

# PERBANDINGAN *K-MEANS* DAN *K-MEDOIDS* UNTUK KLASSTERING TINGKAT STRES PADA MANUSIA

Arya Triansyah<sup>1)</sup> Dyah Erny Herwindiati<sup>2)</sup> Janson Hendryli<sup>3)</sup>

<sup>1)2)3)</sup> Teknik Informatika, FTI, Universitas Tarumanagara

Jl. Letjen S Parman no 1, Jakarta 11440, Indonesia

<sup>1)</sup>Email: [arya.535170104@stu.untar.ac.id](mailto:arya.535170104@stu.untar.ac.id) <sup>2)</sup>Email: [dyahh@fti.untar.ac.id](mailto:dyahh@fti.untar.ac.id) <sup>3)</sup>Email: [jansonh@fti.untar.ac.id](mailto:jansonh@fti.untar.ac.id)

## ABSTRACT

*Society is faced with various problems as a result of progress and development of the times. Things in social relationships and demands from an expectation in achievement but not being met, from these inability and demands can cause stress in a person. Stress is the body's response caused by demands from outside the individual that exceed the ability to meet the demands to overcome and resolve the problem. The need to respond and manage stress well so that the quality of life becomes better, with clustering it can make it easier to group data. The clustering technique used is the K-Means and K-Medoids methods which partition the data into clusters. Comparison of cluster results used the use of a covariance matrix. So that in the comparison of the K-Means method  $k=2$  and  $k=3$ , the best one is  $k=3$  because the determinant of the covariance matrix is smaller, namely  $-1.4709e-11$ . In the comparison of the K-Medoids method  $k=2$  and  $k=3$ , the best one is  $k=3$  because the determinant of the covariance matrix is smaller, namely  $-1.4285e-11$ . Continued comparison of the two methods, namely K-Means and K-Medoids, the best is K-Medoids with a smaller covariance determinant than K-Means.*

## Key words

*Clustering, K-Means, K-Medoids*

## 1. Pendahuluan

Stres merupakan istilah yang banyak diartikan bermacam-macam pendapat. Dalam arti umum stres merupakan pola reaksi serta adaptasi umum, dalam arti pola reaksi menghadapi stresor, yang dapat berasal dari dalam maupun luar individu yang bersangkutan, dapat nyata maupun tidak nyata sifatnya. Masyarakat dibenturkan persoalan bermacam-macam sebagai akibat kemajuan dan perkembangan zaman. Perkara dalam hubungan sosial, target pencapaian dalam suatu kebutuhan, dan harapan seringkali juga ada rintangan dan hambatan, dari ketidaksanggupan dan tuntutan

tersebut bisa menimbulkan stress dalam diri seseorang. Beberapa sumber yang jadi pemicu stress sering disangkutkan dengan berbagai masalah, dimulai dari masalah pekerjaan, tugas, tekanan, krisis dan hubungan sosial atau dengan orang lain. Sehingga tekanan dan tuntutan itulah yang membawa manusia dengan kondisi stress ditingkat yang berbeda-beda.

Pembuatan rancangan ini bertujuan membuat aplikasi perangkat lunak klastering dataset tingkat stres pada manusia yang ditujukan untuk dapat menyikapi dan mengelola stres dengan baik.

## 2. Metodologi Penelitian

### 2.1 K-Means

Algoritma *K-Means* masuk ke penerapan data mining klastering. Clustering membagi kumpulan data ke dalam beberapa kelompok dimana kesamaan dalam sebuah kelompok adalah lebih besar dari pada diantara kelompok-kelompok. Tujuan dari klastering adalah pengelompokan data ke dalam beberapa klaster berdasarkan tingkat kemiripan.[1]

Langkah-langkah algoritma *K-Means* :

1. Menyiapkan dataset.
2. Menentukan jumlah cluster ( $k$ ).
3. Pilih titik centroid secara acak
4. Mengelompokkan data sehingga terbentuk  $K$  cluster dengan titik centroid .

Dengan rumus :

$$\sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Dimana:

$d(x, y)$  adalah jarak antara data  $x$  ke data  $y$

$x_i$  adalah data testing ke- $i$

$y_i$  adalah data training ke- $i$

5. Perbarui nilai titik centroid.

Berikut rumus yang digunakan :

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \quad (2)$$

Keterangan :

$\mu_k$  adalah titik *centroid* dari klaster ke-K

$N_k$  banyaknya data pada klaster ke-K

$x_i$  adalah data ke-i pada klaster ke-K

6. Pengulangan langkah 3 hingga 5 sampai nilai dari titik centroid tidak lagi berubah.

### 2.2 K-Medoids

Algoritma *K-Medoids* atau dikenal dengan PAM (*Partitioning Around Medoids*) merupakan salah satu algoritma yang digunakan untuk mengelompokkan data. Algoritma ini menggunakan objek pada kumpulan objek untuk mewakili sebuah klaster. Dengan objek yang terpilih untuk mewakili sebuah klaster disebut medoids. Klaster dibangun dengan menghitung kedekatan yang dimiliki antara medoids dengan objek non-medoids.[2]

Langkah-langkah algoritma *K-Medoids*

1. Menyiapkan dataset
2. Menentukan nilai K (Jumlah Klaster)
3. Pilih secara acak medoid awal sebanyak k dari n data
4. Hitung jarak masing-masing objek ke medoid sementara.

Berikut rumus yang digunakan :

$$\sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Keterangan :

$\mu_k$  adalah titik *centroid* dari klaster ke-K

$N_k$  banyaknya data pada klaster ke-K

$x_i$  adalah data ke-i pada klaster ke-K

5. Tandai jarak terdekat objek ke medoid dan hitung total nya.
6. Tentukan anggota cluster terhadap medoid sementara
7. Lakukan iterasi medoid
8. Hitung total simpangan (S) dengan menghitung nilai total jarak baru – total jarak lama. Jika didapatkan  $S < 0$ , tukarlah objek dengan data *cluster* untuk membuat sekumpulan k objek baru sebagai medoid.

### 3. Pengujian Dan Evaluasi

Dari tahap pengujian ini bertujuan untuk mengetahui kualitas dari program apakah program berbasis web untuk klastering tingkat stres pada

manusia dengan perbandingan metode *K-Means* dan *K-Medoids* ini sudah dibuat sesuai dengan tujuan dari perancangan aplikasi.

Dari pengujian yang dilakukan dengan menggunakan 2001 dataset. Dataset *training* yang diperoleh ini dari sebuah penelitian atau studi kasus yang mana penggunaannya membutuhkan alat sensor untuk menentukan data tiga parameter pada dataset. dengan melakukan klastering dengan algoritma *K-Means* dan *K-Medoids* dengan menentukan banyaknya klaster tujuh.

Dengan nilai centroid awal acak :

#### 3.1 Pengujian Metode K-Means

Pengujian terhadap metode *K-Means* yang terdapat di sistem aplikasi berbasis web menggunakan data yang telah disediakan dalam forum kaggle dan merupakan hasil dari studi kasus. Berdasarkan data yang diperoleh yaitu 2001 data yang digunakan. Pengujian yang dilakukan kali ini dengan menggunakan k=2 untuk tidak stress dan stress, k= 3 untuk stres rendah, stres sedang, dan stres tinggi. Dan untuk nilai centroid awal acak yang digunakan pada *K-Means* menggunakan sistem random.

Diketahui dalam pemilihan nilai centroid awal secara acak pada sistem dengan K=2 yaitu :

Tabel 1 Data Centroid Awal Acak

	Data ke-	Humidity	Temperature	Step count
C1	1064	24.45	93.45	142
C2	471	20.97	89.97	125

Diketahui dalam pemilihan nilai centroid awal secara acak pada sistem dengan K=3 yaitu :

Tabel 2 Data Centroid Awal Acak

	Data ke-	Humidity	Temperature	Step count
C1	1981	22.69	91.69	115
C2	1494	28.96	97.96	167
C3	1684	13.23	82.23	50

Kemudian melanjutkan dengan langkah-langkah algoritma *K-Means* hingga pemberhentian iterasi terjadi sampai nilai dari titik centroid tidak lagi berubah dari iterasi sebelumnya. Total iterasi yang dilakukan k=2 sebanyak 12 dan total iterasi k=3 sebanyak 18. Diperoleh dari hasil klastering yang dilakukan:

Tabel 3 Hasil Klaster K-Means

	Banyak Anggota
Tidak Stres	1024 data

Stres	977 data
-------	----------

Tabel 4 Hasil Kluster *K-Means*

	Banyak Anggota
Stres Rendah	648 data
Stres Sedang	676 data
Stres Tinggi	677 data

### 3.2 Pengujian Metode *K-Means*

Pada pengujian terhadap metode *K-Medoids* yang terdapat di sistem aplikasi berbasis web menggunakan data yang sama terdapat pada pengujian metode *K-Means*. Berdasarkan data yang diperoleh yaitu 2001 data yang digunakan. Pengujian yang dilakukan kali ini dengan menggunakan  $k=2$  untuk tidak stress dan stress, lalu  $k=3$  untuk stres rendah, stres sedang, dan stres tinggi. Dan untuk nilai centroid awal acak pada *K-Medoids*  $k=2$  dan  $k=3$  juga menggunakan nilai centroid acak dari data yang sama digunakan di *K-Means*. Lihat pada **Tabel 1** dan **Tabel 2**.

Kemudian melanjutkan dengan langkah-langkah algoritma *K-Medoids* hingga pemberhentian iterasi dengan perhitungan total simpangan yaitu : Jumlah jarak kedekatan iterasi baru – jumlah jarak kedekatan iterasi sebelumnya, jika selisih jarak kedekatan  $>0$ , maka iterasi berhasil dan kluster ada pada iterasi sebelumnya. Total iterasi yang dilakukan  $k=2$  sebanyak 3 dan total iterasi  $k=3$  sebanyak 2. Diperoleh dari hasil klastering yang dilakukan: Diperoleh hasil klastering dari *K-Medoids*:

Tabel 5 Hasil Kluster *K-Medoids*

	Banyak Anggota
Tidak Stres	540 data
Stres	1461 data

Tabel 6 Hasil Kluster *K-Medoids*

	Banyak anggota
Klaster1	566 data
Klaster2	596 data
Klaster3	839 data

### 3.3 Pengujian Kovarian *K-Means*

Pengujian ini dilakukan untuk mendapatkan matriks kovarian dari  $K=2$ , dan  $K=3$ . Pada pengujian pertama yang dilakukan dengan  $K=2$  yaitu kelas : tidak stres, dan stres. Diperoleh :

Tabel 7 Matriks kovarian tidak stres *K-Means*

9.414248	9.405055	61.00535
----------	----------	----------

9.405055	9.414248	61.00535
61.00535	61.00535	861.0299

Dengan nilai determinan dari kelas tidak stres =  $-5.4426e-11$ .

Tabel 8 Matriks kovarian stres *K-Means*

8.836797	8.827752	25.72101
8.827752	8.836797	25.72101
25.72101	25.72101	803.5136

Dengan nilai determinan dari kelas stres =  $-2.2871e-11$ .

Pengujian kedua yang dilakukan dengan  $k=3$  yaitu : kelas stres rendah, stres sedang, dan stres tinggi. Diperoleh :

Tabel 9 Matriks kovarian stres rendah *K-Means*

14.2191	14.2191	56.5718
14.2191	14.2191	56.5718
56.5718	56.5718	370.6645

Dengan nilai determinan dari kelas tidak stres =  $-1.4709e-11$ .

Tabel 10 Matriks kovarian stres sedang *K-Means*

4.1885	4.1885	3.3888
4.1885	4.1885	3.3888
3.3888	3.3888	381.4140

Dengan nilai determinan dari kelas tidak stres =  $-8.4522e-12$ .

Tabel 11 Matriks kovarian stres tinggi *K-Means*

6.8533	6.8533	3.3497
6.8533	6.8533	3.3497
3.3497	3.3497	382.5271

Dengan nilai determinan dari kelas tidak stres =  $-1.1592e-11$ .

### 3.4 Pengujian Kovarian *K-Medoids*

Pengujian ini dilakukan untuk mendapatkan matriks kovarian dari  $K=2$ , dan  $K=3$ . Pada pengujian pertama yang dilakukan dengan  $K=2$  yaitu kelas : tidak stres, dan stres. Diperoleh :

Tabel 12 Matriks kovarian tidak stres *K-Medoids*

4.1909	4.1909	2.4337
4.1909	4.1909	2.4337
2.4337	2.4337	243.3673

Dengan nilai determinan dari kelas tidak stres =  $-9.0061e-13$ .

Tabel 13 Matriks kovarian stres *K-Medoids*

22.5167	22.5167	156.9343
22.5167	22.5167	156.9343
156.8269	156.9343	1.8214e+03

Dengan nilai determinan dari kelas tidak stres = 2.9103e-10.

Pengujian kedua yang dilakukan dengan k=3 yaitu : kelas stres rendah, stres sedang, dan stres tinggi. Diperoleh :

Tabel 14 Matriks kovarian stres rendah *K-Medoids*

11.7675	11.7675	43.0836
11.7675	11.7675	43.0836
43.0836	43.0836	272.0715

Dengan nilai determinan dari kelas stres rendah = -4.7798e-12.

Tabel 15. Matriks kovarian stres sedang *K-Medoids*

4.1800	4.1800	2.5402
4.1800	4.1800	2.5402
2.5402	2.5402	296.5473

Dengan nilai determinan dari kelas stres sedang = -2.1904e-12.

Tabel 16. Matriks kovarian stres tinggi *K-Medoids*

6.8835	6.8835	4.8491
6.8835	6.8835	4.8491
4.8491	4.8491	587.5632

Dengan nilai determinan dari kelas stres tinggi = -1.4285e-11.

### 3.5 Evaluasi Dan Hasil

Pada tahap ini akan menampilkan hasil yang didapat dari proses algoritma *K-Means* dan *K-Medoids* dengan menggunakan K=2 dan K=3. Dengan data yang diperoleh dan digunakan berjumlah 2001 dataset.

Tabel 6 Evaluasi Hasil *K-Means*

<i>K-Means</i> K=2	Tidak Stres	Stres	
Banyak Anggota	1024	977	
Determinan kovarian	- <b>5.4426e-11</b>	-2.2871e-11	
Total Iterasi	12		
<i>K-Means</i> K=3	Stres Rendah	Stres Sedang	Stres Tinggi

Banyak Anggota	648	676	677
Determinan kovarian	- <b>1.4709e-11</b>	-8.4522e-12	-1.1592e-11
Total Iterasi	18		

Pada tabel diatas menjelaskan hasil dari proses pada algoritma *K-Means* K=2, dan K=3.

Pada *K-Means* k=2 total iterasi diperoleh = 12.

Tidak Stres

- Memiliki banyak anggota 1024 dataset.
- Determinan dari matriks kovarian kelas tidak stres adalah -5.4426e-11

Stres

- Memiliki banyak anggota 977 dataset.
- Determinan dari matriks kovarian kelas stres adalah -2.2871e-11

Pada *K-Means* k=3 total iterasi diperoleh = 18.

Stres Rendah

- Memiliki banyak anggota 648 dataset.
- Determinan dari matriks kovarian kelas stres rendah adalah -1.4709e-11

Stres Sedang

- Memiliki banyak anggota 676 dataset.
- Determinan dari matriks kovarian kelas stres sedang adalah -8.4522e-12

Stres Tinggi

- Memiliki banyak anggota 677 dataset.
- Determinan dari matriks kovarian kelas stres tinggi adalah -1.1592e-11.

Dari perbandingan nilai determinan kovarian terbesar dari *K-Means* K=2 dan K=3. Diperoleh yang terbaik adalah K=3, karena determinan matriks kovariannya lebih kecil. Yaitu -1.4709e-11. Berarti untuk *K-Means* K=3 lebih baik.

Tabel 6 Evaluasi Hasil *K-Medoids*

<i>K-Medoids</i> K=2	Tidak Stres	Stres	
Banyak Anggota	540	1461	
Determinan kovarian	- 9.0061e-13	2.9103e-10	
Total Iterasi	3		
<i>K-Medoids</i> K=3	Stres Rendah	Stres Sedang	Stres Tinggi
Banyak Anggota	566	596	839
Determinan kovarian	- 4.7798e-12	-2.1904e-12	-1.4285e-11

Total Iterasi	2
---------------	---

Pada tabel diatas menjelaskan hasil dari proses pada algoritma *K-Medoids* K=2, dan K=3.

Pada *K-Medoids* k=2 total iterasi diperoleh = 3.

Tidak Stres

- Memiliki banyak anggota 540 dataset.
- Determinan dari matriks kovarian kelas tidak stres adalah  $-9.0061e-13$

Stres

- Memiliki banyak anggota 1461 dataset.
- Determinan dari matriks kovarian kelas stres adalah  $2.9103e-10$

Pada *K-Medoids* k=3 total iterasi diperoleh = 2.

Stres Rendah

- Memiliki banyak anggota 566 dataset.
- Determinan dari matriks kovarian kelas stres rendah adalah  $-4.7798e-12$

Stres Sedang

- Memiliki banyak anggota 596 dataset.
- Determinan dari matriks kovarian kelas stres sedang adalah  $-2.1904e-12$

Stres Tinggi

- Memiliki banyak anggota 839 dataset.
- Determinan dari matriks kovarian kelas stres tinggi adalah  $-1.4285e-11$ .

Dari perbandingan nilai determinan kovarian terbesar dari *K-Medoids* K=2 dan K=3. Diperoleh yang terbaik adalah K=3, karena determinan matriks kovariannya lebih kecil. Yaitu  $-1.4285e-11$ , sehingga untuk *K-Means* K=3 lebih baik.

Diperoleh hasil terbaik dari dimasing-masing metode *K-Means* dan *K-Medoids* :

- Hasil *K-Means* terbaik berada di K=3 dengan nilai  $-1.4709e-11$
- Hasil *K-Medoids* terbaik berada di K=3 dengan nilai  $-1.4285e-11$

Dari kedua metode yaitu *K-Means* dan *K-Medoids* yang terbaik adalah *K-Medoids* karena determinan matriks kovarian medoid lebih kecil dibandingkan dengan *K-Means*.

Sehingga penggunaan determinan matriks kovarian dapat mengetahui model yang mana terbaik dari kedua metode *K-Means* dan *K-Medoids*. Dari hasil pengujian juga dapat dilihat bahwa hasil klustering dari kedua metode tersebut berjalan cukup baik.

#### 4. Kesimpulan

Berdasarkan rancangan dan pengujian yang telah dibuat pada aplikasi Perbandingan *K-Means* dan *K-Medoids* untuk Klustering Tingkat Stres

Pada Manusia, maka dapat disimpulkan sebagai berikut:

- Metode *K-Means* dan *K-Medoids* dapat digunakan untuk menentukan klustering pada dataset yang di input.
- Aplikasi dapat mengeluarkan hasil kelompok klustering sesuai dengan yang dirancang.
- Pada pengujian *K-Medoids* lebih baik dibandingkan *K-Means*.

#### REFERENSI

- [1] Xiang, Wan-li Xiang; Zhu, Ning; Ma, Shou-feng; Meng, Xue-lei; And An, Mei-qing. "A Dynamic Shuffled Differential Evolution Algorithm For Data Clustering," *Neurocomputing*, Vol. 158, Pp. 144–154, 2015.
- [2] Kaufman, Leonard. dan Rousseeuw, Peter J., *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken : John Wiley & Sons, 1990.
- [3] Alfiah, Febriyanti, Almadayani, Farizi Al, Danial, dan Widodo, Edy. "Analisis Clustering K-Medoids Berdasarkan Indikaator Kemiskinan di Jawa Timur Tahun 2020", *Jurnal Ilmiah Sains*, Vol. 22, Nomor 1, (April 2022), h. 3.

Arya Triansyah, saat ini sebagai mahasiswa Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara.

Dyah Erny Herwindiati, Ir., M.Si, Dr., Prof

Janson Hendryli, S.Kom. M.kom