

PENERAPAN ALGORITME KLASTERISASI K-MEANS KINERJA PRODUK DENGAN ANALISIS *RECENCY FREQUENCY MONETARY* PADA CAFE XYZ

Aldi Kamalludin^{1*}, Utomo Budiyo², Dolly Virgianshaka Yudha Sakti³, Joko Christian⁴

^{1,2,3,4} Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, DKI Jakarta, Indonesia

Email: ^{1*}aldikamalludin99@gmail.com, ²utomo.budiyo@budiluhur.ac.id, ³dolly.virgianshaka@budiluhur.ac.id,
⁴joko.christian@budiluhur.ac.id

Abstrak- XYZ merupakan salah satu *coffee shop* lokal terbesar di Indonesia. Meskipun begitu *Coffee Shop* ini juga memiliki banyak pesaing dengan bisnis yang sama sehingga manajemen dituntut harus selalu memenuhi kebutuhan para konsumen agar tidak ada konsumen yang berpaling ke para pesaing. Maka dari itu penelitian ini bertujuan untuk melakukan klusterisasi penjualan menu produk cafe XYZ ke dalam kluster-kluster yang memiliki kriteria yang sama untuk menghasilkan produk menu kinerja baik, menu kinerja sedang, menu kinerja buruk. Sehingga dapat diketahui produk apa saja yang harus di evaluasi agar mampu bersaing dengan kompetitor lain. Manfaat dari penelitian dapat menganalisa dan evaluasi menu yang kinerja buruk agar bisa menjadikan opsi *management*. Metode yang dipakai pada penelitian ini adalah *Algoritme K-Means Clustering* berdasarkan analisis *Recency*, *Frequency* dan *Monetary* (RFM). Metode ini sangat tepat untuk penelitian karena dalam implementasi menyelesaikan masalah *Algoritme K-Means* sangat *simple* serta *fleksible*. Data yang di gunakan untuk penelitian ini yaitu Februari 2022 sampai Desember 2022. Data transaksi yang diambil dari segment pagi dapat menghasilkan 3 kluster dari data 15632 dengan 322 produk masing-masing nama kluster tersebut yaitu menu kinerja baik terdapat 276 data, menu kinerja sedang terdapat 36 data, dan menu kinerja buruk terdapat 10 data. Penelitian ini berhasil membuat *prototype clustering* menggunakan analisis RFM. Dengan ada nya pengelompokan ini berguna untuk memudahkan *management cafe XYZ* untuk membuat keputusan ketika ingin mengevaluasi menu produk.

Kata Kunci: *K-Means*, *Data mining*, Algoritma Klusterisasi, Kinerja Produk, *Coffee Shop*.

APPLICATION OF K-MEANS CLUSTERING ALGORITHM OF PRODUCT PERFORMANCE WITH RECENCY FREQUENCY MONETARY ANALYSIS AT XYZ CAFE

Abstract- XYZ is one of the biggest local coffee shops in Indonesia. Even so, this Coffee Shop also has many competitors with the same business so that management is required to always meet the needs of consumers so that no customers turn to competitors. Therefore, this research aims to cluster the sales of XYZ cafe product menus into clusters that have the same criteria to produce good performance menu products, moderate performance menus, bad performance menus. So that it can be known what products must be evaluated in order to be able to compete with other competitors. The benefits of research can analyze and evaluate menus that are underperforming in order to make management options. The method used in this research is the *K-Means Clustering Algorithm* based on *Recency*, *Frequency* and *Monetary* (RFM) analysis. This method is very appropriate for research because in implementing solving problems the *K-Means Algorithm* is very simple and flexible. The data used for this research is February 2022 to December 2022. Transaction data taken from the morning segment can produce 3 clusters of 15632 data with 322 products, each name of the cluster, namely the good performance menu has 276 data, the moderate performance menu has 36 data, and the bad performance menu contains 10 data. This study succeeded in making a clustering prototype using RFM analysis. With this grouping it is useful to make it easier for XYZ cafe management to make decisions when they want to evaluate product menu.

Keywords: *K-Means*, *Data mining*, *Clustering Algorithm*, *Product Performance*, *Coffee Shop*.

1. PENDAHULUAN

Budaya masyarakat Indonesia meminum kopi pada umumnya di pagi hari. Sebelum beraktivitas bahkan kebiasaan meminum kopi salah satu cara merayakan kejadian penting. Dalam rutinitas sehari-hari masyarakat Indonesia tidak bisa lepas dengan meminum kopi maka dari itu cafe XYZ menciptakan suasana baru untuk masyarakat Indonesia yang ingin meminum kopi dengan gaya yang berbeda. Penjualan kopi selalu dipengaruhi oleh jam-jam sibuk beraktivitas, seperti pagi, siang, sore dan malam hari. Setiap jam sibuknya selalu mempunyai produk yang tersedia, maka dari itu banyak menu yang menjadi unggulan di setiap *segment* pagi, siang, sore dan malam.

Untuk membantu kelancaran penelitian ini menggunakan *data mining*. Di zaman yang serba maju ini dengan perkembangan teknologi yang sangat cepat banyak keuntungan yang bisa diambil untuk mengambil keputusan.

Permasalahan yang terjadi di cafe XYZ masih terbatas nya sistem yang dijadikan acuan untuk mengevaluasi dan menganalisis menu yang kurang laku Untuk meningkatkan pelayanan harus bisa menciptakan solusi untuk bersaing era globalisasi seperti sekarang salah satunya dengan menganalisa dan mengavaluasi menu yang kurang laku. Data Mining adalah proses yang menarik informasi dan pengetahuan terkait dari database besar menggunakan pendekatan statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. Menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang sudah kita miliki adalah tujuan inti dari data mining [1].

Penelitian ini merupakan salah satu bagian dalam menganalisa data penjualan dengan menggunakan data mining untuk meningkatkan penjualan pada *coffee shop*. Analisa data penjualan didapatkan dari data hasil penjualan yang sudah ada pada *database cafe XYZ*. Dimana penulis mengimplementasikan *algoritme* yang menggunakan teknik *clustering*. *Algoritme* tersebut yaitu *algoritme K-Means* dengan menggunakan analisis *Recency, Frequency*, dan *Monetary* (RFM). *Algoritme* tersebut diimplementasikan dengan menggunakan bahasa pemrograman PHP dan berbasis *web*. Dimulai dari pengumpulan data penjualan selanjutnya data tersebut dilakukan *preprocessing* dengan memilih data yang dibutuhkan, kemudian menentukan nilai dari transaksi pelanggan dengan menggunakan parameter *Recency, Frequency* dan *Monetary* (RFM). Setelah nilai RFM berhasil didapatkan maka hasil tersebut diolah kembali menjadi *cluster* sebagai penerapan *K-Means Clustering*.

K-Means clustering adalah metode *data mining* atau metode penganalisaan dari data dengan proses pemodelan tanpa adanya supervise dan *K-Means*. Model RFM (*Recency, Frequency* dan *Monetary*) yang berguna untuk mengoptimalkan data untuk proses klasterisasi [2]. Pada penelitian sebelumnya [2] sudah ada permasalahan yang dihadapi dalam menentukan tidak laku, sangat laku, laku dengan metode *clustering* dan *algoritme K-Means* agar bisa membantu *cafe 47° coffee* meningkatkan stratregi penjualan dan mengecek stok barang. Di penelitian lainnya[3] menggunakan metode *clustering* dan *algoritme K-Means* memang cukup efektif untuk mengelompokan produk yang laku dan tidak laku.

2. METODE PENELITIAN

2.1 Data mining

Penambahan data merupakan prosedur menjelajahi dan mempelajari statistik dalam jumlah besar dengan tujuan menemukan pola atau informasi yang menarik dari statistik tersimpan dalam jumlah besar menggunakan teknik atau metode tertentu [4]

2.2 Davies Bouldin Index (DBI)

Strategi untuk mengidentifikasi nilai *cluster* yang ideal adalah dengan menggunakan *Davies Bouldin Indeks*. Metode ini mencoba untuk mengurangi jarak antara masing-masing anggota cluster sekaligus memaksimalkan jarak antara satu cluster dengan *cluster* lainnya. berapa banyak *cluster* yang memiliki nilai DBI minimal terbaik [5]. Perhitungan DBI memiliki tahapan yaitu yang pertama menghitung *Sum of Square Within-klaster* (SSW) menggunakan rumus 1. Pada iterasi terakhir perhitungan *Algoritme K-Means* menggunakan *centroid* pada saat menghitung SSW.

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{M_i} d(x_j, c_i) \quad (1)$$

Dimana:

m_i adalah jumlah nilai dalam klaster -i

c_i adalah *centroid* klaster ke-i

$d(x_j, c_i)$ adalah jarak dari data ke-I ke titik klaster i

j adalah banyak nya data

x adalah data

Setelah mencari nilai SSW selanjutnya mencari nilai *Sum of Between* klaster digunakan untuk mengetahui seaparasi antar klaster yang dihitung menggunakan rumus 2

$$SSB_{i,j} d(c_j, c_j) \quad (2)$$

Setelah menemukan nilai SSB maka selanjutnya menghitung nilai DBI menggunakan rumus 3.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (3)$$

Selanjutnya untuk mengetahui nilai perbandingan antara kluster ke-i dan kluster ke-j menggunakan rumus 4

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (4)$$

Dari perhitungan rumus tersebut jika nilai DBI yang diperoleh (non-negatif ≥ 0), maka semakin baik kluster yang diperoleh.

2.3 Metode Clustering

Sebagai salah satu teknik pengelompokan data menggunakan sistem partisi, clustering merupakan teknik analisis data yang menjalankan proses pemodelan tanpa visi [6]. Yang dilakukan pada tahap pertama yaitu menentukan jumlah kluster (k), selanjutnya menentukan k sebagai *centroid*, cara yang dilakukan untuk menentukan *centroid* awal ialah dilakukan secara acak. Setelah memiliki *centroid* awal, hitung jarak data dengan *centroid* menggunakan rumus formula *Euclidean Distance*, seperti pada rumus 5 [7].

$$d_{ij} = \sqrt{\sum_{k=1}^m X_{ij} - C_{ij}^2} \quad (5)$$

Dimana:

- d_(ij) : Jarak objek antar nilai data dan nilai pusat kluster
- m : Jumlah dimensi data
- X_{ij} : Nilai data dari dimensi ke-k
- C_{ik} : Nilai pusat kluster dari dimensi ke-k

Dengan d_{ij} adalah jarak *object* antar nilai data dan nilai pusat kluster, m merupakan jumlah dimensi data, X_{ij} merupakan nilai data dari dimensi ke-k dan C_{jk} adalah nilai pusat kluster dari dimensi ke-k. Kelompokkan data berdasarkan kedekatan dengan *centroid* kemudian perbaharui nilai *centroid* baru dengan lokasi dari pusat kluster. Gunakan rumus 6 dibawah, untuk menghitung *centroid* baru.

$$C = \frac{\sum m}{n} \quad (6)$$

Menurut [8] Teknik penambangan data seperti pengelompokan digunakan untuk menilai dan menganalisis data untuk mengatasi kesulitan pengelompokan disebut *clustering*.

2.4 Recency, Frequency, dan Monetary (RFM)

Recency Frequency Monetary (RFM) adalah metode yang umum digunakan untuk mengoptimalkan data loyalitas *customer* yaitu dengan memabagi aktifitas ke dalam 3 atribut yakni *Recency Frequency* dan *Monetary*. Atribut *Recency* adalah menghitung waktu transaksi terakhir dari produk yang dibeli dan dibandingkan dengan tanggal penelitian ini dilakukan. Atribut *frequency* hanya perlu mengambil jumlah item yang terjual dari produk dalam satu kali periode data. Lalu *monetary* adalah berapa jumlah pendapatan yang dihasilkan oleh produk tersebut selama satu periode data [9].

2.5 Pengumpulan Data

Dalam melakukan penelitian ini penulis membutuhkan data yang berkaitan dengan topik yang dibahas. Metode pengumpulan data dilakukan dengan tujuan untuk memperoleh data, yaitu:

a. Observasi.

Observasi dilakukan untuk meninjau langsung kegiatan proses transaksi dan tata letak *cafe* pada XYZ. Proses yang diamati pada *cafe XYZ* yaitu pembeli langsung datang ke *cafe* dan melakukan transaksi secara langsung dengan *staff* kasir.

b. Wawancara

Wawancara dilakukan dengan *manager store XYZ* yang terkait dengan data yang diperlukan untuk memperoleh data dan informasi tentang proses transaksi saat ini dan tentang sistem penyimpanan data setelah atau sebelum transaksi selesai. Informasi yang didapat digunakan sebagai acuan dalam pembuatan sistem aplikasi.

c. Dokumentasi

Metode ini penulis meminta kepada *manager store* terkait data yang diperlukan dan tersedia pada cafe *XYZ* yang digunakan sebagai masukan dalam proses pembuatan sistem aplikasi.

d. Tinjauan Pustaka

Tinjauan Pustaka dilakukan untuk mengumpulkan informasi yang tersedia dalam bentuk jurnal yang digunakan penulis sebagai acuan dalam melakukan penelitian ini, dengan jurnal yang terkait *data mining* dan *K-Means clustering*.

2.6 CRISP-DM

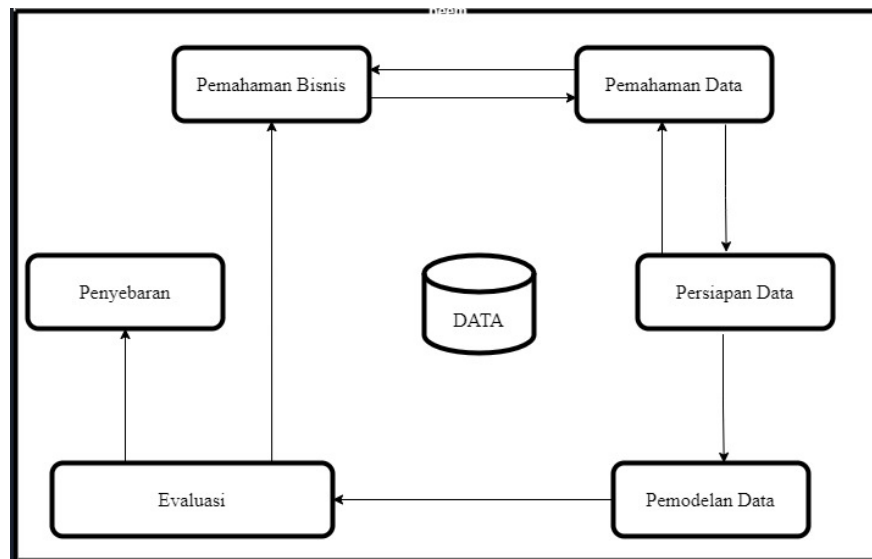
Metode yang diterapkan pada penelitian ini adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*). CRISP-DM merupakan metode yang menggunakan model pengembangan data. Ada enam tahap proses CRISP-DM yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi dan penyebaran [10].

3. HASIL DAN PEMBAHASAN

3.1 Data Penelitian

Data penelitian ini adalah data murni tanpa adanya manipulasi sehingga data tersebut dapat digunakan untuk analisis dalam suatu penelitian. Dengan penelitian di *XYZ* yang merupakan sebuah Coffee Shop yang terletak di Tangerang. *XYZ* memiliki data berupa file excel, dengan jumlah data 48614. Data yang digunakan pada penelitian ini adalah data transaksi pembelian *XYZ* dari tanggal 17/02/2021 sampai dengan 30/12/2021.

3.2 Tahapan CRISP-DM



Gambar 1. CRISP-DM

3.2.1 Pemahaman Bisnis

Pemahaman bisnis mempelajari dan memahami tujuan kebutuhan dari bisnis pada penelitian ini. Proses mempelajari dan memahami bisnis ini selanjutnya penulis dapat mengetahui metode yang tepat untuk menyelesaikan permasalahan pada tempat penelitian. Tempat penelitian yang dimaksud pada penelitian ini adalah bisnis Coffee Shop yang melayani penjualan secara langsung melalui outlet. Data transaksi penjualan *XYZ* selanjutnya di-filter sesuai transaksi per-jam nya. Dari data yang telah ter-filter dikelompokkan atau diklasterisasi. Untuk mempermudah proses klasterisasi, maka penelitian ini menggunakan model RFM (*Recency, Frequency, Monetary*) lalu penelitian ini diimplementasikan ke dalam bentuk aplikasi berbasis web agar memiliki interface yang dapat diakses oleh pengguna.

3.2.2 Pemahaman Data

Pemahaman data dilakukan dengan cara memeriksa kualitas data. Data yang digunakan pada penelitian ini adalah data penjualan XYZ dari tanggal 17/02/2021 sampai dengan 30/12/2021.

3.2.3 Persiapan Data

Dari 48614 data sebelumnya selanjutnya disaring, data yang tidak digunakan pada proses penelitian tidak perlu diikuti untuk proses data mining. Parameter yang dibutuhkan untuk penelitian antara lain: Date, Time, Order, Product, Category, Group, Quantity, Price, Discount, Total. Data seperti tabel 1.

Tabel 1. Penjelasan Parameter

Parameter	Penjelasan
<i>Date</i>	Parameter date ini menjelaskan awal tanggal transaksi itu di lakukan
<i>Time</i>	Parameter time ini menjelaskan transaksi ini masuk pada segement apa
<i>Order</i>	Parameter ini menjelaskan untuk mengetahui banyaknya produk yang sudah dipesan
<i>Product</i>	Parameter ini menjelaskan untuk memberi tahu nama produk
<i>Category</i>	Parameter ini menjelaskan produk yang tersedia masuk di jenis minuman apa
<i>Quantity</i>	Parameter ini menjelaskan banyaknya pesanan
<i>Price</i>	Parameter ini menjelaskan untuk memberi tahu harga dari produk yang dipesan
<i>Discount</i>	Parameter ini menjelaskan untuk memberi tahu potongan harga pada produk yang sudah dipesan
<i>Total</i>	Parameter ini menjelaskan untuk memberi tahu total harga dari produk yang sudah dipesan

3.2.4 Pemodelan Data

Metode yang digunakan dalam proses pemodelan data untuk penelitian ini adalah algoritme K-Means. Data input untuk proses klustering adalah data RFM yang dinormalisasi dari tahap sebelumnya Model RFM digunakan untuk memfasilitasi klustering dalam penelitian ini untuk mensegmentasi pelanggan. Data transaksi diubah menjadi data densitas yang lebih tinggi, menghasilkan data yang meliputi *Recency* (tanggal akhir transaksi), *Frequency* (total produk yang dibeli), *monetary* (jumlah banyak nya produk yang dibeli).

3.2.5 Evaluasi

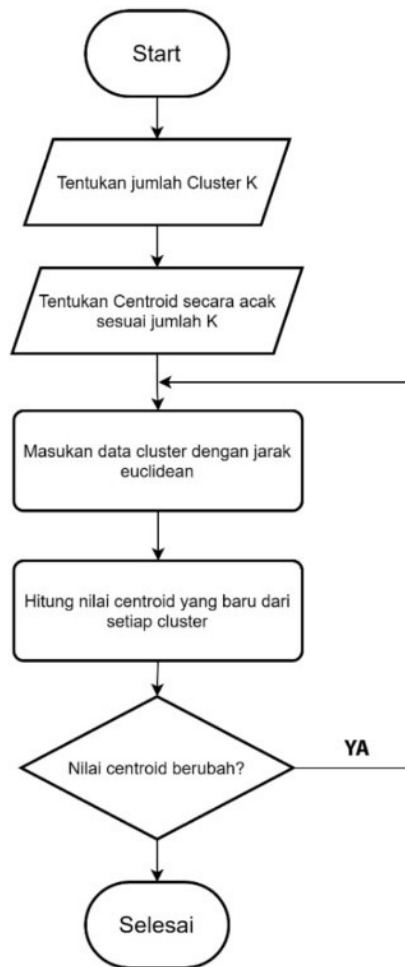
Metode DBI (*Davies Buildin Index*) digunakan untuk menghitung rata-rata nilai setiap titik pada himpunan data. Untuk mengukur validitas klaster pada suatu metode pengelompokan salah satunya menggunakan metode *Davies-Bouldin Index*, kedekatan data terhadap titik pusat kluster didefinisikan sebagai *kohesi*. Sedangkan separasi didasarkan pada jarak antar titik pusat klaster terhadap klasternya [11]

3.2.6 Penyebaran

Di tahap ini pengetahuan atau informasi yang telah dihasilkan algoritme data mining K-Means diimplementasikan menjadi sebuah aplikasi berbasis web sehingga dapat digunakan dengan mudah oleh pihak pengguna.

3.2.7 Implementasi Algoritme K-Means

Setelah pemodelan dari RFM sudah dilakukan, selanjutnya masuk ke perhitungan Algoritme K-Means. Langkah pertama yang dilakukan adalah menentukan jumlah klaster. Dalam penelitian ini klaster yang ditentukan adalah klaster Menu Kinerja Bagus, Menu Kinerja Sedang, dan Menu Kinerja Buruk.



Gambar 2. Flowchart K-Means

Setiap kluster yang ditentukan sebelumnya memiliki pusat kluster atau yang disebut *centroid*. Nilai *centroid* itu dihitung secara berulang-ulang sampai ditemukan *centroid* yang sempurna atau *centroid* yang tidak berubah-ubah lagi, sehingga kluster yang dihasilkan juga tetap atau tidak ada produk yang berpindah-pindah kluster. Untuk menghitung iterasi kluster yang pertama, *centroid* awal ditentukan secara random atau acak. Tabel 3 menunjukkan data *clustering*

Tabel 2. Data Clustering

No	Product Name	Recency	Frequency	Monetary	Cluster
1	HAZELNUT CREAM FRAPPE ICED R	357	10	Rp 36.800	C2
2	BLUEBERRY SCOFFIN	558	1	Rp 32.000	C2
3	AMERICANO ICED MX	362	157	Rp 5.861.000	C1
...
320	HON LEMON CHAMOMILE ICED M	429	2	Rp 72.000	C2
321	CARAMEL FRAPPE ICED MX	369	17	Rp 610.000	C2
322	TUMBLER STRAW BLUE	398	1	Rp 360.000	C2

Tabel 3. Centroid Awal

Klaster	Recency	Frequency	Monetary
C1	369	8	Rp 315.000
C2	609	1	Rp 300.000
C3	563	20	Rp 539.000

Setelah *centroid* awal sudah ditentukan. Langkah selanjutnya adalah menghitung jarak seluruh data dengan masing-masing kluster dengan menggunakan persamaan Euclidean Distance. Berikut adalah perhitungan satu buah data. Menggunakan rumus 5. Pada perhitungan ini menggunakan produk **HAZELNUT CREAM FRAPPE ICED R**

$$\begin{aligned} \text{Jarak Ke C1} &= \sqrt{(560 - 369)^2 + (2 - 8)^2 + (60000 - 315000)^2} = 255000 \\ \text{Jarak Ke C2} &= \sqrt{(560 - 609)^2 + (2 - 1)^2 + (60000 - 300000)^2} = 240000 \\ \text{Jarak Ke C3} &= \sqrt{(560 - 563)^2 + (2 - 20)^2 + (60000 - 539000)^2} = 479000 \end{aligned}$$

Untuk mengetahui **HAZELNUT CREAM FRAPPE ICED R** masuk dalam kluster mana, selanjutnya bandingkan seluruh jarak dan ambil nilai yang paling kecil. Karena jarak ke C2 adalah nilai terkecil maka **HAZELNUT CREAM FRAPPE ICED R** masuk ke kluster C2. Jika seluruh data sudah dihitung dan mendapatkan kluster, pada langkah selanjutnya mencari *centroid* iterasi kedua yaitu dengan cara menghitung rata-rata kluster. Perhitungan rata-rata kluster dapat dilihat seperti rumus di bawah.

Rata-rata pada C1:

$$\begin{aligned} \text{Recency} &= \frac{357 + 558 + 481 + \dots + 390 + 429}{114} = \frac{53610}{114} = 470 \\ \text{Frequency} &= \frac{10 + 1 + 4 + \dots + 3 + 2}{114} = \frac{391}{114} = 3.42982 \\ \text{Monetary} &= \frac{368000 + 32000 + 5861000 + \dots + 610000 + 360000}{114} = \frac{13762500}{114} = 120723.68 \end{aligned}$$

3.2.8 Penentuan Nama Cluster Berdasarkan RFM

Untuk penentuan nama kluster diperlukan nilai dari *centroid* akhir. Seperti pada Tabel 4.

Tabel 4. *Centroid* Akhir RFM

Kluster	Recency	Frequency	Monetary
C1	369.3611	161.3889	5609115.9722
C2	363.7000	437.6000	16445187.5000
C3	430.7101	15.9058	671220.1993

Setelah itu perlu diranking nilai RFM dari setiap kluster. Ketentuan dalam pe-rangking-an adalah jika semakin kecil nilai *recency* maka nilai ranking nya semakin besar. Jika semakin besar nilai *frequency* maka nilai ranking nya semakin besar. Jika semakin besar nilai *monetary* maka nilai ranking nya semakin besar. Nilai rata-rata terbesar dari rangking milik kluster adalah menu kinerja bagus, sedangkan nilai rata-rata terkecil adalah menu kinerja buruk, sisanya adalah menu kinerja sedang. Berikut adalah nilai rangking yang didapatkan oleh setiap kluster berdasarkan *centroid* akhir adalah sebagai berikut. Seperti pada Tabel 5.

Tabel 5. Penentuan Nama Kluster

Kluster	R	F	M	Rata-rata	Nama Kluster
C1	2	2	2	2	Menu Kinerja Sedang
C2	3	3	3	3	Menu Kinerja Bagus
C3	1	1	1	1	Menu Kinerja Buruk

Dari anggota data milik setiap kluster dapat diketahui bahwa setiap anggota data di dalam kluster memiliki kedekatan kriteria. Rata-rata anggota pada kluster kurang diminati memiliki nilai *recency* yang besar, nilai *frequency* yang kecil, dan nilai *monetary* yang kecil. Pada kluster paling diminati memiliki nilai *recency* yang kecil, nilai *frequency* yang besar, dan nilai *monetary* yang besar.

3.2.9 Evaluasi Kluster Menggunakan DBI

Metode *Davies-Bouldin Index* (DBI) digunakan untuk mengukur validitas kluster yang sudah di dapat. Langkah selanjutnya adalah mencari nilai SSW (*Sum of Square Within-kluster*) kepada seluruh data. Berikut adalah perhitungan SSW pada satu data. Menggunakan rumus 1.

HAZELNUT CREAM FRAPPE ICED R

$$SSW1 = \sqrt{(560 - 369.3611)^2 + (2 - 161.3889)^2 + (60000 - 5609115.9722)^2} = 5549115$$

$$SSW2 = \sqrt{(560 - 363.7000)^2 + (2 - 437.6000)^2 + (60000 - 16445187)^2} = 16385187$$

$$SSW3 = \sqrt{(560 - 430.7101)^2 + (2 - 15.9058)^2 + (60000 - 671220.1993)^2} = 611220.2$$

Setelah mendapatkan nilai SSW pada seluruh data. Maka selanjutnya ialah mencari rata-rata nilai SSW1 sampai SSW3 dengan menggunakan perhitungan rumus SSB (*Sum of Square Between*-klaster) pada perhitungan ini dimulai menghitung jarak antar suatu klaster. Menggunakan rumus 2

$$SSB_{1,2} = \sqrt{(369.3611 - 363.7000)^2 + (161.3889 - 437.6000)^2 + (5609115 - 16445187)^2} = 10836072$$

$$SSB_{1,3} = \sqrt{(369.3611 - 430.7101)^2 + (161.3889 - 15.9058)^2} + SSB_{1,3} = 4937895$$

$$SSB_{2,3} = \sqrt{(363.7000 - 430.7101)^2 + (437.6000 - 15.9058)^2 + (16445187 - 671220)^2} = 15773967$$

Setelah nilai SSW dan SSB didapatkan, langkah selanjutnya adalah menghitung rasio dengan cara menghitung jarak antar SSW. Dari perhitungan tersebut menghasilkan nilai rasio terbesar (*R-max*), yang selanjutnya digunakan untuk mendapatkan nilai DBI. Perhitungan dibawah menggunakan rumus 3.

$$R_{1,2} = \frac{20032.9061 + 455144.0833}{10836072} = 0.43851$$

$$R_{1,3} = \frac{20032.9061 + 61122.02}{4937895} = 0.16435$$

$$R_{2,3} = \frac{455144.0833 + 61122.02}{15773967} = 0.4228$$

Tabel 7. Ratio

R	1	2	3	R-MAX
1	0	0.43851	0.16435	0.467
2	0.43851	0	0.4228	0.467
3	0.16435	0.4228	0	0.414

Dari perhitungan di atas mendapatkan nilai rasio terbesar (*R-Max*), selanjutnya nilai *R-Max* tersebut dihitung rata-rata untuk mendapatkan nilai DBI. Dari perhitungan tersebut menghasilkan nilai DBI sebesar 0.449. Maka dapat disimpulkan nilai DBI dapat dikatakan baik karena nilai mendekati 0 (non-negatif).

4. KESIMPULAN

Dalam melalui proses pengerjaan yang dilakukan pada penelitian ini, maka didapatkan kesimpulan penelitian ini berhasil membuat prototype clustering menggunakan analisis RFM, klaster yang di hasilkan ada 3 yaitu menu kinerja baik, menu kinerja sedang, menu kinerja buruk dan memiliki nilai yang di hasilkan 0.449. Dengan adanya pengelompokan data ini, membantu manajemen XYZ dengan mudah mengevaluasi menu. Saran yang dapat dikembangkan lebih lanjut di program dengan harapan dapat menghasilkan penelitian yang lebih baik lagi kedepannya, diharapkan pada penelitian selanjutnya dapat mengeksplorasi metode analisis lain untuk mengklasterisasi data penjualan. Seperti *K-Median*, *Fuzzy C-Means*. Pada penelitian lebih lanjut diharapkan bervariatifnya klaster yang dihasilkan bukan hanya menu kinerja baik, menu kinerja sedang dan menu kinerja buruk.

DAFTAR PUSTAKA

- [1] N. Erlangga, S. Solikhun, and I. Irawan, "Penerapan Data Mining Dalam Mengelompokan Produksi Jagung Menurut Provinsi Menggunakan Algoritma K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 702–709, 2019, doi: 10.30865/komik.v3i1.1681.
- [2] R. Hidayat and H. Kusniyati, "Analisis Clustering Dalam Pengelompokan Penjualan Menggunakan Algoritma K-Means Pada Cafe 47°Coffee Clustering Analysis in Sales Grouping Using The K-Means Algorithm at Cafe 47°Coffee," vol. 7, no. 2, pp. 420–434, 2022, [Online]. Available: www.jurnal.unimed.ac.id.

- [3] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.
- [4] Erliyana, O. Nurdiawan, N. R, A. I. Purnamasari, and I. Ali, "Analisa Penjualan Mobil Dengan Menggunakan Algoritma K-Means Di PT. Mulya Putra Kencana," *J. Data Sci. Inform.*, no. 2, pp. 32–35, 2021, [Online]. Available: <http://publikasi.bigdatascience.id/index.php/jdsi/article/view/12>.
- [5] M. A. Satriawan, R. Andreswari, and O. N. Pratiwi, "Segmentasi Pelanggan Telkomsel Menggunakan Metode Clustering Dengan Rfm Model Dan Algoritma K-Means Telkomsel Customer Segmentation Using Clustering Method With Rfm Model And K-Means Algorithm," vol. 8, no. 2, pp. 2876–2883, 2021.
- [6] Feryanto, F. T. Kesuma, and S. P. Tamba, "Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 67–72, 2020, doi: 10.34012/jusikom.v2i2.376.
- [7] I. R. Mahartika and A. Wibowo, "Data Mining Klasterisasi dengan Algoritme K-Means untuk Pengelompokkan Provinsi Berdasarkan Konsumsi Bahan Bakar Minyak Nasional," *Pros. Semin. Nas. SISFOTEK (Sistem Inf. dan Teknol.*, vol. 3, no. 1, pp. 87–91, 2019, [Online]. Available: <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/108>.
- [8] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Semin. Nas. Teknol. Komput. Sains*, pp. 762–767, 2019.
- [9] D. Yanto, L. Probolinggo, C. Loyal, and K. Loyal, "Analisis RFM dan Algoritma K-Means untuk Clustering Loyalitas Customer," vol. 9, no. 1, pp. 0–8, 2019.
- [10] D. Astuti, "Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 1, no. 2, pp. 60–72, 2019, doi: 10.20895/inista.v1i2.71.
- [11] T. A. A. Sandi, M. Raharjo, J. L. Putra, and R. Ridwan, "20," *J. Pilar Nusa Mandiri*, vol. 14, no. 2, p. 239, 2018, doi: 10.33480/pilar.v14i2.950.