

Analisis Segmentasi Recency dan Customer Value Pada AVANA Indonesia Dengan Algoritma K-Means dan Model RFM (Recency, Frequency and Monetary)

Muhammad Jordy, Agung Triayudi*, Ira Diana Sholihati

Fakultas Teknologi Komunikasi dan Informatika, Sistem Informasi, Universitas Nasional, Jakarta

Jl. Sawo Manila No.61, RW.7, Pejaten Bar., Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta, Indonesia

Email: ¹muhammadjordy98@gmail.com, ^{2,*}agungtriayudi@civitas.unas.ac.id, ³iradiana2803@gmail.com

Email Penulis Korespondensi: agungtriayudi@civitas.unas.ac.id

Submitted: 17/01/2023; Accepted: 28/01/2023; Published: 29/01/2023

Abstrak—Avana Indonesia adalah startup social commerce yang berkantor pusat di Malaysia. Ingin memperluas bisnis mereka dan memasuki pasar Indonesia, mereka masih belum memiliki strategi pemasaran yang terbaik, sehingga kesepakatan penjualan layanan saja tidak cukup. Karena itulah diperlukan strategi pemasaran yang berfokus pada pelanggan dengan customer relationship management, salah satunya adalah segmentasi pelanggan. Segmentasi pelanggan dapat dilakukan dengan mengimplementasikan proses data mining yang dilakukan dengan menggunakan algoritma clustering K-Means berdasarkan model RFM (Recency, Frequency, Monetary). Jumlah cluster dalam proses clustering ditentukan dengan menggunakan metode elbow. Analisis cluster berdasarkan nilai pelanggan dengan metode recency clustering mengungkapkan active customer, warm customers, cold customers dan inactive customers. Kemudian kedua dari recency frequency (customer value) segmentasi menghasilkan cluster yaitu common, ultra-high, low, dan high.

Kata Kunci: Segmentasi Pelanggan; Nilai Pelanggan; Analisis RFM; Klasifikasi, Algoritma K-Means; Metode Elbow; Streamlit.

Abstract—Avana Indonesia is a social commerce startup headquartered in Malaysia. Wanting to expand their business and enter the Indonesian market, they still don't have the best marketing strategy in place, so a service sales deal is not enough. That's why we need a marketing strategy that focuses on customers with customer relationship management, one of which is customer segmentation. Customer segmentation can be done by implementing a data mining process which is carried out using the K-Means clustering algorithm based on the RFM (Recency, Frequency, Monetary) model. The number of clusters in the clustering process is determined using the elbow method. Cluster analysis based on customer value using the recency clustering method reveals active, warm, cold, and inactive customers. Then the two from recency frequency (customer value) segmentation produce common, ultra-high, low, and high clusters.

Keywords: Customer Segmentation; Customer Value; RFM Analysis; Classification, K-Means Algorithm; Elbow Method; Streamlit.

1. PENDAHULUAN

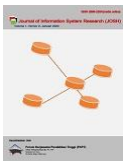
Dunia bisnis saat ini berkembang sangat pesat. Hal ini seiring dengan perkembangan teknologi internet yang dapat memudahkan segala aktivitas bisnis. Internet telah menjadi alat yang memfasilitasi belanja di pasar baru yang dikenal sebagai e-commerce. E-commerce sendiri merupakan sebuah platform atau wadah yang memfasilitasi jual beli barang dan jasa secara online. Peningkatan pengguna internet ini juga mempengaruhi jumlah toko untuk perusahaan e-commerce. Mengingat banyaknya toko e-commerce, penjual sedang mempertimbangkan opsi lain untuk penjualan dan strategi pemasaran lainnya.

Salah satunya adalah mencoba melakukan penjualan langsung menggunakan social media mereka. Kepercayaan dianggap sebagai faktor penting dalam perdagangan sosial karena perannya yang penting dan konten peer-to-peer yang penting. Dengan model yang diusulkan dikembangkan berdasarkan tiga konsep: kehadiran sosial, pencarian informasi bisnis sosial dan keakraban platform. Model ini menjadi salah satu alternatif pencarian informasi, keakraban, kehadiran sosial, dan kepercayaan memengaruhi niat pelanggan untuk membeli dari platform social commerce [1].

Munculnya social commerce telah membawa perubahan besar pada toko dan konsumen, dan itu adalah evolusi perdagangan elektronik. Platform sosial sebagai aktor membantu mempromosikan penjualan barang melalui interaksi toko-ke-konsumen, interaksi konsumen-ke-konsumen atau kampanye konsumen sukarela [2].

PT Niaga Masa Depan (AVANA Indonesia) merupakan sebuah Social Commerce yang populer di Indonesia memiliki beberapa fitur antara lain seperti Facebook Store, Instagram Shopping dan WhatsApp Commerce. Ketika sudah maraknya marketplace yang sudah ada di Indonesia, social commerce memiliki daya tarik sendiri untuk melakukan penjualan. Contoh daya tarik tersebut adalah dapat menjangkau konsumen ditempat-tempat yang penting, menciptakan pengalaman belanja tanpa hambatan dan memanfaatkan hype e-commerce atau kejenuhan pelanggan dalam penggunaannya [3].

Oleh sebab itu dengan pesatnya peningkatan orang berjualan di e-commerce mendorong platform social commerce-pun ikut berkembang juga. Dengan banyaknya customer yang bergabung dengan PT Niaga Masa Depan (AVANA Indonesia) ini dan belum memiliki analisa data pelanggan beserta segmentasinya.



Tujuan segmentasi pelanggan adalah mengelompokkan pelanggan berdasarkan karakteristik yang sama. Pengelompokan didasarkan pada minat atau permintaan pelanggan yang serupa. Dengan melakukan segmentasi pelanggan, perusahaan dapat memasarkan produknya secara individual agar sesuai dengan kebutuhannya.

Metode Recency, Frequency, Monetary (RFM) terbaru menganalisis karakteristik nasabah berdasarkan database transaksi nasabah membentuk tiga variabel utama yaitu recency, frequency, dan monetary sehingga dapat menganalisis data historis peningkatan transaksi nasabah [4].

Untuk menguji model RFM, peneliti menggunakan clustering. Ini mengasumsikan bahwa jika pelanggan termasuk dalam cluster yang sama, perilaku yang diharapkan pelanggan sama dengan perilaku pelanggan di dalam cluster tersebut. Clustering menunjukkan segmentasi pelanggan. Dalam penelitian ini, kami menggunakan metode pengelompokan K-means untuk mengelompokkan data ke dalam segmen-segmen. Algoritma pengelompokan K-Means telah banyak digunakan untuk segmentasi pelanggan dalam penelitian sebelumnya [5].

Dari data yang sudah dimodelkan menggunakan analisis RFM serta menguji dengan menggunakan algoritma clustering K – Means. Menghasilkan beberapa segmentasi pelanggan yaitu recency segmentation dan recency frequency (Customer Value Segmentation) dari segmentasi tersebut menghasilkan sebuah informasi yang dapat membantu manajemen untuk mengetahui kualitas pelanggan. Serta, memberikan sebuah jawaban bagi manajemen perusahaan untuk strategi pemasaran kedepan menggunakan data yang sudah diolah ini [6].

Selain menghasilkan informasi yang berguna data juga harus ditampilkan dengan menarik menggabungkan desain yang diusulkan dengan website Streamlit, menghasilkan aplikasi web yang indah serta informatif yang dapat dipahami dengan mudah oleh semua manajemen dan pemerhati data tersebut [7].

Terdapat beberapa penelitian yang sudah dilakukan diantaranya adalah [8] Ini membantu mengelompokkan pelanggan dan mengukur loyalitas terhadap produk usaha kecil. Kami menggunakan metode CRISP-DM yang terdiri dari enam langkah. Pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi. Algoritma K-Means digunakan untuk pengelompokan dan RapidMiner digunakan sebagai alat untuk mengevaluasi hasil kluster. Pengelompokan didasarkan pada analisis RFM (Recency, Frequency, Monetary). Gunakan Indeks Davies-Bouldin (DBI) untuk menemukan jumlah cluster yang optimal (k). Pelanggan ini terbagi menjadi tiga kluster, jumlah pelanggan pada kluster pertama sebanyak 30 pelanggan yang termasuk dalam kategori pelanggan tipikal, dan jumlah pelanggan pada kluster kedua sebanyak 8 pelanggan yang termasuk dalam kategori pelanggan superstar sebanyak 89 pelanggan. pelanggan yang termasuk dalam kategori pelanggan dan pelanggan superstar. Ketiga cluster tersebut merupakan kelas cluster yang tidak aktif.

Untuk meningkatkan layanan pelanggan, diperlukan penelitian lebih lanjut [9] untuk mempelajari karakteristik loyalitas pelanggan pada suatu perusahaan. Model segmentasi pelanggan berdasarkan data peristiwa dapat memberikan informasi ini. Penelitian ini menggunakan parameter recency, frequency, and monetary (RFM) untuk menentukan segmentasi pelanggan dan mendistribusikan algoritma k-means untuk menentukan jumlah cluster. 588 transaksi penjualan PT Dinar Energi Utama pada tahun 2017 digunakan sebagai data. Cluster yang dibentuk oleh kedua algoritma k-means dan k-means diuji menggunakan metode koefisien siluet. Algoritma k-means membagi keduanya untuk membentuk segmentasi pelanggan terbaik menjadi tiga kelompok: acak, standar, dan emas dengan faktor siluet 0,58132.

Kemudian penelitian dari [10] tujuannya adalah untuk mensegmentasi pelanggan. Model RFM digunakan sebagai atribut kuantitatif dari variabel input. Selanjutnya, gunakan algoritma K-means untuk mengelompokkan pelanggan. Tesis ini menghasilkan segmen pelanggan yang dimiliki perusahaan dan karakteristik masing-masing pelanggan. Berdasarkan akurasi yang diperoleh pada perhitungan yang dilakukan sistem segmentasi pelanggan dalam 29 kali pengujian, 29 pelanggan (100%) memiliki karakteristik yang dihasilkan sistem sesuai dengan pengetahuan pengguna, 0 pelanggan (0%) tidak sesuai dengan pengetahuan pengguna. Dengan nilai akurasi tersebut dapat dikatakan baik. Karakteristik pelanggan ini membantu PT Coversuper Indonesia Global membuat keputusan untuk memprioritaskan personel dan sumber daya untuk pelanggan (potensial) tertentu.

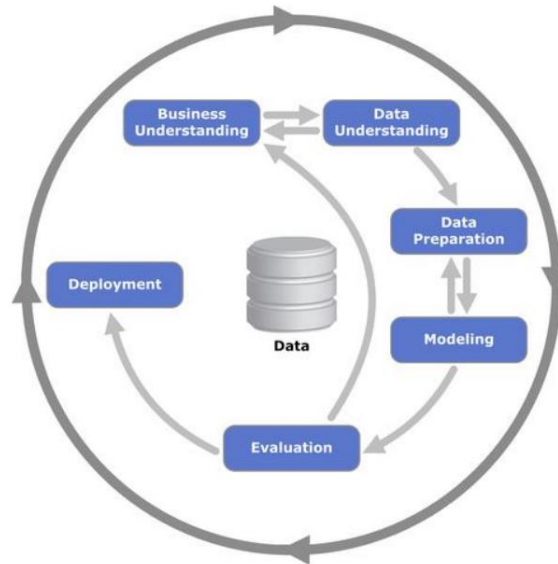
Selanjutnya tujuan penelitian yang dilakukan oleh [11] adalah untuk mengetahui segmentasi pelanggan elektronik dengan menggunakan pengelompokan Fuzzy C-Means dan metode RFM. Proses pengelompokan dilakukan enam kali, dan jumlah kluster bervariasi dari dua hingga tujuh kluster. Hasil penelitian menunjukkan bahwa jumlah kluster optimal yang dibentuk oleh indeks validitas Xie-Beni adalah empat kluster. Cluster-cluster tersebut menjadi segmen pelanggan, dengan karakteristik masing-masing pelanggan berdasarkan recency, frequency dan monetary. Segmen terbaik adalah segmen yang memiliki pelanggan sangat setia yang membeli dari toko online tumbas.in. Segmen yang terbentuk dapat diperhitungkan saat menerapkan strategi pemasaran yang tepat untuk setiap pelanggan.

Dan terakhir penelitian dari [12] Tujuannya adalah mengolah data transaksi dengan total 153.92 data menjadi 10.145 data dalam model RFM untuk mengidentifikasi calon pelanggan. Inisialisasi cluster pertama pada metode K-Medoids, X-Means dan K-Means dilakukan secara acak. Pada penelitian ini nilai k dimulai dari 1 sampai dengan 10. Nilai K diterapkan secara iteratif dan validasi cluster dihitung dengan menggunakan metode David Bouldin Index (DBI) dan centroid dari rata-rata jarak cluster. Hasil penelitian menunjukkan bahwa K-means memiliki nilai validitas yang lebih baik dibandingkan dengan X-means dan K-means. Nilai rata-rata DBI yang dihasilkan oleh metode K-Medoids adalah 0,50778. Jumlah cluster terbaik yang dihasilkan adalah 5 cluster, ditentukan dengan mempertimbangkan distribusi data untuk k = 5 yang memberikan nilai yang sama pada metode

K-Medoids, X-Means dan K-Means. Tingkat pelanggan yang terbentuk adalah pelanggan yang tidak aktif, pelanggan yang menuntut perhatian, pelanggan baru, pelanggan potensial dan pelanggan tetap.

2. METODOLOGI PENELITIAN

Metode penelitian yang digunakan adalah CRISP-DM. Pada Gambar 1 dibawah ini merupakan tahapan -tahapan berikut didasarkan pada tahap CRISP-DM:



Gambar 1. Fase CRISP-DM [13]

2.1 Pemahaman Bisnis (Business Understanding)

Pada tahap memahami perusahaan ini, penulis perlu memahami dan memahami model bisnis PT Niaga Masa Depan (AVANA Indonesia) saat ini. PT Niaga Masa Depan (AVANA Indonesia) adalah perusahaan swasta yang bergerak di bidang social commerce dan menawarkan situs web e-commerce yang terhubung langsung dengan Instagram Shopping, Facebook Store, dan Messenger Store. Ada juga puluhan fitur canggih, seperti manajemen bisnis, untuk membantu perusahaan mendigitalkan bisnisnya dan bersaing dengan kompetitor di industri yang bergerak cepat [14].

2.2 Pemahaman Data (Data Understanding)

Tahap pemahaman data ini berawal dari mengumpulkan informasi, kemudian mengidentifikasi kualitas informasi yang digunakan, mencari informasi tersembunyi yang dapat membentuk hipotesis baru. Data mining dilakukan pada database perusahaan, dimana pada database terdapat 8 tabel, namun yang digunakan hanya 5 tabel yaitu. tabel Shop_ID, Date, Price_Sales, Product_Name, dan Product_Type.

2.3 Data Preparation

Ada beberapa hal yang melatarbelakangi langkah data preparation [15]:

2.3.1 Data Selection

Data Selection adalah data yang dipilih untuk diolah berdasarkan kolom yang tersedia dan sesuai kebutuhan. Data yang tidak terkait dengan metode yang digunakan kemudian akan dihapus.

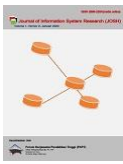
2.3.2 Data Preprocessing

Setelah tahap data selection, data dilanjutkan sehingga data yang disiapkan bersih dari data noise dan missing value. Data diurutkan berdasarkan dengan nilai data yang ada pada setiap kolom.

2.3.3 Data Transformation

Jika data yang dipilih bebas dari data noise dan missing value. Data tersebut kemudian dapat melakukan min-max normalization untuk mengubah data menjadi nilai yang bermakna. Normalisasi digunakan untuk meningkatkan akurasi proses komputasi numerik pada skala data antara 0 dan 1.

$$x^1 = \frac{x - \min_A}{\max_A - \min_A} (new_{\max_A} - new_{\min_A}) + new_{\min_A} \tag{1}$$



X adalah data aktual, minA adalah data aktual terkecil, maxA adalah data aktual terbesar, new_maxA adalah skala data tertinggi yaitu 1, dan new_minA adalah skala data terkecil, dengan skala data terendah adalah 0 [16].

Setelah data dilakukan normalisasi kemudian dilanjutkan pada pada proses clustering algoritma k-means. Algoritma K-Means Clustering bekerja sebagai berikut [17]:

1. Tentukan nilai k sebagai jumlah cluster yang terbentuk.
2. Tetapkan nilai awal centroid atau pusat cluster. Pada langkah ini nilai centroid ditentukan secara acak, namun pada langkah selanjutnya digunakan rumus di bawah ini:

$$V_{ij} = \frac{1}{N_i} = \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

3. Hitung jarak antara titik centroid dengan masing-masing titik objek dengan menggunakan jarak Euclidean menggunakan rumus berikut:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (3)$$

4. Kelompokkan data membentuk cluster sehingga titik centroid pada setiap cluster merupakan centroid terdekat. Anggota cluster ditentukan oleh informasi jarak minimum antar objek.
5. Perbarui nilai centroid untuk setiap cluster.
6. Ulangi langkah 2 sampai akhir sampai nilai center point tidak berubah lagi.

2.4 Modelling

Sesudah data yang dapat digunakan melewati proses pengolahan data, data tersebut diproses dengan bobot sesuai nilai dan tujuan untuk memudahkan pengolahan data. Berdasarkan data penjualan, pengolahan data dilakukan dengan algoritma pengelompokan data mining berbasis bagian yang disesuaikan dengan atribut yang dibuat dalam kumpulan data, dan menentukan jenis pelanggan berdasarkan cluster yang dihasilkan berupa 4 Cluster terdiri dari berbagai macam segmentasi.

2.5 Evaluation

Tahap evaluasi merupakan tahap lanjutan dimana tujuan data mining dievaluasi secara menyeluruh untuk mendapatkan pemodelan yang diinginkan. Beberapa hal yang dilakukan dalam fase ini antara lain mengevaluasi hasil, seberapa jauh pemodelan telah mencapai tujuan yang diinginkan, proses pengecekan atau pengecekan ulang untuk memastikan bahwa semua langkah yang dilakukan tidak meleset, dan pendefinisian langkah-langkah yang dilakukan selanjutnya yaitu melanjutkan ke tahap deployment atau kembali ke tahap awal yaitu business understanding. Pada tahap evaluasi, hasil pengelompokan dari tahap pemodelan dievaluasi menggunakan metode elbow untuk mendapatkan jumlah cluster yang optimal. Pada tahap ini dilakukan pemodelan dengan Model RFM adalah metode analisis untuk mengidentifikasi sikap pelanggan dan menyajikan sikap pelanggan berdasarkan tiga atribut yaitu Recency, Frequency, dan Monetary. Analisis RFM terdiri dari Recency, Frequency dan Monetary yang makna memiliki sebagai berikut[18]:

1. Recency adalah variabel yang mengukur nilai pelanggan berdasarkan periode waktu (tanggal, bulan, tahun) dari transaksi terakhir pelanggan hingga saat ini. Semakin kecil intervalnya, semakin besar nilai akhirnya.
2. Frequency adalah variabel yang mengukur nilai pelanggan berdasarkan jumlah transaksi yang dilakukan pelanggan selama suatu periode. Semakin tinggi jumlah transaksi, semakin tinggi nilai f.
3. Monetary adalah variabel yang mengukur nilai pelanggan berdasarkan jumlah yang dihabiskan pelanggan selama periode waktu tertentu. Semakin tinggi pengeluaran pelanggan, semakin tinggi nilai M.

2.6 Deployment

Pada tahap diseminasi ini, informasi dan pengetahuan yang diperoleh disajikan dalam bentuk yang lebih mudah dipahami oleh masyarakat umum. Pada langkah ini juga digunakan website Streamlit untuk memvisualisasikan hasil yang diperoleh. Dengan diperkenalkannya situs web Streamlit, memudahkan untuk membaca data, berbagi data, berkolaborasi, dan bahkan mengeksplor data. Hasil penelitian ini berupa kelompok pelanggan, data calon konsumen, data kategori dan jenis pembayaran, serta rekomendasi strategi pemasaran yang dapat diterapkan perusahaan berdasarkan data yang dihasilkan untuk meningkatkan pendapatan usaha.

3. HASIL DAN PEMBAHASAN

3.1 Pemahaman Bisnis (Business Understanding)

Tujuan bisnis adalah agar proses bisnis beroperasi secara efisien dan efektif untuk mendukung proses bisnis [14]. Oleh karena itu, agar perusahaan dapat berkembang terlebih dahulu perlu diketahui tujuan usaha, manfaat usaha, tujuan usaha dan hal-hal lain yang berkaitan dengan usaha. Banyak orang perlu mengetahui beberapa hal

ini, terutama mereka yang ingin memulai bisnis. Sebenarnya, kita tidak hanya mengetahui sebagian dari hal-hal ini, tetapi kita perlu memahami hal-hal ini.

3.2 Pemahaman Data (Data Understanding)

Tabel 1 dibawah ini menjelaskan data transaksi penjualan yang diberikan perusahaan adalah untuk periode Januari hingga Desember 2021. Data diperoleh berupa agregat Excel dalam format CSV, dengan total 5 atribut dan total transaksi 1080.

Tabel 1. Data Transaksi (02-01-2021 – 11-01-2021)

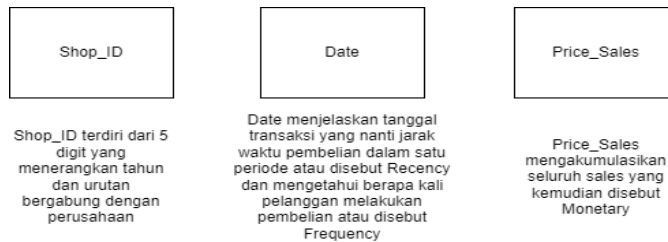
No	Shop ID	Date	Price Sales	Product Name	Product Type
1	185676	02/01/2021	1,499,999	Beginner Yearly	Subscription
2	184673	02/01/2021	399,999	Beginner Quarterly	Subscription
3	184427	03/01/2021	1,499,999	Beginner Yearly	Subscription
4	184511	03/01/2021	1,499,999	Beginner Yearly	Subscription
5	184623	03/01/2021	1,499,999	Beginner Yearly	Subscription
...
1078	184220	30/12/2021	1,499,999	Beginner Yearly	Subscription
1079	185431	30/12/2021	4,999,999	Agent Yearly	Subscription
1080	186675	31/12/2021	1,499,999	Beginner Yearly	Subscription

Kemudian 5 kolom tersebut terdiri dari Shop_ID, Date, Price_Sales, Product_Name dan Product_Type. Data penelitian ini adalah data AVANA Indonesia dengan atribut sesuai penjelasan pada Tabel 2 sebagai berikut.

Tabel 2. Keterangan Atribut

Atribut	Keterangan
Shop_ID	(Tipe Data: Char) berisi nomor ID pelanggan
Date	(Tipe Data: Integer) berisi tanggal transaksi
Price_Sales	(Tipe Data: Integer) berisi total transaksi
Product_Name	(Tipe Data: String) berisi nama produk
Product_Type	(Tipe Data: String) berisi tipe produk

Selanjutnya, file keluaran gabungan diubah dengan mengumpulkan peristiwa transaksi untuk setiap pelanggan. Deskripsi konversi dijelaskan pada Gambar 2 dibawah ini.



Gambar 2. Skema Analisis Model RFM

3.3 Data Preparation

Untuk pra-pemrosesan, penulis menggabungkan semua file yang terpisah menjadi satu file untuk memudahkan pemrosesan. Berikut adalah jumlah total data setelah digabungkan dan siap untuk preprocessing.

3.3.1 Data Selection and Grouping

Data berikut menunjukkan jumlah total data yang digunakan dalam penelitian sebanyak 1080 catatan transaksi. Namun data setelah diakumulasikan akan menjadi 997 record data sesuai penjelasan Gambar 3 dibawah ini.

```

Out[13]:
      Date  Shop_ID  Price_Sales
0  2021-01-02  184673      399999
1  2021-01-02  185676      1499999
2  2021-01-03  160626      399999
3  2021-01-03  184427      1499999
4  2021-01-03  184511      1499999

In [14]: data_new.shape, orders.shape
Out[14]: ((1080, 3), (997, 3))
  
```

Gambar 3. Total Data Keseluruhan

3.3.3 Removing Missing Values

Setelah pemilihan dan pengelompokan data dilakukan, urutkan data di setiap kolom untuk membersihkan dan menghilangkan nilai yang hilang pada data yang dipilih sesuai dengan proses pada Gambar 4 dibawah ini.

```
In [18]: data.drop("Unnamed: 0", axis=1, inplace=True)

KeyError                                Traceback (most recent call last)
--> 1 data.drop("Unnamed: 0", axis=1, inplace=True)

C:\ProgramData\Anaconda3\lib\site-packages\pandas\util\decorators.py in wrapper(*args, **kwargs)
    309         )
    310     )
--> 311     return func(*args, **kwargs)
    312
    313     return wrapper

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\indexes\base.py in drop(self, labels, errors)
    6659         if mask.any():
    6660             if errors != "ignore":
-> 6661                 raise KeyError(f"list(labels[mask]) not found in axis")
    6662             indexer = indexer[mask]
    6663             return self.delete(indexer)

KeyError: "[Unnamed: 0]" not found in axis"
```

Gambar 4. Proses Removing Missing Values

3.4 Modelling

Data yang sudah melalui tahap preparation pada tahap sebelumnya, kemudian akan ditransformasi ke dalam model RFM dan dilakukan clustering berbasis partisi yaitu dengan algoritma K-Means [19].

3.4.1 Pemodelan RFM

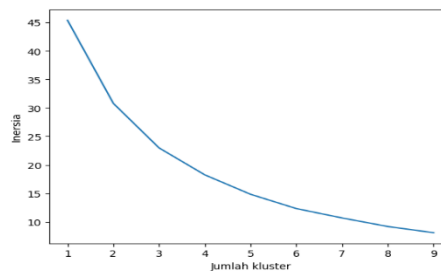
Pada tahaan pemodelan RFM dibawah ini merupakan proses penjumlahan dan penggabungan data dari recency, frequency dan monetary yang kemudian akan diolah untuk mendapatkan scoring RFM [20]. Pemodelan RFM merupakan dimana setiap nilai akan diubah ke dalam rentang nilai 1 sampai 4 sesuai dengan metode kuintil yang membagi data menjadi empat bagian, maka dari itu akan dilakukan tranformasi pada data ke dalam rentang 0 sampa 100 menggunakan teknik min-max normalization. Dibawah ini merupakan Tabel 3 hasil RFM beserta RFM scoring-nya.

Tabel 3. Hasil RFM

No	Shop ID	Recency	Frequency	Monetary	R	F	M	RFM_Score
1	22029	95	2	1,664,999	3	3	3	9
2	28981	284	1	4,999,999	1	1	4	6
3	29846	19	1	4,999,999	4	1	4	9
4	57873	163	2	2,999,998	2	3	3	8
5	65151	149	1	1,499,999	2	1	2	5
...
550	202852	48	1	4,999,999	4	1	4	9
551	202854	46	1	562,5	4	1	1	6
552	202899	44	1	562,5	4	1	1	6

3.4.2 Pemodelan Clustering Algoritma K-Means

Pada tahap pemodelan clustering algoritma k-means dibawah ini merupakan proses clustering berbasis partisi. Dimana pada proses hasil clusterisasinya menggunakan metode elbow untuk menentukan berapa cluster yang tepat pada data yang sedang diolah [21].



Gambar 5. Proses Clustering Algoritma K-Means

Dari Gambar 5 diatas bisa dilihat bahwa clustering yang dihasilkan adalah K=9 cluster. Dikarenakan data yang diolah menunjukkan 9 cluster yang diperoleh. Oleh sebab itu untuk mempermudah clustering penulis menentukan cluster menjadi K=4 Cluster, dimana 4 Cluster ini menjadi banyak rekomendasi bagi para pengguna algoritma K-Means Clustering [22]. Dibawah ini merupakan Tabel 4 yang menampilkan hasil dari Clustering K-Means.

Tabel 4. Hasil Clustering K-Means

No	Shop ID	Recency	Frequency	Monetary	Clusters
1	22029	95	2	1,664,999	1
2	28981	284	1	4,999,999	0
3	29846	19	1	4,999,999	3
4	57873	163	2	2,999,998	1
5	65151	149	1	1,499,999	0
...
550	202852	48	1	4,999,999	0
551	202854	46	1	562,5	3
552	202899	44	1	562,5	3

3.4.3 Pemodelan Recency Clustering

Pada tahap pemodelan Recency clustering algoritma k-means dibawah ini merupakan proses clustering berdasarkan recency. Dimana pada proses hasil clusterisasi-nya menggunakan metode elbow untuk menentukan berapa cluster yang tepat pada data yang sedang diolah dan menghasilkan beberapa cluster yang ditampilkan pada Tabel 5 dibawah ini yaitu warm dengan nilai 30.797101, active dengan nilai 28.804348, inactive dengan nilai 21.014493 terakhir cold dengan nilai 19.384058.

Tabel 5. Hasil Recency Clustering K-Means

No	Shop ID	Recency	Frequency	Monetary	Recency Clusters
1	22029	95	2	1,664,999	Warm
2	28981	284	1	4,999,999	Inactive
3	29846	19	1	4,999,999	Active
4	57873	163	2	2,999,998	Warm
5	65151	149	1	1,499,999	Warm
...
550	202852	48	1	4,999,999	Active
551	202854	46	1	562,5	Active
552	202899	44	1	562,5	Active

3.4.4 Pemodelan Recency Frequency (Customer Value) Clustering

Pada tahap pemodelan Recency Frequency (Customer Value) Clustering algoritma k-means dibawah ini merupakan proses clustering berdasarkan recency. Dimana pada proses hasil clusterisasi-nya menggunakan metode elbow untuk menentukan berapa cluster yang tepat pada data yang sedang diolah dan menghasilkan beberapa cluster yang ditampilkan pada Tabel 6 dibawah ini yaitu common dengan nilai 37.862319, Ultra High dengan nilai 27.355072, Low dengan nilai 20.108696 terakhir High dengan nilai 14.673913.

Tabel 6. Hasil Recency Frequency Clustering K-Means

No	Shop ID	Recency	Frequency	Monetary	Recency Frequency Clusters
1	22029	95	2	1,664,999	Ultra High
2	28981	284	1	4,999,999	Common
3	29846	19	1	4,999,999	Low
4	57873	163	2	2,999,998	Ultra High
5	65151	149	1	1,499,999	Common
...
550	202852	48	1	4,999,999	Common
551	202854	46	1	562,5	Low
552	202899	44	1	562,5	Low

3.5 Evaluation

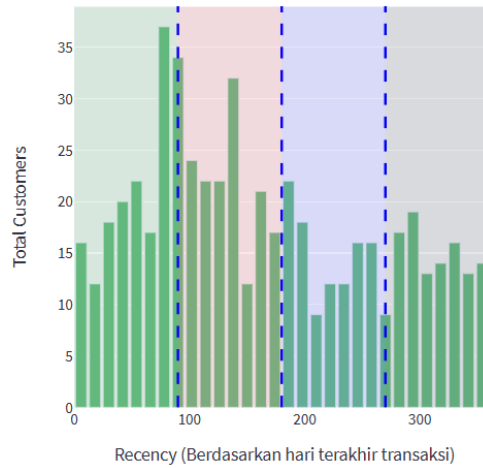
Hasil dari pengujian menampilkan 4 cluster dan segmentasi pelanggan yang terdiri dari recency, frequency, monetary dan kombinasinya. Cluster pertama merupakan recency dan recency frequency clustering yaitu warm, active, inactive, cold dan cluster kedua terdiri dari common, ultra-high, low, high.

3.6 Deployment

Data yang dihasilkan oleh pengelompokan menggunakan atribut recency, frequency dan monetary bersama dengan algoritma penambangan data berbasis partisi yang diterapkan sistem, k-means.

3.6.1 Recency Segmentation

Pada bar chart tampilan recency distribution ini menampilkan recency selama periode 1 tahun. Dan penulis mengkategorikan segmentasi pelanggan kedalam sebuah 4 cluster yaitu active customer, warm customer, cold customer dan inactive customer.



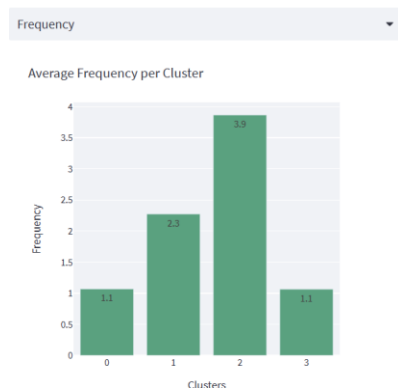
Gambar 6. Tampilan Recency Segmentation

Dari Gambar 6 diatas memberikan penjelasan segmentasi pelanggan berdasarkan recency seperti dibawah ini:

1. Active Customers: Pelanggan yang melakukan transaksi terakhir kurang dari atau sama dengan 90 hari.
2. Warm Customers: Pelanggan yang melakukan transaksi terakhir antara 90 dan 180 hari.
3. Cold Customers: Pelanggan yang melakukan transaksi terakhir antara 180 dan 270 hari.
4. Inactive Customers: Pelanggan yang melakukan transaksi terakhir lebih dari 270 hari.

3.6.2 Cluster Characteristics based on Frequency and Monetary (Customer Value Segmentation)

Pada bar chart tampilan frequency and monetary (customer value segmentation) distribution ini menampilkan recency selama periode 1 tahun. Dan penulis mengkategorikan segmentasi pelanggan kedalam sebuah 4 cluster yaitu active customer, warm customer, cold customer dan inactive customer.



Gambar 7. Diagram Batang Frequency Segmentataion



Gambar 8. Diagram Batang Monetary Segmentation

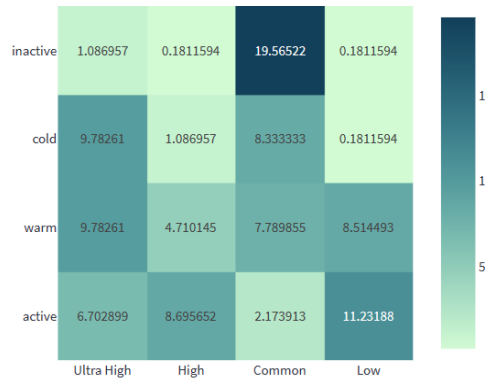
Pada Gambar 7 dan 8 diatas memberikan penjelasan customer value segmentation berdasarkan seperti dibawah ini:

1. Cluster 2 menunjukkan nilai tertinggi baik Frequency maupun Monetary, artinya pelanggan pada cluster ini memiliki daya beli yang sangat tinggi dan sangat sering melakukan transaksi. Jadi kami menamakannya sebagai pelanggan 'Ultra High'.
2. Cluster 1 menunjukkan sedikit lebih rendah dari Cluster 2 Frequency, artinya pelanggan pada cluster ini memiliki daya beli yang tinggi dan cukup sering melakukan transaksi jadi kami menamakannya sebagai pelanggan 'High'. Sedangkan untuk Cluster 0 Monetary menunjukkan nilai sedang baik artinya pelanggan pada cluster ini memiliki daya beli sedang dan jarang melakukan transaksi. Jadi kami menamakannya sebagai pelanggan 'Common'.
3. Cluster 0 dan 3 Frekuensi menunjukkan nilai sedang baik, artinya pelanggan pada cluster ini memiliki daya beli sedang dan jarang melakukan transaksi. Jadi kami menamakannya sebagai pelanggan 'Common'.
4. Cluster 1 monetary, menunjukkan nilai sedang baik artinya pelanggan pada cluster ini memiliki daya beli sedang dan jarang melakukan transaksi. Jadi, kami menamakannya sebagai pelanggan 'Common'.

5. Cluster 3 Monetary menunjukkan nilai terendah, artinya pelanggan pada cluster ini memiliki daya beli yang rendah dan jarang melakukan transaksi. Jadi kami menamakannya sebagai pelanggan 'Low'.

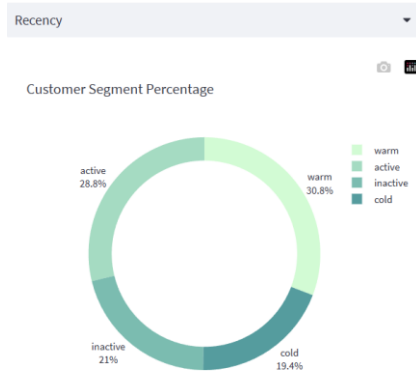
3.6.3 Segmentation Result

Pada diagram pohon persegi dibawah ini menampilkan tampilan recency cluster dan recency frequency (customer value) cluster selama periode 1 tahun. Dan penulis mengkategorikan segmentasi pelanggan kedalam sebuah 4 cluster dalam 2 bagian yaitu warm, active, inactive, cold dan common, ultra-high, low, high.

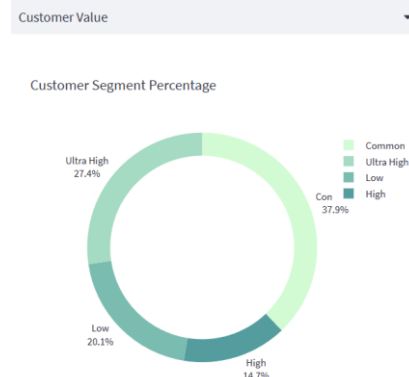


Gambar 9. Tampilan Segmentation Result

Pada Gambar 9 diatas menjelaskan terkait diagram pohon persegi dimana persegi yang warnanya semakin tepat adalah segmentasi customer yang paling banyak melakukan pembelian di AVANA Indonesia. Menjelaskan bahwa "Inactive" dan "Common" menunjukkan hasil segmentasi yang paling tinggi dengan nilai 19.56522.

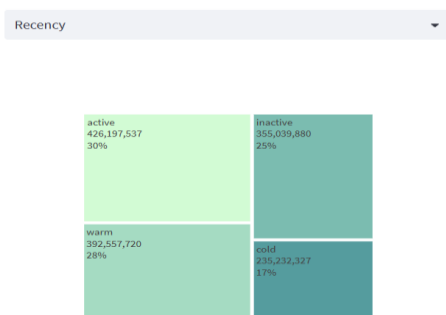


Gambar 10. Diagram Pie Frequency Segmentataion

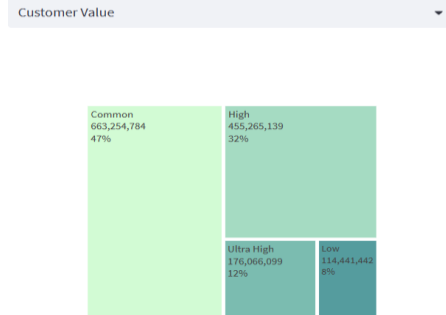


Gambar 11. Diagram Pie Customer Value Segmentation

Sedangkan, Gambar 10 diatas merupakan diagram pie yang menjelaskan persentase dari recency cluster yang terdiri dari beberapa cluster yaitu warm dengan 30.8 %, active dengan 28.8%, inactive dengan 21% dan cold dengan 19.4%. Selain itu ada Gambar 11 diatas merupakan diagram pie lain yang menjelaskan persentase dari recency frequency (customer value) cluster yang terdiri dari beberapa cluster yaitu common dengan 37.9 %, Ultra High dengan 27.4%, Low dengan 20.1% dan High dengan 14.7%.



Gambar 12. Diagram Persegi Panjang Monetary Based Recency Segmentation



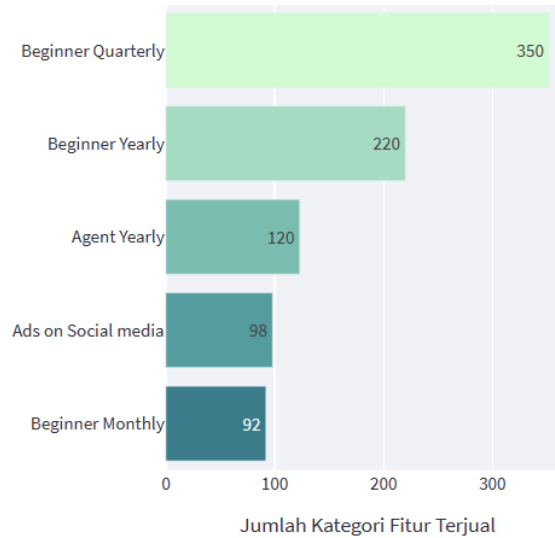
Gambar 13. Diagram Persegi Panjang Monetary Based Customer Value Segmentation

Pada Gambar 12 dan 13 diatas memberikan penjelasan business insight seperti dibawah ini:

1. Hasil segmentasi berdasarkan recency menunjukkan bahwa hampir 42% nasabah dan total pendapatan bersikap 'Cold' dan 'Inactive'. Sementara itu, pelanggan di area 'Active' dan 'Warm' memberikan pendapatan yang lebih baik dengan 58% dari nasabah dan total pendapatan.
2. Berdasarkan segmentasi customer value, hampir 55% pelanggan tetap berada di kategori 'Low' dan 'Common'. Ini sungguh tidak baik, sedangkan hanya 20% dari total pendapatan kami berasal dari pelanggan dengan Segmentasi 'Ultra High' dan 'High'.

3.6.4 Top 5 Products by Customers Segmentation

Dan tampilan yang terakhir pada Gambar 14 dibawah ini adalah diagram batang yang menjelaskan terkait 5 product teratas yang terjual berdasarkan recency dan customer value segmentation pada data yang telah diolah.



Gambar 14. Bar Chart Top 5 Products by Customer Segmentation

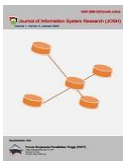
Dengan penjelasan customer value segmentation berdasarkan seperti dibawah ini; berdasarkan visualisasi yang diberikan, baik di halaman business overview maupun RFM. Terlihat bahwa dari segmentasinya, memang jenis fitur ini yang paling banyak diminati (bisa dilihat di halaman business overview). Jadi, dapat disimpulkan bahwa ketika orang menggunakan fitur di AVANA Indonesia, berikut adalah 5 fitur teratas yang paling banyak diminati.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, maka didapat beberapa kesimpulan, penelitian yang dilakukan menggunakan software Jupyter Notebook dengan metode K-Means dan model RFM dapat membantu mengetahui segmentasi recency dan customer value yang dimiliki dan menghasilkan bahwa 9 cluster menggunakan uji metode elbow. Dimana cluster yang ditampilkan tidak ada menunjukkan garis siku, sehingga penulis menentukan cluster dari beberapa penelitian yang paling relevan dengan 4 cluster. Cluster yang dihasilkan berupa 4 Cluster terdiri dari berbagai macam segmentasi. Yang pertama dari recency segmentasi menghasilkan cluster yaitu active customer, warm customers, cold customers dan inactive customers. Kemudian kedua dari recency frequency (customer value) segmentasi menghasilkan cluster yaitu common, ultra-high, low, high. Karakteristik yang dapat diamati yaitu Recency, Frequency, dan Monetary dapat menggambarkan karakteristik dari masing-masing cluster. Sehingga berdasarkan karakteristik yang berbeda dari masing-masing cluster, dapat disusun strategi pemasaran untuk meningkatkan pendapatan perusahaan. Hasil analisis yang divisualisasikan dengan bar chart, diagram pohon persegi, diagram pie dan diagram batang menggunakan Streamlit, aplikasi Python berbasis web. Fitur mana yang memudahkan membaca grafik, mengeksport data, dan mendistribusikan data. Serta informasi yang sudah dibuat dapat memberikan sebuah sistem rekomendasi bagi perusahaan terutama bagian pemasaran dan beberapa saran seperti keanggotaan loyalitas dari informasi yang dihasilkan sekaligus mengaktifkan social media untuk menjangkau pasar yang lebih luas.

REFERENCES

- [1] A. S. Al-Adwan and H. Kokash, "The driving forces of facebook social commerce," J. Theor. Appl. Electron. Commer. Res., vol. 14, no. 2, 2019, doi: 10.4067/S0718-18762019000200103.
- [2] C. H. Lee and C. W. Chen, "An empirical study of social commerce intention: An example of China," Inf., vol. 11, no. 2, 2020, doi: 10.3390/info11020099.



- [3] I. Mubarak, “12+ Strategi Pemasaran Efektif untuk Kemajuan Bisnis Anda,” <https://www.niagahoster.co.id/>, 2021.
- [4] M. Mohammadian and I. Makhani, “International Academic Journal of Accounting and Financial Management RFM-Based customer segmentation as an elaborative analytical tool for enriching the creation of sales and trade marketing strategies,” *Int. Acad. J. Account. Financ. Manag.*, vol. 6, no. 1, 2019.
- [5] 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, doi: 10.30813/jiems.v13i2.2274.
- [6] B. Alma, “Manajemen Pemasaran & Manajemen Pemasaran Jasa,” penerbit Alf., vol. 1, no. 1, 2018.
- [7] S. Subhash, K. Sneha, A. Ullas, and D. Raj, “A Covid-19 Safety Web Application to Monitor Social Distancing and Mask Detection,” in *IEEE Region 10 Humanitarian Technology Conference, R10-HTC, 2021*, vol. 2021-Septe. doi: 10.1109/R10-HTC53172.2021.9641671.
- [8] 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, 2018, doi: 10.21460/jutei.2018.21.76.
- [9] N. Puspitasari, J. A. Widians, and N. B. Setiawan, “Customer segmentation using bisecting k-means algorithm based on recency, frequency, and monetary (RFM) model,” *J. Teknol. dan Sist. Komput.*, vol. 8, no. 2, 2020, doi: 10.14710/jtsiskom.8.2.2020.78-83.
- [10] A. T. Widiyanto and A. Witanti, “Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global),” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 1, no. 1, 2021, doi: 10.24002/konstelasi.v1i1.4293.
- [11] S. S. Prasetyo, M. Mustafid, and A. R. Hakim, “PENERAPAN FUZZY C-MEANS KLUSTER UNTUK SEGMENTASI PELANGGAN E-COMMERCE DENGAN METODE RECENCY FREQUENCY MONETARY (RFM),” *J. Gaussian*, vol. 9, no. 4, 2020, doi: 10.14710/j.gauss.v9i4.29445.
- [12] S. I. Murpratiwi, I. G. Agung Indrawan, and A. Aranta, “ANALISIS PEMILIHAN CLUSTER OPTIMAL DALAM SEGMENTASI PELANGGAN TOKO RETAIL,” *J. Pendidik. Teknol. dan Kejuru.*, vol. 18, no. 2, 2021, doi: 10.23887/jptk-undiksha.v18i2.37426.
- [13] A. P. Pramudiansyah, “SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN MODEL RECENCY FREQUENCY MONETARY,” *J. Ilm. ILMU Komput.*, vol. 7, no. 2, 2021, doi: 10.35329/jiik.v7i2.201.
- [14] N. S. Sasue and A. F. Wijaya, “PERENCANAAN STRATEGIS SISTEM INFORMASI MENGGUNAKAN ENTERPRISE ARCHITECTURE PLANNING (EAP) FRAMEWORK,” *J. Bina Komput.*, vol. 2, no. 2, 2020, doi: 10.33557/binakomputer.v2i2.919.
- [15] I. A. Nikmatun and I. Waspada, “Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor,” *J. SIMETRIS*, vol. 10, no. 2, 2019.
- [16] R. W. Sembiring Brahmata, F. A. Mohammed, and K. Chairuang, “Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, 2020, doi: 10.24843/lkjiti.2020.v11.i01.p04.
- [17] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction To Data Mining Second Edition*. 2019.
- [18] İ. SABUNCU, E. TÜRKAN, and H. POLAT, “CUSTOMER SEGMENTATION AND PROFILING WITH RFM ANALYSIS,” *Turkish J. Mark.*, vol. 5, no. 1, 2020, doi: 10.30685/tujom.v5i1.84.
- [19] U. Rusdiana, I. Ernawati, N. Falih, and A. Arista, “Comparison of Distance Metrics on Fuzzy C-Means Algorithm Through Customer Segmentation,” in *Proceedings - 3rd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2021, 2021*. doi: 10.1109/ICIMCIS53775.2021.9699206.
- [20] D. L. Aditya and D. Fitriana, “COMPARATIVE STUDY OF FUZZY C-MEANS AND K-MEANS ALGORITHM FOR GROUPING CUSTOMER POTENTIAL IN BRAND LIMBACK,” *J. Ris. Inform.*, vol. 3, no. 4, 2021, doi: 10.34288/jri.v3i4.241.
- [21] Y. R. Rahadian and B. Syairudin, “Segmentation Analysis of Students in X Course with RFM Model and Clustering,” *J. Sos. Hum.*, vol. 0, no. 1, 2020, doi: 10.12962/j24433527.v0i1.6776.
- [22] H. Humaira and R. Rasyidah, “Determining The Appropriate Cluster Number Using Elbow Method for K-Means Algorithm,” 2020. doi: 10.4108/eai.24-1-2018.2292388.