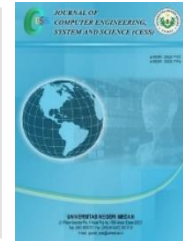


Contents list available at [www.jurnal.unimed.ac.id](http://www.jurnal.unimed.ac.id)

**CESS**  
**(Journal of Computing Engineering, System and Science)**

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



**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**

**Raihan Hidayat<sup>1</sup>, Harni Kusniyati<sup>2</sup>**

*Teknik Informatika, Universitas Mercu Buana, Jakarta*

*Jl. Meruya Selatan No.1, Kembangan, Jakarta Barat, Indonesia*

*email: <sup>1</sup>[41518010096@student.mercubuana.ac.id](mailto:41518010096@student.mercubuana.ac.id), <sup>2</sup>[harni.kusniyati@mercubuana.ac.id](mailto:harni.kusniyati@mercubuana.ac.id)*

Submitted: 15 Juni 2022 | Revision: 07 Juli 2022 | Accepted: 14 Juli 2022

**Abstrak** - Cafe 47°Coffee adalah cafe yang bergerak dibidang kuliner yang menyediakan makanan maupun minuman. Dengan banyaknya pesaing bisnis dibidang cafe maka dari itu pengusaha cafe 47°Coffee diharuskan mampu bersaing dengan memanfaatkan dari data transaksi untuk menentukan suatu strategi penjualan. Data yang akan digunakan pada penelitian ini yaitu data penjualan dari bulan Juli 2021 sampai Februari 2022. Pengusaha cafe 47°Coffee memanfaatkan Data Mining untuk membantu mengambil keputusan strategi penjualan agar dapat mengetahui produk yang harus di tingkatkan dan memberi solusi pengambil keputusan, Metode Clustering dengan menggunakan algoritma K-Means adalah salah satu metode yang akan digunakan pada penelitian ini. Pengusaha Cafe 47°Coffee mengelompokkan produk kedalam 3 kriteria yaitu sangat laku, laku, dan tidak laku. Data diolah dengan perhitungan manual dan menggunakan tools Rapid Miner untuk melakukan pengujian, sehingga dapat hasil akhir berupa 4 item cluster sangat laku yakni V60, Caramel Coffee Latte, Red Velvet, dan Coklat, 7 item cluster laku yakni Cappuccino, Vanilla Coffe Latte, Hazelnut, Taro, Lychee Tea, Japanese, dan Green tea, 14 item cluster tidak laku yakni Americano, Espresso, Moccacino, Kopi tubruk, Vietnam Drip, Kopi sundan, Avocado Coffee Latte, Match, Strawberry Tea, Milkshake Strawberry, Banana, Coffee Lemon, Kopi susu, dan risol. Dari proses clustering menggunakan algoritma k-means diatas diperoleh nilai DBI (Davice Bouldin Index) dengan nilai -0,813. Hasil ini dapat dimanfaatkan pengusaha Cafe 47°Coffee untuk meningkatkan strategi penjualan dan manajemen stok.

**Kata Kunci:** *K-Means; Penjualan; Clustering; Cafe; Data Mining*

**Abstract** - Cafe 47°Coffee is a cafe engaged in the culinary field that provides food and drinks. With so many business competitors in the cafe sector, therefore 47°Coffee cafe entrepreneurs are required to be able to compete by utilizing transaction data to determine a strategy. The data that will be used in this research is sales data from July 2021 to February 2022. 47°Coffee cafe entrepreneurs utilize Data Mining to help make sales strategy decisions in order to find out which products must be improved and provide solutions for decision makers, the Clustering Method using K-Means algorithm is one of the methods that will be used in this research. The 47°Coffee Cafe entrepreneurs group their products into 3 criteria, namely very selling, selling well, and not selling well. The data is processed by manual calculations and using Rapid Miner tools to perform testing, so that the final results are in the form of 4 very selling cluster items, namely V60, Caramel Coffe Latte, Red Velvet, and Chocolate, 7 cluster items selling well namely Cappucino, Vanilla Coffe Latte, Hazelnut, Taro, Lychee Tea, Japanese, and Greentea, 14 cluster items not selling, namely Americano, Espresso, Moccacino, Kopi tubruk, Vietnam Drip, Sundanese coffee, Avocado Coffe Latte, Match, Strawberry Tea, Strawberry Milkshake, Banana, Cofee Lemon, Coffee milk , and risol. From the clustering process using the k-means algorithm above, the DBI (Davice Bouldin Index) value is obtained with a value of -0.813. This result can be used by the 47°Coffee Cafe entrepreneurs to improve their sales strategy and stock management.

**Keywords:** K-Means; Sales; Clustering; Café; Data Mining

---

## 1. PENDAHULUAN

Dalam melakukan rutinitas sehari-hari pasti setiap orang mempunyai rasa jenuh dan bosan. Seperti bekerja, sekolah, dan kesibukan-kesibukan lainnya. Oleh karena ini, biasanya orang melakukan hal yang dapat menghilangkan rasa jenuh karena rutinitas. Hal yang sering dilakukan seperti: menonton film di bioskop maupun di aplikasi penyedia film, menyanyi di tempat karaoke, dan banyak orang juga yang memilih bersantai di cafe. Cafe dipilih sebagai tempat yang paling tepat untuk bersantai karena harganya yang biasanya cukup terjangkau. Di bidang kafe banyaknya pesaing bisnis yang mengharuskan pengusaha kafe untuk mampu bersaing dengan pengusaha cafe lainnya, Pengusaha café dituntut untuk mengambil keputusan yang tepat dalam menentukan strategi penjualan, pengusaha cafe dapat memanfaatkan data transaksi untuk memudahkan dalam menentukan strategi penjualan.

Di era globalisasi saat ini, laju perkembangan teknologi yang semakin meningkat dapat dimanfaatkan untuk memperlancar arus informasi yang siap digunakan untuk membantu mengambil keputusan strategi penjualan agar pengusaha cafe mudah mengetahui produk yang harus ditingkatkan dengan memanfaatkan transaksi penjualan (Data Mining). Data Mining dimanfaatkan untuk memberi solusi pengambil keputusan untuk mengembangkan bisnis. Salah satu metode yang digunakan pada penelitian ini adalah pengelompokan Clustering.

Cafe 47°Coffee adalah salah satu cafe yang bergerak dibidang kuliner yang menyediakan makanan dan minuman. Cafe 47°Coffee terletak di Jl.KH.Hasyim Ashari Gg. Kampung Bakti 2 RT.01 RW.07 No.47. Permasalahan yang terjadi sampai saat ini Cafe 47°Coffee masih kesulitan dalam menentukan makanan dan minuman apa saja yang paling sering dicari atau dipesan oleh konsumen sehingga sering terjadinya penumpukan stok produk yang tidak laku dan habisnya produk yang sering dicari konsumen.

Metode pada penelitian ini yang digunakan adalah Data mining Clustering K-Means. K-Means Clustering adalah metode data mining atau metode penganalisaan dari suatu data dengan proses pemodelan tanpa adanya supervise dan K-Means merupakan salah satu metode dengan proses pengelompokan data yang sistem partisi. Clustering terdapat dua jenis data clustering yang sering digunakan dalam pengelompokan kan dari suatu data, dua jenis data tersebut yaitu Hierarchical dan Non-Hierarchical dan K-Means adalah salah satu metode data clustering non-hierarchical [1].

Dengan mengimplementasikan Algoritma K-Means Clustering dengan optimal diharapkan hasil clustering yang didapat lebih akurat dan bisa digunakan untuk pengelompokan penjualan yang akan dikelompokkan menjadi 3 yaitu: Sangat laku, laku, dan Tidak laku, dan dapat membantu pengusaha Cafe 47°Coffee untuk meningkatkan strategi penjualan dan manajemen stok.

## **2. LANDASAN TEORI**

### **2.1. Data Mining**

Data mining adalah metode yang dapat digunakan untuk pengolahan data guna menemukan pola yang tersembunyi dari data yang ingin diolah. Data yang diolah dengan menggunakan data mining akan menghasilkan suatu pengetahuan baru dari sumber data lama, sehingga pengetahuan baru itu akan digunakan saat menentukan keputusan di masa yang akan datang. Data mining juga salah satu metode yang pengolahannya data berskala yang besar oleh sebab itu data mining sangatlah berguna dalam beberapa bidang kehidupan yaitu dalam sektor teknologi, kesehatan, industri, dan lain-lain[2]

Data mining juga dapat diartikan sebagai gabungan dari suatu nilai, pengalaman, informasi kontekstual dan pandangan pakar untuk menciptakan pengalaman dan informasi baru.

### **2.2. Clustering**

Clustering adalah salah satu teknik data mining yang digunakan untuk menganalisis suatu data untuk memecahkan permasalahan dengan cara mengelompokan data atau mempartisi dari suatu dataset kedalam sebest. Analisis cluster adalah bisa dikatakan data yang terdapat dalam satu cluster tetapi memiliki kesamaan yang rendah.

Pada gambar 1 diatas misalkan data tersebut yaitu data karyawan sederhana yang memiliki dua atribut, yaitu atribut penghasilan dan atribut umur. Sehingga data karyawan tersebut akan dibagi menjadi tiga cluster yaitu karyawan usia muda tetapi penghasilan tinggi (cluster 1), karyawan usia muda dan tua penghasilan tinggi (Cluster 2), dan karyawan usia tua dan penghasilan rendah[3].

### **2.3. K-Means**

K-Means pertama kali dikenalkan pada tahun 1979 oleh MacQueen JB. K-Means adalah suatu algoritma yang digunakan dengan cara mengelompokkan decara partisi yang memisahkan data ke dalam kelompok yang berbeda. Penggunaan algoritma k-means ini dalam proses clustering tergantung kepada data yang ingin diolah dan konklusi yang ingin dicapai di akhir proses[4]. Terdapat aturan dalam penggunaan algoritma k-means sebagai berikut :

- a Berapa jumlah cluster yang ingin dimasukkan
- b Hanya memiliki atribut yang bertipe numerik.

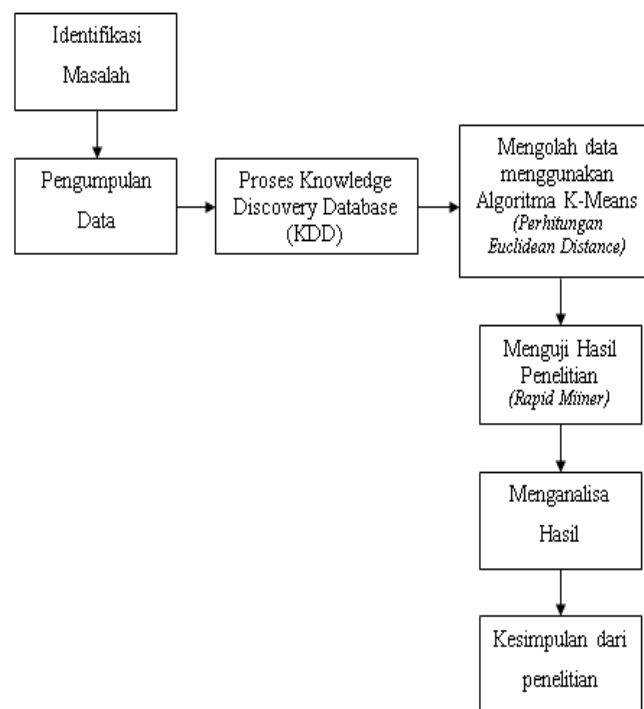
Pada dasarnya algoritma k-means hanya mengambil sebagian saja dari banyaknya data yang didapat kemudian akan dijadikan pusat cluster awal. Menentukan cluster awal ini secara acak dari populasi data. Kemudian algoritma k-means akan menguji masing-masing dari setiap komponendengan ousat cluster tergantung jarak minimum antar komponen dengan stiap-tiap pusat cluster. Selanjutnya posisi pusat akan clusterkan dihitung kembali sampai semua komponen data digolongkan kedalam tiap-tiap cluster dan terakhir akan terbentuk cluster baru[4].

### 3. METODE PENELITIAN

Pada penelitian ini metode yang dipakai adalah metode data mining, ada beberapa tahap metode penelitian, sebagai berikut:

#### 3.1. Tahap Penelitian

Tahapan Penelitian adalah langkah-langkah dari penelitian yang menjadi acuan untuk melakukan tahapan dari penelitian yang sedang dilakukan.



Gambar 1. Tahap Penelitian[5]

#### 3.2. Tahap Pengumpulan Data

Tahap dari pengumpulan data pada penelitian ini menggunakan metode wawancara kepada owner Cafe 47°Coffee dan metode studi dokumentasi dari bulan Juli 2021 sampai Februari 2022 yang sudah ada di Cafe 47°Coffee. Data yang digunakan pada penelitian ini adalah Data penjualan dari bulan Juli 2021 sampai dengan Februari 2022. Data tersebut akan digunakan dalam proses clustering menggunakan Algoritma K-Means. Data tersebut akan diolah menjadi 3 cluster yaitu: Cluster Sangat laku, cluster laku, dan cluster tidak laku.

#### 3.3. Tahap Pengolahan Data

Data yang telah didapatkan dari Cafe 47°Coffee masih berbentuk manual atau data masih berbentuk catatan dalam buku. Data tersebut kemudian akan dipindahkan terlebih

dahulu ke .xlsx atau Microsoft excel sebelum dilakukannya pengolahan data mining menggunakan algoritma K-Means. Data yang telah diperoleh akan di cleaning terlebih dahulu menggunakan tools Rapid Miner, data awal sebelum proses cleaning memiliki atribut seperti menu, harga, dan hasil penjualan dari bulan Juli sampai Februari. Dengan proses cleaning data yang akan di ambil hanya atribut menu dan hasil penjualan sehingga akan di lanjutkan ke proses perhitungan menggunakan perhitungan manual menggunakan Euclidean Distance dan akan dilakukan pengujian menggunakan tools Rapid Miner.

### 3.4. Tahap Clustering

Tahap Clustering adalah klasifikasi tidak adanya suatu pengawasan dan suatu tahap partisi sekumpulan entitas data yang terdiri dari suatu set akan menjadi berbagai kelas. Hal tersebut juga bisa dilakukan dengan penerapan tahap-tahap mengenai jarak algoritma dan berbagai persamaan dengan menggunakan perhitungan Euclidean Distance. Berikut cara kerja algoritma K-Means:

1. Menentukan jumlah cluster yang ingin di pakai.
2. Menentukan centroid sebagai seed points (Prototipe cluster awal). Diperoleh secara acak yang dipilih dari objek data pertama.
3. Setiap komponen data, hitung dan kelompokkan berdasarkan jarak terpendek.
4. Menentukan centroid baru dengan menghitung rata-rata dari seluruh anggota yang telah di kelompokkan.
5. Jika data tiap cluster tidak berubah maka perhitungan berhenti tetapi jika masih berubah akan dihitung Kembali ke Langkah no 2.



Gambar 2. Flowchart K-Means [6]

#### 4. HASIL DAN PEMBAHASAN

##### 4.1. Dataset

Dataset yang ingin diolah perlu dilakukan data cleaning yaitu menghapus subset data yang tidak dipakai atau tidak menjadi perhatian pada penelitian ini, seperti subset harga, maka dataset akan tersisa no data, menu dan hasil penjualan. Berikut data yang diperoleh sebelum dan sesudah di cleaning:

**Tabel 1.** Data Penjualan Belum di Cleaning

NO	Menu	Harga	Penjualan/BULAN								
			Juni	Juli	Agustus	September	Oktober	November	Desember	Januari	Februari
1	Americano	Rp 15.000	2	12	13	11	8	10	24	22	14
2	Espresso	Rp 8.000	3	6	1	4	15	21	18	8	10
3	Cappucino	Rp 15.000	11	37	25	6	6	12	9	17	15
4	Moccacino	Rp 18.000	3	5	3	4	12	13	22	11	22
5	Kopi Tubruk	Rp 10.000	0	4	1	5	7	15	19	18	16
6	Vietnam Drip	Rp 15.000	5	10	5	3	15	11	28	19	19
7	V60	Rp 18.000	17	59	34	16	14	30	12	11	25
8	Caramel Coffee Latte	Rp 18.000	15	42	30	23	17	39	19	29	16
9	Kop Sندان	Rp 18.000	6	5	4	7	25	14	20	28	18
10	Avocado Coffe Latte	Rp 15.000	8	11	4	8	19	19	22	22	24
11	Vanilla Coffee Latte	Rp 18.000	2	34	12	11	5	11	17	19	19
12	Hazelnut	Rp 15.000	18	11	13	16	19	19	10	17	16
13	Taro	Rp 15.000	9	17	25	19	16	9	28	11	22
14	Red Velvet	Rp 15.000	23	59	28	37	10	15	24	18	18
15	Matcha	Rp 15.000	6	12	2	18	16	20	22	15	29
16	Strawberry Tea	Rp 15.000	0	5	5	12	11	14	17	27	17
17	Lychee Tea	Rp 13.000	8	34	23	14	27	27	19	14	27
18	Coklat	Rp 15.000	21	61	55	31	44	36	37	39	31
19	Milshake Strawberry	Rp 15.000	12	8	4	7	13	25	12	19	12
20	Banana	Rp 15.000	2	11	9	4	11	12	23	10	16
21	Coffe Lemon	Rp 15.000	0	0	0	6	17	15	15	24	24
22	Kopi susu	Rp 18.000	3	15	1	13	15	26	16	11	16
23	Japanese	Rp 15.000	9	21	22	8	21	18	11	26	18
24	Greentea	Rp 18.000	7	22	14	15	27	11	19	31	36
25	Risol	Rp 3.000	0	0	28	12	17	28	22	29	22

Data tersebut akan di cleaning menggunakan tools RapidMiner, berikut data yang sudah di cleaning seperti tabel 2 di bawah ini:

**Tabel 2.** Data Penjualan Sudah Di Cleaning

NO	Menu	Penjualan/BULAN								
		Juni	Juli	Agustus	September	Oktober	November	Desember	Januari	Februari
1	Americano	2	12	13	11	8	10	24	22	14
2	Espresso	3	6	1	4	15	21	18	8	10
3	Cappucino	11	37	25	6	6	12	9	17	15
4	Moccacino	3	5	3	4	12	13	22	11	22
5	Kopi Tubruk	0	4	1	5	7	15	19	18	16
6	Vietnam Drip	5	10	5	3	15	11	28	19	19
7	V60	17	59	34	16	14	30	12	11	25
8	Caramel Coffee Latte	15	42	30	23	17	39	19	29	16
9	Kop Sندان	6	5	4	7	25	14	20	28	18
10	Avocado Coffe Latte	8	11	4	8	19	19	22	22	24
11	Vanilla Coffee Latte	2	34	12	11	5	11	17	19	19
12	Hazelnut	18	11	13	16	19	19	10	17	16
13	Taro	9	17	25	19	16	9	28	11	22
14	Red Velvet	23	59	28	37	10	15	24	18	18
15	Matcha	6	12	2	18	16	20	22	15	29
16	Strawberry Tea	0	5	5	12	11	14	17	27	17
17	Lychee Tea	8	34	23	14	27	27	19	14	27
18	Coklat	21	61	55	31	44	36	37	39	31
19	Milshake Strawberry	12	8	4	7	13	25	12	19	12
20	Banana	2	11	9	4	11	12	23	10	16
21	Coffe Lemon	0	0	0	6	17	15	15	24	24
22	Kopi susu	3	15	1	13	15	26	16	11	16
23	Japanese	9	21	22	8	21	18	11	26	18
24	Greentea	7	22	14	15	27	11	19	31	36
25	Risol	0	0	28	12	17	28	22	29	22

Data pada tabel 2 akan masuk kedalam tahap clustering menggunakan perhitungan Euclidean Distance secara manual lalu akan dilakukan pengujian menggunakan tools Rapid Miner[7].

#### 4.2. Pengelompokan Data Manual

##### 1. Menentukan Centroid Awal

Langkah awal untuk pengelompokan data menggunakan Algoritma K-Means adalah menentukan jumlah cluster, pada penelitian ini data akan dikelompokkan menjadi 3 cluster. setelah menentukan jumlah cluster langkah selanjutnya yaitu menentukan centroid awal secara acak dari data yang ada. Pada penelitian ini centroid yang diambil adalah data ke-6 sebagai cluster 1, data ke-12 sebagai cluster 2, dan data ke-16 sebagai cluster 3, seperti tabel 3 dibawah ini:

**Tabel 3.** Centroid Awal Iterasi 1

8	Caramel Coffee Latte	15	42	30	23	17	39	19	29	16
12	Hazelnut	18	11	13	16	19	19	10	17	16
16	Strawberry Tea	0	5	5	12	11	14	17	27	17

##### 2. Menghitung jarak data menggunakan perhitungan Euclidean Distance, berikut rumus dari Euclidean Distance:

$$D(x, y) = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2} \quad [8]$$

Maka didapatkan nilai perhitungan jarak pertama dengan data pertama pada centroid awal, sebagai berikut:

Iterasi 1

a Jarak data ke-1 dan centroid ke-1

$$\begin{aligned}
 &= \sqrt{(2 - 15)^2 + (12 - 42)^2 + (13 - 30)^2 + (11 - 23)^2 + (8 - 17)^2} \\
 &\quad + (10 - 39)^2 + (24 - 19)^2 + (22 - 29)^2 + (14 - 16)^2 \\
 &= 169 + 900 + 289 + 144 + 81 + 841 + 25 + 49 + 4 \\
 &= \sqrt{2.502} = 50,019
 \end{aligned}$$

b Jarak data ke-1 dan centroid ke-2

$$\begin{aligned}
 &= \sqrt{(2 - 18)^2 + (12 - 11)^2 + (13 - 13)^2 + (11 - 16)^2 + (8 - 19)^2} \\
 &\quad + (10 - 19)^2 + (24 - 10)^2 + (22 - 17)^2 + (14 - 16)^2 \\
 &= 256 + 1 + 0 + 25 + 121 + 81 + 196 + 25 + 4 \\
 &= \sqrt{709} = 26,627
 \end{aligned}$$

c Jarak data ke-1 dan centroid ke-3

$$\begin{aligned}
 &= \sqrt{(2 - 0)^2 + (12 - 5)^2 + (13 - 5)^2 + (11 - 12)^2 + (8 - 11)^2} \\
 &\quad + (10 - 14)^2 + (24 - 17)^2 + (22 - 27)^2 + (14 - 17)^2 \\
 &= 4 + 49 + 64 + 1 + 9 + 16 + 49 + 25 + 9 \\
 &= \sqrt{226} = 15,033
 \end{aligned}$$

Data terus akan dihitung sampai data ke-25, Berikut hasil perhitungan jarak data ke-1 sampai dengan data ke-25 untuk iterasi ke-1:

**Tabel 4.** Hasil Perhitungan Jarak Iterasi 1

Menu	C1	C2	C3	JP
Americano	50.019	26.627	15.033	15.033
Espresso	58.719	27.184	23.790	23.790
Cappucino	38.078	34.481	42.166	34.481
Moccacino	60.605	28.390	19.621	19.621
Kopi Tubruk	60.257	29.664	13	13
Vietnam Drip	55.919	28.635	18.574	18.574
V60	30.033	54.890	68.352	30.033
Caramel Coffee Latte	0	44.011	54.863	0
Kop Sندان	55.470	25.059	16.401	16.401
Avocado Coffe Latte	49.365	21.863	17.464	17.464
Vanilla Coffee Latte	41.737	33.660	31.749	31.749
Hazelnut	44.011	0	26.057	0
Taro	45.210	27.838	33.555	27.838
Red Velvet	36.441	57.384	68.789	36.441
Matcha	50.259	24.433	22.449	22.449
Strawberry Tea	54.863	26.057	0	0
Lychee Tea	28.861	32.741	43.943	28.861
Coklat	49.729	82.764	93.546	49.729
Milshake Strawberry	49.739	17.406	20.445	17.406
Banana	55.443	27.331	21.213	21.213
Coffe Lemon	62.120	29.529	13.601	13.601
Kopi susu	48.166	23.043	23.323	23.043
Japanese	36.055	20.420	28.213	20.420
Greentea	45.519	32.388	32.771	32.388
Risol	47.749	33.090	28.982	28.982

### 3. Pengelompokan Data

Hasil dari perhitungan jarak dalam tabel 4 akan dilakukan pengelompokan, jarak yang akan dipilih adalah jarak terdekat antara data bersama pusat cluster. Pengelompokan data dapat dilihat tabel 5 dibawah ini.

**Tabel 5.** Pengelompokan Data Iterasi 1

Menu	C1	C2	C3
Americano			√
Espresso			√
Cappucino		√	
Moccacino			√
Kopi Tubruk			√
Vietnam Drip			√
V60	√		
Caramel Coffee Latte	√		
Kop Sندان			√
Avocado Coffe Latte			√
Vanilla Coffee Latte			√
Hazelnut		√	
Taro		√	
Red Velvet	√		
Matcha			√
Strawberry Tea			√
Lychee Tea	√		
Coklat	√		
Milshake Strawberry		√	
Banana			√
Coffe Lemon			√
Kopi susu		√	
Japanese		√	
Greentea		√	
Risol			√



Berdasarkan pengelompokan yang didapat pada tabel 5 terdapat 5 data untuk cluster 1, 7 data untuk cluster 2, dan 13 data untuk cluster 3.

#### 4. Menentukan Centroid Baru

Setelah didapatkan data dari setiap cluster yang telah di kelompokkan kemudian proses kembali ke langkah 2 yaitu menentukan centroid baru. Menentukan centroid baru dihitung rata-rata dari iterasi pertama dari tiap cluster. Berikut contoh perhitungan untuk menentukan centroid baru:

$$= \frac{\text{Jumlah data}}{\text{banyak data cluster}}$$

$$= \frac{17+15+23+8+21}{5} = 16,8$$

**Tabel 6. Centroid Baru Iterasi ke-2**

C1	16,8	51	34	24.2	22.4	29.4	22.2	22.2	23.4
C2	9.857	18.714	14.857	12	16.714	17.142	15	18.857	19.285
C3	2.846	8.846	6.692	8.076	13.692	15.615	20.692	19.384	19.230

**Tabel 7. Centroid Baru Iterasi ke-3**

C1	19	55.25	36.75	26.75	21.25	30	23	24.25	22.5
C2	9.142	25.142	19.142	12.714	17.285	15.285	16.142	19.285	21.857
C3	3.571	7.428	5.714	8.142	14,357	17.357	20	18.785	19.5

Menentukan centroid baru dilakukannya pengulangan sampai nilai rata-rata tidak berubah. Pada penelitian ini perhitungan pengulangan untuk menentukan centroid baru berhenti pada iterasi ke-3. Centroid yang tidak mengalami perubahan dapat dilihat diatas pada tabel 7.

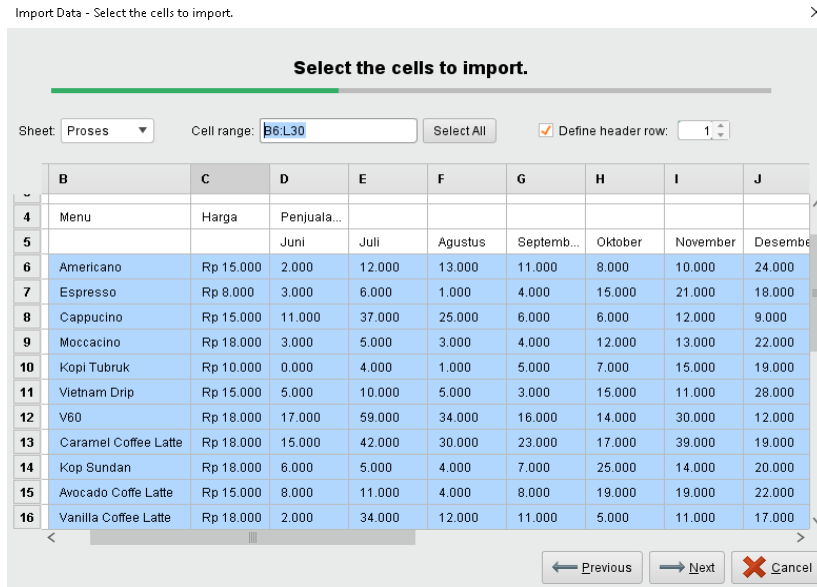
Dibawah ini merupakan hasil dari iterasi terakhir atau iterasi ke-3:

**Tabel 8. Hasil Perhitungan Jarak Iterasi 3**

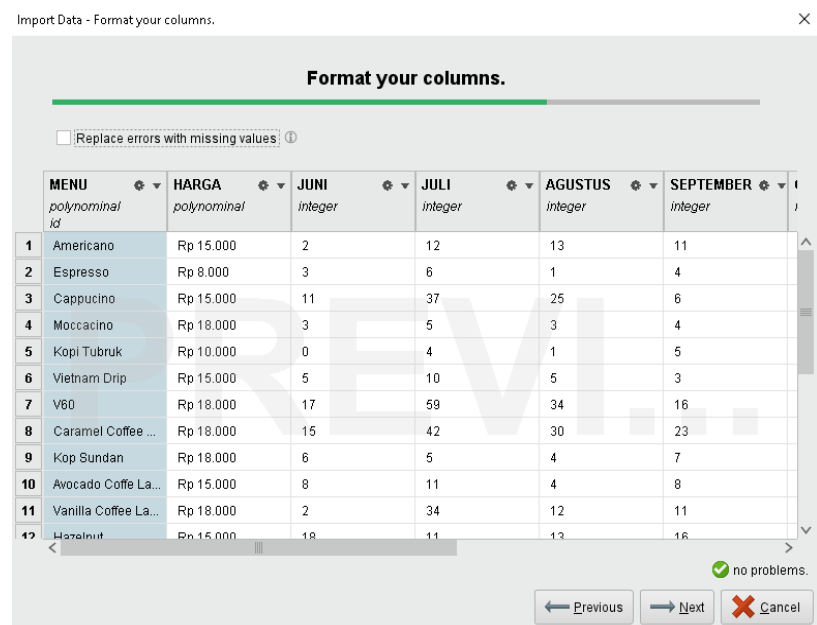
Menu	C1	C2	C3	JT	k
Americano	60.212	22.570	15.024	15.024	3
Espresso	71.011	33.445	15.748	15.748	3
Cappucino	42.744	21.560	39.257	21.560	2
Moccacino	70.633	30.267	11.496	11.496	3
Kopi Tubruk	72.553	32.699	11.142	11.142	3
Vietnam Drip	65.426	26.707	11.863	11.863	3
V60	22.283	41.982	63.471	22.283	1
Caramel Coffee Latte	20.762	35.233	52.185	20.762	1
Kop Sندان	66.858	29.059	15.024	15.024	3
Avocado Coffe Latte	60.312	22.712	10.206	10.206	3
Vanilla Coffee Latte	47.979	19.018	29.890	19.018	2
Hazelnut	55.032	20.518	21.666	20.518	2
Taro	49.300	19.762	28.640	19.762	2
Red Velvet	24.971	46.238	66.391	24.971	1
Matcha	59.833	25.139	16.613	16.613	3
Strawberry Tea	67.428	28.545	11.632	11.632	3
Lychee Tea	33.280	19.759	37.519	19.759	2
Coklat	37.783	71.389	90.583	37.783	1
Milshake Strawberry	64.000	28.127	15.561	15.561	3
Banana	64.877	25.450	13.251	13.251	3
Coffe Lemon	74.189	33.993	14.147	14.147	3
Kopi susu	60.539	26.322	16.169	16.169	3
Japanese	46.406	12.455	25.626	12.455	2
Greentea	50.428	22.456	31.566	22.456	2
Risol	61.266	32.941	28.640	28.640	3

### 4.3. Pengujian menggunakan Software Rapid Miner

Dataset yang akan dianalisis yaitu data penjualan dengan format .xlsx (excel) software yang akan digunakan yaitu RapidMiner Studio. Pada proses open file yang dilakukan di operator Read Excel melalui parameter Import Data untuk memilih file lalu klik next setelah itu pilih sheet yang berisikan sumber data lalu klik next setelah memilih sheet selanjutnya pemberian anotasi, tahapan dari read excel tersebut.

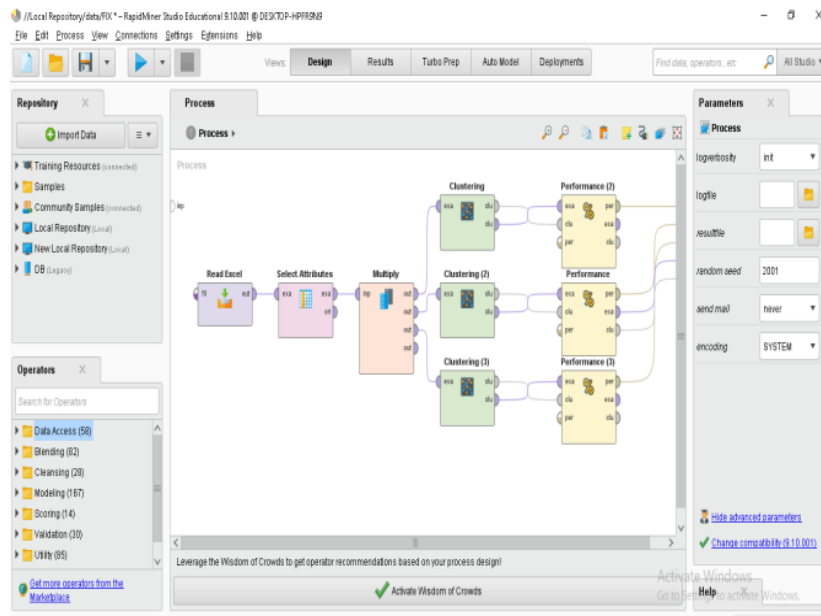


Gambar 3. Tampilan pemilihan sheet



Gambar 4. Tampilan Pemberian Anotasi

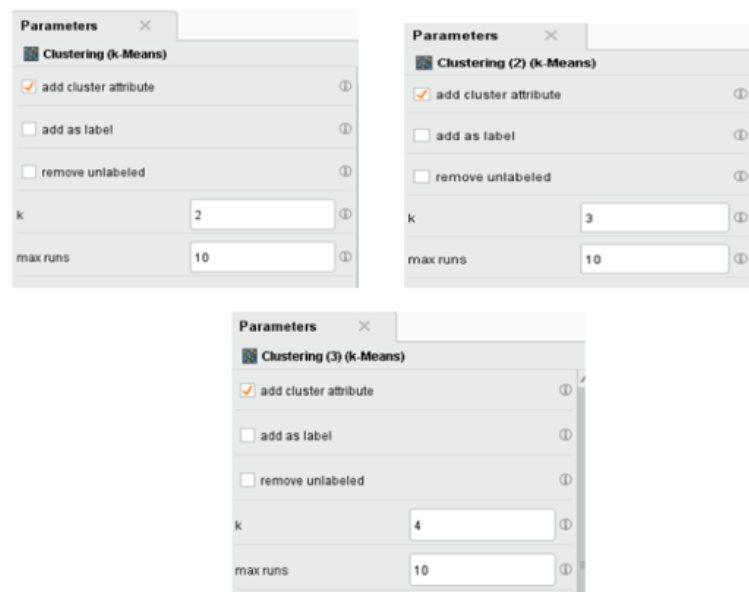
Selanjutnya yaitu drag dan drop parameter yang ingin diproses seperti gambar 5 dibawah parameter lengkap sebagai berikut:



**Gambar 5.** Tampilan Parameter lengkap

Proses yang dilakukan pada tahapan ini, yaitu:

1. Read Excel.  
Tahapan ini dilakukan untuk pengimputan dataset berupa file format .xlsx data penjualan.
2. Select Attributes  
Tahapan ini yaitu pemilihan atribut yang ingin dilakukan proses clustering, pada penelitian ini terdapat atribut Menu, Harga, dan Hasil Penjualan Juni – Februari lalu dilakukan pemilihan atribut menggunakan operator Select Attributes sehingga atribut menjadi Menu dan hasil penjualan bulan Juni - Februari.
3. Multiplay  
mengambil dataset yang telah diinput pada read excel lalu di letakkan pada operator clustering 1, clustering 2, dan clustering 3.
4. Clustering  
Tahapan ini yaitu proses pengclustoran data penjualan dan sebagai algoritma pada penelitian ini, Operator clustering pada penelitian ini terdapat 3 operator yang dimana operator clustering tersebut akan dilakukan perbandingan berapa cluster yang performancenya optimal apakah dengan 2 cluster, 3 cluster, atau 4 kluster. Berikut dibawah ini parameter dengan masing-masing cluster:



**Gambar 6.** Parameters Clustering

#### 5. Performance

Tahapan ini dilakukan operasi pencarian davies bouldin index. Pada operator performance cluster yang optimal ialah memakai 3 cluster , sehingga pada penelitian ini memakai 3 clustering.

Setelah 5 tahapan diatas telah di running makan muncul output hasil pengujian memakai tolls Rapidminer yaitu seperti dibawah ini :

##### 1. ExampleSet

Pada hasil ini menampilkan hasil dari cluster yaitu Data View. Data View adalah hasil data keseluruhan yang telah diinputkan. Setelah itu hasil dari tampilan Scatter, Scatter adalah tampilan berbentuk grafik titik yang terdapat 3 warna untuk tiap-tiap cluster. warna biru yaitu cluster 0 (Tidak Laku), warna hijau yaitu cluster 2 (Laku), dan warna orange yaitu cluster 1 (Sangat Laku).



**Gambar 7.** Tampilan Scatter

## 2. Cluster Model

Pada hasil Cluster Model menampilkan Description Cluster model dan Centroid Table yaitu sebagai berikut[10]:

### Cluster Model

```
Cluster 0: 14 items
Cluster 1: 4 items
Cluster 2: 7 items
Total number of items: 25
```

**Gambar 8.** Text View Cluster Model

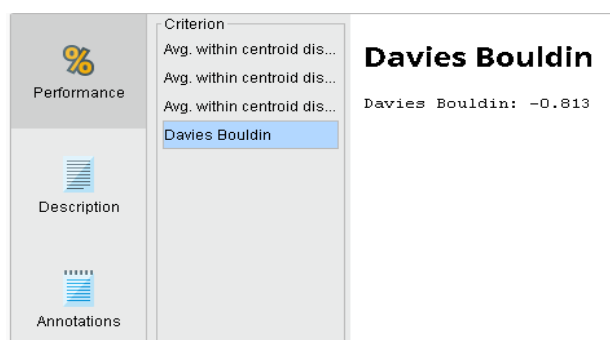
Attribute	cluster_0	cluster_1	cluster_2
JUNI	3.571	19	9.143
JULI	7.429	55.250	25.143
AGUSTUS	5.714	36.750	19.143
SEPTEMBER	8.143	26.750	12.714
OKTOBER	14.357	21.250	17.286
NOVEMBER	17.357	30	15.286
DESEMBER	20	23	16.143
JANUARI	18.786	24.250	19.286
FEBRUARI	18.500	22.500	21.857

**Gambar9.** Centroid Table

Centroid Table adalah nilai titik pusat pada setiap cluster. nilai-nilai tersebut menjadi acuan untuk perhitungan tiap dataset dengan cara mengukur dari masing-masing titik pusat cluster.

## 3. Davies Bouldin Index

Davies Bouldin Index yaitu untuk mengevaluasi kinerja pada proses pengklusteran, Evaluasi ini menghasilkan dalam parameter berupa nilai, Semakin kecil nilai maka semakin baik cluster yang akan diperoleh, berikut Hasil Davies Bouldin Index dari penelitian ini:



**Gambar 10.** Davies Bouldin Index

Setelah dilakukannya pengelompokan dan pengujian secara manual dan pengujian menggunakan tools Rapidminer maka dapat disimpulkan bahwa hasil dari 3 cluster, yaitu : cluster 0 (Tidak Laku) yaituAmericano, Espresso, Moccacino, Kopi tubruk, Vietnam Drip, Kopi sundan, Avocado Coffe Latte, Match, Strawberry Tea, Milkshake Strawberry, Banana, Cofee Lemon, Kopi susu, dan risol, cluster 1 (Sangat Laku) yaitu V60, Caramel Coffe Latte, Red Velvet, dan Coklat , 7 item cluster laku yakni Cappucino, Vanilla Coffe Latte, Hazelnut, Taro, Lychee Tea, Japanese, dan Greentea, dan cluster 2 (Laku) yaitu Cappucino, Vanilla Coffe Latte, Hazelnut, Taro, Lychee Tea, Japanese, dan Greentea. Dan berdasarkan hasil analisa untuk persentase dari tiap-tiap cluster adalah:

**Table 9.** Hasil Persentase Clustering secara manual dan menggunakan Rapidminer

Cluster 0	14	52%	Tidak Laku
Cluster 1	4	16%	Sangat Laku
Cluster 2	7	28%	Laku
Total	25	100%	

## 5. KESIMPULAN

Berdasarkan hasil dari penelitian yang telah dilakukan sebelumnya maka dapat peneliti simpulkan 1). Penerapan dari metode clustering dengan algoritma k-means yang dapat digunakan untuk mengelompokkan data penjualan selama 9 bulan dengan 3 cluster dari 25 item yaitu Cluster 0 Tidak laku, cluster 1 Sangat laku, dan cluster 2 laku. Sehingga pihak cafe dapat melakukan strategi penjualan dengan mengurangi item yang terdapat pada cluster 0 dan meningkatkan item pada cluster 1 dan 2. 2). Berdasarkan penelitian clustering dengan perhitungan secara manual dan dengan bantuan tools Rapidminer mendapatkan hasil yang sama, yaitu didapatkan 3 kali iterasi dan menghasilkan cluster 0 (Tidak laku) jumlah 14 item dengan persentase 52%, cluster 1 (Sangat laku) jumlah 4 item dengan persentase 16%, dan cluster 2 (Laku) jumlah 7 item dengan persentase 28%. dan pada proses ini menghasilkan nilai Davies Bouldin Index sebesar 0,813.

## REFERENSI

- [1] A. Bastian, H. Sujadi, and G. Febrianto, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)."
- [2] I. Nuryani and D. Darwis, "Analisis Clustering Pada Pengguna Brand Hp Menggunakan Metode K-Means," *Proceeding Semin. Nas. Ilmu Komput.*, vol. 1, no. 1, pp. 190–211, 2021.
- [3] E. G. Sihombing, "Klasifikasi Data Mining pada Rumah Tangga Menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa Menggunakan K-Means Clustering Method," *Comput. Eng. Syst. Sci. J.*, vol. 2, no. 2, pp. 74–82, 2017.
- [4] U. Ma'rifatin, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien Pada Puskesmas Warujayeng," *Semin. Nas. Inov. Teknol.*, no. 2549–7952, pp. 285–291, 2020.
- [5] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA J. Inform.*, vol. 7, no. 2, p. 109, 2019, doi: 10.30595/juita.v7i2.5529.
- [6] D. Triyansyah and D. Fitriah, "Analisis Data Mining Menggunakan Algoritma K-

- Means Clustering Untuk Menentukan Strategi Marketing,” *J. Telekomun. dan Komput.*, vol. 8, no. 3, pp. 163–182, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [7] F. Hadi and Y. Diana, “Pengkusteran Penjualan Bahan Bangunan Menggunakan Algoritma K-Means,” *JOISIE (Journal Inf. Syst. Informatics Eng.*, vol. 4, no. 1, pp. 22–28, 2020, doi: 10.35145/joisie.v4i1.629.
- [8] M. H. Siregar, “Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan),” *J. Teknol. Dan Open Source*, vol. 1, no. 2, pp. 83–91, 2018, doi: 10.36378/jtos.v1i2.24.
- [9] S. Butsianto and N. T. Mayangwulan, “Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 3, pp. 187–201, 2020, doi: 10.32672/jnkti.v3i3.2428.
- [10] M. Mardalius, “Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means,” *Jurteks*, vol. 4, no. 2, pp. 123–132, 2018, doi: 10.33330/jurteks.v4i2.36.