

ALGORITMA K-MEANS CLUSTERING UNTUK SEGMENTASI PELANGGAN (STUDI KASUS : FASHION VIRAL SOLO)

Rifal Bayu Ardi¹, Faulinda Ely Nastiti², Sri Sumarlinda³

^{1,2,3}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta
Email: 19010153@fikom.udb.ac.id

ABSTRACT

Fashion Viral Solo is one of the distributors of fashion equipment in Surakarta which has not utilized its data to segment and classify customers based on their similarities. Segmentation is the process of finding out the characteristics of customers based on certain similarities, making it easier to collect information about customers that provide benefits to the company. Segmentation is a strategy to face business competition, to retain customers, and to assist management in developing promotional strategies to increase sales. The purpose of this research is to segment customers who have transacted at Fashion Viral Solo based on their characteristics. In this study, researchers used the K-Means algorithm to perform clustering and apply the Recency, Frequency and Monetary methods. K-Means is a simple algorithm, easy to implement, not slow, easy to adjust, and often used in data mining processes, especially clustering. The value of the Current attribute indicates the last transaction time, Frequency indicates the number of transactions, and Monetary indicates the total nominal value of the transaction. By using K-Means and Recency, Frequency and Monetary methods, researchers can segment Fashion Viral Solo's customers. From this study, the results obtained were 2 customer clusters. Cluster 1 with the highest number of members, namely 343 customers and cluster 2 with 8 customers. To determine the most optimal number of clusters, researchers used the Silhouette Scores method and the results obtained when grouping customers into 2 clusters were the most optimal.

Keywords: Segmentation, Clustering, K-Means, Customers

ABSTRAK

Fashion Viral Solo merupakan salah satu distributor perlengkapan fashion di Surakarta yang belum memanfaatkan data yang dimiliki untuk melakukan segmentasi dan mengelompokkan pelanggan berdasarkan kemiripannya. Segmentasi adalah proses untuk mencari tahu karakteristik dari pelanggan berdasarkan kesamaan tertentu, sehingga memudahkan pengumpulan informasi tentang pelanggan yang memberikan profit bagi perusahaan. Segmentasi menjadi salah satu strategi untuk menghadapi persaingan usaha, untuk mempertahankan pelanggan, dan untuk membantu manajemen dalam menyusun strategi promosi untuk meningkatkan penjualan. Tujuan dari penelitian ini adalah melakukan segmentasi pelanggan yang pernah bertransaksi di Fashion Viral Solo berdasarkan karakteristiknya. Pada penelitian ini, peneliti menggunakan algoritma K-Means untuk melakukan clustering dan penerapan metode Recency, Frequency dan Monetary. K-Means adalah algoritma sederhana, mudah diimplementasikan, tidak lambat, mudah disesuaikan, dan sering digunakan dalam proses data mining khususnya clustering. Nilai atribut Recency menunjukkan waktu terakhir transaksi, Frequency menunjukkan jumlah transaksi, dan Monetary menunjukkan total nominal transaksi. Dengan menggunakan K-Means dan metode Recency, Frequency dan Monetary, peneliti dapat melakukan segmentasi pelanggan Fashion Viral Solo. Dari penelitian ini, diperoleh hasil 2 cluster pelanggan. Cluster 1 dengan jumlah anggota terbanyak, yakni 343 pelanggan dan cluster 2 dengan jumlah anggota 8 pelanggan. Untuk menentukan jumlah cluster yang paling optimal, peneliti menggunakan metode Silhouette Scores dan diperoleh hasil jika membagi pelanggan menjadi 2 cluster adalah yang paling optimal.

Kata Kunci: Segmentasi, Clustering, K-Means, Pelanggan

Riwayat Artikel :

Tanggal diterima : 06-05-2023

Tanggal revisi : 10-05-2023

Tanggal terbit : 10-05-2023

DOI :

<https://doi.org/10.31949/infotech.v9i1.5214>

INFOTECH journal by Informatika UNMA is licensed under CC BY-SA 4.0

Copyright © 2023 By Author



1. PENDAHULUAN

1.1. Latar Belakang

Segmentasi pelanggan adalah strategi untuk membagi pelanggan berdasarkan karakteristik dan ciri yang sama. Pembagian tersebut berdasarkan pada minat atau permintaan pelanggan dengan ciri yang serupa. Dengan melakukan segmentasi pelanggan, dapat membantu perusahaan untuk memasarkan produknya secara pendekatan individual agar sesuai dengan kebutuhannya.

Menurut (Febriani & Putri, 2020) segmentasi pelanggan dipakai untuk mengevaluasi nilai dari setiap pelanggan sehingga perusahaan dapat mengidentifikasi pelanggan yang memberikan keuntungan yang besar dan pelanggan yang tidak. Menurut (Fitriyani & Hendriyani, 2021), dengan mengelola dan memanfaatkan *database* pelanggan dengan tepat, dapat membantu manajemen untuk mendapatkan informasi seperti : mengetahui berbagai kebutuhan pelanggan, memberikan penawaran khusus, memudahkan dalam membuat program loyalitas, dan membantu dalam segmentasi pelanggan. Informasi yang diperoleh dari pengolahan data penjualan, dapat dijadikan sebagai dasar oleh manajemen untuk mengembangkan usaha.

Fashion viral solo merupakan salah satu penyedia perlengkapan fashion di kota surakarta (solo), yang bergerak di bidang distribusi dan penjualan perlengkapan fashion wanita seperti tas, pakaian, sepatu, dan lainnya. Untuk mencapai tujuan bisnisnya, fashion viral solo memiliki bagian-bagian yang mendukung setiap aktivitas bisnis yang berjalan. Salah satu dari bagian yang dimiliki oleh fashion viral solo adalah bagian penjualan. Pada fashion viral solo, bagian penjualan memiliki peranan yang penting yakni melakukan penjualan kepada pelanggan, menerima transaksi order, dan merekap transaksi penjualan.

Dari hasil wawancara dengan pemilik fashion viral solo, saat ini data pelanggan yang dimiliki hanya dimanfaatkan secara terbatas. Data pelanggan yang paling banyak dimanfaatkan adalah nomor pribadi pelanggan yang dipakai untuk membuat grup dan melakukan broadcast promosi. Data lain yang diperoleh dari pelanggan seperti : kapan terakhir pelanggan melakukan pemesanan, jumlah transaksi dalam waktu tertentu, dan total nominal transaksi belum diolah secara optimal, khususnya untuk melakukan segmentasi pelanggan.

Data *mining* adalah sebuah proses menggali dan mengolah informasi penting dari kumpulan data. Proses penggalian informasi melibatkan teknik seperti memanfaatkan kecerdasan buatan, statistik, matematika, *machine learning*, dan lainnya. Teknik yang dipakai memiliki tujuan untuk mengenali dan mengekstrak informasi yang bermanfaat dari suatu *database* yang besar. Saat ini data *mining* telah

menjangkau dan juga dimanfaatkan pada banyak bidang keilmuan lainnya. (Sudarsono Et Al., 2021).

Menurut (Sudarsono et al., 2021) salah satu tugas umum dari data *mining* adalah melakukan *clustering*. Dalam melakukan *clustering*, maka diperlukan algoritma yang mampu melakukan pengelompokan data objek yang memiliki tingkat kesamaan antar kelas objek. Dari pengertian tersebut maka dapat ditarik kesimpulan jika model *clustering* cocok untuk mengelompokkan pelanggan berdasarkan tingkat kesamaannya menjadi beberapa *cluster* atau segmen.

Algoritma yang sering diterapkan untuk melakukan *clustering* salah satunya adalah K-Means. Algoritma K-Means bertujuan mengelompokkan data berdasarkan tingkat kemiripannya. K-means *clustering* adalah salah satu bentuk unsupervised learning, yang artinya tidak memerlukan data *testing*. Salah satu kelebihan algoritma k-means dibandingkan algoritma lainnya adalah sederhana, mudah diimplementasikan, tidak lambat, mudah disesuaikan, dan sering digunakan dalam proses data *mining* (Sari et al., 2020).

Sebagai upaya untuk melakukan segmentasi pelanggan fashion viral solo, maka dibutuhkan model data *mining* yang dapat membantu dalam segmentasi pelanggan. Tujuannya agar data yang telah didapatkan oleh fashion viral solo dapat dimanfaatkan secara lebih optimal, dapat mengelompokkan pelanggan berdasarkan *clustering*, membangun hubungan baik dengan pelanggan berdasarkan segmennya, dan membangun strategi yang tepat untuk meningkatkan penjualan.

Atribut yang digunakan dalam penelitian ini yakni berdasarkan analisis nilai RFM (*Recency*, *Frequency*, dan *Monetary*). Nilai *recency* menunjukkan jumlah hari yang telah berlalu sejak pelanggan terakhir melakukan transaksi. Nilai *frequency* menunjukkan jumlah transaksi yang telah dilakukan pelanggan dalam kurun waktu tertentu. Nilai *monetary* menunjukkan total pembayaran yang telah dilakukan pelanggan.

1.2. Tinjauan Pustaka

Penelitian yang dilakukan oleh (Febriani & Putri, 2020) dengan judul “Segmentasi Konsumen Berdasarkan Model *Recency*, *Frequency*, *Monetary* dengan Metode K-Means” Pada penelitian ini, peneliti melakukan segmentasi kategori konsumen dengan menggunakan metode K-Means clustering dan model RFM (*Recency*, *Frequency*, *Monetary*). Data yang digunakan oleh peneliti adalah data riwayat transaksi di Laundry XYZ. Berdasarkan analisa dan pengolahan data, didapatkan hasil konsumen yang terbagi menjadi 3 *cluster* segmen.

Penelitian yang dilakukan oleh (Setiawan et al., 2021) dengan judul “Perancangan Segmentasi Pelanggan Dengan Metode Clustering K-Means Dan Model RFM Pada Klinik Kecantikan Seoul Secret”. Pada penelitian ini, permasalahan yang dihadapi adalah penurunan jumlah pelanggan yang menggunakan paket yang disediakan objek penelitian dalam kurun waktu 3 tahun terakhir. Dari permasalahan tersebut, peneliti melakukan analisis dengan mengelola data pelanggan menggunakan metode K-Means dengan model RFM. Hasil yang didapatkan berupa segmentasi pelanggan yang terbagi menjadi 4 cluster dan usulan strategi pemasaran yang berbeda pada setiap cluster yang telah disegmentasikan.

Penelitian yang dilakukan oleh (Herdiana & Solihudin, 2023) dengan judul “Segmentasi Pelanggan salon Nuii Beauty Glow menggunakan K-Means Clustering”. Pada penelitian ini, objek penelitian memiliki masalah belum dapat mengklasifikasikan pelanggan yang memberikan keuntungan bagi perusahaan. Perusahaan mengalami kesulitan untuk menentukan strategi pemasaran yang tepat bagi pelanggannya. Peneliti menggunakan metode clustering data mining yang digabungkan dengan model RFM untuk menentukan karakteristik kelompok pelanggan yang terbentuk. Dari penelitian diperoleh 3 kategori kelompok pelanggan dengan karakteristik tertentu.

1.3. Metodologi Penelitian

a. Jenis dan Sumber Data

- a.) Jenis data yang akan dipakai adalah data kuantitatif. Data kuantitatif yakni data yang dapat dihitung, berupa angka atau nominal. Data rekap transaksi penjualan di Fashion Viral Solo termasuk jenis data kuantitatif karena berwujud angka atau nominal dan dapat dihitung.
- b.) Sumber data yang akan dipakai dalam penelitian ini adalah data primer dan data sekunder. Data primer adalah data yang diperoleh peneliti secara langsung dari objek penelitian dan digunakan untuk tujuan khusus. Jenis data primer pada penelitian ini, akan diperoleh melalui pengumpulan dan merekap data transaksi pelanggan Fashion Viral Solo dalam kurun waktu tertentu untuk dijadikan dataset.

b. Metode Pengumpulan Data

- a.) Wawancara
Penulis mengajukan pertanyaan dengan tanya jawab secara langsung kepada responden yang akan memberikan data. Wawancara dilakukan kepada pemilik dan bagian penjualan di Fashion Viral Solo.
- b.) Observasi
Metode observasi dilakukan dengan cara datang langsung pada lokasi objek penelitian, Data-data dan informasi yang diperoleh bersumber dari pengamatan langsung dari proses bisnis yang saat ini

sudah berjalan di objek penelitian terkait, khususnya pada bagian penjualan Fashion Viral Solo.

c.) Dokumentasi

Dokumentasi adalah metode pengumpulan data dengan mengumpulkan catatan atau dokumen yang penulis dapatkan dari metode wawancara dan observasi dengan pihak yang berkaitan dengan penelitian.

d.) Studi Kepustakaan

Studi Kepustakaan yakni penulis memperoleh dan mempelajari data dan informasi dari berbagai penelitian sebelumnya yang mempunyai kemiripan dengan permasalahan yang sedang dihadapi, yakni data mining dengan algoritma K-Means clustering untuk melakukan segmentasi pelanggan.

c. Data Mining Life Cycle

Dalam melakukan data mining, penulis menggunakan metode dari KDD (*Knowledge Discovery in Database*). Menurut (Sari, 2023) tahapan dalam KDD dapat dijelaskan sebagai berikut :

- a.) Data *cleaning*/pembersihan data, yakni tahap data diolah lalu memilih data yang dianggap dapat dipakai.
- b.) Data *integration*/integrasi data, yakni tahap mengintegrasikan data yang berulang dan digabungkan menjadi satu.
- c.) Data *selection*/seleksi data, yakni pemilihan data yang dianggap relevan untuk dilakukan analisis.
- d.) Data *transformation*/transformasi data, yakni proses mengubah data ke dalam bentuk prosedur mining.
- e.) Data *mining*/penggalian data, yakni proses mengekstrak pola potensial untuk menghasilkan data yang bermanfaat. Pada proses ini, penulis menggunakan algoritma K-Means clustering untuk mengelompokkan pelanggan menjadi cluster berdasarkan tingkat kemiripannya. Dalam proses clustering, penulis memakai model data mining yang dibuat menggunakan aplikasi Orange Data Mining. Menurut (Lestari & Sumarlinda, 2021) Algoritma K-Means sensitif terhadap cluster awal yang dipilih secara acak. Algoritma ini dapat dijalankan berulang untuk mendapatkan hasil yang optimal. K-Mean merupakan algoritma sederhana yang telah dipakai ke banyak domain masalah dan kandidat yang baik untuk bekerja pada *data points* yang dibuat secara acak. Menurut (Sulistiyawati & Supriyanto, 2021) langkah dalam K-Means *clustering* dapat dijelaskan sebagai berikut :

1. Terlebih dulu menentukan nilai K sebagai banyaknya kluster yang ingin dibuat.
2. Menginisiasi K pusat *cluster* dengan cara acak dari data yang ada.
3. Melakukan perhitungan jarak setiap data yang ada terhadap masing-masing pusat kluster menggunakan rumus Euclidean Distance sehingga didapatkan jarak yang paling pendek dari setiap data dengan pusat *cluster*. Menurut (Jabbar, 2022) Euclidean Distance dapat dirumuskan dengan persamaan 1 berikut :

$$d(xi, \mu_j) = \sqrt{\sum(xi, \mu_j)^2} \quad (1)$$

d adalah titik data

xi adalah data kriteria

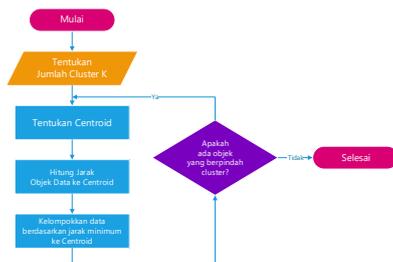
μ_j adalah centroid pada *cluster* ke- j

4. Melakukan klasifikasi setiap data dengan dasar kedekatan dengan pusat *cluster*.
5. Memperbarui nilai pusat *cluster* yang merupakan rata-rata *cluster* yang bersangkutan dengan menggunakan rumus persamaan 2 berikut :

$$d(xi, \mu_j) = \sqrt{\sum(xi, \mu_j)^2} \quad (2)$$

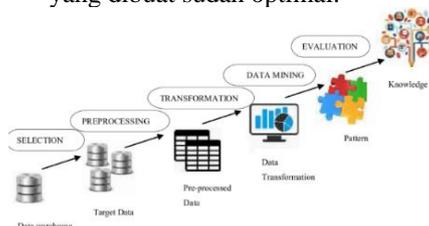
$\mu_j (t+1)$ adalah centroid yang baru pada iterasi ke $(t+1)$ dan N_{sj} merupakan banyaknya data pada *cluster* s_j .

6. Melakukan pengulangan langkah 2 sampai 5 hingga, anggota setiap cluster tidak berubah. Proses pada K-Means *clustering* dapat digambarkan pada diagram alir berikut :



Gambar 1. Diagram Alir Algoritma K-Means

- f.) *Patterns Evaluation*/evaluasi pola, yakni proses dimana pola-pola yang telah diidentifikasi, dilakukan evaluasi. Penulis menggunakan metode Silhouette Score untuk menentukan apakah jumlah *cluster* yang dibuat sudah optimal.



Gambar 2. Proses Knowledge Discovery in Database

2. PEMBAHASAN

2.1. Pengumpulan & Pembersihan Data

Data transaksi dalam kurun waktu tertentu akan dikumpulkan untuk diolah. Penulis memakai data rekap transaksi online bulan Februari 2023 yang diperoleh dari bagian penjualan Fashion Viral Solo. Pembersihan data bertujuan untuk untuk menghapus data dan atribut yang hilang (Sumarlinda & Lestari, 2022). Penulis mengumpulkan sejumlah 851 rekap transaksi selama periode bulan Februari 2023.

	A	B	C	D	E
1	Pelanggan	Tanggal Transaksi	Jumlah Transaksi	Bank	Admin
2	YUNITA	01/02/2023	46000	BCA	MILA
3	ILHAM	01/02/2023	20000	BCA	MILA
4	SONIA	01/02/2023	50000	BCA	MILA
848	PUTRI	28/02/2023	81000	BCA	NANDA
849	VINA	28/02/2023	105000	BCA	NANDA
850	IDAWATI	28/02/2023	145000	MANDIRI	NANDA
851	SRI P	28/02/2023	109000	MANDIRI	NANDA

Gambar 3. Data Transaksi Fashion Viral Solo

2.2. Integrasi Data/Data Integration

Data transaksi akan diintegrasikan dengan cara menggabungkan data transaksi dari pelanggan yang sama, yang melakukan transaksi lebih dari sekali. Sebelum digabungkan, terlebih dahulu penulis mencatat tanggal terakhir pelanggan melakukan transaksi, penulis menghitung jumlah transaksi yang dilakukan oleh masing-masing pelanggan yang bersangkutan, dan menjumlahkan nominal total belanja pelanggan dalam periode yang dianalisis yakni bulan Februari 2023.

	A	B	C	D	E
1	Pelanggan	Tanggal Transaksi	Jumlah Transaksi	Bank	Admin
21	DIAH	01-02-2023	124000	BCA	NANDA
22	DIAH	01-02-2023	38000	BCA	NANDA
196	YUNI	07-02-2023	1096500	MANDIRI	MILA
197	YUNI	07-02-2023	760000	MANDIRI	MILA

Gambar 4. Transaksi Dari Pelanggan Yang Sama

2.3. Seleksi Data/Data Selection

Peneliti melakukan pemilihan data yang dianggap relevan yang bisa dijadikan sebagai variabel untuk dilakukan analisis. Atribut yang akan digunakan dalam penelitian ini diambil dari atribut nama pelanggan, tanggal transaksi, dan jumlah transaksi.

	A	B	C	
1	Pelanggan	Tanggal Transaksi	Jumlah Transaksi	
2	YUNITA	01-02-2023	46000	
3	ILHAM	01-02-2023	20000	
4	SONIA	01-02-2023	50000	
5	DESITA	01-02-2023	211000	
6	NINA	01-02-2023	226000	

Gambar 5. Seleksi Rekap Data Transaksi

2.4. Transformasi Data/Data Transformation

Pada tahap ini, data diubah ke dalam bentuk yang bisa diolah untuk data mining. Penulis membuat 3 atribut baru yakni Recency (hari), Frequency (jumlah transaksi), dan Monetary (total belanja). Nilai Recency menunjukkan jumlah hari yang telah berlalu sejak pelanggan terakhir melakukan

transaksi. Nilai Frequency menunjukkan jumlah transaksi yang telah dilakukan pelanggan dalam kurun waktu tertentu. Nilai Monetary menunjukkan total pembayaran yang telah dilakukan pelanggan. Menurut (Adiana et al., 2018) untuk dapat dilakukan pemodelan RFM pada dataset, maka diperlukan atribut yang dapat mewakili setiap kriteria tersebut. Berikut adalah penjelasan untuk masing-masing atribut :

a. Atribut Recency (R) :

Atribut Recency menghitung waktu yang telah berlalu sejak terakhir pelanggan melakukan transaksi dengan waktu penelitian yang dilakukan. Untuk menggambarkan atribut ini, diperlukan fungsi yang menunjukkan adanya jarak waktu transaksi pelanggan dengan waktu penelitian. Salah satu cara yang dapat dipakai untuk menghitung nilai Recency dengan menghitung selisih hari yang telah berlalu antara tanggal terakhir pelanggan melakukan transaksi dengan waktu penelitian.

b. Atribut Frequency (F) :

Atribut *frequency* menghitung seberapa sering pelanggan melakukan transaksi selama periode penelitian yang dianalisis. Nilai atribut ini didapatkan dengan menghitung banyaknya transaksi yang dilakukan oleh pelanggan selama waktu penelitian. Atribut *frequency* dapat dihitung dengan menjumlahkan transaksi yang dilakukan oleh setiap pelanggan dengan menggunakan rumus CountIF Excel.

c. Atribut Monetary (M) :

Atribut Monetary menghitung seberapa banyak jumlah pembayaran yang telah dikeluarkan oleh pelanggan selama waktu penelitian. Nilai atribut ini didapatkan dengan menghitung total jumlah pembayaran yang dikeluarkan oleh setiap pelanggan dengan menggunakan rumus SumIF Excel.

Pada tahap transformasi data, peneliti juga melakukan penggabungan data transaksi. Data transaksi yang semula berjumlah 851 data dengan pelanggan yang sama dicatat berulang, setelah dilakukan penggabungan menjadi 351 data transaksi dari 351 pelanggan. Penggabungan bertujuan menghindari *redundancy* data dan agar setiap data mewakili satu identitas pelanggan yang unik agar didapatkan hasil yang optimal setelah dilakukan proses *clustering*.

Pelanggan	Recency (hari)	Frequency	Monetary
YUNITA	68	6	538000
ILHAM	68	2	35000
SONIA	68	1	50000
DESITA	68	6	644000
NINA	68	27	4004083

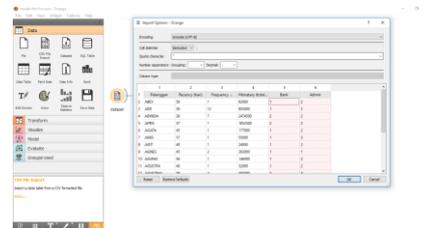
Gambar 6. Pembuatan Atribut Recency, Frequency, dan Monetary

	A	B	C	D
	Pelanggan	Recency (hari)	Frequency (jumlah transaksi)	Monetary (total belanja)
2	ABDI	58	1	92000
3	ADE	38	12	805000
350	ZULFA	38	2	183000
351	ZUROLA	52	1	128000
352	ZUSILA	59	3	212000

Gambar 7. Data Transaksi Setelah Digabungkan

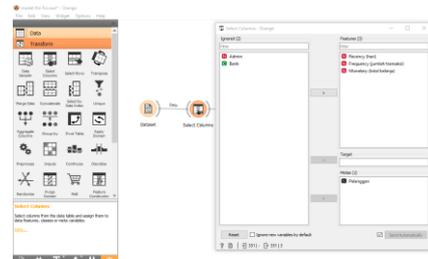
2.5. Penggalan Data/Data Mining

Pada penelitian ini, data *mining* dilakukan menggunakan algoritma K-Means *clustering* dengan model yang dibangun menggunakan aplikasi Orange Data Mining. Hal pertama yang perlu dilakukan adalah menambahkan Widget Dataset pada lembar kerja dan mengimpor dataset yang sebelumnya telah disiapkan ke dalam aplikasi Orange Data Mining seperti yang ditunjukkan pada gambar di bawah ini.



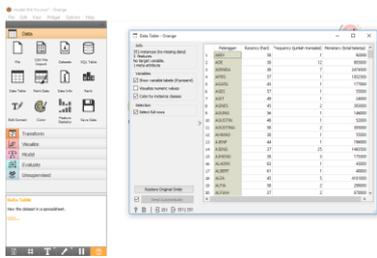
Gambar 8. Proses Import Dataset

Berikutnya peneliti menambahkan widget Select Column. Hubungkan widget Dataset dengan widget Select Column dengan cara menarik garis agar kedua widget terhubung. Pada langkah ini, atribut admin dan bank akan diabaikan karena atribut yang dipakai untuk melakukan *clustering* adalah nilai Recency, Frequency, dan Monetary.

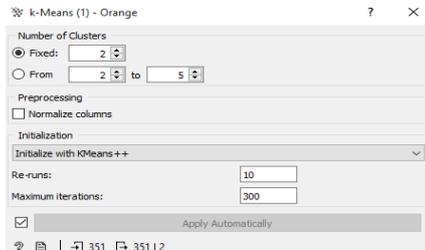


Gambar 9. Proses Select Column

Berikutnya peneliti menambahkan widget Data Table untuk melakukan preview dari data yang akan diolah. Setelah dipastikan tidak ada data yang kosong, berikutnya ditambahkan widget K-Means yang dihubungkan dengan widget Data Table untuk dilakukan *clustering* menggunakan algoritma K-Means. Pilih menu inialisasi dengan pengaturan KMeans+, kemudian aktifkan menu *maximum iterations*, dengan jumlah *maximum iterations* sebanyak 300.



Gambar 10. Preview Data



Gambar 11. Pengaturan Widget K-Means

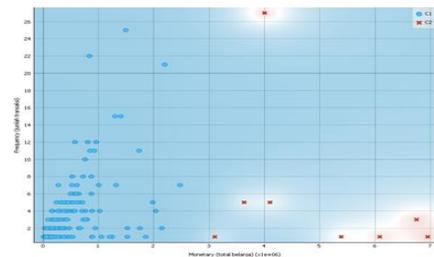
Setelah dilakukan clustering menggunakan Algoritma K-Means dengan pengelompokkan data menjadi 2 kategori cluster dan *maximum iterations* sebanyak 300, didapatkan hasil jika anggota dari cluster 1 sejumlah 343 data, sedangkan anggota dari cluster 2 sejumlah 8 data. Keanggotaan dari setiap cluster dapat dilihat secara lengkap pada tabel 1.

Tabel 1. Keanggotaan Setiap Cluster

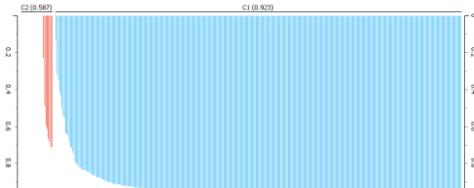
No	Cluster	Anggota Cluster
1	Cluster 1	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,137,138,139,140,141,142,143,144,145,146,147,148,149,150,151,152,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,182,183,184,185,186,187,188,189,190,191,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,208,209,210,211,212,214,215,216,217,218,219,220,221,222,223,224,225,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,243,244,245,246,247,248,249,250,251,252,253,254,255,256,257,258,259,260,261,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277

No	Cluster	Anggota Cluster
		,278,279,280,281,282,283,284,285,286,287,288,289,290,291,292,293,294,295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,331,332,333,334,335,336,337,338,339,340,341,342,343,344,345,346,347,348,349,350,351. (343)
2	Cluster 2	18, 67, 98, 136, 153, 192, 213, 330. (8)

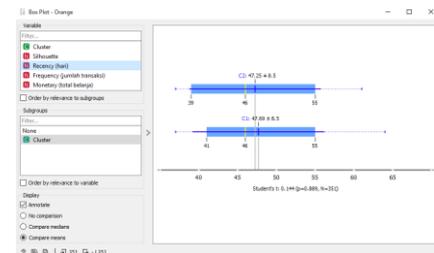
Selanjutnya dilakukan penyajian hasil *clustering* dengan 3 metode yakni Scatter Plot pada gambar 12, Silhouette Plot pada gambar 13, dan Box Plot pada gambar 14, 15, dan 16.



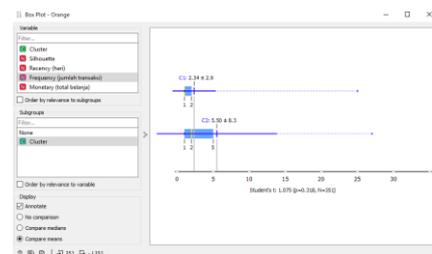
Gambar 12. Scatter Plot Hasil Clustering



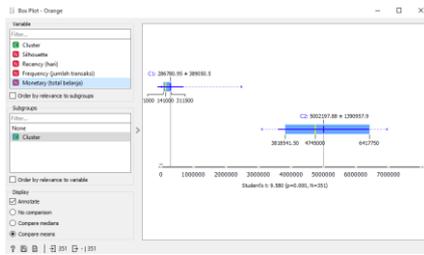
Gambar 13. Silhouette Plot Hasil Clustering



Gambar 14. Box Plot Atribut Recency



Gambar 15. Box Plot Atribut Frequency



Gambar 16. Box Plot Atribut Monetary

Tabel 2. Nilai Centroid Setiap Cluster

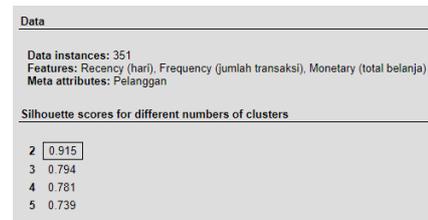
Cluster	Recency	Frequency	Monetary	Rank
C1	47,69	2,34	Rp. 286.780	2
C2	47,25	5,50	Rp. 5.002.197	1

Berdasarkan hasil dari Tabel 2 dapat disimpulkan jika cluster 2 merupakan pelanggan dengan peringkat 1 yang memiliki nilai rata-rata variabel Monetary sebesar Rp. 5.002.197, nilai Frequency 5,50, dan nilai Recency 47,25. Sedangkan cluster 1 merupakan pelanggan dengan peringkat 2 yang memiliki nilai rata-rata variabel Monetary sebesar Rp. 286.780, nilai Frequency 2,34, dan nilai Recency 47,69.

Data transaksi penjualan Fashion Viral Solo sebanyak 351 setelah dilakukan clustering dengan model RFM dihasilkan segmentasi pelanggan pada setiap cluster sesuai dengan karakteristik dari pelanggan tersebut. Berdasarkan model RFM, pelanggan yang termasuk pada kategori loyal terdapat pada cluster 1 dengan jumlah 8 pelanggan karena memiliki ciri nilai Recency yang paling dekat dengan hari dilakukan penelitian, memiliki jumlah Frequency (transaksi) yang paling tinggi, dan Monetary (total belanja) yang paling besar. Sedangkan pada cluster 2 dengan jumlah 343 pelanggan dikategorikan sebagai pelanggan kurang potensial, karena memiliki ciri karakteristik nilai Frequency dan Monetary yang cukup jauh dibandingkan dengan pelanggan pada cluster 1.

2.6. Evaluasi Pola/Pattern Evaluation

Peneliti melakukan evaluasi dengan menggunakan metode Silhouette Score untuk mengetahui apakah jumlah cluster yang telah ditentukan adalah yang paling optimal. Semakin tinggi nilai dari Silhouette Score, maka akan semakin baik hasil cluster-nya. Penulis menentukan jumlah cluster antara 2 hingga 5 cluster dan hasilnya cluster dengan jumlah 2 memiliki score sebesar 0,915 yang merupakan score tertinggi dibandingkan jumlah cluster yang lain. Dari evaluasi cluster tersebut, dapat disimpulkan jika clustering dengan K sejumlah 2 adalah yang paling optimal. Hasil dari evaluasi Silhouette Score dapat dilihat pada gambar di bawah ini :



Gambar 17. Evaluasi Silhouette Score

3. KESIMPULAN

Model clustering menggunakan algoritma K-Means dapat dipakai untuk melakukan segmentasi pelanggan. Data masukan terdiri dari 351 data dengan 3 atribut yang dipakai untuk analisis yakni Recency, Frequency, dan Monetary. Dari hasil analisis dengan Silhouette Score, didapatkan hasil jika pembagian data menjadi 2 kelompok cluster adalah yang paling optimal. Cluster 1 terdiri dengan 343 anggota, sedangkan cluster 2 terdiri dengan 8 anggota. Dari hasil analisis cluster dengan identifikasi anggota setiap cluster dan nilai centroid cluster, menunjukkan bahwa cluster 2 adalah kategori pelanggan dengan atribut Monetary yang relatif tinggi dan memiliki atribut Frequency (jumlah pembelian) 5,5 kali selama waktu penelitian. Sedangkan cluster 1 adalah kategori pelanggan dengan atribut Monetary yang relatif rendah dan memiliki atribut Frequency (jumlah pembelian) 2,34 kali selama waktu penelitian. Data hasil clustering menghasilkan segmentasi pelanggan yang dapat dipakai oleh manajemen untuk pendukung keputusan guna mengembangkan usaha di waktu mendatang.

PUSTAKA

Adiana, B. E., Soesanti, I., & Permanasari, A. E. (2018). ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN KOMBINASI RFM MODEL DAN TEKNIK CLUSTERING. *Jurnal Terapan Teknologi Informasi*. <https://doi.org/10.21460/jutei.2018.21.76>

Febriani, A., & Putri, S. A. (2020). Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means. *JIEMS (Journal of Industrial Engineering and Management Systems)*. <https://doi.org/10.30813/jiems.v13i2.2274>

Fitriyani, I. P., & Hendriyani, C. (2021). Implementasi Customer Data Management Dalam Meningkatkan Retensi Pelanggan IndiHome di PT Telkom. *ICIT Journal*. <https://doi.org/10.33050/icit.v7i2.1645>

Herdiana, R., & Solihudin, D. (2023). SEGMENTASI PELANGGAN SALON NUIH BEAUTY GLOW MENGGUNAKAN K-MEANS CLUSTERING. 7(1), 558–566.

Jabbar, J. (2022). SISTEM INFORMASI STOK BARANG MENGGUNAKAN METODE CLUSTERING KMEANS (STUDI KASUS RMD STORE). *INFOTECH Journal*.

- <https://doi.org/10.31949/infotech.v8i1.2280>
- Lestari, W., & Sumarlinda, S. (2021). Clustering Model of Lecturers Performa in Publication Using K-Means for Decision Support Data. *International Journal of Multisciences*
- Sari, Y. P., Primajaya, A., & Irawan, A. S. Y. (2020). Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang. *INOVTEK Polbeng - Seri Informatika*.
<https://doi.org/10.35314/isi.v5i2.1457>
- Setiawan, Amani, H., & Tripiawan, W. (2021). Perancangan Segmentasi Pelanggan dengan Metode Clustering K- Means dan Model RFM pada Klinik Kecantikan Seoul Secret. *E-Proceeding of Engineering*, 8(2), 2286–2293.
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER. *JBASE - Journal of Business and Audit Information Systems*.
<https://doi.org/10.30813/jbase.v4i1.2729>
- Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*.
<https://doi.org/10.33365/jtk.v15i2.1162>
- Sumarlinda, S., & Lestari, W. (2022). *Decision support system for lecturer publication mapping using k- means clustering method*. 5(4), 140–145.
- Wulan Permata Sari, T. S. (2023). ANALISA CLUSTER DENGAN K-MEAN CLUSTERING UNTUK PENGELOMPOKAN DATA CYBERCRIME. *JINTEKS (Jurnal Informatika Teknologi Dan Sains)*, 5(1), 49–53.