

## Analisis Kluster Dalam Pengelompokan Kabupaten/Kota Di Provinsi Jambi Berdasarkan Penyakit Menular Menggunakan Metode K-Means

Hosni Mubarak<sup>1</sup>, Gusmi Kholijah<sup>2\*</sup>

<sup>1,2</sup>Matematika, Universitas Jambi  
[Hosni1262@gmail.com](mailto:Hosni1262@gmail.com)<sup>1</sup>, [gusmikholijah@unja.ac.id](mailto:gusmikholijah@unja.ac.id)<sup>2\*</sup>

*Diajukan* 6 Mei 2023    *Diperbaiki* 29 Juni 2023    *Diterima* 30 Juni 2023

### Abstrak

**Latar Belakang:** Kesehatan merupakan salah satu indikator pada pembangunan manusia, masalah kesehatan di Provinsi Jambi sedang dihadapkan pada beban ganda yang mana masalah penyakit menular masih menjadi masalah yang belum dapat diselesaikan.

**Tujuan:** Menerapkan analisis kluster metode *K-Means* untuk mengetahui hasil pengelompokan setiap kabupaten/kota di Provinsi Jambi, sehingga memudahkan Dinas Kesehatan Provinsi Jambi dalam mengevaluasi dan mengambil kebijakan terkait masalah penyakit menular.

**Metode:** Pengelompokan menggunakan metode yaitu *K-Means* untuk mengelompokan data. Penelitian ini dilakukan di Provinsi Jambi dengan menggunakan data dari Badan Pusat Statistik Provinsi Jambi dengan data penyakit menular adalah malaria, kusta, pneumonia, IMS dan HIV/AIDS.

**Hasil:** Diperoleh hasil tiga *centroid* dengan masing-masing karakteristik yang berbeda yaitu kluster 1 rendah pada penyakit malaria, pneumonia dan HIV/AIDS, *Klaster 2* tinggi pada penyakit malaria, pneumonia dan HIV/AIDS, dan kluster 3 sedang pada kusta, pneumonia dan HIV/AIDS.

**Kesimpulan:** Kabupaten/kota di Provinsi Jambi berdasarkan penyakit menular dikelompokan menjadi 3 Klaster yaitu Klaster rendah dengan anggota Kabupaten Kerinci, Batang Hari, Muaro Jambi, Tanjung Jabung Timur, Tanjung Jabung Barat dan Kota Sungai Penuh. Klaster sedang dengan anggota Kabupaten Merangin, Sarolangun, dan Tebo. Kemudian klaster tinggi dengan keanggotaan Kabupaten Muaro Bungo dan Kota Jambi.

**Kata kunci:** Analisis Kluster, Metode *K-Means*, Penyakit Menular

### Abstract

**Background:** Health is one of the indicators in human development indicators, health problems in Jambi Province are being faced with a double burden where the problem of infectious diseases is still a problem that cannot be resolved.

**Objective:** Applying kluster analysis of the *K-Means* method to determine the results of grouping each district/city in Jambi Province, making it easier for the Jambi Provincial Health Office to evaluate and take policies related to infectious disease problems.

**Methods:** Grouping uses the *K-Means* method to group data. This study was conducted in Jambi Province using data from the Central Bureau of Statistics of Jambi Province with data on infectious diseases such as malaria, leprosy, pneumonia, STIs and HIV/AIDS.

**Results:** Three centroids were obtained with each different characteristic, namely *Klaster 1* low in malaria, pneumonia and HIV/AIDS, *Klaster 2* high in malaria, pneumonia and HIV/AIDS, and *Klaster 3* moderate in leprosy, pneumonia and HIV/AIDS.

**Conclusion:** Regencies / cities in Jambi Province based on infectious diseases are grouped into 3 Klusters, namely low klaster, medium klaster, and high klaster.

**Keywords :** Cluster Analysis, *K-Means* Methods, Infectious Disease

## PENDAHULUAN

Kesehatan merupakan hal yang penting bagi setiap manusia karena siapa saja bisa mengalami gangguan kesehatan, begitu pula pada manusia yang rentan terhadap berbagai penyakit. Masalah kesehatan di Provinsi Jambi dihadapkan pada beban ganda yang mana kesehatan termasuk bagian dari indikator pembangunan manusia, dimana penyakit menular di Provinsi Jambi masih menjadi masalah yang belum terselesaikan. Berdasarkan data yang dirilis oleh Badan Pusat Statistik Kota Jambi dalam empat tahun terakhir jumlah kasus penyakit menular mengalami peningkatan dimulai pada tahun 2017 untuk penyakit malaria ada 144 kasus dan mengalami peningkatan yang ditunjukkan pada tahun 2021 terdapat 13.751 kasus malaria, untuk penyakit kusta pada tahun 2017 tidak ada kasus yang terjadi dan pada tahun 2021 mengalami peningkatan menjadi 45 kasus, untuk penyakit pneumonia pada tahun 2017 tidak ada kasus yang terjadi dan pada tahun 2021 mengalami peningkatan sebanyak 1160 kasus, untuk penyakit IMS (Infeksi Menular Seksual) pada tahun 2017 terdapat 125 kasus dan pada tahun 2021 mengalami peningkatan menjadi 194 kasus, serta untuk penyakit HIV/AIDS pada tahun 2017 terdapat 94 kasus dan pada tahun 2021 mengalami peningkatan menjadi 800 kasus. Hal ini menyatakan bahwa masalah penyakit menular merupakan masalah yang sangat *urgent* dan harus di atasi (BPS, 2022).

Salah satu upaya yang dilakukan yakni dengan membentuk suatu pengelompokan kabupaten/kota di Provinsi Jambi ke dalam suatu kelompok yang memiliki karakteristik yang hampir sama atau kemiripan. Hal ini bertujuan untuk memberikan informasi terkait

dengan kesehatan pada masing-masing kabupaten/kota. Salah satu cara yang dapat dilakukan dalam pengelompokan ini yaitu menggunakan analisis kluster *K-Means*.

Dalam penelitian ini, Analisis kluster merupakan salah satu teknik multivariat yang bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Jambi berdasarkan penyakit menular. Metode *K-Means* merupakan metode klusterisasi yang mengelompokkan data berdasarkan titik pusat kluster (*centroid*) terdekat dengan data. Metode *K-Means* pernah dilakukan perbandingan dengan metode *K-Medoid* yang dilakukan oleh Arbin et al. (2016), Abbas et al. (2020) dan Yu et al. (2018) menyampaikan bahwa kedua model kluster memiliki *performance* dan kemampuan dalam pengelompokan. Tujuan dari metode ini adalah mengelompokkan data dengan memaksimalkan kemiripan suatu data dalam satu kluster dan meminimalkan kemiripan suatu data antar kluster lainnya (Sobari & Usman, 2013). Metode ini juga pernah digunakan oleh Carvalho et al. (2016) dalam mengelompokkan daerah yang terdampak perubahan iklim berdasarkan temperatur dan curah hujan yang menghasilkan kemampuan *K-Means* dalam mengelompokkan daerah-daerah di Eropa.

Penelitian selanjutnya terkait analisis kluster dengan menggunakan metode hierarki dilakukan untuk mengelompokkan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator kesejahteraan rakyat, menghasilkan tiga kluster dengan tingkat yang berbeda (Yulianto & Hidayatul, 2014). Penelitian lainnya terkait analisis kluster dengan menggunakan metode *K-Means* juga pernah dilakukan untuk mengelompokkan kabupaten/kota di Provinsi Maluku berdasarkan indikator pembangunan manusia, dengan menggunakan metode *K-Means* diperoleh tiga kluster dengan tingkatan yang berbeda

(Talakua et al., 2017). Muzakir (2014) juga menggunakan algoritma *K-Means Clustering* dalam nilai siswa untuk penentuan penerima beasiswa, diperoleh bahwa penentuan *centroid* berpengaruh pada pengklasteran yang menghasilkan hasil kluster yang berbeda. Algoritma *K-Medoids* juga dipakai oleh Triyanto (2015) untuk penentuan strategi pemasaran produk yang menghasilkan pengelompokan dalam 5 kluster. Selanjutnya pengelompokan *K-Modes* dilakukan oleh Faqih (2022) yang menghasilkan 3 kluster pada pengelompokan karakteristik angkutan sungai dan penyeberangan di Kabupaten Bojonegoro.

Analisis kluster memerlukan beberapa ukuran kemiripan dalam mengakses seberapa mirip atau berbeda objek-objek yang dikelompokkan dalam kluster yang sama. Pendekatan yang paling umum digunakan untuk mengukur kemiripan yaitu dengan melihat jarak (*distance*) antara pasangan objek. Apabila nilai jarak antara dua buah objek semakin besar maka semakin besar pula perbedaan yang ada pada dua buah objek tersebut, sehingga semakin kecil kemungkinan untuk menganggapnya ke dalam kelompok yang sama. Sedangkan, semakin kecil nilai jaraknya maka semakin kecil pula perbedaan yang terdapat pada kedua objek tersebut (Yulianto & Hidayatul, 2014).

Pada proses melakukan analisis kluster ada asumsi yang harus terpenuhi yaitu asumsi pertama adalah sampel yang digunakan harus mewakili populasi, hal ini dapat dilihat dari hasil uji KMO yang harus lebih besar dari 0,5 dan asumsi kedua adalah tidak terdapat multikolinieritas, hal ini dapat dilihat dari nilai korelasi antar variabel bebas. Pada penelitian ini apabila nilai korelasi

melebihi 0,8 maka dianjurkan untuk menghilangkan salah satu variabel dari dua variabel yang memiliki nilai korelasi lebih dari 0,8 (Putriana et al., 2016). Menurut Ridwan & Sunendiari (2021) Terdapat beberapa cara mendeteksi multikolinieritas salah satunya menggunakan nilai dari korelasi Pearson dengan cara mendeteksi multikolinieritasnya apabila korelasi kuat atau di atas 0,8 maka ditemukan masalah multikolinieritas.

Kluster *K-Means* merupakan salah satu metode pengklasteran data non hierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kluster dengan demikian metode ini membutuhkan jumlah kluster yang ditentukan terlebih dahulu sebagai  $k$ , dan kemudian membagi  $n$  objek kedalam  $k$  kluster. Kemiripan antar anggota dalam satu kluster sangat tinggi sedangkan kemiripan antar anggota dengan kluster yang lain sangat rendah. Kemiripan anggota pada kluster diukur dengan nilai kedekatan objek terhadap *mean* pada kluster atau *centroid* kluster (Muzakir, 2014)

Analisis kluster *K-Means* sangat simpel serta fleksibel dalam implementasi penyelesaian masalah. Hal ini dikarenakan pada perhitungan komputasinya tidak terlalu rumit dan metode ini dapat diimplementasikan pada segala bidang (Prianto & Bunyamin, 2020).

Menurut Asroni & Adrian (2015) langkah-langkah dalam melakukan pengelompokan dengan menggunakan metode *K-Means* sebagai berikut:

1. Menentukan banyaknya kluster.
  2. Menentukan bahwa *centroid* pertama adalah  $n$  data pertama dari data-data yang akan dikluster.
  3. Menghitung jarak antara *centroid* dengan masing-masing data.
  4. Jika penempatan data sudah sama dengan sebelumnya, maka iterasi berhenti. Jika tidak, kembali ke cara yang ke-b.
- Perhitungan yang dilakukan untuk

menghitung nilai *centroid* baru pada masing-masing *centroid* yaitu dengan menjumlahkan semua nilai pada setiap kelompok kluster yang sama dan membagikannya dengan jumlah data yang ada pada kluster tersebut. Menurut Nurliza (2018), jarak euclidean adalah perhitungan jarak dari dua buah titik dalam euclidean *space*, diperkenalkan oleh matematikawan dari Yunani untuk mempelajari hubungan antara sudut dan jarak. Euclidean ini biasanya diterapkan pada dua dimensi dan tiga dimensi. Tapi juga sederhana jika diterapkan pada dimensi yang lebih tinggi. Jarak euclidean merupakan jarak yang paling umum digunakan untuk data numerik, untuk dua titik data  $x$  dan  $y$  dalam ruang dimensi. Jarak euclidean adalah suatu metode pencarian kedekatan nilai jarak dari dua buah variabel. Penentuan nilai *centroid* ditentukan berdasarkan nilai *range* yang berada pada sumber data yang ada dengan melakukan pemilihan sesuai dengan nilai *centroid* yang dipilih.

Tujuan dalam penelitian ini adalah untuk mengetahui hasil kelompok setiap kabupaten/kota di Provinsi Jambi berdasarkan penyakit menular menggunakan metode *K-Means*. Manfaat dan kontribusi dalam penelitian ini adalah dapat dijadikan sebagai bahan pertimbangan dalam pengambilan atau penentuan kebijakan terkait masalah penyakit penular di Provinsi Jambi serta bermanfaat untuk masyarakat agar mengetahui kelompok kabupaten/kota yang termasuk ke dalam kluster dengan berbagai tingkat penyakit menular.

Pada penelitian ini digunakan analisis kluster dengan metode *K-Means* untuk mengelompokkan setiap kabupaten/kota di Provinsi Jambi berdasarkan penyakit menular. Hasil penelitian ini berguna untuk memudahkan mengetahui dan mengambil kebijakan terkait masalah

penyakit menular di setiap kabupaten/kota di Provinsi Jambi dengan harapan agar dapat mengurangi kejadian penyakit menular di Provinsi Jambi. Dengan demikian perlu diusulkan penelitian dengan judul “Analisis *Kluster* dalam Pengelompokan Kabupaten/Kota di Provinsi Jambi berdasarkan penyakit menular menggunakan metode *K-Means*”.

## METODE

### Desain Penelitian

Desain penelitian yang dipakai pada penelitian ini adalah penelitian kuantitatif, dengan jenis kuantitatif komparatif. Metode statistika yang dipakai pada penelitian ini yaitu metode kluster *K-Means* pada kasus penyakit menular yang menyerang manusia yang bermukim pada kabupaten/kota di Provinsi Jambi. Sumber data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi Jambi BPS (2022) yang memuat data kasus penyakit menular tahun 2021.

### Populasi dan Sampel

Penelitian ini menggunakan populasi kabupaten/kota di Indonesia mengenai kasus penyakit menular, sedangkan sampel yang diambil adalah kabupaten/kota di Provinsi Jambi dengan kasus penyakit menular

### Teknik Sampling

Cara pengambilan data pada penelitian ini disebut dengan mengambil ke pihak lain yaitu BPS (2022) yang termuat dalam *e-book* publikasi Provinsi Jambi dalam angka tahun 2022 yang disebut data sekunder metode kuantitatif.

### Subyek Penelitian

Subyek penelitian ini adalah penduduk yang terserang penyakit menular di Provinsi Jambi berdasarkan kabupaten/kota pada kurun waktu tahun 2021.

## Analisis Kluster Dalam Pengelompokan Kabupaten/Kota Di Provinsi....

Adapun variabel yang digunakan dalam penelitian ini disajikan pada Tabel 1 Seperti berikut:

**Tabel 1.** Variabel Penelitian

No	Variabel	Definisi	Satuan
1	$X_1$	Jumlah kasus malaria	Jiwa
2	$X_2$	Jumlah kasus kusta	Jiwa
3	$X_3$	Jumlah kasus pneumonia	Jiwa
4	$X_4$	Jumlah kasus IMS	Jiwa
5	$X_5$	Jumlah kasus HIV/AIDS	Jiwa

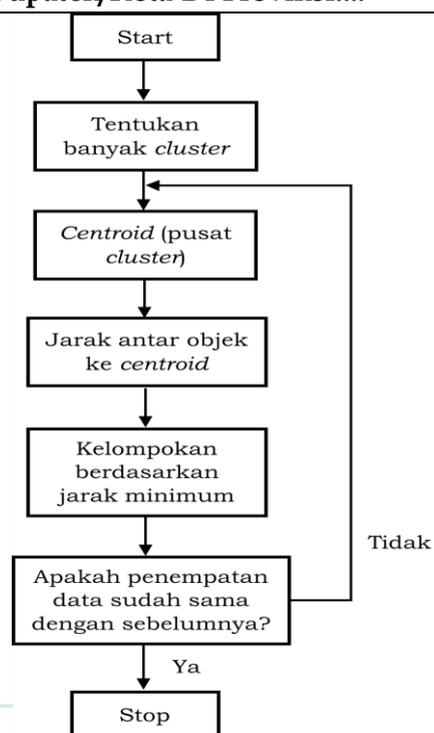
### Teknik Analisis Data

Teknik analisis data pada penyelesaian masalah di penelitian ini adalah analisis kluster *K-Means*.

Adapun Tahapan penyelesaian data dalam penelitian ini menggunakan analisis kluster metode *K-Means* sesuai Asroni & Adrian (2015) yaitu, sebagai berikut:

1. Menentukan banyaknya kluster.
2. Menentukan bahwa *centroid* pertama adalah  $n$  data pertama dari data-data yang akan dikluster.
3. Menghitung jarak antara *centroid* dengan masing-masing data.
4. Jika penempatan data sudah sama dengan sebelumnya, maka stop. Jika tidak, kembali ke cara yang ke-3.

Berdasarkan tahapan penyelesaian data, berikut adalah diagram alir dalam analisis kluster metode *K-Means* yaitu:



**Gambar 1.** Tahapan Analisis Kluster Metode *K-Means*

Penentuan jarak digunakan rumus euclidean sebagai berikut :

$$D(p, q) = \sqrt{\sum_{k=1}^n (p_k - q_k)^2} \quad (1)$$

dengan:

$D(p, q)$  : Jarak antara objek  $p$  dengan objek  $q$

$p_k$  : Nilai objek  $p$  ke  $k$

$q_k$  : Nilai objek  $q$  ke  $k$

$n$  : Banyaknya peubah yang diamati

### HASIL DAN PEMBAHASAN

Analisis kluster menggunakan metode *K-Means*, terlebih dahulu harus terpenuhi asumsi-asumsi analisis kluster, yaitu:

Asumsi pertama, sampel yang diambil harus mewakili populasi yang ada dapat dilihat dari hasil uji *Kaiser-Meyer-Olkin* (KMO). Bentuk hipotesis yang digunakan yaitu:

$H_0$  : Sampel tidak mewakili populasi

$H_1$  : Sampel telah mewakili populasi

Kriteria keputusan  $H_0$  ditolak jika  $KMO > 0,5$ . Berdasarkan hasil uji KMO, diperoleh nilai KMO sebesar 0,548 yang berarti bahwa sampel telah mewakili populasi.

Asumsi kedua, uji multikolinearitas untuk

## Analisis Kluster Dalam Pengelompokan Kabupaten/Kota Di Provinsi....

mengetahui ada tidaknya hubungan atau korelasi yang signifikan antar variabel bebas. Adapun bentuk hipotesis yang digunakan yaitu:

Hipotesis:

$H_0$  : Terdapat korelasi antar variabel bebas

$H_1$  : Tidak terdapat korelasi antar variabel bebas

Adapun kriteria keputusan  $H_0$  ditolak jika korelasi antar variabel lebih kecil dari 0,8

Berdasarkan hasil uji korelasi antar variabel bebas terdapat korelasi yang bernilai lebih dari 0,8 yang berarti bahwa terdapat korelasi antar variabel bebas yaitu ditunjukkan pada Tabel 2 sebagai berikut:

**Tabel 2.** Korelasi Antar Variabel bebas

Variabel	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
$X_1$	1	-0,183	0,282	0,039	0,097
$X_2$	-0,183	1	-0,152	-0,204	-0,138
$X_3$	0,282	-0,152	1	0,442	0,456
$X_4$	0,039	-0,204	0,442	1	0,998
$X_5$	0,097	-0,138	0,456	0,998	1

Berdasarkan Tabel 2 terdapat korelasi antar variabel bebas yaitu antar variabel  $X_4$  dan  $X_5$ . Cara untuk mengatasi masalah multikolinearitas tersebut dengan menghilangkan salah satu variabel yaitu variabel  $X_4$ . Hal pendeteksian dan mengatasi multikolinieritas juga dilakukan oleh (Ridwan & Sunendiari, 2021) dalam penelitiannya untuk mengatasi masalah multikolinieritas pada data diabetes mellitus wanita suku Indian.

Setelah menghilangkan salah satu variabel, dilakukan uji KMO kembali untuk mengetahui bahwa sampel yang digunakan harus mewakili populasi yang ada.

Hipotesis:

$H_0$  : Sampel tidak mewakili populasi

$H_1$  : Sampel telah mewakili populasi

Kriteria keputusan  $H_0$  ditolak jika KMO > 0,5.

Berdasarkan hasil pengujian diperoleh nilai KMO sebesar 0,557 sehingga  $H_0$  ditolak yang berarti bahwa sampel telah mewakili populasi. Setelah menghilangkan variabel  $X_4$  dilakukan uji multikoleniaritas kembali.

Hipotesis:

$H_0$  : Terdapat korelasi antar variable bebas

$H_1$  : Tidak terdapat korelasi antar variabel bebas

Kriteria keputusan  $H_0$  ditolak jika korelasi antar variabel lebih kecil dari 0,8.

**Tabel 3.** Korelasi Antar Variabel Bebas Setelah Menghilangkan Variabel  $X_4$

Variabel	$X_1$	$X_2$	$X_3$	$X_5$
$X_1$	1	-0,183	0,282	0,097
$X_2$	-0,183	1	-0,152	-0,138
$X_3$	0,282	-0,152	1	0,456
$X_5$	0,097	-0,138	0,456	1

Berdasarkan Tabel 3 di atas, nilai korelasi setiap variabel bebas lebih kecil dari 0,8 sehingga  $H_0$  ditolak yang berarti bahwa tidak terdapat korelasi antar variabel bebas atau tidak terjadi multikolinearitas. Sehingga pada penelitian ini dilakukan pengentasan multikolinieritas yang diperoleh variabel yang layak digunakan adalah variabel  $X_1$ ,  $X_2$ ,  $X_3$  dan  $X_5$ .

Setelah kedua asumsi terpenuhi, tahap selanjutnya dapat dilakukan pengelompokkan dengan metode non hierarki yaitu metode *K-Means* Untuk pastinya dapat dilakukan pengecekan asumsi kembali dengan variabel tersebut sehingga pada penelitian ini variabel tersebut dapat dilanjutkan untuk tahap selanjutnya.

Pada analisis kluster dengan metode *K-Means* ini penulis menentukan jumlah kluster yang akan dibentuk sebanyak 3 kluster. Dari jumlah kluster yang ditentukan, terdapat 2 kali proses iterasi klustering yang dilakukan. Proses ini dilakukan agar mendapatkan kluster yang tepat dalam mengelompokkan 11 kabupaten di Provinsi Jambi berdasarkan 4 variabel penyakit menular yaitu malaria, kusta, pneumonia, dan HIV/AIDS. Berikut merupakan karakteristik pengelompokan

kabupaten di Provinsi Jambi:

Tabel 4. Kluster Center Akhir

Variabel	Kluster		
	1	2	3
$X_1$	239	1457,5	3134
$X_2$	5,33	1	3,67
$X_3$	61,17	195,5	134
$X_5$	8	362,5	9

Tabel 4 menyajikan data variabel yang digunakan dalam membedakan isi kluster dari pengelompokan kabupaten yang termasuk ke dalam kluster 1, kluster 2, dan kluster 3. Pada variabel  $X_1$  memiliki rata-rata untuk masing-masing kluster yaitu kluster 1 memiliki rata-rata jumlah kasus malaria sebanyak 239,00 jiwa dan kluster 2 memiliki rata-rata jumlah kasus malaria sebanyak 1457,50 jiwa serta kluster 3 memiliki rata-rata jumlah kasus malaria sebanyak 3134,00 jiwa. Hal ini berarti bahwa kluster 3 merupakan kluster yang memiliki nilai rata-rata variabel  $X_1$  yang tertinggi dan kluster 1 merupakan kluster dengan nilai rata-rata variabel  $X_1$  terendah.

Pada variabel  $X_2$  memiliki rata-rata untuk masing-masing kluster yaitu kluster 1 memiliki rata-rata jumlah kasus kusta sebanyak 5,33 jiwa dan kluster 2 memiliki rata-rata jumlah kasus kusta sebanyak 1 jiwa serta kluster 3 memiliki rata-rata jumlah kasus kusta sebanyak 3,67 jiwa. Hal ini berarti bahwa kluster 1 merupakan kluster yang memiliki nilai rata-rata variabel  $X_2$  yang tertinggi dan kluster 2 merupakan kluster dengan nilai rata-rata variabel  $X_2$  terendah.

Pada variabel  $X_3$  memiliki rata-rata untuk masing-masing kluster yaitu kluster 1 memiliki rata-rata jumlah kasus pneumonia sebanyak 61,17 jiwa dan kluster 2 memiliki rata-rata jumlah kasus pneumonia sebanyak 195,50 jiwa serta kluster 3 memiliki rata-rata jumlah kasus pneumonia sebanyak 134,00 jiwa. Hal ini berarti bahwa kluster 2 merupakan kluster

yang memiliki nilai rata-rata variabel  $X_3$  yang tertinggi dan kluster 1 merupakan kluster dengan nilai rata-rata variabel  $X_3$  terendah.

Pada variabel  $X_5$  memiliki rata-rata untuk masing-masing kluster yaitu kluster 1 memiliki rata-rata jumlah kasus HIV/AIDS sebanyak 8,00 jiwa dan kluster 2 memiliki rata-rata jumlah kasus HIV/AIDS sebanyak 362,50 jiwa serta kluster 3 memiliki rata-rata jumlah kasus HIV/AIDS sebanyak 9,00 jiwa. Hal ini berarti bahwa kluster 2 merupakan kluster yang memiliki nilai rata-rata variabel  $X_5$  yang tertinggi dan kluster 1 merupakan kluster dengan nilai rata-rata variabel  $X_5$  terendah.

Berdasarkan Tabel 4 hasil analisis karakteristik dari masing-masing variabel pada setiap kluster yang terbentuk, dapat ditentukan bahwa kluster 2 memiliki 2 variabel dengan rata-rata penyakit menular tertinggi dan 1 variabel dengan rata-rata penyakit menular sedang, sehingga dapat dinyatakan bahwa kluster 2 merupakan kluster yang terdiri dari kabupaten dengan tingkat variabel penyakit menular tertinggi. Pada kluster 3 memiliki 1 variabel dengan rata-rata penyakit menular tertinggi dan 3 variabel dengan rata-rata variabel penyakit menular sedang, sehingga dapat dinyatakan bahwa kluster 3 merupakan kluster yang terdiri dari kabupaten dengan tingkat variabel penyakit menular menengah atau sedang. Pada kluster 1 memiliki 3 variabel dengan rata-rata penyakit menular terendah dan 1 variabel dengan rata-rata penyakit menular tertinggi, sehingga dapat dinyatakan bahwa kluster 1 merupakan kluster yang terdiri dari kabupaten dengan tingkat variabel penyakit menular terendah.

Setelah kluster dibentuk berdasarkan data kabupaten yang ada, tahapan selanjutnya yaitu melihat perbedaan variabel pada kluster yang terbentuk. Dalam hal ini dapat dilihat dari hasil uji anova.

Hipotesis:

$H_0$  : Minimal ada satu variabel yang berbeda pada kluster yang terbentuk

## Analisis Kluster Dalam Pengelompokan Kabupaten/Kota Di Provinsi....

$H_1$  : Tidak terdapat perbedaan antar variabel pada kluster yang terbentuk dengan taraf signifikansi sebesar 0,05 dan kriteria keputusan semakin besar nilai  $F$  serta signifikansi hasil  $< 0,05$  maka  $H_0$  diterima.

**Tabel 5.** Hasil Uji Anova

Variabel	Klaster		Error		F	Sig,
	Rata-rata kuadrat	Df	Rata-rata kuadrat	df		
$X_1$	8433603,205	2	106931,063	8	78,870	0,000
$X_2$	14,455	2	32,250	8	,448	0,654
$X_3$	15214,697	2	7061,167	8	2,155	0,178
$X_5$	102628,841	2	29880,062	8	3,435	0,084

Berdasarkan Tabel 5, angka  $F$  terbesar ada pada variabel  $X_1$  yaitu 78,870 dengan Sig 0,000 yang artinya  $H_0$  diterima maka ini berarti bahwa variabel  $X_1$  sangat membedakan karakteristik ketiga kluster tersebut.

Jumlah anggota masing-masing kluster yang terbentuk dapat dilihat pada tabel berikut ini:

**Tabel 6.** Jumlah Anggota Setiap Kluster

<b>Klaster</b>	<b>1</b>	<b>6,000</b>
	<b>2</b>	<b>2,000</b>
	<b>3</b>	<b>3,000</b>

Berdasarkan Tabel 6 terlihat bahwa kluster 1 beranggotakan 6 kabupaten di Provinsi Jambi yang memiliki tingkat penyakit menular yang rendah, kluster 2 beranggotakan 2 kabupaten di Provinsi Jambi yang memiliki tingkat penyakit menular yang tinggi, dan pada kluster 3 beranggotakan 3 kabupaten di Provinsi Jambi yang memiliki tingkat penyakit menular yang sedang. Kabupaten mana saja yang masuk dalam kluster 1, kluster 2, dan kluster 3, dapat dilihat pada tabel berikut ini:

**Tabel 7.** Nomor Kluster dan Jarak Kluster

Kabupaten	Kluster	Jarak
Kerinci	1	120,448
Merangin	3	173,297
Sarolangun	3	421,095
Batang Hari	1	411,117
Muaro Jambi	1	285,005
Tanjung Jabung Timur	1	101,763
Tanjung Jabung Barat	1	149,657
Tebo	3	590,437
Bungo	2	384,401
Kota Jambi	2	384,401
Kota Sungai Penuh	1	47,274

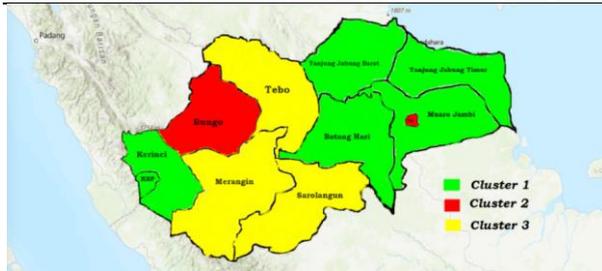
Berdasarkan Tabel 7, maka dapat diperoleh kesimpulan akhir untuk anggota-anggota dari setiap kluster yang terbentuk dari kabupaten yang ada di Provinsi Jambi berdasarkan penyakit menular. Adapun untuk keanggotaan tersebut dapat dilihat pada Tabel 8.

**Tabel 8.** Keanggotaan Kluster

<b>Kluster</b>	<b>1</b>	Kerinci, Batang Hari, Muaro Jambi, Tanjung Jabung Timur, Tanjung Jabung Barat dan Kota Sungai Penuh
	<b>2</b>	Bungo dan Kota Jambi
	<b>3</b>	Merangin, Sarolangun, dan Tebo

Berdasarkan Tabel 8 kluster 1 terdiri dari Kerinci, Batang Hari, Muaro Jambi, Tanjung Jabung Timur, Tanjung Jabung Barat dan Kota Sungai Penuh. Kluster 2 terdiri dari Bungo dan Kota Jambi serta kluster 3 terdiri dari Merangin, Sarolangun, dan Tebo. Peta persebaran kluster-kluster di Kabupaten seluruh Provinsi Jambi berdasarkan variabel penyakit menular dapat disajikan beriku ini:

## Analisis Kluster Dalam Pengelompokan Kabupaten/Kota Di Provinsi....



**Gambar 2.** Persebaran Kluster Berdasarkan Peta Provinsi Jambi

Berdasarkan peta penyebaran penyakit menular di Provinsi Jambi berdasarkan kabupaten/kota diperoleh bahwa kluster 1 terdiri dari 6 kabupaten, yaitu Kerinci, Batang Hari, Muaro Jambi, Tanjung Jabung Timur, Tanjung Jabung Barat dan Kota Sungai Penuh. Kluster ini memiliki nilai variabel penyakit menular yang rendah. Kluster 2 terdiri dari 2 kabupaten, yaitu Bungo dan Kota Jambi. Kluster ini memiliki nilai variabel penyakit menular yang tinggi. Sedangkan kluster 3 terdiri dari 3 kabupaten, yaitu Merangin, Sarolangun, dan Tebo. Kluster ini memiliki nilai variabel penyakit menular yang sedang.

### PENUTUP

Berdasarkan hasil yang telah diperoleh, maka dapat diperoleh kesimpulan adalah sebagai berikut:

1. Kluster 1 terdiri dari 6 kabupaten, yaitu Kerinci, Batang Hari, Muaro Jambi, Tanjung Jabung Timur, Tanjung Jabung Barat dan Kota Sungai Penuh. Kluster 1 merupakan pengelompokan dari kabupaten di Provinsi Jambi yang mempunyai nilai variabel penyakit menular yang rendah. Hal tersebut berarti, kabupaten

yang tergolong ke dalam kluster 1 memiliki tingkat kesehatan yang baik dan tingkat penyakit menular yang rendah.

2. Kluster 2 terdiri dari 2 kabupaten, yaitu Bungo dan Kota Jambi. Kluster 2 merupakan pengelompokan dari kabupaten di Provinsi Jambi yang mempunyai nilai variabel penyakit menular yang tinggi. Hal tersebut berarti, kabupaten yang tergolong ke dalam kluster 2 memiliki tingkat Kesehatan yang kurang baik serta tingkat penyakit menular yang tinggi.
3. Kluster 3 terdiri dari 3 kabupaten, yaitu Merangin, Sarolangun, dan Tebo. Kluster 3 merupakan pengelompokan dari kabupaten di Provinsi Jambi yang mempunyai nilai variabel penyakit menular yang sedang. Hal tersebut berarti, kabupaten yang tergolong ke dalam kluster 3 memiliki tingkat Kesehatan yang cukup baik serta tingkat penyakit menular yang sedang.

Saran agar hasil dari analisis yang peneliti lakukan diharapkan dapat digunakan untuk para pelaku kebijakan sebagaimana mestinya.

### DAFTAR PUSTAKA

- Abbas, S. A., Aslam, A., Rehman, A. U., Abbasi, W. A., Arif, S., & Kazmi, S. Z. H. (2020). K-Means and K-Medoids: Kluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir. *IEEE Access*, *8*, 151847–151855. <https://doi.org/10.1109/ACCESS.2020.3014021>
- Arbin, N., Suhaimi, N. S., Mokhtar, N. Z., & Othman, Z. (2016). Comparative analysis between K-Means and k-medoids for

- statistical klustering. *Proceedings - AIMS 2015, 3rd International Conference on Artificial Intelligence, Modelling and Simulation*, 117–121. <https://doi.org/10.1109/AIMS.2015.82>
- Asroni, & Adrian, R. (2015). Penerapan Metode K-Means Untuk Klustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang (Implementation Method for K-Means Klustering Based Student Value with Weka Interface a Case Study of Department of Information UMM Magelang). *Jurnal Ilmiah Semesta Teknika*, 18(1), 76–82. <https://doi.org/10.18196/st.v18i1.708>
- BPS. (2022a). *Kota Jambi Dalam Angka Jambi Municipality In Figures 2022*. Badan Pusat Statistik Kota Jambi.
- BPS. (2022b). *Provinsi Jambi Dalam Angka 2022*. Badan Pusat Statistik Provinsi Jambi.
- Carvalho, M. J., Melo-Gonçalves, P., Teixeira, J. C., & Rocha, A. (2016). Regionalization of Europe based on a K-Means Kluster Analysis of the climate change of temperatures and precipitation. *Physics and Chemistry of the Earth*, 94, 22–28. <https://doi.org/10.1016/j.pce.2016.05.001>
- Faqih, A. (2022). Implementasi Metode K-modes Untuk Pengelompokan Angkutan Sungai Dan Penyeberangan Di Kabupaten Bojonegoro. *Jurnal Statistika Dan Komputasi*, 1(2), 54–63. <https://doi.org/10.32665/statkom.v1i2.1135>
- Muzakir, A. (2014). Analisa dan Pemanfaatan Algoritma K-Means Klustering pada Data Nilai Siswa sebagai Penentuan Penerima Beasiswa. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014*, 195–200.
- Nurliza, N. N. (2018). *Penerapan Euclidean Distance Pada Pengenalan Pola Citra Sidik Jari. Skripsi*. Universitas Islam Indonesia.
- Prianto, C., & Bunyamin, S. (2020). *Pembuatan Aplikasi Klustering Gangguan Jaringan Menggunakan Metode K-Means Klustering*. Kreatif Industri Nusantara.
- Ridwan, M., & Sunendiari, S. (2021). Mendeteksi dan Mengatasi Multikolinieritas pada Data Penelitian Diabetes Melitus Wanita Suku Indian Tahun 2018. *Prosiding Statistika*, 64–70. <https://doi.org/10.29313/.v7i1.25565>
- Sobari, N., & Usman, H. (2013). *Aplikasi Teknik Multivariate*. Rajawali Pers.
- Talakua, M. W., Leleury, Z. A., & Talluta, A. W. (2017). Analisis Kluster Dengan Menggunakan Metode K-Means untuk Pengelompokan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *Barekeng : Jurnal Ilmu Matematika Dan Terapan*, 11(2), 119–128. <https://doi.org/10.30598/barekengvol11iss2pp119-128>
- Triyanto, W. A. (2015). Algoritma K-Medoids untuk Penentuan Strategi Pemasaran Produk. *Jurnal SIMETRIS*, 6(1), 183–188. <https://doi.org/10.24176/simet.v6i1.254>
- Yu, D., Liu, G., Guo, M., & Liu, X. (2018). An improved K-medoids algorithm based on step increasing and optimizing medoids. *Expert Systems with Applications*, 92, 464–473. <https://doi.org/10.1016/j.eswa.2017.09.052>

Yulianto, S., & Hidayatul, K. H. (2014).  
Analisis Klaster untuk  
Pengelompokan Kabupaten/Kota di  
Provinsi Jawa Tengah Berdasarkan  
Indikator Kesejahteraan Rakyat.  
*Statistika*, 2(1), 56–63.  
[https://doi.org/10.26714/jsunimus.  
2.1.2014.%25p](https://doi.org/10.26714/jsunimus.2.1.2014.%25p)

