

## IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA PENJUALAN MAKANAN (STUDI KASUS : AYAM BETUTU WARUNG WARDANA)

Audi Mayori<sup>1</sup>, Yuyun Tresnawati, M.Kom<sup>2</sup>

<sup>1,2</sup> Sistem Informasi, STMIK “AMIKBANDUNG”, Jl. Jakarta No.28, Bandung, Indonesia, 40272  
Email: audimayori25@gmail.com

**\*Koresponden Author:** Audi Mayori, audimayori25@gmail.com  
Accepted: 10 14, 2023 ; Revised: 11 02, 2023; Published: 12 30, 2023

### Abstrak

*Ayam Betutu Warung Wardana adalah salah satu usaha yang terjun dalam bisnis makanan, non-kopi dan makanan ringan yang banyak diminati di semua kalangan untuk sekedar bersantai atau untuk mengerjakan pekerjaan. Untuk memaksimalkan usaha kafe ini pengelola harus mengetahui kebutuhan dari pelanggan – pelanggan yang pernah berkunjung, sehingga dapat meningkatkan pelayanan yang sesuai dengan apa yang dibutuhkan. Disamping itu ada permasalahan yang kerap terjadi yaitu masalah dalam mengenai menu yang paling berpotensi untuk ditingkatkan serta menu yang paling banyak diminati konsumen. Dari permasalahan yang sudah dijelaskan tersebut, penelitian ini menerapkan metode klasterisasi algoritma K-means untuk mengetahui minat konsumen pada suatu menu makanan dan minuman di Ayam Betutu Warung Wardana. Dalam penelitian ini penulis menggunakan kerangka kerja berdasarkan metode Cross Industry Standard Process for Data Mining (CRISP-DM) Penelitian ini menghasilkan 2 cluster yang terbaik dengan metode validasi Silhouette dan DBI. Pengujian menggunakan Silhouette Coefisient menunjukkan nilai 0.44 dan itu merupakan nilai terbaik yang mendekati angka Davies Bouldin Index (DBI) menunjukkan nilai dengan ratio terkecil pada angka 0.9030707. Hal tersebut baik karena ratio mendekati nilai nol, maka semakin baik yang diperoleh dari pengelompokan K-Means.*

**Kata kunci:** Clustering, K-Means, Silhouette Coefisient, Davin Bouldin Index, CRISP- DM

### Abstract (12pt Bold)

**[IMPLEMENTATION OF THE K-MEANS CLUSTERING ALGORITHM ON FOOD SALES (CASE STUDY: AYAM BETUTU WARUNG WARDANA)]** Ayam Betutu Warung Wardana is one of the small food and beverage businesses which is in great demand to relax or to do work. To maximize this business, managers must know the needs of customers who have visited, therefore they can improve services according to what is needed. Besides that, there are problems that often occur, namely problems regarding the menu that has the most potential to be improved and the menu that is most in demand by consumers. From the problems that have been described above, this study applies the K-means algorithm clustering method to determine consumer interest in a food and beverage menu at Ayam Betutu Warung Wardana. In this study, the authors used a framework based on the Cross Industry Standard Process for Data Mining (CRISP-DM) method. This study produced the best two clusters with the Silhouette and DBI validation methods. Testing using the Silhouette Coefficient shows a value of 0.44 and it is the best value close to the Davies Bouldin Index (DBI) showing the value with the smallest ratio at 0.9030707. This is good because the ratio is close to zero, the better the K-Means grouping

**Keywords:** Clustering, K-Means, Silhouette Coefisient, Davin Bouldin Index, CRISP- DM

## 1. PENDAHULUAN

*Data Mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data. *Data Mining* juga disebut serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak bisa diketahui secara manual dari suatu data yang sudah dikumpulkan. Pada penelitian ini analisa yang digunakan dengan menggunakan metode *K-Means*. Dengan menggunakan metode ini, data-data yang sudah didapatkan dapat dikelompokkan ke dalam beberapa klaster.

Ayam Betutu Warung Wardana adalah salah satu usaha yang terjun dalam bisnis makanan, *non-kopi* dan makanan ringan yang banyak diminati di semua kalangan untuk sekedar bersantai atau untuk mengerjakan pekerjaan. Untuk memaksimalkan usaha kafe ini pengelola harus mengetahui kebutuhan dari pelanggan – pelanggan yang pernah berkunjung, sehingga dapat meningkatkan pelayanan yang sesuai dengan apa yang dibutuhkan. Di samping itu ada permasalahan yang kerap terjadi yaitu masalah mengenai menu yang paling berpotensi untuk ditingkatkan serta menu yang paling banyak diminati konsumen.

Terdapat beberapa metode yang dapat digunakan untuk melakukan pengelompokkan, seperti salah satunya yaitu *clustering*. Metode *Clustering* dapat digunakan untuk mengelompokkan suatu data yang memiliki kesamaan antara satu data dengan data yang lainnya). Untuk menunjang keberhasilan proses analisis, maka diperlukan juga penerapan suatu algoritma. Adapun algoritma yang digunakan pada penelitian ini yaitu algoritma *k-means* [1]. Dari permasalahan yang sudah dijelaskan tersebut, pada penelitian ini akan dilakukan teknik pengklaster dengan menggunakan algoritma *K-Means clustering*. Metode ini merupakan cara yang dapat digunakan untuk menganalisa objek berdasarkan persamaan sifat antar objek tersebut.

Mengacu pada latar belakang di atas dapat disimpulkan bahwa pada penelitian Penerapan Kerangka Kerja *Cross Industry Standrad Process Data Mining (Crisp-Dm)* Pada Penjualan Produk Makanan (*Study Kasus Di Ayam Betutu Warung Wardana*) menerapkan metode klasterisasi algoritma *k-means* untuk mengetahui minat konsumen pada suatu menu makanan dan minuman di Ayam Betutu Warung Wardana.

## 2. METODE

### A. *Data Mining*

*Data mining* adalah proses yang memanfaatkan teknik- teknik statistik, matematika, dan kecerdasan buatan untuk mengekstrak dan mengidentifikasi informasi dan *knowledge* selanjutnya (atau pola-pola) yang berasal dari sekumpulan data yang sangat besar. Berbagai macam pola tersebut bisa dalam bentuk aturan bisnis, kesamaan-kesamaan, korelasi, *trend*, atau model-model prediksi.

### B. *Clustering*

*Clustering* atau klasterisasi adalah metode pengelompokan data. Menurut Tan, 2006 *clustering* adalah sebuah proses untuk mengelompokkan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum. *Clustering* merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster*. Objek yang di dalam cluster memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan *cluster* yang lain.

### C. Algoritma *K-Means*

Metode *K-means* merupakan metode *clustering* yang paling sederhana dan umum. Hal ini dikarenakan *K-means* mempunyai kemampuan dalam mengelompokkan data dengan jumlah yang cukup besar dengan waktu komputasi yang cepat dan efisien.

Algoritma ini pertama kali diusulkan oleh MacQueen (1967) dan dikembangkan oleh Hartigan dan Wong tahun 1975 dengan tujuan untuk dapat membagi M data *point* dalam N dimensi kedalam sejumlah k *cluster* di mana proses klastering dilakukan dengan meminimalkan jarak *sum squares* antara data dengan masing masing pusat *cluster* (*centroid-based*) [1].

#### D. Tableau

Tableau merupakan perangkat lunak untuk menganalisis dan menggambarkan kumpulan data dalam bentuk wawasan dalam pengambilan keputusan serta membantu dalam visualisasi data yang biasanya digunakan untuk membantu industri data *analys*, data *science*, serta *business intelligence*. Pada penelitian ini tableau digunakan untuk membantu beberapa visualisasi dari data. Dimana produk perangkat lunak tableau sendiri termasuk tableau Dekstop, Tableau Prep, Tableau *Online*, dan Tableau Publik.

#### E. Bahasa Pemrograman Rstudio

Bahasa pemrograman Rstudio ini merupakan bahasa pemrograman yang digunakan untuk analisis data dan Data *Science*. Bahasa pemrograman ini membantu Peneliti untuk mengolah data.

#### F. Devise Bouldin Index (DBI)

Metode *Davies Bouldin Index* (DBI) merupakan metode untuk mengecek hasil *clustering*. Pendekatan pengujian nilai DBI berupa nilai separasi dan kohesi. Kohesi berupa jumlah dari kemiripan data terhadap pusat *cluster* dari *cluster* tersebut. Separasi adalah jarak antara pusat *cluster* dari *cluster*. Cluster yang optimal adalah *cluster* yang memiliki nilai separasi yang tinggi dan nilai kohesi yang rendah. Nilai *Davies Bouldin Index* (DBI) yang semakin mendekati nilai 0, menandakan semakin baik *cluster* yang diperoleh, semakin rendah nilai DBI menunjukkan hasil *cluster* yang optimal [3]. Untuk menghitung nilai *Davies*

*Bouldin Index* (DBI) sesuai dengan persamaan (2.1):

$$DBI = \frac{1}{k} \times \sum_{a=1}^k R_a \quad (2.1)$$

Keterangan :

DBI = Indeks *Davies Bouldin*

k = Jumlah klaster

Rab = Ukuran kemiripan antara klaster ke-a dan klaster ke-b

#### G. Silhoutte Coefisient

Metode ini merupakan evaluasi *cluster* yang menggabungkan metode *cohesion* dan *separtion*. *Cohession* diukur dengan menghitung seluruh objek yang terdapat dalam sebuah *cluster* dan *separation* diukur dengan

$$sil(c) = sil(k) \frac{1}{|k|} \sum_{i=1}^k sil(c_i) \quad (2.2)$$

menghitung jarak rata-rata setiap objek dalam sebuah *cluster* dengan *cluster* terdekatnya [4].

Keterangan :

Sil (k) = Nilai *silhoutte* semua *cluster*

|k| = Banyaknya *cluster* k

Sil (ci) = Rata-rata nilai *silhouette*

#### H. Min – Max Normalization

*Min-max Normalization* merupakan metode normalisasi dengan strategi linier yang mentransformasikan data dari satu rentang nilai yang baru, sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses. Data akan diubah menjadi 0 sampai 1. Metode *Min-max Normalization* merupakan salah satu metode mengubah data yang *komplex* dengan tidak menghilangkan isi, sehingga lebih mudah diolah [5].

$$X^I = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2.3)$$

Keterangan :

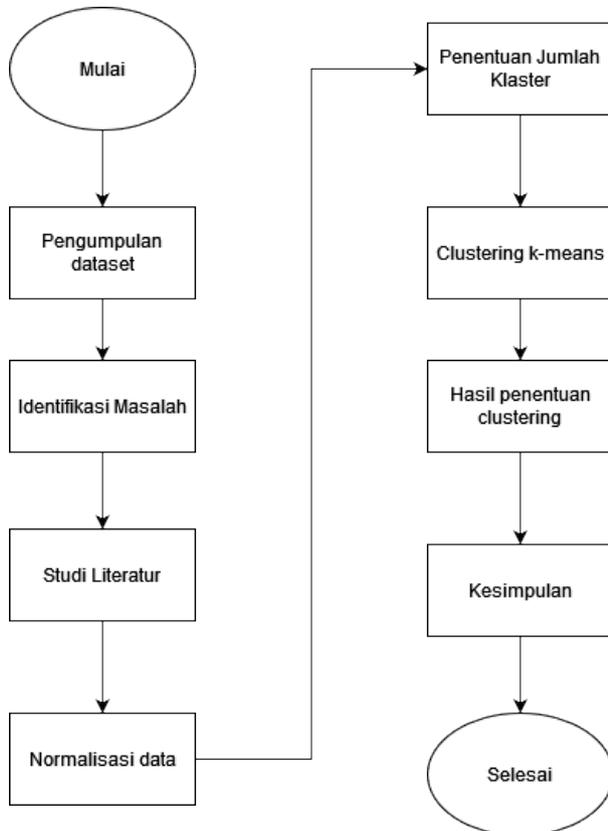
X1 = Data atribut yang akan diolah

Xmin = Nilai terkecil atribut tersebut

Xmax = Nilai tertinggi atribut tersebut

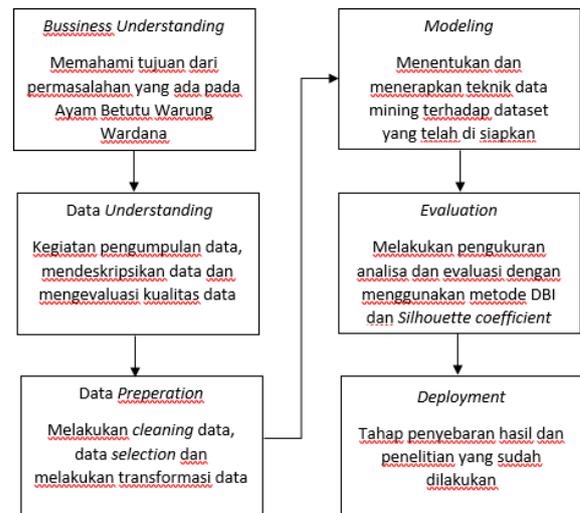
Metode normalisasi ini akan melakukan transformasi linier terhadap data asli

sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses. Berikut merupakan proses tahapan klusterisasi pada penelitian :



Gambar 2.1 Alur Penelitian

Dalam penelitian ini Penulis menggunakan kerangka penelitian *Cross Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM adalah suatu metode proses data *mining* yang berfungsi untuk melakukan tahapan analisis dari suatu dasar strategi pada suatu penelitian. Berikut merupakan tahapan proses data analisa menggunakan (CRISP-DM) :



Gambar 2.2 Tahapan CRISP-DM

Berikut penjelasan dari setiap tahapan pada gambar diatas :

1) *Bussiness Understanding*

Pada awal tahapan ini akan berfokus mengenai pemahaman tujuan yang dimana dari pemahaman tersebut akan diubah menjadi sebuah rencana awal data mining yang akan dirancang demi mencapai tujuan dari permasalahan yang ada pada Ayam Betutu Warung Wardana yakni ingin mengetahui menu manakah yang paling berpotensi untuk ditingkatkan serta menu manakah yang paling banyak diminati konsumen.

2) *Data Understanding*

Pada tahap selanjutnya akan dilakukan kegiatan mengumpulkan data, mendeskripsikan data, serta mengevaluasi data penjualan di Ayam Betutu Warung Wardana kemudian dari data tersebut diidentifikasi data mana yang akan digunakan dalam penelitian. Dikarenakan Ayam Betutu Warung Wardana termasuk UMKM baru yang terjun dunia bisnis ditahun 2023, maka dari itu penulis hanya mengolah dataset dari bulan Februari – Juli 2023, field dari dataset tersebut yaitu no, nama produk, harga dan tanggal jumlah penjualan dari Februari – Juli 2023.

### 3) Data Preparation

Tahapan ini yaitu membangun dataset akhir dari berupa data mentah yang harus dilakukan diantaranya pembersihan data (*Data Cleaning*), melakukan pemilihan data (*Data Selection*), dan juga melakukan transformasi terhadap data (*Data Transformasi*) untuk dijadikan masukan dalam tahap pemodelan. Yang dimana pada penelitian hanya dibutuhkan nama produk dan jumlah penjualan dari dataset yang diberikan oleh owner Ayam Betutu Warung Wardana.

### 4) Modeling

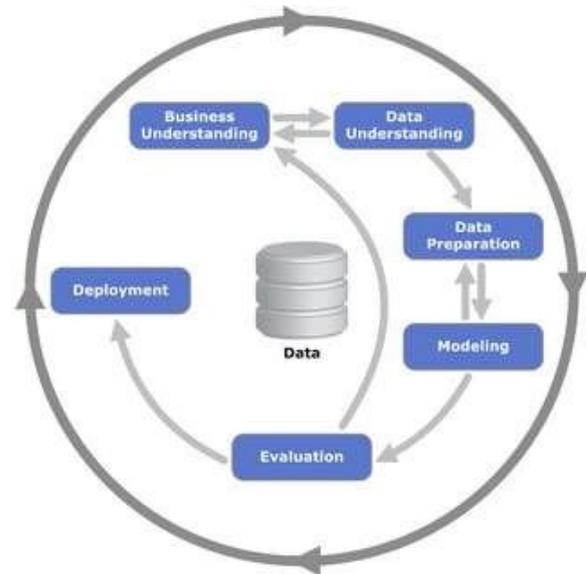
Pada tahap ini melibatkan *Machine Learning*, proses dalam menentukan dan penerapan teknik atau mode *data mining* terhadap dataset yang telah disiapkan untuk pengelompokan menu makanan dan minuman pada Ayam Betutu Warung Wardana. Pada penelitian ini akan menggunakan pemodelan algoritma *K-Means Clustering* karena data tersebut akan terbagi menjadi beberapa kelompok sesuai dengan rata-rata hasil jumlah penjualan produk dan akan diolah menggunakan pemogram python.

### 5) Evaluation

Tahap ini akan dilakukan analisa atau pengukuran terhadap pemodelan yang telah dilakukan. Evaluasi dilakukan dengan menggunakan metode *Davies Bouldin Index (DBI)* dan *Silhouette Coefisient*, yang di mana metode ini merupakan metode pengujian yang berdasarkan klaster yang dihasilkan. Evaluasi ditujukan untuk mengetahui pemodelan yang dilakukan apakah sudah tepat dan sesuai pada kasus penelitian ini. Selanjutnya hasil evaluasi tersebut adalah menentukan langkah berikutnya.

### 6) Deployment

Langkah terakhir yaitu tahap penyebaran hasil dari penelitian yang sudah dilakukan dan dijadikan sebagai laporan atau presentasi dari proses *data mining*.



Gambar 2.3 Alur CRISP-DM

## 3. HASIL DAN PEMBAHASAN

### A. Pengumpulan Dataset

Dataset ini terdiri dari no, menu, data penjualan di setiap bulannya, total penjualan, waktu, umur. Data penjualan ditunjukkan pada gambar di bawah ini:

No	Menu	Februari	Maret	April	Mei	Juni	Juli	Total	Waktu	Umur
1	Nasi ayam betutu	10	30	36	45	52	25	198	siang	dewasa
2	Nasi ayam suwir	4	10	12	20	21	44	111	siang	dewasa
3	Nasi ayam penyet	64	34	45	12	45	23	223	siang	dewasa
4	Nasi sapi lada hitam	3	16	13	16	26	21	95	siang	dewasa
5	Nasi goreng kandang	24	21	18	27	39	34	163	siang	dewasa
6	Nasi goreng sapi lada hitam	43	45	32	43	67	46	276	siang	dewasa
7	Nasi goreng betutu	25	24	15	23	34	21	142	siang	dewasa
8	Mie kandang	29	18	35	41	15	28	166	siang	dewasa
9	Kroketten	67	78	94	75	82	93	489	sore	remaja
10	Risolles	34	55	62	34	22	32	239	sore	remaja
11	Bitterballen	12	21	14	16	20	42	125	sore	remaja
12	Poffertjes	16	23	14	19	8	16	96	sore	remaja
13	waffles original	24	33	12	23	38	29	159	siang	remaja
14	waffles bluberry jam	23	25	34	21	27	46	176	siang	remaja
15	waffles oreo caramel	37	26	32	38	41	24	198	siang	remaja
16	Pisang goreng	25	32	13	23	34	28	155	malam	remaja
17	Lumpia pisang keju	45	38	27	34	31	29	204	malam	remaja
18	Cireng crispy	34	45	11	39	18	43	190	malam	remaja
19	Kentang sosis	32	19	42	30	18	38	179	malam	remaja
20	Roti bakar original	30	34	23	31	45	44	207	malam	remaja
21	Roti bakar keju	44	56	53	32	37	51	273	malam	remaja
22	Roti bakar sarikaya keju	45	34	23	61	32	49	244	malam	remaja
23	Roti bakar peanut butter	23	21	20	12	26	27	129	malam	remaja
24	Roti bakar kacang coklat susu	34	18	42	35	45	64	238	malam	remaja

Gambar 4.1 Pengumpulan Dataset

### B. Data Cleaning

Tahapan ini akan dilakukan *cleaning* atau pembersihan data yang berupa data kosong, agar tidak mengganggu keakuratan hasil pengujian. Berikut merupakan *sourcecode* Pemograman R untuk menghilangkan data *null* :

```
#identifikasi Missing Values
any(is.na(cafe))
```

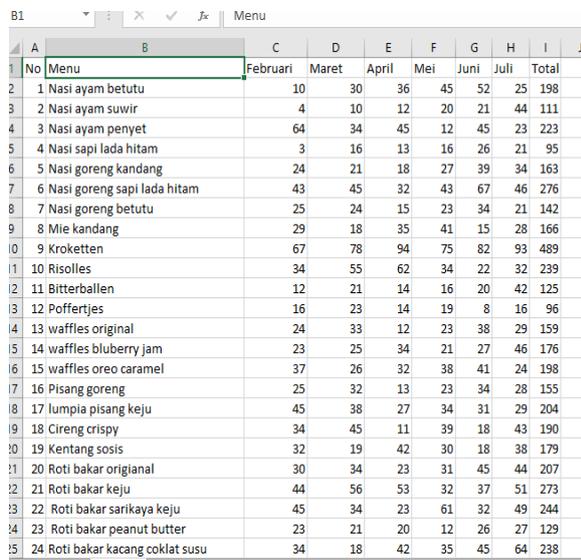
Setelah diidentifikasi *dataset* dengan *source code* di atas, maka berikut hasil dari data *cleaning* :

```
#identifikasi Missing Values
any(is.na(cafe))
[1] FALSE
```

Dari hasil *source code* di atas menunjukkan bahwa tidak ada data kosong (*Null*) didataset penjualan pada Ayam Betutu Warung Wardana.

### C. Data Transformasi

Pada tahapan ini akan dilakukan transformasi data ke dalam bentuk yang sesuai agar dapat dilakukan proses data *mining*. Mentransformasi data dapat menentukan kualitas dari hasil data *mining*. Pada pemrograman R hanya bisa mengolah data murni angka maka dari itu penulis menulis nama bulannya saja.



No	Menu	Februari	Maret	April	Mei	Juni	Juli	Total
1	Nasi ayam betutu	10	30	36	45	52	25	198
2	Nasi ayam suwir	4	10	12	20	21	44	111
3	Nasi ayam penyyet	64	34	45	12	45	23	223
4	Nasi sapi lada hitam	3	16	13	16	26	21	95
5	Nasi goreng kandang	24	21	18	27	39	34	163
6	Nasi goreng sapi lada hitam	43	45	32	43	67	46	276
7	Nasi goreng betutu	25	24	15	23	34	21	142
8	Mie kandang	29	18	35	41	15	28	166
9	Kroketten	67	78	94	75	82	93	489
10	Risolles	34	55	62	34	22	32	239
11	Bitterballen	12	21	14	16	20	42	125
12	Poffertjes	16	23	14	19	8	16	96
13	waffles original	24	33	12	23	38	29	159
14	waffles bluberry jam	23	25	34	21	27	46	176
15	waffles oreo caramel	37	26	32	38	41	24	198
16	Pisang goreng	25	32	13	23	34	28	155
17	lumpia pisang keju	45	38	27	34	31	29	204
18	Cireng crispy	34	45	11	39	18	43	190
19	Kentang sosis	32	19	42	30	18	38	179
20	Roti bakar original	30	34	23	31	45	44	207
21	Roti bakar keju	44	56	53	32	37	51	273
22	Roti bakar sarikaya keju	45	34	23	61	32	49	244
23	Roti bakar peanut butter	23	21	20	12	26	27	129
24	Roti bakar kacang coklat susu	34	18	42	35	45	64	238

**Gambar 3.2** Data Transformasi

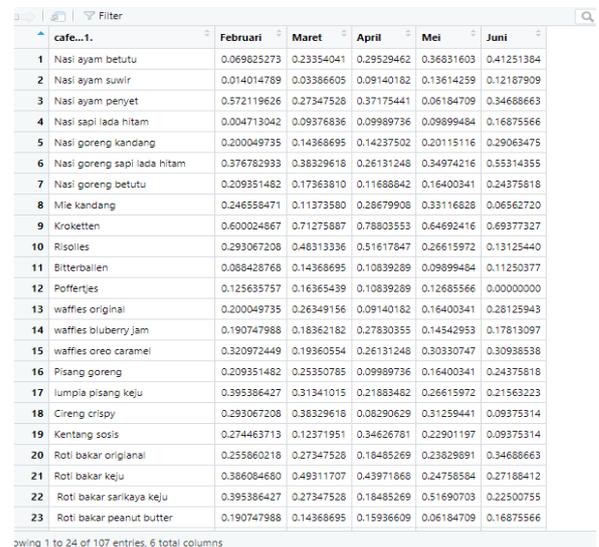
### D. Normalisasi Data Penjualan

Hasil transformasi kemudian melakukan normalisasi, dikarenakan agar tidak mengganggu keakuratan hasil pengujian. Normalisasi *Min-Max* digunakan dengan range nilai antara 0-1. Berikut merupakan

*source code* pemrograman R untuk normalisasikan data penjualan :

```
norm_clv=scale(cafe[,2:6])
norm_data <- (norm_clv - min(norm_clv))
/ (max(norm_clv) -min(norm_clv))
gabungan= data.frame(data_gb, norm_data)
View(gabungan)
summary(cafe)
```

Setelah Normalisasi data penjualan, berikut merupakan hasil data normalisasi yang ditunjukkan pada gambar 3.3:



	Februari	Maret	April	Mei	Juni
1 Nasi ayam betutu	0.069825273	0.23354041	0.29529462	0.36831603	0.41251384
2 Nasi ayam suwir	0.014014789	0.03366605	0.09140182	0.13614259	0.12187909
3 Nasi ayam penyset	0.572119626	0.27347528	0.37175441	0.06184709	0.34688663
4 Nasi sapi lada hitam	0.004713042	0.09376836	0.09899736	0.09899484	0.16875566
5 Nasi goreng kandang	0.200049735	0.14368695	0.14237502	0.20115116	0.29063475
6 Nasi goreng sapi lada hitam	0.376782933	0.38329618	0.26131248	0.34974216	0.55314355
7 Nasi goreng betutu	0.209351482	0.17363810	0.11688842	0.16400341	0.24375818
8 Mie kandang	0.246556471	0.11373580	0.28679908	0.33116828	0.06562720
9 Kroketten	0.600024867	0.71275887	0.78803553	0.64692416	0.69377327
10 Risolles	0.293067206	0.48313336	0.51617847	0.26615972	0.13125440
11 Bitterballen	0.088428768	0.14366695	0.10839289	0.09899484	0.11250377
12 Poffertjes	0.125635757	0.16365439	0.10839289	0.12685566	0.00000000
13 waffles original	0.200049735	0.26349156	0.09140182	0.16400341	0.28125943
14 waffles bluberry jam	0.190747988	0.18362182	0.27830355	0.14542953	0.17813097
15 waffles oreo caramel	0.320972449	0.19360554	0.26131248	0.30330747	0.30938538
16 Pisang goreng	0.209351482	0.25350785	0.09899736	0.16400341	0.24375818
17 lumpia pisang keju	0.395386427	0.31341015	0.21883482	0.26615972	0.21563223
18 Cireng crispy	0.293067206	0.38329618	0.08290629	0.31259441	0.09375314
19 Kentang sosis	0.274463713	0.12371951	0.34626781	0.22901197	0.09375314
20 Roti bakar original	0.255860216	0.27347528	0.18485269	0.23829891	0.34688663
21 Roti bakar keju	0.386084680	0.49311707	0.43971868	0.24758584	0.27188412
22 Roti bakar sarikaya keju	0.395386427	0.27347528	0.18485269	0.51690703	0.22500755
23 Roti bakar peanut butter	0.190747988	0.14368695	0.15936609	0.06184709	0.16875566

**Gambar 3.3** Hasil Normalisasi

### E. Cluster Optimal dengan Silhouette Coefisient

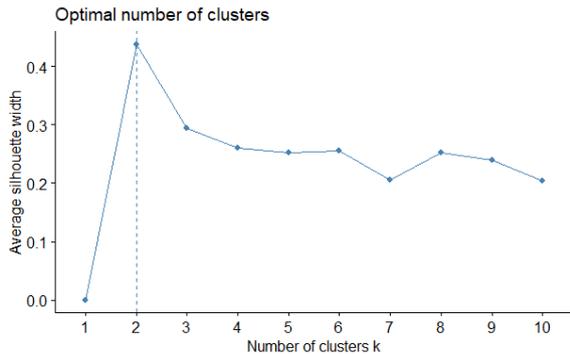
Setelah tahapan normalisasi dilakukan penentuan jumlah *cluster* optimal dengan pendekatan rata-rata nilai dengan metode *silhouette coefisien* untuk menguji kualitas dari *cluster* yang terbentuk, semakin tinggi nilai rata-ratanya maka akan semakin baik.

Berdasarkan grafik pada gambar di bawah ini banyak *cluster* optimal yang terbentuk pada K= 2. Berikut merupakan *source code* pemrograman R untuk melihat *silhouette coefisient dataset* :

```
fviz_nbclust(norm_data, kmeans,method = "silhouette")
sil5 <- silhouette(Klaster1$cluster, dist (norm_data))
fviz_silhouette(sil5)
sil5 <- silhouette(klaster2$cluster, dist (norm_data))
fviz_silhouette(sil5)
```

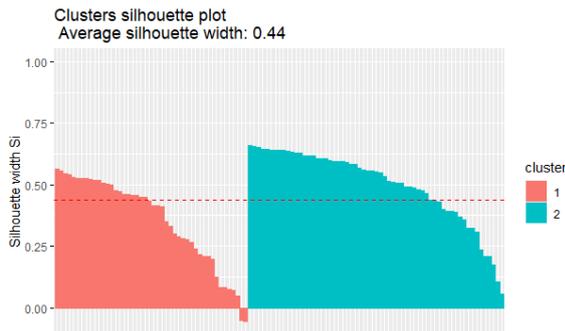


Berikut hasil nilai *cluster* dengan *silhouette* ditunjukkan pada gambar di bawah ini:

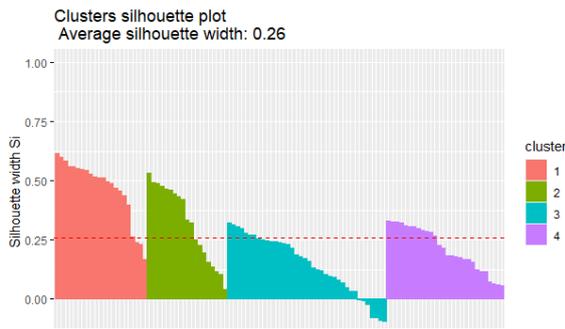


**Gambar 3.4 Silhouette Coefisient**

Pada validasi akhir menggunakan *silhouette coefisient* dapat menggunakan fungsi *fviz\_silhouette* dapat langsung divisualisasikan dengan hasil visualisasi pada contoh gambar 3.5 berikut:



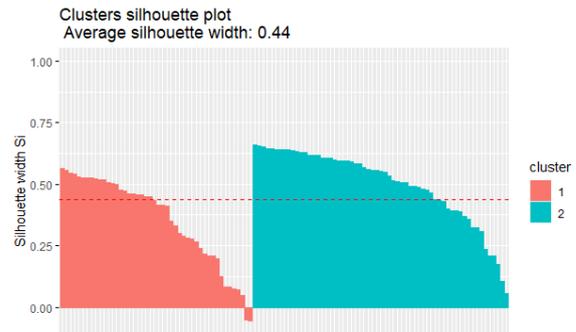
**Gambar 3.5 Silhoutte dengan 2 Cluster**



**Gambar 3.6 Silhoutte dengan 4 Cluster**

Dari hasil 2-4 *cluster* yang ditunjukkan pada gambar menunjukan hasil dari *silhouette coefisient* pada 2 *cluster* menghasilkan nilai rata-rata 0.44, *cluster* tersebut yang paling

mendekati 0,75 dalam artian mendekati angka 1 yaitu dengan hasil *cluster* sangat baik. Hasil validasi akhir *silhoutte* dengan 2 *cluster* nilai rata-rata tiap *cluster* bisa dilihat pada gambar 3.7 di bawah :



**Gambar 3.7 Hasil validasi akhir silhoutte**

F. Evaluasi Nilai *Devise Bouldin Index* (DBI)

Untuk menentukan nilai *Davies Bouldin Indeks* (DBI), berikut *source code* pemogram R :

```
d <- dist(norm_data)
dbi1 <- print(index.DB(norm_data,
Klaster1$cluster, d, centrotypes
= 'centroids'))
dbi2 <- print(index.DB(norm_data,
klaster2$cluster, d, centrotypes
= 'centroids'))
dbi1$DB
dbi2$DB
> dbi1$DB
[1] 0.9030707
> dbi2$DB
[1] 1.269624
```

Hasil dari validasi akhir menggunakan DBI seperti pada tabel yang hasilnya adalah 0.9030707 pada *cluster* 2.

**Tabel 1.** Evaluasi Nilai DBI

Cluster	DBI
2	0.9030707
4	1.269624

G. Perbandingan *Silhouette* dan DBI

Untuk mempermudah dalam melihat validasi nilai akhir, maka perbandingan evaluasi hasil antara *silhouette coefficient* dengan *Davies Bouldin Index* (DBI).

**Tabel 2.** Hasil Perbandingan *Silhouette* dengan DBI

Perbandingan <i>Silhouette</i> dan DBI			
Cluster	<i>Silhouette</i>	DBI	Evaluasi
2	0.44	0.9030707	2K yang paling optimal
4	0.26	1.269624	

Berdasarkan tabel 2 diketahui bahwa nilai pada metode *silhouette coefficient* terbaik terdapat pada K optimal = 2 pada metode *Davies Bouldin Index* (DBI) menunjukkan nilai ratio yang paling kecil terdapat pada K optimal = 2 maka clustering data penjualan terbaik di cluster 2 karena nilai *silhouette coefficient* mencapai 0.44 dan nilai *Davies Bouldin Index* yang paling kecil di angka 0.9030707.

H. Cluster K-means

Setelah menentukan cluster optimal menggunakan *silhouette* dan DBI tahapan selanjutnya adalah pengklasteran data penjualan dengan menggunakan algoritma K-means, data penjualan ditentukan menjadi 2 cluster. Berikut merupakan source code pemograman R untuk mengidentifikasi cluster K-means:

```
#melihat data klaster
Klaster1 <- kmeans(norm data, 2, nstart = 25)

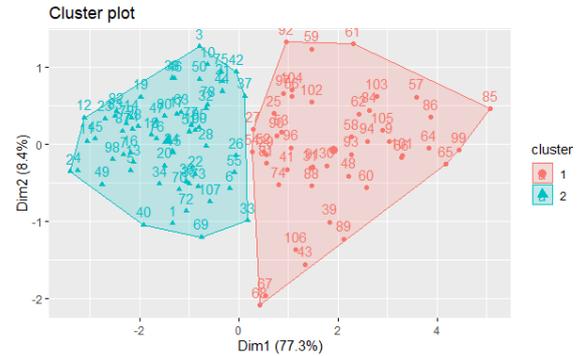
print(Klaster1)

hasil = data.frame(gabungan, Klaster1$cluster)
view(hasil)
summary(cafe)

#sebaran data
fviz_cluster(Klaster1, norm data)
```

Berdasarkan hasil dari penentuan *cluster* menggunakan *silhouette* dan DBI. Hasil

*clustering* menggunakan K-means dapat dilihat pada gambar 3.8 :



**Gambar 3.8** Hasil Cluster Plot

I. Hasil Proses K-Means

Setelah didapatkan hasil akhir dari proses K-means *clustering* diperoleh hasil pengelompokan pada masing-masing cluster yang dapat dilihat pada gambar 3.9 di bawah ini :

	Februari	Maret	April	Mei	Juni	klaster1.cluster
1 Nasi ayam betutu	0.069825273	0.23354041	0.29529462	0.36831603	0.41251384	2
2 Nasi ayam suwir	0.014014789	0.03366605	0.09140182	0.13614259	0.12187909	2
3 Nasi ayam penyot	0.572119626	0.27347528	0.37175441	0.06184709	0.34688663	2
4 Nasi sapi lada hitam	0.004713042	0.09376836	0.09989736	0.08999484	0.16875566	2
5 Nasi goreng kandang	0.200049735	0.14368695	0.14237502	0.20115116	0.29063475	2
6 Nasi goreng sapi lada hitam	0.376782933	0.38329618	0.26131248	0.34974216	0.55314355	2
7 Nasi goreng betutu	0.209351482	0.17363810	0.11688842	0.16400341	0.24375818	2
8 Mie kandang	0.246558471	0.11373580	0.28679906	0.33116828	0.06562720	2
9 Krokette	0.600024867	0.71275867	0.76803553	0.64692416	0.69377327	1
10 Risoles	0.293067208	0.48313336	0.51617847	0.26615972	0.13125440	2
11 Bitterbalen	0.088428768	0.14368695	0.10839289	0.08999484	0.11230377	2
12 Poffertjes	0.125635757	0.16365439	0.10839289	0.12685566	0.00000000	2
13 waffles original	0.200049735	0.26349156	0.09140182	0.16400341	0.28125943	2
14 waffles blueberry jam	0.190747988	0.18362182	0.27830355	0.14542853	0.17813097	2
15 waffles oreo caramell	0.320972449	0.19360554	0.26131248	0.30330747	0.30938538	2
16 Pisang goreng	0.209351482	0.23550765	0.09989736	0.16400341	0.24375818	2
17 lumpia pisang keju	0.395386427	0.31341015	0.21883482	0.26615972	0.21693223	2
18 Cireng crispy	0.293067208	0.38329618	0.08290629	0.31259441	0.09375314	2
19 Kentang sosis	0.274463713	0.12371951	0.34626781	0.22801197	0.09375314	2
20 Roti bakar original	0.255860218	0.27347528	0.18485269	0.23828891	0.34688663	2
21 Roti bakar keju	0.386084680	0.49311707	0.43971868	0.24758884	0.27188412	2
22 Roti bakar sarikaya keju	0.395386427	0.27347528	0.18485269	0.16907033	0.22500755	2
23 Roti bakar peanut butter	0.190747988	0.14368695	0.15936609	0.06184709	0.16875566	2

**Gambar 3.9** Hasil proses K-means

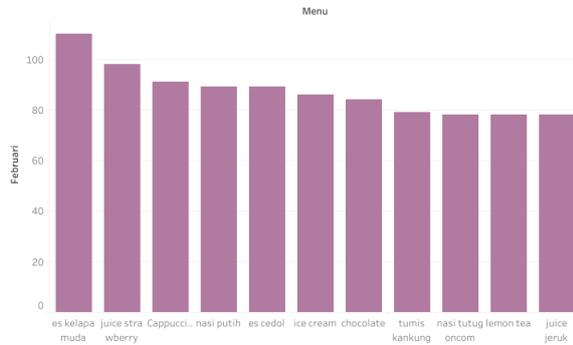
J. Hasil Visualisasi Penjualan

Untuk melihat detail dari produk yang banyak diminati sampai kurang diminati pelanggan, penulis menggunakan pemograman tableau agar memperlihatkan hasil visualisasi data dari rekap penjualan Februari sampai Juli 2023.

Berikut merupakan hasil visualisasi penjualan Ayam Betutu Warung Wardana dengan menggunakan pemograman Tableau yang berwarna ungu merupakan penjualan

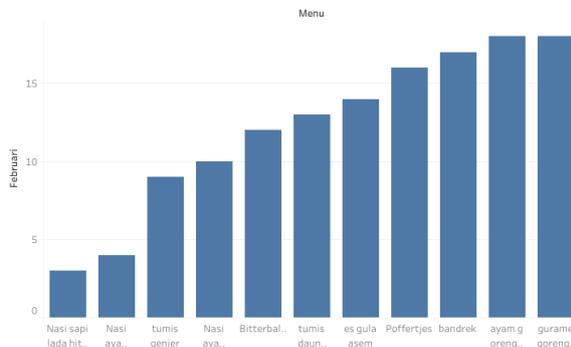
tertinggi dan yang biru merupakan penjualan terendah dalam setiap bulannya :

Menu Paling Banyak Terjual Pada Bulan Februari



**Gambar 3.10** Visualisasi Hasil Penjualan Tertinggi Februari

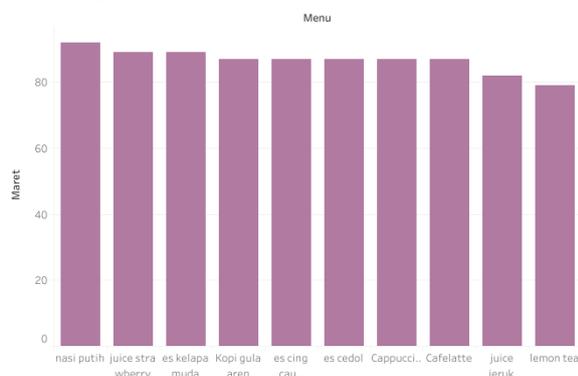
Menu Paling Sedikit Terjual Pada Bulan Februari



**Gambar 3.11** Visualisasi Hasil Penjualan Terendah Februari

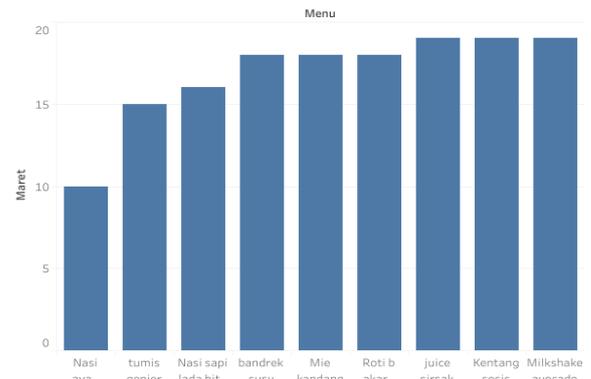
Pada gambar 3.10 dan gambar 3.11 dapat dilihat bahwa pada bulan Februari yang paling banyak terjual yaitu es kelapa muda dan menu yang paling sedikit terjual nasi sapi lada hitam.

Menu Paling Banyak Terjual Pada Bulan Maret



**Gambar 3.12** Visualisasi Hasil Penjualan Tertinggi Maret

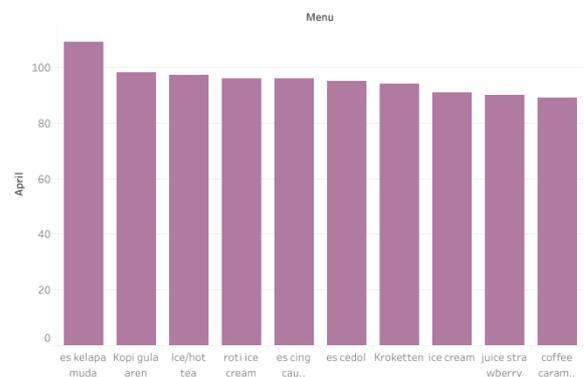
Menu Paling Sedikit Terjual Pada Bulan Maret



**Gambar 3.13** Visualisasi Hasil Penjualan Terendah Maret

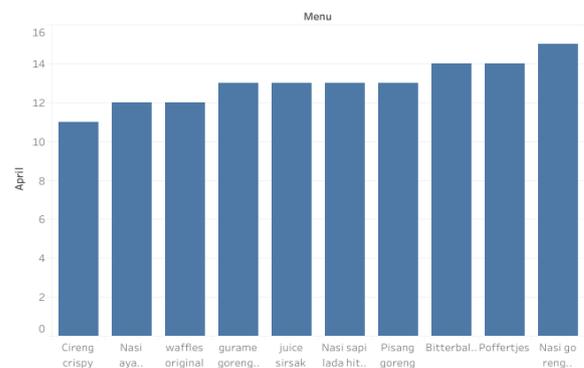
Selanjutnya pada gambar 3.12 dan gambar 3.13 merupakan penjualan pada bulan Maret di mana menu yang paling banyak dipesan pelanggan yaitu nasi putih, sedangkan menu yang paling sedikit terjual yaitu nasi ayam suwir.

Menu Paling Banyak Terjual Pada Bulan April



**Gambar 3.14** Visualisasi Hasil Penjualan Tertinggi April

Menu Paling Sedikit Terjual Pada Bulan April

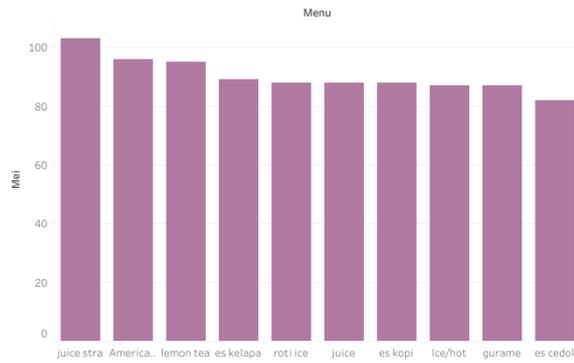


**Gambar 3.15** Visualisasi Hasil Penjualan Terendah April



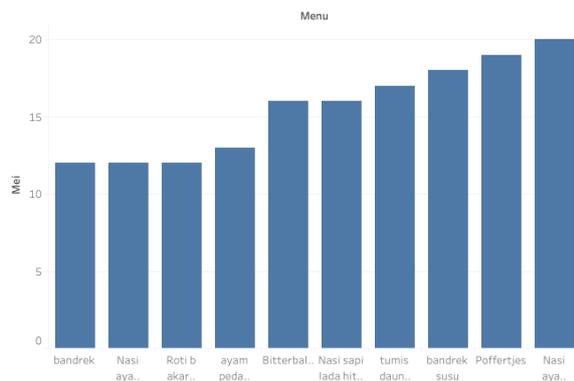
Selanjutnya pada gambar 3.14 dan gambar 3.15 merupakan penjualan pada bulan April dimana menu yang paling banyak dipesan yaitu es kelapa muda, serta menu yang paling sedikit dipesan yaitu cireng krispi.

Menu Paling Banyak Terjual Pada Bulan Mei



**Gambar 3.16** Visualisasi Hasil Penjualan Tertinggi Mei

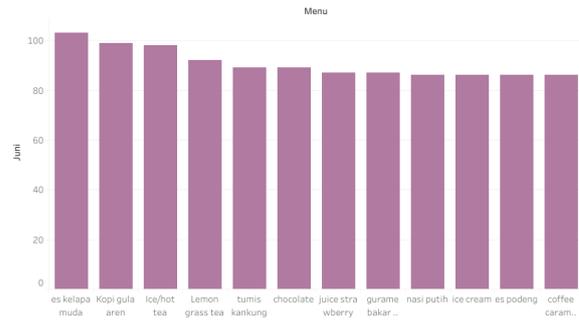
Menu Paling Sedikit Terjual Pada Bulan Mei



**Gambar 3.17** Visualisasi Hasil Penjualan Terendah Mei

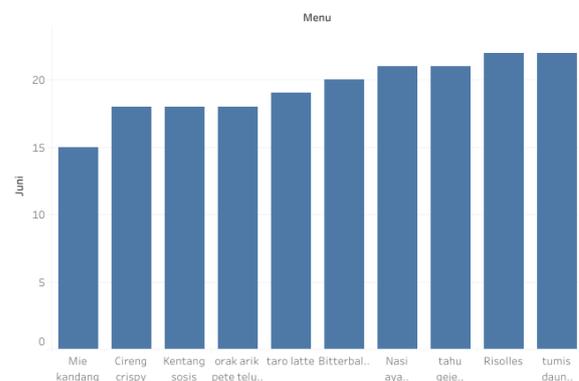
Selanjutnya pada gambar 3.16 dan gambar 3.17 merupakan penjualan pada bulan Mei di mana menu yang paling banyak dipesan yaitu *juice strawberry*, dan menu yang paling sedikit terjual bandrek.

Menu Paling Banyak Terjual Pada Bulan Juni



**Gambar 3.18** Visualisasi Hasil Penjualan Tertinggi Juni

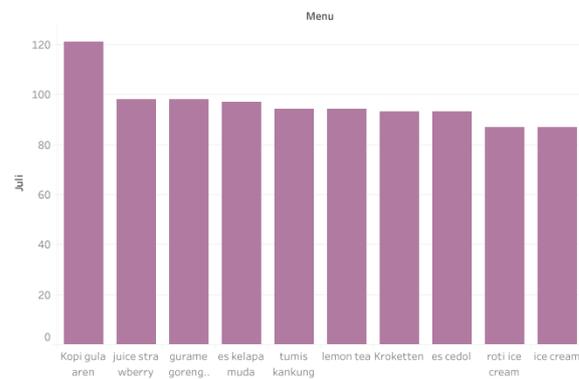
Menu Paling Sedikit Terjual Pada Bulan Juni



**Gambar 3.19** Visualisasi Hasil Penjualan Terendah Juni

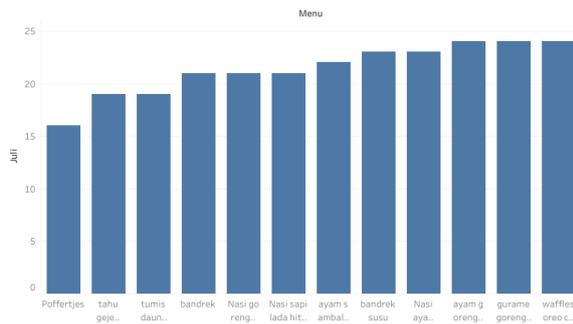
Selanjutnya pada gambar 3.18 dan gambar 3.19 merupakan penjualan pada bulan Juni dimana menu yang paling banyak terjual yaitu es kelapa muda, sedangkan menu yang paling sedikit di pesan yaitu mie kandang.

Menu Paling Banyak Terjual Pada Bulan Juli



**Gambar 3.20** Visualisasi Hasil Penjualan Tertinggi Juli

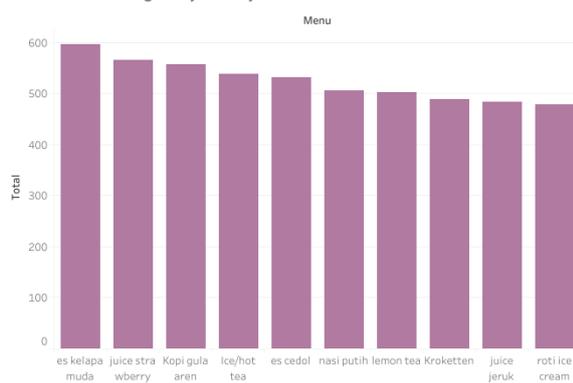
Menu Paling Sedikit Terjual Pada Bulan Juli



**Gambar 3.21** Visualisasi Hasil Penjualan Terendah Juli

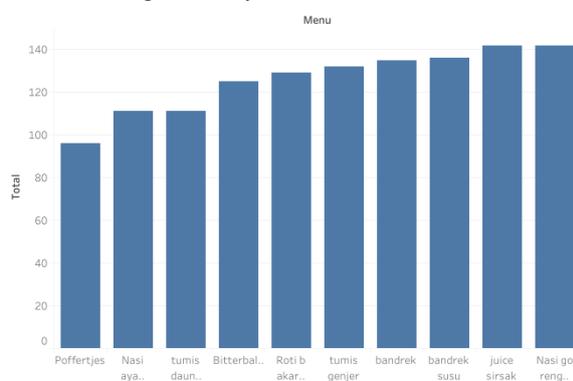
Selanjutnya pada gambar 3.20 dan gambar 3.21 merupakan penjualan pada bulan Juli dimana menu yang paling banyak dipesan pelanggan yaitu kopi gula aren, sedangkan menu yang paling sedikit diminati pelanggan yaitu *poffertjes*.

Total Menu Paling Banyak Terjual



**Gambar 3.22** Visualisasi Hasil Penjualan Total Tertinggi

Total Menu Paling Sedikit Terjual



**Gambar 3.23** Visualisasi Hasil Penjualan Total Terendah

Pada gambar 3.22 dan gambar 3.23 merupakan total penjualan dari bulan Februari sampai Juli di mana menu yang paling banyak dipesan pelanggan yaitu es kelapa muda, sedangkan menu yang paling sedikit dipesan pelanggan yaitu *poffertjes*.

#### 4. KESIMPULAN

Berdasarkan analisis yang telah dilakukan maka diperoleh beberapa kesimpulan sebagai berikut: 1) Penelitian ini menghasilkan 2 *cluster* yang terbaik dengan metode validasi *silhouette* dan DBI. Pengujian menggunakan *Silhouette Coefisient* menunjukkan nilai 0.44 dan itu merupakan nilai terbaik yang mendekati angka 1. *Davies Bouldin Index* (DBI) menunjukkan nilai dengan *ratio* terkecil pada angka 0.9030707. Hal tersebut baik karena *ratio* mendekati nilai nol, maka semakin baik yang diperoleh dari pengelompokan K-Means; 2) Berdasarkan hasil penelitian K-means *clustering* untuk pengelompokan minat pelanggan pada produk Ayam Batutu Warung Wardana yang telah diuraikan maka disimpulkan bahwa penerapan sebuah informasi mengenai produk yang diminati dan produk yang kurang diminati. Dari 107 produk yang dapat menghasilkan 2 kelompok, yaitu *clustering* 1 merupakan produk yang paling banyak diminati atau yang paling banyak dibeli sedangkan *clustering* 2 merupakan produk yang kurang diminati atau sedikit pembeli. Untuk penelitian selanjutnya bisa menggunakan algoritma lain agar lebih baik mengetahui data penjual dan pembeli setiap bulannya seperti menggunakan algoritma klasifikasi. Selain itu, juga dapat menambahkan sumber data lain seperti waktu pembelian agar lebih memahami produk mana saja yang paling sering terjual pada waktu tertentu.

#### 5. DAFTAR PUSTAKA

- [1] Aji, Farell Apuilino Iman Seno,etal.“PENERAPAN METODE CLUSTERING PADA ANALISIS REALISASI PENDAPATAN ASLI

- DAERAH DENGAN ALGORITMA K-MEANS.” JATI (Jurnal Mahasiswa Teknik Informatika), vol. 5, no. 2, 21 Oct. 2021, pp. 443-451, [ejournal.itn.ac.id/index.php/jati/article/view/3741](http://ejournal.itn.ac.id/index.php/jati/article/view/3741), (<https://doi.org/10.36040/jati.v5i2.3741>). Accessed 13 Feb. 2023.
- [2] Poltak Sianturi, et al. “Analisis Data Mining Menggunakan Algoritma K-Means Dalam Mengklusterkan Menu Makan Potensial Di Cafe Minum Kopi.” Jurnal Cyber Tech, vol. 3, no. 8, 2020, pp. 1412–1419, ([ojs.trigunadharma.ac.id/index.php/jct/article/view/4657](http://ojs.trigunadharma.ac.id/index.php/jct/article/view/4657)). Accessed 13 Feb. 2023.
- [3] Badruttamam, Ahmad, et al. “PENERAPAN ANALISIS KLASTER K- MODES DENGAN VALIDASI DAVIES BOULDIN INDEX DALAM MENENTUKAN KARAKTERISTIK KANAL YOUTUBE DI INDONESIA (Studi Kasus: 250 Kanal YouTube Indonesia Teratas Menurut Socialblade).” Jurnal Gaussian, vol. 9, no. 3, 30 Aug. 2020, pp. 263–272, ([ejournal3.undip.ac.id/index.php/gaussian/article/view/28907](http://ejournal3.undip.ac.id/index.php/gaussian/article/view/28907), <https://doi.org/10.14710/j.gauss.9.3.263-272>). Accessed 13 Feb. 2023.
- [4] Hidayati, Rahmatina, et al. “Analisis Silhouette Coefficient Pada 6 Perhitungan Jarak K-Means Clustering.” Techno.com, vol. 20, no. 2, 19 May 2021, pp. 186–197, (<https://doi.org/10.33633/tc.v20i2.4556>). Accessed 24 Nov. 2021.
- [5] Permana, Inggih, and Febi Nur Salisah. “Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation: The Effect of Data Normalization on the Performance of the Classification Results of the Backpropagation Algorithm.” Indonesian Journal of Informatic Research and Software Engineering (IJIRSE), vol. 2, no. 1, 31 Mar. 2022, pp. 67–72, ([journal.irpi.or.id/index.php/ijirse/article/view/311](http://journal.irpi.or.id/index.php/ijirse/article/view/311), <https://doi.org/10.57152/ijirse.v2i1.311>). Accessed 13 Feb. 2023.
- [6] Puteri, Suci Rahma. “Analisis Visualisasi Data Kecamatan Kertapati Menggunakan Tableau Public.” JUPITER (Jurnal Penelitian Ilmu Dan Teknik Komputer), vol. 14, no. 2-b, 2022, pp. 366–373, [jurnal.polsri.ac.id/index.php/jupiter/article/view/5141/2197](http://jurnal.polsri.ac.id/index.php/jupiter/article/view/5141/2197). Accessed 20