

# Klasterisasi Outlet Berdasarkan Data Penjualan Dengan Menggunakan Algoritma K-Medoids

**SRI MULYANI AZHAR<sup>1</sup>, TACBIR HENDRO PUDJIANTORO<sup>2</sup>, IRMA SANTI KARAMA<sup>3</sup>**

<sup>1 2 3</sup>Program Studi Teknik Informatika, Fakultas Sains dan Informatika  
Universitas Jenderal Achmad Yani  
Jl. Terusan Sudirman, Cimahi  
srimulyaniazhar17@if.unjani.ac.id

## ABSTRAK

*Data mining kini digunakan diberbagai bidang termasuk bisnis. Data mining memiliki beberapa teknik, salah satunya yang populer yaitu teknik Clustering. Ketersediaan data yang melimpah, kebutuhan akan ketersediaan informasi (atau pengetahuan) sebagai acuan bagi pelaku bisnis dalam proses pengambilan keputusan untuk membuat solusi bisnis merupakan cikal-bakal dari lahirnya teknologi data mining. PT. Gondowangi salah satu bisnis yang memiliki ketersediaan data yang melimpah, menghasilkan data penjualan dari setiap outlet yang beragam sehingga pendistribusian produk dan penentuan strategi penjualan. Analisis terhadap data penjualan dengan jumlah yang besar dan data yang beragam tidak efektif dan efisien. Proses analisis data yang dilakukan manual memerlukan waktu yang lama, sehingga ketersediaan informasi atau pengetahuan mendalam dari data penjualan sulit ditemukan. Penelitian ini mengarah pada penggunaan teknik clustering dengan algoritma K-Medoids untuk mengeksplorasi data guna mendapatkan pengelompokan outlet berdasarkan data penjualan. Penggunaan teknik clustering dengan algoritma K-Medoids mampu menunjukkan pengelompokan terhadap data penjualan yang menghasilkan pengelompokan outlet. Hasil pengujian pada hasil clustering dengan beberapa nilai K didapatkan jumlah K=3 memiliki nilai silhoutte Coefficient yang paling tinggi dan mendekasi  $Si = 1$  yaitu dengan nilai 0,92669616303752 yang merupakan struktur kuat.*

**Kata kunci:** Data mining, clustering, K-medoids.

## ABSTRACT

*Data mining is now used in various fields including business. Data mining has several techniques, one of which is popular, namely the Clustering technique. The availability of abundant data, the need for the availability of information (or knowledge) as a reference for business people in the decision-making process to create business solutions is the forerunner of the birth of data mining technology. PT. Gondowangi is one of the businesses that have abundant data availability, generating sales data from various outlets so that product distribution and sales strategy determination. Analysis of sales data with large amounts and diverse data is not effective and efficient. The manual data analysis process takes a long time, so the availability of information or in-depth knowledge of sales data is difficult to find. This research leads to the use of clustering techniques with the K-Medoids*

*algorithm to explore the data to obtain a grouping of outlets based on sales data. The use of clustering techniques with the K-Medoids algorithm can show the grouping of sales data which results in the grouping of outlets. The test results on the results of clustering with several K values obtained the number of K = 3 has the highest Silhouette Coefficient value and dedicates  $S_i = 1$  with a value of 0.92669616303752 which is a strong structure.*

**Keywords:** *Data mining, clustering, K-medoids.*

## 1. PENDAHULUAN

Hadirnya data mining dilihat sebagai salah satu hasil dari evolusi alami perkembangan teknologi informasi. Industri database dan manajemen data berkembang dalam pengembangan beberapa fungsi penting (Agarwal, 2014). Data mining bertujuan untuk menemukan pola data yang berguna dari jumlah data yang besar (Fu, 1997). Data mining kini digunakan diberbagai bidang termasuk bisnis, ketersediaan data yang melimpah, kebutuhan akan ketersediaan informasi (atau pengetahuan) sebagai acuan bagi pelaku bisnis dalam proses pengambilan keputusan untuk membuat solusi bisnis, dan sebagai pengembangan di bidang teknologi informasi merupakan cikal-bakal dari lahirnya teknologi data mining (Deker, G. W., Pechenizkiy, M., Vleeshouwers, 2009). Teknik data mining yang populer dalam bidang bisnis diantaranya prediksi penjualan (Ribeiro & Seruca, n.d.), Klasifikasi pelanggan (SULISTIANI, 2016), *Association rule* untuk menemukan pola penjualan dan pembelian suatu produk (Djamaludin & Nursikuwagus, 2017) serta *Clustering* untuk segmentasi pasar (Chiu et al., 2009).

Salah satu bidang bisnis yang memiliki ketersediaan data yang melimpah adalah industri kosmetik & skincare, seperti PT. Gondowangi Tradisional Kosmetika yang bergerak dibidang distribusi produk kosmetik dan skincare yang melakukan distribusi ke outlet diberbagai daerah. Dari kegiatan distribusi menghasilkan data penjualan yang beragam setiap outletnya, proses analisis data yang dilakukan manual terhadap data penjualan dengan jumlah yang besar dan data yang beragam kurang efektif dan efisien karena memerlukan waktu yang lama. Sehingga eksplorasi data guna menemukan pengetahuan mendalam dari data penjualan sulit ditemukan. Seperti informasi mengenai karakteristik penjualan dari setiap outlet yang dapat dijadikan sebagai acuan bahan analisis selanjutnya dalam penentuan strategi penjualan. Sehingga diperlukan eksplorasi terhadap data penjualan untuk memunculkan pola pengelompokan outlet yang tersebar diberbagai daerah berdasarkan data penjualan

Dalam bidang bisnis ada beberapa penelitian yang telah menggunakan data mining dengan teknik *clustering* untuk mengelompokan data diantaranya teknik *clustering* Sebagai Strategi Pemasaran pada Penjualan Buku Di Tokopedia dan Shopee (Baihaqi et al., 2019), *clustering* pelanggan listrik perumahan menggunakan load time series (Motlagh et al., 2019) , dan Segmentasi Pelanggan Berdasarkan Produk (Prakasawati et al., 2019). Algoritma *clustering* yang digunakan pada beberapa penelitian tersebut salah satunya yaitu K-Medoids, algoritma K-Medoids menggunakan objek representatif dengan secara acak K objek sebagai titik acuan (medoid) (Baihaqi et al., 2019). K-Medoids melakukan pengelompokan data yang serupa sehingga data dalam satu kelompok memiliki kesamaan yang sama tetapi memiliki perbedaan dengan data dalam kelompok lain (Fu, 1997).

Berdasarkan pembahasan diatas maka penelitian ini berfokus untuk melakukan eksplorasi terhadap data penjualan dengan teknik *clustering* untuk klasterisasi outlet berdasarkan data penjualan dengan menggunakan Algoritma K-Medoids guna membantu mengelompokan outlet yang dimana hasilnya dapat digunakan sebagai acuan analisis dan pengambilan keputusan dalam mendukung penentuan strategi penjualan.

## 2. METODOLOGI

Dalam penelitian ini terdapat beberapa tahapan, pertama menyiapkan data sebagai objek penelitian, data yang digunakan yaitu data penjualan. Tahap kedua yaitu melakukan Proses KDD serta melakukan proses data mining dengan menggunakan algoritma K-Medoids. Tahapan-tahapan yang dimaksud diantaranya:

### 3.1. Data Collections

Pengumpulan data yang dilakukan pada penelitian ini yaitu dengan melakukan wawancara dengan pihak organisai terkait untuk mendapatkan data yang akan menjadi bahan pada penilitian ini.

### 3.2. Data Selections

Tahapan ini merupakan proses pemilihan data, data yang telah dipilih dipisahkan dibuat menjadi *data warehouse* (Yahya & Puspita Hidayanti, 2020). Dari banyaknya kegiatan operasional menghasilkan jenis data yang beragam. Data yang digunakan pada penelitian ini yaitu data transaksi penjualan milik perusahaan PT. Gondowangi Tradisional Kosmetika produk azalea. Pada tahapan selections dilakukan juga pemilihan variable data yang akan digunakan untuk proses *clustering*.

### 3.3. Data Pre-processing

Tahapan pre-processing data melibatkan pembersihan data, integrasi data, transformasi data, reduksi data atau kompresi data untuk representasi ringkas (Gürbüz et al., 2011). Pre-processing yg dilakukan pada penelitian ini yaitu pembersihan data membuang duplikasi data, memeriksa data yang tidak konsisten. Data hasil proses *cleaning* akan menyajikan data tanpa noise, tanpa record data yang duplikat dan data yang tidak konsisten.

### 3.4. Data Transformation

Pada tahap ini merupakan proses transformasi untuk data yang telah melewati proses pembersihan data, sehingga data tersebut sesuai untuk proses data mining. Transformasi data pada penilitian ini melakukan normalisasi data menggunakan teknik min-max normalisasi dengan merubah data menjadi skala lebih kecil bernilai 0-1. Hasil data yang telah di normalisasi disajikan kembali sebagai dataset untuk pemrosesan data mining. Persamaan (1) untuk min-max normalisasi sebagai berikut (Agarwal, 2014):

$$V_i' = \frac{V_i - \min_A}{\max_A - \min_A} (\text{new\_max}_A - \text{new\_min}_A) + \text{new\_min}_A \quad (1)$$

Keterangan :

$V_i'$  = Nilai setelah normalisasi

$V_i$  = Nilai sebelum normalisasi

$\min_A$  = Nilai minimal setiap variable

$\max_A$  = Nilai maksimal setiap variable

$\text{new\_max}_A$  = Rentang maksimal dengan 1

$\text{new\_min}_A$  = Retang minimal dengan 0

### 3.5. Data Mining Clustering

Teknik atau metode pada data mining meliputi klasifikasi, pengelompokan, peringkasan, regresi, menemukan aturan asosiasi dan dependensi fungsional, dan ekstraksi aturan (Gürbüz et al., 2011). Proses Data mining merupakan suatu proses mencari pola atau informasi yang sangat menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik data mining yang digunakan dalam penelitian ini yaitu *clustering* dengan menggunakan algoritma K-Medoids. K-Medoids hadir menajdi penanggulangan untuk kekurangan pada K-Means. lebih baik dalam hal waktu eksekusi, tidak sensitif terhadap outlier dan mengurangi

noise dibandingkan dengan K-Means karena meminimalkan jumlah perbedaan objek data (Arora & Varshney, 2016). Adapun tahapan proses dari algoritma K-Medoid dalam membentuk *cluster* yaitu sebagai berikut:

- a. Pertama pilih secara random K objek pada sekumpulan data objek sesuai jumlah titik pusat cluster sebagai medoid awal.
- b. Setiap objek atau data yang bukan merupakan medoid di masukan ke dalam cluster berdasarkan jarak terdekat dengan menggunakan persamaan Euclidean distance.
- c. Pilih objek secara acak yang diambil dari anggota di masing-masing cluster sebagai kandidat medoid baru.
- d. Hitung jarak objek yang terdapat pada masing - masing *cluster* dengan medoid baru.
- e. Hitung total cost  $S$ , dari pertukaran medoid dengan jumlah cost baru – jumlah cost lama.
- f. Jika  $S < 0$  maka tukar objek secara acak untuk membentuk sekumpulan k objek baru sebagai medoid.
- g. Lakukan kembali langkah ke 3 sampai 5 hingga tidak ada perubahan (Baihaqi et al., 2019).

Secara umum untuk mendefinisikan jarak antara dua objek  $x$  dan  $y$ , digunakan Persamaan (2) *Euclidean distance* (Larose & Larose, 2014).

$$d_{Euclidean}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Keterangan :

$d_{xy}$  : *Euclidean distance*, jarak objek  $x$  dan  $y$

$n$  : jumlah data

$x_i$  : nilai data variabel ke  $i$

$y_i$  : nilai data variabel sentroid ke  $i$

### 3.6. Evaluasi

Evaluasi yang dilakukan yaitu melakukan pengujian pada hasil cluster dengan menggunakan persamaan Silhouette Coefficient. Metode penggabungan dari pada dua metode yaitu metode *Cohesion* yang berguna mengukur kedekatan relasi antar satu objek dengan objek lainnya dalam satu cluster dan metode *Separation* yang berguna untuk menghitung seberapa jauh sebuah cluster berpisah dari *cluster* yang lain atau sejauh apa sebuah cluster dengan cluster yang lainnya. Sehingga dengan menggunakan Silhouette Coefficient akan dihitung kedekatan anggota dalam *cluster* dan luar *cluster*. Tahapan pada perhitungan Silhouette Coefficient terdiri dari 3 tahapan diantaranya :

1. Pada setiap objek  $i$  dihitung nilai rata-rata titik satu dengan semua objek yang berada dalam satu cluster. Maka akan diperoleh nilai rata-rata yang disebut dengan  $a_j$
2. Pada setiap objek  $i$  dihitung nilai minimal jarak rata-rata dari titik satu ke titik yang lain yang berbeda cluster. Maka akan diperoleh nilai rata-rata minimum yang disebut dengan  $b_j$

Setelah semua nilai diketahui maka nilai Silhouette Coefficient dapat ditentukan dengan menggunakan (3):

$$Si = \frac{b_i - a_i}{\max(a_i - b_i)} \quad (3)$$

Keterangan :

$b$  : Jarak rata-rata medoid dengan obyek di luar *cluster*

$a$  : Jarak rata-rata antara medoid dengan objek di dalam

Nilai Silhoutte Coefficient terdapat pada angka antara nilai -1 sampai 1. Berikut adalah Silhoutte Coeffisien menurut Kaufman dan Rousseeuw:

$0.7 < SC \leq 1$  merupakan cluster Strong stucture

$0.5 < SC \leq 0.7$  merupakan cluster Medium structure

$0.25 < SC \leq 0.5$  merupakan cluster Weak structure

$SC \leq 0.25$  merupakan cluster No structure.

### 3. HASIL PENELITIAN DAN ANALISIS

Hasil dari penelitian ini yaitu klasterisasi outlet yang di proses oleh sistem yang telah berhasil dibangun untuk eksplorasi terhadap data. Proses data mining hingga membentuk *cluster* outlet dengan menggunakan algoritma K-Medoids terdiri dari beberapa tahapan sebagai berikut:

#### 3.1. Data Selection

Dari banyak data yang dihasilkan pada kegiatan operasional di PT.Gondowangi, data yang digunakan dalam penelitian ini yaitu data penjualan produk merk Azalea dari tahun 2017 sampai 2020 dari dua regional berbeda dengan jumlah 3000 record data. Data penjualan ini berisi catatan transaksi penjualan ke setiap outlet diberbagai daerah yang berupa dokumen excel yang kemudian disimpan terpisah dari data operasional lainnya. Data transaksi penjualan yang digunakan memiliki 20 variable, pada proses selections dilakukan juga pemilihan variable data yang akan digunakan untuk proses clustering outlet. Sehingga variable yang digunakan menjadi 4 variable yang menunjang dalam proses clustering outlet dengan menggunakan algoritma K-Medoids. Variabel yang dipilih tersebut menggambarkan jumlah penjualan produk ke setiap outlet. Adapun contoh dari data transaksi penjualan produk azalea yang utuh dan belum melewati proses *selection* untuk pemilihan variable dapat dilihat pada Tabel 1.

**Tabel 1. Data Hasil Selections**

No.	Kd_Outlet	Ttl_sls_val	Ttl_Sls_Unit	Sales(Rp)	Sales (Pc)
1	BDG1-9869	30056	2	30056	2
2	BDG1-6864	45084	3	45084	3
3	BDG1-6864	45084	3	45084	3
4	BDG1-6521	1803360	120	1803360	120
5	BDG1-6282	360672	24	360672	24
6	BDG1-6282	45084	3	45084	3
7	BDG1-6282	180336	12	180336	12
8	BDG1-4540	180336	12	180336	12
9	BKS1-2648	30384	2	45576	3
10	BKS1-2648	30056	2	30056	2

No.	Kd_Outlet	Ttl_sls_val	Ttl_Sls_Unit	Sales(Rp)	Sales (Pc)
11	BKS1-2622	45576	3	45576	3
12	BKS1-2622	45084	3	45084	3
.....	.....	.....	.....	.....	....
.....	.....	.....	.....	.....	....
2993	4402.N002	840000	24	840000	24
2994	4402.N002	840000	24	840000	24
2995	4402.N002	216228	12	216228	12
2996	4402.N002	108111	6	108111	6
2997	4402.L101	864912	48	864912	48
2998	4402.L101	864912	48	864912	48
2999	4402.L101	432456	24	432456	24
3000	4402.L101	1430400	96	1430400	96

### 3.2. Data Pre-processing

Proses setelah melakukan data *selection* yaitu melakukan *cleaning* data. Proses *cleaning* data ini terdiri dari beberapa operasi dasar, seperti normalisasi, penghilangan noise dan penanganan data yang hilang, pengurangan data yang tidak diperlukan dan pengurangan duplikasi data. Proses *cleaning* yang dilakukan terhadap data penjualan ini yaitu melakukan pengurangan duplikasi data. Pengelompokan yang dilakukan yaitu untuk mengidentifikasi *cluster* dari outlet. Sehingga data yang semula berupa record transaksi dari setiap outlet memungkinkan setiap outlet memiliki beberapa record transaksi yang artinya dapat ditemukan Kd\_outlet yang sama. Maka pembersihan data ini mengambil record transaksi berdasarkan Kd\_outlet, hasilnya tidak ada Kd\_outlet yang duplikat. Data yang semula berjumlah 3000 *record* menjadi 437 *record*, contoh hasil pembersihan data ditunjukkan pada Tabel 2.

Tabel 2. Data Hasil Cleaning

No.	Kd_Outlet	Ttl_sls_val	Ttl_Sls_Unit	Sales(Rp)	Sales (Pc)
1	3300.G001	29799	62	1527315	62
2	3301.A1067	25300	2	50600	2
3	3301.A1430	126500	10	253000	10
4	3301.A2055	420000	24	840000	24
5	3301.B1015	151800	48	1109260	48
6	3301.B1342	51250	2	51250	2
7	3301.B1371	50600	4	101200	4
8	3301.B330	18019	2	18019	1
9	3301.B344	54056	21	378395	21
10	3301.B346	89397	35	598854	27
11	3301.B347	108114	116	2409776	108
12	3301.B348	108111	173	503133	30
13	3301.B349	324334	246	5216683	246
14	3301.B350	75900	73	1499132	66
15	3301.B351	54056	51	1121731	51
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
437	JKT1-0631	360672	474	7159602	474

### 3.3. Data Transformations

Tahapan selanjutnya yaitu data transformasi yang berkaitan dengan proses konversi data mengubah sumber data menjadi bentuk data yang disesuaikan dengan kebutuhan pemrosesan metode yang digunakan. Terdapat beberapa teknik transformasi data, pada penelitian ini

menggunakan teknik normalisasi dengan menggunakan persamaan min-max normalisasi data untuk proses data transformasi. Normalisasi mengubah data menjadi skala yang lebih kecil, yaitu dengan batas terendah 0 dan batas tertinggi 1. Hasil data yang telah di normalisasi ditunjukkan pada Tabel 3.

**Tabel 3. Data Hasil Normalisasi**

No.	Kd_Outlet	Ttl_sls_val	Ttl_Sls_Unit	Sales(Rp)	Sales (Pc)
1	3300.G001	0,000252	0,009223	0,012827	0,009226
2	3301.A1067	0,000212	0,000297	0,000424	0,000297
3	3301.A1430	0,001062	0,001487	0,002124	0,001488
4	3301.A2055	0,003526	0,003571	0,007054	0,003571
5	3301.B1015	0,001274	0,007140	0,009316	0,007142
6	3301.B1342	0,000433	0,000297	0,00043	0,000297
7	3301.B1371	0,000424	0,000595	0,00084	0,000595
8	3301.B330	0,000151	0,000297	0,000151	0,000148
9	3301.B344	0,000453	0,003124	0,003177	0,003125
10	3301.B346	0,00075	0,005206	0,005029	0,004017
11	3301.B347	0,000972	0,017256	0,020238	0,016071
12	3301.B348	0,000913	0,025736	0,004225	0,004464
13	3301.B349	0,002722	0,036596	0,043812	0,036607
...	.....	.....	.....	.....	....
...	.....	.....	.....	.....	....
437	JKT1-0631	0,003028	0,070514	0,060137	0,070535

### 3.4. Hasil Clustering

*Clustering* yang dilakukan untuk mengelompokan data outlet ini menggunakan algoritma K-Medoids. Setelah data melewati *data selections*, *data cleaning*, dan *data trasformations* berikut adalah hasil *clustering* dengan menggunakan titik pusat *cluster* K=3 ditunjukkan pada Tabel 4.

**Tabel 4. Hasil Clustering**

Data-Ke	C-1	C-2	C-3	Kedekatan	Cluster
1	0,017182	0,033727	0,017389	0,017182	1
2	0,000775	0,051009	0,017182	0,017182	3
3	0,001957	0,048666	0,002267	0,001957	1
4	0,008152	0,043650	0,008494	0,008152	1
5	0,012633	0,038055	0,012857	0,012633	1
6	0,000679	0,051006	0,000626	0,000626	3
7	0,000448	0,050422	0,000613	0,000448	1
8	0,000956	0,051251	0,000717	0,000717	3
9	0,004311	0,046155	0,004495	0,004311	1
10	0,007118	0,043384	0,007315	0,007118	1
11	0,029936	0,020996	0,030135	0,020996	2
12	0,025422	0,036059	0,025470	0,025422	1
13	0,066714	0,017414	0,066916	0,017414	2
...	.....	.....	.....	.....	.....
...	.....	.....	.....	.....	.....
435	0,000547	0,050842	0,000457	0,000457	3
436	0,013089	0,037420	0,013289	0,013089	1
437	0,115346	0,065086	0,115506	0,065086	2

### 3.5. Pengujian Clustering

Pada penelitian ini metode pengujian yang digunakan yaitu *silhouette Coefficient* dengan menggunakan *silhouette Coefficient* akan dihitung kedekatan anggota dalam *cluster* dan luar *cluster*. Pengujian dilakukan terhadap seluruh data outlet sebanyak 437. Pengujian terhadap hasil clustering dengan titik pusat cluster = 3 maka hasilnya menunjukkan nilai *silhouette coefficient* mendekati  $S_i=1$  yaitu dengan nilai 0,92669616303752. Pada Gambar 1 menunjukkan hasil perhitungan *silhouette* oleh sistem yang telah dibangun.

No	Cluster	Jumlah Anggota	Rata-rata Cluster
1	k-1	171	0.8923274697081
2	k-2	249	0.89323638894103
3	k-3	17	0.89561737320073

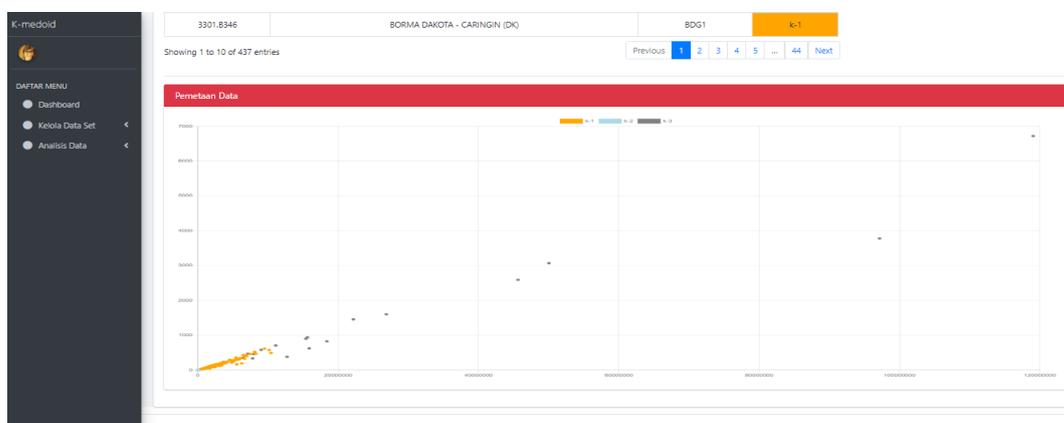
Showing 1 to 3 of 3 entries

Setinggi hasil silhouette pada K = 3 adalah 0,92669616303752 (Struktur Kuat)

Gambar 1. Interpretasi Hasil Pengujian Cluster

### 3.6. Interpretasi

Tahap interpretasi ini menafsirkan pola yang ditemukan dari hasil penggalan data dengan teknik *clustering*. Dan memungkinkan visualisasi dimensi rendah dari hasil clustering ini disajikan dalam bentuk grafik. Sumbu X dan Y pada grafik menggambarkan jumlah penjualan. Dari hasil pengelompokan outlet dengan algoritma K-Medoids maka terbentuk 3 kelompok outlet dari jumlah data outlet sebanyak 437. Jumlah outlet tersebut bersal dari dua regional. Pola yang ditemukan dari pengelompokan outlet dengan variable yang digunakan yaitu jumlah penjualan, maka dapat di tafsirkan outlet pada *cluster-1* merupakan outlet dengan tingkat penjualan yang dominan sedang, outlet pada *cluster-2* merupakan outlet dengan tingkat penjualan yang dominan tinggi dan outlet pada *cluster-3* merupakan outlet dengan tingkat penjualan yang dominan rendah, Gambar 2 menunjukkan grafik pada sistem klasterisasi yang telah dibangun.



Gambar 2. Grafik

#### 4. KESIMPULAN

Hasil dari penggunaa teknik *Clustering* dengan algoritma K-Medoids, menunjukkan hasil teknik *Clustering* dengan algoritma K-Medoids mampu melakukan pengelompokan data terhadap data penjualan yang menghasilkan pengelompokan outlet. Dengan hasil pengujian pada *cluster* dengan beberapa nilai K didapatkan jumlah K=3 memiliki nilai Silhoutte Coefficient yang paling tinggi dan mendekasi Si= 1 yaitu dengan nilai 0,92669616303752 yang merupakan struktur kuat menurut Kaufman dan Rousseeuw. Sehingga pola pengelompokan outlet yang terbentuk dapat menjadi pengetahuan baru yang berguna bagi perusahaan dalam menentukan strategi penjualan.

#### DAFTAR RUJUKAN

##### Rujukan Jurnal:

Ribeiro, A., & Seruca, I. (n.d.). *Farmacêutica : Uma Aproximação Baseada Em Data Company : a Data Mining Based Approach*.

SULISTIANI, H. (2016). *Tesis – ks142501 pemilihan fitur untuk klasifikasi loyalitas pelanggan terhadap merek produk*.

##### Rujukan Jurnal:

Baihaqi, W. M., Indartono, K., & Banat, S. (2019). Penerapan Teknik Clustering Sebagai Strategi Pemasaran pada Penjualan Buku Di Tokopedia dan Shopee. *Paradigma - Jurnal Komputer Dan Informatika*, 21(2), 243–248. <https://doi.org/10.31294/p.v21i2.6149>

Djamaludin, I., & Nursikuwagus, A. (2017). Analisis Pola Pembelian Konsumen Pada Transaksi Penjualan Menggunakan Algoritma Apriori. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 8(2), 671.

Yahya, Y., & Puspita Hidayanti, W. (2020). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada "Lombok Vape On." *Infotek : Jurnal Informatika Dan Teknologi*, 3(2), 104–114. <https://doi.org/10.29408/jit.v3i2.2279>.

##### Rujukan Prosiding:

Agarwal, S. (2014). S. Agarwal, Data mining: Data mining concepts and techniques. 2014. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>

Arora, P., Deepali, & Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm for Big Data. *Physics Procedia*, 78(December 2015), 507–512. <https://doi.org/10.1016/j.procs.2016.02.095>

Arora, P., & Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm For Big Data. *Procedia - Procedia Computer Science*, 78(December 2015), 507–512. <https://doi.org/10.1016/j.procs.2016.02.095>

Chiu, C. Y., Chen, Y. F., Kuo, I. T., & Ku, H. C. (2009). An intelligent market segmentation system using k-means and particle swarm optimization. *Expert Systems with Applications*, 36(3 PART 1), 4558–4565. <https://doi.org/10.1016/j.eswa.2008.05.029>

Deker, G. W., Pechenizkiy, M., Vleeshouwers, J. M. (2009). Predicting Students Drop Out. *The 2nd International Conference on Educational Data Mining*, 1(1), 44–56. <https://eric.ed.gov/?id=ED539082%0Ahttp://www.win.tue.nl/~mpechen/research/edu.html>.

- Fu, Y. (1997). Data mining Theories, Algorithms, and Examples [Ye 2013-07-26]. In *IEEE Potentials* (Vol. 16, Issue 4). <https://doi.org/10.1109/45.624335>
- Gürbüz, F., Özbakir, L., & Yapici, H. (2011). Data mining and preprocessing application on component reports of an airline company in Turkey. *Expert Systems with Applications*, *38*(6), 6618–6626. <https://doi.org/10.1016/j.eswa.2010.11.076>
- Larose, D. T., & Larose, C. D. (2014). Discovering Knowledge in Data. In *Discovering Knowledge in Data*. <https://doi.org/10.1002/9781118874059>
- Motlagh, O., Berry, A., & O'Neil, L. (2019). Clustering of residential electricity customers using load time series. *Applied Energy*, *237*(December 2018), 11–24. <https://doi.org/10.1016/j.apenergy.2018.12.063>
- Prakasawati, P. E., Chrisnanto, Y. H., & Hadiana, A. I. (2019). Segmentasi Pelanggan Berdasarkan Produk Menggunakan Metode K- Medoids. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, *3*(1), 335–339. <https://doi.org/10.30865/komik.v3i1.1610>