

Data Mining Menggunakan Algoritma K-Means Pengelompokan Penyebaran Diare Di Kota Makassar

Hasyrif SY, Rismayani, Asrul Syam

STMIK Dipanegara Makassar

Jl. Perintis Kemerdekaan Km 9 Makassar, telp 0411-587194 fax. 0411-588284

e-mail: hasyrif@gmail.com, maya_setya@ymail.com, assyams03@gmail.com

Abstrak

Sebagai kota besar, makassar merupakan sebuah ikon yang penuh dengan dinamika, di kota ini banyak sekali aktifitas ekonomi yang berdampak positif pada pendapatan daerah, namun layaknya kota lainnya di Indonesia kota ini juga memiliki masalah, salah satunya adalah masalah kesehatan. Berdasarkan data profil kesehatan 2016 masih banyak ditemukan kasus kasus diare yang di alami oleh warganya. Salah satu aspek pendukung dalam penyelesaian masalah ini yaitu dengan melakukan analisa untuk mempelajari sebaran diare tersebut. Dalam kasus ini akan digunakan algoritma K-Means Clustering yang merupakan salah satu algoritma yang akan mengelompokkan data yang sama pada kelompok tertentu, dan menempatkan data dengan karakteristik berbeda dalam kelompok lain. Hasil dari penelitian ini ditemukan bahwa ada dua cluster daerah sebaran yang muncul berdasarkan algoritma diatas.

Kata kunci: K-Means Algoritma, Data Mining, Diare.

Abstract

As a big city, Makassar is an icon full of dynamics, in this city a lot of economic activities that have a positive impact on local revenue, but like other cities in Indonesia the city also has problems, one of which is a health problem. Based on health profile data 2016 is still found many cases of diarrhea cases experienced by the citizens. One aspect of supporters in this case is to conduct an analysis to start the spread of diarrhea. In this case we will use the K-Means Clustering algorithm which is one algorithm that will group the same data in a particular group, and put the data with different people in other groups. The results of this study found that there are two clusters of regional distribution that appear based on the above algorithm.

Keywords: K-Means Algorithm, Data Mining, Diarrhea

1. Pendahuluan

Makassar merupakan kota metropolitan terbesar di kawasan Indonesia Timur, kota ini merupakan ibukota dari provinsi sulawesi selatan. Sebagai kota besar, tentu banyak dinamika yang terjadi selain banyaknya aktifitas perekonomian yang berdampak positif bagi daerah namun layaknya kota lain juga punya masalah tersendiri, salah satunya adalah terkait aspek kesehatan, terkhusus penyakit diare. Berdasarkan data dari profil kesehatan kota makassar, masih banyak ditemukan kasus kasus diare yang tersebar di berbagai kecamatan, setidaknya ada sekitar 30 ribuan data temuan kasus diare yang tercatat di puskesmas, sebagai kota besar tentu ini merupakan sebuah masalah dan tantangan tersendiri bagi pemerintah setempat.

Tentu pemerintah dalam hal ini sudah melakukan usaha-usaha untuk meminimalisir masalah ini, secara umum dengan memaksimalkan layanan-layanan kesehatan dan tenaga kesehatan, dan tentu juga aspek aspek lain yang secara langsung/tidak langsung sebagai faktor pendukung. Aspek lain yang bisa menjadi pendukung adalah terkait data, kita melihat kadang data tersebut digunakan hanya sebagai bahan untuk menampilkan kondisi yang telah/sedang terjadi. Namun sebenarnya ada potensi tersembunyi yang bisa digali, yaitu bagaimana data tersebut bisa diolah untuk mendukung penyelesaian kasus. Salah satu contoh pemanfaatannya adalah menganalisa data data tersebut sehingga bisa menghasilkan informasi

yang berguna. Dalam kasus ini kami akan mencoba untuk mengolah data laporan dari buku profil kesehatan kota Makassar untuk tujuan tertentu, dimana data akan diolah untuk menghasilkan klasifikasi / pengelompokan wilayah penyebaran diare.

K-means merupakan salah satu metode clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster. Metode ini mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam cluster yang lain[1]. K-Means Clustering adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Terdapat dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu Hierarchical dan Non-Hierarchical, dan K-Means merupakan salah satu metode data clustering non-hierarchical atau Partitional Clustering. Metode K-Means Clustering berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain.

Diare (bahasa Inggris: *diarrhea*) adalah sebuah penyakit di saat tinja atau feses berubah menjadi lembek atau cair yang biasanya terjadi paling sedikit tiga kali dalam 24 jam. Diare kebanyakan disebabkan oleh beberapa infeksi virus tetapi juga seringkali akibat dari racun bakteria. Dalam kondisi hidup yang bersih dan dengan makanan mencukupi dan air tersedia, pasien yang sehat biasanya sembuh dari infeksi virus umum dalam beberapa hari dan paling lama satu minggu. Namun untuk individu yang sakit atau kurang gizi, diare dapat menyebabkan dehidrasi yang parah dan dapat mengancam jiwa bila tanpa perawatan. Diare dapat menjadi gejala penyakit yang lebih serius, seperti disentri, kolera atau botulisme, dan juga dapat menjadi indikasi sindrom kronis seperti penyakit Crohn. Meskipun penderita apendisitis umumnya tidak mengalami diare, diare menjadi gejala umum radang usus buntu. Diare juga dapat disebabkan oleh konsumsi alkohol yang berlebihan, terutama dalam seseorang yang tidak cukup makan. jadi apabila mau mengkonsumsi alkohol lebih baik makan terlebih dahulu. Kondisi cuaca yang tidak stabil, sanitasi tempat pengungsian yang buruk serta kondisi rumah yang masih kotor terkena genangan air, juga sulitnya mendapat air bersih menyebabkan mudahnya terjadi wabah diare setelah banjir. Penyakit diare yang terlihat ringan justru bisa membahayakan jiwa, karena saat tubuh kekurangan cairan, maka semua organ akan mengalami gangguan. Diare akan semakin berbahaya jika terjadi pada anak-anak[2].

Ada beberapa penelitian terkait dalam penelitian ini adalah penelitian yang dilakukan oleh Ade Bastian dan dengan judul Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka) yang membahas mengenai penggunaan Algoritma *K-Means Cluster Analysis* dalam studi kasus penyakit menular manusia pada suatu objek. Penelitian ini mengkaji metode *K-Means Cluster Analysis* dalam penyakit menular pada manusia berdasarkan set variabel yang dibentuk per kecamatan tiap puskesmas yang jumlahnya ada 32 Kantor Puskesmas di Kabupaten Majalengka[3]. Penelitian yang dilakukan oleh ediyanto dkk dengan judul Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis yang membahas mengenai Algoritma K-Means Cluster Analysis pada dasarnya dapat diterapkan pada permasalahan dalam memahami perilaku konsumen, mengidentifikasi peluang produk baru dipasaran dan algoritma K-Means ini juga dapat digunakan untuk meringkas objek dari jumlah besar sehingga lebih memudahkan untuk mendiskripsikan sifat-sifat atau karakteristik dari masing-masing kelompok[4]. Kemudian penelitian yang dilakukan oleh Ardilla dkk dengan judul Deteksi Penyakit Epilepsi dengan Menggunakan Entropi Permutasi, K-means Clustering, dan Multilayer Perceptron yang membahas mengenai sistem yang bisa memprediksi apakah seseorang menderita epilepsi bebas kejang, atau epilepsi kejang. Dalam artikel ini diimplementasikan perangkat lunak pendeteksi penyakit epilepsi dengan menggunakan entropi permutasi, K-means clustering, dan multilayer perceptron[5]. Selanjutnya penelitian yang dilakukan oleh Novita Meisida dkk dengan judul K-Means Untuk Klasifikasi Penyakit Karies Gigi yang membahas mengenai engklasifikasian metode Clustering K-Means dibandingkan hasilnya dengan laporan Poli Gigi Medical Center ITS 2009, untuk membandingkan keakuratan Clustering K-Means[6].

2. Metode Penelitian

Adapun lokasi dari penelitian adalah wilayah kota Makassar, dengan mengambil beberapa sampel data mengenai penyakit diare di beberapa kecamatan kota Makassar.

Alat penelitian yang digunakan dalam penelitian adalah :

1. 1 Unit PC dengan Spesifikasi Processor Core i3 HDD 520 GB
2. Dreamweaver
3. Database Mysql XAMP
4. Sistem Operasi Windows 10
5. Alat Perancangan UML

Bahan Penelitian sebagai berikut :

1. Data-data lokasi penyebaran Diare
2. Data-data informasi mengenai penyakit diare

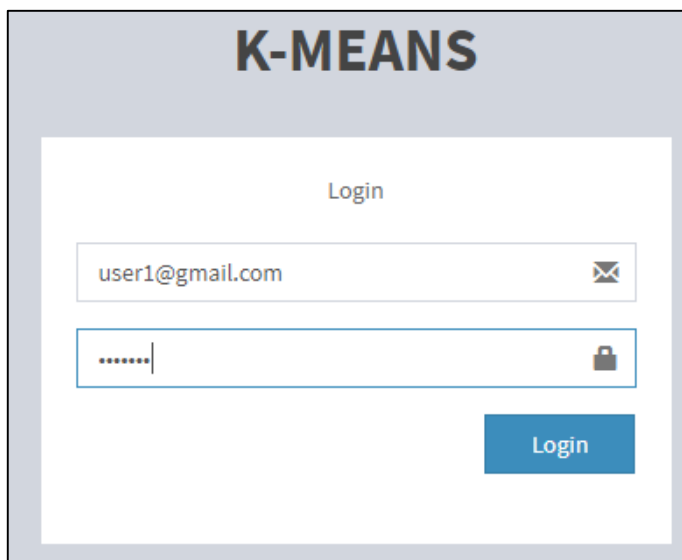
Tahapan penelitian yang dilakukan oleh penulis adalah :

1. Pengumpulan Data: Tahapan dimana penulis terlebih dahulu melakukan pengumpulan data-data yang berhubungan dengan penelitian yang akan di lakukan, baik data dari tempat penelitian, observasi, wawancara dan pustaka.
2. Analisis Sistem: Tahapan dimana penulis merancang sistem maka terlebih dahulu melakukan penganalisaan terhadap sistem yang sedang berjalan saat itu.
3. Desain Sistem: Tahapan dimana penulis melakukan perancangan sistem setelah melakukan analisis sistem. Bagaimana suatu sistem akan menyelesaikan apa yang akan di selesaikan.
4. Pengujian Sistem: Setelah proses coding selesai, dilakukan pengujian system menggunakan metode black box. Serta dilakukan pengujian kuisisioner untuk mengetahui tingkat kegunaan dari sistem.
5. Implementasi Sistem: Tahap ini di gunakan untuk penerapan dan pengujian system kedalam kondisi sebenarnya agar dapat diketahui kekurangan dan kelebihanannya.

3. Hasil dan Pembahasan

3.1 Hasil

Halaman Login



Gambar 1. Halaman Login

Data Awal

DATA AWAL			
NO	KECAMATAN	JUMLAH PUSKESMAS	JUMLAH PENDERITA
1	UJUNG TANAH	4	1319
2	TALLO	3	37815
3	BONTOALA	2	1518
4	WAJO	2	830
5	UJUNG PANDANG	1	764
6	MAKASSAR	3	2279
7	MAMAJANG	2	1641
8	MARISO	3	1588
9	TAMALATE	4	4002
10	RAPPOCINI	4	5532
11	PANAKKUKANG	5	5125
12	MANGGALA	4	2490
13	BIRINGKANAYA	4	5309
14	TAMALANREA	5	2992

Gambar 2. Data Awal

Iterasi 1

ITERASI 1						
NO	KECAMATAN	JUMLAH PUSKESMAS	JUMLAH PENDERITA	CLUSTER 1	CLUSTER 2	KELOMPOK CLUSTER
1	UJUNG TANAH	4	1319	681.00	2,681.00	C1
2	TALLO	3	37815	35,815.00	33,815.00	C2
3	BONTOALA	2	1518	482.00	2,482.00	C1
4	WAJO	2	830	1,170.00	3,170.00	C1
5	UJUNG PANDANG	1	764	1,236.00	3,236.00	C1
6	MAKASSAR	3	2279	279.00	1,721.00	C1
7	MAMAJANG	2	1641	359.00	2,359.00	C1
8	MARISO	3	1588	412.00	2,412.00	C1
9	TAMALATE	4	4002	2,002.00	2.00	C2
10	RAPPOCINI	4	5532	3,532.00	1,532.00	C2
11	PANAKKUKANG	5	5125	3,125.00	1,125.00	C2
12	MANGGALA	4	2490	490.00	1,510.00	C1
13	BIRINGKANAYA	4	5309	3,309.00	1,309.00	C2
14	TAMALANREA	5	2992	992.00	1,008.00	C1

Gambar 3. Iterasi 1

Iterasi 2

ITERASI 2						
NO	KECAMATAN	JUMLAH PUSKESMAS	JUMLAH PENDERITA	CLUSTER 1	CLUSTER 2	KELOMPOK CLUSTER
1	UJUNG TANAH	4	1319	394.45	10,237.60	C1
2	TALLO	3	37815	36,101.56	26,258.40	C2
3	BONTOALA	2	1518	195.45	10,038.60	C1
4	WAJO	2	830	883.44	10,726.60	C1
5	UJUNG PANDANG	1	764	949.45	10,792.60	C1
6	MAKASSAR	3	2279	565.56	9,277.60	C1
7	MAMAJANG	2	1641	72.45	9,915.60	C1
8	MARISO	3	1588	125.44	9,968.60	C1
9	TAMALATE	4	4002	2,288.56	7,554.60	C1
10	RAPPOCINI	4	5532	3,818.56	6,024.60	C1
11	PANAKKUKANG	5	5125	3,411.56	6,431.00	C1
12	MANGGALA	4	2490	776.56	9,066.60	C1
13	BIRINGKANAYA	4	5309	3,595.56	6,247.60	C1
14	TAMALANREA	5	2992	1,278.56	8,564.60	C1

Gambar 4. Iterasi 2

Iterasi 3

ITERASI 3						
NO	KECAMATAN	JUMLAH PUSKESMAS	JUMLAH PENDERITA	CLUSTER 1	CLUSTER 2	KELOMPOK CLUSTER
1	UJUNG TANAH	4	1319	1,403.23	36,496.00	C1
2	TALLO	3	37815	35,092.77	-	C2
3	BONTOALA	2	1518	1,204.23	36,297.00	C1
4	WAJO	2	830	1,892.23	36,985.00	C1
5	UJUNG PANDANG	1	764	1,958.23	37,051.00	C1
6	MAKASSAR	3	2279	443.23	35,536.00	C1
7	MAMAJANG	2	1641	1,081.23	36,174.00	C1
8	MARISO	3	1588	1,134.23	36,227.00	C1
9	TAMALATE	4	4002	1,279.77	33,813.00	C1
10	RAPPOCINI	4	5532	2,809.77	32,283.00	C1
11	PANAKKUKANG	5	5125	2,402.77	32,690.00	C1
12	MANGGALA	4	2490	232.23	35,325.00	C1
13	BIRINGKANAYA	4	5309	2,586.77	32,506.00	C1
14	TAMALANREA	5	2992	269.77	34,823.00	C1

Gambar 5. Iterasi 3

3.2. Pembahasan

Data Awal

Data yang akan diolah ini merupakan data kasus temuan diare perkecamatan di kota Makassar, Sulawesi Selatan. Data tersebut menampilkan data jumlah puskesmas dan jumlah penderita diare perkecamatan.

Tabel 1. Data Awal

NO	KECAMATAN	JUMLAH PUSKESMAS	JUMLAH PENDERITA
1	UJUNG TANAH	4	1319
2	TALLO	3	37815
3	BONTOALA	2	1518
4	WAJO	2	830
5	UJUNG PANDANG	1	764
6	MAKASSAR	3	2279
7	MAMAJANG	2	1641
8	MARISO	3	1588
9	TAMALATE	4	4002
10	RAPPOCINI	4	5532
11	PANAKKUKANG	5	5125
12	MANGGALA	4	2490
13	BIRINGKANAYA	4	5309
14	TAMALANREA	5	2992

Tahap Awal

Pengelompokan data menggunakan Algoritma K-Means terdiri dari beberapa tahapan. Tahapan yang paling awal dari proses ini adalah penentuan jumlah cluster. Cluster menentukan berapa banyak kelompok data yang ingin kita buat atau hasilkan. Dalam penelitian ini kami menggunakan 2 cluster.

Penentuan Pusat Cluster

Tentukan pusat cluster (centroid). Jumlah pusat cluster yang dibuat disesuaikan dengan jumlah cluster yang di tentukan sebelumnya (ada dua cluster). Pusat cluster/centroid berupa angka yang di pilih secara acak. Untuk kasus ini kita akan memilih pusat cluster dengan data sebagai berikut :

- a) (2,2000)
- b) (4,4000)

Menghitung Distance Space

Langkah berikutnya yaitu melakukan perhitungan jarak antara titik pusat cluster dengan data yang kita miliki (data sebaran penyakit diare). Perhitungan ini dilakukan dengan menggunakan persamaan *Euclidean Distance*. Metode *Euclidean Distance* yaitu metode klasifikasi tetangga terdekatnya dengan menghitung jarak antara dua buah objek, metode ini disebut juga jarak Euclidean. Berikut merupakan formula yang dipakai menggunakan persamaan *Euclidean Distance Space* :

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

Ket :

- Dij = Jarak Objek antara nilai data nilai pusat cluster
- p = Jumlah dimensi data
- Xik = Nilai data dari dimensi ke-k
- Xjk = Nilai pusat cluster dari dimensi ke-k

Berdasarkan data pada tabel diatas (tabel 1. data awal) kita melihat ada dua data yang akan di hitung jaraknya menggunakan rumus Euclian Distance yaitu data jumlah puskesmas untuk dimensi 1 dan jumlah penderita untuk dimensi 2, berikut akan diperlihatkan proses perhitungan untuk baris pertama.

$$d(\text{cluster1, data1}) = \sqrt{(\text{data1 dimensi 1} - \text{centroid1 dimensi 1})^2 + (\text{data 2 dimensi 2} - \text{centroid1 dimensi 2})^2}$$

$$d(\text{cluster1, data1}) = \sqrt{(4-2)^2 + (1319 - 2000)^2} = 681.00$$

$$d(\text{cluster1, data1}) = \sqrt{(\text{data1 dimensi 1} - \text{centroid2 dimensi 1})^2 + (\text{data 1 dimensi 2} - \text{centroid2 dimensi 2})^2}$$

$$d(\text{cluster1, data1}) = \sqrt{(4-4)^2 + (1319 - 4000)^2} = 2,681$$

Berdasarkan rumus hitungan diatas maka kita mendapatkan nilai jarak (berdasar persamaan *Euclidean Distance*) perhitungan antara data dan centroid, berikut nilainya :

- Cluster 1 = 681.00
- Cluster 2 = 2,681

NO	KECAMATAN	JUMLAH PUSKESMAS	JUMLAH PENDERITA	CLUSTER 1	CLUSTER 2
1	UJUNG TANAH	4	1319	681.00	2,681.00

Gambar 6. hasil perhitungan euclidean distance untuk baris 1

Dengan menggunakan rumus yang sama kita akan menerapkannya pada baris/record setelahnya (jumlah puskesmas ujung tanah =3,penderita= 37815;jumlah puskesmas bontoala=2,penderita 1518, Dst). Pada tabel X diperlihatkan hasil perhitungan dari semua record.

Tabel 2. hasil perhitungan tahap I

NO	KECAMATAN	JUMLAH PUSKESMAS	JUMLAH PENDERITA	CLUSTER 1	CLUSTER 2	KELOMPOK CLUSTER
1	UJUNG TANAH	4	1319	681.00	2,681.00	C1
2	TALLO	3	37815	35,815.00	33,815.00	C2
3	BONTOALA	2	1518	482.00	2,482.00	C1
4	WAJO	2	830	1,170.00	3,170.00	C1
5	UJUNG PANDANG	1	764	1,236.00	3,236.00	C1
6	MAKASSAR	3	2279	279.00	1,721.00	C1
7	MAMAJANG	2	1641	359.00	2,359.00	C1
8	MARISO	3	1588	412.00	2,412.00	C1
9	TAMALATE	4	4002	2,002.00	2.00	C2
10	RAPPOCINI	4	5532	3,532.00	1,532.00	C2
11	PANAKKUKANG	5	5125	3,125.00	1,125.00	C2
12	MANGGALA	4	2490	490.00	1,510.00	C1
13	BIRINGKANAYA	4	5309	3,309.00	1,309.00	C2
14	TAMALANREA	5	2992	992.00	1,008.00	C1

Pembangkitan Nilai Centroid Baru

Setelah melakukan perhitungan jarak, dilanjutkan dengan membentuk nilai *centroid* baru, caranya yaitu dengan mengelompokkan data berdasar dimensi, pengelompokan diambil dengan memilih jarak terpendek terhadap centroid. Contohnya yaitu pada data kecamatan ujung tanah kita kelompokkan pada C1 karena nilai Cluster 1 (681.00) lebih kecil dari Cluster 2 (2,681.00). Setelah menemukan kelompok cluster berikutnya kita tinggal menghitung nilai rata-ratanya, nilai rata-rata inilah yang akan menentukan nilai dari centroid berikutnya. Berikut dasar perhitungannya:

$$C_{(cluster1, dimensi1)} = \frac{4 + 2 + 2 + 1 + 3 + 2 + 3 + 4 + 5}{9} = 2.89$$

$$C_{(cluster1, dimensi2)} = \frac{1,319 + 1,518 + 830 + 764 + 2,279 + 1,641 + 1,588 + 2,490 + 2,992}{9} = 21,713.44$$

Untuk cluster 2 memiliki 5 data. Berikut perhitungannya:

$$C_{(cluster2, dimensi1)} = \frac{3 + 4 + 4 + 5 + 4}{5} = 4$$

$$C_{(cluster2, dimensi2)} = \frac{37,815 + 4,002 + 5,532 + 5,125 + 5,309.00}{5} = 11,556.60$$

Hingga tahap ini kita telah mendapatkan nilai centroid yang baru yaitu :

- (2.89, 21,713.44)
- (4, 11,556.60)

Perhitungan Iterasi Tahap II

Setelah mendapatkan nilai centroid baru, maka dilakukan tahapan yang sama dengan proses sebelumnya, yaitu dengan melakukan perhitungan jarak terdekat menggunakan persamaan *Euclidean Distance* untuk cluster 1 dan cluster 2, namun menggunakan titik centroid yang baru, berikut diperlihatkan hasil proses perhitungan untuk iterasi ke 2.

Tabel 3. hasil perhitungan tahap II

NO	KECAMATAN	JUMLAH PUSKESM	JUMLAH PENDERITA	CLUSTER 1	CLUSTER 2	KELOMPOK CLUSTER
1	UJUNG TANAH	4	1319	394.45	10,237.60	C1
2	TALLO	3	37815	36,101.56	26,258.40	C2
3	BONTOALA	2	1518	195.45	10,038.60	C1
4	WAJO	2	830	883.44	10,726.60	C1
5	UJUNG PANDANG	1	764	949.45	10,792.60	C1
6	MAKASSAR	3	2279	565.56	9,277.60	C1
7	MAMAJANG	2	1641	72.45	9,915.60	C1
8	MARISO	3	1588	125.44	9,968.60	C1
9	TAMALATE	4	4002	2,288.56	7,554.60	C1
10	RAPPOCINI	4	5532	3,818.56	6,024.60	C1
11	PANAKKUKANG	5	5125	3,411.56	6,431.60	C1
12	MANGGALA	4	2490	776.56	9,066.60	C1
13	BIRINGKANAYA	4	5309	3,595.56	6,247.60	C1
14	TAMALANREA	5	2992	1,278.56	8,564.60	C1

- Centroid 1 = 3.31, 2,722.23
- Centroid 2 = 3.00, 37,815.00

Iterasi Tahap III

Berikut diperlihatkan tabel hasil perhitungan untuk iterasi ke 3

Tabel 3. hasil perhitungan tahap III

NO	KECAMATAN	JUMLAH PUSKESMAS	JUMLAH PENDERITA	CLUSTER 1	CLUSTER 2	KELOMPOK CLUSTER
1	UJUNG TAN	4	1319	1,403.23	36,496.00	C1
2	TALLO	3	37815	35,092.77	-	C2
3	BONTOALA	2	1518	1,204.23	36,297.00	C1
4	WAJO	2	830	1,892.23	36,985.00	C1
5	UJUNG PAN	1	764	1,958.23	37,051.00	C1
6	MAKASSAR	3	2279	443.23	35,536.00	C1
7	MAMAJANG	2	1641	1,081.23	36,174.00	C1
8	MARISO	3	1588	1,134.23	36,227.00	C1
9	TAMALATE	4	4002	1,279.77	33,813.00	C1
10	RAPPOCINI	4	5532	2,809.77	32,283.00	C1
11	PANAKKUKU	5	5125	2,402.77	32,690.00	C1
12	MANGGALA	4	2490	232.23	35,325.00	C1
13	BIRINGKAN	4	5309	2,586.77	32,506.00	C1
14	TAMALANRI	5	2992	269.77	34,823.00	C1

- Centroid 1 = 3.31, 2,722.23
- Centroid 2 = 3.00, 37,815.00

Hingga tahap ini kita menemukan satu hal yaitu nilai centroid pada iterasi ke 2 dan ke 3 memiliki nilai yang sama, proses perhitungan dianggap cukup jika nilai centroid tidak mengalami perubahan. Untuk kasus ini proses iterasi/perulangan perhitungan dilakukan sebanyak 3 kali.

4. Kesimpulan

Dari hasil perhitungan menggunakan algoritma K-Means didapatkan nilai centroid akhir untuk centeroid 1 (3.31, 2,722.23) dan centroid 2 (3.00, 37,815.00), kelompok cluster dua hanya 1 yaitu kecamatan Tallo, dan selebihnya masuk pada cluster ke satu. Pusat cluster ke dua merupakan daerah dengan tingkat penderita diare tertinggi.

Referensi

- [1] informatikalogi, "Algoritma K-Means Clustering," *INFORMATIKALOGI*, 12-Nov-2016.
- [2] - Suyadi and S. N. Aini, "FAKTOR-FAKTOR YANG MEMPENGARUHI TERJADINYA DIARE PADA BALITA USIA (1-3 TAHUN) DI DESA BLIMBING, SAMBIREJO, SRAGEN," vol. 6, no. 2, Jul. 2018.
- [3] Ade Bastian, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," *J. Sist. Inf.*, vol. 14, no. 1, pp. 28–34, Apr. 2018.

- [4] M. N. M. Ediyanto, "PENGKLASIFIKASIAN KARAKTERISTIK DENGAN METODE K-MEANS CLUSTER ANALYSIS," *BIMASTER*, vol. 2, no. 02, Aug. 2013.
- [5] Y. Ardilla, H. Tjandrasa, and I. Ariesianti, "Deteksi Penyakit Epilepsi dengan Menggunakan Entropi Permutasi, K-means Clustering, dan Multilayer Perceptron," *J. Tek. ITS*, vol. 3, no. 1, pp. A70–A74, Mar. 2014.
- [6] N. Meisida, O. Soesanto, and H. K. Candra, "K-MEANS UNTUK KLASIFIKASI PENYAKIT KARIES GIGI," *KLIK - Kumpul. J. ILMU Komput.*, vol. 1, no. 1, pp. 12–22, Jun. 2017.