Penerapan Algoritma K-means Pada Pengelompokan Daerah Penderita Filariasis

¹Nai Siti Jenab, ²Hanny Hikmayanti Handayani, ³Yana Cahyana
¹if15.naijenab@mhs.ubpkarawang.ac.id, ²hanny.hikmayanti@ubpkarawang.ac.id,
³yana.cahyana@ubpkarawang.ac.id

Universitas Buana Perjuangan
Karawang, Jawa Barat

ABSTRAK

Banyak jenis penyakit yang disebabkan oleh gigitan nyamuk salah satu penyakitnya yaitu filariasis. Filariasis menyebar di seluruh daerah di Indonesia, dan mengalami peningkatan setiap tahunnya semenjak tahun 2002 sampai tahun 2014. Penyakit ini diprediksi akan terus meningkat setiap tahunnya apabila tidak dilakukan tindakan lebih lanjut. Maka penelitian ini dilakukan untuk mencegah atau penanggulangan dengan menggunakan teknik komputasi yaitu algoritma k-means. Algoritma k-means berfungsi untuk mengelompokkan data sehingga mampu mengelompokkan daerah yang terdapat penderita filariasis untuk menunjang suatu keputusan dalam pencegahan atau penanggulangan. Dari hasil pengelompokan daerah penderita filariasis berdasarkan provinsi menghasilkan tiga clustering. Clustering tinggi menghasilkan dua provinsi (Provinsi Aceh dan provinsi Nusa Tenggara Timur), pada clustering sedang terdapat dua provinsi (Provinsi Papua dan Provinsi Papua Barat) sedangkan yang 29 provinsi lainnya termasuk ke dalam clustering rendah.

Kata kunci : *clustering, filariasis, k-means*

1. PENDAHULUAN

Penyakit filariasis yang lebih diketahui orang-orang sebagai penyakit kaki gajah merupakan salah satu penyakit yang disebabkan oleh nyamuk. [1]. Kasus penderita filariasis di Indonesia terus mengalami peningkatan setiap tahunnya (tahun 2002 – 2014). Namun, berdasarkan laporan dinas kesehatan pada tahun 2015 penyakit filariasis mengalami pengurangan karena beberapa penderita filariasis meninggal dunia. Kemudian jika tidak ada tindakan lebih lanjut maka penderita filariasis akan terus meningkat [2]. Salah satu cara penanggulangannya agar jumlah penderita penyakit filariasis tidak terus meningkat adalah dengan teknik komputasi. Teknik komputasi yang terbukti berhasil menyelesaikan beberapa masalah berdasarkan pengelompokan salah satunya adalah k-means [1,2,3,4].

Penelitian algoritma k-means oleh Fatmawati dan Windarto [3] mengelompokkan daerah terjangkau Demam Berdarah Dengue (DBD) yang mampu menghasilkan tiga *cluster* yaitu tinggi, sedang dan rendah. Penelitian tersebut, sebagai masukan bagi pemerintahan untuk memberikan kebijakan terhadap masyarakat dalam mengembangkan pola hidup sehat. Penelitian berikutnya tentang *clustering* terhadap penentuan alokasi pemberian dana pendidikan yang dilakukan oleh Salulolo, Tanaamah dan Wowor [4]. Penelitian tersebut telah membuktikan bahwa algoritma k-means berhasil menentukan kelompok sangat penting, penting dan kurang penting. Penelitian ini digunakan untuk menunjang keputusan dalam melakukan pembagian dana pendidikan. Kemudian penelitian data nilai siswa sebagai penentuan beasiswa yang dilakukan oleh Muzakir [5] dengan algoritma k-means. Hasil penelitian digunakan untuk pengelompokan penerima beasiswa berdasarkan nilai siswa. Penelitian Algoritma k-means juga digunakan untuk pengelompokan penjualan produk yang dilakukan oleh Darmi dan Setiawan [6]. Pengelompokan penjualan pada penelitian tersebut mampu menggolongkan barang yang laku dan tidak laku.

Berdasarkan penelitian-penelitian yang telah membuktikan pengelompokan dengan algoritma k-means. Maka kinerja k-means dalam pengelompokan mampu digunakan untuk menunjang keputusan dalam menentukan *clustering*. Pengelompokan ini bertujuan untuk menunjang keputusan dalam pengelompokan daerah penderita filariasis dan mengetahui pengelompokan penderita filariasis. Sehingga manfaat dari hasil *clustering* untuk menunjang keputusan dalam menentukan tindakan. Tindakan yang dilakukan yaitu pencegahan atau penanggulangan bagi daerah yang terdapat penderita filariasis berdasarkan kondisi yaitu tinggi, sedang atau rendah.

Penelitian ini disusun dengan beberapa tahapan diantaranya menjelaskan algoritma k-means pada bagian 2, implementasi algoritma k-means dengan menjelaskan data, cara hitung dan evaluasi pada bagian 3, lalu hasil dan pembahasan pada bagian 4, selanjutnya kesimpulan dari penelitian ini pada bagian 5.

2. ALGORITMA K-MEANS

James MacQueen menggunakan algoritma k-means pada tahun 1967 [7]. Algoritma k-means sering digunakan dalam melakukan pengelompokan karena kelebihannya dalam waktu komputasi yang cepat [8]. Cara kerja algoritma k-means langkah pertama menentukan K (jumlah cluster) sebagai jumlah *clustering* yang akan dibentuk. Setelah itu menentukan *centroid* (titik pusat *cluster*) awal, yang diambil secara acak. Kemudian melakukan perhitungan jarak setiap data ke masing-masing *centroid* yang telah ditentukan sebelumnya. Lalu setiap data memilih *centroid* yang berdekatan selanjutnya tentukan posisi *centroid* pada *centroid* yang sama. Jika posisi *centroid* yang baru dengan yang lama tidak sama maka melakukan pengulangan dari perhitungan jarak setiap data ke *centroid*, pengulangan terus dilakukan sampai nilai *centroid* tidak berubah [6].

Rumus pengelompokan untuk menghitung dalam menentukan jarak data dari masing-masing *centroid* dapat dilihat dari persamaan 1 [9].

$$d(P,Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(C))^2} \quad (1)$$

Keterangan :

D = Titik dokumen

P = Data *record*

C = Data *Centroid*

Setelah melakukan perhitungan persamaan 1 selanjutnya menentukan posisi *cluster* dengan menghitung jarak yang terpendek. Adapun rumus untuk menghitungnya dengan cara persamaan 2.

$$C(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x} \quad (2)$$

Keterangan :

C = *Centroid*

x_1 = Nilai data *record* ke 1

x_2 = Nilai data *record* ke 2

$\sum x$ = jumlah data *record*

3. IMPLEMENTASI ALGORITMA K-MEANS

A. Data

Data penderita filariasis diperoleh dari situs *web* Badan Pusat Statistik (BPS). Data terdiri dari 33 provinsi di Indonesia dari tahun 2008 sampai 2014. Data jumlah penderita filariasis diseleksi agar menjadi *dataset* sesuai dengan kebutuhan. Dari data penderita filariasis tersebut semua data digunakan dari 33 provinsi selama lima tahun dan merubah atribut berdasarkan tahun. Sehingga menjadi enam atribut terdiri dari nama provinsi, jumlah penderita filariasis sejak tahun 2008 hingga tahun 2012. Hasil yang telah seleksi ditunjukkan pada Tabel 1

Tabel 1 hasil seleksi

Nama Provinsi	Tahun 2008	Tahun 2009	Tahun 2010	Tahun 2011	Tahun 2012
Prov. Aceh	2359	2359	2359	2359	2359
Prov. Sumatra Utara	141	141	141	148	186
Prov. Sumatra Barat	274	274	274	274	193
Prov. Riau	532	532	532	532	310
Prov. Jambi	257	257	221	222	300
Prov. Sumatera Selatan	210	210	210	210	185
Prov. Bengkulu	94	94	94	94	85
Prov. Lampung	74	74	74	74	74
Prov. Kepulauan Bangka Belitung	207	207	207	207	207
Prov. Kepulauan Riau	31	31	31	31	39
.					
.					
.					

Nama Provinsi	Tahun 2008	Tahun 2009	Tahun 2010	Tahun 2011	Tahun 2012
Prov. Papua	1127	1158	1343	1343	1346

B. Penerapan Algoritma K-means

Pengolahan data penderita filariasis sesuai dengan tahapan - tahapan pada algoritma k-means, untuk pengelompokan ini terdapat tiga *cluster*. Kemudian akan menghasilkan provinsi mana saja penderita filariasis yang termasuk *cluster* tinggi, *cluster* sedang dan *cluster* rendah. Pengelompokan ini dilakukan dengan menentukan jumlah *cluster* lalu menentukan titik awal dan perhitungan jarak. Contoh perhitungan setiap data penderita filariasis ke *centroid* 1 dengan menggunakan persamaan 1. Perhitungan *centroid* 1 (C_1) untuk menghitung jarak *centroid* 1

$$C_1 = \sqrt{((2359-2359)^2+(2359-2359)^2+(2359-2359)^2+(2359-2359)^2+(2359-2359)^2)} = 714,2991$$

$$C_1 = \sqrt{((141-2359)^2+(141-2359)^2+(141-2359)^2+(148-2359)^2+(186-2359)^2)} = 4222,544$$

Selanjutnya, perhitungan setiap data penderita filariasis ke *centroid* 2 dengan menggunakan persamaan 1. Perhitungan *centroid* 2 (C_2) untuk menghitung jarak *centroid* 2

$$C_2 = \sqrt{((2359-985)^2+(52359-988)^2+(2359-988)^2+(2359-988)^2+(2359-988)^2)} = 2760,68$$

$$C_2 = \sqrt{((141-985)^2+(141-988)^2+(141-988)^2+(148-988)^2+(186-988)^2)} = 2180,164$$

Agar dapat mengetahui *centroid* 3 pada data penderita penyakit filariasis maka data dihitung menggunakan persamaan 1. Perhitungan *centroid* 3 (C_3) untuk menghitung jarak *centroid* 3

$$C_3 = \sqrt{((2359-18)^2+(52359-18)^2+(2359-18)^2+(2359-18)^2+(2359-18)^2)} = 4841,522$$

$$C_3 = \sqrt{((141-18)^2+(141-18)^2+(141-18)^2+(148-18)^2+(186-18)^2)} = 102,9393$$

Setelah melakukan perhitungan jarak, kemudian menentukan *cluster*, dan dari setiap *clustering* dihitung untuk memperoleh *centroid* baru. Cara untuk menghitung *centroid* baru dari masing-masing *clustering* menggunakan persamaan 2. Contoh perhitungan *centroid* 1 (C_1) tahun 2008 dari *cluster* 1, perhitungan ini untuk menghasilkan nilai *centroid* baru 1 pada tahun 2008

$$C_1(2008) = \frac{2359+2359}{2} = 2020,5$$

Kemudian perhitungan untuk menghasilkan nilai *centroid* 1 pada tahun 2009 dengan perhitungan *centroid* 1 (C_1) tahun 2009 dari *cluster* 1,

$$C_1(2009) = \frac{2359+1730}{2} = 2044,5$$

Lalu dari *cluster* 1 untuk perhitungan *centroid* 1 (C_1) tahun 2010, perhitungan ini untuk menghasilkan nilai *centroid* baru 1 pada tahun 2010 sebagai berikut

$$C_1(2010) = \frac{2359+1730}{2} = 2044,5$$

Perhitungan selanjutnya *centroid* baru 1 pada tahun 2011 dari *cluster* 1 untuk perhitungan *centroid* 1 (C_1) tahun 2011.

$$C_1(2011) = \frac{2359+1730}{2} = 2044,5$$

Berikutnya perhitungan *centroid* 1 (C_1) tahun 2012 didapat dari *cluster* 1, untuk menghasilkan nilai *centroid* baru 1 pada tahun 2012

$$C_1(2012) = \frac{2359+1730}{2} = 2044,5$$

Berdasarkan persamaan 1 dan persamaan 2, maka cara sederhananya algoritma k-means dirujuk pada Algoritma 1 [10].

Algoritma 1
Input:
 $D = \{x_1, x_2, \dots, x_n\}$ // kumpulan data
 K // jumlah *cluster* yang akan dibentuk
Output:
 K // kumpulan *cluster*
Algoritma k-means:
menentukan *centroid* (Titik awal) $C_1, (C_1, \dots, C_1$
repeat
menentukan setiap data ke *cluster* yang memiliki rata-rata jarak terdekat;
menghitung rata-rata setiap *cluster* untuk *centroid baru*;
until
pengelompokan berhasil

C. Evaluasi

Penelitian ini, menghasilkan pengelompokan daerah penderita filariasis berdasarkan provinsi di Indonesia dengan algoritma k-means yang dilakukan perhitungan manual menggunakan Microsoft Excel 2016. Lalu evaluasi menggunakan teknik komputasi sebagai pembandingan dengan perhitungan manual.

4. HASIL DAN PEMBAHASAN

Pengelompokan daerah penderita filariasis dibentuk menjadi tiga *clustering*. Agar menghasilkan *clustering* salah satunya melakukan perhitungan jarak. Hasil perhitungan mulai dari setiap data filariasis kesemua *centroid* yaitu *centroid* 1 (C_1), *centroid* 2 (C_2), dan *centroid* 3 (C_3). Pada provinsi Aceh termasuk *Cluster* 1 sebab antara C_1, C_2, C_3 C_1, C_2 dan C_3 nilai terkecil terdapat pada C_2 . Kemudian provinsi Papua terdapat pada *Cluster* 2 karena dari C_1, C_2 dan C_3 , nilai yang terdapat C_2 lebih kecil dari yang lainnya. Provinsi Sumatra Utara termasuk *clustering* 3 karena nilai C_3 lebih kecil dari pada C_2 dan C_3 . Hasil dari perhitungan *centroid* terdapat pada Tabel 2

Tabel 2 Menghitung ke setiap centroid

Nama Provinsi	2008	2009	2010	2011	2012	C_1	C_2	C_3
Prov. Aceh	2359	2359	2359	2359	2359	714,2991	2760,68	4841,522
Prov. Sumatra Utara	141	141	141	148	186	4222,544	2180,164	102,9393
Prov. Sumatra Barat	274	274	274	274	193	3985,264	1946,316	159,919
Prov. Riau	532	532	532	532	310	3476,649	1450,638	685,9617
Prov. Jambi	257	257	221	222	300	3999,378	1959,137	145,1495
Prov. Sumatera Selatan	210	210	210	210	185	4102,66	2061,739	33,12702
Prov. Bengkulu	94	94	94	94	85	4354,813	2313,187	227,3548
Prov. Lampung	74	74	74	74	74	4395,491	2353,661	267,9432
Prov. Kepulauan Bangka Belitung	207	207	207	207	207	4098,098	2056,648	29,92323
Prov. Kepulauan Riau	31	31	31	31	39	4488,061	2446,008	360,5994
.								
.								
.								
Prov. Papua	1127	1158	1343	1343	1346	1748,237	327,5966	2401,835

Dari perhitungan pengelompokan sampai *centroid* baru dan *centroid* lama nilainya sama melakukan dua kali iterasi, namun dilakukan terus pengulangan sampai 30 iterasi untuk pembuktian nilai *centroid* tidak berubah. Hasil *centroid* baru pada setiap iterasi yaitu iterasi 1 dan iterasi 2 terdapat pada Tabel 3

Tabel 3 Centroid setiap iterasi

Iterasi	Centroid	Tahun 2008	Tahun 2009	Tahun 2010	Tahun 2011	Tahun 2012
1	Centroid baru 1	2020,5	2044,5	2044,5	2044,5	2044,5
	Centroid baru 2	1056	1073	1165,5	1165,5	1167
	Centroid baru 3	191,2413793	195,2413793	192,3793103	197,4482759	192,7241379
2	Centroid baru 1	2020,5	2044,5	2044,5	2044,5	2044,5
	Centroid baru 2	1056	1073	1165,5	1165,5	1167
	Centroid baru 3	191,2413793	195,2413793	192,3793103	197,4482759	192,7241379

Hasil pengelompokan penderita filariasis menggunakan algoritma k-means melakukan perhitungan dari setiap *centroid* untuk mengelompokkan berdasarkan *cluster*. *Cluster* tinggi menghasilkan dua provinsi, *cluster* sedang ada dua provinsi dan 29 provinsi lainnya termasuk *cluster* rendah. Nama-nama provinsi berdasarkan tingkatan *cluster* dilihat pada Tabel 4

Tabel 4 Cluster berdasarkan provinsi

Cluster	Provinsi	Jumlah
Tinggi	Aceh, Nusa Tenggara Timur	2 Provinsi
Sedang	Papua Barat, Papua	2 Provinsi
Rendah	Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara	29 Provinsi

5. KESIMPULAN DAN SARAN

Berdasarkan dari hasil penelitian ini kesimpulan yang diperoleh adalah pengelompokan daerah penderita filariasis dengan algoritma k-means pada cluster tinggi yaitu *clustering* 1 terdapat dua provinsi. kemudian cluster sedang terdapat pada *clustering* 2 terdiri dari dua provinsi. 29 provinsi lainnya termasuk kedalam *clustering* 1 yaitu cluster rendah. Sehingga hasil penelitian ini dapat menjadi acuan untuk mengambil keputusan dalam melakukan sebuah tindakan pada penyakit filariasis. Tindakan dapat berupa pencegahan dan penanggulangan penderita filariasis berdasarkan jumlah pasien yang lebih banyak (C_1) agar tidak mengalami peningkatan.

Peneliti selanjutnya disarankan pengelompokan daerah penderita filariasis berdasarkan provinsi menggunakan teknologi digital lain seperti sistem pakar untuk mendiagnosa penyakit penderita filariasis. Sehingga mengurangi terjadinya peningkatan penderita filariasis.

DAFTAR PUSTAKA

- [1] D. Agustiantiningsih, "Praktik Pencegahan Filariasis," *Jurnal Kesehatan Masyarakat*, p. 191, 2013.
- [2] DEPKES RI, 2016. [Online]. Available: www.depkes.go.id/download.php?file=download/pusdatin/infodatin/Infodatin-Filariasis-2016.pdf.
- [3] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan RapidMiner dengan K-Means Cluster pada daerah terjangkau Demam Berdarah Dengue berdasarkan provinsi," *Jurnal of Computer Engineering System and Science*, p. 175, 2018.
- [4] Y. H. Salulolo, A. R. Tanaamah and A. D. Wowor, "Analisis dan Perbandingan Algoritma Clustering dalam Penentuan Alokasi Bantuan Dana Pendidikan Propinsi Jawa Tengah," *Konferensi Nasional Teknologi Informasi dan Komunikasi*, p. 179, 2016.
- [5] A. Muzakir, "Analisa dan Pemanfaatan Algoritma K-means Clustering Pada Data Nilai Siswa Sebagai Penentuan Penerima Beasiswa," *Prosiding Seminar Nasional Aplikasi Sains & Teknologi*, p. 195, 2014.

- [6] Y. Darmi and A. Setiawan, "PENERAPAN METODE CLUSTERING K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK," *Jurnal Media Infotama*, pp. 149-150, 2016.
- [7] F. Effendy, C. Faticah and D. Purwitasari, "Jurnal Informatika Mulawarman," *Implementasi Fitur Geometri dan K-means pada Perhitungan dan Segmentasi sel Darah Merah Bertumpuk*, vol. 9 No.3, p. 13, 2014.
- [8] A. Yasid, "Implementasi Automatic Clustering Menggunakan Differential Evolution dan CS Measure untuk Analisis Data Kemahasiswaan," *Jurnal Ilmiah NERO*, vol. 1, p. 47, 2014.
- [9] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering pada Penerimaan Mahasiswa Baru," *Jurnal Sistemasi*, p. 240, 2018.
- [10] W. M. W. Mohd, A. H. Beg, T. Herawan and K. F. Rabbi, "MaxD K-means: A clustering algorithm for Auto-generation of centroids and distance of data points in clusters," *Communications in Computer and Information Science*, pp. 192-199, 2012.