

Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient

Riva Arsyad Farissa¹, Rini Mayasari², Yuyun Umidah³

^{1,2,3}Prodi Teknik Informatika Universitas Singaperbangsa Karawang

riva.arsyad17184@student.unsika.ac.id¹, rini.mayasari@staff.unsika.ac.id², yuyun.umidah@staff.unsika.ac.id³

Article Info

Article history:

Received 2021-08-07

Revised 2021-09-20

Accepted 2021-09-25

Keyword:

K-Means,
K-Medoids,
Silhouette Coefficient.

ABSTRACT

Perencanaan kebutuhan obat yang tidak efektif dan efisien menyebabkan masalah tentang tidak meratanya distribusi obat-obatan di setiap puskesmas. Dengan penggunaan *data mining* kebutuhan obat-obatan dapat dikendalikan agar tidak terjadi penumpukan stok serta kehabisan stok obat. Teknik *clustering* ini sangat cocok digunakan dalam perencanaan kebutuhan obat-obatan agar diketahui stok obat yang stabil dan tidak stabil. Metode yang akan digunakan untuk *clustering* data obat-obatan adalah algoritma *K-Means* dan *K-Medoids*. Tujuan dari penelitian ini adalah untuk mengelompokkan data obat-obatan di Puskesmas Karangsambung yang dapat digunakan sebagai referensi untuk perencanaan obat yang akan datang di puskesmas tersebut. Pengelompokan data dibagi menjadi tiga yaitu lambat, sedang dan cepat. Hasil yang didapatkan yaitu kedua algoritma tersebut menunjukkan bahwa algoritma *K-Means* mendapatkan hasil *Silhouette Coefficient* lebih tinggi yaitu sebesar 0,627 sedangkan *K-Medoids* sebesar 0,536 yang artinya hasil *clustering K-Means* lebih berkualitas dibandingkan *K-Medoids*. Dengan hasil dari pengelompokan obat tersebut puskesmas dapat melakukan pemerataan stok obat agar tidak terdapat kelebihan dan kekurangan stok obat.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Puskesmas merupakan unit pelaksana fungsional yang berperan sebagai pusat pembangunan kesehatan, pusat partisipasi masyarakat bidang kesehatan dan pusat pelayanan kesehatan primer yang menyelenggarakan kegiatan secara komprehensif, terpadu, dan berkelanjutan di suatu wilayah tertentu. Dalam proses pengadaan, bagian obat melakukan pengecekan deposit puskesmas, kemudian melihat tingkat fungsionalnya dengan sisa stok, tingkat kebutuhan dan kasus yang sering ditemukan. Kemudian petugas obat meminta barang-barang yang harus di stok ulang, untuk diteruskan ke bagian gudang obat. Sehubungan dengan semakin meningkatnya aktivitas medis khususnya di bagian obat-obatan, ditemukan beberapa permasalahan. Permasalahan yang ditemukan adalah tidak terkendalinya stok obat-obatan. Pengendalian obat adalah kegiatan untuk memastikan tercapainya tujuan yang diharapkan sesuai dengan strategi dan rencana yang telah dirumuskan, serta memastikan tidak terjadi kelebihan dan kekurangan obat [1]. Jika cara

pengelolaan obat dilakukan secara efektif, efisien dan rasional maka ketersediaan dan keterjangkauan obat akan terjamin [2].

Oleh sebab itu, setiap puskesmas harus memiliki sistem pengolahan data yang baik agar data-data yang dihasilkan dari penggunaan obat tersebut dapat berguna untuk dibuat menjadi sebuah laporan bulanan atau tahunan. Data ini tidak hanya digunakan sebagai arsip, tetapi juga dapat digunakan dan diolah menjadi informasi yang berguna untuk meningkatkan pelayanan kesehatan. Dengan dilakukannya pengelompokan obat maka dapat dilakukan pemerataan stok obat-obatan di setiap puskesmas.

Penerapan *data mining* di bidang kesehatan banyak dikemukakan peneliti, karena memiliki kemampuan mengekstrak data dalam jumlah besar untuk memperoleh informasi yang berguna [3]. Teknik *data mining* dapat digunakan secara luas di rumah sakit, klinik dan apotek oleh penyedia layanan kesehatan untuk memberikan layanan kesehatan yang lebih baik dan lebih terjangkau kepada pasien [4]. Penggunaan *data mining* ini dapat mengendalikan stok obat agar tidak terjadi penumpukan stok serta kehabisan stok

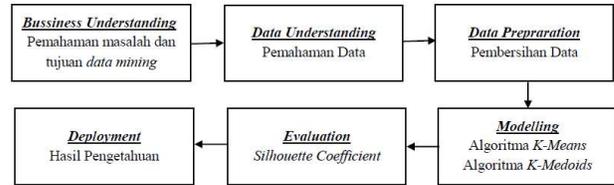
obat. *Clustering* adalah teknik pengelompokan *record* dalam database berdasarkan kondisi tertentu. Konsep dasar *clustering* adalah mengelompokkan beberapa objek ke dalam *cluster*, dimana *cluster* yang baik adalah *cluster* yang memiliki tingkat kemiripan yang tinggi antara objek-objek dalam *cluster* tersebut dan tingkat ketidaksamaan yang tinggi dengan objek *cluster* lainnya. Alasan penggunaan algoritma *K-means* dan *K-medoids clustering* ini dikarenakan algoritma ini memiliki akurasi yang cukup tinggi dan efisien dalam memproses obyek dalam jumlah besar. Selain itu kelebihan metode *K-Means* dan *K-Medoids* mempunyai sifat yang fleksibel karena *user* dapat menentukan jumlah *cluster* yang akan dibuat [5].

Penelitian tentang perbandingan clustering *k-means* dan *hierarchical clustering single linkage* pada persediaan obat. Penelitian ini menganalisis tentang penggunaan obat sehingga dapat dilakukan perencanaan stok obat. Hasil dari penelitian ini didapatkan nilai evaluasi algoritma *K-means* sebesar 0.8014 dan *HCC Single* sebesar 0.8629 [6]. Penelitian lain tentang analisis *K-medoids Clustering* untuk mengelompokkan data imunisasi campak balita di Indonesia dengan 2 kali iterasi dengan hasil 3 *cluster* yakni *cluster* rendah sebanyak 2 provinsi *cluster* sedang sebanyak 30 provinsi dan *cluster* tinggi sebanyak 2 provinsi [7]. Penelitian lainnya yang berjudul *Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data* mengenai perbandingan algoritma *K-means* dan *K-medoids* menggunakan dataset tanaman iris hasil yang diperoleh *K-medoids* lebih baik daripada *K-means* untuk akurasi pengelompokan terhadap kelas-kelas yang sebenarnya sehingga nilai *K-means* adalah 88,7% dan untuk *K-medoids* yaitu 92% [8].

Berdasarkan latar belakang di atas, maka penelitian ini akan melakukan pengelompokan data obat berdasarkan jenis obat, pemasukan dan penggunaan obat. Dengan melakukan analisa data obat di Puskesmas Karangsambung, diharapkan Puskesmas Karangsambung dapat menentukan keputusan yang tepat dalam pengendalian stok obat. Berdasarkan dari akurasi dan nilai akhir yang baik dari metode yang digunakan pada penelitian sebelumnya, maka metode yang digunakan adalah metode *cluster analysis* menggunakan algoritma *K-means* dan *K-medoids*. Pemrosesan pada penelitian ini akan dilakukan dengan bantuan Bahasa pemrograman *Python*.

II. METODOLOGI PENELITIAN

Metode *data mining* yang digunakan dalam penelitian ini adalah *Cross-Industry Standard Process for Data mining* (CRISP-DM). CRISP-DM adalah *framework data mining* yang secara garis besar terdiri dari 6 alur proses yang dimulai dari pemahaman bisnis hingga penerapan pengetahuan [9]. Tahapan yang akan dilakukan dengan CRISP-DM seperti gambar 1 berikut:



Gambar 1. Rancangan Penelitian CRISP-DM

Berikut ini adalah penjelasan dari tahapan metode CRISP-DM, yaitu:

A. Business Understanding

Business Understanding adalah tahapan pendefinisian masalah dan objektif dari *data mining* yang akan dilakukan. Pada fase ini dibutuhkan pemahaman terkait masalah dan tujuan yang akan diselesaikan oleh data mining.

B. Data Understanding

Data Understanding adalah tahap pengumpulan koleksi data awal, dan melakukan proses pengenalan terhadap data tersebut dengan tujuan untuk lebih mengenal *nature* dari data yang akan dipakai

C. Data Preparation

Aktivitas-aktivitas yang dilakukan pada fase ini meliputi pemilihan data yang akan digunakan untuk proses modelling, proses *data preprocessing* seperti mengatasi *outlier*, *noisy data*, *inconsistent* dan *missing value* yang ada dan *data transformation*, melakukan normalisasi data. Hasil dari tahap ini akan digunakan untuk proses modelling.

D. Modelling

Tahap ini mengerjakan aktivitas terkait algoritma *data mining* yang akan digunakan. Berikut ini metode yang akan digunakan dalam tahap *modelling* yaitu:

1) K-Means

K-means adalah sebuah algoritma *data mining* unsupervised yang menggunakan centroid sebagai titik pusat dalam setiap cluster. *K-means* ialah algoritma yang tidak rumit, ukuran kesamaan berperan penting dalam proses clustering menggunakan *K-Means* [10]. Algoritma clustering *K-means* banyak digunakan karena algoritmanya yang sederhana dan konvergensi yang cepat. Namun, nilai *K* clustering perlu diberikan terlebih dahulu dan pilihan nilai *K* secara langsung mempengaruhi hasil konvergensi [11]. Algoritma *K-means* merupakan algoritma yang diwakili oleh *cluster* yaitu *centroid* atau nilai rata-rata sebagai titik pusat *cluster*.

2) K-Medoids

Algoritma PAM (*Partitioning Around Medoid*) merupakan algoritma yang diwakili oleh *cluster* yaitu medoid. Algoritma ini lebih populer dengan sebutan Algoritma *K-Medoids*. Algoritma *K-Medoids* lebih baik daripada *K-Means*, karena pada *K-Medoids* kita menemukan *k* sebagai representasi objek untuk meminimalkan jumlah pertidaksamaan objek data, sedangkan pada *K-Means* kita menggunakan jumlah jarak *Euclidean* pada objek data [12]. Jika dibandingkan dengan algoritma *K-Means* perbedaan terletak pada titik pusat cluster, *K-Means* menggunakan nilai rata-rata (mean) sebagai pusat *cluster* sedangkan *K-Medoids* menggunakan sebuah

objek sebagai pusat *cluster* dan objek tersebut harus mewakili masing-masing dari *cluster*.

E. Evaluation

Pada tahap ini evaluasi yang digunakan yaitu *Silhouette Coefficient* yang berfungsi untuk menguji kualitas dari kluster lalu ditentukan keputusan apakah hasil *data mining* akan digunakan atau tidak.

Silhouette Coefficient terletak pada rentang antara nilai -1 sampai 1. Rentang nilai tersebut digunakan untuk menunjukkan kedekatan kemiripan data yang dikelompokkan didalam suatu *cluster*. Jika nilai rata-rata *Silhouette Coefficient* mendekati nilai 1, maka cluster tersebut semakin baik. Sebaliknya jika nilai rata-rata *Silhouette Coefficient* mendekati -1, maka *cluster* tersebut tidak baik. Berikut kriteria pengukuran nilai *Silhouette Coefficient* [13] pada tabel I.

TABEL I
KRITERIA SILHOUETTE COEFFICIENT

<i>Silhouette Coefficient</i>	Kriteria Penilaian
$0.7 < SC \leq 1.0$	<i>Stronge Structure</i>
$0.5 < SC \leq 0.7$	<i>Medium Structure</i>
$0.25 < SC \leq 0.5$	<i>Weak Structure</i>
$SC \leq 0.25$	<i>No Structure</i>

F. Deployment

Deployment adalah tahapan penggunaan hasil dari proses *data mining*.

III. HASIL DAN PEMBAHASAN

Hasil penelitian yang dilakukan berupa hasil *clustering* melalui proses *CRISP-DM* dengan menggunakan algoritma *K-Means* dan *K-Medoids* yang didukung oleh *tools GoogleColab* dengan bahasa pemrograman *Python*.

A. Bussiness Understanding

Pada tahapan ini diperlukan pemahaman tentang pentingnya pemanfaatan data distribusi obat, agar dapat digunakan untuk mengetahui pengelompokan obat berdasarkan pengeluaran dan pemasukan obat yang akan diterapkan pada setiap puskesmas. Tujuan dalam penerapan *data mining* ini adalah agar setiap puskesmas mengetahui pengelompokan obat mana saja yang termasuk obat lambat habis, sedang dan cepat habis. Dengan mengetahui pengelompokan tersebut distribusi obat pada setiap puskesmas dapat terkendali.

B. Data Understanding

Data yang digunakan merupakan data Laporan Pemakaian dan Lembar Permintaan Obat (LPLPO) yang bersumber dari Puskesmas Karangsambung. Laporan Pemakaian dan Lembar Permintaan Obat (LPLPO) merupakan suatu pngelolaan terhadap obat yang pemakaian, distribusi, tingkatan stok, kebutuhan obat dibatasi dengan tujuan agar pemakaian yang ada dapat terkendali dengan baik.

Data yang akan digunakan sebanyak 374 data dari bulan januari 2019 sampai dengan desember 2020. Data yang digunakan menggunakan format *.xlsx* (Excel). Atribut yang terdapat dalam data tersebut terdiri dari 8 atribut yaitu nama

obat, stok awal, penerimaan, persediaan, pemakaian, stok akhir, stok optimal dan permintaan. Terlihat pada gambar 2.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 374 entries, 0 to 373
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   NAMA OBAT       374 non-null    object
1   STOK AWAL       374 non-null    int64
2   PENERIMAAN     374 non-null    int64
3   PERSEDIAAN     374 non-null    int64
4   PEMAKAIAN      374 non-null    int64
5   STOK AKHIR     374 non-null    int64
6   STOK OPTIMAL   374 non-null    float64
7   PERMINTAAN     374 non-null    float64
dtypes: float64(2), int64(5), object(1)
memory usage: 23.5+ KB
```

Gambar 2. Deskripsi Data

C. Data Preparation

Pada tahap ini hal pertama yang dilakukan adalah memilih atribut yang akan digunakan untuk proses *modelling*. Hal ini dilakukan karena atribut yang tidak berhubungan akan berpengaruh terhadap hasil dari *modelling*. Tahap ini dilakukan dengan cara menghapus atribut stok optimal dan permintaan, dikarenakan terdapat nilai yang tidak konsisten di dalam atribut tersebut. Berikut dari hasilnya dapat dilihat pada gambar 3.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 374 entries, 0 to 373
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   NO               374 non-null    int64
1   NAMA OBAT       374 non-null    object
2   STOK AWAL       374 non-null    int64
3   PENERIMAAN     374 non-null    int64
4   PERSEDIAAN     374 non-null    int64
5   PEMAKAIAN      374 non-null    int64
6   STOK AKHIR     374 non-null    int64
dtypes: int64(6), object(1)
memory usage: 20.6+ KB
```

Gambar 3. Hasil *Data Selection*

Setelah memilih atribut yang akan digunakan untuk proses *modelling* tahap selanjutnya yaitu mengatasi *missing value* seperti menghapus kolom yang hilang dan menghapus data yang tidak relevan seperti data yang pada semua atributnya bernilai "0". Hal ini dilakukan agar proses *modelling* mendapatkan hasil yang baik. Setelah proses ini dilakukan jumlah kolom berkurang dari 373 menjadi 198 kolom. Adapun hasil dari prosesnya dapat dilihat pada tabel II.

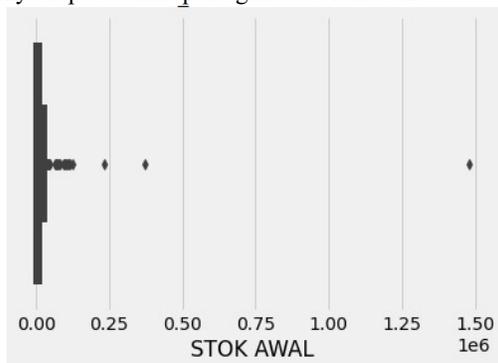
TABEL II
HASIL MENGATASI MISSING VALUE

No	Nama Obat	Stok Awal	Penerimaan	Persediaan	Pemakaian	Stok Akhir
0	Albendazol Tablet 400 Mg.	14500	6300	20800	12700	8100
1	Albendazol Suspens/5ml 200 Mg	0	568	568	0	568
2	Alopurinol Tablet 100 Mg.	11400	3400	14800	3000	11800
...
195	Lidokain Jelly 2% Syringe'10	120	30	150	30	120
196	Bisakodil Tab 5 Mg	2440	300	2740	90	2650
197	Omeprazole Sodium 40 Mg Iv Inj	860	100	960	96	864

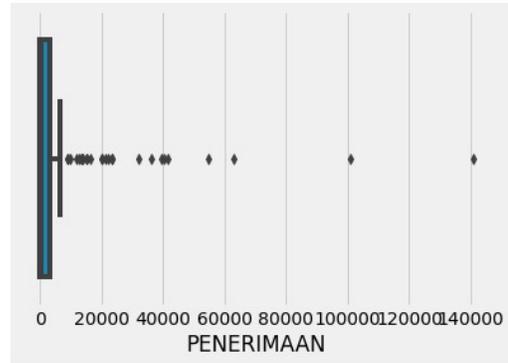
Tahap selanjutnya yaitu mengatasi *outlier* dengan cara menghapus *outlier* dengan perbandingan sebanyak 75 % data yang akan digunakan dan 25 % *outlier* dihapus dari data tersebut. Berikut merupakan cara mengatasi *outlier* dengan *python*.

```
for x in ['STOK AKHIR']:
    q75,q25 = np.percentile(X.loc[:,x],[75,25])
    intr_qr = q75-q25
    max = q75+(1.5*intr_qr)
    min = q25-(1.5*intr_qr)
    X.loc[X[x] < min,x] = np.nan
    X.loc[X[x] > max,x] = np.nan
X = X.dropna(axis = 0)
X.isnull().sum()
```

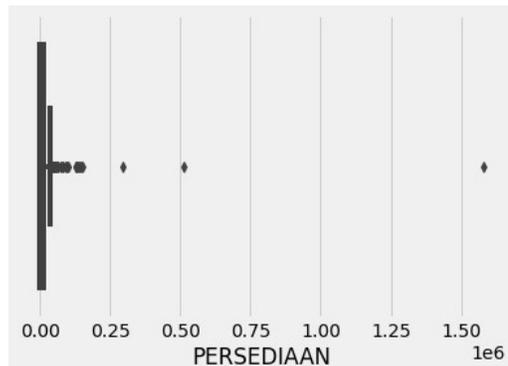
Hasilnya dapat terlihat pada gambar berikut ini.



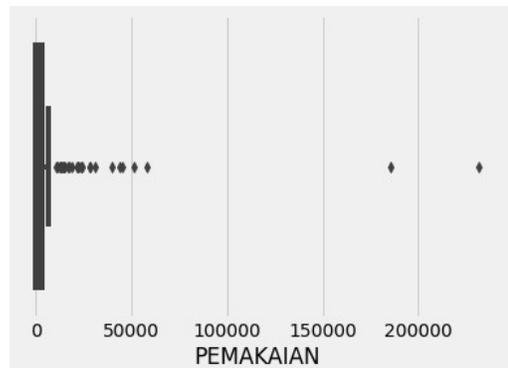
Gambar 4 Cek Outlier Atribut Stok Awal



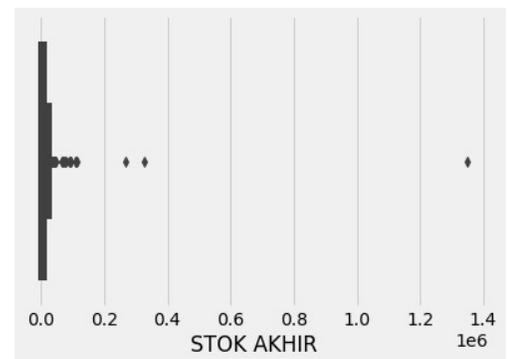
Gambar 5 Cek Outlier Atribut Penerimaan



Gambar 6 Cek Outlier Atribut Persediaan



Gambar 7 Cek Outlier Atribut Pemakaian



Gambar 8 Cek Outlier Atribut Stok Akhir

Penghapusan outlier diperlukan karena *outlier* dapat menimbulkan hasil bias. Selain itu, hal ini berfungsi untuk mengatasi algoritma *K-Means* yang sensitif terhadap *outlier*. Hasil setelah penghapusan *outlier* jumlah data berkurang menjadi 163 kolom, terlihat pada tabel III.

TABEL III
HASIL DATASET SETELAH MENGATASI OUTLIER

No	Stok Awal	Penerimaan	Persediaan	Pemakaian	Stok Akhir
0	0	568	568	0	568
1	11400	3400	14800	3000	11800
2	8827	1000	9827	42	9785
...
160	120	30	150	30	120
161	2440	300	2740	90	2650
162	860	100	960	96	864

Tahap selanjutnya yaitu transformasi data dengan cara melakukan normalisasi pada data. Proses normalisasi yang akan digunakan yaitu menggunakan cara *Min-Max*. *Min-Max* adalah metode normalisasi dengan melakukan transformasi linier terhadap data asli. Adapun hasil dari normalisasi data dapat dilihat pada tabel IV.

TABEL IV
HASIL NORMALISASI MIN-MAX

No	Stok Awal	Penerimaan	Persediaan	Pemakaian	Stok Akhir
0	0	0.08875	0.017197	0	0.020505
1	0.408602	0.53125	0.451153	0.441696	0.425993
2	0.31638	0.15625	0.299518	0.006184	0.353249
...
160	0.004301	0.004687	0.004452	0.004417	0.004332
161	0.087455	0.046875	0.083425	0.013251	0.095668
162	0.030824	0.015625	0.02915	0.014134	0.031191

D. Modelling

1) *K-Means*

Untuk dapat melakukan pengelompokan data dengan *K-Means* perlu menentukan jumlah *cluster* yang diperlukan. Data obat tersebut akan dikelompokkan menjadi tiga *cluster* yaitu Pemakaian Obat Lambat, Sedang dan Cepat atau $k=3$. Pengelompokan tiga *cluster* dilakukan agar diketahui obat mana saja yang stabil atau sedang dan tidak stabil. Jarak yang digunakan untuk menghitung jarak terpendek yaitu menggunakan *Euclidean distance*. Hasilnya terlihat sebagai berikut.

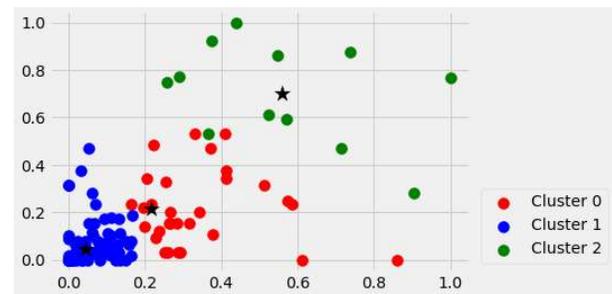
```
KMeans(algorithm='auto', copy_x=True, init='random',
max_iter=300, n_clusters=3, n_init=10, n_jobs=None,
precompute_distances='auto', random_state=42,
tol=0.0001, verbose=0)
```

Selanjutnya output dari program akan menampilkan *centroid* terakhir dan jumlah iterasi yang sudah di proses. Hasilnya seperti berikut ini.

Centroid Iterasi Terakhir :
 [0.340654 0.2169895 0.3320222 0.2197143 0.3392551]
 [0.044107 0.0476691 0.0467031 0.0375174 0.0462389]
 [0.560367 0.7024739 0.6136749 0.7612877 0.5400511]

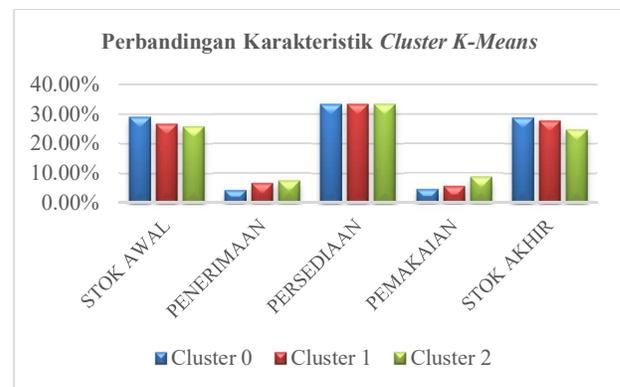
Jumlah Iterasi : 6

Hasil pengelompokan iterasi terakhir tersebut yaitu 30 obat pada *cluster* 0, 121 obat pada *cluster* 1 dan 12 obat pada *cluster* 2. Gambar 10 bawah ini merupakan hasil visualisasi *cluster* dengan menggunakan algoritma *K-Means*.



Gambar 9. Visualisasi *Cluster K-Means*

Selanjutnya melakukan analisis terhadap 3 *cluster* tersebut untuk penentuan label obat cepat habis, sedang dan lambat habis yang dilakukan berdasarkan karakteristik jumlah data stok obat dalam setiap *cluster*. Gambar 11 dan tabel V di bawah ini merupakan perbandingan karakteristik setiap *cluster* berdasarkan rata-rata dari keseluruhan atribut yang di tampilkan dalam bentuk persentase.



Gambar 10 Visualisasi Karakteristik *Cluster K-Means*

TABEL V
PERBANDINGAN KARAKTERISTIK CLUSTER K-MEANS

	Stok Awal	Penerimaan	Persediaan	Pemakaian	Stok Akhir
Cluster 0	29.09%	4.25%	33.34%	4.57%	28.76%
Cluster 1	26.71%	6.62%	33.33%	5.53%	27.80%
Cluster 2	25.89%	7.44%	33.33%	8.56%	24.77%

Dari perbandingan karakteristik tersebut, hal yang paling disoroti adalah perbedaan signifikan antara persentase rata-rata pemakaian dan stok akhir. Dalam penelitian ini pelabelan akan diambil dari selisih antara rata-rata pemakaian dengan stok akhir yang artinya semakin besar selisih maka pemakaian obat akan semakin lambat habis begitu pula sebaliknya semakin kecil selisih maka pemakaian obat akan semakin cepat habis. Berikut ini adalah hasil dari analisis pelabelan *clustering*.

- Data obat cluster 0 dapat dikategorikan ke dalam *cluster* pemakaian obat lambat lambat. Hal ini dibuktikan dengan selisih antara rata-rata pemakaian dengan stok akhir yaitu 24.19%.
- Data obat *cluster* 1 dapat dikategorikan pemakaian obat sedang. Hal ini dibuktikan dengan selisih antara rata-rata pemakaian dengan stok akhir yaitu 22.27%.
- Data obat *cluster* 2 dapat dikategorikan pemakaian obat cepat. Hal ini dibuktikan dengan selisih antara rata-rata pemakaian dengan stok akhir yaitu 16.21%.

2) K-Medoids

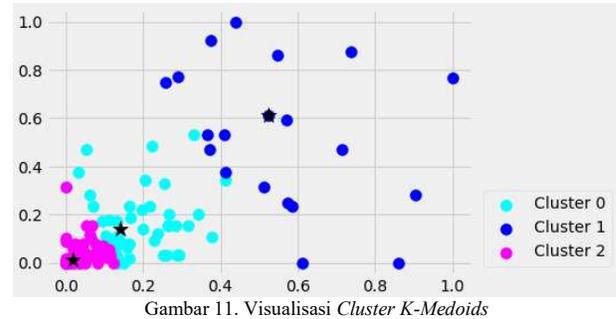
Untuk dapat melakukan pengelompokan data dengan *K-Means* perlu menentukan jumlah *cluster* yang diperlukan. Data obat tersebut akan dikelompokkan menjadi tiga *cluster* yaitu Pemakaian Obat Lambat, Sedang dan Cepat atau $k=3$. Pengelompokan tiga *cluster* dilakukan agar diketahui obat mana saja yang stabil atau sedang dan tidak stabil. Jarak yang digunakan untuk menghitung jarak terpendek yaitu menggunakan *Euclidean distance*. Hasilnya seperti berikut.

```
kmedoids(init='random', n_cluster=3, random_state=42
```

Selanjutnya output dari program akan menampilkan *centroid* terakhir dan jumlah iterasi yang sudah di proses. Hasilnya seperti berikut.

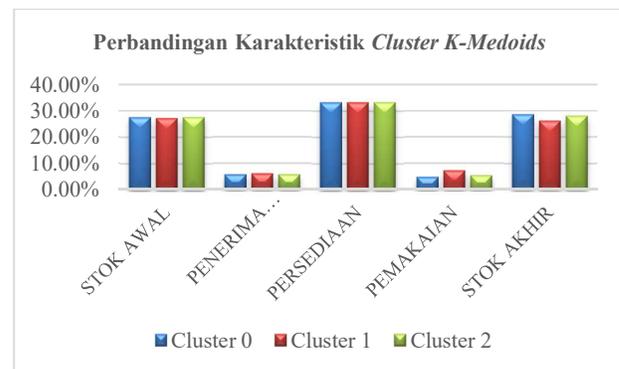
```
Medoid Iterasi Terakhir :
[0.1996415 0.140625 0.1971581 0.1236749 0.2032491]
[0.5232974 0.609375 0.5639712 0.6625441 0.5054151]
[0.0164874 0.014062 0.0166483 0.0103062 0.0173285]
Jumlah Iterasi : 3
```

Hasil pengelompokan iterasi terakhir tersebut yaitu 43 obat pada *cluster* 0, 20 obat pada *cluster* 1 dan 100 obat pada *cluster* 2. Gambar 14 merupakan hasil visualisasi *cluster* dengan menggunakan algoritma *K-Medoids*.



Gambar 11. Visualisasi Cluster K-Medoids

Selanjutnya melakukan analisis terhadap 3 *cluster* tersebut untuk penentuan label obat cepat habis, sedang dan lambat habis yang dilakukan berdasarkan karakteristik jumlah data stok obat dalam setiap *cluster*. Gambar 15 dan tabel VI di bawah ini merupakan perbandingan karakteristik setiap *cluster* berdasarkan rata-rata dari keseluruhan atribut yang di tampilkan dalam bentuk persentase.



Gambar 12 Visualisasi Cluster K-Medoids

TABEL VI
PERBANDINGAN KARAKTERISTIK CLUSTER K-MEDOIDS

	Stok Awal	Penerimaan	Persediaan	Pemakaian	Stok Akhir
Cluster 0	27.64%	5.70%	33.34%	4.82%	28.51%
Cluster 1	27.33%	6.01%	33.33%	7.22%	26.12%
Cluster 2	27.60%	5.73%	33.33%	5.35%	27.98%

Dari perbandingan karakteristik tersebut, hal yang paling disoroti adalah perbedaan signifikan antara persentase rata-rata pemakaian dan stok akhir. Dalam penelitian ini pelabelan akan diambil dari selisih antara rata-rata pemakaian dengan stok akhir yang artinya semakin besar selisih maka pemakaian obat akan semakin lambat habis begitu pula sebaliknya semakin kecil selisih maka pemakaian obat akan semakin cepat habis. Berikut ini adalah hasil dari analisis pelabelan *clustering*.

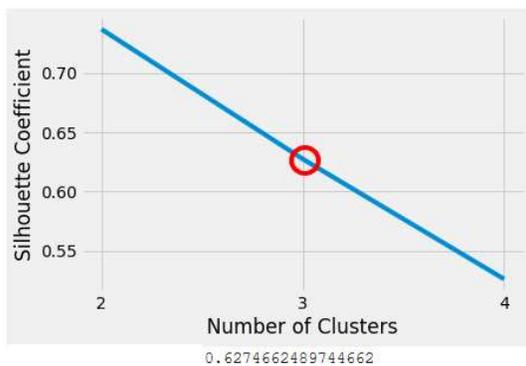
- Data obat cluster 0 dapat dikategorikan ke dalam *cluster* pemakaian obat lambat lambat. Hal ini dibuktikan

dengan selisih antara rata-rata pemakaian dengan stok akhir yaitu 23.69%.

- b. Data obat *cluster* 1 dapat dikategorikan pemakaian obat cepat. Hal ini dibuktikan dengan selisih antara rata-rata pemakaian dengan stok akhir yaitu 18.90%.
- c. Data obat *cluster* 2 dapat dikategorikan pemakaian obat sedang. Hal ini dibuktikan dengan selisih antara rata-rata pemakaian dengan stok akhir yaitu 22.63%.

E. Evaluation

Evaluation berguna untuk mengukur kualitas hasil *clustering*. Teknik yang digunakan yaitu *Silhouette Coefficient*. Proses evaluasi *Silhouette* dilakukan sesuai jumlah *K* dan setelah hasil perhitungan dari proses *K-Means* dan *K-Medoids* selesai. Berikut ini merupakan hasil dari perhitungan kualitas *cluster* dengan menggunakan nilai $k = 3$ dapat dilihat pada gambar 16.



Gambar 13. Visualisasi *Silhouette Coefficient K-Means*

Berdasarkan gambar diatas dapat dilihat bahwa hasil akurasi menggunakan metode *K-Means* dengan nilai $k = 3$ didapatkan *Silhouette Index* sebesar 0,627 termasuk ke dalam kriteria *Medium Structure*. Untuk hasil algoritma *K-Medoids* dapat dilihat pada gambar 17.

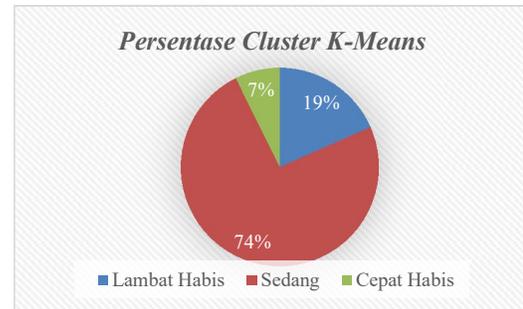


Gambar 14. Visualisasi *Silhouette Coefficient K-Medoids*

Berdasarkan gambar diatas hasil evaluasi dengan menggunakan metode *K-Medoids* didapatkan *Silhouette Index* sebesar 0,536 termasuk dalam kriteria *Medium Structure*.

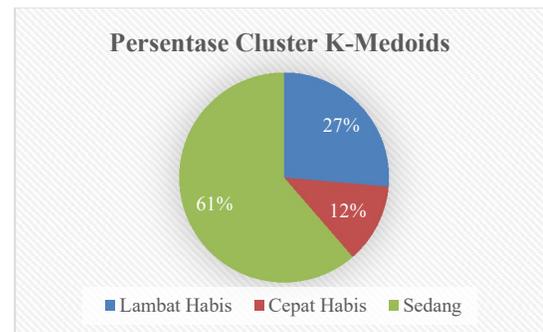
F. Deployment

Pada tahap ini dilakukan proses pembuatan laporan hasil penelitian sehingga lebih mudah dipahami oleh orang awam dan sebagai bahan kajian untuk puskesmas dalam proses pengendalian obat. Gambar 18 merupakan persentase hasil dari *cluster K-Means*.



Gambar 15. Hasil *Deployment K-Means*

Dari hasil *cluster K-Means* tersebut diketahui bahwa anggota data obat yang berada di *cluster* lambat habis sebanyak 19 % atau 30 obat, *cluster* sedang sebanyak 74 % atau 121 obat dan *cluster* cepat habis sebanyak 7 % atau 12 obat. Artinya penggunaan algoritma untuk *K-Means* didominasi oleh *cluster* sedang. Selanjutnya untuk algoritma *K-Medoids* dapat dilihat pada gambar 19.



Gambar 16. Hasil *Deployment K-Medoids*

Dari hasil *cluster K-Medoids* tersebut diketahui bahwa anggota data obat yang berada di *cluster* lambat habis sebanyak 27 % atau 43 obat, *cluster* sedang sebanyak 61 % atau 100 obat dan *cluster* cepat habis sebanyak 12 % atau 20 obat. Artinya penggunaan algoritma untuk *K-Medoids* didominasi oleh *cluster* sedang.

V. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan beberapa hal antara metode *K-Means* dan *K-Medoids* *clustering* telah berhasil diimplementasikan untuk mengelompokkan obat lambat habis, sedang, dan obat cepat habis. Penggunaan metode *K-Means* dan *K-Medoids* *clustering* ini menggunakan nilai $K = 3$. Pelabelan dilakukan berdasarkan karakteristik jumlah data stok obat dalam setiap *cluster*. Penelitian ini dibuat dengan menggunakan bahasa pemrograman *Python*. Dari perhitungan Algoritma *K-Means*

dan *K-Medoids* menghasilkan informasi baru berupa informasi pemakaian obat yang dibagi menjadi 3, yaitu

Pemakaian obat cepat yaitu pemakaian obat yang dimana pemakaiannya membuat persediaan yang tersedia hampir habis dari analisis *data mining* menggunakan metode *K-Means* terdapat 12 obat dan *K-Medoids* terdapat 20 obat. Hasil tersebut dapat menjadi acuan rekomendasi puskesmas untuk lebih memperbanyak stock obat khusus nya pada pemakaian obat cepat.

Pemakaian obat sedang yaitu pemakaian obat yang dimana persediaannya masih normal dari analisis *data mining* menggunakan metode *K-Means* terdapat 121 obat dan *K-Medoids* terdapat 100 obat.

Pemakaian obat lambat yaitu pemakaian obat yang dimana pemakaiannya membuat tidak terlalu tinggi dan cenderung jarang digunakan dari analisis *data mining* menggunakan metode *K-Means* terdapat 30 obat dan *K-Medoids* terdapat 43 obat. Hasil tersebut dapat menjadi acuan rekomendasi puskesmas untuk tidak melakukan persediaan pada obat-obat tersebut dikarenakan obat yang masih tersedia banyak.

Evaluasi algoritma *K-Means* dan *K-Medoids* untuk pengelompokan obat di puskesmas karangsambung dilakukan dengan menggunakan *Silhouette Coefficient* didapatkan hasil untuk *K-Means* sebesar 0.627 dan *K-Medoids* sebesar 0.536 sehingga dapat disimpulkan hasil *cluster* cukup baik dan masuk ke dalam kategori *medium structure*.

DAFTAR PUSTAKA

- [1] J. J. Tinangon, F. Ekonomi, and J. Akuntansi, "Analisis Pengendalian Internal Atas Persediaan Obat Pada Organisasi Sektor Publik Di Puskesmas Bahu," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 7, no. 3, pp. 3099–3108, 2019, doi: 10.35794/emba.v7i3.24062.
- [2] S. Chaira, E. Zaini, and T. Augia, "Drugs Management Evaluation at Community Health Centers in Pariaman City, Indonesia," *J. Sains Farm. Klin.*, vol. 3, no. 1, pp. 35–41, 2016, [Online]. Available: <http://jsfkonline.org/index.php/jsfk/article/view/97>.
- [3] M. Harahap, A. M. Husein, S. Aisyah, F. R. Lubis, and B. A. Wijaya, "Mining association rule based on the diseases population for recommendation of medicine need," *J. Phys. Conf. Ser.*, vol. 1007, no. 1, pp. 0–11, 2018, doi: 10.1088/1742-6596/1007/1/012017.
- [4] Z. Ceylan, S. Gürsev, and S. Bulkan, "An Application of Data Mining in Individual Pension Savings and Investment System," no. January, pp. 7–11, 2018.
- [5] M. Mardalius, "Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means," *Jurteksi*, vol. 4, no. 2, pp. 123–132, 2018, doi: 10.33330/jurteksi.v4i2.36.
- [6] R. D. Firdaus, T. G. Laksana, and R. D. Ramadhani, "Pengelompokan Data Persediaan Obat Menggunakan Perbandingan Metode K-Means Dengan Hierarchical Clustering Single Linkage," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 2, no. 1, pp. 33–48, 2019, doi: 10.20895/inista.v2i1.87.
- [7] S. Sundari, I. S. Damanik, A. P. Windarto, H. S. Tambunan, J. Jalaluddin, and A. Wanto, "Analisis K-Medoids Clustering Dalam Pengelompokan Data Imunisasi Campak Balita di Indonesia," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 687, 2019, doi: 10.30645/senaris.v1i0.75.
- [8] P. Kumar and D. Sirohi, "Comparative analysis of FCM and HCM algorithm on Iris data set," *Int. J. Comput. Appl.*, vol. 5, no. 2, pp. 33–37, 2010, doi: 10.5120/888-1261.
- [9] A. Azevedo and M. F. Santos, "Kdd , Semma And Crisp-Dm : A Parallel Overview Ana Azevedo and M . F . Santos," *IADIS Eur. Conf. Data Min.*, pp. 182–185, 2008, [Online]. Available: <http://recipp.ipp.pt/handle/10400.22/136%0Ahttp://recipp.ipp.pt/bitstream/10400.22/136/3/KDD-CRISP-SEMMA.pdf>.
- [10] M. Robani and A. Widodo, "Algoritma K-Means Clustering Untuk Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia," *J. Sist. Inf. Bisnis*, vol. 6, no. 2, p. 164, 2016, doi: 10.21456/vol6iss2pp164-176.
- [11] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J.*, vol. 2, no. 2, pp. 226–235, 2019, doi: 10.3390/j2020016.
- [12] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- [13] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.