

# Pengelompokan Data Persediaan Obat Menggunakan Perbandingan Metode *K-Means* Dengan *Hierarchical Clustering Single Linkage*

Rahmatika Diana Firdaus<sup>1</sup>, Tri Ginanjar Laksana<sup>2</sup>, Rima Dias Ramadani<sup>3</sup>

Institut Teknologi Telkom Purwokerto  
Jl. D.I Panjaitan No. 128 Purwokerto Selatan Jawa Tengah 53147

<sup>1</sup> 15102072@ittelkom-pwt.ac.id

<sup>2</sup> anjarlaksana@gmail.com

<sup>3</sup> rimadiaz@ittelkom-pwt.ac.id

\*Fakultas Teknologi Industri Dan Informatika

Accepted on 09-08-2019

## Abstract

Kesehatan merupakan hak asasi manusia sekaligus investasi bagi keberhasilan pembangunan bangsa Indonesia. Salah satu faktor penting di dunia kesehatan adalah tersedianya obat-obatan untuk nantinya disalurkan ke seluruh wilayah Indonesia melalui badan organisasi kesehatan milik pemerintah secara merata dan berkelanjutan. Fungsi obat yaitu sebagai upaya pencegahan, penyembuhan, maupun peningkatan kesehatan bagi manusia. Obat juga merupakan bahan yang diatur oleh pemerintah dalam hal ini adalah Badan Pengawasan Obat dan Makanan (BPOM). Di era modern seperti saat ini, kita mengenal dengan istilah *Data Mining*. Dalam perkembangannya, *data mining* berhubungan erat dengan analisa data, maka dari itu *data mining* mampu mengolah dan mengelompokan data dalam jumlah yang besar berdasarkan kesamaan dalam sekumpulan data. Algoritma *K-Means* merupakan metode pengelompokan yang mudah digunakan. Pada proses penentuan titik pusat klaster (*centroid*) awal merupakan kelemahan bagi *K-Means* karena sifatnya yang acak. Algoritma *Hierarchical Clustering (HCC) Single Linkage* pada penentuan titik pusat klaster (*centroid*) memiliki sifat yang konsisten dan kompleks. Dari 204 data dan variabel yang akan diolah, kedua algoritma tersebut akan mendapatkan klaster optimal data pada kelompok klaster C1 yaitu obat dengan pemakaian lambat dan klaster C2 yaitu obat dengan pemakaian cepat dan membandingkan nilai validitasnya. Hasil dari penelitian ini menunjukkan bahwa algoritma *HCC Single linkage* mampu memberikan hasil yang terbaik dengan validitas *Sillhouette Index (SI)* sebesar 0.8629 sedangkan algoritma *K-Means* mendapatkan nilai validitas *SI* sebesar 0.8414.

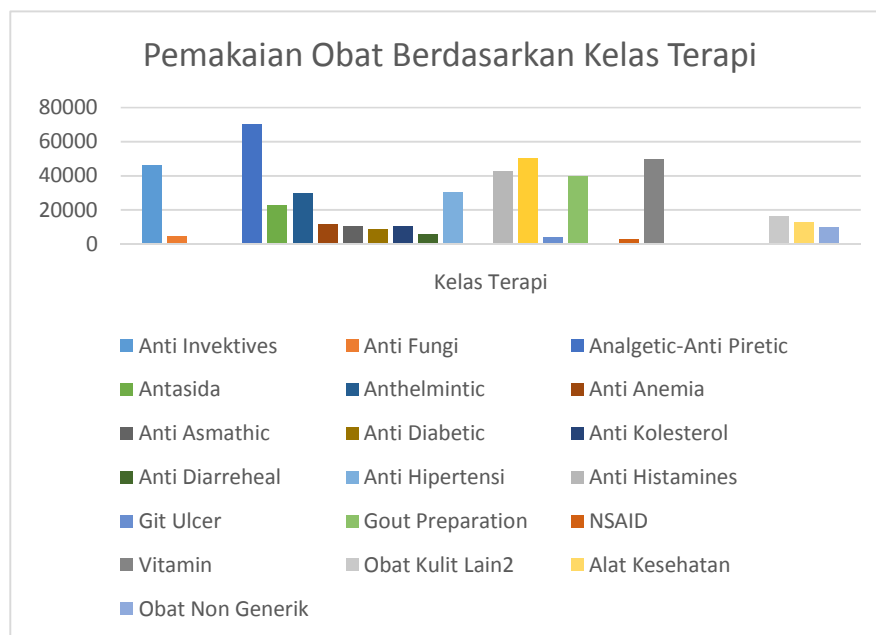
**Kata Kunci:** Algoritma, *Data Mining*, *Hierarchical Clustering*, *K-Means*, Obat, *Sillhouette Index*

## I. INTRODUCTION

**K**esehatan merupakan hak asasi manusia dan sekaligus merupakan investasi untuk keberhasilan pembangunan Bangsa Indonesia[1]. Maka dari itu, perlu diadakannya pembangunan kesehatan secara merata

dan *continue*. Salah satu unsur penting dalam penjaminan mutu kesehatan adalah dengan adanya ketersediaan obat agar mampu memenuhi kebutuhan kesehatan di masyarakat.

Obat merupakan salah satu komponen yang tak tergantikan dalam pelayanan kesehatan[2]. Fungsi obat yaitu upaya pencegahan, penyembuhan, maupun peningkatan kesehatan bagi manusia. Obat juga merupakan bahan yang diatur oleh pemerintah dalam hal ini adalah Badan Pengawasan Obat dan Makanan (BPOM). Tujuan dari adanya peraturan mengenai obat ini adalah untuk melindungi konsumen dari dampak merugikan karena kualitas dan keamanannya[3]. Berbagai macam obat yang dikonsumsi oleh masyarakat harus sesuai dengan standar BPOM dan menyesuaikan kelas terapi (sesuai keluhan, diagnosa dokter) agar tepat sasaran dan tepat guna.



Gambar 1. Jumlah Pemakaian Obat Tahun 2018

Berdasarkan Gambar 1 Merujuk ke Form Laporan Pemakaian Dan Lembar Permintaan Obat (LPLPO) Puskesmas II Ajibarang tahun 2018 tingkat pemakaian obat mencapai angka hampir 80.000 pada obat kelas terapi Analgetic-Antipiretic. Hal ini menandakan bahwa masyarakat tidak bisa lepas dari obat karena termasuk dalam kebutuhan sehari-hari dalam bidang kesehatan. Namun, sistem penanganan pengelompokan obat di Puskesmas II Ajibarang berdasarkan jenis obat yang memiliki varian yang sangat banyak masih dilakukan secara manual dan membutuhkan waktu yang lebih lama. Proses pengelompokan tersebut mengakibatkan terjadinya kesalahan dalam *output* yang diperoleh sehingga hasil pengelompokan menjadi kurang optimal. Besarnya jumlah pemakaian obat harus diimbangi dengan tersedianya pasokan obat dan tersalurkannya secara merata ke setiap wilayah. Upaya kesehatan dilakukan dengan pendekatan, pemeliharaan, peningkatan kesehatan, pencegahan penyakit, penyembuhan penyakit serta pemulihan kesehatan yang dilakukan secara menyeluruh, terpadu dan berkesinambungan[4]. Peraturan Menteri Kesehatan Nomor 75 Tahun 2014 menyebutkan bahwa Sistem Informasi Kesehatan wajib dilakukan oleh Dinas Kesehatan dan Puskesmas[5]. Berdasarkan form LPLPO Puskesmas II Ajibarang, data didalamnya akan di proses menggunakan algoritma *Hierarchical Clustering (HCC) Single Linkage* dengan *K-Means* untuk mendapatkan konvergen (mendekati titik *centroid*) klaster obat dengan pemakaian lambat (C1) dan obat dengan pemakaian cepat (C2). Sedangkan *K-Means* bekerja untuk mengelompokkan obat kedalam k kelompok (dimana k adalah jumlah kelompok yang diinginkan). Algoritma *HCC single linkage* bekerja untuk menentukan sekumpulan objek dalam hal ini adalah obat menjadi sebuah hierarki sehingga terbentuk kelompok yang memiliki kesamaan menggunakan jarak terdekat[6]. Setelah konvergen klaster dari kedua

algoritma didapatkan, maka akan dilakukan uji validitas kluster dari tiap-tiap algoritma untuk dibandingkan manakah yang mendapatkan validitas terbaik.

Penelitian yang relevan mengenai pemanfaatan metode *K-Means* pernah dilakukan oleh peneliti Joanna Ardhiyanti M.N, Yupie Kusumawati. Penelitian ini dilakukan untuk menganalisa pemakaian obat sehingga dapat dilakukan perencanaan dan pengendalian persediaan obat dengan mendapatkan jenis obat yang masuk dalam kluster pemakaian cepat dalam waktu 3 tahun terakhir. Hasil dari penelitian ini didapatkan 10 jenis obat pada tahun 2011, 5 jenis obat pada tahun 2012 dan 6 jenis obat pada tahun 2013 serta 2 jenis obat yaitu Klorfeniramin Maleat tb 4 mg, Vitamin B kompleks tablet pada gabungan ketiga tahun tersebut. Metode *clustering* dapat digunakan untuk mengendalikan persediaan obat di Puskesmas Pandanaran[1].

Penelitian lain mengenai analisa perbandingan metode *Hierarchical Clustering Single Linkage* dan *K-Means* serta gabungan keduanya pernah dilakukan oleh peneliti Tahta Alfina dkk. Latar belakang dari penelitian ini adalah untuk mengelompokkan sejumlah data atau objek ke dalam sebuah kluster dimana kluster tersebut akan berisi data semirip mungkin dan berbeda dengan objek dalam kluster lainnya, penggabungan keduanya dengan harapan mendapatkan hasil yang lebih baik. Hasil yang diperoleh adalah gabungan antara metode *K-Means* dan *Hierarchical Clustering Single Linkage* memberikan hasil yang terbaik berdasarkan *cluster variance* yang dibagi menjadi 5 kluster dan mendapatkan nilai *V* terkecil dan penjumlahan terbesar nilai *s* pada nilai *Sillhoutte*[7].

Selanjutnya Peneliti lain Rendy Handoyo dkk dalam membandingkan metode *clustering* menggunakan *Single Linkage* dengan *K-Means* dalam pengelompokan dokumen. *Clustering* bertujuan untuk mengelompokkan dokumen berita berdasarkan tingkat kemiripan dari dokumen tersebut. Hasil yang diperoleh adalah algoritma *Single Linkage* lebih optimal dalam pengelompokan dokumen dibandingkan *K-Means* dilihat dari nilai *Sillhoutte* dan *Purity* dimana algoritma *Single Linkage* selalu lebih tinggi dari *K-Means*. Kurang optimalnya algoritma *K-Means* adalah karena proses inialisasi kluster awal yang dilakukan secara acak. Meskipun dilakukan sebanyak 30 kali iterasi, hasil pengujian-pengujian tersebut belum tentu merepresentasikan suatu kluster yang baik[8].

Dari berbagai kekurangan dan kelebihan yang dimiliki algoritma *K-Means* dan *HCC single Linkage*, penulis ingin mengetahui hasil kluster yang didapat dari keduanya dan algoritma mana yang terbaik dilihat dari hasil validitas kluster untuk studi kasus persediaan obat di Puskesmas II Ajibarang. Oleh sebab itu, penelitian ini akan dituangkan ke dalam bentuk tugas akhir dan diberi judul "Pengelompokan Data Persediaan Obat Menggunakan Perbandingan Metode *K-Means* dan *Hierarchical Clustering Single Linkage* (Studi Kasus di Puskesmas II Ajibarang).

## II. LITERATURE REVIEW

### A. Penelitian Joanna Ardhiyanti M.N dan Yupie Kusumawati

Permasalahan yang diangkat dalam penelitian yang dilakukan oleh Joanna Ardhiyanti M.N, Yupie Kusumawati adalah penelitian belum adanya metode baku untuk menganalisa pemakaian obat yang bermanfaat sebagai perencanaan dan pengendalian persediaan obat. Tujuan dari penelitian ini adalah menganalisa pemakaian obat sehingga dapat dilakukan perencanaan dan pengendalian persediaan obat dengan mendapatkan jenis obat yang masuk dalam kluster pemakaian cepat dalam waktu 3 tahun terakhir. Alur penelitian yang digunakan oleh peneliti mengacu kepada tahapan *K-Means* pada umumnya. Penelitian tersebut menghasilkan beberapa jenis obat pada kluster pemakaian lambat dan cepat, namun tujuan dari penelitian ini adalah mendapatkan jenis obat hanya pada pemakaian cepat untuk dilakukan pengendalian persediaan obat pada tahun berikutnya. Kesimpulannya adalah didapatkan 2 jenis obat yang ada pada pemakaian cepat dalam gabungan 3 tahun. Dari penjelasan tersebut, diketahui bahwa metode *K-Means clustering* merupakan metode yang dapat digunakan untuk pengendalian persediaan obat di Puskesmas Pandanaran Semarang. Data yang digunakan adalah data persediaan obat di Puskesmas Pandanaran tahun 2011-2013 sebanyak 9627 data namun setelahnya dilakukan *pre-processing*. Atribut yang digunakan diantaranya nama obat, permintaan, pemakaian, stok awal, stok akhir, persediaan, penerimaan. Hasil dari penelitian ini didapatkan 10 jenis obat pada tahun 2011, 5 jenis obat pada tahun 2012 dan 6 jenis obat pada tahun 2013 serta 2 jenis obat yaitu Klorfeniramin

Maleat tb 4 mg, Vitamin B kompleks tablet pada gabungan ketiga tahun tersebut. Metode klastering dapat digunakan untuk mengendalikan persediaan obat di Puskesmas Pandanaran[1].

### B. Penelitian Tahta Alfina, dkk

Penelitian oleh Tahta Alfina, dkk pada tahun 2012 di latar belakang oleh mengelompokkan sejumlah data atau objek ke dalam sebuah kluster dimana kluster tersebut akan berisi data semirip mungkin dan berbeda dengan objek dalam kluster lainnya dan penggabungan keduanya dengan harapan mendapatkan hasil yang lebih baik. Penentuan *centroid* awal dilakukan oleh algoritma *Hierarchical* dan setelah didapat maka dilanjutkan dengan pemrosesan data menggunakan algoritma *K-Means*. Pada pengujian performansi algoritma menggunakan *Sillhoutte Index* menghasilkan algoritma gabungan *Hierarchical Clustering* dengan metode *single linkage* dengan *K-Means* optimal dengan nilai *Sillhoutte* terbesar yaitu 300 pada pembagian 5 kluster dan gabungan *Hierarchical Clustering* metode *centroid linkage* dengan *K-Means* optimal dengan nilai *Sillhoutte* terbesar dengan nilai 300 pada pembagian 7 kluster. Hasil yang diperoleh adalah gabungan antara metode *K-Means* dan *Hierarchical* klastering *single linkage* memberikan hasil yang terbaik berdasarkan kluster *variance* yang dibagi menjadi 5 kluster dan mendapatkan nilai *V* terkecil dan penjumlahan terbesar nilai *s* pada nilai *Sillhoutte*[7].

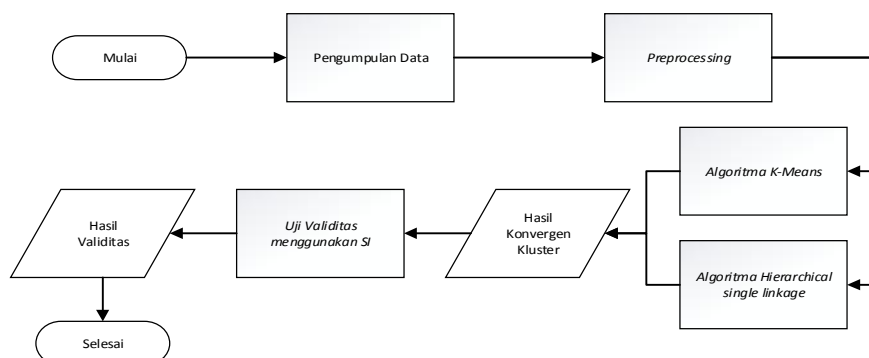
### C. Penelitian Rendy Handoyo, dkk

Menurut Penelitian oleh Rendy Handoyo, dkk yang memiliki tujuan untuk mengelompokkan dokumen berita berdasarkan tingkat kemiripan dari dokumen tersebut. Hasil yang diperoleh pada semua pengujian menggunakan uji validitas *Sillhoutte Index*, algoritma *HCC Single linkage* unggul dibanding *K-Means*. Pengujian tersebut berupa menguji performansi *Clustering* dengan dibagi 4 kategori yaitu 25 dokumen, 50 dokumen, 75 dokumen dan 100 dokumen. Pengujian metode *HCC Single linkage* lebih unggul dibandingkan dengan *K-Means* dengan mendapatkan nilai *Silhouette* masing-masing sebesar 0.92, 0.52, 0.33 dan 0.24. Kurang optimalnya algoritma *K-Means* adalah karena proses inialisasi kluster awal yang dilakukan secara acak. Meskipun dilakukan sebanyak 30 kali iterasi, hasil pengujian-pengujian tersebut belum tentu merepresentasikan suatu kluster yang baik[8].

## III. RESEARCH METHOD

Studi literatur pada penelitian ini adalah merujuk kepada penelitian sebelumnya yang berkaitan. Studi literatur dilakukan dengan membaca jurnal terkait untuk menentukan metode dan studi kasus yang akan digunakan serta cara penggunaan metode yang ada sebelumnya. Data yang digunakan adalah data form Laporan Pemakaian dan Laporan Permintaan Obat (LPLPO) tahun 2017-2018. Teknik yang digunakan adalah observasi dan wawancara.

Pada penelitian ini terdapat tahapan-tahapan dalam menyelesaikan masalah yang dituangkan dalam *flowchart* penelitian:



Gambar 3.2. 1 Alur Metode Penelitian

Berdasarkan *flowchart* penelitian pada gambar 1 dapat dijelaskan sebagai berikut:

### Mulai

#### Pengumpulan Data

Pengumpulan data yaitu menyiapkan data awal form LPLPO di Puskesmas II Ajibarang. Pada penelitian ini, pengumpulan data dilakukan dengan cara mendapatkan data dari Puskesmas II Ajibarang dan menggunakan wawancara. Data yang telah didapatkan berupa form Laporan Pemakaian dan Laporan Permintaan Obat (LPLPO) tahun 2017-2018 yang berformat .excel. Terdapat 642 data untuk tahun 2017-2018 tetapi yang digunakan sejumlah 204 karena banyak terdapat data kosong. Data form LPLPO terdapat 9 atribut diantaranya nama obat, satuan, stok awal, penerimaan, persediaan, pemakaian, sisa stok, stok optimal dan permintaan. Atribut yang akan dipakai dalam penelitian ini hanya sebanyak 7 karena apabila stok optimal dan penerimaan dipakai maka akan mempengaruhi hasil klaster.

Tabel 3. 1 Data Obat

Data set ke-i	ATRIBUT						
	Nama Obat	Satuan	Stok Awal	Penerimaan	Persediaan	Pemakaian	Sisa Stok
1	Amoksilin Syr Ker	Botol	1440.00	552.00	1992.00	1349.00	643.00
2	Amoksilin Kap 500 mg	Tablet	26900.00	29000.00	55900.00	33600.00	22300.00
3	Ampisilin Inj Im/Iv 1000mg/ml	Vial	30.00	30.00	60.00	30.00	30.00
Ke-n	...	...	...	...	...	...	...

Tabel 3.1 menjelaskan bahwa data yang akan digunakan memiliki 7 atribut. Tidak semua atribut yang ada pada tabel 3.1 digunakan. Pada proses selanjutnya, atribut yang digunakan berjumlah 5 diantaranya stok awal, penerimaan, persediaan, pemakaian dan sisa stok.

#### Preprocessing

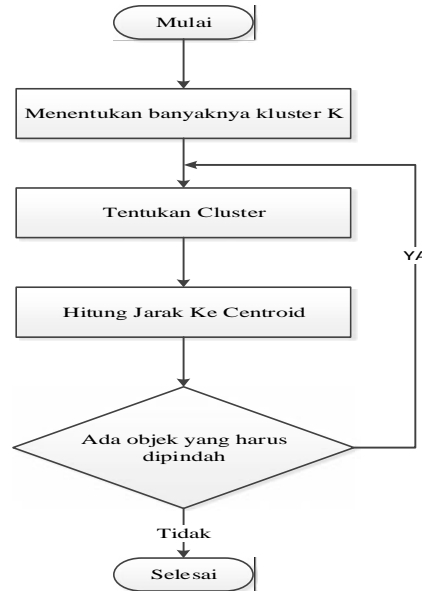
*Preprocessing* yaitu pemeriksaan data agar terhindar dari redundansi dan data inkonsisten, menghilangkan variabel yang tidak dapat dihitung. Tujuan dilakukannya *preprocessing* karena metode *clustering* tidak dapat mengolah data selain angka. Dari tujuh atribut, atribut yang akan diolah sebanyak lima yaitu stok awal, penerimaan, persediaan, Seperti pada tabel 3.2 dibawah ini:

Tabel 3. 2 Data Obat Setelah *Preprocessing*

Data set ke-i	ATRIBUT				
	Stok Awal	Penerimaan	Persediaan	Pemakaian	Sisa Stok
1	1440.00	552.00	1992.00	1349.00	643.00
2	26900.00	29000.00	55900.00	33600.00	22300.00
3	30.00	30.00	60.00	30.00	30.00
Ke-n	...	...	...	...	...

#### Implementasi algoritma *K-Means*

Setelah data siap diolah, proses algoritma *K-Means* tertuang dalam *flowchart* dibawah ini:



Gambar 3.2.2 Tahapan algoritma *K-Means*

Berdasarkan dari gambar 3.2.2 Tahapan algoritma *K-Means* dapat dijabarkan sebagai berikut:

a. Mulai

b. Menentukan banyaknya *cluster* k

Penentuan banyaknya klaster pada penelitian ini, menurut penelitian sebelumnya klaster dibagi menjadi dua yaitu obat dengan pemakaian lambat dan obat dengan pemakaian cepat (C1 dan C2) serta menyesuaikan kebutuhan.

c. Menentukan *cluster* (titik *centroid*)

Menurut peneliti Joanna Ardhiyanti M.N dan Yupie Kusumawati, penentuan *centroid* awal *K-Means* adalah secara acak. Maka dari itu, *centroid* awal pada penelitian ini diambil 2 data secara acak untuk klaster C1 dan C2 berturut-turut atribut masing-masing Antifungi Doen, 123, 24, 147, 55, 92 dan Paracetamol 500 Mg 11800, 54000, 65800, 54200, 11600

d. Menghitung jarak data ke *centroid*

Tahap selanjutnya adalah menghitung jarak data ke titik *centroid* dengan menggunakan *Euclidean Distance* atau jarak terdekat untuk mendapatkan hasil data masuk ke klaster C1 atau klaster C2.

Iterasi pertama jarak data ke *centroid* C1

(d1,c1)

$$= \sqrt{((1440-123)^2 + (552-24)^2 + (1992-147)^2 + (1349-55)^2 + (643-92)^2)}$$

$$= 2719.436523$$

(d2,c1)

$$= \sqrt{((26900-123)^2 + (29000-24)^2 + (55900-147)^2 + (33600-55)^2 + (22300-92)^2)}$$

$$= 79268.37707$$

(d3,c1)

$$= \sqrt{((30-123)^2 + (30-24)^2 + (60-147)^2 + (30-55)^2 + (30-92)^2)}$$

$$= 143.954854$$

(d4,c1)

$$= \sqrt{((11200-123)^2 + (6000-24)^2 + (17200-147)^2 + (5200-55)^2 + (12000-92)^2)}$$

$$\begin{aligned}
 &= 24849.32198 \\
 &(d5,c1) \\
 &= \sqrt{((607-123)^2+(400-24)^2+(1007-147)^2+(377-55)^2+(630-92)^2)} \\
 &= 1228.153085 \\
 &\text{(Seterusnya sampai n data).}
 \end{aligned}$$

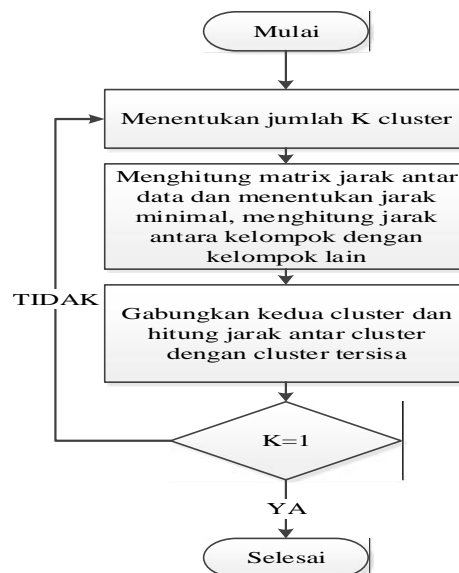
Iterasi pertama jarak data ke centroid C2

$$\begin{aligned}
 &(d1,c2) \\
 &= \sqrt{((1440-11800)^2+(552-54000)^2+(1992-65800)^2+(1349-54200)^2+(643-11600)^2)} \\
 &= 84214.13816 \\
 &(d2,c2) \\
 &= \sqrt{((26900-11800)^2+(29000-54000)^2+(55900-65800)^2+(33600-54200)^2+(22300-11600)^2)} \\
 &= 29408.24714 \\
 &(d3,c2) \\
 &= \sqrt{((30-11800)^2+(30-54000)^2+(60-65800)^2+(30-54200)^2+(30-11600)^2)} \\
 &= 86766.81583 \\
 &(d4,c2) \\
 &= \sqrt{((11200-11800)^2+(6000-54000)^2+(17200-65800)^2+(5200-54200)^2+(12000-11600)^2)} \\
 &= 24849.32198 \\
 &(d5,c2) \\
 &= \sqrt{((607-11800)^2+(400-54000)^2+(1007-65800)^2+(377-54200)^2+(630-11600)^2)} \\
 &= 1228.153085 \\
 &\text{(Seterusnya sampai n data).}
 \end{aligned}$$

- e. Jika masih ada objek atau data yang berpindah maka ulangi iterasi sampai tidak ada lagi yang berpindah.
- f. Selesai

### Implementasi algoritma HCC Single Linkage

Data sampel yang akan digunakan untuk perhitungan manual HCC Single Linkage adalah 20% dari populasi yaitu sebanyak 20 data. Proses algoritma HCC Single Linkage tertuang dalam flowchart dibawah ini:



Gambar 3.2. 3 Tahapan HCC Single Linkage

Berdasarkan *flowchart* penelitian pada gambar 1 dapat dijelaskan sebagai berikut:

- a. Mulai
- b. Menentukan jumlah *k cluster*  
 Klaster yang dibentuk pada penelitian ini berdasarkan kebutuhan sebanyak dua yaitu C1 untuk obat dengan pemakaian lambat dan C2 untuk obat dengan pemakaian cepat.
- c. Menghitung matriks jarak antar data dan menentukan jarak minimal, menghitung jarak antar kelompok dengan kelompok lain

Matriks Jarak antar data

$$(1,2) = \sqrt{(1440-26900)^2+(552-29000)^2+(1992-55900)^2+(1349-33600)^2+(643-22300)^2}$$

$$= 69719.66$$

$$(1,3) = \sqrt{(1440-30)^2+(552-30)^2+(1992-60)^2+(1349-30)^2+(643-30)^2}$$

$$= 2523.94$$

$$(1,4) = \sqrt{(1440-11200)^2+(552-6000)^2+(1992-17200)^2+(1349-5200)^2+(643-12.000)^2}$$

$$= 22027.24$$

$$(1,5) = \sqrt{(1440-607)^2+(552-400)^2+(1992+1007)^2+(1349-377)^2+(643-630)^2}$$

$$= 1299.37$$

$$(2,3) = \sqrt{(26900-30)^2+(29000-30)^2+(55900-60)^2+(33600-30)^2+(22300-30)^2}$$

$$= 79386.78$$

$$(2,4) = \sqrt{(26900-11200)^2+(29000-6000)^2+(55900-17200)^2+(33600-5200)^2+(22300-12000)^2}$$

$$= 56443.16$$

$$(2,5) = \sqrt{(26900-607)^2+(29000-400)^2+(55900-1007)^2+(33600-377)^2+(22300-630)^2}$$

$$= 78076.12$$

$$(3,4) = \sqrt{(30-11200)^2+(30-6000)^2+(60-17200)^2+(30-5200)^2+(30-12000)^2}$$

$$= 24983.98$$

$$(3,5) = \sqrt{(30-607)^2+(30-400)^2+(60-1007)^2+(30-377)^2+(30-630)^2}$$

$$= 1359.06$$

$$(4,5) = \sqrt{(11200-607)^2+(6000-400)^2+(17200-1007)^2+(5200-377)^2+(12000-630)^2}$$

$$= 23628.86$$

Seterusnya sampai matriks data (n,n)

- d. Menggabungkan kedua klaster dan hitung jarak antar klaster dengan klaster tersisa. Penggabungan klaster diambil dari nilai kelompok paling minimum pada matriks (*single linkage*) sehingga tersisa hanya satu klaster tunggal (K=1).
- e. Selesai

### Hasil klaster optimal kedua algoritma

Hasil klaster optimal akan didapatkan dari dari algoritma *K-Means* dan *HCC Single Linkage*.

#### Uji validitas

Uji validitas klaster optimal bertujuan untuk menghitung nilai kesesuaian dari klaster yang telah didapat. Pada penelitian ini menggunakan validitas *Sillhoutte Index*. *Centroid* yang digunakan diambil dari baris pertama klaster C1 dan klaster C2

#### *Sillhoutte Index K-Means:*

Jarak *centroid* C1 ke antar data C1

(cd1, d1)

$$= \sqrt{((3307-3307)^2+(6-6)^2)}$$

$$= 0$$

(cd1, d2)



$$= \sqrt{((3307-7353)^2+(6-3)^2)}$$

$$= 4046.01112$$

(cd1, d3)

$$= \sqrt{((3307-5892)^2+(6-4)^2)}$$

$$= 2585.000774$$

Seterusnya sampai data ke- n.

Jarak centroid C2 ke antar data C2

(cd2, d1)

$$= \sqrt{((6529-6529)^2+(9-9)^2)}$$

$$= 0$$

(cd2, d2)

$$= \sqrt{((6529-1328)^2+(9-4)^2)}$$

$$= 5201.002403$$

(cd2, d3)

$$= \sqrt{((6529-6798)^2+(9-4)^2)}$$

$$= 269.0464644$$

Seterusnya sampai data ke- n.

Jarak centroid C1 ke antar data C2

(c1,1)

$$= \sqrt{((3307-6529)^2+(6-9)^2)}$$

$$= 3222.001397$$

(c1,2)

$$= \sqrt{((3307-1328)^2+(6-4)^2)}$$

$$= 1979.001011$$

(c1,3)

$$= \sqrt{((3307-6798)^2+(6-4)^2)}$$

$$= 3491.000573$$

Seterusnya sampai data ke- n.

Jarak centroid C2 ke antar data C1

(c2,1)

$$= \sqrt{((6529-3307)^2+(9-6)^2)}$$

$$= 3222.001.397$$

(c2,2)

$$= \sqrt{((6529-7353)^2+(9-3)^2)}$$

$$= 824.0218444$$

(c1,1)

$$= \sqrt{((6529-5892)^2+(9-4)^2)}$$

$$= 637.0196229$$

Seterusnya sampai data ke-n

Menjumlahkan, mencari rata-rata dari nilai perhitungan yang diperoleh.

Hasil yang telah diperoleh dari perhitungan sebelumnya, selanjutnya dijumlahkan dan dihitung rata-ratanya. Perhitungan diambil per kolom yang hasilnya akan diinisialisasi menjadi a1, a2, b1 dan b2. Kemudian *Sillhoutte Index* akhir menyesuaikan dengan persamaan *Sillhoutte Index*.

***Sillhoutte Index HCC Single Linkage:***

Sama seperti perhitungan *Sillhoutte Index* pada *K-Means*, langkah pertama adalah dengan menghitung jarak masing-masing *centroid* ke data 3,10 atau data 8,18,5,...n kemudian menghitung jarak *centroid* data 3,10 ke data 8,18,5,...n dan sebaliknya.

Menghitung jarak *centroid* C1 ke data C2

(cd1,d1)

$$= \sqrt{((143-143)^2+(9-9)^2)}$$

$$= 0$$

(cd1,d2)

$$= \sqrt{((143-187)^2+(9-8)^2)}$$

$$= 44.01136217$$

(cd1,d3)

$$= \sqrt{((143-194)^2+(9-9)^2)}$$

$$= 51$$

Seterusnya sampai data ke-n

Menghitung jarak *centroid* C2 ke data C1

(cd2,d1)

$$= \sqrt{((3859-3859)^2+(8-8)^2)}$$

$$= 0$$

(cd2,d2)

$$= \sqrt{((3859-5644)^2+(8-1)^2)}$$

$$= 1785.013725$$

Seterusnya sampai data ke-n.

Menghitung *centroid* C1 ke data C2

(c1,1)

$$= \sqrt{((143-3859)^2+(9-8)^2)}$$

$$= 3716.000135$$

(c1,2)

$$= \sqrt{((143-5644)^2+(9-1)^2)}$$

$$= 5501.005817$$

Menghitung *centroid* C2 ke data C1

(c2,1)

$$= \sqrt{((3859-143)^2+(8-9)^2)}$$

$$= 3716.000135$$

(c2,2)

$$= \sqrt{(3859-187)^2+(8-8)^2}$$

$$= 3672$$

(c2,3)

$$= \sqrt{(3859-194)^2+(8-9)^2}$$

$$= 3665.000136$$

Seterusnya sampai data ke-n

Menjumlahkan, mencari rata-rata dari nilai perhitungan yang diperoleh.

Hasil yang telah diperoleh dari perhitungan sebelumnya, selanjutnya dijumlahkan dan dihitung rata-ratanya. Perhitungan diambil per kolom yang hasilnya akan diinisialisasi menjadi a1, a2, b1 dan b2. Kemudian perhitungan *Sillhoutte Index* akhir menyesuaikan dengan persamaan *Sillhoutte Index*.

### Hasil Validitas

Hasil validitas menunjukkan bahwa algoritma *K-Means* mendapatkan nilai validitas sebesar 0.8414 dan algoritma *HCC Single Linkage* memiliki nilai validitas lebih tinggi dibandingkan *K-Means* yaitu sebesar 0.8629.

Selesai.

## IV. RESULT AND DISCUSSION

### Hasil Penelitian

Pada sub bab hasil penelitian ini, akan dijelaskan tahapan pengumpulan data, *preprocessing*, hasil perhitungan manual algoritma *K-Means*, hasil perhitungan manual *HCC Single Linkage*, hasil kluster optimal, dan penentuan algoritma kluster terbaik berdasarkan uji validitas.

### Pengumpulan Data

Tahap pertama dalam penelitian ini adalah melakukan pengumpulan data. Pengambilan data dilakukan di Puskesmas II Ajibarang berupa data Laporan Pemakaian dan Laporan Permintaan Obat (LPLPO) tahun 2017-2018 yang bersifat *private*. Di dalam form LPLPO terdapat 11 atribut diantaranya nama obat, satuan, stok awal, penerimaan, persediaan, pemakaian, sisa stok, stok optimal, permintaan pemberian, nomor *batch* dan *end* serta keterangan.

LAPORAN PEMAKAIAN DAN LEMBAR PERMINTAAN OBAT (LPLPO)												
PUSKESMAS/BALAI KESEHATAN												
PUSKESMAS : II AJIBARANG		PERIODE BULAN/PERIODE : SEPTEMBER					DOKUMEN / NOMOR/KODE		TANGGAL			
KECAMATAN : AJIBARANG		PERMINTAAN BULAN/PERIODE: Bulan Oktober, November dan					/ ITRR/ITLA					
KAB/KODYA : Banyuwangi		Desember tahun 2017										
PUSKESMAS II AJIBARANG												
DAFTAR DISUSUN BERDASARKAN KELAS TERAPI												
1	2	3	4	5	6	7	8	9	10	11	12	13
			STOK AWAL	PENERIMAAN	PERSEDIaan	PEMAKAIAN	SISA STOK	STOK OPTIMAL	PERMINTAAN	PEMBERIAN	NOMOR BATCH DAN CC	KETERANGAN
01	ANTI INFEKTIFES											
1	AMOKSILIN SYR KER	BOTOL	295	152	453	407	81	427	350			
2	AMOKSILIN KAP 500 MG	TABLET	5400	7000	13400	11400	2000	11400	13000			
3	AMOKSILIN TAB	TABLET	0	-	-	-	0	0	0			
4	AMPISILIN INJ IM/IV 1000MG/ML	VIAL	10	10	10	10	0	10	10			
5	GENTAMIN DEKUIL 2,4 JUTA IUMAL	VIAL	0	-	-	-	0	0	0			
6	GENTAMIN 300 MG	TABLET	3500	3500	300	2900	300	-	-			
7	DIDICLON 100 MG	CAPSUL	0	-	-	-	0	0	0			
8	ERYTHROMISIN 300 MG/5ML	BOTOL	0	-	-	-	0	0	0			
9	ERYTHROMISIN 500 MG	TABLET	0	-	-	-	0	0	0			
10	GENTAMIN INJEKSI 800 MG/2ML	AMPUL	0	-	-	-	0	0	0			

Gambar 4. 1 Dataset form LPLPO tahun 2017

Gambar 4.1 merupakan *dataset* form LPLPO tahun 2017 yang berisi tentang nama obat, satuan, stok awal, penerimaan, persediaan, pemakaian, sisa stok, stok optimal, permintaan, pemberian, nomor *batch* dan *end* serta keterangan. Pada penelitian ini atribut yang digunakan sebanyak 7 yaitu nama obat, satuan, stok awal, penerimaan, persediaan, pemakaian dan sisa stok. Penggunaan 7 atribut ini dikarenakan terdapat banyak data yang memiliki *missing value* yang mempengaruhi atribut stok optimal dan atribut penerimaan. Kemudian pada atribut pemberian, nomor *batch* dan *end* serta keterangan adalah kolom kosong. Algoritma *clustering* hanya dapat memproses data yang berbentuk angka, maka dari itu atribut yang akan diperkecil kembali menjadi 5 berupa stok awal, penerimaan, persediaan, pemakaian, sisa stok dan nama obat hanya sebagai *id*.

Tabel 4. 1 Dataset form LPLPO tahun 2017-2018

No	NAMA OBAT	STOK AWAL	PENERIMAAN	PERSEDIaan	PEMAKAIAN	SISA STOK
1.	Amoksilin Syr Ker	1440.00	552.00	1992.00	1349.00	643.00
2.	Amoksilin Kap 500 Mg	26900.00	29000.00	55900.00	33600.00	22300.00
3.	Ampisilin Inj IM/IV 1000 Mg/ML	30.00	30.00	60.00	30.00	30.00

4.	Ciprofloxacin 500 MG	11200.00	6000.00	17200.00	5200.00	12000.00
5.	Kotrimoksazol Suspensi 60 MI	607.00	400.00	1007.00	377.00	630.00
...	...	...	...	...	...	...
642.	Zinkkid Syrup	0.00	528.00	528.00	308.00	220.00

Tabel 4.1 merupakan *dataset* form LPLPO tahun 2017-2018. *Dataset* tersebut akan dibersihkan dari data redundansi maupun data inkonsisten.

#### **Preprocessing**

Sebelum melakukan pengujian pada data, data yang sudah tersedia akan diperiksa agar bersih dari redundansi dan data inkonsisten. Pada penelitian ini akan diambil 204 data yang siap olah.

Pada Tabel 4.2 merupakan *dataset* setelah proses *preprocessing* dan siap siap olah pengambilan 204 data dikarenakan pembersihan dari banyaknya data inkonsisten dan *missing value*.

Tabel 4. 2 *Dataset* setelah proses *preprocessing*

No	NAMA OBAT	STOK AWAL	PENERIMAAN	PERSEDIAAN	PEMAKAIAN	SISA STOK
1.	Amoksilin Syr Ker	1440.00	552.00	1992.00	1349.00	643.00
2.	Amoksilin Kap 500 Mg	26900.00	29000.00	55900.00	33600.00	22300.00
3.	Ampisilin Inj IM/IV 1000 Mg/MI	30.00	30.00	60.00	30.00	30.00
4.	Ciprofloxacin 500 MG	11200.00	6000.00	17200.00	5200.00	12000.00
5.	Kotrimoksazol Suspensi 60 MI	607.00	400.00	1007.00	377.00	630.00
...	...	...	...	...	...	...
204.	Zinkkid Syrup	0.00	528.00	528.00	308.00	220.00

#### **Hasil perhitungan manual algoritma K-Means**

##### Penentuan *centroid* awal

Pada penelitian ini, penentuan *centroid* awal diambil secara acak.

Tabel 4. 3 *Centroid* Awal

Klaster	<i>Centroid</i> Awal				
<b>C1</b>	123.00	24.00	147.00	55.00	92.00
<b>C2</b>	11800.00	5400.00	65800.00	54200.00	11600.00

Pada Tabel 4.3 *centroid* awal klaster C1 bernilai 123.00 , 24.00 , 147.00 , 55.00 , 92.00 dan untuk klaster C2 bernilai 11800.00 , 5400.00 , 65800.00 , 54200.00 , 11600.00.

##### Perhitungan Jarak data ke *centroid*

Perhitungan jarak data ke *centroid* menggunakan *Euclidean Distance* iterasi pertama menghasilkan nilai dan klaster seperti pada Tabel 4.4

Tabel 4. 4 Hasil jarak data iterasi pertama

No	C1	C2	Terpendek	Klaster
1.	2719.436523	84214.13816	2719.436523	C1
2.	79268.37707	29408.24714	29408.24714	C2

3.	143.954854	86766.81583	143.954854	C1
4.	24849.32198	69017.62094	24849.32198	C1
5.	1228.153085	85677.41083	1228.153085	C1
6.	17942.88725	73198.84562	17942.88725	C1
7.	1304.91494	85501.26666	1304.91494	C1
8.	1044.654967	85924.81152	1044.654967	C1
9.	3291.535052	84472.81693	3291.535052	C1
10.	0	86664.46706	0	C1
...	...	...	...	...
204	703.3484201	86217.58855	703.3484201	C1

Tabel 4.4 merupakan hasil perhitungan jarak antar data pada iterasi pertama. Pada iterasi pertama, jumlah anggota kluster C1 sebanyak 186 data dan anggota kluster C2 sebanyak 18 data. Perhitungan *K-Means* terhenti jika anggota kluster sudah tidak ada yang berpindah tata letaknya, untuk memastikan apakah data sudah tidak ada lagi yang berpindah maka perlu dilakukan iterasi kedua. Jika sudah tidak ada lagi data yang berpindah maka iterasi selesai.

**Hasil perhitungan manual algoritma HCC Single Linkage**

Menghitung jarak antar data

Pada tahap ini, menghitung jarak antar data menggunakan *Euclidean Distance*.

Tabel 4. 5 Matriks Jarak Antar Data

Data	1	2	3	...n
1	0.00	76633.71	2847.58	...
2	76633.71	0.00	79386.78	...
3	2847.58	79386.78	0.00	...
4	22361.42	56443.16	24983.98	...
5	1622.40	78076.12	1359.06	...
6	15445.36	62297.27	18074.93	...
7	1601.75	78019.68	1402.57	...
8	1880.62	78318.98	1177.65	...
9	1761.20	76382.85	3424.21	...
10	2719.44	79268.38	143.95	...
...n	...	...	...	...

Pada Tabel 4.5 merupakan matriks jarak antar data yang dihitung menggunakan rumus *Euclidean Distance*. Pada tahap ini akan diambil nilai terkecil dari perhitungan *single linkage* atau jarak *minimum*. Dari tabel diatas dapat diambil nilai pada kelompok 3 dan 10 sebagai jarak minimum sebesar 143.95 dan kelompok tersebut akan digabungkan menjadi satu kelompok yaitu (3,10). Selanjutnya akan dihitung kembali kelompok yang sudah didapat (3,10) dengan kelompok lain yang tersisa.

Tabel 4. 6 Hasil perhitungan data

Data	Nilai
(3,10)	143.95
(10,3)	143.95
(8,18)	187.85
(5;8,18)	194.99
(13,16)	218.47
(5,8,18;3,16)	253.84
(5,8,18,13,16;17)	297.26
(5,8,18,13,16,17;11)	308.67

(3,10;5,8,18,13,16,17,11)	4219.91
(3,10,5,8,18,13,16,17,11;7)	588.50
(3,10,5,8,18,13,16,17,11,7;15)	629.53
(1;3,10,5,8,18,13,16,17,11,7,15)	1601.75
(1,3,10,5,8,18,13,16,17,11,7,15;9)	1761.20
(1,3,10,5,8,18,13,16,17,11,7,15,9;14)	1967.23
(1,3,10,5,8,18,13,16,17,11,7,15,9,14;20)	2605.99
(6,12)	4885.53
(4;6,12)	7005.71
(1,3,10,5,8,18,13,16,17,11,7,15,9,14,20;4,6,12)	21613.65
(2;19)	38598.83
(2,19)	56443.16

Tabel 4.6 merupakan hasil perhitungan data menggunakan *Single Linkage*. Perhitungan berhenti setelah mendapatkan nilai kelompok kluster tunggal dan pada tabel diatas dapat diketahui kelompok luster terakhir adalah kelompok (2,19). Dari 20 data yang menjadi sampel sebanyak 18 data masuk ke dalam kluster C1 dan 2 data masuk ke kluster C2.

**Hasil kluster optimal**

Kluster optimal dari proses sebelumnya didapatkan hasil untuk algoritma *K-Means* kluster C1 sebanyak 180 data, kluster C2 sebanyak 24 data. Algoritma *HCC Single Linkage* kluster optimal C1 sebanyak 203 data dan kluster 2 sebanyak 1 data.

**Hasil Validitas terbaik**

Hasil uji validitas, algoritma *HCC Single Linkage* memperoleh hasil yang lebih optimal dibandingkan dengan algoritma *K-Means* dengan nilai sebesar 0.8629 banding 0.8414 berdasarkan perhitungan *Sillhoutte Index*.

**Hasil Pengujian**

Pada penelitian ini, penulis menguji nilai kluster optimal dan nilai validitas menggunakan 4 kategori data yaitu 100 data, 150 data, 200 data dan 204 data untuk menguji performansi *clustering* dengan *data set* tersebut menghasilkan

Tabel 4. 7 Hasil pengujian *K-Means*

Data	Klaster		Validitas
	C1	C2	
100	11	89	0.7883
150	132	18	0.8343
200	177	23	0.8393
204	180	24	0.8414

Tabel 4.7 merupakan hasil pengujian menggunakan 100 data, 150 data, 200 data dan 204 data. Pada tabel tersebut menunjukkan bahwa jumlah data pada algoritma *K-Means* mempengaruhi kluster optimal dari kluster C1 dan kluster C2. Jumlah data juga mempengaruhi dari nilai validitas yang didapat, semakin banyak data yang diolah hasil validitasnya semakin baik.

Tabel 4. 8 Hasil Pengujian *HCC Single Linkage*

Data	Klaster		Validitas
	C1	C2	
100	99	1	0.7838
150	149	1	0.8573
200	199	1	0.8611
204	203	1	0.8629

Tabel 4.8 merupakan hasil pengujian menggunakan 100 data, 150 data, 200 data dan 204 data. Pada tabel tersebut menunjukkan bahwa jumlah data tidak terlalu mempengaruhi karena metodenya yang menggunakan *single linkage* atau menghitung sampai kluster tunggal ( $K=1$ ). Dapat dilihat dari kluster C2 dari 4 pengujian adalah jumlah kluster konvergenya yang sama yaitu 1 (tunggal). Semakin banyak data yang diolah hasil validitasnya semakin optimal.

### B. Analisis

Pada proses Algoritma *K-Means*, penentuan *centroid* awal dilakukan secara acak. Maka dari itu setiap data memiliki probabilitas (kemungkinan) yang sama untuk dipilih sebagai *centroid* awal. Kurang optimalnya algoritma *K-Means* terletak pada penentuan *centroid* awal tersebut. Jika *centroid* awalnya berubah maka akan mempengaruhi nilai secara keseluruhan sampai pada hasil kluster optimal dan validasi yang akan ikut berubah. Pada hasil pengujian menggunakan 4 kategori data, terlihat jumlah data mempengaruhi hasil kluster optimal dan nilai validitasnya. Pada hasil pengujian dengan 204 data, kluster optimal mendapatkan 180 data berada pada kluster C1 (obat dengan pemakaian lambat) dan 24 data berada pada kluster C2 (obat dengan pemakaian cepat) menandakan bahwa data-data tersebut mendekati atau menuju ke titik pusat kluster (*centroid*). Hasil kluster optimal direpresentasikan dengan plot yang berupa titik-titik menyesuaikan dari tiap kluster yang dapat dengan mudah dibaca dan dipahami.

Pada proses algoritma *HCC Single Linkage*, penentuan *centroid* awal dilakukan secara konsisten dengan menghitung jarak antar data kemudian diambil nilai terkecil dari nilai jarak antar data tersebut. Semua kelompok yang dibentuk berdasarkan *entities* individu dan menggabungkan jarak dari tiap kelompok dengan nilai paling kecil atau *similarities* (kemiripan) data yang paling besar. Penggabungan dua kelompok tersebut merupakan hasil akhir dari proses algoritma ini karena dua kelompok yang telah digabung tidak bisa dikembalikan seperti semula. Algoritma *HCC Single Linkage* merupakan algoritma yang kompleks, maka dari itu pengerjaannya membutuhkan waktu sedikit lebih lama dibandingkan dengan *K-Means* yang sederhana dan mudah. Pada hasil pengujian menggunakan 4 kategori data, terlihat jumlah data tidak begitu mempengaruhi hasil kluster optimal tetapi semakin banyak data yang diolah, hasil dari nilai validitasnya akan semakin optimal. Pada hasil kluster optimal C1 mendapatkan 203 data berada pada kluster C1 (obat dengan pemakaian lambat) dan 1 data berada pada kluster C2 (obat dengan pemakaian cepat) menandakan bahwa data-data tersebut sudah berada pada tingkat kemiripan yang paling besar. Berbeda dengan algoritma *K-Means*, hasil dari kluster optimal di representasikan dengan dendogram berupa bagan kelompok kluster yang tujuannya sama seperti plot pada *K-Means* untuk lebih mudah dibaca dari proses algoritma *HCC Single Linkage*.

Dari hasil kluster optimal, setelahnya dihitung berapa nilai validitas dari kedua algoritma tersebut dan selisih nilai validitas kedua algoritma tidak terlalu jauh hanya sebesar 0.0215. Nilai validitas tertinggi didapatkan oleh algoritma *HCC Single Linkage*. Hal ini disebabkan karena pembangkitan titik pusat kluster (*centroid*) secara acak sehingga *centroid* yang diperoleh belum tentu *centroid* yang didominasi oleh data yang lebih banyak kemiripannya. Sedangkan pada *HCC Single Linkage*, titik pusat kluster (*centroid*) dilakukan dengan konsisten, sehingga *centroid* didominasi dengan data yang lebih banyak kemiripannya.

Hasil dari uji validitas algoritma *K-Means* dan *HCC Single Linkage* memiliki tingkat validitas yang tinggi karena mendapatkan nilai masing-masing sebesar 0.8414 dan 0.8629. Nilai ukuran dari validitas kedua algoritma tersebut adalah *Strong Structure* karena memiliki nilai *Sillhouette Index* lebih dari 0.7.

Penelitian ini relevan dengan peneliti Joanna Ardiyanti M.N dan Yupie Kusumawati bahwa dari proses algoritma *K-Means* hasil yang didapat adalah obat dengan pemakaian lambat dan obat dengan pemakaian cepat. Sejalan dengan penelitian milik Undang Syaripudin, dkk bahwa algoritma *Hierarchical Clustering* lebih kompleks dan lama dalam pengerjaannya dibanding algoritma *K-Means* yang sederhana dan cepat. Peneliti Rendy Handoyo, dkk menyebutkan bahwa kurang optimalnya *K-Means* disebabkan karena proses inisialisasi kluster yang dilakukan secara acak pada pembangkitan kluster awal. Oleh karena itu, sulit untuk mendapatkan kluster awal yang unik. Meskipun dilakukan iterasi sebanyak beberapa kali, namun hasil iterasi tersebut belum tentu merepresentasikan suatu kluster yang baik dianggap searah dengan penelitian ini.

## V. CONCLUSION

Merujuk ke rumusan masalah dan tujuan penelitian, kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut:

1. Algoritma *K-Means* dan *HCC Single Linkage* mampu mendapatkan kluster optimal sesuai dengan data kemiripan. Kluster optimal yang didapat dari kedua algoritma yaitu algoritma *K-Means* 180 data berada pada C1 atau obat dengan pemakaian lambat dan 24 data berada pada C2 atau obat dengan pemakaian cepat. Algoritma *HCC Single Linkage* dengan hasil 203 data berada pada C1 atau obat dengan pemakaian lambat dan 1 data berada pada C2 atau obat dengan pemakaian cepat.
2. Nilai validitas algoritma *k-Means* dan *HCC Single Linkage* tergolong tinggi (*strong structure*) dibuktikan dengan pengujian validitas *Sillhoutte Index* masing-masing mendapatkan nilai validitas SI sebesar 0.8014 dan 0.8629. Oleh karena itu, nilai validitas terbaik berdasarkan pengujian SI diantara keduanya diperoleh algoritma *HCC Single Linkage*.

## REFERENCES

- [1] J. Ardhyanti, M. Nugraha, Y. Kusumawati, S. Informasi, F. I. Komputer, and U. D. Nuswantoro, "Data Mining Dengan Metode Clustering Untuk Pengolahan Informasi Persediaan Obat Pada Puskesmas Pandanaran Semarang," *UDiNus Repos.*, vol. 1, no. 1, p. 1, 2014.
- [2] BPOM, "Materi Edukasi tentang Peduli Obat dan Pangan Aman," *GNPOPA (Gerakan Nas. Peduli Obat Dan Pangan Aman)*, vol. 1, no. 1, p. 5, 2015.
- [3] P. F. Sujati Woro Indijah, *Farmakologi*. 2016.
- [4] Supriani, "Evaluasi Siklus Pengelolaan Obat Tentang Distribusi Obat Askes di IFRS RSUD Ajibarang Banyumas," *2-Trik*, vol. 8, no. 1, p. 1, 2018.
- [5] M. K. Sadiyanto, SKM., "Dinas Kesehatan Kabupaten Banyumas," vol. 1, no. 1, pp. 1–180, 2015.
- [6] U. Syaripudin, I. Badruzaman, E. Yani, and M. Ramdhani, "Studi Komparatif Penerapan Metode Hierarchical, K-Means Dan Self Organizing Maps (SOM) Clustering Pada Basis Data," vol. VII, no. 1, pp. 132–149, 2013.
- [7] T. Alfina, B. Santosa, and R. Barakbah, "Analisa Perbandingan Metode Hierarchical Clustering , K-means dan Gabungan Keduanya dalam Cluster Data ( Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS )," *J. Tek. ITS*, vol. 1, no. Data Mining, pp. 1–5, 2012.
- [8] R. Handoyo, S. M. Nasution, P. Studi, S. Komputer, S. Linkage, S. Coefficient, R. Handoyo, and S. M. Nasution, "Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K-Means Pada Pengelompokan Dokumen," *JSM STMIK Mikroskil*, vol. 15, no. 2, pp. 73–82, 2014.
- [9] A. S. A. Ali Ridho Barakbah, Tita Karlita, *Logika dan algoritma*. 2013.
- [10] A. P. Windarto, D. Hartama, and S. Solikhun, "Komparasi Kinerja Algoritma Fuzzy C-Means Dan K-Means Dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai Akademik Siswa ( Studi Kasus : SMP Negeri 2 Pematangsiantar )," *JURASIK (Jurnal Ris. Sist. Inf. Tek. Inform.*, vol. 1, no. 1, pp. 1–9, 2016.
- [11] Bertalya, "Konsep Data Mining," *Data Min. Knowl. Discov. Databases*, 2009.
- [12] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [13] Amirulloh, *Populasi dan sampel*. 2015.
- [14] R. Nasution, "Teknik Sampling," *Usu Digit. Libr.*, vol. 1, no. 2, pp. 1–7, 2003.
- [15] M. Tania, "Hubungan Pengetahuan Remaja Dengan Konsumsi Minuman Ringan di SMKN Baleendah Bandung," *J. Ilmu Keperawatan*, vol. IV, no. 1, 2016.
- [16] W. Budiharto, *Pengantar Praktis Pemrograman R untuk Ilmu Komputer*. 2013.
- [17] A. Segonds, *Introduction to ggplot2*. 2015.
- [18] K. Ram, "Data Visualization Using R & ggplot2 Some housekeeping," p. 4, 2013.
- [19] N. Ghazzali, "NbClust: An R Package for Determining the," *J. Stat. Softw.*, vol. 61, no. 6, 2014.
- [20] M. S. Wafa, "Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma K-Means Clustering," *Univ. Islam Negeri Maulana Malik Ibrahim Malang*, vol. 1, pp. 1–117, 2013.
- [21] S. Agustina, D. Yhudo, H. Santoso, N. Marnasusanto, A. Tirtana, and F. Khusnu, "Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means," *Clust. K-Means*, vol. 1, no. 1, pp. 1–7, 2012.
- [22] Asroni and R. Adrian, "Penerapan Metode K-Means untuk Clustering Mahasiswa Berdasarkan Nilai Akademik dengan Weka Interface Studi Kasus pada Jurusan Teknik Informatika UMM Magelang," *J. Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015.
- [23] I. Wahyuni, Y. A. Auliya, A. Rahmi, and W. F. Mahmudy, "Clustering Nasabah Bank Berdasarkan Tingkat Likuiditas Menggunakan Hybrid Particle Swarm Optimization dengan K-Means," *J. Ilm. Teknol. dan Inf. ASIA*, vol. 10, no. 2, pp. 24–33, 2016.
- [24] S. Mashfuufah and D. Istiawan, "Penerapan Partition Entropy Index , Partition Coefficient Index dan Xie BeniIndex untuk Penentuan Jumlah Kluster Optimal pada Algoritma Fuzzy C-Means dalam Pemetaan Tingkat Kesejahteraan Penduduk Jawa Tengah," *URECOL*, vol. 1, no. 1, pp. 51–60, 2018.
- [25] A. Ramadhan and Z. Efendi, "Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling," *SNTIKI*, vol. 1, no. 2, pp. 18–19, 2017.
- [26] G. P. Trayasiwi, "Penerapan Metode Klastering Dengan Algoritma K-Means Untuk Prediksi Kelulusan Mahasiswa Pasa Program Studi Teknik Informatika Strata Satu," *UDiNus Repos.*, vol. 1, no. 1, pp. 1–11, 2017.
- [27] Kemenkes RI, "Keputusan Menteri Kesehatan Republik Indonesia Nomor 189/MENKES/SK/III/2006 Tentang Kebijakan Obat Nasional," *Kementrian Kesehat.*, pp. 4–8, 2010.