



Segmentasi pelanggan menggunakan algoritme bisecting k-means berdasarkan model recency, frequency, dan monetary (RFM)

Customer segmentation using bisecting k-means algorithm based on recency, frequency, and monetary (RFM) model

Novianti Puspitasari^{*)}, Joan Angelina Widiyans, Noval Bayu Setiawan

Program Studi Teknik Informatika, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Mulawarman
Jl. Panajam Kampus Gn. Kelua Universitas Mulawarman Samarinda, Kalimantan Timur, Indonesia 75123

Cara sitasi: N. Puspitasari, J. A. Widiyans, and N. B. Setiawan "Segmentasi pelanggan menggunakan algoritme bisecting k-means berdasarkan model recency, frequency, dan monetary (RFM)," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 2, pp. 78-83, 2020. doi: [10.14710/jtsiskom.8.2.2020.78-83](https://doi.org/10.14710/jtsiskom.8.2.2020.78-83), [Online].

Abstract – Information on customer loyalty characteristics in a company is needed to improve service to customers. A customer segmentation model based on transaction data can provide this information. This study used parameters from the recency, frequency, and monetary (RFM) model in determining customer segmentation and bisecting k-means algorithm to determine the number of clusters. The dataset used 588 sales transactions for PT Dinar Energi Utama in 2017. The clusters formed by the bisecting k-means and k-means algorithm were tested using the silhouette coefficient method. The bisecting k-means algorithm can form the best customer segmentation into three groups, namely Occasional, Typical, and Gold, with a silhouette coefficient of 0.58132.

Keywords – bisecting k-means; customer segmentation; RFM; best cluster; silhouette coefficient

Abstrak – Informasi tentang karakteristik loyalitas pelanggan di sebuah perusahaan sangat dibutuhkan dalam rangka meningkatkan pelayanan kepada pelanggan. Sebuah model segmentasi pelanggan berdasarkan data transaksi dapat memberikan informasi ini. Penelitian ini menggunakan parameter dari model recency, frequency, dan monetary (RFM) dalam menentukan segmentasi pelanggan dan algoritme bisecting k-means untuk menentukan jumlah kluster. Dataset yang digunakan adalah data 588 transaksi penjualan PT Dinar Energi Utama selama tahun 2017. Kluster yang dibentuk oleh algoritme bisecting k-means dan k-means diuji menggunakan metode koefisien silhouette. Algoritme bisecting k-means mampu membentuk model segmentasi pelanggan terbaik ke dalam tiga kelompok, yaitu Occasional, Typical, dan Gold, dengan koefisien silhouette 0,58132.

Kata kunci – bisecting k-means; segmentasi pelanggan; RFM; kluster terbaik; koefisien silhouette

I. PENDAHULUAN

Persaingan dalam dunia bisnis yang saat ini semakin ketat menuntut para pelaku bisnis untuk terus berinovasi. Teknologi informasi dapat dimanfaatkan sebagai sarana untuk melakukan pembaharuan strategi pemasaran dalam menganalisa kebutuhan pelanggan agar dapat mencetak nilai yang lebih berkualitas dalam penjualan produk maupun jasa yang dimiliki oleh suatu perusahaan. Salah satu strategi untuk memenangkan persaingan adalah strategi membangun kualitas pelayanan (*service quality*). Kualitas adalah totalitas fitur dan karakteristik produk atau jasa yang bergantung pada kemampuannya untuk memuaskan kebutuhan yang dinyatakan atau tersirat [1].

Upaya untuk menerapkan strategi membangun kualitas layanan ini seringkali terkendala pada minimnya informasi mengenai karakteristik loyalitas pelanggan terhadap perusahaan. Hal ini menyebabkan perusahaan kesulitan untuk mengelompokkan pelanggan dengan jumlah yang sangat banyak secara konvensional. Akibatnya, ketidaktepatan pelayanan yang diberikan kepada pelanggan rentan terjadi sehingga menyebabkan tingkat kepuasan pelanggan terhadap perusahaan menurun. Informasi mengenai segmentasi pelanggan sangat dibutuhkan untuk mengenali dan mengkaji perilaku pada suatu kelompok pelanggan serta memudahkan perusahaan dalam menentukan strategi pemasaran yang lebih tepat dan efektif.

Informasi segmentasi pelanggan di sebuah perusahaan bertujuan untuk mengenali perilaku loyalitas pelanggan. Lebih lanjut, segmentasi pelanggan merupakan strategi pemasaran yang tepat dalam rangka meningkatkan kepuasan layanan pelanggan terhadap perusahaan yang akan berdampak pada meningkatnya keuntungan perusahaan. Segmentasi pelanggan dengan jumlah data yang besar dapat diselesaikan dengan proses data mining dan dilakukan pemodelan berdasarkan parameter *recency, frequency, and*

^{*)} Penulis korespondensi (Novianti Puspitasari)
Email: novipuspita@fkti.unmul.ac.id

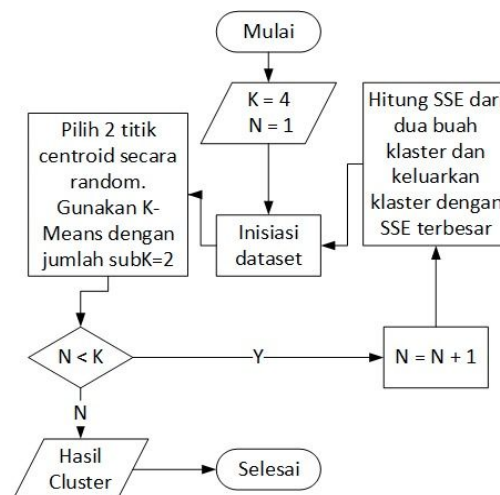
monetary (RFM). Model RFM membedakan tingkat loyalitas atau kepentingan pelanggan berdasarkan tiga variabel yaitu *recency*, *frequency*, dan *monetary*.

Model RFM telah diterapkan dalam berbagai aplikasi segmentasi yang terkait di bidang industri dan pemasaran, di antaranya adalah menentukan nasabah yang loyal terhadap perusahaan [2], membentuk segmentasi pelanggan berdasarkan data transaksi pada industri makanan [3], menentukan produk yang paling unggul berdasarkan indeks produk tertinggi [4], segmentasi pelanggan untuk rumah sakit hewan [5], dan perusahaan furnitur [6]. Model RFM memiliki keuntungan yang terletak pada relevansinya selama beroperasi di beberapa variabel yang dapat diamati dan memiliki objektivitas dalam menentukan segmentasi pelanggan [3].

Segmentasi dapat dilakukan dengan mengelompokkan sejumlah data menjadi kelompok-kelompok data tertentu (klaster) [7]. Objek data yang terletak dalam satu klaster harus mempunyai kemiripan. Aktifitas klasterisasi membutuhkan sebuah algoritme untuk menyelesaikan tugasnya. Beberapa variasi algoritme yang digunakan untuk membuat klaster di antaranya adalah k-means [8]-[11] dan fuzzy c-means (FCM) [12]-[17].

Selain algoritme tersebut, algoritme bisecting k-means dapat juga digunakan untuk melakukan pengelompokan. Algoritme ini merupakan variasi dari algoritme pengelompokan k-means. Proses *data mining* menggunakan algoritma ini bertujuan untuk mengelompokkan data ke dalam setiap klaster (*clustering*) yang sesuai dengan titik pusat (*centroid*) dari masing masing klaster. Algoritme ini telah digunakan untuk mengetahui pola perilaku pengunjung halaman situs web dalam [18]. Algoritme bisecting k-means juga diterapkan dalam menentukan segmentasi warna gambar untuk menghasilkan analisis gambar yang akurat [19]. Algoritme bisecting k-means telah teruji lebih baik daripada metode standar k-means karena memiliki waktu komputasi yang cepat [18]. Lebih lanjut, algoritme ini lebih baik dari metode pendekatan klaster hirarkis untuk tiga metode pengukuran kualitas klaster, yaitu *entropy*, *f-measure* dan *overall similarity* [20].

Kelebihan model RFM dalam relevansi dan objektivitas parameter pelanggan suatu perusahaan dapat diterapkan bersama algoritme bisecting k-means untuk melakukan segmentasi pelanggan suatu perusahaan berdasarkan parameter *recency*, *frequency*, dan *monetary*. Berbeda dengan kajian sebelumnya, penelitian ini secara khusus bertujuan untuk mengkaji penerapan algoritme *bisecting k-means* untuk membentuk segmentasi pelanggan di PT Dinar Energi Utama (PT. DEU), sebuah perusahaan penjualan alat survey di Kalimantan. Parameter yang digunakan dalam penentuan segmentasi pelanggan didasarkan pada model RFM dari data transaksi penjualan selama tahun 2017.



Gambar 1. Diagram alir algoritme bisecting k-means

II. METODE PENELITIAN

Penelitian ini menggunakan tahapan dari CRISP-DM [21], [22], model RFM, algoritme bisecting k-means, data transaksi pelanggan sebagai dataset, dan analisis pengujian klaster menggunakan metode koefisien *silhouette*.

A. Model RFM (*recency, frequency, and monetary*)

Segmentasi pelanggan dalam penelitian ini menggunakan 4 (empat) karakteristik menurut model RFM, yaitu pelanggan *Superstar*, *Golden*, *Typical*, dan *Occasional* [23]. Dua karakteristik terbawah yang tidak digunakan dalam penelitian ini adalah *Everyday Shopper* dan *Dorment*. Karakteristik *Everyday Shopper* tidak digunakan karena transaksi penjualan di PT. DEU bersifat periodik, bukan bersifat secara terus menerus (*continue*) seperti transaksi yang terjadi di tempat perbelanjaan atau minimarket. Karakteristik *Dorment* tidak digunakan karena memiliki nilai frekuensi dan *monetary* paling rendah. Hal ini tidak sesuai dengan jumlah transaksi penjualan pada PT. DEU yang memiliki jumlah nilai transaksi cukup besar.

B. Algoritme bisecting k-means

Algoritme bisecting k-means digunakan dalam penelitian ini karena mampu menginisialisasi *centroid* secara acak dan melakukan proses *bisecting* (pembagian menjadi dua) pada klaster dengan nilai *sum of squared error* (SSE) maksimum atau klaster terbesar. Gambar 1 menunjukkan diagram alir algoritme bisecting k-means untuk menemukan *k* klaster pada sebuah koleksi dokumen. Tahapan dari algoritme bisecting k-means secara berurutan adalah menentukan klaster yang akan dipisah (*split*), menemukan 2 sub-klaster menggunakan k-means tipe dasar (tahap *bisecting*), membagi dua ITER waktu dan ambil hasil *split clustering* yang memiliki SSE tertinggi, dan mengulangi langkah-langkah sebelumnya hingga jumlah klaster tercapai.

C. Koefisien silhouette

Pengujian kualitas dan kekuatan kluster yang dihasilkan oleh algoritme bisecting k-means dilakukan dengan menggunakan koefisien silhouette (SC) [24]. Untuk menghitung nilai SC diperlukan jarak antar dokumen dengan menggunakan rumus jarak Euclidean yang dinyatakan dalam Persamaan 1. Parameter S_i menunjukkan nilai akhir silhouette data tertentu, a_i nilai rata-rata jarak data tertentu, dan b_i nilai terkecil dari jarak rata-rata. Rentang nilai SC adalah 0 hingga 1. Dari nilai SC yang telah diperoleh, ditentukan kategori kualitas kluster berdasarkan standarisasi nilai silhouette seperti dinyatakan pada Tabel 1.

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (1)$$

D. Dataset

Penelitian ini menggunakan data transaksi penjualan yang terdapat di PT. DEU selama tahun 2017. Data transaksi penjualan dinyatakan pada Tabel 2. Data transaksi penjualan berisi informasi dari riwayat transaksi pelanggan. Kolom *Cust_id* menyatakan nomor identitas pelanggan yang melakukan transaksi. Kolom *Cust_name* menyatakan nama pelanggan. Kolom *Date* merupakan tanggal transaksi. Kolom *Invoice_no* memuat informasi tentang keterangan jenis atau nomor invoice transaksi. Kolom *Nama Barang* memiliki informasi nama barang. Kolom *Pembayaran* berisi data jumlah nominal transaksi dalam satuan Rupiah. Pengolahan data pada penelitian ini dilakukan dengan program Rapidminer.

III. HASIL DAN PEMBAHASAN

Proses segmentasi pelanggan menggunakan tahapan dari metode CRISP-DM. Tahapan ini diawali dengan *business understanding* (pemahaman bisnis). Dalam tahapan ini, dilakukan penggalian pemahaman bisnis di PT. DEU. PT. DEU memerlukan segmentasi pelanggan

Tabel 1. Nilai silhouette berdasarkan Kaufman dan Rousseeuw [25]

Nilai SC	Kualitas	Interpretasi
0,71 – 1,00	<i>Strong</i>	Kluster terbaik sudah ditemukan
0,51 – 0,70	<i>Medium</i>	Penempatan kluster yang wajar
0,26 – 0,50	<i>Weak</i>	Strukturnya lemah, coba metode tambahan
<= 0,25	<i>No Structure</i>	Tidak ada struktur yang ditemukan

untuk mengenali karakteristik pelanggan yang telah melakukan transaksi berdasarkan data riwayat transaksi pelanggan pada PT. DEU dan menggunakan parameter dari model RFM untuk proses pengelompokan (segmentasi).

Tahapan berikutnya adalah *data understanding* (pemahaman data). Pada tahap ini, dilakukan pembersihan data dari data asli yang terdapat pada Tabel 2. Dari hasil proses pada tahapan pemahaman data, diperoleh tiga variabel data yang digunakan seperti dinyatakan dalam Tabel 3 yaitu variabel *cust_id*, *date*, dan *value*. Variabel *cust_id* merepresentasikan nomor identitas pelanggan yang melakukan transaksi. Variabel *date* merupakan tanggal transaksi, dan *value* merupakan jumlah nominal pembayaran.

Tahapan *data preparation* (persiapan data) diawali dengan melakukan pemilihan data berdasarkan variabel dari model RFM. Hasil proses persiapan dan perhitungan RFM pada data transaksi pelanggan dinyatakan pada Tabel 4 yang menunjukkan nilai R, F, dan M dari setiap pelanggan. Waktu periode analisis penelitian adalah pekan ke-2 (kedua) dimulainya waktu pengumpulan dan penafsiran data penelitian, yaitu pada tanggal 23-08-2018.

Tahap selanjutnya adalah *data transformation* (transformasi data) dengan menggunakan metode Min-Max [26]. Perhitungan metode Min-Max menggunakan Persamaan 2. Parameter X menunjukkan nilai baru, p data lama, min nilai minimal lama, max nilai maksimal lama, min_{new} nilai minimal baru, dan max_{new} nilai maksimal baru. Hasil normalisasi data untuk hasil

Tabel 2. Data transaksi penjualan tahun 2017

No	Cust_id	Cust_name	Date	Invoice_no	Nama Barang	Pembayaran (Rp.)
1	170001	PT. Bara Mega Quantum	9-Feb-17	-	Pembelian Alat Total Station Topcon ES 105 (SN: GZ8328)	87.500.000
2	170002	PT. Putra Cilegon Mandiri	22-Feb-17	KALIBRASI	Service Kalibrasi Automatic Level Nikon AZ-2S (SN: 763330)	300.000
3	170002	PT. Putra Cilegon Mandiri	22-Feb-17	KALIBRASI	Service Kalibrasi Automatic Level Carlzeiss NI040A (SN: 53707)	300.000
4	170002	PT. Putra Cilegon Mandiri	22-Feb-17	KALIBRASI	Service Kalibrasi Theodolite H Trans (SN: 53707)	500.000
5	170002	PT. Putra Cilegon Mandiri	22-Feb-17	PENJUALAN	Pembelian Alat Theodolite Minds CDT-2 (SN : 177603)	13.000.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
585	170574	Sapto, Bp	6-Dec-17	17SA17026	TS FOCUS	20.000.000
586	170575	PDAM Wonoiri	16-Nov-17	17SA17028	GARMIN OREGON 750	17.700.000
587	170576	Ace	11-Dec-17	17RE17033	TS CX-102 SOKKIA	6.930.000
588	170577	Wisma Sarana Teknik, PT	19-Dec-17	17SA17029	HAMER TEST HT-225	3.300.000

perhitungan model RFM dengan metode Min-Max dinyatakan dalam Tabel 5.

$$X = \frac{(p - \min) \times (\max_{new} - \min_{new})}{(\max - \min) + \min_{new}} \quad (2)$$

Tahap setelah melakukan normalisasi adalah tahap *modeling* (pemodelan). Pada tahap pemodelan, hal yang dilakukan adalah mencari kluster terbaik dengan melakukan pemodelan dari jumlah $k = 2$ sampai dengan $k = 5$ dengan algoritme bisecting k-means dan k-means. Evaluasi dilakukan untuk mengukur kualitas kluster dengan menggunakan metode koefisien *silhouette*. Hasil pengujian kualitas kluster yang dibentuk oleh kedua algoritme tersebut dinyatakan pada Tabel 6. Kluster terbaik didapatkan pada algoritme bisecting k-means dengan jumlah kluster sebanyak tiga ($k = 3$) dengan nilai koefisien *silhouette* (SC) sebesar 0,58132. Algoritme k-means memiliki nilai tertinggi sebesar 0,55812 dengan 2 (dua) kluster.

Algoritme bisecting k-means menghasilkan nilai koefisien *silhouette* yang paling tinggi atau mempunyai kualitas kluster lebih baik daripada k-means seperti yang digunakan dalam [8]-[11]. Keunggulan bisecting k-means ini sesuai dengan [20] yang mengukur perbandingan algoritme menggunakan parameter *entropy*, *f-measure*, dan *overall similarity*. Pengujian lain perlu dilakukan untuk mendapatkan analisis kinerja algoritme bisecting k-means dengan parameter RFM, misalnya waktu komputasi seperti dalam [18].

Hal ini menunjukkan bahwa pembentukan kluster dengan algoritme bisecting k-means adalah pilihan yang tepat. Nilai *centroid* dari ketiga kluster yang dihasilkan oleh algoritme bisecting k-means ditunjukkan dalam Tabel 7. Dari hasil *centroid* dari masing-masing kluster tersebut, *grade* atau jenis / karakteristik pelanggan untuk masing-masing kluster diidentifikasi. Kluster pertama (C1) memiliki nilai *recency* (R) yang paling tinggi. Nilai tersebut merupakan karakteristik jenis pelanggan *Occasional*. Kluster kedua (C2) memiliki nilai *monetary* (M) dan *frequency* (F) tertinggi yang merupakan karakteristik dari jenis pelanggan *Golden*. Kluster ketiga (C3) memiliki nilai rata-rata tidak terlalu tinggi dan tidak terlalu rendah. Hal tersebut merupakan ciri dari jenis pelanggan *Typical*. Jenis pelanggan *Superstar* yang memiliki karakteristik tertinggi untuk semua nilai variabel RFM tidak ditemukan di dalam hasil *clustering*.

Hasil identifikasi jenis pelanggan ditunjukkan Gambar 2. Nilai *centroid* yang teridentifikasi oleh jenis pelanggan untuk masing-masing kluster adalah kelompok pelanggan *Occasional* (C1) berjumlah 106 orang, kelompok pelanggan *Typical* (C2) 199 orang, dan kelompok pelanggan *Golden* (C3) 8 orang. Tidak ada kelompok pelanggan *Superstar* yang teridentifikasi.

Tahapan terakhir adalah tahapan *deployment* untuk menampilkan hasil pengelompokan data yang divisualisasikan dengan *3D scatter plot*. Hal ini bertujuan agar lebih mudah memahami pola penyebaran data dari kluster yang dibentuk seperti ditunjukkan pada

Tabel 3. Hasil pemahaman data

Cust_id	Date	Value
170001	9-Feb-17	87.500.000
170002	22-Feb-17	300.000
170002	22-Feb-17	300.000
170002	22-Feb-17	500.000
170002	22-Feb-17	13.000.000
170003	5-Apr-17	1.000.000
170003	10-Apr-17	1.000.000
170003	15-Apr-17	1.000.000
⋮	⋮	⋮
170576	11-Dec-17	6.930.000
170577	19-Dec-17	3.300.000

Tabel 4. Hasil perhitungan dengan model RFM

Cust_d	R	F	M
170001	560	1	87.500.000
170002	547	4	14.100.000
170003	495	3	3.000.000
170004	474	3	3.300.000
170005	505	1	26.000.000
170006	434	1	1.000.000
170007	311	2	5.800.000
170008	288	4	2.150.000
170009	380	1	5.600.000
170010	370	1	66.500.000
170574	260	1	20.000.000
⋮	⋮	⋮	⋮
170575	280	1	17.699.999
170577	247	1	3.300.000

Tabel 5. Hasil normalisasi metode Min-Max

Cust_Id	R	F	M
170001	0,9	0	0,4
170002	0,86	0,1	0,06
170003	0,71	0,07	0,01
170004	0,65	0,07	0,01
170005	0,74	0	0,12
170006	0,54	0	0
170007	0,19	0,03	0,03
170008	0,13	0,1	0,01
170009	0,39	0	0,03
170010	0,36	0	0,3
170574	0,05	0	0,09
⋮	⋮	⋮	⋮
170575	0,1	0	0,08
170577	0,01	0	0,01

Tabel 6. Hasil pembentukan kluster

Algoritma	Jumlah kluster, k	Nilai SC
Bisecting K-Means	2	0,55812
	3	0,58132
	4	0,468827
	5	0,470172
K-Means	2	0,55812
	3	0,451443
	4	0,477819
	5	0,431487

Gambar 3. Pengukuran kualitas kluster dilakukan dengan menggunakan tiga metode yang berbeda, yaitu

Dunn index, koefisien *silhouette*, dan *elbow method* (WSS). Hasil pengujian tiap metode menghasilkan masing-masing nilai indeks sebesar 0,027967, 0,58132, dan 5,414225 yang ditunjukkan pada Gambar 4. Kualitas kluster yang dihasilkan oleh algoritme bisecting k-means termasuk ke dalam kategori *medium structure* dengan interpretasi penempatan kluster yang wajar [24], [25].

Berdasarkan proses segmentasi yang telah dilakukan menggunakan algoritme bisecting k-means dengan parameter RFM, diperoleh sebuah model segmentasi pelanggan ke dalam tiga karakteristik jenis pelanggan di PT. DEU, yaitu *Occasional*, *Golden*, dan *Typical*. Hal ini menunjukkan bahwa algoritme bisecting k-means mampu membentuk segmentasi pelanggan secara objektif karena jumlah kluster *k* tidak ditentukan terlebih dahulu sesuai [19]. Model RFM telah berhasil diterapkan, seperti halnya dalam [2]-[6], untuk segmentasi pelanggan menggunakan algoritme bisecting k-means dalam [18]-[20]. Hasil segmentasi ini dapat membantu perusahaan untuk meningkatkan layanannya sesuai [1], misalnya menawarkan produk dan layanan baru untuk jenis pelanggan *Gold*, memberikan diskon atau promosi kepada jenis pelanggan *Typical*, dan menjalankan kampanye pemasaran kesadaran merek (*brand awareness*) dengan target jenis pelanggan *Occasional*.

IV. KESIMPULAN

Algoritme bisecting k-means dengan parameter RFM mampu membentuk sebuah model segmentasi pelanggan bagi perusahaan berdasarkan data transaksi penjualan serta menghasilkan kluster yang termasuk *medium structure* dan mempunyai kualitas yang lebih baik daripada algoritme *k-means*.

UCAPAN TERIMA KASIH

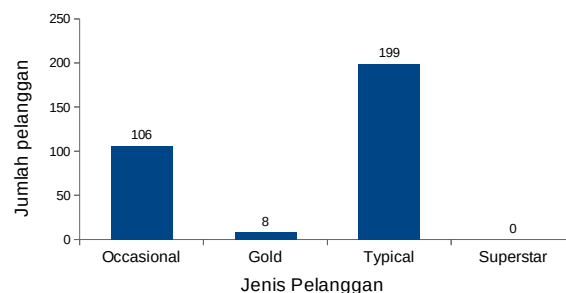
Terima kasih disampaikan kepada perusahaan PT. Dinar Energi Utama yang mendukung dan membantu dalam pelaksanaan penelitian ini.

DAFTAR PUSTAKA

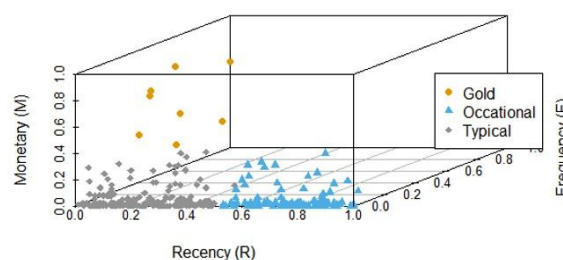
- [1] A. Ali, P. D. Paramita, and A. Fathoni, "Analisis faktor-faktor yang mempengaruhi kepuasan pelanggan di perusahaan galangan kapal (studi kasus pada PT. Janata Marina Indah Semarang)," *Journal of Management*, vol. 2, no. 2, pp. 1-12, 2016.
- [2] T. Hardiani and R. Hartanto, "Segmentasi nasabah tabungan menggunakan model RFM (recency, frequency, monetary) dan k-means pada lembaga keuangan mikro," in *Seminar Nasional Teknologi Informasi dan Komunikasi Terapan (SEMATIK)*, Semarang, Indonesia, Nov. 2015, pp. 463-468.
- [3] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis segmentasi pelanggan menggunakan kombinasi RFM model dan teknik clustering,"

Tabel 7. Nilai *centroid* dari bisecting k-means

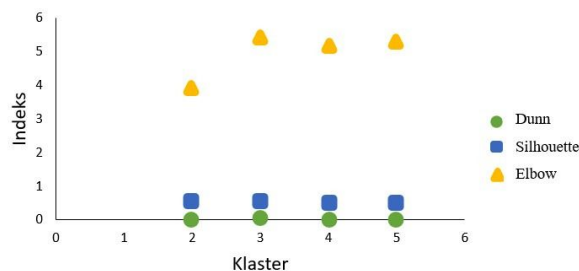
Kluster	R	F	M
C1	0,753113	0,015	0,041321
C2	0,1925	0,315	0,6375
C3	0,137188	0,021979	0,036667



Gambar 2. Hasil segmentasi jenis pelanggan



Gambar 3. Sebaran data untuk tiap jenis pelanggan berdasarkan RFM



Gambar 4. Pengujian kualitas kluster dengan beragam metode

- [4] M. A. Yaqin, S. Naja, S. K. Al-Azhar, and K. Mahbullah, "Penerapan metode fuzzy sugeno pada analisis RFM untuk menentukan indeks produk pada permainan hayday," *Prosiding SENIATI*, vol. 5, no. 1, pp. 50-56, 2019.
- [5] J. T. Wei, S.-Y. Lin, Y.-Z. Yang, and H.-H. Wu, "Applying data mining and RFM model to analyze customers' values of a veterinary hospital," in *2016 International Symposium on Computer, Consumer and Control (IS3C)*, Xian, China, Jul. 2016, pp. 481-484. doi: 10.1109/IS3C.2016.126
- [6] N. P. P. Yuliari, I. K. G. D. Putra, and N. K. D. Rusjayanti, "Customer segmentation through fuzzy c-means and fuzzy RFM method," *Journal of*

- Theoretical and Applied Information Technology*, vol. 78, no. 3, pp. 380-385, 2015.
- [7] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques, 3rd edition*. MA: Morgan Kaufman Publishers, 2012.
- [8] C. Slamet, A. Rahman, M. A. Ramdhani, and W. Darmalaksana, "Clustering the verses of the Holy Qur'an using k-means algorithm," *Asian Journal of Information Technology*, vol. 15, no. 24, pp. 5159-5162, 2016.
- [9] A. P. Windarto, "Penerapan data mining pada ekspor buah-buahan menurut negara tujuan menggunakan k-means clustering method," *Techno. Com*, vol. 16, no. 4, pp. 348-357, 2017. doi: [10.33633/tc.v16i4.1447](https://doi.org/10.33633/tc.v16i4.1447)
- [10] A. Bastian, "Penerapan algoritma k-means clustering analysis pada penyakit menular manusia (studi kasus kabupaten Majalengka)," *Jurnal Sistem Informasi*, vol. 14, no. 1, pp. 28-34, 2018.
- [11] P. Purnawansyah, H. Havaluddin, A. F. O. Gafar, and I. Tahyudin, "Comparison between k-means and fuzzy c-means clustering in network traffic activities," in *International Conference on Management Science and Engineering Management*, Kanazawa, Japan, Jul. 2017, pp. 300-310. doi: [10.1007/978-3-319-59280-0_24](https://doi.org/10.1007/978-3-319-59280-0_24)
- [12] T. Hardiani, "Segmentasi nasabah simpanan menggunakan fuzzy c means dan fuzzy RFM (recency, frequency, monetary) pada BMT xyz," *Network Engineering Research Operation (NERO)*, vol. 3, no. 3, pp. 185-192, 2018.
- [13] N. Puspitasari, R. Rosmasari, and S. Stefanie, "Penentuan prioritas perbaikan jalan menggunakan fuzzy c-means: studi kasus perbaikan jalan di kota Samarinda," *Jurnal Teknologi dan Sistem Komputer*, vol. 5, no. 1, pp. 7-14, 2017. doi: [10.14710/jtsiskom.5.1.2017.7-14](https://doi.org/10.14710/jtsiskom.5.1.2017.7-14)
- [14] M. N. Sutoyo and A. T. Sumpala, "Penerapan fuzzy c-means untuk deteksi dini kemampuan penalaran matematis," *Scientific Journal of Informatics*, vol. 2, no. 2, pp. 129-135, 2016. doi: [10.15294/sji.v2i2.5080](https://doi.org/10.15294/sji.v2i2.5080)
- [15] R. J. E. Putra, N. Nasution, and Y. Yummastian, "Aplikasi e-zakat penerimaan dan penyaluran menggunakan fuzzy c-means (studi kasus: LAZISMU Pekanbaru)," *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, vol. 6, no. 2, pp. 42-54, 2015.
- [16] A. A. Rachman and Z. Rustam, "Cancer classification using fuzzy c-means with feature selection," in *12th International Conference Mathematics, Statistics, and Their Applications*, Banda Aceh, Indonesia, Oct. 2016, pp. 31-34. doi: [10.1109/ICMSA.2016.7954302](https://doi.org/10.1109/ICMSA.2016.7954302)
- [17] N. Puspitasari, J. A. Widians, and P. Pohny, "A clustering of generative and infectious diseases using fuzzy c-means," *ITSMART: Jurnal Teknologi dan Informasi*, vol. 7, no. 1, pp. 22-28, 2018.
- [18] R. Patil and A. Khan, "Bisecting K-means for clustering web log data," *International Journal of Computer Applications*, vol. 116, no. 19, pp. 36-41, 2015. doi: [10.5120/20448-2799](https://doi.org/10.5120/20448-2799)
- [19] Q. Zhang, Y. Chi, and N. He, "Color image segmentation based on a modified k-means algorithm," in *7th International Conference on Internet Multimedia Computing and Service*, New York, USA, Aug. 2015, pp. 1-4. doi: [10.1145/2808492.2808538](https://doi.org/10.1145/2808492.2808538)
- [20] M. S. G. Karypis, V. Kumar, and M. Steinbach, "A comparison of document clustering techniques," in *KDD Workshop on Text Mining*, Boston, USA, Aug. 2000, pp. 1-20.
- [21] M. Wati, W. Indrawan, J. A. Widians, and N. Puspitasari, "Data mining for predicting students' learning result," in *4th International Conference on Computer Applications and Information Processing Technology*, Bali, Indonesia, Aug. 2017, pp. 1-4. doi: [10.1109/CAIPT.2017.8320666](https://doi.org/10.1109/CAIPT.2017.8320666)
- [22] P. Chapman et al., *CRISP-DM 1.0 Step-by-step data mining guide*. CRISP-DM consortium, 2000.
- [23] C.-H. Cheng and Y.-S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert systems with applications*, vol. 36, no. 3, pp. 4176-4184, 2009. doi: [10.1016/j.eswa.2008.04.003](https://doi.org/10.1016/j.eswa.2008.04.003)
- [24] P. Tan, M. Steinbach, and V. Kumar, "Introduction to Data Mining, Boston: Person Education," ed: Inc, 2006.
- [24] E. Prasetyo, *Data mining mengolah data menjadi informasi menggunakan Matlab*. Yogyakarta: Andi Offset, 2014.
- [25] L. Kaufman and P. J. Rousseeuw, *Finding groups in data*. New York: John Wiley & Sons, 1990.
- [26] Y. K. Jain and S. K. Bhandare, "Min max normalization based data perturbation method for privacy protection," *International Journal of Computer & Communication Technology*, vol. 2, no. 8, pp. 45-50, 2011.