

## SEGMENTASI PELANGGAN BERDASARKAN ANALISIS *RECENCY, FREQUENCY, MONETARY* MENGUNAKAN ALGORITMA *K-MEANS* *CUSTOMER SEGMENTATION BASED ON RECENCY, FREQUENCY, MONETARY ANALYSIS USING K-MEANS ALGORITHM*

<sup>1</sup>Sri Wahyuni, <sup>2</sup>Fahrullah\*,

<sup>1,2,3</sup>Sistem Informasi (Kampus Kota Samarinda), Fakultas Ilmu Komputer, Universitas  
Mulia

Jalan Pahlawan No. 2A Kota Samarinda, Provinsi Kalimantan Timur, Indonesia

\*e-mail: [fahrullah@universitasmulia.ac.id](mailto:fahrullah@universitasmulia.ac.id)

### Abstrak

CV. Toedjoe Sinar Group merupakan salah satu percetakan di Samarinda dan bergerak di bidang digital printing dengan beragam produk, tidak hanya fokus pada strategi yang mengutamakan produk, tetapi juga melakukan strategi yang mengutamakan pelanggan. Hal ini diperlukan untuk mengetahui perilaku pelanggan sehingga akan membantu dalam penerapan strategi pemasaran yang tepat untuk meningkatkan pendapatan perusahaan. 8020 data transaksi pelanggan dari Bulan Januari hingga Maret tahun 2022 digunakan dalam penelitian ini. Pemodelan data yang digunakan dalam pengelompokan pelanggan adalah RFM (Recency, Frequency, Monetary) dan teknik data mining dengan menggunakan metode clustering menggunakan algoritma k-means yang terdiri dari *business understanding, data understanding, data preparation, modeling, evaluation, dan deployment*. Menentukan jumlah cluster dengan mencari nilai K menggunakan metode elbow dan hasil nilai  $k = 7$  sehingga cluster yang digunakan sebanyak 7 cluster dan cluster terbaik = 5 dengan nilai rata-rata recency 0,0, frequency 140,33 dan monetary 43328833.

**Kata kunci:** Algoritma K-Means, Model RFM (Recency, Frequency, Monetary), Python, Segmentasi Pelanggan.

### Abstract

CV. Toedjoe Sinar Group is one of the printing houses in Samarinda and is engaged in digital printing with a variety of products, not only focusing on strategies that prioritize products, but also carrying out strategies that put customers first. It is necessary to know the behavior of customers so that it will help in the implementation of the right marketing strategy to increase the company's revenue. 8020 customer transaction data from January to March of 2022 was used in this study. The data Modeling used in customer grouping is RFM (Recency, Frequency, Monetary) and data mining techniques using clustering methods using the k-means algorithm. Determine the number of clusters by looking for the value of K using the elbow method and the result of the value of  $k = 7$  so that the clusters used are 7 clusters and the best clusters = 5 with an average recency value of 0.0, frequency 140.33 and monetary 43328833.

**Keywords:** Customer Segmentation, K-Means Algorithm, Python, RFM Model (Recency, Frequency, Monetary).

## 1 PENDAHULUAN

CV. Toedjoe Sinar Group merupakan salah satu percetakan yang ada di Samarinda dan bergerak di bidang digital printing dengan berbagai macam produk seperti banner, baliho, stiker, poster, kartu nama, brosur, id card, stempel dan lain-lain. Saat ini, CV. Toedjoe Sinar Group tidak hanya fokus pada strategi yang mengutamakan produk, tetapi juga melakukan strategi yang mengutamakan pelanggan. Perusahaan ini menyadari akan pentingnya hubungan antara pelanggan yang loyal dengan keberhasilan bisnis. Hal ini tentu penting dilakukan untuk perkembangan perusahaan dengan menciptakan hubungan baik antara perusahaan dan pelanggan. Dalam situasi ekonomi yang semakin berubah membuat persaingan industri percetakan semakin ketat dan ditandai dengan munculnya perusahaan percetakan baru.

Segmentasi Pelanggan adalah suatu cara untuk mengelompokkan pelanggan ke dalam beberapa cluster (kelompok) dan setiap cluster memiliki beberapa anggota dengan karakteristik yang sama [1]. Segmentasi pelanggan melakukan pengelompokan karakteristik, kebutuhan, dan minat konsumen/pelanggan yang sama ke dalam satu kelompok dengan tujuan untuk menerapkan model bisnis yang mampu memenuhi setiap kebutuhan dari segmen-segmen tersebut. Dilakukannya hal ini agar strategi pemasaran dari perusahaan lebih tepat sasaran dan efisien. Dengan dilakukannya segmentasi, pelaku usaha dapat memahami bahwa pelanggan pun memiliki karakteristik khusus saat berinteraksi dengan produk [2]. Model RFM merupakan model untuk menentukan segmentasi konsumen berdasarkan kapan transaksi terakhir dilakukan, frequency yaitu tingkat keseringan pelanggan transaksi dan monetary ialah besarnya nilai transaksi yang dilakukan [3]. Analisa RFM adalah proses menganalisis perilaku pelanggan. Hal ini umumnya digunakan dalam pemasaran database dan pemasaran langsung. Analisis RFM merupakan suatu model perhitungan yang terdiri dari tiga atribut domain yang memperhatikan transaksi pelanggan berdasarkan transaksi terakhir/keterkinian (Recency), jumlah transaksi (Frequency), serta nominal dari transaksi (Monetary). Tujuan dari RFM adalah untuk meramalkan perilaku pelanggan di masa depan agar dapat mengarahkan keputusan segmentasi yang lebih baik. Model RFM ini merupakan metode yang sudah lama dan populer untuk mengukur hubungan dengan pelanggan [4].

Penggunaan teknik data mining merupakan salah satu solusi untuk persoalan segmentasi. Data mining merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari artificial intelligence, machine learning, statistics, dan database systems [5]. Data Mining merupakan bidang ilmu yang digunakan untuk menangani masalah pengambilan informasi dari database yang besar dengan menggabungkan teknik dari statistik, pembelajaran mesin, visualisasi data, pengenalan pola, dan database [6].

Clustering adalah teknik menemukan sekelompok data dari pemecahan atau pemisahan sekumpulan data menurut karakteristik tertentu yang telah ditentukan. Dalam pengelompokan tersebut nilai label nya belum diketahui label dari data tersebut. Metode clustering juga sering disebut tahapan awal sebelum melakukan metode lain seperti klasifikasi [7][8]. Clustering merupakan kumpulan objek data yang memiliki kemiripan antara satu dengan yang lain dalam kelompok. Clustering atau lebih dikenal dengan analisis cluster merupakan proses pengelompokkan satu set benda fisik ataupun abstrak kedalam satu kelas objek yang sama [9]. Semakin kecil jumlah cluster yang digunakan, maka akan mempermudah pemahaman terhadap struktur data di dalamnya, tetapi pola informasi penting yang ada akan terabaikan. Sebaliknya dengan memberikan jumlah cluster akhir yang lebih besar, maka kesamaan yang dimiliki antar kelompok akan semakin meningkat pula dan mengabaikan struktur data yang ada.

Salah satu metode clustering yang sangat populer dan banyak dipelajari untuk meminimalkan kesalahan clustering untuk titik ruang Euclidean disebut K-Means

clustering. K-Means adalah metode clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numerik. Algoritma k-means partitioning clustering yang memisahkan data ke K daerah yang terpisah. Algoritma k-means sangat terkenal karena kemudahannya dan kemampuannya untuk mengklaster data yang besar dan data yang outlier dengan sangat cepat. dalam algoritma k-means, setiap data harus termasuk ke cluster tertentu dan bisa dimungkinkan bagi setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster lainnya [10] [11].

Metode Elbow adalah untuk menghasilkan informasi dengan cara melihat perbandingan hasil antara jumlah cluster yang akan membentuk suatu titik terakhir, cluster nilai akan digunakan sebagai model data untuk cluster terbaik. Selain itu, hasil perhitungan akan digunakan untuk membandingkan jumlah cluster.

## 2 TINJAUAN PUSTAKA

Penelitian yang dilakukan dengan metode K-Means menggunakan data transaksi selama enam bulan, dengan total transaksi sebanyak 464 dari 25 data member. Data tersebut selanjutnya dicari nilai minimal dan maksimal dari masing-masing model RFM. Algoritma yang digunakan adalah algoritma k-means dengan metode Euclidean. Dilakukan penentuan jumlah klaster pada segmentasi pelanggan menggunakan SPSS. Pengolahan data menggunakan k-means clustering pada SPSS dibentuk menjadi 3 klaster, berdasarkan pengolahan data tersebut, diketahui anggota klaster 1 berjumlah 13 pelanggan, klaster 2 berjumlah 10 pelanggan dan anggota klaster 3 berjumlah 2 pelanggan. Dan hasil tersebut dapat disimpulkan bahwa perilaku pembelian pelanggan memiliki pola yang berbeda-beda. Klaster yang termasuk dalam segmen pelanggan potensial adalah klaster 1. Klaster tersebut perlu dipertahankan agar pelanggan tersebut tetap memiliki nilai R yang rendah, F yang tinggi dan M yang tinggi. Sementara itu, klaster yang perlu ditingkatkan lagi strateginya adalah klaster 2. Perlu dilakukan kajian lebih mendalam terkait peta demografis dan profil pelanggan untuk dapat menentukan strategi yang lebih tepat. Perbedaan penelitian ini dengan penelitian yang penulis lakukan yaitu objek penelitian yang dilakukan berbeda, serta aplikasi yang digunakan untuk berbeda penulis menggunakan bahasa program python dalam melakukan clustering menggunakan K-Means [12].

Penelitian yang dilakukan pada salah satu perusahaan digital printing di Yogyakarta, data dikumpulkan melalui observasi dan interview pada pemilik perusahaan. 500 data pelanggan dari periode maret sampai dengan mei 2020 di modelkan dengan RFM dan algoritma yang digunakan yaitu k-means dengan menggunakan software python. Penerapan analisis RFM dan metode k-means menggunakan software python menghasilkan 4 cluster dari 500 data pelanggan. Pada hasil clustering untuk 4 cluster dapat dijelaskan bahwa hasil presentasinya masih rendah. Untuk cluster 0 hanya 28% dengan frequency dan monetary-value nya berada pada tingkat yang rendah, yang berarti pada cluster tersebut termasuk cluster untuk kriteria customer yang berbelanja pada waktu terdekat, tetapi tidak membeli produk banyak. Selanjutnya pada cluster tertentu total untuk presentase data nya dapat mencapai 101%, dikarenakan fungsi pada python. Perbedaan penelitian ini dengan penelitian yang dilakukan penulis adalah objek penelitian berbeda [13].

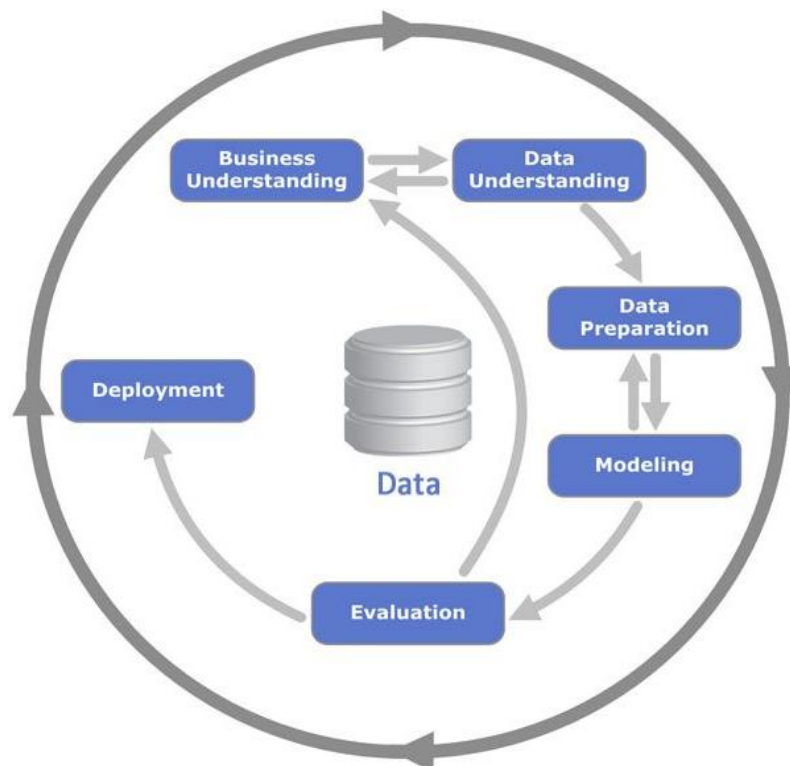
Adapun penelitian menggunakan metode Clustering Fuzzy C-Means untuk menentukan jumlah K, metode yang digunakan adalah metode Elbow. Dalam proses clustering, diterapkan variabel RFM yang menggambarkan perilaku pelanggan dalam melakukan transaksi. Setelah hasil clustering muncul maka akan divisualisasikan sehingga mudah untuk menganalisa setiap cluster. Diperoleh data sejumlah 123 pelanggan yang melakukan transaksi sejak Januari 2017 hingga Juni 2017 Hasil dari clustering dengan Fuzzy C-Means dan model RFM adalah tiga segmen pelanggan. Segmen 1 adalah segmen terburuk dengan rata-rata nilai CLV 0.0538. Segmen 2 adalah segmen menengah dengan rata-rata nilai CLV 0.3516. Segmen 1 memiliki 42 anggota, segmen 2 memiliki 64 anggota, dan segmen 3

memiliki 17 anggota. Perbedaan penelitian ini dengan penelitian yang dilakukan penulis yaitu algoritma yang digunakan berbeda, penulis menggunakan algoritma k-means dan menggunakan bahasa program python dalam melakukan clustering menggunakan k-means [14].

### 3 METODE PENELITIAN

Penelitian ini dilakukan di CV. Toedjoe Sinar Group yang beralamat di Jalan K.H. Wahid Hasyim I Nomor 40, Sempaja Selatan, Kecamatan Samarinda Utara, Kota Samarinda, Kalimantan Timur. Penelitian ini dilaksanakan selama 3 bulan yaitu pada bulan Januari 2022 sampai dengan bulan Maret 2022. Scope penelitian ini adalah pelanggan toedjoe sinar group berdasarkan kelompok transaksi pelanggan. Adapun pengambilan data menggunakan teknik wawancara pada pimpinan CV. Toedjoe Sinar Group dan menganalisis database aplikasi di CV. Toedjoe Sinar Group yang berupa data excel yang diambil pada masa penelitian.

Berikut merupakan alur penelitian dari penerapan algoritma K-Means dengan metode analisis data menggunakan CRISP-DM (Cross-Industry Standard Process for Data Mining) dalam segmentasi pelanggan berdasarkan analisis Recency, Frequency, Monetary menggunakan algoritma K-Means yang dapat dilihat pada gambar 1 sebagai berikut:



**Gambar 1.** Crisp-DM

1) *Business Understanding*

Pada fase ini berfokus pada pemahaman dan perspektif bisnis proses dari suatu sistem atau penelitian. Yaitu penentuan tujuan penelitian, menentukan tujuan bisnis, dan menilai situasi, dan menentukan tujuan data mining

2) *Data Understanding*

Pada tahap ini yaitu berfokus pada pembelajaran data yang sudah ada, pengumpulan dan penyeleksian data.

3) *Data Preparation*

Pada fase persiapan data terdiri dari pemilihan data, pembersihan data, mengintegrasikan data, dan transformasi data agar dapat dilanjutkan kedalam tahap pemodelan. Oleh sebab itu, dibutuhkan teknik data preprocessing untuk mempersiapkan data yang akan diproses.

4) *Modeling*

fase ini proses yang terjadi adalah pemilihan model yang sesuai agar mengoptimalkan hasil. Model clustering dimulai dari dataset akan dilakukan pemodelan dengan algoritma K-Means sehingga dihasilkan model clustering dan memunculkan parameter evaluasi

5) *Evaluation*

Pada fase ini akan dilakukan proses evaluasi dari phase sebelumnya. Fase evaluasi ini akan dilakukan uji validasi dengan sum of square error (SEE). SEE digunakan untuk menganalisis variasi pada jarak antar pusat kluster yang terbentuk

Berikut ini merupakan rumus SSE:

$$SSE = \sum_{k=1}^k \sum_{xi \in Sk} \|Xi - Ck\|^2 \tag{1}$$

Keterangan:

$Xi$  : nilai atribut dari data ke- $i$

$Ck$  : nilai atribut titik pusat Cluster ke- $i$

Metode Elbow untuk menentukan jumlah cluster yang paling optimum atau yang terbaik. Langkah-langkah metode elbow:

1. awal nilai cluster (2 cluster)
2. Menaikkan nilai cluster sampai jumlah cluster (4 cluster)
3. Menghitung nilai SSE (Sum of Squares Error) dari setiap cluster 6

$$SSE = \sum_{k=1}^k \sum_{xi \in Sk} \|Xi - Ck\|_2^2 \tag{2}$$

$Xi$  : nilai atribut dari data ke- $i$

$Ck$  : nilai atribut titik pusat Cluster ke- $i$

4. Melakukan perhitungan SSE (Sum of Squares Error) sampai cluster yang ditentukan.

5. Melihat hasil SSE (Sum of Squares Error) dari nilai cluster yang turun secara drastis atau nilai yang berubah signifikan.

6. Menetapkan nilai cluster yang berbentuk siku [15].

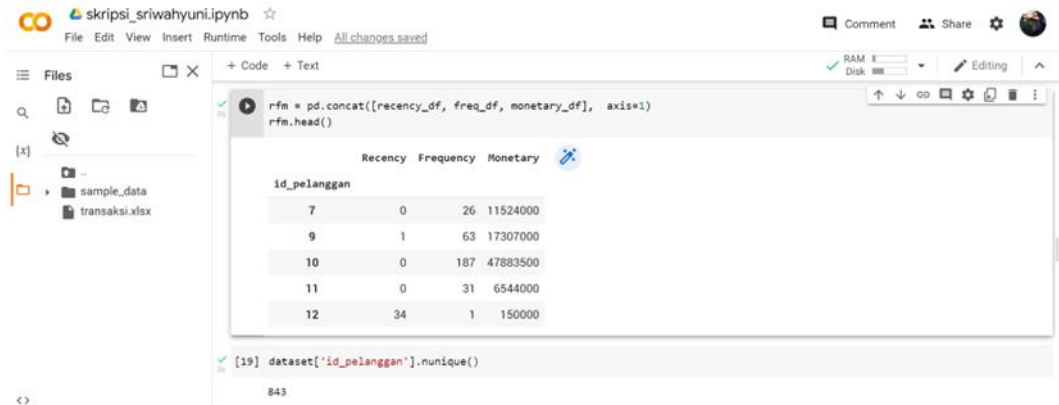
6) *Deployment*

Pada fase ini proses yang terjadi adalah penyusunan laporan atau presentasi dari pengetahuan yang didapat dari analisis masing-masing atribut RFM.

#### 4 HASIL DAN PEMBAHASAN

Objek dalam penelitian ini adalah pelanggan CV. Toedjoe sinar group, data yang diperoleh dari database aplikasi CV. Toedjoe Sinar Group sebanyak 8020 data transaksi pelanggan dimulai dari bulan Januari 2022 sampai dengan bulan Maret 2022.

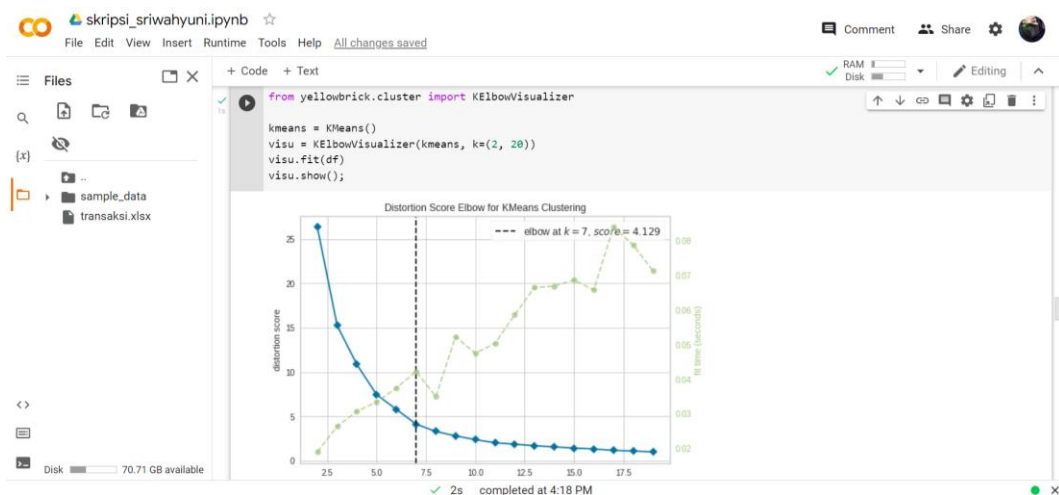
Setelah memperoleh masing-masing nilai recency, frequency, dan monetary. Selanjutnya menggabungkan tabel recency, frequency, dan monetary menjadi satu tampilan atau satu tabel seperti pada gambar 2 berikut.



Gambar 2. Tampilan Data Model RFM pada Google Colab

Jumlah data pada hasil pembersihan data sebanyak 7895 data, dan dilakukan pemodelan RFM sehingga data berubah menjadi 843 data, dan data ini menjadi dataset yang akan digunakan dalam proses selanjutnya yaitu proses clustering ke algoritma k-means.

Selanjutnya menentukan jumlah cluster atau mencari nilai K dengan menggunakan metode elbow.



Gambar 3. Metode Elbow Dalam Penentuan Jumlah Cluster

Dapat dilihat pada gambar 3 diatas, tampilan grafik untuk mencari nilai K dengan menggunakan metode elbow, dan hasilnya menunjukkan nilai K = 7. Dengan demikian cluster yang akan digunakan adalah sebanyak 7 cluster

Setelah analisis pelanggan dari cluster 1 hingga 7 telah dilakukan, selanjutnya dilakukan perbandingan hasil cluster analisis pelanggan dengan menampilkan nilai RFM semua cluster yang dapat dilihat pada tabel 1 berikut.

Tabel 1. Perbandingan Hasil Cluster Analisis Pelanggan

	REGENCY	FREQUENCY	MONETARY
CLUSTER 1	42,79	1,76	484533
CLUSTER 2	5,64	6,16	1266482
CLUSTER 3	76,42	1,16	225955
CLUSTER 4	22,43	2,28	456420
CLUSTER 5	0,0	140,33	43328833
CLUSTER 6	59,85	1,41	298318

CLUSTER 7	0,96	45,24	10510712
-----------	------	-------	----------

Hasil nilai RFM pada semua cluster terlihat cluster dengan nilai terbaik yaitu pada cluster 5, cluster terbaik kedua yaitu pada cluster 7 dan cluster dengan nilai RFM terendah yaitu cluster 3.

## 5 KESIMPULAN

Berdasarkan analisis yang dilakukan didapatkan kesimpulan bahwa segmentasi pelanggan pada CV. Toedjoe Sinar Group menghasilkan nilai  $K=7$  dan cluster terbaik = 5 dengan nilai rata-rata recency 0,0, frequency 140,33 dan monetary 43328833

Pada penelitian selanjutnya diharapkan dapat mengembangkan penelitian ini dengan menambahkan atribut produk untuk mengidentifikasi dan mencari karakteristik pembeda agar setiap kelompok dapat diidentifikasi berdasarkan dengan produk yang paling laris dan atau mampu melakukan komparasi untuk mencari metode clustering terbaik dengan menggunakan algoritma clustering lainnya seperti K-Medoids dan Hierarchical Clustering.

## DAFTAR PUSTAKA

- [1] N. H. Harani and F. A. Nugraha, *Segmentasi Pelanggan Menggunakan Python*. Kreatif, 2020.
- [2] R. P. Justitia, N. Hidayat, and E. Santoso, "Implementasi Metode Agglomerative Hierarchical Clustering Pada Segmentasi Pelanggan Barbershop (Studi Kasus: RichDjoe Barbershop Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, pp. 1048–1054, 2021.
- [3] F. Hadi, M. Mustakim, D. O. Rahmadia, F. H. Nugraha, N. P. Bulan, and S. Monalisa, "Penerapan K-Means Clustering Berdasarkan RFM Mofek Sebagai Pemetaan dan Pendukung Strategi Pengelolaan Pelanggan (Studi Kasus: PT. Herbal Penawar Alwahidah Indonesia Pekanbaru)," *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 15, no. 1, pp. 69–76, 2017.
- [4] B. E. Adiana, I. Soesanti, and A. E. Permasari, "Analisis segmentasi pelanggan menggunakan kombinasi RFM model dan teknik clustering," *J. Terap. Teknol. Inf.*, vol. 2, no. 1, pp. 23–32, 2018.
- [5] D. Suyanto, "Data Mining untuk klasifikasi dan klusterisasi data," *Bandung Inform. Bandung*, 2017.
- [6] I. Werdiningsih, M. Kom, B. Nuqoba, M. Kom, and S. S. Muhammadun, *Data Mining Menggunakan Android, Weka, dan SPSS*. Airlangga University Press, 2020.
- [7] E. Prasetyo, "Data mining konsep dan aplikasi menggunakan matlab," *Yogyakarta Andi*, vol. 1, 2012.
- [8] R. P. Aditya, F. Fahrullah, and N. W. W. Sari, "Implementasi Algoritma Apriori Untuk Rekomendasi Paket Menu Pada Cafe Abc Berbasis Website," *J. Simantec*, vol. 11, no. 2, pp. 223–230, 2023.
- [9] E. Irfiani and S. S. Rani, "Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita," *J. Sist. dan Teknol. Inf*, vol. 6, no. 4, p. 161, 2018.
- [10] M. Wahyudi, M. Masitha, R. Saragih, and S. Solikhun, *Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*. Yayasan Kita Menulis, 2020.
- [11] E. Prasetyowati, I. R. NR, and S. Rachmatullah, "PENERAPAN K-MEANS ALGORITHM UNTUK MENGIDENTIFIKASI SUPPLIER BAHAN BAKU PADA KOMODITAS AGRIKULTUR DI KABUPATEN PAMEKASAN," *J. Simantec*, vol. 11, no. 2, pp. 147–156, 2023.
- [12] A. Febriani and S. A. Putri, "Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means," *JIEMS (Journal Ind. Eng. Manag. Syst.*, vol. 13, no. 2, 2020.

- [13] I. Istiningsih, “IMPLEMENTASI DATA MINING PENGELOMPOKKAN PELANGGAN MENGGUNAKAN RFM DAN K-MEANS CLUSTERING STUDI KASUS DI INDONESIA DIGITAL PRINTING YOGYAKARTA.” STMIK AKAKOM YOGYAKARTA, 2020.
- [14] D. B. Saputra and E. Riksakomara, “Implementasi Fuzzy C-Means dan Model RFM untuk Segmentasi Pelanggan (Studi Kasus: PT. XYZ),” *J. Tek. ITS*, vol. 7, no. 1, pp. A119–A124, 2018.
- [15] N. P. E. Merliana and A. J. Santoso, “Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering,” 2015.