

## IMPLEMENTASI ALGORITMA CLUSTERING PARTITIONING AROUND MEDOID (PAM) DALAM CLUSTERING VIRUS MERS-CoV

Septian Wulandari<sup>1</sup>, Nurfidah Dwitianti<sup>2</sup>  
Informatika, Universitas Indraprasta PGRI Jakarta  
septian.pmb09@rocketmail.com

*Submitted June 13, 2020; Revised July 25, 2020; Accepted August 1, 2020*

### Abstrak

*Middle East Respiratory Syndrome Coronavirus (MERS-CoV)* merupakan suatu penyakit saluran pernapasan yang disebabkan oleh virus corona. Virus ini merupakan virus yang mudah menular, namun penularannya tidak semudah flu biasa, melainkan virus MERS-CoV lebih rentan menular melalui kontak langsung, misalnya pada orang yang merawat penderita virus MERS-CoV tanpa menerapkan prosedur perlindungan diri terhadap virus dengan baik. Untuk mengetahui karakteristik, penyakit virus MERS-CoV dapat diketahui dengan mengidentifikasi DNA (*deoxyribonucleic acid*). Salah satu teknik dalam mengidentifikasi karakteristik makhluk hidup adalah dengan melakukan *clustering*. *Clustering* dapat dilakukan dengan mengelompokkan DNA menjadi kelompok-kelompok yang memiliki kesamaan sifat dan fungsinya. Algoritma *clustering* yang digunakan pada penelitian ini adalah *Partitioning Around Medoid (PAM)*. Algoritma ini memiliki kelebihan yaitu hasil proses *clustering* tidak bergantung pada urutan masuk dataset dan mengatasi *sensitive* terhadap *noise* dan *outlier*. Tujuan penelitian ini adalah untuk melakukan implementasi algoritma *clustering Partitioning Around Medoid (PAM)* dalam *clustering* virus MERS-CoV. Penelitian ini dilakukan melalui pendekatan studi literatur deskriptif kuantitatif. Implementasi algoritma PAM terhadap barisan DNA MERS-CoV diperoleh 2 *cluster* dengan nilai *silhouette coefficient* yang tertinggi berada pada jumlah *cluster* 2 yaitu 0,61534. Jumlah anggota pada *Cluster* 1 adalah 84 barisan DNA MERS-CoV dan jumlah anggota pada *Cluster* 2 adalah 16 barisan DNA MERS-CoV.

**Kata Kunci :** *MERS-CoV, Clustering, Partitioning Around Medoid (PAM)*

### Abstract

*The Middle East Respiratory Coronavirus (MERS-CoV)* is a disease caused by a coronavirus. This virus is contagious, but its transmission is not as easy as the common cold, MERS-CoV virus is better susceptible to transmitting through direct contact, for example in people who care about the MERS-CoV virus without the need for virus protection. To determine the characteristics, the MERS-CoV disease virus can be identified by identifying DNA (*deoxyribonucleic acid*). One technique in understanding the characteristics of life is by grouping. Grouping can be done by grouping DNA into groups that have attributes and functions. The Clustering algorithm used in this study is *Partitioning Around Medoid (PAM)*. This algorithm has the advantage that the results of the grouping process are not by following the order of entering the dataset and overcome sensitivity to noise and outliers. The purpose of this study is to implement the *Partitioning Around Medoid (PAM)* clustering algorithm in clustering the MERS-CoV virus. This research was conducted through a quantitative descriptive literature study. The implementation of the PAM algorithm on the MERS-CoV DNA sequence obtained 2 clusters with the highest *silhouette coefficient* value on the number of clusters 2, namely 0.61534. The number of members in *Cluster* 1 is 84 MERS-CoV DNA sequences and the number of members in *Cluster* 2 is 16 MERS-CoV DNA sequences.

**Key Words :** *MERS-CoV, Clustering, Partitioning Around Medoid (PAM)*

### 1. PENDAHULUAN

*Middle East Respiratory Syndrome Coronavirus (MERS-CoV)* merupakan

suatu penyakit saluran pernapasan yang disebabkan oleh corona virus. MERS-CoV merupakan virus yang menular dari unta ke

manusia, bahkan baru-baru ini ditemukan bahwa penyakit ini dapat menular dari manusia ke manusia. Berdasarkan laporan WHO, sejak September 2012 sampai dengan September 2013 telah ditemukan 130 kasus konfirmasi MERS-CoV dengan 58 kematian. Artinya sebanyak 44,6% manusia yang terjangkit virus ini meninggal dunia [1]. Lebih lanjut, WHO menyatakan bahwa selama 2015, National Focal Point (NFP) Arab Saudi mencatat sebuah tambahan 149 kasus infeksi MERS-CoV, termasuk 64 orang yang meninggal, dari total 1.090 pasien infeksi MERS-CoV dengan total 412 pasien telah meninggal [2]. MERS-CoV diduga awalnya berasal dari unta yang hidup di negara-negara Timur Tengah, seperti Arab Saudi, Yordania, dan Yaman. Virus ini merupakan virus yang mudah menular, namun penularannya tidak semudah flu biasa, melainkan virus MERS-CoV lebih rentan menular melalui kontak langsung, misalnya pada orang yang merawat penderita virus MERS-CoV tanpa menerapkan prosedur perlindungan diri terhadap virus dengan baik. Meskipun Indonesia sampai saat ini tercatat belum ada yang menderita virus MERS-CoV, namun pemerintah Indonesia rasanya perlu waspada pada penyakit ini. Mengingat bahwa Indonesia merupakan salah satu negara di dunia yang melakukan ibadah haji ataupun umrah di Arab Saudi, sehingga bukan tidak mungkin jamaah Haji ataupun Umrah dapat terinfeksi virus MERS-CoV ini.

Untuk mengetahui karakteristik penyakit virus MERS-CoV dapat diketahui dengan mengidentifikasi DNA (deoxyribonucleic acid). DNA merupakan polimer linier yang tersusun dari monomer nukleotida. Komponen penyusun nukleotida terdiri dari tiga jenis molekul, yaitu gula pentosa (deoksiribosa pada DNA), basa nitrogen, dan gugus fosfat. Basa yang ditemukan pada nukleotida adalah basa purin (adenin = A, guanin = G)

dan basa pirimidin yaitu cytosin = C, tymin = T, (urasil = U untuk RNA) [3]. Perbedaan susunan keempat basa itulah yang menyebabkan perbedaan karakteristik di semua makhluk hidup.

Salah satu teknik dalam mengidentifikasi karakteristik makhluk hidup adalah dengan melakukan *clustering* atau pengelompokan. *Clustering* data merupakan salah satu metode dalam data *mining* yang dapat digunakan untuk memetakan data kedalam kelompok-kelompok yang lebih kecil berdasarkan kesamaan karakteristik yang dimilikinya[4]. *Clustering* dapat dilakukan dengan mengelompokkan DNA menjadi kelompok-kelompok yang memiliki kesamaan sifat dan fungsinya. Salah satu algoritma *clustering* yang terkenal adalah *Partitioning Around Medoid* (PAM). Penelitian terdahulu yang menggunakan analisis *clustering* telah dilakukan oleh A. Bustamam, et, all yang melakukan *clustering* DNA virus MERS-CoV dengan menggunakan metode K-Mer [2].

Alasan dalam penelitian ini menggunakan *Partitioning Around Medoid* (PAM) dalam menganalisis kluster adalah karena algoritma ini memiliki kelebihan yaitu hasil proses *clustering* tidak bergantung pada urutan masuk dataset [5]. Disamping itu, algoritma ini mengatasi sensitif terhadap *noise* dan *outlier*, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data.

Tujuan dari penelitian ini adalah untuk mengimplementasikan algoritma *clustering Partitioning Around Medoid* (PAM) dalam mengelompokkan virus MERS-CoV. Diharapkan penelitian ini dapat memberikan informasi kesamaan karakteristik virus MERS-CoV yang terjangkit di beberapa negara. Sehingga dapat memberikan informasi dalam pembuatan antivirus atau vaksin untuk

menyembuhkan manusia yang terjangkit virus MERS-CoV.

## 2. METODE PENELITIAN

Pendekatan penelitian yang digunakan dalam penelitian ini adalah pendekatan studi literatur deskriptif kuantitatif. Pada studi literatur, dilakukan pengumpulan bahan-bahan pustaka yang mendukung dalam penyelesaian topik penelitian ini. Sedangkan untuk pendekatan deskriptif kuantitatif mencakup mengolah, menganalisa dan menginterpretasikan data sesuai dengan kebutuhan peneliti.

Pada penelitian ini data yang digunakan adalah 100 data DNA virus MERS-CoV yang diperoleh dari *National Center for Biotechnology Information* (NCBI) dalam laman [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). Barisan-barisan DNA yang diperoleh berupa FASTA dan berjenis *complete genome*. *Complete genome* dipilih karena diharapkan mampu memberikan informasi genetik yang lebih banyak.

Sistem pengkodean protein berdasarkan kombinasi dari nukleotida yang ada karena 20 asam amino tidak mungkin dikendalikan oleh empat jenis nukleotida. Terdapat empat jenis basa nitrogen pada DNA (A, C, G, dan T). Jika sebuah kodon hanya dapat mengkode empat asam amino. Jika sebuah kodon memiliki dua basa, maka hanya 16 asam amino yang dapat dikodekan sehingga jumlah tersebut belum dapat mewakili 20 asam amino.

Akan tetapi, jika kodon merupakan tiga basa, maka terdapat 64 kode yang mungkin berkombinasi sehingga jumlah tersebut diduga sudah dapat mewakili 20 asam amino. Variabel penelitian yang digunakan pada penelitian ini adalah 64 kode asam amino yang diperoleh dari ekstraksi ciri DNA MERS-CoV dengan menggunakan metode *n-mers frequency*. 64 kode asam amino tersebut dapat dilihat pada Gambar 1.

1st base	Standard genetic code				3rd base
	T	C	A	G	
T	TTT (Phe/F) Phenylalanine TTC TTA TTG	TCT TCC TCA TCG	TAT (Tyr/Y) Tyrosine TAC TAA Stop (Ochre) TAG Stop (Amber)	TGT (Cys/C) Cysteine TGC TGA Stop (Opal) TGG (Trp/W) Tryptophan	T C A G
C	CTT (Leu/L) Leucine CTC CTA CTG	CCT CCC CCA CCG	CAT (His/H) Histidine CAC CAA CAG	CGT (Arg/R) Arginine CGC CGA CGG	T C A G
A	ATT (Ile/I) Isoleucine ATC ATA ATG <sup>M</sup> (Met/M) Methionine	ACT ACC (Thr/T) Threonine ACA ACG	AAT (Asn/H) Asparagine AAC AAA (Lys/K) Lysine AAG	AGT (Ser/S) Serine AGC AGA AGG	T C A G
G	GTT (Val/V) Valine GTC GTA GTG	GCT GCC GCA GCG	GAT (Asp/D) Aspartic acid GAC GAA (Glu/E) Glutamic acid GAG	GGT (Gly/G) Glycine GGC GGG GGG	T C A G

Gambar 1. 64 Kode Asam Amino

Sumber: nafium.com

Metode atau algoritma yang digunakan pada penelitian ini adalah algoritma *clustering Partitioning Around Medoid* (PAM). Algoritma *clustering Partitioning Around Medoid* (PAM) bertujuan untuk clustering DNA MERS-CoV berdasarkan 64 asam amino yang diperoleh dari ekstraksi ciri DNA MERS-CoV. Tahapan pertama yang dilakukan dalam *clustering* DNA MERS-CoV adalah melakukan ekstraksi ciri pada DNA MERS-CoV dengan menggunakan metode *n-mers frequency*.

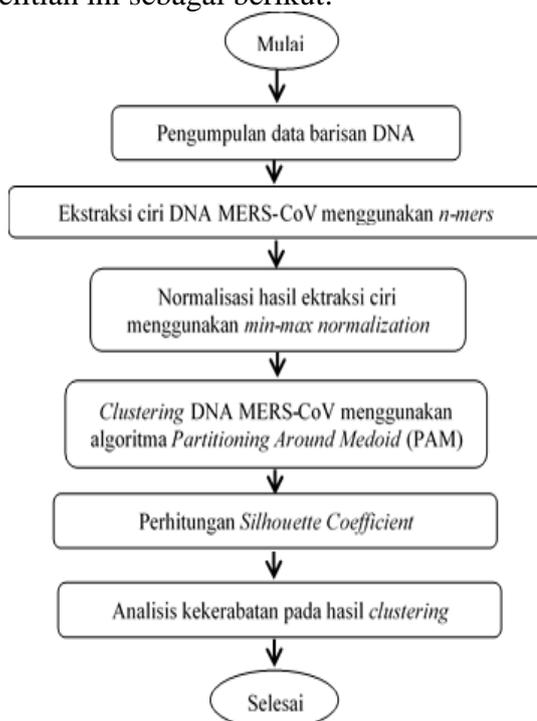
Pada proses ekstraksi ciri  $n = 3$  sehingga pola kemunculannya adalah  $4^3 = 64$  sehingga diperoleh 64 asam amino yang menjadi variabel pada penelitian ini. Langkah selanjutnya adalah melakukan proses normalisasi data untuk menghindari dimensi data yang terlalu besar ataupun terlalu kecil. Proses normalisasi data dilakukan dengan menggunakan *min-max normalization*. Setelah data sudah normal, maka langkah selanjutnya dilakukan proses *clustering* dengan menggunakan algoritma *clustering Partitioning Around Medoid* (PAM).

Pada proses *clustering* dengan algoritma PAM langkah pertama yang dilakukan adalah menginisialisasi pusat *cluster* sebanyak  $k$  (jumlah *cluster*) dimana nilai  $k$  diperoleh dengan menggunakan *average silhouette*. Kemudian dilakukan pengukuran jarak dengan menggunakan *Euclidean distance* serta menentukan secara acak objek pada masing-masing cluster sebagai kandidat medoid baru.

Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat medoid baru.

Kemudian, hitung total simpangan ( $S$ ) dengan menghitung nilai total *distance* baru dikurang total *distance* lama. Jika  $S < 0$ , maka tukar objek dengan data *cluster* untuk membentuk sekumpulan  $k$  objek baru sebagai medoid. Perulangan didasarkan pada terjadinya perubahan medoid. Jika tidak terjadi perubahan medoid maka iterasi berhenti dan didapatkan *cluster* beserta anggota *cluster* masing-masing.

Pada penelitian ini, alat bantu yang digunakan yaitu *software* RStudio 1.1.456 dan *notepad*. Adapun diagram alur penelitian ini sebagai berikut:



Sumber : Dokumentasi Pribadi

**Gambar 2. Diagram Alur Penelitian**

### 3. HASIL DAN PEMBAHASAN

Penerapan algoritma *clustering Partitioning Around Medoid* (PAM) dalam *clustering* virus MERS-CoV dilakukan dengan langkah-langkah sebagai berikut:

1. Pada tahap pertama, dilakukan pengambilan data berupa 100 data DNA virus MERS-CoV yang diperoleh dari *National Center for Biotechnology Information* (NCBI) dalam laman [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). Barisan-barisan DNA yang diperoleh berupa FASTA dan berjenis *complete genome*, *complete genome* dipilih karena diharapkan mampu memberikan informasi genetik yang lebih banyak.
2. Pada tahap kedua, barisan DNA yang diperoleh pada *Gen Bank* NCBI di situs <http://www.ncbi.nlm.nih.gov> memiliki bentuk dengan format FASTA (*string*), sehingga perlu untuk mengubah data tersebut ke dalam bentuk numerik. Mengubah data dalam bentuk *string* menjadi numerik dilakukan dengan ekstraksi ciri menggunakan *n-mers frequency*. Pada penelitian ini, ekstraksi ciri DNA akan dilakukan dengan melihat pola setiap susunan 4 pasangan nukleotida yang terdiri dari *Adenine* (A), *Guanine* (G), *Thymine* (T), dan *Cytosine* (C). Pola kemunculannya adalah  $4^n$  dengan  $n \geq 1$  dimana  $n$  elemen bilangan bulat positif (Liu, *et.al.*, 2012). Pada proses ekstraksi ciri,  $n = 3$  dengan pola kemunculannya adalah  $4^3 = 64$ , sehingga diperoleh 64 asam amino yang menjadi variabel pada penelitian ini. Hasil ekstraksi ciri barisan DNA MERS-CoV dengan menggunakan *n-mers frequency* yang dibantu oleh *software* RStudio dapat dilihat pada Gambar 3 merupakan potongan hasil ekstraksi dari barisan DNA virus MERS-CoV.

X	aaa	aac	aag	aat	aca	acc	acg	act	aga	a
1	581	449	516	619	567	349	199	635	436	
2	581	449	516	619	567	349	199	635	436	
3	571	438	511	603	555	344	201	626	427	
4	571	438	511	603	555	344	201	626	427	
5	571	438	511	602	554	345	201	626	426	
6	571	437	511	604	554	345	201	624	428	
7	571	438	512	601	554	345	201	626	427	
8	570	437	512	602	554	345	201	626	425	
9	571	437	512	603	554	345	201	626	427	
10	574	437	512	603	554	344	201	628	427	
11	574	438	509	597	561	343	198	622	426	
12	579	439	516	613	558	346	202	632	434	
13	578	441	515	612	556	345	203	632	434	
14	571	436	508	600	551	341	202	629	424	
15	567	436	508	602	552	342	202	628	425	
16	587	447	517	625	563	349	204	632	438	
17	588	447	517	624	563	347	204	632	437	

Sumber: Dokumentasi Pribadi

**Gambar 3. Potongan Hasil Ekstraksi Ciri 100 Barisan DNA MERS-CoV Menggunakan Software RStudio**

Nilai minimum hasil ekstraksi ciri DNA MERS-CoV adalah 116 dan nilai maksimum hasil ekstraksi ciri DNA MERS-CoV adalah 887. Hasil perhitungan ekstraksi ciri dengan menggunakan *n-mers frequency* pada 100 barisan DNA MERS-CoV.

3. Pada tahap ketiga, dilakukan proses normalisasi dengan menggunakan *min-max normalization* dari data 100 barisan DNA MERS-CoV yang telah dilakukan ekstraksi ciri DNA MERS-CoV. Proses perhitungan normalisasi pada penelitian ini menggunakan bantuan software *RStudio* dengan nilai maksimum terbaru adalah 1 dan nilai minimum terbaru adalah 0. Hasil normalisasi dari ekstraksi ciri DNA yang dibantu dengan menggunakan software *RStudio* dapat dilihat pada Gambar 4.

X	aaa	aac	aag	aat	aca	acc
0.000000000	0.6546275	0.5056433	0.5812641	0.6975169	0.6388262	0.3927765
0.001128668	0.6546275	0.5056433	0.5812641	0.6975169	0.6388262	0.3927765
0.002257336	0.6433409	0.4932280	0.5756208	0.6794582	0.6252822	0.3871332
0.003386005	0.6433409	0.4932280	0.5756208	0.6794582	0.6252822	0.3871332
0.004514673	0.6433409	0.4932280	0.5756208	0.6783296	0.6241535	0.3882619
0.005643341	0.6433409	0.4920993	0.5756208	0.6805869	0.6241535	0.3882619
0.006772009	0.6433409	0.4932280	0.5767494	0.6772009	0.6241535	0.3882619
0.007900677	0.6422122	0.4920993	0.5767494	0.6783296	0.6241535	0.3882619
0.009029345	0.6433409	0.4920993	0.5767494	0.6794582	0.6241535	0.3882619
0.010158014	0.6467269	0.4920993	0.5767494	0.6794582	0.6241535	0.3871332
0.011286682	0.6467269	0.4932280	0.5733634	0.6726862	0.6320542	0.3860045
0.012415350	0.6523702	0.4943567	0.5812641	0.6907449	0.6286682	0.3893905
0.013544018	0.6512415	0.4966140	0.5801354	0.6896163	0.6264108	0.3882619
0.014672686	0.6433409	0.4909707	0.5722348	0.6760722	0.6207675	0.3837472
0.015801354	0.6388262	0.4909707	0.5722348	0.6783296	0.6218962	0.3848758
0.016930023	0.6613995	0.5033860	0.5823928	0.7042889	0.6343115	0.3927765
0.018058691	0.6625282	0.5033860	0.5823928	0.7031603	0.6343115	0.3905192

Sumber: Dokumentasi Pribadi

**Gambar 4. Potongan Hasil Normalisasi dari Ekstraksi Ciri DNA MERS-CoV yang dibantu dengan Menggunakan Software RStudio**

Hasil 100 barisan DNA MERS-CoV yang telah diekstraksi ciri dan dinormalisasi kemudian akan menjadi matriks objek ( $x$ ) yang menjadi input pada saat algoritma PAM.

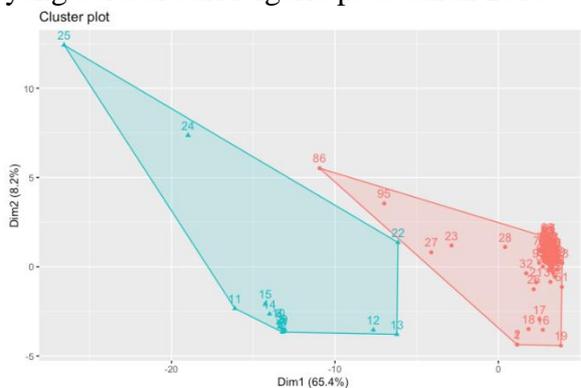
4. Pada tahap keempat, dilakukan *clustering* barisan DNA MERS-CoV menggunakan algoritma PAM *clustering*. Hasil *clustering* barisan DNA MERS-CoV menggunakan algoritma PAM *clustering* diperoleh 2 *cluster* dengan *medoid* dari masing-masing *cluster* dapat dilihat pada Gambar 5.

ID	X	aaa	aac	aag	aat	aca
[1,]	58	0.064334086	0.6693002	0.5045147	0.5869074	0.7020316
[2,]	8	0.007900677	0.6422122	0.4920993	0.5767494	0.6783296
	acc	acg	act	aga	agc	agg
[1,]	0.3905192	0.2291196	0.7167043	0.5022573	0.3837472	0.3984199
[2,]	0.3882619	0.2257336	0.7054176	0.4785553	0.3837472	0.3927765
	ata	atc	atg	att	caa	cac
[1,]	0.5428894	0.3905192	0.8002257	0.8397291	0.6467269	0.4819413
[2,]	0.5248307	0.3679458	0.7889391	0.8194131	0.6320542	0.4796840
	cat	cca	ccc	ccg	cct	cga
[1,]	0.5485327	0.4322799	0.2573363	0.1354402	0.4650113	0.1455982
[2,]	0.5316027	0.4243792	0.2426637	0.1354402	0.4604966	0.1501129
	cgg	cgt	cta	ctc	ctg	ctt
[1,]	0.1422122	0.3081264	0.6410835	0.4740406	0.6444695	0.8227991
[2,]	0.1331828	0.3036117	0.6264108	0.4627540	0.6399549	0.8069977
	gac	gag	gat	gca	gcc	gcg
[1,]	0.3555305	0.3318284	0.5225734	0.5135440	0.2765237	0.217833
[2,]	0.3566591	0.3216704	0.5056433	0.5056433	0.2731377	0.214447
	gga	ggc	ggg	ggt	gta	gtc
[1,]	0.3115124	0.3702032	0.1794582	0.5158014	0.5270880	0.3860045
[2,]	0.3103837	0.3600451	0.1783296	0.5124153	0.5124153	0.3814898
	gtt	taa	tac	tag	tat	tca
[1,]	0.8352144	0.7065463	0.6410835	0.4537246	0.8002257	0.5711061
[2,]	0.8171558	0.6817156	0.6151242	0.4390519	0.7866817	0.5609481
	tcg	tct	tga	tgc	tgg	tgt
[1,]	0.2121896	0.7562077	0.6918736	0.6986456	0.6568849	0.9909707
[2,]	0.2076749	0.7381490	0.6772009	0.6851016	0.6568849	0.9672686
	ttc	ttg	ttt			
[1,]	0.6546275	0.9525959	0.9920993			
[2,]	0.6534989	0.9300226	0.9853273			

Sumber: Dokumentasi Pribadi

**Gambar 5. Medoid dari Hasil Clustering Barisan DNA MERS-CoV**

Pada Gambar 5. terlihat bahwa *medoid* atau pusat *cluster* dari *cluster* 1 adalah barisan DNA MERS-CoV M58 yaitu barisan DNA MERS-CoV yang berasal dari Qasim pada tahun 2016 dan *medoid* dari *cluster* 2 adalah barisan DNA MERS-CoV M8 yaitu barisan DNA MERS-CoV yang berasal dari Nigeria pada tahun 2016.

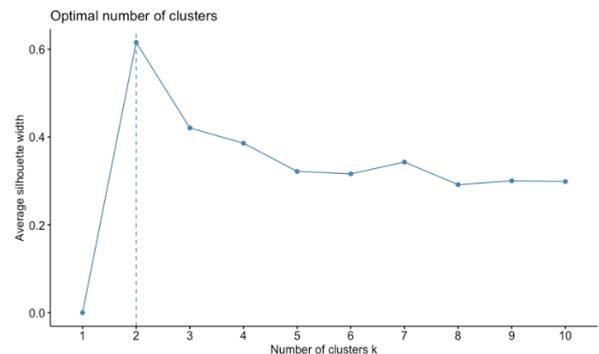


Sumber: Dokumentasi Pribadi

**Gambar 6. Hasil Clustering Barisan DNA MERS-CoV menggunakan Algoritma PAM**

5. Pada tahap kelima, dilakukan perhitungan *Silhouette Coefficient* dan diperoleh lebar *average silhouette* pada masing-masing *cluster* yaitu untuk *cluster* 1 diperoleh nilai *average silhouette* 0.6140725, sedangkan nilai

*average silhouette* pada *cluster* 2 yaitu 0.6219940. Pengujian pengaruh jumlah *cluster* dilakukan dengan memasukkan nilai *k* yang berbeda-beda yaitu  $k = 1, 2, \dots, 10$ . Nilai *silhouette coefficient* yang tertinggi berada pada jumlah *cluster* 2 yaitu 0,61534 dapat dilihat pada Gambar 7.



Sumber: Dokumentasi Pribadi

**Gambar 7. Grafik hasil pengujian pengaruh jumlah cluster terhadap Silhouette Coefficient**

6. Pada tahap keenam, dilakukan analisis dari hasil *clustering* dengan menggunakan algoritma PAM *clustering*. Pada Gambar 8 terlihat bahwa *medoid* atau pusat *cluster* dari *cluster* 1 adalah barisan DNA MERS-CoV m58 yaitu barisan DNA MERS-CoV yang berasal dari Qasim pada tahun 2016 dan *medoid* dari *cluster* 2 adalah barisan DNA MERS-CoV m8 yaitu barisan DNA MERS-CoV yang berasal dari Nigeria pada tahun 2016. *Maximum dissimilarity* adalah nilai ketidakmiripan antar *cluster*. Nilai *maximum dissimilarity* dari *cluster* 1 adalah 0,09208162, sedangkan nilai *maximum dissimilarity* dari *cluster* 2 adalah 0,09354338. *Average dissimilarity* merupakan nilai ketidakmiripan antar objek di dalam *cluster* yang sama. Semakin kecil nilai *average dissimilarity* maka semakin baik *cluster* yang terbentuk. Nilai *average*

*dissimilarity* dari *cluster* 1 adalah 0,3518736, sedangkan nilai *average dissimilarity* dari *cluster* 2 adalah 0,3044489. Kemudian, diameter menunjukkan panjang diameter dari masing-masing *cluster*. Diameter dari *cluster* 1 adalah 0,1291228, sedangkan diameter dari *cluster* 2 adalah 0,1206725. *Separation* atau pemisah pada *cluster* 1 dan 2 memiliki nilai yang sama yaitu 0,0262764.

	size	max_diss	av_diss	diameter	separation
[1,]	84	0.09208162	0.03518736	0.1291228	0.0262764
[2,]	16	0.09354338	0.03044489	0.1206725	0.0262764

Sumber: Dokumentasi Pribadi

**Gambar 8. Karakteristik Hasil Clustering menggunakan Algoritma PAM Clustering**

**Tabel 1. Hasil Clustering Barisan DNA MERS-CoV menggunakan Algoritma PAM**

Cluster	Anggota	Jumlah
1	M1, M2, M16, M17, M18, M19, M20, M21, M23, M26, M27, M28, M29, M30, M31, M32, M33, M34, M35, M36, M37, M38, M39, M40, M41, M42, M43, M44, M45, M46, M47, M48, M49, M50, M51, M52, M53, M54, M55, M56, M57, M58, M59, M60, M61, M62, M63, M64, M65, M66, M67, M68, M69, M70, M71, M72, M73, M74, M75, M76, M77, M78, M79, M80, M81, M82, M83, M84, M85, M86, M87, M88, M89, M90, M91, M92, M93, M94, M95, M96, M97, M98, M99, M100	84
2	M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15, M22, M24, M25	16

Pada Tabel 1 terlihat bahwa jumlah anggota pada *Cluster* 1 adalah 84 barisan DNA MERS-CoV dan jumlah anggota pada *Cluster* 2 adalah 16 barisan DNA MERS-CoV. *Cluster* 1 terdiri dari barisan DNA MERS-CoV yang berasal Kenya tahun 2018, Ethiopia tahun 2017, Egypt tahun 2018, China tahun 2018, Hu-France

tahun 2013, England tahun 2013, Thailand tahun 2015, Korea Selatan tahun 2015, Qatar tahun 2015, Riyadh tahun 2018, Riyadh tahun 2016, Riyadh tahun 2015, Jeddah tahun 2017, Riyadh tahun 2016, Qasim tahun 2016, Madinah tahun 2016, Riyadh tahun 2017, Artawiyah tahun 2016, Jeddah tahun 2016, Taif tahun 2014, Riyadh tahun 2015, Jeddah tahun 2015, Najran tahun 2015, Kharj tahun 2015, dan Khobar tahun 2015. Sedangkan, *Cluster* 2 terdiri dari barisan DNA MERS-CoV yang berasal Nigeria tahun 2016, Nigeria tahun 2015, Burkina tahun 2015, Morocco tahun 2015, dan Hu-France tahun 2013.

#### 4. SIMPULAN

Pada penelitian ini dapat disimpulkan:

1. Algoritma PAM dapat diimplementasikan pada *clustering* barisan DNA MERS-CoV. Data barisan DNA MERS-CoV yang digunakan sebanyak 100 data set. Proses awal yang dilakukan adalah dengan data set yang memiliki bentuk dengan format FASTA (*string* ke dalam bentuk numerik. Mengubah data dalam bentuk *string* menjadi numerik dilakukan dengan ekstraksi ciri menggunakan *n-mers frequency*. Kemudian, dilakukan proses normalisasi dengan menggunakan *min-max normalization*, selanjutnya dilakukan *clustering* barisan DNA MERS-CoV menggunakan algoritma PAM *clustering*. Setelah itu, menghitung nilai *silhouette coefficient* untuk mengetahui kualitas *cluster* yang dihasilkan.
2. Implementasi algoritma PAM terhadap barisan DNA MERS-CoV diperoleh 2 *cluster* dengan nilai *silhouette coefficient* yang tertinggi berada pada jumlah *cluster* 2 yaitu 0,61534. Jumlah anggota pada *Cluster* 1 adalah 84 barisan DNA MERS-CoV dan jumlah anggota pada *Cluster* 2 adalah 16

barisan DNA MERS-CoV. Kemudian, *medoid* atau pusat *cluster* dari *cluster* 1 adalah barisan DNA MERS-CoV M58 yaitu barisan DNA MERS-CoV yang berasal dari Qasim pada tahun 2016 dan *medoid* dari *cluster* 2 adalah barisan DNA MERS-CoV M8 yaitu barisan DNA MERS-CoV yang berasal dari Nigeria pada tahun 2016.

#### DAFTAR PUSTAKA

- [1] K. Kamal, K. A. Sugeng, and T. Siswantining, "Implementation algorithm modification maximum standard deviation reduction in graph clustering using matrix complement as input," *AIP Conf. Proc.*, vol. 2023, 2018.
- [2] A. Bustamam, E. D. Ulul, H. F. A. Hura, and T. Siswantining, "Implementation of hierarchical clustering using k-mer sparse matrix to analyze MERS-CoV genetic relationship," *AIP Conf. Proc.*, vol. 1862, no. July, 2017.
- [3] R. V. S. A. Morihito *et al.*, "Identification of Changes of DNA Structures on Cancer CII Form Using Graph Decomposition," *J. Ilm. Sains*, vol. 17 No.2, pp. 153–160, 2017.
- [4] T. Taslim and F. Fajrizal, "Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 7, no. 2, pp. 108–114, 2016.
- [5] L. P. Rizby, Marji, and L. Muflikhah, "Clustering Pasien Kanker Berdasarkan Struktur Protein Dalam Tubuh," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 10, pp. 3810–3816, 2018.