

# Penerapan Algoritma K-Means Menggunakan Model Recency Frequency Monetary Dalam Klasterisasi Penjualan Produk Sepatu

*by Hari Soetanto*

---

**Submission date:** 28-Aug-2023 09:20AM (UTC+0700)

**Submission ID:** 2152402680

**File name:** jurnal\_k\_means\_produk\_sepatu.pdf (281.82K)

**Word count:** 3559

**Character count:** 21179

## Penerapan Algoritma K-Means Menggunakan Model *Recensy Frequency Monetary* Dalam Klasterisasi Penjualan Produk Sepatu

Arief Rachman Hakim<sup>1\*</sup>, Hari Soetanto<sup>2</sup>

<sup>1,2</sup>Fakultas Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia

Email: <sup>1\*</sup>ariefrachmanhakim98@gmail.com, <sup>2</sup>hari.soetanto@budiluhur.ac.id  
(\* : corresponding author)

**Abstrak** - Selama masa pandemi Covid-19 penjualan barang atau jasa banyak yang mengalami penurunan. Seperti yang terjadi pada toko *online* Indoshoes yang mengalami penurunan penjualan sejak tahun 2021 serta penyebaran penjualan pada setiap wilayah tidak terjual secara merata. Toko Indoshoes memerlukan strategi yang tepat untuk mengetahui pengelompokan peminatan pelanggan pada suatu produk. Penerapan algoritma *K-Means* sangat membantu dalam pengelompokan produk berdasarkan analisis *Recensy*, *Frequency*, dan *Monetary* (RFM). Berdasarkan pengujian pada wilayah Jawa Barat didapatkan 3 data *cluster*, pengelompokan C1 dengan 7 produk sangat favorit, C2 dengan 83 produk tidak favorit, dan C3 dengan 16 produk lumayan favorit. Hasil penelitian ini diharapkan dapat membantu toko Indoshoes dalam mengambil keputusan dalam menentukan strategi *marketing* untuk meningkatkan penjualan toko Indoshoes.

**Kata Kunci:** *K-Means*, *Clustering*, RFM, Penjualan

### *Application Of The K-Means Algorithm Using The Recensy Frequency Monetary Model In The Clustering Of Shoe Product Sales*

**Abstract** - During the Covid-19 pandemic, sales of goods or services have decreased. As happened to the Indoshoes online store which has experienced a decline in sales since 2021 and the distribution of sales in each region is not sold evenly. Indoshoes stores need the right strategy to find out the grouping of customer interest in a product. The application of the *K-Means* algorithm is very helpful in grouping products based on *Recensy*, *Frequency*, and *Monetary* (RFM) analysis. Based on testing in the West Java region, 3 cluster data were obtained, grouping C1 with 7 very favorite products, C2 with 83 non-favorite products, and C3 with 16 fairly favorite products. The results of this study are expected to help Indoshoes stores in making decisions in determining marketing strategies to increase Indoshoes store sales.

**Keywords:** *K-Means*, *Clustering*, RFM, Sales

## 1. PENDAHULUAN

Transaksi jual beli suatu barang atau jasa selalu memiliki data yang disimpan. Data tersebut mengenai informasi setiap transaksi yang terjadi setiap waktu dan data tersebut hanya berisi informasi tertentu, sebagai contoh dalam penjualan barang berupa nama barang, jumlah yang terjual dan harga barang. Informasi yang penting harus disimpan karena bisa saja informasi ini akan digunakan pada masa mendatang dan harus disimpan pada sebuah sistem yang baik. Sistem ini dinamakan *database*. *Database* merupakan bagian dari sistem informasi secara keseluruhan [1]. *Database* tidak hanya untuk menyimpan sebuah data, juga mampu membantu kita untuk memperoleh data dengan mudah jika sewaktu saat kita membutuhkan data tersebut. Namun, *database* tidak dapat melakukan analisa suatu data dengan baik, oleh karena itu dibutuhkanlah *data mining*. *Data mining* merupakan suatu proses iterasi menganalisa data untuk mencari suatu informasi [2].

Toko *online* Indoshoes merupakan toko *online* yang menjual berbagai jenis sepatu wanita sejak tahun 2015 di *marketplace* Lazada. Sejak satu tahun terakhir penjualan pada toko Indoshoes mengalami ketidakstabilan yang disebabkan oleh pandemi serta persaingan antar toko *online* semakin ketat. Strategi pemasaran yang dijalankan oleh toko Indoshoes saat ini adalah dengan cara beriklan pada Lazada itu sendiri. Toko Indoshoes memerlukan strategi pemasaran yang lebih efektif agar dapat meningkatkan kembali tingkat penjualan toko Indoshoes.

Berbagai penelitian yang terkait dengan penentuan strategi penjualan menggunakan metode *data mining* sudah banyak dilakukan. Seperti yang dilakukan Normah, Siti Nurajizah, dan Arinda Salbinda [3] dengan judul Penerapan *Data Mining* Metode *K-Means Clustering* Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten. Pada penelitian tersebut peneliti menggunakan metode *clustering K-Means* sebagai *data mining*. Penelitian ini menggunakan dataset stok awal, stok terjual, dan stok akhir. Hasil dari penelitian ini diperoleh 3 *cluster* yaitu ada 11 artikel sangat laris, artikel laris 55 artikel, dan 34 artikel kurang laris. Parameter dari dataset tersebut dirasa masih kurang sebagai faktor analisa untuk penentuan *cluster* penjualan. Karena hal tersebut pada penelitian kali ini peneliti menggunakan tambahan metode pemodelan RFM karena metode tersebut sangat efektif untuk menganalisa segmentasi pelanggan [4]

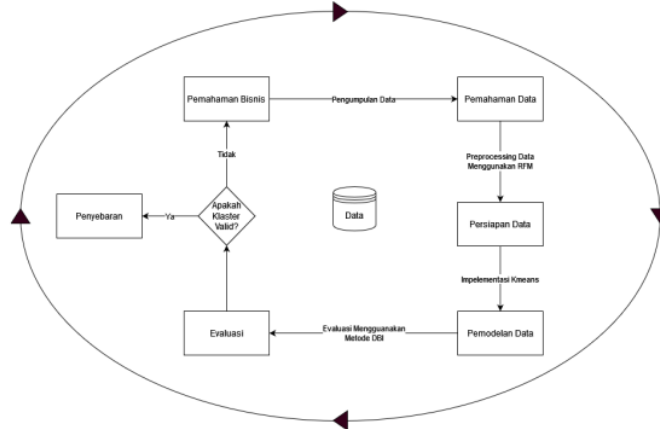
Penelitian lain dilakukan oleh Felix Toknady Kesuma dan Feryanto [5] dengan judul Penerapan *Data Mining* Untuk Menentukan Penjualan *Sparepart* Toyota Dengan Metode *K-Means Clustering*. Penelitian ini menggunakan dataset stock awal dan stock terjual. Penelitian ini berhasil menghasilkan 3 *cluster* produk *sparepart* toyota yaitu barang paling laris sebanyak 15, barang laris sebanyak 45, dan 13 barang kurang laris. Tapi hasil *cluster* tersebut masih belum bisa dipastikan keabsahannya karena pada penelitian tersebut tidak melakukan pengujian validasi. Karena hal tersebut pada penelitian kali ini peneliti menggunakan metode DBI untuk memvalidasi hasil *cluster* dari metode *K-Means*, sehingga *cluster* tersebut terbukti valid [6].

Berdasarkan penelitian sebelumnya, peneliti dapat mengidentifikasi strategi pemasaran yang baik dengan mempromosikan produk berdasarkan preferensi konsumen di masing-masing daerah. Oleh karena itu, perlu dilakukan analisis pengolahan data menggunakan model RFM. RFM, digunakan untuk menunjukkan perilaku pelanggan [7], kapan pelanggan terakhir bertransaksi (*recency*), seberapa sering pelanggan bertransaksi (*frequency*) dan seberapa besar jumlah uang dari pelanggan atas transaksinya (*monetary*) [8]. Kemudian perlu dilakukan analisis data transaksi penjualan, agar dapat diketahui pola penjualan dari informasi data tersebut. Salah satu teknik analisis *data mining* yang digunakan adalah teknik *clustering* [9]

*K-Means Clustering* adalah algoritma yang paling baik digunakan dalam mengimplementasikan *clustering* [10]. Algoritma ini cukup mudah diimplementasikan dan dijalankan, relatif cepat, mudah disesuaikan, dan banyak digunakan. Algoritma *K-Means* dimulai dengan menentukan banyaknya *K*, *K* adalah jumlah *cluster* yang akan dibentuk, kemudian tetapkan titik pusat *cluster* secara acak atau random, titik pusat *cluster* sering disebut dengan nama *centroid*, Hitung setiap jarak data yang ada terhadap masing-masing *centroid* dengan menggunakan rumus *Euclidean Distance*. Kemudian alokasikan masing-masing objek ke dalam *cluster* berdasarkan jarak minimum [11]. Lakukan langkah tersebut hingga nilai *centroid* tidak berubah. Selanjutnya hasil dari kedua metode antara RFM dan *clustering K-Means* kemudian dilakukan evaluasi *clustering* dengan menggunakan *Davies-Bouldin Index* (DBI) yang digunakan untuk mengevaluasi *cluster* [12]. Semakin kecil nilai *Davies-Bouldin Index* (DBI) yang diperoleh ( $\text{non-negatif } \geq 0$ ), maka semakin baik *cluster* yang diperoleh dari pengelompokan menggunakan algoritma *clustering* [13].

## 2. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah CRISP-DM. CRISP-DM adalah metode yang menggunakan model dari proses pengembangan data. Ada enam tahapan proses CRISP-DM yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi dan penyebaran [14]. Seperti yang disajikan pada Gambar 1.



Gambar 1. Proses Diagram CRISP-DM

### 2.1 Pemahaman Bisnis

Bisnis yang dilakukan oleh toko *online* Indoshoes merupakan toko *online* yang bergerak di bidang penjualan sepatu wanita yang berada di *marketplace* Lazada. Dari proses mempelajari dan memahami bisnis ini selanjutnya peneliti dapat mengetahui metode yang tepat untuk menyelesaikan permasalahan penjualan pada toko *online* Indoshoes. Data penjualan milik toko *online* Indoshoes selanjutnya akan difilter sesuai provinsi pengirimannya. Dari data yang telah terfilter akan dikelompokkan atau diklasterisasi. Untuk mempermudah proses klasterisasi, maka penelitian ini akan menggunakan model RFM (*recency, frequency, monetary*)

## 2.2 Pemahaman Data

Pemahaman data dilakukan dengan cara memeriksa kualitas data. Data yang digunakan pada penelitian ini adalah data penjualan Toko *Online Indoshoes* selama 29/11/2020 sampai dengan 09/04/2022 dengan jumlah data sebanyak 12.976 data.

## 2.3 Persiapan Data

Dari 12.976 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: *order\_item\_id*, *payment\_date*, *product\_name*, *price*, *quantity*, *amount*, *buyer\_city*.

## 2.4 Pemodelan Data

Metode yang digunakan pada proses pemodelan data di penelitian ini adalah algoritma *K-Means*. Data input untuk proses klusterisasi yaitu data RFM yang sudah dinormalisasi pada tahap sebelumnya. Model RFM dipergunakan untuk mempermudah dalam melakukan klusterisasi penelitian ini untuk mengsegmentasi pelanggan dari data transaksi [15]. Dari data transaksi akan diubah menjadi data yang lebih padat dan menghasilkan data dengan parameter *recency* (tanggal terakhir produk), *frequency* (jumlah produk terjual), *monetary* (jumlah pendapatan dari produk terjual).

## 2.5 Evaluasi

Tahapan ini menggunakan metode DBI (*Davies Buildin Index*) untuk melakukan evaluasi *cluster* dan menghitung nilai rata-rata untuk setiap titik dalam kumpulan data. *Davies Buildin Index* merupakan salah satu cara untuk mengukur efektivitas *cluster* dalam teknik pengelompokan. Kohesi didefinisikan sebagai jumlah kedekatan data dengan pusat *cluster* dari *cluster* yang diikuti. Pemisahan didasarkan pada jarak dari pusat *cluster* ke *cluster*.

## 2.6 Penyebaran

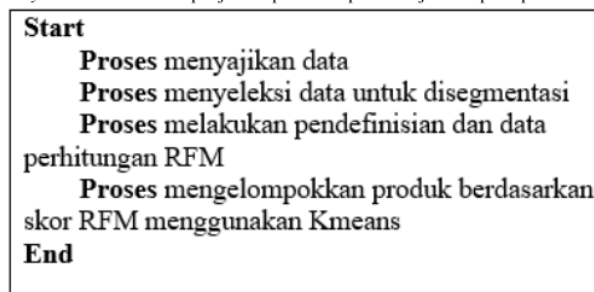
Di tahap ini pengetahuan atau informasi yang telah dihasilkan algoritma *data mining K-Means* diimplementasikan menjadi sebuah aplikasi berbasis web sehingga dapat digunakan dengan mudah oleh pihak pengguna.

## 3. HASIL DAN PEMBAHASAN

Dari 12.976 data penjualan *Indoshoes* itu terdapat banyak transaksi yang dilakukan oleh pembeli dari beragam provinsi di Indonesia. Karena pada penelitian ini peneliti ingin meneliti pembelian dari Jawa Barat maka data tersebut harus di-filter terlebih dahulu dengan hanya memilih data pembelian dari Jawa Barat. Setelah difilter data yang ditemukan berjumlah 4.916 data pembelian dari Jawa Barat. Selanjutnya 4.916 data tersebut yang akan diproses menggunakan metode RFM kemudian dilanjutkan dengan menggunakan algoritma *clustering K-Means* dan dievaluasi dengan metode *Davies-Bouldin Index*

### 3.1 Implementasi Metode RFM

Algoritma penerapan preprocessing metode RFM pada penelitian penerapan algoritma *K-Means* menggunakan model *recency frequency monetary* dalam klusterisasi penjualan produk sepatu disajikan seperti pada Gambar 2.



Gambar 2. Algoritma RFM

- Mencari nilai *Recency* dengan menghitung waktu transaksi terakhir dari produk yang dibeli dan dibandingkan dengan tanggal penelitian ini dilakukan. Penelitian ini dilakukan pada tanggal 17 Juni 2022. Jumlah hari dari selisih tanggal tersebut akan dijadikan nilai *recency* dapat dilihat pada Tabel 1.

**Tabel 1.** Tabel *Recency*

No.	Nama Produk	Tanggal Terakhir	Recency
1	1760 Mulan Sepatu Sandal Casual	12/02/2022	125
2	1944 Wedges Selop Bordir Mulan	20/03/2022	89
3	1980 Sandal Wedges Strappy Mulan	06/04/2022	72
...	...	...	...
105	Wedges Sandal Mulan 782	05/04/2022	73
106	Wedges Sepatu Sandal Kokop Mulan 2171	07/04/2022	71

- b. Mencari nilai *frequency*, nilai *frequency* hanya perlu mengambil jumlah item yang terjual dari produk dalam satu kali periode data, dapat dilihat pada Tabel 2.

**Tabel 2.** Tabel *Frequency*

No.	Nama Produk	Frequency
1	1760 Mulan Sepatu Sandal Casual	3
2	1944 Wedges Selop Bordir Mulan	28
3	1980 Sandal Wedges Strappy Mulan	7
...	...	...
105	Wedges Sandal Mulan 782	13
106	Wedges Sepatu Sandal Kokop Mulan 2171	245

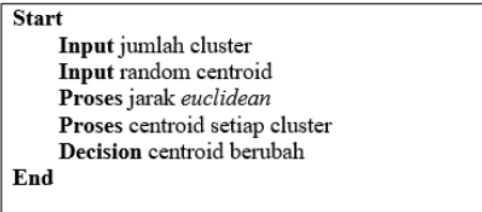
- c. Mencari nilai *monetary*, nilai *monetary* adalah berapa jumlah pendapatan yang dihasilkan oleh produk tersebut selama satu periode data. Lakukan perkalian nilai *frequency* produk dengan harga dari produk tersebut seperti pada Tabel 3.

**Tabel 3.** Tabel *Monetary*

No.	Nama Produk	Monetary
1	1760 Mulan Sepatu Sandal Casual	Rp201,000
2	1944 Wedges Selop Bordir Mulan	Rp2,240,700
3	1980 Sandal Wedges Strappy Mulan	Rp575,500
...	...	...
105	Wedges Sandal Mulan 782	Rp708,900
106	Wedges Sepatu Sandal Kokop Mulan 2171	Rp19,890,802

### 3.2 Implementasi Algoritma *K-Means*

Setelah pemodelan dari RFM sudah dilakukan, selanjutnya masuk ke perhitungan Algoritma *K-Means* yang disajikan pada gambar 3.



**Gambar 3** Algoritma *K-Means*

Langkah pertama yang dilakukan adalah menentukan jumlah *cluster*. Dalam penelitian ini *cluster* yang ditentukan adalah *cluster* paling favorit, *cluster* lumayan favorit, dan *cluster* kurang favorit. Setiap *cluster* yang ditentukan sebelumnya akan memiliki pusat *cluster* atau yang disebut *centroid*. Nilai *centroid* itu akan dihitung secara berulang-ulang sampai ditemukan *centroid* yang sempurna atau *centroid* yang tidak akan berubah-ubah lagi, sehingga *cluster* yang dihasilkan juga tetap atau tidak ada produk yang berpindah-pindah cluster. Untuk menghitung iterasi *cluster* yang pertama, *centroid* awal akan ditentukan secara random atau acak.

**Tabel 4.** Nilai *Centroid* Awal

Cluster	Recency	Frequency	Monetary
C1	161	63	Rp3,966,000
C2	77	40	Rp3,587,700
C3	265	30	Rp1,898,900

Setelah *centroid* awal sudah ditentukan seperti yang disajikan pada Tabel 4. Langkah selanjutnya adalah menghitung jarak seluruh data dengan masing-masing *cluster* dengan menggunakan persamaan *Euclidean Distance* disajikan pada Persamaan (1).

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (1)$$

Keterangan:

- $D_e$  = *Euclidean Distance*
- $i$  = banyaknya objek
- $(x, y)$  = merupakan koordinat objek
- $(s, t)$  = merupakan koordinat *centroid* (titik pusat *cluster*)

#### 1760 Mulan Sepatu Sandal Casual

$$\begin{aligned} \text{Jarak Ke C1} &= \sqrt{\frac{(161 - 125)^2 + (63 - 2)^2 +}{(3966000 - 201000)^2}} \\ &= 3765000 \end{aligned}$$

$$\begin{aligned} \text{Jarak Ke C2} &= \sqrt{\frac{(77 - 125)^2 + (40 - 2)^2 +}{(3587700 - 201000)^2}} \\ &= 3386700 \end{aligned}$$

$$\begin{aligned} \text{Jarak Ke C3} &= \sqrt{\frac{(265 - 125)^2 + (30 - 2)^2 +}{(1898900 - 201000)^2}} \\ &= 1697900 \end{aligned}$$

Untuk mengetahui **1760 Mulan Sepatu Sandal Casual** masuk dalam *cluster* mana, selanjutnya bandingkan seluruh jarak dan ambil nilai yang paling kecil. Karena jarak ke C3 adalah nilai terkecil maka **1760 Mulan Sepatu Sandal Casual** masuk ke *cluster* C3. Setelah semua data dihitung dan data *cluster* telah diperoleh, langkah selanjutnya adalah menghitung jumlah *mean cluster* yang diperoleh dari RFM untuk mendapatkan *centroid* untuk iterasi kedua yang disajikan pada Persamaan (2).

$$\bar{v}_{i,j} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \quad (2)$$

Keterangan:

- $\bar{v}_{i,j}$  = *centroid/rata-rata cluster ke-i untuk variabel ke-j*
- $N_i$  = jumlah data yang menjadi anggota *cluster ke-i*
- $i, k$  = indeks dari *cluster*
- $j$  = indeks dari variabel
- $x_{kj}$  = nilai data ke-k yang ada di dalam *cluster* tersebut untuk variabel ke-j

#### Rata-rata pada C1

$$\text{Rencency} = \frac{161+73+71+\dots+71+71}{21} = \frac{2550}{21} = 121.4285714$$

$$\text{Frequency} = \frac{63+52+50+\dots+156+245}{21} = \frac{3845}{21} = 183.0952381$$

$$\begin{aligned} \text{Monetary} &= \frac{3966000 + 4492218 + 4402501 + \dots + 10911650 + 19890802}{21} \\ &= \frac{219659178}{21} = 10459961 \end{aligned}$$

Tabel 5. *Centroid* Iterasi Kedua

Cluster	Rencency	Frequency	Monetary
C1	121.4285714	183.0952381	Rp10,459,961
C2	119.3333333	47.66666667	Rp3,435,940
C3	220.9756098	11.31707317	Rp561,661

Persamaan *Euclidean Distance* pada penelitian ini berhenti sampai iterasi ke 7 seperti disajikan pada Tabel 5. karena nilai *centroid* sudah akhir atau bernilai sama dengan iterasi selanjutnya dapat dilihat pada tabel 6.

Tabel 6. *Centroid* Akhir

Cluster	Recency	Frequency	Monetary
C1	121.4285714	183.0952381	Rp10,459,961
C2	119.3333333	47.66666667	Rp3,435,940
C3	220.9756098	11.31707317	Rp561,661

Menghitung data untuk menentukan *cluster* menggunakan *centroid* akhir menghasilkan data berikut:

Tabel 7. Tabel Hasil Klasterisasi

No.	Nama Produk	C1	C2	C3	Cluster
1	1760 Mulan Sepatu Sandal Casual	3562781	296827	72226500	C3
2	1944 Wedges Selop Bordir Mulan	3552781	286827	72216500	C3
3	1980 Sandal Wedges Strappy Mulan	3412781	146827	72076500	C3
...	...	...	...	...	...
105	Wedges Sandal Mulan 782	3472781	206827	72136500	C3
106	Wedges Sepatu Sandal Kokop Mulan 2171	3522781	256827	72186500	C1

Dari hasil klasterisasi data yang diperoleh seperti disajikan pada Tabel 7. terhitung *cluster* C1 sebanyak 7 data, *cluster* C2 sebanyak 16 data, *cluster* C3 sebanyak 83 data.

### 3.3 Penentuan Nama Cluster Berdasarkan RFM

Untuk penentuan nama *cluster* diperlukan nilai dari *centroid* akhir seperti Tabel 8.

Tabel 8. *Centroid* Akhir RFM

Cluster	Recency	Frequency	Monetary
C1	121.4285714	183.0952381	Rp10,459,961
C2	119.3333333	47.66666667	Rp3,435,940
C3	220.9756098	11.31707317	Rp561,661

Setelah itu perlu diranking nilai RFM dari setiap *cluster*. Ketentuan dalam pe-rangking-an adalah sebagai berikut:

- Semakin kecil nilai *recency* maka nilai *ranking*nya semakin besar.
- Semakin besar nilai *frequency* maka nilai *ranking*nya semakin besar.
- Semakin besar nilai *monetary* maka nilai *ranking*nya semakin besar.
- Nilai rata-rata terbesar dari *ranking* milik klaster adalah klaster paling favorit, sedangkan nilai rata-rata terkecil adalah klaster kurang favorit, sisanya adalah klaster lumayan favorit.

Berikut adalah nilai *ranking* yang didapatkan oleh setiap klaster berdasarkan *centroid* akhir dapat dilihat pada Tabel 9.

Tabel 9. Penentuan Nama Cluster

Cluster	R	F	M	Rata-rata	Nama Cluster
C1	1	1	1	1	Kurang Favorit
C2	3	3	3	3	Paling Favorit
C3	2	2	2	2	Lumayan Favorit

### 3.4 Evaluasi Cluster Menggunakan DBI

Peneliti menggunakan metode *Davies-Bouldin Index* (DBI) untuk mengukur validitas *cluster* yang diperoleh dengan algoritma *K-Means* dalam penelitian ini. Langkah pertama dari metode DBI adalah menghitung nilai SSW (*Sum of Square Within-Cluster*) dari semua data. Berikut adalah perhitungan SSW di *database*.

#### 1760 Mulan Sepatu Sandal Casual

SSW1

$$= \sqrt{(125 - 132.7143)^2 + (3 - 355.1429)^2 + (201000 - 18374982.2857)^2}$$

$$= 18173982$$

SSW2

$$= \sqrt{(125 - 111.3750)^2 + (3 - 90.8750)^2 + (201000 - 6139787.6250)^2}$$

$$= 5938788$$

SSW3

$$= \sqrt{(125 - 220.6867)^2 + (3 - 11.7590)^2 + (201000 - 592310.2289)^2}$$

$$= 391310$$

Setelah menghitung SSW pada seluruh data. Selanjutnya adalah menghitung rata-rata nilai SSW1 sampai SSW3. Maka perhitungannya seperti pada Tabel 10.

**Tabel 10.** Hasil Perhitungan SSW

Centroid	Jarak	Anggota	Rata-Rata
SSW1	31765683.76	7	4537954.822
SSW2	28881194.56	16	1805074.66
SSW3	44841782.23	83	540262.4365

Kemudian menghitung jarak antar *cluster* menggunakan rumus SSB (*Sum of Square Between-cluster*) dengan menghitung jarak antar suatu *cluster*.

SSB1,2

$$= \sqrt{(132.7143 - 111.3750)^2 + (355.1429 - 90.8750)^2 + (18374982.2857 - 6139787.6250)^2}$$

$$= 12235194.66$$

SSB1,3

$$= \sqrt{(132.7143 - 220.6867)^2 + (355.1429 - 11.7590)^2 + (18374982.2857 - 592310.2289)^2}$$

$$= 17782672.06$$

SSB2,3

$$= \sqrt{(111.3750 - 220.6867)^2 + (90.8750 - 11.7590)^2 + (18374982.2857 - 592310.2289)^2}$$

$$= 5547477.40$$

Setelah mendapatkan nilai SSW dan SSB, tahap selanjutnya adalah menghitung jarak antar SSW dan menghitung rasio untuk mendapatkan nilai rasio maksimum (R-max). Gunakan ini untuk menghitung nilai DBI.

$$R_{1,2} = \frac{4537954.822 + 1805074.66}{12235194.66} = 0.5184$$

$$R_{1,3} = \frac{4537954.822 + 540262.4365}{17782672.06} = 0.2856$$

$$R_{2,3} = \frac{1805074.66 + 540262.4365}{5547477.40} = 0.4228$$

**Tabel 11.** Nilai Rasio R-Max

R	1	2	3	R-MAX
1	0	0.467	0.414	0.467
2	0.467	0	0.380	0.467
3	0.414	0.380	0	0.414

Setelah nilai rasio maksimum (*R-Max*) diperoleh seperti disajikan pada Tabel 11. Langkah selanjutnya adalah menghitung nilai DBI dengan cara menghitung rata-rata dari nilai *R-Max*. Dari perhitungan tersebut didapatkan nilai DBI sebesar 0.449. Nilai DBI dapat dikatakan baik karena nilainya mendekati 0 (non-negatif).

#### 4. KESIMPULAN

Berdasarkan penelitian yang dilakukan, peneliti mendapatkan hasil pengujian analisa produk menggunakan analisis *Recency Frequency Monetary* berhasil mendapatkan data dari provinsi jawa barat dengan kriteria pengelompokan C1 dengan 7 produk sangat favorit, C2 dengan 83 produk tidak favorit, dan C3 dengan 16 produk lumayan favorit. Pengimplementasian *data mining* berhasil dilakukan dengan menggunakan algoritma *K-Means* serta evaluasi *cluster K-Means* menggunakan DBI mendapatkan nilai 0.449, nilai DBI dapat dikatakan baik karena nilai mendekati 0 (non-negatif).

Adapun saran yang ingin disampaikan terkait dengan penelitian yang dilakukan Pada pengembangan selanjutnya kiranya aplikasi *data mining* ini dapat melakukan proses analisis tanpa harus melakukan *upload* ulang dataset, Kombinasi dengan metode atau pendekatan lain untuk menentukan dan memilih pusat *cluster* yang lebih tepat untuk tipe data yang sama atau lainnya.

#### DAFTAR PUSTAKA

- [1] A. D. Hardiansyah, D. C. Nugrahaeni, dan P. Dewi, "Perancangan basis data sistem informasi perwira tugas belajar (sipatubel) pada kementerian pertahanan," *Senamika*, vol. 1, no. 2, pp. 222–233, 2020.
- [2] Kantardzic, M. 2003. *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons.
- [3] N. Normah, S. Nurajizah, and A. Salbinda, "Penerapan data mining metode k-means clustering untuk analisa penjualan pada toko fashion hijab Banten," *J. Tek. Komput.*, vol. 7, no. 2, 2021,
- [4] P. A. Wicaksana, I. Bagus, A. Swamardika, and R. S. Hartati, "Literature review analisis perilaku pelanggan menggunakan rfm model," vol. 21, no. 1, 2022.
- [5] F., 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.
- [6] A. F. Khairati, A. Adlina, G. Hertono, and B. Handari, "kajian indeks validitas pada algoritma k-means enhanced dan k-means mmca," *Prism. Pros. Semin. Nas. Mat.*, vol. 2, pp. 161–170, 2019.
- [7] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis segmentasi pelanggan menggunakan kombinasi rfm model dan teknik clustering," *J. Terap. Teknol. Inf.*, vol. 2, no. 1, pp. 23–32, 2018.
- [8] Birant, D. 2011. *Data Mining Using RFM Analysis, in Knowledge-Oriented Applications in Data Mining*, Turkey: Kimito Funatso, pp. 91–108.
- [9] A. N. M. Hadi Yansah, E. Y. Puspaningrum, and A. M. Rizky, "Implementasi kombinasi metode rfm dan algoritma kmeans untuk klasterisasi penjualan minuman kemasan," *Pros. Semin. Nas. Inform. Bela Negara*, vol. 2, pp. 122–126, 2021.
- [10] W. M. Baihaqi, K. Indartono, and S. Banat, "Penerapan teknik clustering sebagai strategi pemasaran pada penjualan buku di tokopedia dan shopee," *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 2, 2019.
- [11] B. Ginting and F. Riandari, "Implementasi metode k-means clustering dalam pengelompokan bibit tanaman kopi arabika," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 2, pp. 151–157, 2020.
- [12] Sitompul, Bernad Jumadi Dehotman. 2018. "peningkatan hasil evaluasi clustering davies bouldin dengan penentuan titik pusat cluster awal k means".
- [13] S. Butsianto and N. Saepudin, "penerapan data mining terhadap minat siswa dalam mata pelajaran matematika dengan metode k-means," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 1, pp. 51–59, 2020.
- [14] 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.
- [15] A. P. Pramudiansyah, "Segmentasi Pelanggan Menggunakan Algoritma K-Means Berdasarkan Model Recency Frequency Monetary," *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 06–19, 2021.

# Penerapan Algoritma K-Means Menggunakan Model Recensy Frequency Monetary Dalam Klasterisasi Penjualan Produk Sepatu

---

ORIGINALITY REPORT

---

**10**%  
SIMILARITY INDEX

%  
INTERNET SOURCES

**9**%  
PUBLICATIONS

**6**%  
STUDENT PAPERS

---

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

---

3%  
★ Submitted to Universitas Pamulang  
Student Paper

---

---

Exclude quotes      On  
Exclude bibliography      On

Exclude matches      < 1%