



Artikel Penelitian

## Analisis *Cluster* Penyakit Malaria Provinsi Papua Menggunakan Metode *Single Linkage* Dan *K-Means*

Alvian Sroyer<sup>a</sup>, Samuel A. Mandowen<sup>b</sup>, Felix Reba<sup>c\*</sup>

<sup>a,c</sup>Program Studi Matematika, Fakultas MIPA, Universitas Cenderawasih, Papua, Indonesia

<sup>b</sup>Program Studi Sistem Informasi, Fakultas MIPA, Universitas Cenderawasih, Papua, Indonesia

### INFORMASI ARTIKEL

#### Sejarah Artikel:

Diterima Redaksi: 12 Juli 2021

Revisi Akhir: 01 Januari 2022

Diterbitkan Online: 03 Januari 2022

### KATA KUNCI

Jenis Malaria,

Rentang Usia,

Bulan Kejadian

*Single Linkage*,

*K-Means*

### KORESPONDENSI

E-mail: [felix.reba85@gmail.com](mailto:felix.reba85@gmail.com)\*

### A B S T R A C T

Malaria adalah penyakit yang disebabkan oleh parasite bernama *Plasmodium*. Tercatat keseluruhan kasus malaria yang terjadi di Indonesia pada tahun 2019 adalah sebanyak 250.644 kasus. Dan kasus malaria tertinggi terjadi di provinsi Papua, yaitu sebesar 86% atau sebanyak 216.380 kasus. Di Provinsi Papua, penyakit malaria dialami oleh semua usia dan bulan-bulan terjadi peningkatan pasien penderita malaria juga sangat bervariasi. Hal ini mengakibatkan dinas Kesehatan mengalami kesulitan dalam mengelompokkan jenis malaria berdasarkan usia pasien dan bulan-bulan kejadian. Sebenarnya sudah ada penelitian yang menjelaskan pengelompokan jenis-jenis malaria, namun belum dijelaskan secara terperinci masing-masing kelompok malaria seperti Malaria *Tropika*, Malaria *Tertiana*, Malaria *Quartana*, Malaria *Ovale*. Tujuan dari penelitian ini adalah, melakukan analisis *cluster* terhadap beberapa jenis malaria, usia dan bulan kejadian. Metode *cluster* yang digunakan dalam penelitian ini adalah metode *Single Linkage* dan *K-Means*. Selanjutnya kedua metode akan di evaluasi menggunakan standar deviasi. Metode terbaik yang dapat digunakan untuk analisis *cluster* adalah metode yang memiliki nilai standar deviasi lebih kecil. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Dinas Kesehatan Provinsi Papua. Hasil penelitian menunjukkan bahwa, metode *Single Linkage* lebih akurat dibandingkan dengan *K-Means*. Dimana dari 50 pasien terdapat 47 pasien lebih dominan terkena penyakit malaria tertiana yaitu pada rentang usia remaja dan dewasa pada bulan juni. Sehingga diharapkan pemerintah Provinsi Papua dapat memberikan sosialisasi kepada masyarakat, khususnya mereka yang pada rentang usia remaja dan dewasa. Karena hampir 94% penyakit malaria tertiana di derita oleh mereka yang berusia remaja dan dewasa.

## 1. PENDAHULUAN

Parasit *Plasmodium* merupakan penyebab munculnya penyakit malaria. Parasit ini dapat menyebar di dalam tubuh manusia melalui gigitan nyamuk yang disebut *Anopheles* betina. Jika seseorang digigit oleh nyamuk tersebut dan orang tersebut tidak memiliki kekebalan tubuh, maka gejala yang akan di timbulkan biasanya muncul 10-15 hari setelah gigitan nyamuk. Gejala yang biasanya pertama kali muncul adalah : terasa demam, sakit kepala, dan mengigil. Namun gejala ini sangat sulit dikenali sebagai gejala malaria. Apabila gejala ini di alami oleh seseorang dan ditangani dengan pemberian obat dalam waktu 24 jam, penyakit malaria justru akan berkembang menjadi penyakit yang

parah, bahkan penyakit malaria seringkali menyebabkan kematian. ([who.int](http://who.int)).

Penyakit malaria dapat ditemukan di seluruh wilayah di Indonesia, namun lebih dominan malaria banyak di temukan di Indonesia Timur. Stratifikasi sedang dapat ditemukan juga di daerah Kalimantan, Sumatera, dan Sulawesi. Berdasarkan laporan WHO, pada tahun 2008–2009, Provinsi Nusa Tenggara Timur (NTT), Papua Barat dan Papua adalah Provinsi dengan *Annual Parasite Incidence* (API) tertinggi di Indonesia. Lebih khusus wilayah Papua, pada tahun 2017 ditemukan lebih dari 100 kasus *Plasmodium falciparum* yang terkonfirmasi per 1000 penduduk. Berdasarkan penelitian yang dilakukan oleh Riset Kesehatan Dasar (Riskesdas) pada tahun 2018, hasil

pemeriksaan *Rapid Diagnostic Test* (RDT) terhadap penyakit malaria adalah sebesar 0,4%. Dimana jenis *Plasmodium falciparum* sendiri ditemukan sebanyak 57% kasus. (malaria.id). Tercatat keseluruhan kasus malaria yang terjadi di Indonesia pada tahun 2019 adalah sebanyak 250.644 kasus. Dan kasus malaria tertinggi terjadi di provinsi Papua, yaitu sebesar 86% atau sebanyak 216.380 kasus. (kompas.com).

Selama ini informasi berupa data penyakit penderita malaria di provinsi Papua yang disajikan dalam Profil Kesehatan Provinsi Papua masih secara umum, tanpa membedakan jenis-jenis malaria berdasarkan usia penderita dan juga bulan-bulan yang memiliki kasus tinggi. Akibatnya masyarakat menganggap bahwa kasus malaria terjadi merata disetiap usia. Bahkan bulan-bulan dimana terjadi peningkatan kasus malaria juga tidak diketahui oleh masyarakat.

Penelitian terkait malaria telah dilakukan oleh [1], penelitian ini dilakukan di provinsi Papua yaitu di kabupaten Jayapura khususnya wilayah danau sentani. Kasus yang diteliti adalah pengaruh risiko lingkungan fisik terhadap kasus malaria. Penelitian [2] juga dilakukan di provinsi Papua, karena banyak hal yang mempengaruhi kejadian malaria maka dalam penelitian ini dikaji faktor-faktor yang bersifat potensial. Namun dalam penelitian ini tidak disajikan secara jelas kasus malaria berdasarkan usia dan bulan terjadinya. Karena jika diketahui pasien yang menderita penyakit malaria berdasarkan kelompok usia dan bulan, maka akan sangat memudahkan dinas Kesehatan dalam hal penanganan. Sebenarnya sudah ada penelitian yang menjelaskan pengelompokan jenis-jenis malaria, seperti yang dilakukan oleh [3] penelitian ini juga dilakukan di provinsi Papua, hal yang diteliti adalah Kelompok *Anopheles punctulatus* (*Anopheles farauti*, *Anopheles koliensis*, *Anopheles punctulatus*) berdasarkan Bionomi Vektor Malaria. Penelitian ini juga belum dijelaskan secara terperinci masing-masing kelompok malaria seperti Malaria *Tropika*, Malaria *Tertiana*, Malaria *Quartana*, Malaria *Ovale*.

Tujuan dari penelitian ini adalah, melakukan analisis *cluster* terhadap beberapa jenis malaria, usia dan bulan kejadian. Dimana data jenis penyakit malaria seperti : malaria *Tropika*, malaria *Tertiana*, malaria *Quartana*, malaria *Ovale* dikelompokkan ke dalam beberapa kelompok berdasarkan usia dan bulan kejadian. Metode *cluster* yang digunakan dalam penelitian ini adalah metode *Single Linkage* dan *K-Means*. Selanjutnya kedua metode akan dievaluasi menggunakan standar deviasi. Metode terbaik yang dapat digunakan untuk analisis *cluster* adalah metode yang memiliki nilai standar deviasi lebih kecil. Ada begitu banyak penelitian menggunakan *Single Linkage* dan *K-Means* dalam penelitian untuk membandingkan metode dengan akurasi terbaik dalam pengelompokan [4-7]. Namun dalam penelitian ini, perbandingan metode tidak hanya untuk mengatakan bahwa satu metode lebih unggul, tetapi juga untuk melihat suatu kumpulan data yang sama dikelompokkan dengan cara yang berbeda. Dengan cara pandang yang berbeda pada data yang sama, peneliti dapat memiliki kemungkinan yang lebih besar untuk menemukan wawasan disetiap khusus dan mengetahui cara yang lebih efektif dalam menangani kasus *cluster*. Hal lainnya adalah, khususnya wilayah Papua masih banyak kasus pengelompokan terhadap data yang selama ini dilakukan dengan metode-metode konvensional di bidang Kesehatan, sehingga kedua metode ini

masih merupakan hal yang baru yang perlu diperkenalkan kepada dinas Kesehatan.

## 2. METODE

### 2.1. Studi Literatur

Sapriyanti, S., & Rianto, Y tahun 2020 membandingkan metode *K-Means* dan *Single Linkage* menggunakan software Rapidminer sebagai alat bantu, dalam penelitian ini kasus yang diteliti adalah data *agent contact center* pada PT XYZ. Dari hasil pengujian disimpulkan bahwa metode *K-Means* lebih baik jika dibanding metode *Single Linkage* [4]. Gustini, R., & Aziz, R. A. 2019 membuat sistem pendukung keputusan yang dapat digunakan untuk pemberian bantuan KIP. Sistem ini dibuat dengan mempertimbangkan beberapa kriteria yang telah ditentukan. Metode yang digunakan adalah *K-Means Clustering* dan *Average Linkage Clustering*. Kesimpulan yang diperoleh metode *Average Linkage* lebih baik dibandingkan dengan *K-Means* [5]. Arifin, Z., Santosa, S., & Soeleman, M. A. 2017 Menggunakan metode *Hierarchical Clustering*, yaitu Metode *Single Linkage Clustering* untuk melakukan *cluster* cerpen berdasarkan tingkat kemiripan dari dokumen cerpen. Kesimpulan yang diperoleh Metode *Hierarchical Agglomerative Clustering* memiliki performansi yang lebih baik dibandingkan *K-Means* [6]. Handoko, K. 2016 menggunakan *K-Means* dalam peningkatan mutu pembelajaran di Akademi Komunitas Solok Selatan, salah satu prodi yang dijadikan sampel penelitian yaitu Program Studi TKJ. Selanjutnya pengujian digunakan software RapidMiner 5.3. Hasilnya mempresentasikan data mahasiswa dengan mutu pembelajaran sangat baik, cukup baik, dan kurang baik [7].

### 2.2. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder, dimana peneliti hanya mengambil 50 data kasus malaria sebagai sampel penelitian dari keseluruhan data. Juga data ini hanya berupa data sampel malaria pada satu tahun terakhir. Selanjutnya dari data 50 sampel yang terpilih diambil beberapa atribut yang berkaitan dengan penelitian ini, seperti : jenis malaria, usia dan bulan kejadian. Sumber data yang digunakan pada penelitian ini adalah data kesakitan dan kematian akibat malaria yang diambil dari profil dinas Kesehatan Provinsi Papua.

### 2.3. Proses Cluster dengan Single Linkage dan K-Means

Tahap cluster data penyakit malaria provinsi Papua berdasarkan usia dan bulan dengan metode *Single Linkage* dan *K-Means* menggunakan software Minitab 16, Weka 3.9.5, RStudio. Prosedur cluster sebagai berikut :

#### 2.3.1. Kategori Data

Karena metode *Single Linkage* dan *K-Means* menggunakan data angka, maka data yang masih berbentuk nominal harus dikategorikan ke dalam bentuk angka. Sehingga variabel yang digunakan dalam penelitian ini seperti : bulan dan jenis malaria yang masih berbentuk nominal, perlu dikategorikan ke dalam bentuk angka [8].

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j)$$

### 2.3.2. Standarisasi dan Transformasi Data

Jika terdapat perbedaan satuan nilai dalam setiap variabel penelitian, maka data tersebut perlu dilakukan standarisasi. Perbedaan yang dimaksudkan adalah, misalkan terdapat 3 variabel : Variabel pertama memiliki nilai dalam satuan juta, Variabel kedua memiliki nilai dalam satuan puluhan dan variabel ketiga memiliki nilai lebih kecil dari kedua variabel tersebut. Ketiga variabel di atas jika tidak dilakukan standarisasi data, maka pada saat dilakukan pengelompokan akan ada perbedaan yang besar dalam skala data. Perbedaan ini akan menyebabkan perhitungan jarak menjadi tidak valid. Transformasi data dapat dilakukan dengan menggunakan uji *z-score* [9], artinya transformasi data menggunakan distribusi normal baru yaitu  $N(0,1)$  sebagai berikut :

$$z_i = \frac{x_i - \bar{x}}{s}, \quad \text{untuk } i = 1, 2, \dots, n$$

Dimana :

- $x_i$  = data ke -  $i$
- $\bar{x}$  = rata - rata data
- $s$  = simpangan baku

### 2.3.3. Cluster Terbaik dengan Dendrogram untuk Single Linkage

Dendrogram adalah diagram yang menunjukkan hubungan hirarkis antara objek. Dendrogram juga paling sering dibuat sebagai output dari pengelompokan hirarkis. Kegunaan utama dendrogram adalah untuk menentukan cara terbaik dalam mengelompokan suatu objek [10], [11].

### 2.3.4. Cluster Terbaik dengan Silhouette untuk K-Means

Terdapat begitu banyak metode yang dapat digunakan untuk menentukan melihat kualita dan kekuatan suatu pengelompokan, salah satunya metode Silhouette Coefficient. Metode *Silhouette* lebih mengarah kepada pada metode interpretasi dan kebaikan dalam *cluster* data. Hasil analisis dari metode ini disajikan dalam bentuk Grafik yang sangat rinci tentang seberapa valid setiap objek kelompokan [12], [13]:

- 1) Hitung nilai rerata jarak dari data ke- $i$  dengan semua data lain yang ada dalam satu *cluster* dengan rumus berikut :

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j)$$

Dimana :

- $j$  : Data dalam satu *cluster*  $A$
- $d(i, j)$  : Jarak antara data  $i$  dengan  $j$ .

- 2) Selanjutnya dihitung rerata jarak dari data ke -  $i$  tersebut dengan semua data di cluster lain. Selanjutnya diambil nilai yang paling kecil. Rumusnya sebagai berikut :

Dimana :

$d(i, C)$  : Jarak rata-rata data ke -  $i$  dengan semua objek pada cluster lain  $C$  dimana  $A \neq C$ .

- 3) Untuk menghitung nilai Silhouette adalah sebagai berikut :

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Dengan :

$$b(i) = \min_{C \neq A} d(i, C)$$

### 2.3.5. Metode Single Linkage

Metode Single Linkage adalah salah satu dari beberapa metode pengelompokan hirarkis. Metode ini didasarkan pada pengelompokan *cluster* secara aglomerasi, pada setiap langkah menggabungkan dua klaster yang berisi pasangan elemen terdekat yang belum termasuk dalam *cluster* yang sama satu sama lain. Metode ini bisanya diawali dengan menentukan objek yang memiliki nilai jarak terkecil dalam matriks proximity  $D = \{d_{ik}\}$ . Objek-objek yang memiliki jarak terkecil disatukan dalam satu kelompok, dimisal  $U$  dan  $V$  adalah dua objek yang akan dilakukan pengelompokan. Sehingga akan diperoleh kelompok  $(UV)$ . Untuk mencari jarak antara kelompok  $(UV)$  dan kelompok  $W$  atau kelompok lainnya digunakan rumus sebagai berikut [14]:

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\}$$

Dimana :

$d_{UW}$  : Jarak terdekat dari kelompok  $U$  dan  $W$

$d_{VW}$  : Jarak terdekat dari kelompok  $V$  dan  $W$

### 2.3.6. Metode K-Means

K-Means clustering adalah pendekatan sederhana untuk mempartisi kumpulan data menjadi  $k$  *cluster* yang berbeda dan tidak tumpang tindih. Untuk melakukan K-Means clustering, pertama-tama kita harus menentukan jumlah cluster  $k$  yang diinginkan. Metode K-Means akan menetapkan setiap pengamatan tepat satu dari  $k$  *cluster*. Metode K-Means *cluster* mencoba memisahkan sampel dalam  $n$  kelompok dengan varians yang sama, meminimalkan kriteria yang dikenal sebagai inersia atau jumlah kuadrat dalam cluster. Metode ini bertujuan untuk memilih centroid yang meminimalkan inersia, atau kriteria jumlah kuadrat dalam *cluster* [15]. Berikut ini langkah-langkah menggunakan metode *K-Means* [16] :

- 1) Tentukan nilai  $k$ , nilai  $k$  mewakili banyaknya kelompok yang akan dibentuk.
- 2) Pilih secara acak titik pusat ( $k$  *centroid*) *cluster*
- 3) Selanjutnya pengukuran dilakukan pada jarak euclidean antara tiap titik dan centroid.
- 4) Menetapkan tiap titik ke kelompok yang terdekat.
- 5) Hitung rata-rata setiap cluster sebagai centroid baru menggunakan rumus berikut :

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i$$

Dimana :

$n_k$  : Jumlah rentang usia dan bulan terjadinya dalam cluster k

$d_i$  : Berdasarkan rentang usia dan bulan terjadinya dalam cluster k.

- 6) Ulangi langkah 3-5 dengan pusat cluster baru. Ulangi sampai Konvergensi. (Tidak ada perubahan lebih lanjut).

### 7. Evaluasi Metode

Agar dapat mengetahui kebaikan suatu metode cluster, dapat dihitung nilai rerata simpangan baku (standar deviasi) dalam cluster ( $S_w$ ) dan juga harus di hitung standar deviasi antar cluster ( $S_b$ ). Berikut adalah rumus untuk menghitung nilai ( $S_w$ ) [17] :

$$S_w = \frac{1}{K} \sum_{k=1}^K S_k$$

Jika diberikan cluster  $c_k$  dimana ( $k = 1, \dots, p$ ). Selanjutnya untuk setiap cluster yang memiliki anggota  $x_i$  dimana ( $i = 1, \dots, n$ ) dan  $\bar{x}_k$  adalah rata-rata dari cluster k, Maka untuk menghitung nilai standar deviasi ke-k ( $S_k$ ) adalah sebagai berikut:

$$S_k = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_i - \bar{x}_k)^2}$$

Jika terdapat rerata variabel dalam suatu cluster adalah  $k(\bar{x}_k)$ , maka item dari setiap kelompok akan berbeda, serta standar deviasi antar kelompok  $S_b$  adalah seperti dibawah ini :

$$S_b = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (\bar{X}_k - \bar{X})^2}$$

Sehingga untuk menentukan metode cluster terbaik, harus di hitung rasio (S) dari masing-masing metode cluster. Cara untuk menghitung rasio adalah nilai  $S_w/S_b$ . Jika rasio  $S_w$  terhadap  $S_b$  menghasilkan nilai yang kecil, maka dapat dikatakan metode cluster tersebut mempunyai kinerja yang baik. Secara umum rumus untuk menghitung rasio di sajikan dibawah ini :

$$S = \frac{S_w}{S_b}$$

Berikut ini adalah beberapa hal yang dapat menyebabkan hasil cluster dengan metode *Single Link* dan *K-Means* berbeda [18]:

- 1) Analisis cluster adalah metode statistik yang tidak memiliki asumsi atau pengetahuan awal tentang data. Sehingga data yang dikelompokkan hanya berdasarkan kedekatan dan hanya melihat murni bilangan, bukan apa yang diwakilinya;
- 2) Penentuan titik awal disetiap metode cluster;
- 3) Inisialisasi terhadap data;
- 4) Tidak dilakukan iterasi beberapa kali;
- 5) Kesalahan kuadrat rata-rata dan mungkin entropi jika data diberi label;
- 6) Misalnya, jika jumlah cluster dalam penelitian relatif  $\geq 50$ , dan kumpulan data yang digunakan memiliki dimensi tinggi  $\geq 10$ , kemungkinan besar metode *k-means* apa pun yang digunakan, bahkan setelah beberapa kali pengulangan, hasilnya akan sangat berbeda.

### 8. Kategori Usia menurut Departemen Kesehatan RI

Departemen Kesehatan RI pada tahun 2019 memberikan kategori usia menjadi beberapa kelompok, yang kemudian masing-masing kelompok usia menggambarkan jenjang pertumbuhan. Berikut adalah pembagian jenjang berdasarkan kategori usia :

Tabel 1. Usia menurut Departemen Kesehatan RI

Usia (Tahun)	Jenjang
0 – 5	Masa balita
6 – 11	Masa kanak-kanak
12 – 16	Masa remaja awal
17 – 25	Masa remaja akhir
26 – 35	Masa dewasa awal
36 – 45	Masa dewasa akhir
46 – 55	Masa lansia awal
56 – 65	Masa lansia akhir
> 65	Masa Manula

(Sumber : Departemen Kesehatan RI, 2019)

## 3. HASIL DAN PEMBAHASAN

### 3.1 Data Malaria dalam bantuk Kategori

Data yang diperoleh dari Profil Kesehatan Provinsi Papua, selanjutnya dipilih beberapa variabel input yang akan digunakan dalam pengelompokan. Variabel yang dipilih diantaranya : usia pasien, jenis penyakit malaria yang diderita pasien, dan bulan kejadian. Selanjutnya data berbentuk nominal seperti jenis penyakit malaria dan bulan harus dikategorikan dalam bentuk angka. Berikut adalah data yang telah dikategorikan :

#### 3.1.1. Jenjang Usia

Variabel usia yang diperoleh dari dinas Kesehatan pada tabel 2 dipisahkan berdasarkan jenjang usia:

Tabel 2. Kelompok usia berdasarkan jenjang

Usia	Jenjang
< 17 tahun	Masa bayi dan Kanak-kanak
17 – 45 tahun	Masa Remaja dan dewasa
$\geq 46$ tahun	Masa tua

(Sumber : Departemen Kesehatan RI pada tahun 2019)

#### 3.1.2. Kategori berdasarkan Jenis Malaria

Ada 4 jenis penyakit malaria dalam data dinas Kesehatan, sehingga setelah dikategorikan hasilnya seperti di bawah ini:

Tabel 3. Kelompok usia yang digunakan dalam penelitian

Jenis Penyakit Malaria	Kategori
Malaria Tertiana	1
Malaria Quartana	2
Malaria Tropika	3
Malaria Ovale	4

### 3.1.3. Kategori berdasarkan Bulan

Sama halnya dengan data jenis malaria yang telah dikategorikan, data bulan terjadinya kasus malaria setelah dikategorikan hasilnya seperti di bawah ini:

Tabel 4. Kelompok usia yang digunakan dalam penelitian

Bulan	Kategori	Bulan	Kategori
Januari	1	Juli	7
Februari	2	Agustus	8
Maret	3	September	9
April	4	Oktober	10
Mey	5	November	11
Juni	6	Desember	12

### 3.2. Standarisasi dan Transformasi Data

Data pasien yang telah dikategorikan berdasarkan usia dan bulan atas, selanjutnya akan dilakukan transformasi. Transformasi data dilakukan untuk menyamakan skala dari setiap variabel. Tabel 5 dibawah ini merupakan data hasil transformasi:

Tabel 5. Kelompok usia yang digunakan dalam penelitian

No	Jenis Malaria	Usia	Bulan	No	Jenis Malaria	Usia	Bulan
1	-1.08	-0.83	-1.62	26	-1.08	1.81	1.57
2	-1.08	-0.89	-1.62	27	0.00	1.17	1.57
3	0.00	-1.36	-1.62	28	1.08	0.41	0.70
4	0.00	0.35	-1.04	29	1.08	0.23	-0.75
5	1.08	-0.30	-1.04	30	-1.08	0.17	-0.75
6	0.00	-0.95	-0.17	31	1.08	-0.48	-0.46
7	-1.08	-1.47	-1.33	32	1.08	-0.36	-0.17
8	1.08	0.99	-0.75	33	-1.08	-1.59	-0.46
9	0.00	-1.30	1.57	34	0.00	1.87	-1.04
10	0.00	-0.65	0.41	35	-1.08	1.17	-1.04
11	1.08	-0.77	-0.46	36	-1.08	-0.24	1.57
12	-1.08	-0.24	0.99	37	-1.08	-0.36	1.57
13	1.08	-0.18	0.99	38	-1.08	-0.36	1.28
14	2.16	0.64	0.99	39	0.00	-0.30	-1.04
15	-1.08	1.64	0.99	40	1.08	-0.48	-1.04
16	-1.08	-0.48	1.28	41	0.00	-0.95	-0.46
17	1.08	-0.65	0.12	42	1.08	2.75	0.70
18	1.08	-1.00	0.12	43	0.00	0.99	0.99
19	-1.08	0.82	0.12	44	-1.08	1.87	0.99
20	-1.08	-0.01	0.12	45	-1.08	1.99	0.12
21	-1.08	0.05	-0.17	46	0.00	-0.59	0.12
22	1.08	-0.53	-1.62	47	1.08	-0.83	-0.46
23	1.08	-0.36	0.70	48	1.08	-0.65	-0.75
24	-1.08	-0.42	0.70	49	1.08	-0.18	-0.75
25	-1.08	0.93	1.28	50	1.08	-0.12	-1.04

(Sumber : Departeman Kesehatan RI pada tahun 2019)

### 3.3. Penentuan k Terbaik dengan Dendrogram dan Silhouette

#### 3.3.1. Menentukan k Terbaik dengan Dendrogram

Setelah data jenis malaria, usia dan bulan diimport ke *software* RStudio dan kemudian dilakukan pengujian menggunakan dendrogram, maka hasilnya sebagai berikut :

#### Cluster Dendrogram

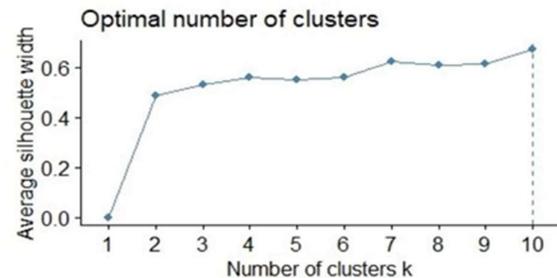


Gambar 1. Dendrogram untuk *Single Linkage*

Dendrogram yang ditunjukkan pada gambar 1 di atas, dapat disimpulkan bahwa banyaknya cluster untuk metode *Single Linkage* adalah k=4.

#### 3.3.2. Menentukan k Terbaik dengan Silhouette

Sama seperti langkah yang dilakukan pada *Single Linkage*, data jenis malaria, usia dan bulan diimport ke *software* RStudio dan kemudian dilakukan pengujian menggunakan metode *Silhouette*, maka hasilnya sebagai berikut:

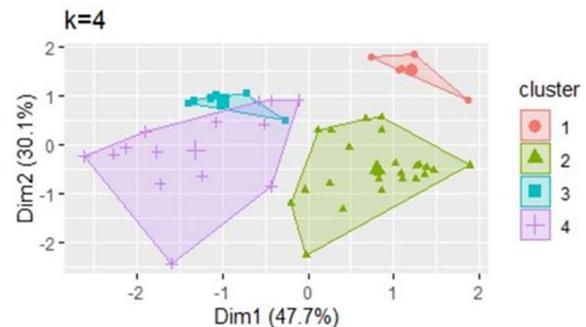


Gambar 2. Silhouette untuk *K-Means*

Berdasarkan gambar *Silhouette* untuk *K-Means* di atas, dapat disimpulkan bahwa banyaknya cluster untuk metode *K-Means* adalah k=9.

### 3.4. Single Linkage

Berdasarkan hasil pengujian menggunakan dendrogram, maka data pasien sebanyak (50 pasien) dibuat menjadi 4 kelompok dan 3 atribut (jenis malaria, usia dan bulan). Pengolahan data menggunakan *Software Weka* dan Minitab dapat dilihat di Gambar 3 dan Tabel 6.



Gambar 3. Cluster Pasien dengan Metode *Single Linkage*

Tabel 6. Cluster Pasien dengan Metode *Single Linkage*

Atribut	Cluster Pasien 50 (Pasien)			
	1 (47 Pasien)	2 (1 Pasien)	3 (1 Pasien)	4 (1 Pasien)
Jenis Malaria	-0,02	0,00	0,00	1,08
Usia	-0,07	-1,29	1,87	2,75
Bulan	-0,02	1,56	-1,03	0,70

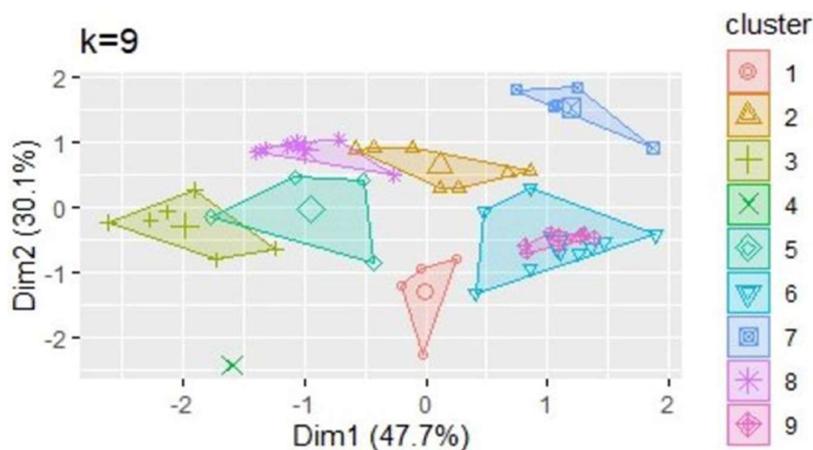
Berdasarkan Gambar 3 dan Tabel 6 di atas, dapat dijelaskan sebagai berikut :

- 1) *Cluster* pertama, dari 50 pasien terdapat 47 pasien lebih dominan terkena penyakit malaria tertiana yaitu pada rentang usia remaja dan dewasa pada bulan juni.

- 2) *Cluster* kedua, dari 50 pasien terdapat 1 pasien lebih dominan terkena penyakit malaria quartana yaitu pada rentang usia bayi dan anak-anak serta terjadi pada bulan desember.
- 3) *Cluster* ketiga, dari 50 pasien terdapat 1 pasien lebih dominan terkena penyakit malaria quartana yaitu pada rentang usia tua dan pada bulan maret.
- 4) *Cluster* keempat, dari 50 pasien terdapat 1 pasien lebih dominan terkena penyakit tropika yaitu pada rentang usia tua dan pada bulan september.

### 3.5. *K-Means Cluster*

Berdasarkan hasil pengujian menggunakan *Silhouette*, maka data pasien sebanyak (50 pasien) dibuat menjadi 9 kelompok dan 3 atribut. Pengolahan data menggunakan *Software Weka* dan *Minitab* dapat dilihat di Gambar 4 dan Tabel7.

Gambar 4. Cluster Pasien dengan Metode *K-Means*Tabel 7. Cluster Pasien dengan Metode *K-Means*

Atribut	Cluster Pasien 50 (Pasien)				
	1 (3 Pasien)	2 (11 Pasien)	3 (8 Pasien)	4 (12 Pasien)	
Jenis Malaria	1.4402	-0.9819	-0.405	1.0801	
Usia	1.2666	0.4853	0.2775	-0.2897	
Bulan	0.7974	1.2538	-0.4214	-0.7713	
Atribut	Cluster Pasien 50 (Pasien)				
	5 (1 Pasien)	6 (4 Pasien)	7 (4 Pasien)	8 (4 Pasien)	9 (3 Pasien)
Jenis Malaria	-1.0801	-1.0801	-1.0801	-1.0801	0
Usia	-1.5916	1.4624	-0.5491	-1.1364	-0.3191
Bulan	-0.4576	-0.4576	0.4837	-1.5437	0.9905

Berdasarkan Gambar 4 dan Tabel 7 di atas, dapat dijelaskan sebagai berikut :

- 1) *Cluster* pertama, dari 50 pasien terdapat 3 pasien lebih dominan terkena penyakit malaria tropika yaitu pada rentang usia tua dan pada bulan september.

- 2) *Cluster* kedua, dari 50 pasien terdapat 11 pasien lebih dominan terkena penyakit malaria tertiana yaitu pada rentang usia remaja dan dewasa dan pada bulan oktober.
- 3) *Cluster* ketiga, dari 50 pasien terdapat 8 pasien lebih dominan terkena penyakit malaria tertiana yaitu pada rentang usia remaja dan dewasa dan pada bulan mei.

- 4) *Cluster* keempat, dari 50 pasien terdapat 12 pasien lebih dominan terkena penyakit malaria tropika yaitu pada rentang usia remaja dan dewasa dan pada bulan maret.
- 5) *Cluster* kelima, dari 50 pasien terdapat 1 pasien lebih dominan terkena penyakit malaria tertiana yaitu pada rentang usia bayi dan anak-anak serta terjadi pada bulan mei.
- 6) *Cluster* keenam, dari 50 pasien terdapat 4 pasien lebih dominan terkena penyakit malaria tertiana yaitu pada rentang usia tua serta terjadi pada bulan mei.
- 7) *Cluster* ketujuh, dari 50 pasien terdapat 4 pasien lebih dominan terkena penyakit malaria tropika yaitu pada rentang usia remaja dan dewasa serta terjadi pada bulan agustus.
- 8) *Cluster* kedelapan, dari 50 pasien terdapat 4 pasien lebih dominan terkena penyakit malaria tertiana yaitu pada rentang usia bayi dan anak-anak serta terjadi pada bulan januari.
- 9) *Cluster* kesembilan, dari 50 pasien terdapat 3 pasien lebih dominan terkena penyakit malaria quartana yaitu pada rentang usia remaja dan dewasa serta terjadi pada bulan oktober.

Hasil penelitian menggunakan single linkage menunjukkan bahwa, dari 50 pasien terdapat 47 pasien (tertinggi) lebih dominan terkena penyakit malaria tertiana yaitu pada rentang usia remaja dan dewasa pada bulan juni. Jika dibandingkan dengan hasil penelitian yang dilakukan oleh [1] dengan uji bivariat menggunakan *Chi-Square* menunjukkan bahwa, curah hujan berpengaruh langsung terhadap kejadian malaria. Dimana bulan bulan dengan kejadian malaria tertinggi terjadi pada bulan Maret yakni 471,7 mm, sedangkan bulan dengan kejadian malaria terendah terjadi pada bulan Juni sebesar 92,4 mm. Artinya bahwa metode *Chi-Square* hanya dapat menunjukan secara umum bulan kejadian tertinggi atau terendah dan tidak secara langsung menunjukan pada usia berapa tahun pasien yang mendarita penyakit malaria. Juga tidak di tunjukan jenis penyakit malaria yang dominan pada bulan kejadian malaria. Demikian juga penelitian yang dilakukan oleh [2], dimana dalam penelitian ini digunakan analisis model SAR dan analisis model CAR untuk menentukan faktor-faktor potensial yang mempengaruhi kejadian Malaria di Provinsi Papua.

### 3.6. Evaluasi Metode

Berdasarkan hasil perhitungan menggunakan *software* rstudio diperoleh hasil  $S_w$ ,  $S_b$ . Selanjutnya dihitung rasio  $S_w$  terhadap  $S_b$ , maka hasil yang diperoleh berturut-turut sebagai berikut : S (metode *Single Linkage*) = 76.2% dan S untuk (metode *K-Means*) = 85.1 %.

## 4. KESIMPULAN

Berdasarkan penelitian yang telah dibahas tentang analisis cluster menggunakan *Single Linkage* dan *K-Means* pada data penyakit malaria berdasarkan rentang usia dan bulan terjadinya, maka beberapa hal dapat di simpulkan bahwa dengan menggunakan Dendrogram untuk metode *Single Linkage*, data penyakit malaria harus dibuat menjadi 5 cluster. Dimana cluster 1 menunjukan bahwa, dari 50 pasien terdapat 47 pasien (tertinggi) lebih dominan terkena penyakit malaria tertiana yaitu pada rentang usia remaja dan dewasa pada bulan juni. Selanjutnya dengan menggunakan *Silhouette* untuk metode *K-Means*, data penyakit malaria harus dibuat menjadi 9 *cluster*. Dimana cluster kedua

menunjukan bahwa, dari 50 pasien terdapat 11 pasien lebih dominan terkena penyakit malaria tertiana yaitu pada rentang usia remaja dan dewasa dan pada bulan oktober.

Berdasarkan hasil evaluasi terhadap kebaikan suatu metode, maka S(metode *Single Linkage*) = 76.2% dan S(metode *K-Means*) = 85.1 %. Artinya *Single Linkage* lebih akurat dibandingkan dengan *K-Means*. Namun jika dilihat dari hasil *cluster*, kedua metode memberikan hasil yang hampir mirip.

Algoritma *Single Linkage clustering* untuk pengelompokan data penyakit malaria, khususnya dinas Kesehatan Provinsi Papua. Selanjutnya diharapkan pemerintah Provinsi Papua dapat memberikan sosialisasi kepada masyarakat, khususnya mereka yang pada rentang usia remaja dan dewasa. Karena hampir 94% penyakit malaria tertiana di derita oleh mereka yang berusia remaja dan dewasa.

## UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih yang sebesar-besarnya kepada Kementerian Pendidikan dan Kebudayaan, yang telah memberikan Hibah Penelitian PNPB (Penerimaan Negara Bukan Pajak) pada tahun anggaran 2021, melalui Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Cenderawasih.

## DAFTAR PUSTAKA

- [1] A. F. Watofa, A. H. Husodo, S. Sudarmadji, and O. Setiani, "Risiko Lingkungan Fisik Terhadap Kejadian Malaria Di Wilayah Danau Sentani Kabupaten Jayapura Provinsi Papua," *J. Mns. dan Lingkung.*, 2018.
- [2] S. Siswanto and S. A. Thamrin, "Penentuan Faktor-Faktor Potensial Yang Mempengaruhi Kejadian Malaria Di Provinsi Papua Dengan Epidemiologi Spasial," *Indones. J. Stat. Its Appl.*, 2020, doi: 10.29244/ijsa.v4i3.681.
- [3] S. Sandy, "Bionomi Vektor Malaria Kelompok *Anopheles punctulatus* (*Anopheles farauti*, *Anopheles koliensis*, *Anopheles punctulatus*) di Provinsi Papua," *Balaba*, 2014.
- [4] P. Magister, I. Komputer, P. Studi, and I. Komputer, "Komparasi Metode Clustering K-Means Dan Single Linkage Untuk Penentuan Kelompok Agent Pada Call Center Jisamar ( Journal of Information System , Applied , Management , Accounting and Researh ) p-ISSN : 2598-8700 ( Printed ) JISAMAR ( Journal of Informatio," vol. 4, no. 3, pp. 1–7, 2020.
- [5] R. Gustini and R. Z. A. Aziz, "Pengembangan Model Pengambilan Keputusan Penerima Kartu Indonesia Pintar (Kip) Dengan Metode K-Means Dan Average Linkage Clustering (Studi ....," *JTKSI (Jurnal Teknol. Komput. dan Sist. ....*, 2019.
- [6] Z. Arifin, S. Stefanus, and A. M. Soeleman, "Klasterisasi Genre Cerpen Kompas Menggunakan Agglomerative Hierarchical Clustering- Single Linkage," *J. Teknol. Inf.*, 2017.
- [7] K. Handoko, "Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Di Program Studi Tkj Akademi Komunitas Solok Selatan)," *J. Teknol. dan Sist. Inf.*, 2016.
- [8] S. Rony, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru ( Studi Kasus :

- Politeknik LP3i Jakarta ),” *J. Lentera Ict*, 2016.
- [9] I. F. Anshori and Y. Nuraini, “Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Tasikmalaya Menggunakan Algoritma K-Means,” *J. Responsif*, 2020.
- [10] M. Schmidt, A. Kutzner, and K. Heese, “A novel specialized single-linkage clustering algorithm for taxonomically ordered data,” *J. Theor. Biol.*, 2017, doi: 10.1016/j.jtbi.2017.05.008.
- [11] J. L. Lin, J. C. Kuo, and H. W. Chuang, “Improving density peak clustering by automatic peak selection and single linkage clustering,” *Symmetry (Basel)*, 2020, doi: 10.3390/sym12071168.
- [12] D. M. SAPUTRA, D. SAPUTRA, and L. D. OSWARI, “Effect of Distance Metrics in Determining K-Value in K-Means Clustering Using Elbow and Silhouette Method,” 2020, doi: 10.2991/aisr.k.200424.051.
- [13] A. R. Mamat, F. S. Mohamed, M. A. Mohamed, N. M. Rawi, and M. I. Awang, “Silhouette index for determining optimal k-means clustering on images in different color models,” *Int. J. Eng. Technol.*, 2018, doi: 10.14419/ijet.v7i2.14.11464.
- [14] Z. Arifin, S. Stefanus, and A. M. Soelean, “Klasterisasi Genre Cerpen Kompas Menggunakan Agglomerative Hierarchical Clustering- Single Linkage,” *J. Teknol. Inf.*, vol. 13, pp. 92–100, 2017.
- [15] M. Mardalius, “Implementasi Algoritma K-Means Clustering Untuk Menentukan Kelas Kelompok Bimbingan Belajar Tambahan (Studi Kasus : Siswa Sma Negeri 1 Ranah Pesisir),” 2018, doi: 10.31219/osf.io/6mec3.
- [16] R. Handoyo, R. Rumani, and S. M. Nasution, “Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K-Means Pada Pengelompokan Dokumen,” *JSM STMIK Mikroskil*, vol. 15, no. 2, pp. 73–82, 2014, [Online]. Available: <https://mikroskil.ac.id/ejurnal/index.php/jsm/article/view/161>.
- [17] N. Ulinuh and R. Veriani, “Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage , Average Linkage dan Ward,” *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, 2020.
- [18] Christou, I.T. (2011) "Koordinasi Cluster Ensembles melalui Metode Persis", *Transaksi IEEE pada Analisis Pola dan Kecerdasan Mesin*, 33(2)279-293.



Felix Reba  
Dosen di Program Studi Matematika.  
Universitas Cenderawasih, Papua. Lulusan S2  
dari Universitas Gadjja Mada. Bidang  
penelitian tentang: Matematika Terapan,  
Statistik Komputasi dan Terapan .

## BIODATA PENULIS



Alvian Sroyer  
Dosen di Program Studi Matematika.  
Universitas Cenderawasih, Papua. Lulusan S2  
dari Institut Teknologi Bandung. Bidang  
penelitian tentang: Matematika Terapan.



Samuel A. Mandowen  
Dosen di Program Studi Sistem Informasi.  
Universitas Cenderawasih, Papua. Lulusan S2  
dari Queensland University of Technology.  
Bidang penelitian tentang: Teknologi  
Informasi.