
PENERAPAN METODE ALGORITMA K-MEANS CLUSTERING UNTUK PEMETAAN PENYEBARAN PENYAKIT DEMAM BERDARAH DENGUE (DBD)

**Muhammad Ardiansyah Sembiring, Raja Tama Andri Agus,
Mustika Fitri Larasati Sibuea
Sekolah Tinggi Manajemen Informatika dan Komputer Royal
e-mail: adinmantap88@gmail.com**

Abstract: *Dengue hemorrhagic fever is generally caused by the Aedes Aegypti mosquito. These mosquitoes usually live in dirty environments. This study aims to apply the concept of data mining using the K-Means algorithm (clustering) for mapping villages affected by Dengue Hemorrhagic Fever. Based on the results of the study, there were 3 clusters for the level of spread of Dengue hemorrhagic fever disease, namely cluster 1 with medium category, cluster 2 with high category and cluster 3 with low category. The resulting performance using Devies Bouldin Index (DBI) is 1.044 which shows K=3 as the most optimum number of groups compared to other K. The results of the mapping carried out can help the government act quickly in making decisions for areas prone to dengue fever in Setia Promise District.*

Keywords: *Demam Berdarah Dengue (DBD), Data Mining, K-Means Clustering.*

Abstrak: DBD umumnya disebabkan oleh nyamuk Aedes Aegypti. Nyamuk tersebut biasa hidup di lingkungan yang kotor. Penelitian ini bertujuan untuk menerapkan konsep data mining menggunakan algoritma K-Means (clustering) untuk pemetaan desa yang terjangkit Demam Berdarah Dengue. Pemetaan yang dilakukan untuk mencari daerah yang banyak terjangkit penyakit DBD, dan daerah yang jarang terjangkit penyakit DBD Berdasarkan hasil penelitian terdapat 3 cluster untuk tingkat penyebaran penyakit DBD yakni cluster 1 dengan kategori sedang, cluster 2 dengan kategori tinggi dan cluster 3 dengan kategori rendah. Performance yang dihasilkan menggunakan Devies Bouldin Index (DBI) adalah sebesar 1.044 yang menunjukkan K=3 sebagai jumlah kelompok yang paling optimum dibandingkan K lainnya. Hasil pemetaan yang dilakukan dapat membantu pemerintah bertindak cepat dalam mengambil keputusan bagi daerah yang rawan terjangkit penyakit DBD di Kecamatan Setia Janji.

Kata Kunci: Demam Berdarah Dengue (DBD), Data Mining, K-Means Clustering.

PENDAHULUAN

Demam Berdarah Dengue (DBD) adalah penyakit yang terdapat pada anak-anak dan dewasa dengan gejala utama demam, nyeri otot dan sendi, yang biasanya memburuk setelah dua hari pertama. Penyakit DBD merupakan salah satu masalah kesehatan masyarakat di Indonesia yang jumlah penderitanya cenderung meningkat dan penyebarannya

semakin luas, sejalan dengan meningkatnya *mobilitas* dan kepadatan penduduk. DBD umumnya ditemukan di daerah tropis dan sub-tropis.

Penyakit DBD adalah penyakit yang disebabkan oleh virus dengue yang dibawa oleh nyamuk *Aedes Aegypti* betina lewat air liur gigitan saat menghisap darah manusia yang menyebabkan gangguan pada sistem pembekuan darah, sehingga

mengakibatkan pendarahan. Penyakit DBD ditandai dengan ciri demam tinggi mendadak selama 2-7 hari tanpa sebab yang jelas disertai manifestasi seperti petekie, epistaxis kadang disertai muntah darah, berak darah, kesadaran menurun, bertandensi menimbulkan senjata (shock) dan kematian.

Virus *Dengue* dapat masuk ke tubuh manusia melalui gigitan nyamuk dari genus *Aedes* seperti *Aedes Aegypti* betina dan *Aedes Albopictus*. *Aedes Aegypti* adalah vektor yang paling banyak ditemukan penyebab penyakit DBD. Penyakit DBD merupakan penyakit endis di Indonesia dan sampai saat ini masih merupakan masalah utama kesehatan masyarakat. Hampir di seluruh pelosok Indonesia sering terjangkau penyakit DBD. Sebagian menganggap hal ini terjadi karena kurangnya kesadaran akan kebersihan lingkungan dan sebagian lagi menganggap karena pemerintah lambat dalam mengantisifasi dan merespon kasus ini.

Seiring dengan banyaknya daerah yang terjangkau penyakit DBD di Kecamatan Setia Janji maka perlu dilakukan proses pengumpulan data untuk menentukan daerah yang sering terjangkau penyakit DBD. Dengan demikian dibutuhkan sistem yang dapat membantu pengelompokan daerah yang sering terjangkau penyakit DBD. Penyakit DBD menyerang mulai dari anak-anak sampai dengan lansia.

Metode yang digunakan yaitu metode *Algoritma K-Means Clustering*. Metode K-means Clustering merupakan salah satu metode data *clusterig* non-hirarki yang mengelompokan data dalam bentuk satu atau lebih cluster/kelompok. Metode ini adalah sebuah metode pengelompokan yang dapat dengan cepat dan tepat sasaran dalam penentuan daerah yang sering terjangkau penyakit DBD.

Penyakit *Demam Berdarah Dengue* (DBD) adalah penyakit infeksi oleh virus *Dengue* yang ditularkan melalui gigitan nyamuk *Aedes*, dengan ciri demam tinggi mendadak disertai manifestasi pendarahan dan bertandensi

menimbulkan senjata (shock) dan kematian. Vektor utama penyakit DBD di Indonesia adalah nyamuk *Aedes Aegypti* (Fathi&dkk, 2005).

Aedes Aeypti adalah salah satu nyamuk spesies tropis yang sangat efisien sebagai penampung virus, karena sangat ganas dan berkembang biak secara subur pada lingkungan manusia. Satu nyamuk dapat menjangkiti beberapa orang dalam waktu singkat dan lebih dari 1 kali. Penyakit DBD pertama kali ditemukan di Surabaya pada tahun 1958 dimana saat itu sebanyak 58 orang terinfeksi dan 24 orang diantaranya meninggal dunia. Mulai saat itu, penyakit ini pun menyebar luas ke seluruh penjuru Indonesia Perkembangan hidup nyamuk *Aedes Aegypti* dari telur hingga dewasa memerlukan waktu sekitar 10 – 12 hari dan umur nyamuk *Aedes Aegypti* betina berkisar antara 1,5 bulan, tergantung dari suhu kelembaban udara sekelilingnya .

Musim hujan tiba maka perlu diwaspadai adanya genangan genangan – genangan air yang terjadi pada selokan yang buntuh, gorong – gorong yang tidak lancar serta adanya banjir yang berkepanjangan. Selama ini semua manusia pasti mengetahui dan mengenal serangga yang disebut nyamuk. Antara nyamuk dan manusia bisa dikatakan hidupberdampingan bahkan nyaris tanpa batas. Namun, berdampinganya manusia dengan nyamuk bukan dalam makna positif. Tetapi nyamuk dianggap mengganggu kehidupan manusia. Meski jumlah nyamuk yang dibunuh oleh manusia jauh lebih banyak dari pada jumlah manusia yang meninggal karna nyamuk.

Maka perlu dilakukannya pemberantasan vektor penyakit karena vaksin yang efektif masih belum tersedia. Pemberantasan vektor ini meliputi pemberantasan sarang nyamuk dan jentik. Pemberantasan ini meliputi pembersihan tempat penampungan air yang bersih yang merupakan sarana utama perkembangbiakan nyamuk. Tempat air bersih perlu dilindungi dengan ditutup yang baik.

Gejala Penyakit DBD

selama 3 – 15 hari sejak seseorang terinfeksi virus *Dengue*, maka akan timbul gejala sebagai berikut :

1. Demam tinggi yang mendadak 2 - 7 hari (38 – 40 derajat Celsius).
2. Pada pemeriksaan uji tourniquet, tampak adanya jentik (*puspura*) pendarahan.
3. Adanya bentuk pendarahan dikelopak mata bagian dalam (*konjungvita*), Mimisan (*epitaksis*), Buang air besar dengan kotoran (*peaces*) berupa lendir bercampur darah (*melena*), dan lain-lainnya.
4. Terjadi pembesaran hati (*Hepatomegali*).
5. Tekanan darah menurun sehingga menyebabkan syok.
6. Pada pemeriksaan laboratorium (darah) hari ke 3 – 7 terjadi penurunan trombosit dibawah 100.000 /mm³ (*Trombositopeni*), terjadi peningkatan nilai Hematokrit diatas 20% dari nilai normal (*Hemokonsentrasi*).
7. Timbulnya beberapa gejala klinik yang menyertai seperti mual, muntah, penurunan nafsu makan (*Anoreksia*), sakit perut, diare, menggigil, kejang dan sakit kepala.
8. Mengalami pendarahan pada hidung (mimisan) dan gusi.
9. Demam yang dirasakan penderita menyebabkan keluhan pegal/sakit pada persendihan.
10. Munculnya bintik- bintik merah pada kulit akibat pecahnya pembuluh darah.

Data Mining

Data mining merupakan salah satu disiplin ilmu yang digunakan untuk menentukan suatu informasi tertentu dalam sekumpulan data sebagai pendukung pengambilan keputusan. Data mining juga sebagai salah satu proses untuk memperoleh informasi yang memiliki nilai guna dari sekumpulan data (Sembiring, 2018).

Data mining telah dipakai diberbagai bidang seperti ilmu sains, bisnis dan industri, teknik, kesehatan, serta pertahanan dan keamanan. Ketersediaan data yang berlimpah yang dihasilkan dari penggunaan teknologi informasi di hampir semua bidang kehidupan, menimbulkan kebutuhan untuk dapat memanfaatkan informasi dan pengetahuan yang terkandung didalam limbah data tersebut, yang kemudian melahirkan data mining. data mining merupakan proses untuk menemukan pengetahuan yang ditambang dari sekumpulan data yang volumenya sangat besar.

Menurut prasetyo (2012), data mining terbagi menjadi beberapa kelompok yaitu sebagai berikut:

1. Model Prediksi
2. Pemodelan yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang telah didapat.
3. Analisis Kelompok
4. Mengelompokkan sekumpulan data ke dalam kelompok (cluster) tertentu berdasarkan karakteristik yang sama pada masing-masing data.
5. Analisis Asosiasi
6. Menentukan pola yang menggambarkan kekuatan hubungan fitur dalam data.
7. Deteksi Anomaly
8. Pengamatan suatu data dari sekumpulan data yang secara signifikan memiliki karakteristik yang berbeda dari sisa data yang lain.

Algoritma K-Means

K-Means merupakan salah satu metode data *clustering* non hirarki yang berusaha berpartisipasi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok (Mohammad, 2013). Metode ini berpartisipasi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai

karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain (Agusta, 2007).

METODE

Ada banyak metode yang dapat digunakan dalam pengelompokan contohnya metode K-means. Pengelompokan yang dapat digunakan seperti pengelompokan nonhierarki yang membagi data kedalam bentuk dua atau lebih kelompok. K-means merupakan metode analisis kelompok yang mengarah pada pembagian N objek pengamatan kedalam K kelompok (cluster) dan setiap objek pengamatan dimiliki oleh suatu kelompok dengan rata-rata (mean) terdekat (prasetyo, 2012).

Algoritma K-Means dalam pengelompokan data seperti berikut:

1. Menentukan jumlah kelompok
2. Alokasi data ke dalam kelompok secara acak
3. Hitung pusat kelompok (rata-rata) dari data yang terdapat pada masing-masing kelompok.

Dalam menentukan nilai centroid untuk awal iterasi, nilai awal centroid dilakukan secara acak. Sedangkan jika menentukan nilai centroid yang merupakan tahap dari iterasi, maka digunakan rumus sebagai berikut ini :

$$\bar{v}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} ,$$

1. Menghitung jarak antara titik centroid dengan titik tiap objek.

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

2. Pengelompokan object untuk menentukan anggota cluster adalah dengan memperhitungkan jarak minimum objek.
3. Kembali ke tahap ke-2, lakukan perulangan hingga nilai centroid

yang dihasilkan tetap dan anggota cluster tidak berpindah ke cluster lain.

Kelemahan K-Means ;

1. Bila jumlah data tidak terlalu banyak, mudah untuk menentukan cluster awal.
2. Jumlah cluster, sebanyak K, harus ditentukan sebelum dilakukan perhitungan.
3. Tidak pernah mengetahui real cluster dengan menggunakan data yang sama, namun jika dimasukan dengan cara yang berbeda mungkin dapat memproduksi cluster yang berbeda jika jumlah datanya sedikit.
4. Tidak tahu kontribusi dari atribut dalam proses pengelompokan karena dianggap bahwa setiap atribut memiliki bobot yang sama.

HASIL DAN PEMBAHASAN

Adapun langkah-langkah yang harus dilakukan dalam pemetaan wilayah yang sering terjangkit penyakit DBD adalah sebagai berikut:

1. Pemilihan Data
 Dalam tahap pemilihan data yang dijadikan objek penelitian adalah penyebaran penyakit DBD di Kecamatan Setia Janji.

Tabel 1. Data Penyebaran Penyakit DBD Kecamatan Setia Janji

Lokasi	JUMLAH PENDERITA				
	ANAK-ANAK	REMAJA	DEWASA	BALITA	LANSIA
A1	16	13	10	4	5
A2	11	20	7	3	4
A3	10	4	5	3	3
A4	15	9	6	2	0
A5	11	6	2	3	2
A6	15	17	2	1	2
A7	8	5	3	4	2
A8	9	4	2	2	3
A9	11	8	4	2	2
A10	12	4	3	3	4

Jumlah *cluster* merupakan jumlah kelompok yang akan dihasilkan. Dalam penelitian kali ini *cluster* yang akan dibuat adalah 3 *cluster*. Untuk menentukan nilai *centroid* awal dapat diambil secara acak dari objek data yang ada. *Centroid* 1 diambil dari lokasi 2, *Centroid* 2 diambil dari lokasi 1, dan *Centroid* 3 diambil dari lokasi 7. Menghitung jarak data ke *centroid* baru untuk iterasi berikutnya. Dengan menggunakan nilai *cluster* baru, jarak ke masing-masing *centroid* dihitung sampai group baru sama dengan group lama atau group sebelumnya.

Adapun nilai *centroid* yang digunakan untuk mendapatkan *cluster* yang stabil adalah sebagai berikut:

Tabel 2. Centroid

Centroid 1	13.0	18.50	4.50	2.00	3.00
Centroid 2	15.5	11.00	8.00	3.00	2.50
Centroid 3	10.2	5.17	3.17	2.83	2.67

Berdasarkan proses iterasi yang telah dilakukan, *centroid* yang digunakan dalam melakukan proses *cluster* pada penelitian ini.

Tabel 3. Hasil Perhitungan Jarak Data Pada Masing-masing Centroid

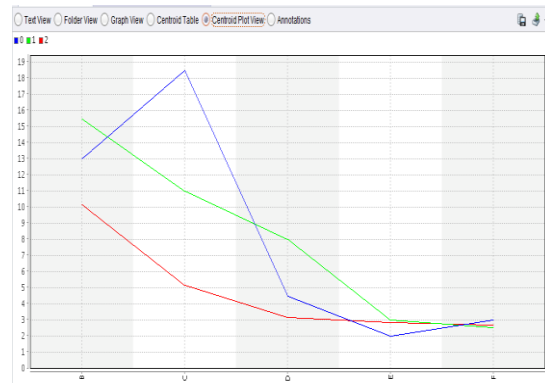
	C1	C2	C3
	9,22	0	13,64
	0	9,22	15,97
	16,19	12,12	3,32
	12,45	7,87	9,06
	15,00	12,17	3,46
	7,62	9,95	14,25
	15,97	13,64	0
	16,94	14,21	2,83
	12,57	9,95	4,90
	16,52	12,17	3,46

Pengelompokan data *cluster* setelah perhitungan jarak data pada *centroid*. Berikut adalah hasil pengelompokan :

Tabel 4. Hasil Pengelompokkan Cluster

Lokasi	C1	C2	C3
1	1	2	0
2	2	1	0
3	0	1	2
4	0	2	1
5	0	1	2
6	2	1	0
7	0	1	2
8	0	1	2
9	0	1	2
10	0	1	2

Dari hasil yang di atas dapat di kelompokkan bahwa *cluster* 1 yaitu desa 2 dan 6. selanjutnya *Cluster* 2 yaitu desa 1 dan 4 dan *cluster* 3 yaitu desa 3,5, 7, 8, 9, 10. Pada gambar 1 dapat menjelaskan di atas sudah terlihat jelas daerah yang terdapat pada setiap *cluster* nya.



Gambar 1. Centroid Plot View

Dari gambar Centroid Plot View atau diagram penggambaran pada *cluster* dapat dilihat dengan jelas daerah mana yang sering terjangkit penyakit Demam Berdarah Dengue (DBD). Daerah yang sering terjangkit DBD yaitu terdapat pada garis warna hijau atau terdapat pada *Cluster* 2, dan daerah yang sedang terdapat pada garis biru atau *Cluster* 1, dan daerah yang jarang terjangkit penyakit DBD yaitu terdapat pada garis berwarna merah atau *Cluster* 3.

KESIMPULAN

1. Dengan menggunakan Metode Algoritma K-Means Clustering, dapat ditemukan desa di Kecamatan Setia janji yang paling sering dan paling banyak masyarakatnya terjangkit penyakit DBD yaitu desa 2 dan 4.
2. Berdasarkan devies bouldin index diketahui performance adalah sebesar 1.044 yang menunjukkan 3 cluster (K=3) adalah jumlah yang optimum dibandingkan dengan jumlah K lainnya.

DAFTAR PUSTAKA

Fithi, & dkk. (2005). Peran Faktor Lingkungan dan Perilaku Terhadap Penularan Demam Berdarah Dengue Di Kota Mataram. *Jurnal Kesehatan Lingkungan*, vol2, no.1, 1-10

- Agusta , Y. (2007). K-Means - Penerapan Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*, vol2, no.1, 1-10
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Andi. Yogyakarta.
- Kusrini dan Luthfi, Emha T, 2009.” *Algoritma Data Mining*”, Penerbit Andi, Yogyakarta.
- Mohamad, I. B., & Usman, D. (2013). Standardization and its effects on K-means clustering algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, 6(17), 3299-3303.
- Sembiring, M. A., Sibuea, M. F. L., & Sapta, A. (2018). Analisa Kinerja Algoritma C. 45 Dalam Memprediksi Hasil Belajar. *Journal Of Science and Social Research*, 1(1), 73-79.