

Implementasi *K-Means* dalam Segmentasi Pelanggan Usaha Aluminium dan Kaca Berdasarkan Perilaku Pembelian

<http://dx.doi.org/10.28932/jutisi.v11i2.9533>

Riwayat Artikel

Received: 23 Juli 2024 | Final Revision: 23 Juli 2025 | Accepted: 23 Juli 2025

Creative Commons License 4.0 (CC BY – NC)



Salsabilla Ramadhani^{✉#1}, Priza Pandunata^{#2}, Fajrin Nurman Arifin^{#3}

[#] Program Studi Sistem Informasi, Universitas Jember

Jl. Kalimantan Tegalboto No.37, Kecamatan Sumbersari, Kabupaten Jember, 68121, Indonesia

¹salsa.ramadhani512@gmail.com

²priza@unej.ac.id

³fajrin.pssi@unej.ac.id

✉Corresponding author: salsa.ramadhani512@gmail.com

Abstrak— Mulia Jasa Aluminium dan Kaca adalah Usaha di sektor ritel dan jasa, yang menawarkan bahan dan layanan aluminium dan kaca untuk pembuatan, pemasangan, dan perbaikan. Saat ini, persaingan usaha pada bidang tersebut cukup ketat, sehingga pemilik mengaku kesulitan meningkatkan penjualannya. Oleh sebab itu, pemilik usaha perlu menerapkan strategi pemasaran dan pelayanan untuk meningkatkan penjualannya. Namun, banyaknya pelanggan dengan karakteristik dan perilaku yang berbeda membuat sulitnya penetapan strategi pemasaran maupun pelayanan. Sehingga, pada penelitian ini dilakukan segmentasi pelanggan berdasarkan perilaku pembelian. Hal tersebut bertujuan mengetahui perilaku serta loyalitas pelanggan menggunakan data laporan penjualan usaha tersebut. Variabel yang digunakan untuk menilai seberapa bernilai seorang pelanggan adalah variabel *Length*, *Recency*, *Frequency* dan *Monetary* atau yang dikenal dengan model *LRFM*. Variabel-variabel tersebut dikelompokkan menggunakan algoritma pengelompokan *K-Means*. Tujuan penelitian ini adalah untuk mengelompokkan pelanggan berdasarkan perilaku pembelian mereka, sehingga membantu bisnis dalam menyusun strategi pemasaran dan layanan yang lebih efektif, meningkatkan kepuasan pelanggan, dan pada akhirnya meningkatkan penjualan dan loyalitas. Dengan menggunakan metode *Silhouette* untuk menentukan jumlah cluster optimal, tiga kelompok pelanggan diidentifikasi, dengan nilai koefisien tertinggi sebesar 0,663063. *Cluster 0* adalah “*Lost Customer Group*”, *Cluster 1* adalah “*New Customer Group*” dan *Cluster 2* adalah “*Core Customer Group*”.

Kata kunci— *K-Means*; Model *LRFM*; Segmentasi Pelanggan; *Silhouette*.

K-Means Implementation for Customer Segmentation in Aluminium and Glass Business by Purchasing Behavior

Abstract— *Mulia Jasa Aluminium dan Kaca is a business in the retail and service sector, offering Aluminium and glass materials and services for manufacturing, installation, and repair. Currently, competition in this field is quite intense, leading the business owner to admit difficulties in increasing sales. Therefore, the business owner needs to implement marketing and service strategies to boost sales. However, the diversity of customers with varying characteristics and behaviors makes it challenging to establish effective marketing and service strategies. Thus, this study conducts customer segmentation based on purchasing behavior. The aim is to understand customer behavior and loyalty using sales report data from the business. The variables used to assess a customer's value are Length, Recency, Frequency, and Monetary (LRFM). These variables are grouped using the K-means clustering algorithm. The objective of*

this study is to group customers based on their purchasing behavior, thereby assisting the business in developing more effective marketing and service strategies, enhancing customer satisfaction, and ultimately increasing sales and loyalty. Using the Silhouette method to determine the optimal number of clusters, three customer groups were identified, with the highest coefficient value of 0.663063. Cluster 0 is the “Lost Customer Group”, Cluster 1 is the “New Customer Group”, and Cluster 2 is the “Core Customer Group”.

Keywords— Customer Segmentation; K-Means; LRFM Model; Silhouette.

I. PENDAHULUAN

Usaha dalam bisnis adalah kegiatan yang dilakukan untuk menghasilkan barang atau jasa dengan tujuan mendapatkan keuntungan. Berdasarkan skala, usaha dapat bervariasi, salah satunya adalah Usaha Mikro, Kecil, dan Menengah atau UMKM. UMKM adalah usaha produktif milik badan usaha atau perorangan yang telah memenuhi kriteria tertentu sesuai dengan UU Republik Indonesia No.20 Tahun 2008 Tentang UMKM [1]. UMKM memiliki peran dalam pertumbuhan ekonomi di Indonesia. Berdasarkan data Kementerian Koperasi dan UKM, pada tahun 2023 jumlah pelaku usaha UMKM mencapai 66 juta. UMKM juga berkontribusi dalam Pendapatan Domestik Bruto (PDB) sebesar 61%. Selain itu, UMKM membantu sekitar 117 juta orang mendapatkan lapangan pekerjaan [2].

Salah satu UMKM yang memiliki potensi adalah Usaha Aluminium dan Kaca Mulia Jasa yang bergerak di bidang ritel serta jasa produk aluminium dan kaca. Usaha tersebut fokus menjual bahan serta menyediakan jasa membuat, memasang hingga servis produk aluminium dan kaca. Namun berdasarkan wawancara bersama pemilik usaha, persaingan usaha dalam bidang tersebut sangat ketat. Pada wawancara tersebut, pemilik mengungkapkan bahwa terdapat lebih dari 10 kompetitor di kecamatan yang sama. Selain itu, usaha tersebut saat ini sedang dalam masa pemulihan pasca pandemi *Covid-19*. Dengan adanya persaingan usaha yang ketat, pemilik mengaku kesulitan meningkatkan penjualan. Berdasarkan penelitian sebelumnya, untuk mempertahankan bisnis, diperlukan strategi pemasaran dan pelayanan yang tepat agar penjualan dapat meningkat [3]. Perusahaan perlu memahami setiap pelanggan dari segi kebutuhan maupun perilaku untuk bisa memberikan pelayanan yang maksimal [4].

Banyaknya pelanggan yang memiliki perbedaan karakteristik dan perilaku pada usaha Mulia Jasa membuat sulitnya penentuan strategi pemasaran dan pelayanan. Oleh sebab itu, penting untuk mengenali dan memahami kelompok pelanggan agar strategi pemasaran dan pelayanan yang dilakukan lebih tepat sasaran [5]. Dalam upaya mengenali kelompok pelanggan, salah satu caranya adalah dengan memanfaatkan histori data penjualan dan pelanggan. Pelanggan merupakan aset yang dapat dimanfaatkan perusahaan untuk mempertahankan bisnis. Segmentasi pelanggan dilakukan untuk mengenali kelompok pelanggan tersebut, sehingga strategi pemasaran dan pelayanan yang diberikan lebih maksimal untuk meningkatkan penjualan [6]. Segmentasi pelanggan yang efektif memerlukan pemahaman mendalam tentang perilaku pembelian pelanggan.

Salah satu model yang dapat digunakan untuk segmentasi pelanggan adalah model *LRFM*. Model *LRFM* adalah model yang dimanfaatkan untuk menilai seberapa bernilai seorang pelanggan berdasarkan riwayat transaksi. *LRFM* merupakan singkatan dari 4 variabel yang digunakan yaitu *Length*, *Recency*, *Frequency* dan *Monetary*. Model *LRFM* digunakan untuk mengetahui perilaku pembelian pelanggan yang dilihat dari nilai 4 variabel tersebut [4]. *LRFM* akan mengukur nilai perilaku pelanggan berdasarkan panjang waktu hubungan dengan usaha, waktu sejak transaksi terakhir, frekuensi transaksi, dan total pengeluaran pelanggan. Dibandingkan dengan model *RFM*, Model *LRFM* dipilih karena adanya tambahan variabel *Length* untuk mengetahui durasi hubungan yang telah terjalin antara usaha dengan pelanggan [4]. Pada penelitian sebelumnya, model *LRFM* dapat digunakan untuk mengelompokkan pelanggan UKM server pulsa all operator AR-Pulsa Biz Malang ke dalam 2 segmen pelanggan [7]. Pada penelitian lain, hasil model *LRFM* memiliki dampak yang baik terhadap pertumbuhan penjualan [8]. Setelah mendapatkan hasil model *LRFM*, dapat dilanjutkan ke tahap pengelompokan atau *Clustering*.

Clustering merupakan metode mengelompokkan data dari himpunan data yang tidak diketahui kelompok atau kelasnya [9]. Data akan dikelompokkan menjadi beberapa *Cluster* atau kelas yang memiliki kemiripan karakteristik dengan data lain dan memiliki perbedaan karakteristik dengan *Cluster* lain. Salah satu algoritma yang dapat digunakan dalam *Clustering* adalah algoritma *K-Means*. Algoritma *K-Means* merupakan metode *Clustering* partisi yang banyak digunakan dalam pemasaran karena algoritma tersebut sederhana [10]. Algoritma *K-Means* digunakan karena dapat membantu menentukan kelompok pelanggan usaha aluminium dan kaca Mulia Jasa berdasarkan nilai variabel *LRFM* [4]. Algoritma *K-Means* dinilai memiliki kinerja yang baik dibandingkan dengan algoritma *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*) dan *Optics algorithms* dalam melakukan segmentasi pelanggan pada penelitian sebelumnya [8]. Penelitian serupa juga dilakukan dan menunjukkan hasil bahwa algoritma *K-Means* bekerja lebih baik dibandingkan algoritma *K-Medoids*, dan algoritma *Fuzzy C-Means* dalam pengelompokan data pelanggan dengan model *LRFM* [11]. Tantangan dalam proses algoritma *K-Means* adalah penentuan jumlah *Cluster* optimal agar hasil *Clustering* lebih akurat. Oleh karena itu diperlukan evaluasi dalam penentuan pusat *Cluster* paling optimal. Salah satu metode dalam penentuan jumlah *Cluster* paling optimal adalah metode Metode *Silhouette*. Dalam penelitian sebelumnya, model *K-Means* dengan menggunakan metode *Silhouette*

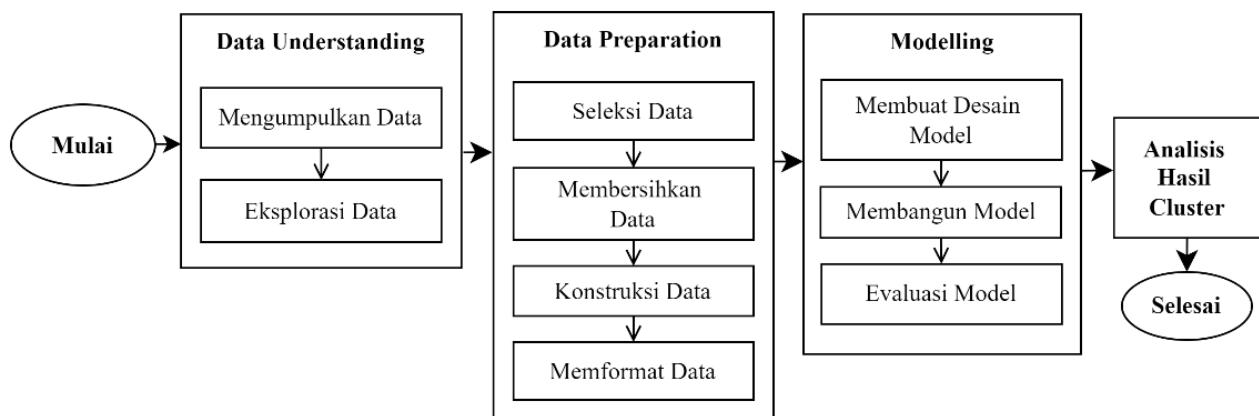
memiliki performa lebih baik jika dibandingkan dengan *Elbow* dalam *Clustering* dengan nilai DBI pada metode *Silhouette* yang lebih rendah dibanding *Elbow* [12].

Penelitian sebelumnya yang digunakan sebagai acuan berjudul “*Business Intelligence Using Machine Learning Algorithms*” yang dilakukan oleh Hamzehi dan Hosseini pada tahun 2022 [8], membahas pengenalan model yang efisien untuk mengoptimalkan penjualan produk pada perusahaan. Penelitian selanjutnya berjudul “*An LRFM Model to Analyze Outpatient Loyalty From a Medical Center in Taiwan*” yang dilakukan oleh Chao et al pada tahun 2021 [13] membahas tentang model *LRFM* dan algoritma *Clustering* dalam pengelompokan pelanggan sebuah pusat medis. Terdapat penelitian lain dengan judul “*RFM Model for Customer Purchase Behavior Using K-Means Algorithm*” yang dilakukan oleh Anitha dan Patil pada tahun 2019 [14], membahas bagaimana implementasi model *RFM* dan metode *Clustering* menggunakan algoritma *K-Means* untuk mengidentifikasi pelanggan potensial di industri ritel dan dievaluasi menggunakan metode *Silhouette*.

Berdasarkan hasil analisis yang dilakukan, segmentasi pelanggan menggunakan integrasi model *LRFM* dan algoritma *K-Means* pada sektor usaha aluminium dan kaca masih jarang dijadikan fokus dalam penelitian sebelumnya sehingga akan dilakukan segmentasi pelanggan dengan memanfaatkan model *LRFM* dan algoritma *K-Means* pada usaha aluminium dan kaca Mulia Jasa untuk membantu usaha tersebut menentukan kelompok pelanggan dengan memanfaatkan data penjualan dan pelanggan agar mengetahui perilaku pembelian pelanggannya. Dengan pemahaman yang lebih baik tentang pelanggan, diharapkan dapat membantu usaha dalam menentukan strategi pemasaran dan pelayanan yang lebih efektif, meningkatkan kepuasan pelanggan, dan pada akhirnya, meningkatkan penjualan serta loyalitas pelanggan.

II. METODE PENELITIAN

Tahapan penelitian yang dilakukan sejak awal penelitian dapat dilihat pada gambar 1.



Gambar 1 Tahapan Penelitian

A. Data Understanding

Berikut tahapan pada proses *Data Understanding* :

1. *Mengumpulkan data*: Data yang dikumpulkan untuk penelitian ini didapatkan dari Toko Aluminium dan Kaca Mulia Jasa. Data yang diperoleh adalah data laporan penjualan dari tahun 2019 hingga 2023.
2. *Eksplorasi data*: Pada tahap eksplorasi data, dijelaskan mengenai informasi data yang didapatkan. Data tersebut merupakan data laporan penjualan usaha aluminium dan kaca terdiri dari 539 baris dan 9 kolom atribut berisi informasi penjualan dan pelanggan.

B. Data Preparation

Berikut tahapan pada proses *Data Preparation* :

1. *Seleksi data*: Pada tahap seleksi data, dipilih atribut data yang relevan dan berguna dalam penelitian. Dalam penelitian ini, atribut dari data laporan penjualan yang digunakan adalah atribut No, Tanggal, Total dan Telepon. Atribut No dipilih karena berisi nomor transaksi yang digunakan untuk memproses variabel *Frequency*, atribut Tanggal merupakan tanggal terjadinya transaksi digunakan untuk memproses variabel *Length* dan *Recency*. Atribut Total digunakan untuk memproses

nilai *Monetary* dan atribut Telepon digunakan untuk mengidentifikasi pelanggan karena atribut tersebut paling unik dibandingkan atribut yang lain dalam mengidentifikasi pelanggan.

2. *Membersihkan data*: Tahap pembersihan data dilakukan pengecekan *missing value*, mengoreksi kesalahan, menghapus data duplikat, dan menghapus atribut yang tidak relevan.
3. *Konstruksi data*: Tahap ini merupakan tahap persiapan data konstruktif seperti membuat atribut turunan. Pada tahap konstruksi data, dibangun atribut Id pelanggan yang dibentuk dari atribut Telepon dan dibangun juga atribut model *LRFM* yaitu atribut *Length*, *Recency*, *Frequency*, dan *Monetary*. Nilai *Length* diambil dari selisih transaksi penjualan yang diambil dari tanggal maksimum atau terakhir dan tanggal minimum atau awal pelanggan melakukan transaksi. Nilai *Recency* diambil dari selisih tanggal acuan analisis yaitu 31 Desember 2023 dengan tanggal maksimum atau terakhir pelanggan melakukan transaksi. Nilai *Frequency* merupakan jumlah transaksi yang dilakukan pelanggan diambil dari atribut 'Nomor Transaksi'. Lalu Nilai *Monetary* diambil dari total pengeluaran pelanggan dari atribut 'Total'. Sehingga, variabel yang digunakan dalam tahap selanjutnya adalah Id Pelanggan, *Length*, *Recency*, *Frequency*, dan *Monetary* dengan tipe data numerik.
4. *Memformat data*: Pada tahap ini dilakukan transformasi format data dengan normalisasi terhadap data hasil model *LRFM*. Normalisasi membuat rentang nilai yang sama pada beberapa variabel agar data yang memiliki dimensi berbeda tidak mengganggu hasil penelitian. Normalisasi yang dilakukan menggunakan *Min-max Scaler* dengan menggunakan persamaan 1 [15].

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

x' = hasil nilai normalisasi.

x = nilai data yang dinormalisasikan.

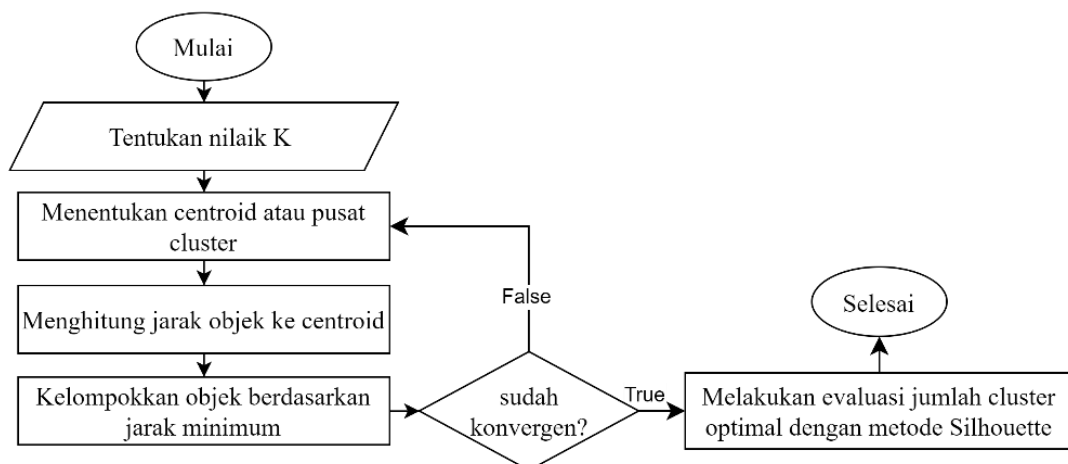
x_{\min} = nilai minimum data pada variabel yang dinormalisasi.

x_{\max} = nilai maksimum data pada variabel yang dinormalisasi.

C. Modelling

Berikut tahapan pada proses Modelling :

1. *Membuat desain model*: Membuat desain model menggunakan algoritma *K-Means* untuk mengelompokkan pelanggan berdasarkan nilai *LRFM* hingga tahap evaluasi. Desain model tersebut terdapat pada gambar 2.



Gambar 2. Desain Model

Berdasarkan gambar 2, penjelasan alur desain model tersebut adalah sebagai berikut:

- a. Pertama menentukan jumlah *cluster* (K) yang akan digunakan yaitu mulai percobaan jumlah *cluster* $K=2$ hingga $K=10$.
- b. Lalu menentukan *centroid* awal atau pusat *cluster* secara acak.
- c. Menghitung jarak objek ke *centroid* menggunakan rumus persamaan *Euclidean Distance* sesuai persamaan 2 berikut [16].

$$d(x_i, y_j) = \left(\sum (x_i - y_j)^2 \right)^{1/2} \quad (2)$$

$d(x_i, y_j)$ = jarak *Euclidean*.

x_i = jarak objek ke- i pada variabel ke- j .
 y_j = nilai *centroid* pada *Cluster* ke- j .

- d. Kelompokkan objek berdasarkan jarak minimum objek dengan *centroid*.
e. Hitung *centroid* baru dengan rumus rata-rata semua titik data pada *cluster* sesuai dengan persamaan 3 [16].

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j \quad (3)$$

$\mu_j(t+1)$ = *Centroid* baru ke $(t+1)$.

N_{sj} = Jumlah data pada *Cluster* s_j .

Jika *centroid* sudah tidak berubah atau konvergen, maka dapat dilanjutkan ke tahap evaluasi. Namun, jika *centroid* masih berubah atau masih belum konvergen ulangi tahap 2 hingga 4 sampai hasil *cluster* konvergen.

- f. Melakukan evaluasi jumlah *cluster* optimal dengan menggunakan metode *Silhouette*.
2. *Membangun model*: Selanjutnya adalah tahap membangun model. Tahap tersebut menjalankan desain model yang sudah siapkan pada dataset laporan penjualan usaha aluminium dan kaca Mulia Jasa.
3. *Evaluasi model*: Tahap ini, dilakukan evaluasi hasil *Clustering* dengan mencari nilai k paling optimal. Metode yang akan digunakan dalam mengevaluasi hasil *cluster* dan menentukan nilai k optimal adalah metode *Silhouette* menggunakan persamaan 4 berikut [10].

$$S_i = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (4)$$

S_i = Nilai koefisien *Silhouette*.

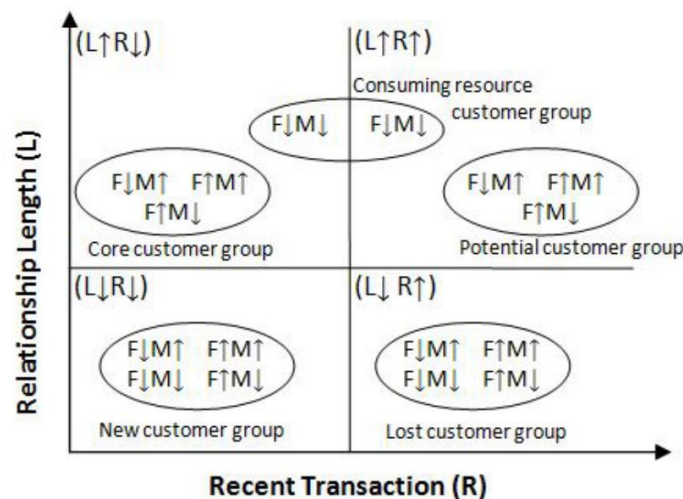
a_i = Rata-rata jarak objek i dengan seluruh objek dalam satu *Cluster*.

b_i = rata-rata jarak dari objek i ke semua objek dalam *cluster* terdekat.

Hasil koefisien *Silhouette* berada pada range -1 hingga 1. Semakin tinggi nilai koefisien *Silhouette*, semakin optimal jumlah *cluster* tersebut. Hal tersebut didasarkan pada kenyataan bahwa *cluster* yang lebih homogen menghasilkan tingkat korelasi yang lebih tinggi di antara objek-objek di dalamnya, sehingga meningkatkan nilai koefisien *Silhouette* [17]. Nilai koefisien *Silhouette* dibagi menjadi 4 kategori yaitu nilai $0.7 < S_i \leq 1$ termasuk dalam kategori *Strong Structure* dimana *cluster* berkelompok dengan sangat baik, Nilai $0.5 < S_i \leq 0.7$ termasuk dalam kategori *Medium Structure* dimana *cluster* cukup jelas tapi masih ada beberapa kemiripan antar *cluster*, Nilai $0.25 < S_i \leq 0.5$ termasuk dalam kategori *Weak Structure* dimana *cluster* yang terbentuk kurang jelas dan Nilai $S_i \leq 0.25$ termasuk dalam kategori *No structure* dimana tidak ada struktur *cluster* yang jelas sehingga menunjukkan hasil *clustering* yang buruk [10].

D. Analisis Hasil Cluster

Analisis hasil segmentasi pelanggan dilakukan untuk mengetahui tipe dan karakteristik pelanggan berdasarkan hasil *cluster*. Untuk mengetahui tipe pelanggan tersebut, dapat dilakukan pemetaan berdasarkan matrik loyalitas pelanggan dari penelitian Chang dan Tsay pada tahun 2004 [18]. Matrik tersebut dapat dilihat pada gambar 3.



Gambar 3 Matrik Loyalitas Pelanggan

Pada gambar 3, matrik loyalitas pelanggan mengklasifikasikan perilaku pelanggan berdasarkan segmen hasil model *LRFM*. Simbol (↑) menunjukkan nilai rata-rata *cluster* yang terbentuk lebih tinggi dari nilai rata-rata keseluruhan. Sebaliknya, simbol Simbol (↓) menunjukkan nilai rata-rata *cluster* yang terbentuk lebih rendah dari nilai rata-rata keseluruhan [19].

III. HASIL DAN PEMBAHASAN

Data yang dikumpulkan merupakan rekap laporan penjualan dari tahun 2019 hingga 2023 dengan 539 baris data dan 9 kolom. Kolom tersebut berisi informasi penjualan dan informasi pelanggan yang terlibat dalam penjualan. Contoh data hasil pengumpulan dapat dilihat pada Tabel 1.

TABEL 1
CONTOH HASIL PENGUMPULAN DATA

No	Tanggal	Nama Barang	Harga	Qty	Total	Pelanggan	Alamat	Telepon
1	5/1/2019	KACA POLOS 5 MM 102 X 203	154,000	2	308,000	I*****L	POJOK CAMPURDARAT	082*****34
		KACA RIBEN 5 MM 122 X 153	156,000	3	468,000			
2	5/1/2019	RAK ROKOK	250,000	1	250,000	B*****I	TANGGUNG	081*****21
3	9/1/2019	ALMARI 2 PINTU KACA BENING	1,500,000	1	1,500,000	A***	MALASAN TRENGGALEK	085*****89
4	11/1/2019	HL 1x11/2 SEDANG	158,000	1	158,000	S*****I	SECANG POJOK	085*****63
		KACA POLOS 5 MM 102 X 203	154,000	3	462,000			
...

Tahap selanjutnya merupakan seleksi data dengan menyeleksi atribut yang akan digunakan dalam pembentukan model *LRFM*. Atribut yang akan digunakan dalam proses model *LRFM* adalah atribut No, Tanggal, Total dan Telepon. Atribut No berisi nomor transaksi digunakan untuk memproses nilai *Frequency*. Atribut Tanggal berisi tanggal transaksi digunakan untuk memproses nilai *Length* dan *Recency*. Atribut Total berisi total pembelian digunakan untuk memproses nilai *Monetary*. Atribut Telepon digunakan karena merupakan atribut yang paling unik untuk mengidentifikasi pelanggan.

Selanjutnya tahap pembersihan data dengan melakukan pengecekan *missing value*. Hasil menunjukkan 152 *missing value* pada atribut No, Tanggal, dan Telepon. Setelah dianalisis, *missing value* tersebut merupakan data transaksi yang jumlah barangnya lebih dari satu. Agar analisis data dapat dilakukan, dalam mengatasi *missing value* tersebut, *value* yang kosong dilengkapi dengan mengambil informasi dari data barang pertama yang dicatat. Selanjutnya melakukan pengecekan data duplikat dan tidak ada data yang duplikat dan menghapus atribut atau kolom yang tidak digunakan. Contoh tampilan data setelah dilakukan pembersihan data dapat dilihat pada Tabel 2.

TABEL 2
CONTOH TAMPILAN HASIL PEMBERSIHAN DATA

No	Tanggal	Total	Telepon
1	2019-01-05	308000	082*****34
1	2019-01-05	468000	082*****34
2	2019-01-05	250000	081*****21
3	2019-01-09	1500000	085*****89
4	2019-01-11	158000	085*****63
...

Pada tahap konstruksi data, dibuat atribut turunan Id pelanggan dari atribut yang paling unik yaitu Telepon untuk menjaga data pribadi yang merupakan privasi pelanggan. Setelah itu memproses atribut *LRFM* dan contoh hasil tahap konstruksi data dapat dilihat pada Tabel 3.

TABEL 3
CONTOH HASIL KONSTRUKSI DATA

Length	Recency	Frequency	Monetary	Id Pelanggan
227	1594	2	500000	1
1412	409	5	1826000	2
527	1290	4	2557000	3
1222	593	7	2312000	4
1374	440	4	4520000	5
...

Pada Tabel 3, nilai masing-masing variabel memiliki dimensi yang berbeda. Nilai *Monetary* memiliki selisih nilai yang cukup jauh dibandingkan variabel yang lain sehingga perlu dibuat rentang nilai yang sama agar data yang memiliki dimensi berbeda tidak mengganggu hasil penelitian. Oleh sebab itu dilakukan tahapan memformat data, dengan cara transformasi format data menggunakan metode normalisasi *Min-max Scaler* terhadap atribut *LRFM* (*Length*, *Recency*, *Frequency* dan *Monetary*). Contoh tampilan hasil normalisasi pada tahap memformat data dapat dilihat Tabel 4.

TABEL 4
CONTOH HASIL NORMALISASI

Length	Recency	Frequency	Monetary	Id Pelanggan
0.132284	0.880110	0.125	0.088423	1
0.822844	0.225414	0.500	0.330173	2
0.307110	0.712155	0.375	0.463446	3
0.712121	0.327072	0.750	0.418778	4
0.800699	0.242541	0.375	0.821331	5
...

Pada Tabel 4 rentang data semua atribut *LRFM* sudah berada di rentang 0 hingga 1 sehingga nilai yang dimiliki data tersebut tidak terlalu jauh. Selanjutnya pada tahap membangun model, data hasil normalisasi dikelompokkan dengan menjalankan desain model pada gambar 2. Algoritma *K-Means* diimplementasikan dengan percobaan pencarian *cluster* dengan jumlah $K=2$ hingga $K=10$. Hasil percobaan *clustering* tersebut dapat dilihat pada Tabel 5 Setiap baris pada Tabel 5 merupakan hasil dari jumlah anggota tiap jumlah *cluster* (K).

TABEL 5
HASIL CLUSTERING DAN JUMLAH ANGGOTA TIAP CLUSTER

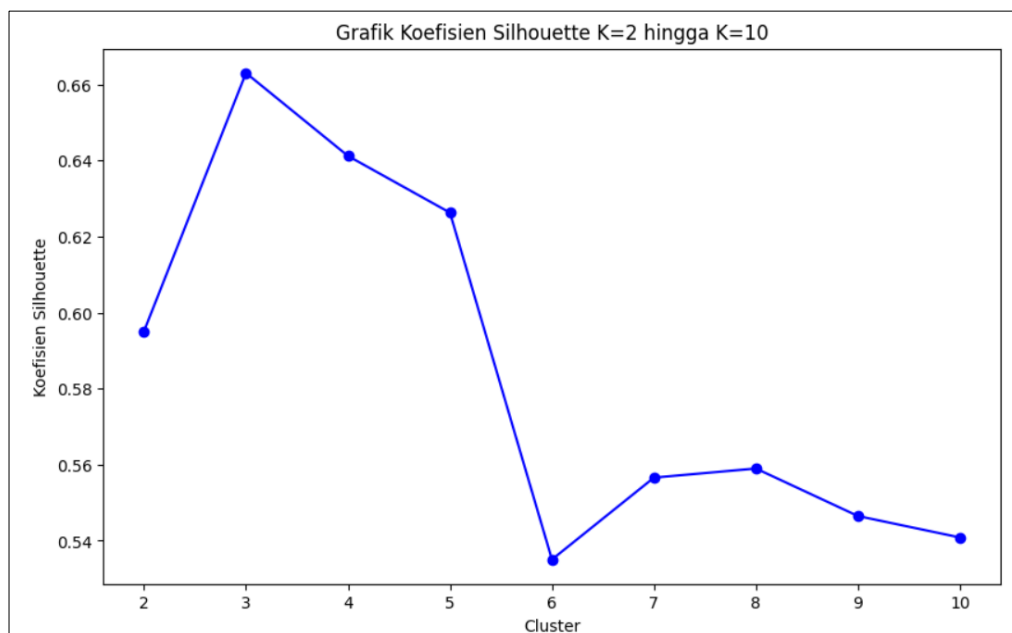
Percobaan	Cluster0	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7	Cluster8	Cluster9
K = 2	132	172	-	-	-	-	-	-	-	-
K = 3	132	144	28	-	-	-	-	-	-	-
K = 4	105	143	28	28	-	-	-	-	-	-
K = 5	28	141	14	105	16	-	-	-	-	-
K = 6	89	27	16	53	105	14	-	-	-	-
K = 7	14	105	82	27	16	13	47	-	-	-
K = 8	27	48	6	105	14	10	80	14	-	-
K = 9	9	93	19	6	14	36	10	14	103	-
K = 10	103	14	9	9	1	14	19	41	88	6

Selanjutnya tahap evaluasi kualitas hasil *Clustering* dengan mencari nilai k paling optimal. Metode yang digunakan dalam mengevaluasi hasil *cluster* dan menentukan nilai k optimal adalah metode *Silhouette*. Jumlah *cluster* paling optimal menghasilkan nilai koefisien *Silhouette* yang paling tinggi, hal tersebut menunjukkan bahwa *cluster* paling homogen atau data dalam *cluster* tersebut paling mirip satu sama lain dan terpisah dengan baik dari *cluster* lain. Hasil perhitungan koefisien di masing-masing percobaan jumlah *cluster* (K) terdapat pada tabel 6 dan grafik perhitungan tersebut dapat dilihat pada gambar 4.

TABEL 6
HASIL PERHITUNGAN KOEFISIEN SILHOUETTE

Jumlah Cluster (K)	Koefisien Silhouette
2	0.595115
3	0.663063
4	0.641235
5	0.626308
6	0.535038

Jumlah Cluster (K)	Koefisien Silhouette
7	0.556587
8	0.558987
9	0.546483
10	0.540837



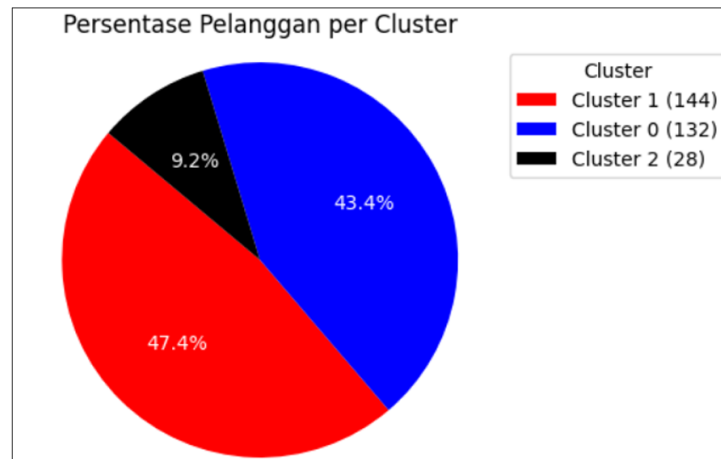
Gambar 4 Grafik Koefisien Silhouette

Berdasarkan hasil perhitungan dan grafik koefisien *Silhouette* pada tabel 6 dan gambar 4, dari percobaan K=2 hingga K=10 dapat diketahui jumlah *cluster* paling optimalnya adalah 3 *cluster* karena pada jumlah *cluster* 3 menunjukkan nilai koefisien *Silhouette* tertinggi dengan nilai koefisien *Silhouette* sebesar 0,663063. Nilai tersebut termasuk ke dalam kategori kualitas *cluster* “Medium structure” dimana nilai koefisien *Silhouette* lebih dari 5 dan kurang dari sama dengan 7. Hasil *cluster* dari metode *silhouette* adalah 3 *cluster* dan contoh data hasil 3 *cluster* dapat dilihat pada Tabel 7.

TABEL 7
HASIL CLUSTER DENGAN K=3

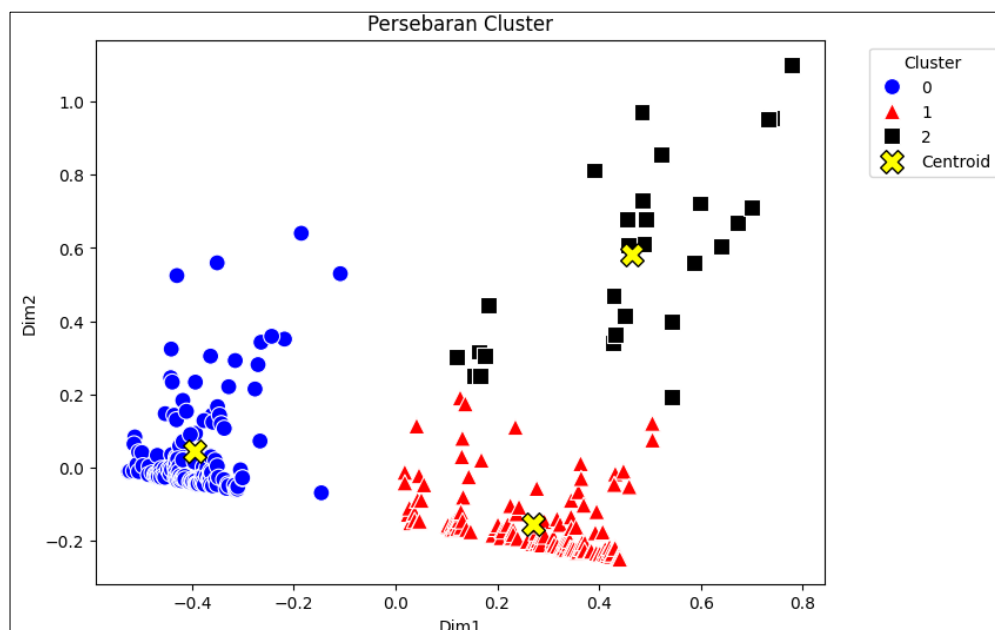
Length	Recency	Frequency	Monetary	Id Pelanggan	Cluster
227	1594	2	500000	1	0
1412	409	5	1826000	2	2
527	1290	4	2557000	3	0
1222	593	7	2312000	4	2
1374	440	4	4520000	5	2
...

Berdasarkan data hasil *cluster* dengan jumlah *cluster* (K) 3 pada Tabel 7, dapat diketahui jumlah data pada masing-masing *cluster*. Hasil persentase jumlah data pelanggan pada masing-masing *cluster* dapat dilihat pada gambar 5.



Gambar 5 Visualisasi persentase pelanggan tiap cluster

Berdasarkan visualisasi persentase pelanggan tiap *cluster* pada gambar, persentase pelanggan dalam *cluster* 0 sebesar 43,4% dengan jumlah data 132 pelanggan. Persentase pelanggan dalam *cluster* 1 sebesar 47,4% dengan jumlah data sebanyak 144 pelanggan dan persentase pelanggan dalam *cluster* 2 sebesar 9,2%. Dengan jumlah data sebanyak 28 pelanggan. Visualisasi persebaran *cluster-cluster* tersebut dapat dilihat pada visualisasi *Scatter plot* Gambar 6.



Gambar 6 Visualisasi persebaran cluster

Berdasarkan visualisasi persebaran *cluster* pada Gambar 6, dapat disimpulkan bahwa dataset terbagi menjadi tiga *cluster* yang terpisah dengan jelas, masing-masing dengan karakteristik dan persebaran yang berbeda serta tidak ada warna yang tumpang tindih. *Cluster* 0 berwarna biru berkelompok di sebelah kiri dengan persebaran yang cukup rapat di sekitar *centroid*. Hal tersebut menunjukkan bahwa pelanggan dalam *cluster* 0 memiliki karakteristik yang mirip. *Cluster* 1 berwarna merah berada di bagian kanan bawah gambar 6, dimana menunjukkan lebih banyak penyebaran dibandingkan *Cluster* 0 tetapi masih cukup rapat di area tertentu. Ini berarti pelanggan dalam *cluster* 1 memiliki beberapa variasi karakteristik, namun masih relatif mirip. *Centroid* berada di tengah titik yang bergerombol pada bagian bawah. *Cluster* 2 berwarna hitam menunjukkan penyebaran yang paling luas diantara 2 *cluster* yang lain dengan *centroid* di bagian tengah *cluster*, artinya *centroid* masih

mewakili *cluster-cluster* tersebut. Pelanggan pada *cluster* tersebut memiliki variasi yang lebih besar. Meskipun *cluster* 1 dan 2 berdekatan, namun *centroid* kedua *cluster* tersebut berjauhan sehingga masih menunjukkan perbedaan.

Analisis hasil *cluster* dilakukan dengan cara mencari nilai rata-rata tiap variabel dalam masing-masing *cluster* lalu dibandingkan dengan nilai keseluruhan pada tiap variabel. Jika rata-rata *cluster* lebih tinggi dibandingkan rata-rata keseluruhan pada tiap variabel, maka akan ditandai dengan simbol (\uparrow) dan jika rata-rata *cluster* lebih rendah dibandingkan rata-rata keseluruhan pada tiap variabel, maka akan ditandai dengan simbol (\downarrow). Setelah mendapat hasil perbandingan rata-rata *cluster* pada masing-masing variabel dengan rata-rata keseluruhan, maka *cluster* tersebut dikategorikan berdasarkan matrix loyalitas pelanggan pada gambar 3 dan dianalisis karakteristik perilaku pembelian pelanggan di setiap *clusternya*. Hasil rata-rata tiap variabel dari masing-masing *cluster* dan rata-rata keseluruhan masing-masing variabel dapat dilihat pada Tabel 8.

TABEL 8
PERBANDINGAN RATA-RATA TIAP VARIABEL LRFM PER *CLUSTER*

Cluster	L	R	F	M	Jumlah
0	0.009461	0.877394	0.013258	0.128836	132
1	0.009927	0.183705	0.006076	0.078355	144
2	0.683920	0.194081	0.276786	0.377354	28
Rata-rata	0.071803	0.485868	0.034128	0.127814	304

Rata-rata keseluruhan dari variabel *Length* adalah 0,071803. Rata-rata keseluruhan variabel *Recency* adalah 0,485868. Rata-rata keseluruhan variabel *Frequency* adalah 0,034128. Rata-rata keseluruhan variabel *Monetary* adalah 0.127814. Selanjutnya rata-rata keseluruhan masing-masing variabel tersebut akan dianalisis dengan cara dibandingkan dengan rata-rata variabel di setiap *clusternya*. Berikut analisis hasil *cluster* yang terbentuk :

1. *Cluster 0*

Cluster 0 terdiri dari 132 pelanggan atau 43,4% dari seluruh pelanggan. Berdasarkan Gambar 6 rata-rata nilai *Length* sebesar 0,009461 lebih rendah dibandingkan rata-rata keseluruhan ($L\downarrow$), rata-rata nilai *Recency* sebesar 0,877394 lebih tinggi dari rata-rata keseluruhan ($R\uparrow$), rata-rata nilai *Frequency* sebesar 0,013258 lebih rendah dari rata-rata keseluruhan ($F\downarrow$) dan rata-rata nilai *Monetary* sebesar 0,128836 lebih tinggi dari rata-rata keseluruhan ($M\uparrow$).

Dengan demikian, hasil perbandingan rata-rata masing-masing variabel pada *cluster 0* adalah ($L\downarrow R\uparrow F\downarrow M\uparrow$) dan berdasarkan matrik loyalitas pelanggan pada gambar 3, *cluster* tersebut termasuk ke dalam segmen “*Lost Customer Group*” yaitu kelompok pelanggan yang sudah hilang dengan karakteristik sebagian besar pelanggan pada *cluster 0* memiliki durasi hubungan yang rendah, merupakan pelanggan yang sudah lama tidak melakukan transaksi dan jarang melakukan transaksi namun potensi tinggi dalam hal pengeluaran.

2. *Cluster 1*

Cluster 1 terdiri dari 144 pelanggan dengan persentase 47,4% dari seluruh pelanggan. Berdasarkan Gambar 6 rata-rata nilai *Length* sebesar 0,009927 lebih rendah dari rata-rata keseluruhan ($L\downarrow$), rata-rata nilai *Recency* sebesar 0.183705 lebih rendah dari rata-rata keseluruhan ($R\downarrow$), rata-rata nilai *Frequency* sebesar 0,006076 lebih rendah dari rata-rata keseluruhan ($F\downarrow$) dan rata-rata nilai *Monetary* sebesar 0,078355 lebih rendah dari rata-rata keseluruhan ($M\downarrow$).

Dengan demikian, hasil perbandingan rata-rata masing-masing variabel pada *cluster 1* adalah ($L\downarrow R\downarrow F\downarrow M\downarrow$) dan berdasarkan matrik loyalitas pelanggan pada gambar 3 *cluster* tersebut termasuk ke dalam segmen “*New Customer Group*” yaitu kelompok pelanggan baru dengan karakteristik pelanggan tersebut memiliki durasi hubungan dengan usaha Mulia Jasa yang rendah dan perilaku pembelian terbaru yang dilakukan masih belum lama. Pelanggan tersebut masih belum banyak melakukan transaksi di usaha Mulia Jasa.

3. *Cluster 2*

Cluster 2 terdiri dari 28 pelanggan dengan persentase 9,2% dari seluruh pelanggan. Berdasarkan Gambar 6 rata-rata nilai *Length* sebesar 0,683920 lebih tinggi dari rata-rata keseluruhan ($L\uparrow$), rata-rata nilai *Recency* sebesar 0.194081 lebih rendah dari rata-rata keseluruhan ($R\downarrow$), rata-rata nilai *Frequency* sebesar 0,276786 lebih tinggi dari rata-rata keseluruhan ($F\uparrow$) dan rata-rata nilai *Monetary* sebesar 0.377354 lebih tinggi dari rata-rata keseluruhan ($M\uparrow$).

Dengan demikian, hasil perbandingan rata-rata masing-masing variabel pada *cluster 2* adalah ($L\uparrow R\downarrow F\uparrow M\uparrow$) dan berdasarkan matrik loyalitas pelanggan pada gambar 3, *cluster* tersebut termasuk ke dalam kategori “*Core Customer Group*” yaitu Kelompok pelanggan inti yang berpotensi dengan karakteristik tingginya durasi hubungan dengan usaha, sebagian besar pelanggan baru melakukan transaksi dan tingginya *Frequency* pembelian. Selain itu, nilai uang yang dikeluarkan pelanggan ke usaha Mulia Jasa termasuk tinggi.

IV. SIMPULAN

Hasil proses implementasi model *LRFM* dan Algoritma *K-Means* pada usaha Aluminium dan Kaca Mulia Jasa menghasilkan *cluster* paling optimal dari metode Silhouette adalah 3 *cluster* yaitu *cluster 0* atau “*Lost Customer Group*”

dimana pelanggan tersebut merupakan pelanggan baru yang menunjukkan potensi tinggi dalam hal pengeluaran. Namun, pelanggan tersebut sudah lama tidak bertransaksi di usaha Mulia Jasa. Pelanggan pada *cluster* 1 atau “*New Customer Group*” karena pelanggan tersebut menunjukkan perilaku pembelian terbaru yang masih belum lama. Pelanggan tersebut masih belum banyak melakukan transaksi di usaha Mulia Jasa. Pelanggan pada *cluster* 2 atau “*Core Customer Group*” yang berisi pelanggan berpotensi karena tingginya durasi hubungan dengan usaha dan tingginya *Frequency* pembelian. Selain itu, nilai uang yang dikeluarkan pelanggan ke usaha Mulia Jasa termasuk tinggi. Meskipun penelitian ini tidak mengembangkan strategi pemasaran secara langsung, namun hasil segmentasi tersebut dapat digunakan pemilik usaha untuk merancang strategi pemasaran yang lebih terarah berdasarkan karakteristik masing-masing *cluster* sehingga dapat meningkatkan penjualan. Selain itu, penelitian lanjutan juga dapat dilakukan untuk menentukan dan menguji strategi yang diterapkan pada masing-masing *cluster* untuk melihat dampak nyata terhadap peningkatan penjualan.

DAFTAR PUSTAKA

- [1] S. Muheramtohad, “Peran Lembaga Keuangan Syariah dalam Pemberdayaan UMKM di Indonesia,” *MUQTASID J. Ekon. dan Perbank. Syariah*, vol. 8, no. 1, p. 95, 2017.
- [2] (Kemenko Perekonomian RI), “Dorong UMKM Naik Kelas dan Go Export, Pemerintah Siapkan Ekosistem Pembiayaan yang Terintegrasi,” Kementerian Koordinator Bidang Perekonomian Republik Indonesia. [Online]. Available: <https://www.ekon.go.id/publikasi/detail/5318/dorong-umkm-naik-kelas-dan-go-export-pemerintah-siapkan-ekosistem-pembiayaan-yang-terintegrasi>
- [3] E. Suryanti, H. Lesmana, and H. Mubarak, “Analisis Strategi Pemasaran untuk Meningkatkan Pendapatan UMKM,” *Permana J. Perpajakan, Manajemen, dan Akunt.*, vol. 13, no. 1, pp. 60–72, 2021.
- [4] N. P. V. Viandari, I. M. A. D. Suarjaya, and I. N. Piarsa, “Pemetaan Pelanggan dengan LRFM dan Two Stage Clustering untuk Memenuhi Strategi Pengelolaan,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 130–139, 2022.
- [5] R. Siagian, P. S. Pahala Sirait, and A. Halima, “E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model,” *J. Informatics Telecommun. Eng.*, vol. 5, no. 1, pp. 21–30, 2021.
- [6] A. Wibowo and A. R. Handoko, “Segmentasi Pelanggan Ritel Produk Farmasi Obat Menggunakan Metode Data Mining Klasterisasi Dengan Analisis Recency Frequency Monetary (RFM) Termomodifikasi,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, pp. 573–580, 2020.
- [7] F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, Fachrudin, and T. M. A. Aziz, “Segmentation model of customer lifetime value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM model,” *Int. J. Integr. Eng.*, vol. 11, no. 3, pp. 169–180, 2019.
- [8] M. Hamzahi and S. Hosseini, “Business Intelligence Using Machine Learning Algorithms,” *Multimed. Tools Appl.*, vol. 81, pp. 33233–33251, 2022.
- [9] H. Priyatman, F. Sajid, and D. Haldivany, “Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa,” *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, p. 62, 2019.
- [10] S. Monalisa, “Klasterisasi Customer Lifetime Value dengan Model LRFM menggunakan Algoritma K-Means,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 2, pp. 247–252, 2018.
- [11] H. Syukron, M. Fauzi Fayyad, F. Junita Fauzan, Y. Ikhsani, and U. Rizkya Gurning, “Perbandingan K-Means K-Medoids dan Fuzzy C-Means untuk Pengelompokan Data Pelanggan dengan Model LRFM,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 76–83, 2022.
- [12] S. Juanita and R. D. Cahyono, “K-Means Clustering With Comparison of Elbow and Silhouette Methods for Medicines Clustering Based on User Reviews,” *J. Tek. Inform.*, vol. 5, no. 1, pp. 283–289, 2024.
- [13] S. H. Chao, M. K. Chen, and H. H. Wu, “An LRFM Model to Analyze Outpatient Loyalty From a Medical Center in Taiwan,” *SAGE Open*, vol. 11, no. 3, 2021.
- [14] P. Anitha and M. M. Patil, “RFM model for customer purchase behavior using K-Means algorithm,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1785–1792, 2019.
- [15] A. Rausanfitra, P. P. Adikara, and S. Adinugroho, “Analisis Sentimen Twitter Menggunakan Ensemble Feature dan Metode Extreme Learning Machine (ELM) (Studi Kasus: Samsung Indonesia),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 6409–6417, 2018.
- [16] F. Indriyani and E. Irfiani, “Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means,” *JUITA J. Inform.*, vol. 7, no. 2, p. 109, 2019.
- [17] R. Novidianto and K. Fithiasari, “Algoritma ClusterMix K-Prototypes Untuk Menangkap Karakteristik Pasien Berdasarkan Variabel Penciri Mortalitas Pasien Dengan Gagal Jantung,” *Inferensi*, vol. 4, no. 1, p. 37, 2021.
- [18] D. A. Kandeil, A. A. Saad, and S. M. Youssef, “A two-phase clustering analysis for B2B customer segmentation,” *Proc. - 2014 Int. Conf. Intell. Netw. Collab. Syst. IEEE INCoS 2014*, pp. 221–228, 2014.
- [19] A. A. D. Sulistyawati and M. Sadikin, “Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan,” *Sist. J. Sist. Inf.*, vol. 10, no. 3, pp. 516–526, 2021.