuk pelanggan dalam klaster ini adalah mempertahankan kerja sama mereka dengan perusahaan, mengingat biaya untuk menarik pelanggan baru jauh lebih tinggi dibandingkan dengan mempertahankan pelanggan yang sudah ada. Selain itu, beberapa pendekatan yang dapat diterapkan meliputi:

1) Mengadakan pertemuan secara berkala untuk membangun hubungan yang lebih erat.
2) Memberikan pemberitahuan produk secara tepat waktu guna meningkatkan keterlibatan pelanggan.
3) Melakukan negosiasi yang konstruktif dan berkelanjutan untuk menjaga kepuasan mereka.
4) Menawarkan diskon khusus sebagai insentif agar mereka tetap loyal.

Strategi lainnya mencakup profilisasi pelanggan, yaitu memahami kebutuhan spesifik mereka dan menawarkan produk yang sesuai.

IV. SIMPULAN

Penelitian ini dilakukan di Esindo Multi Tata menggunakan data transaksi peserta pelatihan dalam rentang waktu 2017 hingga 2022. Pada tahap pengelompokan, algoritma K-Means dibandingkan dengan algoritma K-Means dinamis untuk menentukan metode yang paling sesuai. Berdasarkan hasil evaluasi menggunakan *Silhouette Coefficient Score* (SCS) dan *Sum of Squared Error* (SSE), algoritma K-Means dinamis terbukti lebih baik dan lebih stabil, sehingga dipilih untuk proses klastering. Dengan menggunakan algoritma tersebut, pelanggan berhasil disegmentasikan ke dalam tiga klaster optimal (K=3), yaitu *Gold* (17%), *Diamond* (64%), dan *Silver* (17%).

Selanjutnya, analisis LRFM (*Length*, *Recency*, *Frequency*, dan *Monetary*) digunakan untuk memprofilkan setiap klaster. Hasil analisis menunjukkan bahwa klaster *Gold* merupakan segmen *loyal customers*, klaster *Diamond* merupakan lost customers, dan klaster *Silver* merupakan *new customers*. Dengan memanfaatkan model LRFM dan algoritma K-Means dinamis, perusahaan mampu mengidentifikasi dan memahami karakteristik setiap segmen pelanggan secara lebih akurat. Hasil segmentasi ini memberikan kontribusi praktis dan strategis, khususnya dalam mendukung perusahaan untuk merancang program retensi, personalisasi layanan, program loyalitas, dan strategi pemasaran yang lebih sesuai kebutuhan dan perilaku masing-masing klaster pelanggan. Dengan penerapan hasil penelitian ini, diharapkan kepuasan pelanggan meningkat, retensi lebih terjaga, dan pertumbuhan bisnis berkelanjutan dapat dicapai.

Saran untuk Pengembangan Lanjutan, Meskipun penelitian ini sudah mampu memberikan segmentasi pelanggan secara jelas dan terukur, terdapat beberapa peluang pengembangan untuk penelitian selanjutnya. Pertama, proses implementasi hasil segmentasi dapat diotomatisasi agar lebih efisien dan terintegrasi langsung ke dalam sistem internal perusahaan, sehingga meminimalkan intervensi manual. Kedua, model segmentasi bisa dikembangkan lebih lanjut dengan melibatkan variabel lain, seperti data perilaku pelanggan (misalnya tingkat kepuasan, partisipasi dalam program promosi, lokasi), untuk meningkatkan akurasi dan kedalaman profil segmen. Ketiga, pengujian model segmentasi ini juga bisa diterapkan pada jenis bisnis lain agar relevansinya lebih luas dan mampu memberikan perbandingan efektivitas di berbagai industri. Selain itu, penggunaan metode klastering alternatif (misalnya DBSCAN, *Gaussian Mixture Model*, *hierarchical clustering*) dan evaluasi jangka panjang terhadap stabilitas klaster juga menjadi topik menarik untuk pengembangan riset berikutnya.

DAFTAR PUSTAKA

[1] A. Dhini, L. R. Budiani, and E. Laoh, "Segmenting and Targeting the Potential Markets of a Muslim Fashion Company," in *7th International Conference on ICT for Smart Society: AIoT for Smart Society, ICISS 2020 - Proceeding*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/ICISS50791.2020.9307604.

[2] M. Alvandi, S. Fazli, and F. S. Abdoli, "K-Mean Clustering Method For Analysis *Customer Lifetime Value* With LRFM Relationship Model In Banking Services," *International Research Journal of Applied and Basic Sciences*, 2012, [Online]. Available: www.irjabs.com

[3] N. Fitrianti Fahrudin and R. Rindiyani, "Comparison of K-Medoids and K-Means Algorithms in Segmenting Customers based on RFM Criteria," *E3S Web of Conferences*, vol. 484, p. 02008, Feb. 2024, doi: 10.1051/e3sconf/202448402008.

[4] K. Auliasari and M. Kertaningtyas, "Penerapan Algoritma K-Means untuk Segmentasi Konsumen Menggunakan R," *Jurnal Teknologi dan Manajemen Informatika*, vol. 5, no. 2, Dec. 2019, doi: 10.26905/jtmi.v5i2.3644.

[5] Y. P. Anggodo, W. Cahyaningrum, A. N. Fauziyah, I. L. Khoiriyah, O. Kartikasari, and I. Cholissodin, "Hybrid K-means Dan Particle Swarm Optimization Untuk Clustering Nasabah Kredit," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 2, p. 104, May 2017, doi: 10.25126/jtiik.201742303.

[6] M. Aryuni, E. Didik Madyatmadja, and E. Miranda, "Customer Segmentation in XYZ Bank Using K-Means and K-Medoids Clustering," in *2018 International Conference on Information Management and Technology (ICIMTech)*, IEEE, Sep. 2018, pp. 412–416. doi: 10.1109/ICIMTech.2018.8528086.

[7] Z.-J. Lee, C.-Y. Lee, L.-Y. Chang, and N. Sano, "Clustering and Classification Based on Distributed Automatic Feature Engineering for Customer Segmentation," *Symmetry (Basel)*, vol. 13, no. 9, p. 1557, Aug. 2021, doi: 10.3390/sym13091557.

[8] M. Tavakoli, M. Molavi, V. Masoumi, M. Mobini, S. Etemad, and R. Rahmani, "Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study," in *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*, IEEE, Oct. 2018, pp. 119–126. doi: 10.1109/ICEBE.2018.00027.

[9] J. Zhou, J. Wei, and B. Xu, "Customer segmentation by web content mining," *Journal of Retailing and Consumer Services*, vol. 61, p. 102588, Jul. 2021, doi: 10.1016/j.jretconser.2021.102588.

[10] A. Solichin and G. Wibowo, "Customer Segmentation Based on Recency Frequency Monetary (RFM) and User Event Tracking (UET) Using K-Means Algorithm," in *2022 IEEE 8th Information Technology International Seminar (ITIS)*, IEEE, Oct. 2022, pp. 257–262. doi: 10.1109/ITIS57155.2022.10009981.

[11] S. Peker, A. Kocyigit, and P. E. Eren, "LRFMP model for customer segmentation in the grocery retail industry: a case study," *Marketing Intelligence & Planning*, vol. 35, no. 4, pp. 544–559, May 2017, doi: 10.1108/MIP-11-2016-0210.

[12] H. Astuti, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Untuk Pengelompokkan Data Pelanggan (Studi Kasus : PT. Pinus Merah Abadi)," 2021.

[13] Md. Z. Hossain, Md. N. Akhtar, R. B. Ahmad, and M. Rahman, "A dynamic K-means clustering for data mining," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 2, p. 521, Feb. 2019, doi: 10.11591/ijeecs.v13.i2.pp521-526.

[14] M. R. K. Ibrahim and R. Tyasnurita, "LRFM Model Analysis for Customer Segmentation Using K-Means Clustering," in *2022 International Conference on Electrical and Information Technology (IEIT)*, IEEE, Sep. 2022, pp. 383–391. doi: 10.1109/IEIT56384.2022.9967896.

[15] H. Soetanto, Rusdah, Painem, and Ameliana, "Reseller Segmentation for a Strategy to Increase The Purchase Deposit of a Telecommunications Company Using The LRFM and K-Means Algorithm," in *2023 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, IEEE, Dec. 2023, pp. 180–185. doi: 10.1109/ISRITI60336.2023.10467378.

[16] H. P. Kurniawan and L. Farhatuaini, "Identifikasi Pola Kepuasan Mahasiswa Terhadap Proses Pembelajaran Menggunakan Algoritma K-Means Clustering.," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 9, no. 2, pp. 164–172, Aug. 2024, doi: 10.30591/jpit.v9i2.6740.

[17] E. Bakhshizadeh, H. Aliasghari, R. Noorossana, and R. Ghousi, "Customer Clustering Based on Factors of *Customer Lifetime Value* with Data Mining Technique (Case Study: Software Industry)," *International Journal of Industrial Engineering and Production Research*, vol. 33, no. 1, Mar. 2022, doi: 10.22068/ijiepr.33.1.1.