ANALISIS CLUSTERING TINGKAT KERAWANAN WILAYAH TERHADAP

KASUS PENYAKIT DI KABUPATEN SLEMAN DENGAN METODE K-MEANS

*CLUSTERING ANALYSIS OF REGION'S VULNERABILITY LEVELS TO DISEASE CASES USING K-MEANS METHOD IN THE DISTRICT OF SLEMAN*

Wirnaning Rokhimah, Prodi Matematika FMIPA UNY

Kismiantini\*, Prodi Matematika FMIPA UNY

\*e-mail: kismi@uny.ac.id

**Abstrak**

Penyakit adalah sebuah keadaan yang menyebabkan terganggunya fungsi ataupun bentuk salah satu bagian tubuh yang menjadikan tubuh menjadi tidak dapat bekerja dengan normal. Tujuan dari penelitian ini adalah mendiskripsikan hasil pengelompokan kecamatan berdasarkan kasus penyakit dan karakteristik dari cluster menggunakan metode k-means. Pengelompokan data penelitian jurnal kasus dilakukan pada 17 kecamatan di Kabupaten Sleman menggunakan variabel Demam Berdarah Dengue (DBD), Diare, Tuberkulosis paru (TB paru), Hepatitis B, dan Campak. Data penelitian tersebut diperoleh dari Badan Pusat Statistika (BPS) dan Profil Kesehatan Sleman tahun 2020. Penelitian ini menghasilkan 3 cluster terbaik dengan menggunakan indeks validasi metode elbow, silhouette coefficient, dunn index. Karakteristik jurnal kasus dari setiap cluster dapat dilihat dari rata-rata variabel tiap cluster yaitu variabel DBD, Diare, TB paru, Hepatitis B, Campak. Cluster 1 memiliki nilai rata-rata variabel paling tinggi yang didominasi oleh penyakit diare dan DBD, cluster 2 memiliki rata-rata variabel lebih rendah dari cluster 1 yang didominasi oleh penyakit diare dan Hepatitis B, cluster 3 memiliki rata-rata cluster paling rendah yang didominasi oleh penyakit diare dan campak.

**Kata kunci:** Kabupaten Sleman, Rawan Wilayah, Penyakit, K-Means

***Abstract***

*Disease is a condition that disrupts the function or shape of one part of the body that makes the body unable to work the way it should. K-means is a non-hierarchical grouping method that trying to patition which exists data into one or more clusters, so it can be expected visible points of the location of the spread of infectious diseases. Study case journal data clustering is done at 17 districts in Sleman Regency using the variables of Dengue Homorrhagic Fever (DHF), Diarrhea, Tuberculosis (TB), Hepatitis B, and Measles. The variabel data was obtained from the Central Bureau of Statistics and Health Profile of Sleman in 2020. The results of the validation index using the elbow method, silhouette coefficient, dunn index calculation show that the best number of clusters is 3 clusters. The characteristics case journal of each cluster can be seen from the average variabel of each cluster, namely the variables of Dengue Homorrhagic Fever (DHF), Diarrhea, Tuberculosis (TB), Hepatitis B, and Measles. Cluster 1 in the high cluster mean which is dominated by Dengue Homorrhagic Fever (DHF) and Diarrhea. Cluster 2 in the medium cluster mean which is dominated by Hepatitis B and Diarrhea. Cluster 3 in the low cluster mean which is dominated by Measles and Diarrhea.*

***Keywords****: Sleman Regency, Prone Area, Disease, K-Means*

**PENDAHULUAN**

Penyakit adalah suatu kondisi patologis berupa kelainan fungsi atau morfologi suatu organ atau jar tubuh (Achmadi, 2012). Ada berbagai macam penyakit di dunia, salah satunya adalah penyakit menular. Penyakit menular adalah penyakit yang disebabkan oleh bakteri, virus, atau parasit yang dapat ditularkan melalui media tertentu. (Vatimatunnimah, 2013).

Kabupaten Sleman adalah salah satu kabupaten di Provinsi Daerah Istimewa Yogyakarta yang tidak lepas dari kasus penyakit menular. Wilayah tersebut berpotensial bagi penularan kasus penyakit menular karena jumlah dan kepadatan penduduknya yang semakin padat. Berdasarkan data yang dipeoleh dari Profil Kesehatan Kabupaten Sleman, tercatat pada tahun 2017 ada 844 kasus tuberkulosis (TB) yang terjadi di Kabupaten Sleman. Jumlah kasus demam berdarah dengue (DBD) pada tahun 2017 mencapai 427 kasus dengan jumlah kasus kematian sebanyak 3 orang, tahun 2018 tercatat 144 kasus turun 31,78 % dari tahun 2017 dengan kematian 1 kasus. Pada tahun 2019 ditemukan sebanyak 728 dengan kematian 1 kasus. Jumlah kasus hepatitis B pada tahun 2017 tercatat ada 139 kasus. Jumlah kasus campak di Kabupaten Sleman ada 177 kasus pada tahun 2017, tahun 2018 tercatat kasus turun menjadi 115 kasus (Denkes, 2020).

Banyaknya kasus penyakit di wilayah tersebut menjadikan wilayah tersebut sangat rawan terhadap penyakit. Pengelompokan wilayah berdasarkan kasus penyakit dapat dilakukan dengan metode pengelompokan *(clustering)* untuk mendapatkan gambaran umum dari situasi penyakit berbasis lingkungan di tiap wilayah. Pengelompokan *(clustering)* merupakan metode yang digunakan dalam *data mining* yang cara kerjanya mencari dan mengelompokan data yang mempunyai kemiripan karakteristik antara satu dan yang lainnya. Dalam *data mining* ada dua jenis metode *clustering* yaitu metode hierarki dan metode non-hierarki (Santosa, 2007).

*K-means* merupakan salah satu metode pengelompokan non hirarki yang berusaha memartisi data yang ada ke dalam bentuk satu atau lebih *cluster* (Sibarani & Chafid, 2018). Metode *K-Means Clustering* memiliki konsep pengelompokan berdasarkan ukuran kedekatan dan kemudian menggunakan karakteristik tersebut sebagai vektor karakteristik atau *centroid*, sehingga metode *K-Means* bisa menjadi solusi untuk pengklasifikasian karakteristik dari objek dan memiliki ketelitian yang cukup tinggi terhadap ukuran objek (Abrianto, 2018).

Penelitian ini menggunakan metode *K-Means Clustering.* *K-means* di gunakan untuk pengelompokkan tingkat kerawanan terhadap kasus penyakit sehingga menghasilkan kategori dari setiap kecamatan. *K-means* *Clustering* akan membagi kecamatan di Kabupaten Sleman dalam kelompok-kelompok sesuai dengan data variabelnya, yaitu jumlah kasus DBD, kasus diare, kasus tuberkulosis, kasus hepatitis b, dan kasus campak.

Berdasarkan latar belakang yang telah diuraikan diatas, penelitian ini bertujuan untuk mengelompokkan kasus penyakit di wilayah Kabupaten Sleman dengan Metode *K-Means* untuk data tahun 2020. Hasil pengelompokan dengan metode *K-means* dianalisis untuk menentukan kecamatan mana yang memiliki wilayah paling rawan.

**Analisis Cluster.**

Analisis *cluster* merupakan teknik multivariat yang digunakan untuk mengklasifiksi objek atau kasus (responden) ke dalam kelompok yang relatif homogen, yang disebut *cluster*. Objek atau kasus dalam setiap kelompok cenderung mirip satu sama lain dan berbeda jauh atau tidak samadengan objek dari *cluster* lainnya (Supranto, 2004).

*Clustering* merupakan suatu metode untuk mengelompokkan dan mencari data dengan karakteristik yang memiliki kemiripan antara satu data dengan yang lain. Dalam *data mining* ada dua jenis metode *clustering* yang digunakan dalam mengelompokan data, yaitu *hierarchical clustering method* dan *non-hierarchical clustering method* (Santosa, 2007)*.*

Pada penelitian ini menggunakan ukuran kedekatan jarak *Euclidean*. Jarak *Euclidean* merupakan besarnya jarak suatu garis lurus yang menghubungkan antar obyek yang diteliti.

Jika dan maka adalah jarak objek *x* dengan objek *y* (Johnson and Wichern, 2002).

|  |  |
| --- | --- |
|  | (2.1) |

Keterangan:

= Jarak antara obyek *x* dan *y*

= nilai obyek *x* pada variabel *i*

= nilai obyek *y* pada variabel *i*

*p* = banyak variabel yang diamati

## 

## Metode *K-Means*

Metode *K-Means* merupakan salah satu algoritma dalam *data mining* yang bisa digunakan untuk melakukan pengelompokan suatu data. Metode *K-Means* adalah metode yang termasuk dalam algoritma *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan algoritma ini hanya bekerja pada numerik (Bailey, 2005).

C*luster* memiliki tingkat variasi yang kecil (Ong, 2013). Adapun langkah-langkah melakukan *clustering* dengan metode *K-Means* (Rahmayani, 2018) adalah sebagai berikut :

1. Menentukan jumlah *cluster k* yang akan dibentuk.

Misalkan diberikan matriks data dengan *i=1,2,...,n, j=1,2, ..., p* dan asumsikan jumlah *cluster* awal *K.*

1. Inisialisasi *k centroid* atau titik pusat *cluster* awal dapat dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random atau acak. Pusat-pusat *cluster* diberikan nilai awal dengan angka-angka random. Rumus yang digunakan untuk menghitung *centroid cluster* ke- *i* sebagai berikut:

|  |  |
| --- | --- |
|  | (2.2) |

*centroid cluster* ke-*i*

objek ke-*i*

banyaknya objek/ jumlah objek yang menjadi anggota *cluster*

1. Hitung jarak setiap objek ke *centroid* dari masing-masing *cluster*.

Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling dekat antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana.

1. Melakukan proses standarisasi data.

Proses menyamakan bobot dari variabel independen, agar memiliki skala yang sama.

1. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.
2. Hitung lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

## Validasi *cluster* Metode *K-Means*

Validari *cluster* adalah metode yang digunakan untuk melakukan evaluasi hasil dari sebuah algoritma *clustering* yang telah dilakukan dengan tujuan untuk memperoleh jumlah *cluster* yang terbaik (Susilowati, 2020). Validasi hasil *clustering* dapat menggunakan beberapa indeks diantaranya yaitu:

1. Metode *Elbow*

Metode *Elbow* merupakan salah satu metode untuk menentukan jumlah *cluster* yang tepat melalui persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik (Putu & Santoso, 2015).

Jika nilai *cluster* pertama dengan nilai *cluster* kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka jumlah nilai *cluster* tersebut yang tepat (Dewi & Pramita, 2019).

Nilai dari metode *elbow* ini menggunakanperhitungan *Sum Squared Error (SSE).* SSE merupakan kriteria termudah untuk mengukur *clustering.* Adapun perhitungan SSE (Irwanto & Soelaiman, 2012) sebagai berikut:

|  |  |
| --- | --- |
|  | 2.3 |

Keterangan

banyak *cluster*

data ke-*i*

himpunan dari elemen *cluster k*

rata-rata (pusat) dalam *cluster k*

1. *Silhouette Coefficient*

*Silhouette Coefficient* digunakan untuk melihat kualitas dan kekuatan *cluster,* seberapa baik atau buruknya suatu obyek ditempatkan dalam suatu *cluster.*

Nilai *Silhouette index* dari sebuah data ke*-i* diperlukan untuk menghitung nilai *silhouette coefficient.* Nilai *Silhouette Coefficient* didapatkan dengan mencari nilai maksimum dari nilai *Silhouette index Global* dari jumlah *cluster* 2 sampai jumlah *cluster n-1,* seperti pada persamaan (2.4) berikut.

|  |  |
| --- | --- |
|  | (2.4) |

Keterangan:

*Silhouette Coefficient*

*Silhouette index Global*

jumlah *cluster*

Perhitungan nilai dari sebuah data ke-*i,* ada 2 komponen yaitu dan . Nilai adalah rata-rata jarak ke-*i* terhadap semua data lainnya dalam satu *cluster,* sedangkan didapatkan dengan menghitung rata-rata jarak data ke-*i* terhadap semua data dari *cluster* lainnya yang tidak satu *cluster* dengan data ke-*i,* lalu diambil yang terkecil (Petrovic, 2006). Berikut persamaan (2.5) untuk menghitung nilai

|  |  |
| --- | --- |
|  | (2.5) |

Keterangan:

rata-rata jarak data ke-*i* terhadap semua data dalam satu *cluster k.*

jumlah data dalam *cluster* ke-*k*

jarak data ke-*i* dengan data ke-*r* dalam satu *cluster k.*

Berikut ini adalah rumus perhitungan untuk mendapatkan nilai dapat dilihat pada persamaan (2.6) berikut.

|  |  |
| --- | --- |
|  | (2.6) |

Keterangan:

rata-rata jarak data ke-*i* terhadap semua data yang tidak dalam satu *cluster* dengan data ke-*i.*

jarak data ke-*i* dengan data ke-*r* dalam satu *cluster k.*

data ke*-i* pada *cluster k*

data ke*-r* pada *cluster k*

Berikut ini adalah rumus perhitungan untuk mendapatkan nilai dapat dilihat pada persamaan (2.7) berikut.

|  |  |
| --- | --- |
|  | (2.7) |

Keterangan:

*Silhouette index* data ke-*i* dalam satu *cluster k*

rata-rata jarak data ke-*i* terhadap semua data yang tidak dalam satu *cluster* dengan data ke-*i*

rata-rata jarak data ke-*i* terhadap semua data dalam satu *cluster* dengan data ke-*i*

Berikut ini adalah rumus perhitungan untuk mendapatkan nilai dapat dilihat pada persamaan (2.8) berikut.

|  |  |
| --- | --- |
|  | (2.8) |

Keterangan:

rata-rata *Silhouette index cluster j*

*Silhouette index* data ke-*i* dalam satu *cluster*

jumlah data dalam *cluster* ke-*j*

*index* data

Berikut ini adalah rumus perhitungan untuk mendapatkan nilai *SI global* sesuai dengan persamaan (2.9) berikut.

|  |  |
| --- | --- |
|  | (2.9) |

Keterangan:

Rata-rata *Silhouette index* dari dataset

jumlah *cluster*

1. *Dunn Index*

*Dunn index* digunakan untuk validasi hasil kelompok yang ditentukan dengan menggunakan diameter kelompok dan jarak antara dua kelompok. Indeks ini dapat dihitung dengan mengukur rasio jarak terkecil antar kelompokdengan jarak terbesar di dalam kelompok. Nilai *dunn index* yang tertinggi merupakan jumlah *cluster* yang optimal (Fluegemann, Davies, & Aguirre, 2011). Perhitungan *dunn index* sebagai berikut:

|  |  |
| --- | --- |
|  |  |

(2.10)

Keterangan:

*D =* Nilai *Dunn Index*

k = Jumlah *cluster*

*=* Jarak antara *cluster i* dan *cluster j.*

= Diameter *cluster i*

**METODE**

Penelitian ini menggunakan data sekunder. Sumber data yang digunakan dalam penelitian ini adalah data yang telah dipublikasikan oleh Badan Pusat Statistik (BPS) Kabupaten Sleman pada tahun 2020 dengan judul buku Sleman Dalam Angka 2020 berupa Profil Kesehatan Sleman 2020. Data yang digunakan pada penelitian ini terdiri dari 17 Kecamatan. Variabel yang digunakan pada penelitian ini adalah jumlah kasus DBD (Demam Berdarah Dengue), jumlah kasus Diare, jumlah kasus TB Paru, jumlah kasus Hepatitis B, dan jumlah kasus Campak. Langkah-langkah yang digunakan pada penelitian ini sebagai berikut:

1. Menentukan variabel-variabel penelitian yaitu variabel jumlah kasus DBD, jumlah kasus Diare, jumlah kasus Tuberkulosis, jumlah kasus Hepatitis B, dan jumlah kasus Campak.
2. Melakukan proses standarisasi data.
3. *Clustering* data menggunakan *K-means* dengan jarak *euclidean* setiap objek ke *centroid* dari masing-masing *cluster*.
4. Validasi untuk memperoleh *k* optimal dengan menggunakan metode *Elbow Method, Silhouette Coefficient* dan, *Dunn Index.*
5. Menganalisis hasil *cluster* dengan metode *k-means.*
6. Interpretasi dan kesimpulan hasil pengelompokan.

**HASIL DAN PEMBAHASAN**

Dalam penelitian ini data yang digunakan adalah data kasus penyakit menular berbasis lingkungan yang ada di kabupaten Sleman dengan jumlah kecamatan sebanyak 17 kecamatan.

Standarisasi data perlu dilakukan apabila data yang digunakan di dalam sebuah penelitian mempunyai satuan yang bervariasi atau berbeda. Pada penelitian ini data yang dikumpulkan mempunyai variabilitas satuan yang berbeda, maka perlu dilakukan langkah standarisasi terhadap variabel.

Tabel 1 Standarisasi Data Kasus Penyakit

### Graphical user interface, table, Excel Description automatically generated

### Metode *K-Means*

1. Menentukan jumlah *cluster,* dalam penelitian ini data dikelompokan menjadi 3 *cluster.*
2. Melakukan proses standarisasi data.
3. Inisialisasi Koordinat *Centroid*

Pada Inisialisasi Koordinat *Centroid* terbentuk tiga buah *centroid* pada tahap inisialisasi ini yaitu kecamatan Cangkringan, kecamatan Depok dan kecamatan Ngemplak.

Tabel 2 Inisialisasi Pusat (*Centroid*) awal Metode *K-Means*.

Table

Description automatically generated with medium confidence

1. Menghitung jarak dari *Centroid*

Perhitungan jarak setiap data yang ada terhadap setiap pusat *cluster*. Berikut perhitungan jarak setiap data ke-1 dan data ke-2 dengan *centroid* pertama (c1) menggunakan rumus jarak *Euclidean* :

Graphical user interface, text, application

Description automatically generated

Lakukan perhitungan sampai kecamatan ke-17. Selanjutnya dilakukan perhitungan jarak pusat *cluster* kedua (c2\*). Berikut ini adalah perhitungan jarak setiap variabel pada kecamatan (obyek) pusat *cluster* kedua (c2\*).

Graphical user interface, application, Word

Description automatically generated

Lakulan perhitungan sampai kecamatan ke-17. Selanjutnya dilakukan perhitungan jarak pusat *cluster* ketiga (c3\*). Berikut ini adalah perhitungan jarak setiap variabel pada kecamatan (obyek) pusat *cluster* kedua (c3\*).

Graphical user interface, application, Word

Description automatically generated

dan seterusnya perhitungan jarak dengan pusat *cluster*  ketiga (c3\*) dilakukan sampai kecamatan ke-17.

Lakukan langkah diatas sampai semua *centroid* konvergen atau stabil, dalam arti semua *centroid* yang dihasilkan dalam iterasi saat ini sama dengan semua *centroid*  pada iterasi sebelumnya.

Kemudian dilakukan pengelompokan data berdasarkan *cluster* masing-masing untuk membentuk *centroid* baru untuk iterasi kedua.

Tabel 3 Pusat Cluster (*Centroid*) Baru

Graphical user interface, application

Description automatically generated

Pada penelitian ini iterasi berhenti sampai iterasi ke-4, karena pada iterasi ke-3 dan ke-4 posisi *cluster* tidak berubah. Sehingga iterasi dihentikan yang menghasikan rata-rata *cluster* atau *centroid* pada Tabel 4

**Tabel 4 Rata-rata setiap *cluster***

Graphical user interface, application, table, Excel

Description automatically generated

## 

## Validasi Hasil *Clustering*

Adapun penentuan *cluster* optimal pada metode *k-means* menggunakan metode elbow, koefisien silhoutte, dan indeks Dunn.

1. Metode *elbow method*

*Elbow Method* menampilkan suatu grafik penurunan dan *cluster* terbaik dilihat dari lekukan yang membentuk siku pada grafik tersebut (Bholowalia, 2014).

Tabel 5 Nilai elbow untuk k=2,3,4.

|  |  |  |  |
| --- | --- | --- | --- |
| Indeks Validasi | k=2 | k=3 | k=4 |
| *Elbow Method* | 44,1749 | 27,6936 | 18,2598 |

Berdasarkan Tabel 5 akan ditampilkan plot metode *elbow* dapat dilihat pada Gambar 4.1.

Chart, line chart

Description automatically generated

Gambar 1 Grafik *Elbow Method*

Penentuan nilai 𝑘 dengan metode wss atau *elbow method* dengan melihat grafik yang pergerakannya mulai landai dari yng pertama setelah curam. Dari Gambar 4.1 dapat dilihat bahwa grafik yang pergerakannya landai pertama yaitu pada angka 3 dan yang membentuk seperti *elbow* (lipatan siku) terdapat pada titik 3, sehingga nilai 𝑘 optimum metode *elbow* adalah 3.

1. *Silhouette coefficient*

Penentuan *k* optimal pada metode *Silhouette coefficient* dengan melihat pergerakan grafik yang landai setelah grafik yang curam.

Tabel 6 Nilai Silhouette Coefficient untuk 𝑘=2,3,4

|  |  |  |  |
| --- | --- | --- | --- |
| Indeks Validasi | k=2 | k=3 | k=4 |
| *Silhouette coefficient* | 0,3940 | 0,3384 | 0,3747 |

Berdasarkan Tabel 6 hasil dari *Silhouette coefficient* memiliki jumlah cluster sebanyak 2 segmen karena memiliki nilai tertinggi diantara yang lain yaitu 0,3940.

Chart, line chart

Description automatically generated

Gambar 2 Grafik *silhouette coefficient*

Dari Gambar 4.2 titik tertinggi atau yang paling optimum adalah 3. Berdasarkan *output* dan Gambartersebut, maka diketahui nilai *silhouette coefficient* yang mendekati 1 berada pada 𝑘=3, sehingga dengan indeks ini pembentukan 3 *cluster* adalah *cluster* terbaik.

1. *Dunn Index*

Nilai *dunn index* yang paling besar merupakan jumlah *cluster* yang optimal.

Tabel 7 Output nilai dunn index untuk 𝑘=2,3,4

|  |  |  |  |
| --- | --- | --- | --- |
| Indeks Validasi | k=2 | k=3 | k=4 |
| *Dunn index* | 0,2912 | 0,3084 | 0,3061 |

Berdasarkan Tabel 4.8 maka nilai *dunn index* yang paling besar berada pada 𝑘=3 yaitu sebesar 0,3084, sehingga dengan indeks ini *cluster* terbaik adalah pada pembentukan 3 *cluster*.

## Analisis Hasil *Clustering*

Hasil analisis *cluster* pada metode *k-means* dengan menggunkan indeks validasi metode *elbow* dan *silhouette coefficient* menunjukkan bahwa jumlah *cluster* terbaik yaitu *k* = 3.

Graphical user interface, application, table, Excel

Description automatically generated

Gambar 3 Rata-rata *cluster* pada masing-masing variabel

Karakteristik dari setiap klaster dapat diketahui dari rata-rata variabel tiap *cluster*, sehingga diperoleh interpretasi *cluster* yaitu *cluster* 1 didominasi oleh penyakit diare dan DBD. Hasil *cluster* 2 di dominasi oleh penyakit diare dan Hepatitis B, *cluster* 3 didominasi oleh penyakit diare dan campak. *Cluster* 1 menempati urutan pertama untuk semua jenis penyakit, *cluster* 2 menepati unrutan kedua kecuali penyakit campak yang berada pada urutan ketiga.

Berdasarkan hasil penelitian menggunakan algoritma k-means diketahui bahwa penyebaran kasus penyakit terbagi dalam tiga kelompok atau tiga *cluster*.

1. *Cluster* 1 terdiri dari 7 kecamatan yaitu kecamatan Godean, Gamping, Mlati, Kalasan,Ngaglik, Ngemplak, dan Sleman.
2. *Cluster* 2 terdiri dari 9 kecamatan yaitu kecamatan Moyudan, Minggir, Turi, Pakem, Cangkringan, Seyegan, Berbh, Prambanan, dan Tempel.
3. *Cluster* 3 terdiri dari 1 kecamatan yaitu kecamatan Depok.

Berdasarkan hasil penelitian yang telah dilakukan oleh Sari dan Nugraha (2016), dengan hasil penelitian yang telah penulis lakukan yaitu wilayah rawan di Kabupaten Sleman Kecamatan Ngaglik merupakan *cluster* 1, kecamatan Minggir, Berbah, Prambanan, Tempel, Turi, Pakem, dan Cangkringan merupakan *cluster* 2, kecamata Depok merupakan *cluster* 3.

Penelitian ini sejalan dengan penelitian Rahmayani (2018) yang menunjukkan bahwa tingkat keparahan penyakit Penyakit yang paling sering diderita adalah penyakit demam berdarah dengue (DBD).

**SIMPULAN**

Berdasarkan hasil pembahasan *Clustering* daerah rawan kasus penyakit di Kabupaten Sleman menggunakan metode *k-means* yang telah dilakukan, maka diperoleh kesimpulan sebagai berikut:

* + - 1. Terdapat 3 *cluster* yang diperoleh dari analisis k-means yang dilakukan. Adapun *cluster* tersebut adalah:
         1. *Cluster* 1 terdiri dari 7 kecamatan yaitu kecamatan Godean, Gamping, Mlati, Kalasan, Ngaglik, Ngemplak, dan Sleman. Pada *cluster* ini dikategorikan daerah dengan tingkat kerawanan rendah.
         2. *Cluster* 2 terdiri dari 9 kecamatan yaitu kecamatan Moyudan, Minggir, turi, Pakem, Cangkringan, Seyegan, Berbah, Prambanan, dan Tempel. Pada *cluster* ini dikategorikan daerah dengan tingkat kerawanan sedang.
         3. *Cluster* 3 terdiri dari 1 kecamatan yaitu kecamatan Depok. Pada *cluster* ini dikategorikan daerah dengan tingkat kerawanan tinggi.
      2. Karakteristik dari setiap *cluster* dapat dilihat dari rata-rata variabel tiap *cluster* yaitu variabel DBD, Diare, TB paru, Hepatitis B, Campak. C*luster* 1 memiliki nilai rata-rata variabelpaling tinggi yang didominasi oleh penyakit diare dan DBD, *cluster* 2 memiliki rata-rata variabel lebih rendah dari *cluster* 1yang didominasi oleh penyakit diare dan Hepatitis B, c*luster* 3 memiliki rata-rata *cluster* paling rendah yang didominasi oleh penyakit diare dan campak.

**UCAPAN TERIMA KASIH**

Terimakasih kepada koordinator Prodi Maatematika dan seluruh Dosen Prodi Matematika yang telah memberikan ilmu dan bimbingan hingga terselesainya artikel ini.

**DAFTAR PUSTAKA**

Achmadi, U. F. (2012). *Manajemen Penyakit berbasis Wilayah*. Jakarta: Rajawali Pres.

Abrianto, P. M. C. (2018). Penerapan Metode K-Means Clustering untuk Pengelompokan Pasien Penyakit Liver. *Jurnal Mahasiswa Teknik Informatika,* 2(2), 247-250

Badan Pusat Statistik ( 2018) *Kabupaten Sleman dalam Angka 2020*. Kabupaten Sleman : Badan Pusat Statistika.

Bailey, L., Vardulaki, K., Langham, J. & Chandramohan, D. (2005). *Introduction to Epidemiology.* Maidenhead: Open University Press.

Depkes Kabupaten Sleman (2020)*. Profil Kesehatan Sleman 2020*. Sleman.

Dewi, I. C, & Pramita, K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Jurnal Matrix.* 9 (3), 102-109

Fluegemann, J. K., Davies, M. D., & Aguirre, N. D. (2011).Determining the Optimal Number of ClustersWith the Clustergram*. NASA USRP – Internship Final Report,* 1-9.

Irwanto, I, Purwananto, Y & Soelaiman, R. (2012). Optimasi Kinerja Algoritma Klasterisasi *K-means* untuk Kuantisasi Warna Citra. *Jurnal Teknik ITS*, 1 (1), 198.

Johnson, Richard A., & Wichern, Dean W. (2002). *Applied Multivariate Statistical Analysis Fifth Edition. Upper Saddle River,* New Jersey: Prentice Hall.

N. Putu, E. M., & Santoso, A. (2015). Analisis Penentuan Jumlah Cluster terbaik pada Metode *K-Means.* *Seminar Nasional Multi Disiplin Ilmu*, 978-979.

Ong, J. O (2013). Implementasi Algoritma *K-Means Clustering* untuk menentukan Strategi Marketing President University. *Jurnal Ilmiah industri,* 12(1), 10.

Petrovic, S. (2006). A comparison between the silhouette index and the davies-bouldin index in labeling IDS clusters. *In 11 th Nordic Workshop on Secure IT-systems*.

Rahmayani, M. T. (2018). Analisis *Clustering* Tingkat Keparahan Penyakit Pasien Menggunkan Algoritma *K-means. Jurusan Informatika FST UPIT*, 1 (2), 42-43.

Santosa, B. C. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis.* Yogyakarta: Graha Ilmu.

Sari, D. K., & Nugraha, J. (2016). Analisis Kelompok Berdasarkan Wilayah Rawan Penyakit Tuberkulosis di Kabupaten Sleman. *Seminar Nasional Pendidikan Matematika Ahmad Dahlan 2016.* 249-254

Sibarani, R., &Chafid. (2018). Algorithma K-means Clustering Strategi Pemasaran Penerimaan Mahasiswa Baru Universitas Satya Negara Indonesia: *Seminar Nasional Cendekiawan,*

Supranto. (2004). *Analisis Multivariat: Arti dan Interpretas.* Jakarta: PT. Asdi Mahasatya.

Susilowati, T. Sugiarto, G. & Mardianto, I. (2020). Uji Validasi *Algoritme* *Self-Organizing Map (SOM)* dan *K-means* untuk Pengelompokan Pengawai. *Jurnal RESTI,* 4 (6), 1771-1178.

Vatimatunnimah, V. N. (2013). *Epidemiologi Penyakit Menular dan Penyakit Tidak Menular*. Jakarta : Rineka Cipta.