



# Penerapan K-Means Clustering Untuk Pengelompokan Penyebaran Demam Berdarah Dengue (DBD) di Kabupaten Deli Serdang

Erica Melysa Sembiring

Program Studi Teknik Informatika STMIK Budi Darma, Medan, Indonesia

Email: erica.89@gmail.com

**Abstrak**—DBD merupakan penyakit yang penularannya cepat meluas. Biasanya jika disuatu daerah ada yang terkena demam berdarah, besar kemungkinan akan menyebar ke orang lain didaerah tersebut. Karena banyaknya penderita DBD, maka banyak pula data yang terkumpul dan perlu dilakukan pengolahan pada data tersebut, seperti pengelompokan data penderita DBD dengan tujuan memfokuskan pengendalian vektor diwilayah yang rawan DBD. Daerah tersebut akan menjadi prioritas utama untuk melaksanakan sosialisasi terkait penanggulangan DBD. Data mining merupakan serangkaian proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Salah satu fungsi dari proses data mining untuk menemukan kelompok atau identifikasi kelompok adalah Pengklasteran. Terdapat dua jenis metode clustering, yaitu hierarchical clustering dan non-hierarchical clustering yang biasa disebut Kmeans. K-Means Clustering dimulai dengan menentukan jumlah cluster terlebih dahulu.. Hasil yang didapatkan dari mengelompokan data tersebut adalah berupa daerah yang memiliki potensi penyakit DBD tertinggi di Deli Serdang. Dengan menggunakan metode clustering untuk melakukan perhitungan, mampu membantu menyelesaikan permasalahan di Kabupaten Deli Serdang dalam mengelompokan data penderita DBD yang masih dilakukan secara manual.

**Kata Kunci:** Data Mining; Demam Berdarah Dengue; K-Means Clustering

**Abstract**—DBD is a disease that spread rapidly. Usually if there is an area affected by dengue fever, it is likely to spread to other people in the area. Due to the large number of DBD sufferers, so much data is collected and processing needs to be done on these data, such as the grouping of DBD sufferers data with the aim of focusing vector control in areas that are vulnerable to DBD. The area will be the main priority to carry out socialization related to the handling of DBD. Data mining is a series of processes to get useful information from large database warehouses. One function of the data mining process for finding groups or group identification is Clustering. There are two types of clustering methods, namely hierarchical clustering and non-hierarchical clustering, commonly called Kmeans. K-Means Clustering begins by determining the number of clusters first. The results obtained from grouping the data are in the form of areas that have the highest DBD potential in Deli Serdang. By using clustering method to do calculations, it can help solve problems in Deli Serdang Regency in classifying the data of DBD sufferers which is still done manually.

**Keywords:** Data Mining; Dengue Hemorrhagic Fever; K-Means Clustering

## 1. PENDAHULUAN

Pada saat ini, kita hidup dizaman modernisasi. Moredenisasi sendiri merujuk pada bentuk transformasi dari keadaan yang kurang maju atau kurang berkembang kearah yang lebih baik dengan harapan membuat kehidupan menjadi lebih baik. Dizaman modrenisasi seperti sekarang, manusia sangat bergantung pada teknologi.perkembangan teknologi yang begitu pesat merambahke berbagai sektor termasuk kesehatan. Dengan bantuan teknologi, dapat membantu dalambanyak pekejaan dibidang medis. Selain memiliki potensi dalam memfilter data dan mengolah menjadi informasi, juga mampu menyimpan dalm jumlah kapasitas jauh lebih besar. Konvergensi dengan teknologi informasi juga memungkinkan data kesehatan di-share dengan mudah dan cepat

Demam Berdarah Dengue (DBD) merupakan penyakit infeksi yang disebabkan oleh virus Dengue yang menyebabkan demam dan menimbulkan pendarahan.Seseorang bisa terkena Demam Berdarah Dengue (DBD) apabila digigit oleh nyamuk (terutama) Aedes Aegypti yang mengandung virus Dengue.Masyarakat awam kadang sulit mendeteksi lebih awal diagnosis DBD ini karena gejalanya amat luas.DBD sering dikelirukan dengan penyakit lainnya.Gejala khas seperti pendarahan pada kulit atau pendarahan pada kulit atau tanda pendarahan lainnya kadang terjadi di akhir periode penyakit.Tragisnya bila penyait ini terlambat didiagnosis, maka dapat berujung kematian.Menurut data Badan Kesehatan Dunia (WHO), Indonesia termasuk dalam negara dengan kasus DBD terbesar di dunia.Kejadian demam berdarah cenderung megalami peningkatan dan menyebar dengan cepat.Pasien yang terjangkit DBD bukanlah yang sepele karna bisa meluas dengan cepat.

DBD merupakan penyakit yang menjadi masalah kesehatan masyarakat dihampir seluruh Kabupaten Deli Sredang. Dari data kasus yang ada di Dinas Kesehatan Deli Serdang, angka kesakitan di Deli Serdang cukup tinggi, meskipun angka kematian masih rendah.Karena begitu banyak penderita DBD tentu banyak pula data yang terkumpul.Pengelompokkan dilakukan dengan tujuan memfokuskan pengendalian vektor di wilayah tertentu yang memiliki potensi penyakit DBD tertinggi di Deli Serdang. Daerah tersebut dapat menjadi sasaran untuk melaksanakan sosialisasi terkait kesadaran hidup bersih dan sehat.

Analisis untuk mengelompokkan data yang dapat dilakukan pada data terhadap penyakit DBD adalah Clustering. Clustering mencari kemiripan suatu data terhadap data lainnya selanjutnya dimasukkan dalam suatu cluster yang sama. Clustering mengelompokkan data berdasarkan kesamaan karakteristik. Clustering memiliki beberapa metode diantaranya metode hierarchical dan metode non- hierarchical [1]. Salah satu metode hierarchical antara lain adalah metode K-Means. K-Means adalah algoritma yang populer dengan teknik yang sederhana. K-Means melakukan pencarian kemiripan dengan melakukan iterasi sampai terjadi kovergen[2].



## 2. METODOLOGI PENELITIAN

### 2.1 Data Mining

Data mining sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Data Mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode Data Mining ini dapat digunakan untuk mengambil keputusan di masa depan. Data Mining ini juga dikenal dengan istilah pattern recognition. Data mining merupakan metode pengolahan data berskala besar oleh karena itu data mining ini memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi. Secara umum kajian data mining membahas metode-metode seperti, clustering, klasifikasi, regresi, seleksi variable, dan market basket analisis [1], [3], [4].

### 2.2 Algoritma K-Means Clustering

K-means clustering merupakan salah satu metode data clustering non-hirarkiyang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-datayang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompokdan data yang memiliki karakteristik yang berbeda dikelompokkan dengancluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompokmemiliki tingkat variasi yang kecil [5]–[7].

Implementasi K=Means Clusteringtelah digunakan untuk beberapa penelitian. Diantaranya untuk menentukan potensi siswa dalam dunia kerja berdasarkan kecerdasan emosi.

Langkah-langkah melakukan clustering denganmetode K-Means [5], [8], sebagai berikut:

- Pilih jumlah cluster k.
- Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat clusterdiberiduberi nilai awal dengan angka-angka random.
- Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objekditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatansuatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster.Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak palingantara satu data dengan satu cluster tertentu akan menentukan suatu data masukdalam cluster mana. Untuk menghiutng jarak semua data ke setiap titik pusatcluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagaiberikut:

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \dots (1)$$

Dimana[3]:

$D(i, j)$  = Jarak data ke-i kepusat cluster j

$X_{ki}$  = data ke-i pada atribut data ke-k

$X_{kj}$  = titik pusat ke-j pada atribut ke-k

- 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 di pakai.

Tugaskan lagi setiap objek memakai kcluster yang bary. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi.

## 3. HASIL DAN PEMBAHASAN

Demam Berdarah merupakan infeksi yang disebabkan oleh virus dengue dan ditularkan melalui nyamuk Aedes. Penyebaran DBD dipengaruhi oleh kondisi lingkungan. Kondisi lingkungan tersebut mencakup kepadatan penduduk, mobilitas penduduk, dn juga cuaca

Deli Serdang masuk dalam daerah endemis DBD karena terdapat banyak penderita yang terserang penyakit DBD. Penyebaran penyakit DBD ini terbesar di Kecamatan Sunggal dengan 197 penderita dan 12 meninggal dunia, Tanjung Morawa 177 penderita dengan dua meninggal dunia, Percut Seituan 163 penderita dengan empat meninggal dunia, Hamparanperak 81 penderita dengan dua meninggal dunia, Pancurbatu 80 penderita dengan satu meninggal dunia, Lubukpakam 45 penderita dengan dua meninggal dunia, Batangkuis 45 penderita namun tidak ada yang meninggal dunia, Patumbak 37 penderita dengan satu meninggal dunia dan Namorambe 30 penderita dengan satu meninggal dunia.

Dalam mengolah data kasus demam berdarah, Dinas kesehatan perlu mengelompokkan data-data tersebut agar mudah diolah. Dari data yang terkumpul, setiap bulannya kasus demam berdarah mengalami peningkatan dan juga penurunan pada setiap daerah. Dengan banyaknya data yang tercatat, pegawai sulit untuk menentukan daerah dengan jumlah penderita demam berdarah terbanyak, dalam dunia medis disebut dengan daerah atau wilayah endemis. Maka dalam hal ini diperlukan pengolahan data yang lebih lanjut untuk menemukan pengetahuan baru yng dapat membantu para pegawai dinas kesehatan Deli Serdang dalam mengolah data.

Data yang digunakan untuk membentuk cluster adalah data kasus demam berdarah tahun 2016



Tabel 1. Data Kasus DBD Kabupaten Deli Serdang tahun2016

Table with 13 columns (Nama Puskesmas, Jan, Feb, Mar, Apr, Mei, Jun, Jul, Ags, Sep, Okt, Nov, Des) and 20 rows of data for various health centers.

Melakukan penyelesaian untuk mengelompokkan data kasus demam berdarah dengan menggunakan algoritma K-Means Clustering. Proses penyelesaian pengelompokkan demam berdarah adalah sebagai berikut:

1. Tentukan pusat awal cluster "centroid"

Diasumsikan:

Data ke-3 sebagai pusat cluster ke-1:

(3,0,6,0,4,0,1,0,2,0,1,0,2,0,4,0,0,0,3,0,2,0,0,0)

Data ke-10 sebagai pusat cluster ke-2:

(9,0,0,0,1,0,5,0,1,0,6,0,7,0,4,0,2,0,7,0,12,0,10,0)

2. Perhitungan jarak pusat cluster

Digunakan Euclidian distancesehingga didapatkan matrik jarak sebagai berikut:

a. Perhitungan jarak dari data ke-1 terhadap cluster

C1=sqrt((3-4)^2+(0-0)^2+(6-0)^2+(0-0)^2+(4-0)^2+(0-0)^2+(1-2)^2+(0-0)^2+(2-7)^2+(0-0)^2+(1-2)^2+(0-0)^2+(2-6)^2+(0-0)^2+(4-8)^2+(0-0)^2+(0-0)^2+(0-0)^2+(3-0)^2+(0-0)^2+(2-0)^2+(0-0)^2+(0-0)^2+(0-0)^2 = 11,18

C2=sqrt((9-4)^2+(0-0)^2+(0-0)^2+(0-0)^2+(1-0)^2+(0-0)^2+(5-2)^2+(0-0)^2+(1-7)^2+(0-0)^2+(6-2)^2+(0-0)^2+(7-6)^2+(0-0)^2+(4-8)^2+(0-0)^2+(2-0)^2+(0-0)^2+(7-0)^2+(0-0)^2+(12-0)^2+(0-0)^2+(10-0)^2+(0-0)^2 = 20,02

b. Perhitungan jarak dari data ke-2 terhadap cluster

C1=sqrt((3-6)^2+(0-0)^2+(6-9)^2+(0-0)^2+(4-7)^2+(0-0)^2+(1-4)^2+(0-0)^2+(2-2)^2+(0-0)^2+(1-3)^2+(0-0)^2+(2-1)^2+(0-0)^2+(4-7)^2+(0-0)^2+(0-6)^2+(0-0)^2+(3-4)^2+(0-0)^2+(2-4)^2+(0-0)^2+(0-0)^2+(0-0)^2 = 9,53

C2=sqrt((9-6)^2+(0-0)^2+(0-9)^2+(0-0)^2+(1-7)^2+(0-0)^2+(5-4)^2+(0-0)^2+(1-2)^2+(0-0)^2+(6-3)^2+(0-0)^2+(7-1)^2+(0-0)^2+(4-7)^2+(0-0)^2+(2-6)^2+(0-0)^2+(7-4)^2+(0-0)^2+(12-4)^2+(0-0)^2+(10-0)^2+(0-0)^2 = 19,26

c. Perhitungan jarak dari data ke-3 terhadap cluster

C1=sqrt((3-3)^2+(0-0)^2+(6-6)^2+(0-0)^2+(4-4)^2+(0-0)^2+(1-1)^2+(0-0)^2+(2-2)^2+(0-0)^2+(1-1)^2+(0-0)^2+(2-2)^2+(0-0)^2+(4-4)^2+(0-0)^2+(0-0)^2+(0-0)^2+(3-3)^2+(0-0)^2+(2-2)^2+(0-0)^2+(0-0)^2+(0-0)^2 = 0

C2=sqrt((9-3)^2+(0-0)^2+(0-6)^2+(0-0)^2+(1-4)^2+(0-0)^2+(5-1)^2+(0-0)^2+(1-2)^2+(0-0)^2+(6-1)^2+(0-0)^2+(7-2)^2+(0-0)^2+(4-4)^2+(0-0)^2+(2-0)^2+(0-0)^2+(7-3)^2+(0-0)^2+(12-2)^2+(0-0)^2+(10-0)^2+(0-0)^2 = 19,18

d. Perhitungan jarak dari data ke-4 terhadap cluster

C1=sqrt((3-0)^2+(0-0)^2+(6-2)^2+(0-0)^2+(4-8)^2+(0-0)^2+(1-2)^2+(0-0)^2+(2-0)^2+(0-0)^2+(1-1)^2+(0-0)^2+(2-0)^2+(0-0)^2+(4-0)^2+(0-0)^2+(0-4)^2+(0-0)^2+(3-0)^2+(0-0)^2+(2-0)^2+(0-0)^2+(0-0)^2+(0-0)^2 = 9,74

C2=sqrt((9-0)^2+(0-0)^2+(0-2)^2+(0-0)^2+(1-8)^2+(0-0)^2+(5-2)^2+(0-0)^2+(1-0)^2+(0-0)^2+(6-1)^2+(0-0)^2+(7-0)^2+(0-0)^2+(4-0)^2+(0-0)^2+(2-4)^2+(0-0)^2+(7-0)^2+(0-0)^2+(12-0)^2+(0-0)^2+(10-0)^2+(0-0)^2 = 23,04

e. Perhitungan jarak dari data ke-5 terhadap cluster

C1=sqrt((3-0)^2+(0-0)^2+(6-0)^2+(0-0)^2+(4-0)^2+(0-0)^2+(1-6)^2+(0-0)^2+(2-0)^2+(0-0)^2+(1-3)^2+(0-0)^2+(2-2)^2+(0-0)^2+(4-4)^2+(0-0)^2+(0-0)^2+(0-0)^2+(3-0)^2+(0-0)^2+(2-0)^2+(0-0)^2+(0-0)^2+(0-0)^2 = 10,34

C2=sqrt((9-0)^2+(0-0)^2+(0-9)^2+(0-0)^2+(1-0)^2+(0-0)^2+(5-6)^2+(0-0)^2+(1-0)^2+(0-0)^2+(6-3)^2+(0-0)^2+(7-2)^2+(0-0)^2+(4-4)^2+(0-0)^2+(2-0)^2+(0-0)^2+(7-0)^2+(0-0)^2+(12-0)^2+(0-0)^2+(10-0)^2+(0-0)^2 = 20,37

f. Perhitungan jarak dari data ke-6 terhadap cluster

C1=sqrt((3-15)^2+(0-0)^2+(6-9)^2+(0-0)^2+(4-11)^2+(0-0)^2+(1-10)^2+(0-0)^2+(2-5)^2+(0-0)^2+(1-3)^2+(0-0)^2+(2-10)^2+(0-0)^2+(4-0)^2+(0-0)^2+(0-23)^2+(0-1)^2+(3-10)^2+(0-0)^2+(2-15)^2+(0-0)^2+(0-16)^2+(0-0)^2 = 37,38

C2=sqrt((9-15)^2+(0-0)^2+(0-9)^2+(0-0)^2+(1-11)^2+(0-0)^2+(5-10)^2+(0-0)^2+(1-5)^2+(0-0)^2+(6-3)^2+(0-0)^2+(7-10)^2+(0-0)^2+(4-0)^2+(0-0)^2+(0-23)^2+(0-1)^2+(7-10)^2+(0-0)^2+(12-15)^2+(0-0)^2+(10-16)^2+(0-0)^2 = 28,07

g. Perhitungan jarak dari data ke-7 terhadap cluster



- $C1 = \sqrt{(3-11)^2 + (0-0)^2 + (6-7)^2 + (0-0)^2 + (4-7)^2 + (0-0)^2 + (1-6)^2 + (0-0)^2 + (2-7)^2 + (0-0)^2 + (1-6)^2 + (0-0)^2 + (2-5)^2 + (0-0)^2 + (4-9)^2 + (0-0)^2 + (0-8)^2 + (0-0)^2 + (3-4)^2 + (0-0)^2 + (2-3)^2 + (0-0)^2 + (0-3)^2 + (0-0)^2} = 16,06$   
 $C2 = \sqrt{(9-11)^2 + (0-0)^2 + (0-7)^2 + (0-0)^2 + (1-7)^2 + (0-0)^2 + (5-6)^2 + (0-0)^2 + (1-7)^2 + (0-0)^2 + (6-6)^2 + (0-0)^2 + (7-5)^2 + (0-0)^2 + (4-9)^2 + (0-0)^2 + (2-8)^2 + (0-0)^2 + (7-4)^2 + (0-0)^2 + (12-3)^2 + (0-0)^2 + (10-3)^2 + (0-0)^2} = 18,16$
- h. Perhitungan jarak dari data ke-8 terhadap cluster
- $C1 = \sqrt{(3-7)^2 + (0-0)^2 + (6-0)^2 + (0-0)^2 + (4-12)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (2-5)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (2-5)^2 + (0-0)^2 + (4-0)^2 + (0-0)^2 + (0-3)^2 + (0-0)^2 + (3-4)^2 + (0-0)^2 + (2-5)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2} = 13,07$   
 $C2 = \sqrt{(9-7)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-12)^2 + (0-0)^2 + (5-0)^2 + (0-0)^2 + (1-5)^2 + (0-0)^2 + (6-0)^2 + (0-0)^2 + (7-5)^2 + (0-0)^2 + (4-0)^2 + (0-0)^2 + (2-3)^2 + (0-0)^2 + (7-4)^2 + (0-0)^2 + (12-5)^2 + (0-0)^2 + (10-0)^2 + (0-0)^2} = 19,51$
- i. Perhitungan jarak dari data ke-9 terhadap cluster
- $C1 = \sqrt{(3-0)^2 + (0-0)^2 + (6-0)^2 + (0-0)^2 + (4-0)^2 + (0-0)^2 + (1-2)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (4-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (3-5)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2} = 9,74$   
 $C2 = \sqrt{(9-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (5-2)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (6-0)^2 + (0-0)^2 + (7-0)^2 + (0-0)^2 + (4-0)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (7-5)^2 + (0-0)^2 + (12-0)^2 + (0-0)^2 + (10-0)^2 + (0-0)^2} = 21,09$
- j. Perhitungan jarak dari data ke-10 terhadap cluster
- $C1 = \sqrt{(3-9)^2 + (0-0)^2 + (6-0)^2 + (0-0)^2 + (4-1)^2 + (0-0)^2 + (1-5)^2 + (0-0)^2 + (2-1)^2 + (0-0)^2 + (1-6)^2 + (0-0)^2 + (2-7)^2 + (0-0)^2 + (4-4)^2 + (0-0)^2 + (0-2)^2 + (0-0)^2 + (3-7)^2 + (0-0)^2 + (2-12)^2 + (0-0)^2 + (0-10)^2 + (0-0)^2} = 19,18$   
 $C2 = \sqrt{(9-9)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (5-5)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (6-6)^2 + (0-0)^2 + (7-7)^2 + (0-0)^2 + (4-4)^2 + (0-0)^2 + (2-2)^2 + (0-0)^2 + (7-7)^2 + (0-0)^2 + (12-12)^2 + (0-0)^2 + (10-10)^2 + (0-0)^2} = 0$
- k. Perhitungan jarak dari data ke-11 terhadap cluster
- $C1 = \sqrt{(3-0)^2 + (0-0)^2 + (6-1)^2 + (0-0)^2 + (4-2)^2 + (0-0)^2 + (1-2)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (4-11)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (3-4)^2 + (0-0)^2 + (2-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2} = 9,89$   
 $C2 = \sqrt{(9-0)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (1-2)^2 + (0-0)^2 + (5-2)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (6-1)^2 + (0-0)^2 + (7-0)^2 + (0-0)^2 + (4-11)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (7-4)^2 + (0-0)^2 + (12-1)^2 + (0-0)^2 + (10-0)^2 + (0-0)^2} = 21,21$
- l. Perhitungan jarak dari data ke-12 terhadap cluster
- $C1 = \sqrt{(3-0)^2 + (0-0)^2 + (6-0)^2 + (0-0)^2 + (4-0)^2 + (0-0)^2 + (1-3)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (4-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (3-0)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2} = 10,14$   
 $C2 = \sqrt{(9-0)^2 + (0-0)^2 + (0-9)^2 + (0-0)^2 + (1-7)^2 + (0-0)^2 + (5-4)^2 + (0-0)^2 + (1-2)^2 + (0-0)^2 + (6-3)^2 + (0-0)^2 + (7-1)^2 + (0-0)^2 + (4-7)^2 + (0-0)^2 + (2-6)^2 + (0-0)^2 + (7-4)^2 + (0-0)^2 + (12-4)^2 + (0-0)^2 + (10-0)^2 + (0-0)^2} = 22,02$
- m. Perhitungan jarak dari data ke-13 terhadap cluster
- $C1 = \sqrt{(3-0)^2 + (0-0)^2 + (6-0)^2 + (0-0)^2 + (4-0)^2 + (0-0)^2 + (1-0)^2 + (1-0)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (4-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (3-0)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2} = 10$   
 $C2 = \sqrt{(9-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (5-0)^2 + (0-0)^2 + (1-0)^2 + (1-0)^2 + (0-0)^2 + (6-0)^2 + (0-0)^2 + (7-0)^2 + (0-0)^2 + (4-0)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (7-0)^2 + (0-0)^2 + (12-0)^2 + (0-0)^2 + (10-0)^2 + (0-0)^2} = 22,49$
- n. Perhitungan jarak dari data ke-14 terhadap cluster
- $C1 = \sqrt{(3-0)^2 + (0-0)^2 + (6-3)^2 + (0-0)^2 + (4-0)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (2-3)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (4-6)^2 + (0-0)^2 + (0-3)^2 + (0-0)^2 + (3-0)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2} = 8,12$   
 $C2 = \sqrt{(9-0)^2 + (0-0)^2 + (0-3)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (5-1)^2 + (0-0)^2 + (1-3)^2 + (0-0)^2 + (6-0)^2 + (0-0)^2 + (7-0)^2 + (0-0)^2 + (4-6)^2 + (0-0)^2 + (2-3)^2 + (0-0)^2 + (7-0)^2 + (0-0)^2 + (12-0)^2 + (0-0)^2 + (10-0)^2 + (0-0)^2} = 22,22$
- o. Perhitungan jarak dari data ke-15 terhadap cluster
- $C1 = \sqrt{(3-4)^2 + (0-0)^2 + (6-12)^2 + (0-0)^2 + (4-10)^2 + (0-0)^2 + (1-6)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (1-6)^2 + (0-0)^2 + (2-4)^2 + (0-0)^2 + (4-3)^2 + (0-0)^2 + (0-8)^2 + (0-0)^2 + (3-5)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (0-3)^2 + (0-0)^2} = 14,59$   
 $C2 = \sqrt{(9-4)^2 + (0-0)^2 + (0-12)^2 + (0-0)^2 + (1-10)^2 + (0-0)^2 + (5-6)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (6-6)^2 + (0-0)^2 + (7-4)^2 + (0-0)^2 + (4-3)^2 + (0-0)^2 + (2-8)^2 + (0-0)^2 + (7-5)^2 + (0-0)^2 + (12-0)^2 + (0-0)^2 + (10-3)^2 + (0-0)^2} = 22,24$

**Tabel 2.** Hasil perhitungan jarak

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
C	11,1				10,3	37,3	16,0	13,0		19,1		10,1			14,5
1	8	9,53	0	9,74	4	8	6	7	9,74	8	9,89	4	10	8,12	9
C	20,0	19,2	19,1	23,0	20,3	28,0	18,1	19,5	21,0		21,2	22,0	22,4	22,2	22,2
2	2	6	8	4	7	7	6	1	9	0	1	2	9	2	4

**Tabel 3.** Pengelompokan data

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
C1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
C2	1	0	0	1	1	1	0	0	1	0	1	1	1	1	1

Menentukan pusat Cluster baru, karena C1 memiliki 1 anggota barumaka perhitungan cluster baru menjadi:

$C1 = \frac{15}{1}, \frac{0}{1}, \frac{9}{1}, \frac{0}{1}, \frac{11}{1}, \frac{0}{1}, \frac{10}{1}, \frac{0}{1}, \frac{5}{1}, \frac{0}{1}, \frac{3}{1}, \frac{0}{1}, \frac{10}{1}, \frac{0}{1}, \frac{0}{1}, \frac{23}{1}, \frac{1}{1}, \frac{10}{1}, \frac{0}{1}, \frac{15}{1}, \frac{0}{1}, \frac{16}{1}, \frac{0}{1}$

1 1

$C1 = 15;0;9;0;11;0;10;0;5;0;3;0;10;0;0;23;1;10;0;15;0;16;0$



Karena C2 memiliki 10 anggota maka perhitungan cluster baru menjadi:

$$C2 = \frac{(4+0+0+15+0+0+0+0+0+4)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

$$\frac{(0+2+0+9+0+1+0+0+3+12)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

$$\frac{(0+8+0+11+0+2+0+0+0+10)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

$$\frac{(2+2+6+10+2+2+3+0+1+6)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

$$\frac{(7+0+0+5+0+0+0+0+3+0)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

$$\frac{(2+1+3+3+0+1+0+0+0+6)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

$$\frac{(6+0+2+10+0+0+0+0+0+4)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

$$\frac{(8+0+4+0+0+11+0+0+6+3)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

$$\frac{(0+4+0+23+0+0+0+0+3+8)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

$$\frac{(0+0+0+10+5+4+0+0+0+5)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

$$\frac{(0+0+0+15+0+1+0+0+0+0)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

$$\frac{(0+0+0+16+0+0+0+0+0+3)}{10} ; \frac{(0+0+0+0+0+0+0+0+0)}{10} ;$$

C2 = 2,3;0;2,7;0;3,1;0;3,4;0;1,5;0;1,6;0;2,2;0;3,2;0;3,8;0;2,4;0;1,6;0;1,9;0

### 3.1 Pengujian

Pada bagian ini merupakan hasil pengujian untuk mengelompokkan data penderita Demam Berdarah Dengue (DBD) dengan aplikasi Weka 3.8.1. Hasil pengujian ini sama hasilnya seperti cara manual. Sampai di sini hasil langkah-langkah penyelesaian untuk mengelompokkan data Penderita Demam Berdarah Dengue (DBD). Pada tabel 4 dibawah menjelaskan hasil penyelesaian secara manual.

**Tabel 4.** Data Hasil Pengelompokan Demam Berdarah

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
C1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
C2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

## 4. KESIMPULAN

Dari penelitian yang penulis lakukan, maka disimpulkan penggunaan metode clastering untuk melakukan perhitungan, ternyata mampu menyelesaikan permasalahan pada Kabupaten Deli Serdang dalam mengelompokkan data penderita Demam Berdarah Dengue (DBD). Aplikasi weka 3.8.1 digunakan sebagai aplikasi pendukung dalam menguji hasil dari pengelompokan data penderita DBD yang didapatkan secara manual.

## REFERENCES

- [1] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish, 2020.
- [2] F. Tambunan, "Implementation of Data Mining using the Clustering Method (Case: Region of the Actors of Theft Crime by Province)," *IJISTECH (International J. Inf. Syst. Technol.)*, vol. 2, no. 2, p. 75, 2019.
- [3] A. P. Windarto, U. Indriani, M. R. Raharjo, and L. S. Dewi, "Bagian 1 : Kombinasi Metode Klastering dan Klasifikasi ( Kasus Pandemi Covid-19 di Indonesia )," *J. Media Inform. Budidarma*, vol. 4, no. 3, pp. 855-862, 2020.
- [4] A. Widiastari, Solikhun, and Irawan, "Analisa Datamining dengan Metode Klasifikasi C4 . 5 Sebagai Faktor Penyebab Tanah Longsor," *J. Comput. Syst. Informatics*, vol. 2, no. 3, pp. 247-255, 2021.
- [5] H. Sulastris and A. I. Gufroni, "Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia," *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 299-305, 2017.
- [6] R. W. Sari and D. Hartama, "Data Mining : Algoritma K-Means Pada Pengelompokkan Wisata Asing ke Indonesia Menurut Provinsi," *Semin. Nas. Sains Teknol. Inf.*, pp. 322-326, 2018.
- [7] A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, p. 26, 2017.
- [8] K. D. R. Sianipar, S. W. Siahaan, M. Siregar, and P. P. P. A. N. W. F. I. R. H. Zer, "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Online," *Infomatek*, vol. 22, no. 1, pp. 23-30, 2020.