

# I. PENDAHULUAN

## 1.1 Latar Belakang

Penyakit menular adalah penyakit yang dapat ditularkan secara langsung atau melalui perantara dari orang yang sakit ke orang yang sehat atau orang yang tidak tertular penyakit menular. Peningkatan penyakit menular mengindikasikan bahwa kurangnya kesadaran akan kesehatan pada suatu wilayah. Oleh karena itu harus dilakukan upaya penurunan kasus dari penyakit menular, dimana upaya yang dilakukan mempunyai peranan penting dalam mengukur derajat kesehatan masyarakat (Kemenkes, 2022). Pada peraturan Gubernur Jambi No. 36 tahun 2016 pasal 27 menyebutkan bahwa terdapat beberapa penyakit menular yang menjadi program pengendalian penyakit meliputi diare, malaria, tuberkulosis, pneumonia, HIV/AIDS, DBD, kusta, IMS, dan campak.

Berdasarkan data yang dikeluarkan oleh Badan Pusat Statistik Provinsi Jambi untuk kasus penyakit menular pada tahun 2019 hingga tahun 2020, terdapat kenaikan dan penurunan jumlah kasus penyakit menular meliputi penyakit menular diare mengalami penurunan dari 70.882 jiwa menjadi 46.379 jiwa, penyakit menular malaria mengalami kenaikan dari 26.806 jiwa menjadi 30.088 jiwa, penyakit menular tuberkulosis mengalami penurunan dari 5.444 jiwa menjadi 4.838 jiwa, penyakit menular pneumonia mengalami penurunan dari 3.329 jiwa menjadi 1.812 jiwa, penyakit menular HIV/AIDS mengalami penurunan dari 1.238 jiwa menjadi 765 jiwa, penyakit menular DBD mengalami penurunan dari 2.229 jiwa menjadi 2.049 jiwa, penyakit menular kusta mengalami penurunan dari 93 jiwa menjadi 61 jiwa, penyakit menular IMS pada tahun 2019 tidak memiliki kasus sama sekali hingga pada tahun 2020 terdapat kasus penyakit menular IMS sebanyak 275 jiwa, dan penyakit menular campak mengalami penurunan dari 173 jiwa menjadi 62 jiwa. Selanjutnya, berdasarkan data yang dikeluarkan oleh Badan Pusat Statistik Indonesia tahun 2020 Provinsi Jambi menempati urutan ketujuh dengan jumlah 46.379 jiwa untuk penyakit menular diare terbanyak di Pulau Sumatera dan urutan pertama dengan jumlah 30.088 jiwa untuk penyakit menular malaria terbanyak di Pulau Sumatera. Kemudian untuk penyakit menular tuberkulosis, pneumonia, HIV/AIDS dan IMS Provinsi Jambi menempati urutan ketujuh untuk penyakit menular tuberkulosis berjumlah 4.838 jiwa, penyakit menular pneumonia berjumlah 1.812 jiwa, penyakit menular HIV/AIDS berjumlah 765 jiwa dan penyakit menular IMS berjumlah 275 jiwa. Provinsi Jambi menempati urutan kelima dengan jumlah 2.049 jiwa untuk penyakit menular DBD terbanyak di Pulau Sumatera serta Provinsi Jambi menempati urutan keenam untuk penyakit menular kusta

berjumlah 61 jiwa dan campak berjumlah 62 jiwa. Berdasarkan data yang dikeluarkan oleh Badan Pusat Statistik Provinsi Jambi tahun 2020, angka kejadian penyakit menular diare terbanyak berada di wilayah Kabupaten Merangin dengan jumlah 6.653 jiwa. Kemudian Kota Jambi mempunyai angka kejadian penyakit menular malaria, tuberkulosis, HIV/AIDS, DBD, IMS, dan campak terbanyak dengan penyakit menular malaria berjumlah 6.619 jiwa, penyakit menular tuberkulosis berjumlah 1.021 jiwa, penyakit menular HIV/AIDS berjumlah 683 jiwa, penyakit menular DBD berjumlah 717 jiwa, penyakit menular IMS berjumlah 254 jiwa, dan penyakit menular campak berjumlah 40 jiwa. Kemudian angka kejadian penyakit menular pneumonia terbanyak berada di wilayah Kabupaten Muaro Jambi dengan jumlah 536 jiwa serta angka kejadian penyakit menular kusta terbanyak berada di wilayah Kabupaten Tanjung Jabung Timur dengan jumlah 30 jiwa. Berdasarkan angka kejadian penyakit menular, dibutuhkan pengelompokan sebagai upaya yang dapat dilakukan dalam menentukan prioritas wilayah kabupaten/kota di Provinsi Jambi dalam mengatasi masalah penyakit menular. Angka kejadian penyakit menular tersebut digunakan sebagai ukuran karakteristik dalam suatu pengelompokan, yang mana pengelompokan ini dalam statistika dikenal sebagai analisis *cluster*.

Analisis *cluster* merupakan suatu teknik dalam analisis multivariat yang bertujuan untuk mengelompokkan objek atau individu berdasarkan pada karakteristik tertentu dengan data kuantitatif (Widayat, 2018). Analisis *cluster* terbagi ke dalam beberapa metode diantaranya adalah metode hierarki dan non-hierarki. Metode non-hierarki merupakan metode yang mempunyai proses penyelesaian yang lebih cepat, mudah diimplementasikan serta menghasilkan *cluster* yang lebih padat bila dibandingkan dengan metode hierarki (Lasheng dan Yuqiang, 2017). Menurut Dean (2014), terdapat beberapa teknik dalam analisis *cluster* non-hierarki yaitu *k-means* dan *k-medoid*. *K-Means* merupakan analisis *cluster* yang menggunakan *mean* sebagai pusat *cluster* dan *k-medoid* merupakan analisis *cluster* yang menggunakan *medoid* sebagai pusat *cluster*.

Berdasarkan penelitian yang dilakukan oleh Irwan, Sanusi dan Saman (2022) yang berjudul Pengelompokan Daerah Rawan Kriminalitas di Sulawesi Selatan Menggunakan Metode *K-Means Clustering*, pada penelitian tersebut diperoleh 4 *cluster* dengan urutan tingkat kejahatan cukup rendah, tinggi, rendah, dan cukup tinggi. Selanjutnya berdasarkan penelitian yang dilakukan oleh Agustian dan Dermawan (2022) yang berjudul Analisis *Clustering* Demam Berdarah Dengue Dengan Algoritma *K-Medoid* (Studi Kasus Kabupaten Karawang), diperoleh *cluster* pertama terdiri dari 6 kecamatan dengan tingkatan

DBD tertinggi dan *cluster* kedua terdiri dari 22 kecamatan dengan tingkatan DBD terendah.

Penelitian yang dilakukan oleh Harahap (2021) yang berjudul Perbandingan Algoritma *K-Means* dan *K-Medoid* Untuk *Clustering* Kelas Siswa Tunagrahita, pada penelitian tersebut diperoleh hasil bahwa algoritma *k-means* lebih baik dari pada algoritma *k-medoid* berdasarkan evaluasi menggunakan *Davies Bouldin Index*. Penelitian lainnya dilakukan oleh Siagian, Sirait dan Halim (2022) yang berjudul Penerapan Algoritma *K-Means* dan *K-Medoid* untuk Segmentasi Pelanggan pada Data Transaksi *E-Commerce*, berdasarkan penelitian tersebut dengan menggunakan evaluasi *Davies Bouldin Index* diperoleh bahwa algoritma *k-medoid* lebih baik dari pada algoritma *k-means*.

Pada penelitian ini, pengelompokan tersebut dapat dilakukan dengan mengelompokkan kabupaten/kota di Provinsi Jambi berdasarkan data penyakit menular yaitu diare, malaria, tuberkulosis, pneumonia, HIV/AIDS, DBD, kusta, IMS, dan campak sebagai ukuran karakteristiknya dan data yang digunakan telah memenuhi syarat dilakukannya analisis *cluster* yaitu data kuantitatif berskala rasio. Berdasarkan permasalahan serta beberapa penelitian terdahulu tersebut, penulis tertarik untuk menguji dan membandingkan kedua algoritma dari analisis *cluster* yaitu *K-Means* dan *K-Medoid* dalam mengelompokkan kabupaten/kota di Provinsi Jambi berdasarkan data penyakit menular diare, malaria, tuberkulosis, pneumonia, HIV/AIDS, DBD, kusta, IMS, dan campak. Adapun judul dari penelitian ini adalah **“Perbandingan Algoritma *K-Means* dan *K-Medoid* dalam Pengelompokan Kabupaten/Kota di Provinsi Jambi Berdasarkan Data Penyakit Menular dengan *Calinski Harabasz Index*”**.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah yang akan dikaji dalam penelitian ini adalah sebagai berikut:

1. Bagaimana pengelompokan setiap kabupaten/kota di Provinsi Jambi berdasarkan data penyakit menular menggunakan metode *k-means* dan *k-medoid*?
2. Manakah algoritma terbaik dari algoritma *k-means* dan algoritma *k-medoid* dalam pengelompokan kabupaten/kota di Provinsi Jambi berdasarkan data penyakit menular dengan *calinski harabasz index*?

### 1.3 Tujuan

Berdasarkan rumusan masalah yang akan dikaji, maka tujuan yang akan dicapai dalam penelitian ini adalah sebagai berikut:

1. Untuk mengetahui hasil pengelompokan setiap kabupaten/kota di Provinsi Jambi berdasarkan data penyakit menular menggunakan metode *k-means* dan *k-medoid*.
2. Untuk mengetahui algoritma mana yang terbaik antara algoritma *k-means* dan algoritma *k-medoid* dalam pengelompokan kabupaten/kota di Provinsi Jambi berdasarkan data penyakit menular dengan *calinski harabasz index*.

### 1.4 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Bagi penulis adalah untuk menerapkan teori matematika dalam bidang statistika khususnya pengelompokan suatu objek.
2. Bagi pembaca adalah untuk menambah wawasan dan sebagai bahan referensi untuk penelitian selanjutnya.
3. Bagi pemerintah adalah sebagai bahan informasi mengenai pengelompokan kabupaten/kota di Provinsi Jambi berdasarkan data penyakit menular dalam mengambil kebijakan-kebijakan untuk mengatasi penyakit menular.

### 1.5 Batasan Masalah

Pada penelitian ini, penulis membatasi materi dan pembahasan agar tidak meluas sehingga digunakan batasan masalah sebagai berikut:

1. Menggunakan data penyakit menular yang terjadi pada setiap kabupaten/kota di Provinsi Jambi
2. Terdapat 9 data penyakit menular yang terdiri atas diare, malaria, tuberkulosis, pneumonia, HIV/AIDS, DBD, kusta, IMS, dan campak.
3. Menggunakan jarak *Euclidean* sebagai ukuran jarak dalam analisis yang dilakukan.

## II. TINJAUAN PUSTAKA

### 2.1 Penyakit Menular

Penyakit menular ialah suatu penyakit yang diakibatkan oleh agen biologi seperti bakteri, virus, dan parasit serta bukan diakibatkan oleh faktor fisik lainnya. Penyakit menular dapat menyebar melalui perantara baik secara langsung maupun tak langsung, baik dari orang yang terjangkit penyakit menular kepada orang yang belum terjangkit atau dari orang yang telah selesai mengalami penyakit (*carrier*) kepada orang yang belum terjangkit. Apabila suatu tempat atau daerah mengalami penyakit atau agen infeksi tertentu secara terus menerus maka dapat dikatakan bahwa terjadi endemis. *Hyperendemis* merupakan suatu kondisi dimana penyakit tertentu selalu ditemukan atau muncul di suatu wilayah dengan kejadian kasus yang tinggi. Sedangkan, *holoendemis* merupakan suatu keadaan dimana penyakit tersebut selalu ditemukan di suatu wilayah dengan kejadian kasus yang tinggi dan menimpa sebagian besar penduduk (Chin, 2000). Adapun penyakit menular yang menjadi variabel penelitian ini merupakan penyakit menular yang memiliki angka kasus atau kejadian yang banyak terjadi di antaranya, yaitu:

#### 1. Diare

Diare adalah penyakit menular yang diakibatkan oleh *agent* penyebab berupa bakteri, virus, jamur, serta akibat konsumsi dari makanan dan minuman yang mengandung *agent* penyebab diare tersebut, serta juga dari akibat penurunan imun tubuh. Penderita akan mengalami buang air besar yang sering dengan kandungan air yang berlebihan (Irwan, 2017). Menurut Kemenkes RI (2022) diare ialah penyakit endemis dimana penyakit ini memiliki potensi untuk menimbulkan kejadian yang luar biasa serta masih menjadi penyumbang angka kematian di Indonesia terutama pada balita.

#### 2. Malaria

Malaria ialah penyakit menular yang diakibatkan suatu parasit (*protozoa*) dari genus *plasmodium*, penyakit ini dapat ditularkan melalui kontak langsung dengan nyamuk *Anopheles* (Irwan, 2017). Penyakit malaria bisa menyebabkan kematian apabila orang yang terjangkit mendapatkan pengobatan yang terlambat. Pertolongan dari tim medis harus segera didapatkan jika yang seseorang telah dicurigai menderita penyakit ini (Chin, 2000). Menurut Kemenkes RI (2022) malaria merupakan penyakit menular yang menjadi bagian dari tujuan global terkait dengan pengendalian malaria di setiap wilayah yang ada di Indonesia agar dapat terbebas dari malaria.

### 3. Tuberkulosis Paru

Tuberkulosis (TB) merupakan kategori penyakit yang bisa menular dan diakibatkan oleh bakteri *Mycobacterium tuberculosis*. Penyakit yang menyerang paru, akan tetapi bisa juga menyerang ke semua organ atau jaringan tubuh. Penularan penyakit ini melalui udara disaat seorang yang menderita tuberkulosis batuk sehingga percikan air liur yang mengandung bakteri penyebab terhirup oleh individu lain ketika bernapas (Irwan, 2017). Menurut Kemenkes RI (2022) tuberkulosis saat ini merupakan masalah kesehatan masyarakat dengan penyebab kematian tertinggi.

### 4. Pneumonia

Pneumonia adalah penyakit menular yang berasal dari suatu infeksi bakteri yang akut dengan tanda serangan yaitu mendadak demam menggigil dan batuk yang dialami disertai dahak berwarna merah-kemerahan. Pada bayi dan anak kecil akan mengalami demam, muntah serta kejang yang termasuk indikasi awal dari penyakit ini. Pneumonia termasuk penyakit yang menetap berkelanjutan khususnya dalam menyerang bayi dan usia lanjut (Irwan, 2017). Pneumonia ialah suatu infeksi akut yang juga menyerang jaringan paru-paru (*alveoli*) yang diakibatkan oleh berbagai mikroorganisme lainnya (Kemenkes, 2022).

### 5. HIV/AIDS

HIV (*Human Immunodeficiency Virus*) ialah suatu virus yang akan menyerang sistem kekebalan tubuh dimana infeksi yang terjadi akan menyebabkan penderitanya mengalami drop kekebalan tubuh sehingga akan sangat mudah penyakit lain untuk masuk ke dalam tubuh . Sedangkan, AIDS atau *Acquired Immuno Deficiency Syndrome* merupakan sekumpulan gejala dimana akan terjadi penurunan kemampuan pertahanan diri diakibatkan dengan masuknya virus HIV. Adapun program pengendalian HIV di Indonesia bertujuan dalam menurunkan hingga menghilangkan infeksi baru, menurunkan dan menghilangkan angka kematian terkait AIDS, serta menurunkan angka stigma dan diskriminasi yang disebabkan oleh HIV/AIDS (Kemenkes, 2022).

### 6. DBD

DBD (Demam Berdarah Dengue) merupakan penyakit menular yang diakibatkan oleh virus dengue dan disebarkan oleh perantara nyamuk dari jenis nyamuk *Aedes aegypti* dan *Aedes albopictus*. Karakteristik dari perantara yaitu menyukai genangan air seperti vas bunga, tempat sampah dan selokan sebagai tempat peridukan (Kemenkes, 2022).

## 7. Kusta

Kusta ialah suatu penyakit yang menular ditimbulkan dari bakteri *Mycobakterium leprae*. Penyakit menular kusta bersifat kronis, menyerang kulit, syaraf tepi dan organ tubuh lain kecuali syaraf pusat. penularan dari penyakit menular ini melalui udara disekitar yang tercemar percikan air dari hidung hingga tenggorokan orang yang mengalami infeksi pernafasan atas atau juga melalui kulit yang terluka (Irwan, 2017).

## 8. IMS

IMS (Infeksi Menular Seksual) adalah penyakit yang menular terdiri dari berbagai infeksi yang diakibatkan menularnya dari orang satu ke orang yang lain melalui kontak langsung, penanganan penyakit ini masih menjadi tantangan karena merupakan suatu penyakit yang sangat erat kaitannya dengan perilaku individu yang berisiko. Bila tidak dilakukan upaya pencegahan akan sangat berdampak pada peningkatan angka kesakitan, kecacatan dan bahkan kematian (Kemenkes, 2022).

## 9. Campak

Campak adalah penyakit menular yang diakibatkan oleh virus yang berasal dari *genus Morbillivirus* serta termasuk ke golongan Paramyxovirus. Campak juga dikenal sebagai *morbili* atau *measles*, penularan dari penyakit ini melalui udara yang tercemar air percikan dari hidung, mulut atau tenggorokan orang yang mengalami infeksi dengan gejala yang biasanya muncul 10-20 hari setelah infeksi ditandai dengan demam suhu tinggi, pilek, mata berwarna merah hingga bintik putih kecil pada mulut bagian dalam(Kemenkes, 2022).

Secara umum, pencegahan adalah mengambil tindakan awal sebelum terjadinya kejadian tersebut. Pencegahan penyakit ini ialah suatu upaya yang dilakukan dalam mencegah terjadinya penyakit yang mencakup semua kalangan. Salah satunya pencegahan primordial merupakan usaha yang dilakukan dalam menghindari terjadinya faktor resiko, serta diperlukannya keterlibatan dari instansi-instansi terkait sehingga cepat terlaksana (Irwan, 2017). Contohnya membuat pengelompokan kabupaten/kota berdasarkan penyakit menular agar dapat memberikan informasi mengenai kabupaten/kota yang rawan mengalami penyakit menular sehingga dapat dilakukan upaya pencegahan penyakit menular. Adapun kebijakan yang telah diupayakan oleh pemerintah kabupaten/kota di Provinsi Jambi untuk mengatasi masalah penyakit menular tersebut adalah sebagai berikut:

### 1. Kebijakan yang dilakukan oleh Kabupaten Kerinci

Kebijakan yang dilakukan oleh Kabupaten Kerinci dalam mengatasi masalah penyakit menular adalah dengan meningkatkan kualitas dan kuantitas baik

dari segi sarana dan prasarana kesehatan, meningkatkan pelayanan kesehatan bagi masyarakat, memberikan sosialisasi dan penyuluhan kepada lapisan masyarakat dalam upaya menurunkan angka kasus penyakit menular, dan meningkatkan upaya-upaya pencegahan yang memungkinkan terjadinya penyakit menular (Pemkab Kerinci, 2018).

2. Kebijakan yang dilakukan oleh Kabupaten Merangin

Kebijakan yang dilakukan oleh Kabupaten Merangin dalam mengatasi masalah penyakit menular adalah dengan meningkatkan operasional kesehatan di Kabupaten Merangin diantaranya meningkatkan sumber daya di bidang kesehatan termasuk tenaga kesehatan serta dana kesehatan, persediaan alat dan obat dalam mengatasi masalah penyakit menular serta memanfaatkan teknologi dalam menyelenggarakan upaya penanggulangan penyakit menular (Pemkab Merangin, 2019).

3. Kebijakan yang dilakukan oleh Kabupaten Sarolangun

Kebijakan yang dilakukan oleh Kabupaten Sarolangun dalam mengatasi masalah penyakit menular adalah dengan upaya pemeriksaan, pengobatan, perawatan dan isolasi bagi penderita penyakit menular, melakukan pemusnahan dengan cara yang dilakukan tanpa merusak lingkungan hidup dan tidak mengakibatkan wabah penyakit menular tersebar serta menggunakan teknologi tepat guna untuk mendukung upaya kebijakan pemerintah Kabupaten Sarolangun dalam mengatasi masalah penyakit menular (Pemkab Sarolangun, 2018)

4. Kebijakan yang dilakukan oleh Kabupaten Batanghari

Kebijakan yang dilakukan oleh Kabupaten Batanghari dalam mengatasi masalah penyakit menular adalah dengan meningkatkan sumber daya pada bidang kesehatan baik dari persediaan obat dan tenaga kesehatan, memberikan informasi kepada masyarakat mengenai penularan dan pencegahan masalah penyakit menular serta melakukan pembatasan kegiatan masyarakat apabila terjadi wabah (Pemkab Batanghari, 2017)

5. Kebijakan yang dilakukan oleh Kabupaten Muaro Jambi

Kebijakan yang dilakukan oleh Kabupaten Muaro Jambi dalam mengatasi masalah penyakit menular adalah dengan meningkatkan upaya pelayanan kesehatan yang didukung oleh semua pihak mulai dari kalangan yang lebih kecil seperti keluarga hingga ke kalangan yang lebih tinggi dalam suatu kabupaten untuk menentukan kebijakan pada bidang kesehatan, memberikan pelayanan kesehatan dasar, akses dan meningkatkan mutu pelayanan kesehatan serta memberikan sosialisasi mengenai perilaku hidup sehat di masyarakat (Pemkab Muaro Jambi, 2015)



6. Kebijakan yang dilakukan oleh Kabupaten Tanjung Jabung Timur  
Kebijakan yang dilakukan oleh Kabupaten Tanjung Jabung Timur dalam mengatasi masalah penyakit menular adalah dengan meningkatkan kualitas dan kuantitas baik dari segi sarana dan prasarana kesehatan, meningkatkan pelayanan kesehatan bagi masyarakat, memberikan sosialisasi dan penyuluhan kepada lapisan masyarakat dalam upaya menurunkan angka kasus penyakit menular, dan melakukan antisipasif sebagai bentuk pencegahan yang memungkinkan terjadinya penyakit menular (Pemkab Tanjung Jabung Timur, 2018).
7. Kebijakan yang dilakukan oleh Kabupaten Tanjung Jabung Barat  
Kebijakan yang dilakukan oleh Kabupaten Tanjung Jabung Barat dalam mengatasi masalah penyakit menular adalah meningkatkan fasilitas dan sumber daya kesehatan, meningkatkan pelayanan kesehatan bagi masyarakat, memberikan sosialisasi dan penyuluhan kepada lapisan masyarakat dalam upaya menurunkan angka kasus penyakit menular, dan melakukan implementasi pembatasan aktifitas masyarakat apabila terjadi wabah guna bentuk pencegahan yang memungkinkan mengurangi terjadinya penyakit menular (Pemkab Tanjung Jabung Barat, 2018).
8. Kebijakan yang dilakukan oleh Kabupaten Tebo  
Kebijakan yang dilakukan oleh Kabupaten Tebo dalam mengatasi masalah penyakit menular adalah dengan meningkatkan sumber daya pada bidang kesehatan baik dari tenaga kesehatan dan obat-obatan, memberikan informasi serta meningkatkan pemberdayaan masyarakat mengenai penularan dan pencegahan masalah penyakit menular serta melakukan pembatasan kegiatan masyarakat apabila terjadi wabah (Pemkab Tebo, 2017).
9. Kebijakan yang dilakukan oleh Kabupaten Bungo  
Kebijakan yang dilakukan oleh Kabupaten Bungo dalam mengatasi masalah penyakit menular adalah menyediakan sarana dan pra-sarana kesehatan, melakukan upaya percepatan dalam penanganan masalah penyakit menular, memberikan sosialisasi mengenai perilaku hidup bersih dan sehat kepada semua lapisan masyarakat, serta memberikan protokol dan penerpan sanksi dalam mengatasi masalah penyakit menular (Pemkab Bungo, 2017).
10. Kebijakan yang dilakukan oleh Kota Jambi  
Kebijakan yang dilakukan oleh Kota Jambi dalam mengatasi masalah penyakit menular adalah penggunaan Nomor Induk Kependudukan (NIK) sebagai identitas tunggal guna memberikan pelayanan publik termasuk pelayanan kesehatan, melakukan peningkatan fasilitas pelayanan kesehatan, melaksanakan upaya untuk memberikan pengetahuan dan

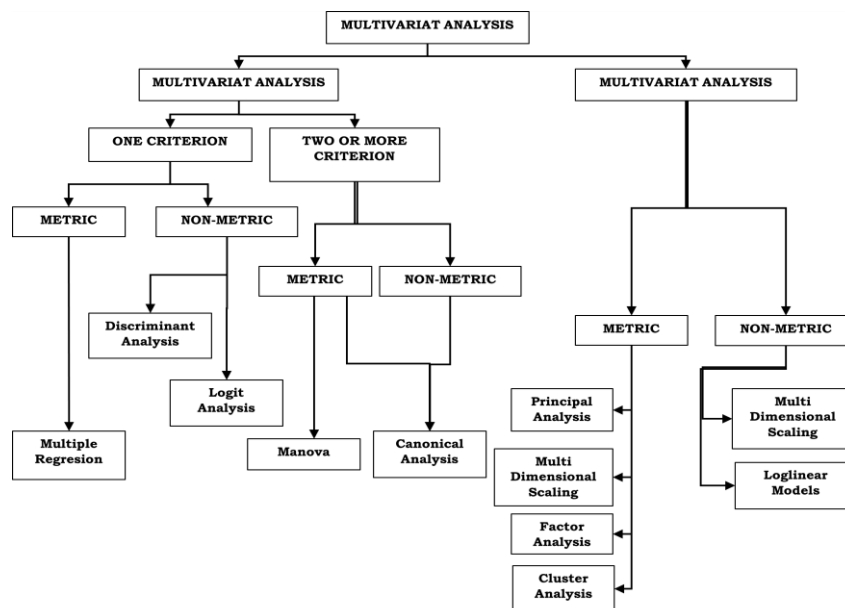
kepedulian, pencegahan, deteksi dini dan penanganan mengenai masalah penyakit menular (Pemkot Jambi, 2018).

#### 11. Kebijakan yang dilakukan oleh Kota Sungai Penuh

Kebijakan yang dilakukan oleh Kota Sungai Penuh dalam mengatasi masalah penyakit menular adalah dengan melakukan pembinaan, pemeriksaan, dan pengawasan terkait masalah penyakit menular, mengupayakan fasilitas sanitasi yang memenuhi persyaratan kesehatan, serta mengikutsertakan asosiasi, organisasi profesional instansi lainnya seperti laboratorium kesehatan dalam melakukan pembinaan dan pengawasan mengenai masalah penyakit menular (Pemkot Sungai Penuh, 2018).

### 2.2 Analisis Multivariat

Analisis multivariat adalah salah satu metode statistik yang dapat digunakan dalam melakukan penelitian terhadap satu atau lebih dari dua variabel secara bersamaan. Keuntungan dalam menggunakan teknik analisis ini yaitu bisa menganalisis hubungan atau perbedaan antar beberapa variabel terhadap variabel lainnya untuk waktu yang bersamaan (Riswan dan Khairudin, 2019). Data atau informasi multivariat tidak hanya terdiri atas satu variabel saja melainkan dapat terdiri atas lebih dari satu variabel. Analisis multivariat artinya analisis statistika yang melibatkan lebih berasal satu variabel bebas. Analisis data multivariat melibatkan identifikasi dan pengukuran variasi dalam satu set variabel, baik diantaranya variabel itu sendiri ataupun antara variabel dependen serta satu atau lebih dari variabel *independent* (Hair et al., 2010). Adapun jenis-jenis analisis dalam analisis multivariat dapat dilihat dalam skema berikut:



Sumber: Hair, et al (2010)

**Gambar 1.** Skema Analisis Multivariat

Menurut Wijaya dan Budiman (2016), Teknik dalam analisis multivariat secara umum terbagi ke dalam dua macam metode, diantaranya yaitu:

1. Metode dependensi, yaitu metode analisis multivariat yang digunakan untuk menerangkan atau memprediksi variabel terikat (*dependent variable*) dengan berdasarkan dua atau lebih variabel bebas (*independent variable*). Pada analisis multivariat dengan metode dependensi, jika jika variabel terikat yang digunakan hanya satu dan menggunakan pengukuran data bersifat metrik, maka teknik analisis yang digunakan adalah analisis regresi berganda (*multiple regression*). Jika terdapat satu variabel terikat dengan pengukuran data bersifat non metrik, maka dapat menggunakan analisis diskriminan ataupun analisis logit sebagai teknik analisisnya. Adapun jika variabel terikat lebih dari satu dan pengukurannya bersifat metrik, maka dapat menggunakan analisis varian multivariat (MANOVA). Jika variabelnya lebih dari satu dan bersifat non metrik, maka teknik analisisnya menggunakan analisis kanonikal.
2. Analisis interdependensi, yaitu metode analisis multivariat yang digunakan dalam memberikan makna terhadap beberapa variabel atau membuat kelompok-kelompok yang didasarkan dari suatu objek atau individu. Metode interdependensi diklasifikasikan dengan berdasarkan jenis variabel yang digunakan dengan skala pengukuran bersifat metrik dan non metrik. Apabila data yang digunakan bersifat metrik maka teknik analisis yang dapat digunakan adalah analisis faktor, analisis komponen utama, analisis *cluster* dan multidimensional scaling. Sedangkan jika data berskala non metrik, maka teknik analisis yang dapat digunakan adalah *multidimensional scaling* dan *loglinear models*.

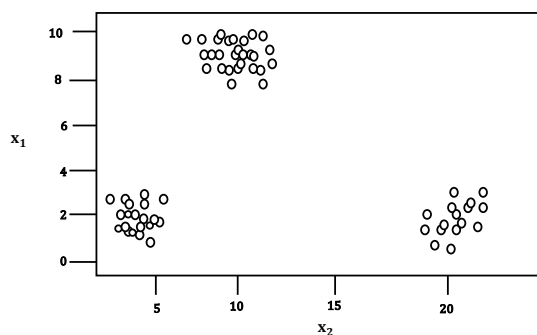
Contoh dari data atau informasi multivariat adalah dimisalkan kasus kejadian penyakit diabetes dari tahun 2012 sampai 2022 yang dipengaruhi oleh usia, berat badan, banyaknya konsumsi makanan cepat saji, dan keturunan. Penyakit diabetes dapat disimbolkan dengan  $y$  lalu usia, berat badan, banyaknya konsumsi makanan cepat saji, dan keturunan merupakan serangkaian variabel yang dapat disimbolkan secara matematis sebagai  $x_1$ ,  $x_2$ ,  $x_3$ , dan  $x_4$ . Hal tersebut menunjukkan bahwa data atau informasi dari penyakit diabetes yang terdiri lebih dari satu variabel.

### **2.3 Analisis Cluster**

Analisis *cluster* ialah salah satu bagian metode dari analisis teknik interpedensi dalam analisis multivariat. Data multivariat yang terdiri atas lebih dari satu variabel dan bersifat metrik (kuantitatif) dapat dilakukan pengelompokan secara statistika dengan menggunakan analisis *cluster*. Analisis

*cluster* merupakan salah satu bagian dari metode dalam analisis multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* ini mengelompokkan individu atau objek penelitian, sehingga setiap objek yang paling dekat kesamaannya dengan objek lain akan berada dalam *cluster* yang sama (Sobari dan Usman, 2013).

Menurut Everitt dan Hothorn (2011), analisis klaster adalah istilah umum untuk berbagai metode numerik yang ditunjukkan untuk menemukan kelompok atau klaster yang homogen dan berbeda dengan kelompok lainnya. Hal ini artinya objek dari setiap klaster yang terbentuk akan mirip satu sama lain dan berbeda atau tidak identik dengan objek dari kelompok lain. Hal ini ditunjukkan pada gambar berikut:



Sumber: Everitt dan Hothorn (2011)

**Gambar 2.** Data yang Menunjukkan Keberadaan dari Ketiga *Cluster*

Menurut Sobari dan Usman (2013), suatu *cluster* yang baik merupakan suatu *cluster* yang terbentuk dan memiliki kondisi diantaranya, yaitu:

- 1) Homogenitas internal (*within cluster*), yaitu memiliki kesamaan antar anggota di dalam satu klaster.
- 2) Heterogenitas eksternal (*between cluster*), yaitu terdapat perbedaan antara satu klaster dengan klaster lainnya.

## 2.4 Metode Analisis Cluster

Analisis klaster mempunyai dua metode yang dapat dilakukan dalam proses klasterisasi. Kedua metode klaster yang digunakan akan menentukan hasil klasterisasi. Analisis klaster tersebut terdiri atas metode hierarki dan metode non-hierarki. Hasil klaster yang didapatkan kemudian diinterpretasikan untuk menggambarkan dan menjelaskan kondisi dari setiap klaster yang diperoleh. Ada dua jenis pengelompokan dalam analisis klaster, antara lain sebagai berikut:

### 1. Metode Hierarki

Metode analisis klaster dengan menggunakan metode hierarki ini ialah metode yang biasa digunakan untuk jumlah klaster yang akan dibentuk belum

diketahui. Hasil dari pengelompokan ini disajikan ke dalam bentuk *dendogram*, yang mirip dengan struktur diagram pohon atau *tree diagram*. Dalam metode hierarki terdapat alat yang digunakan untuk membantu memperjelas proses hierarki, alat tersebut disebut dengan “*dendogram*” (Sobari dan Usman, 2013). *Dendogram* dapat membantu dalam memperjelas hierarki atau tingkatan dalam suatu klaster yang terbentuk. Posisi garis pada skala digunakan dalam menunjukkan jarak (*distance*) untuk mana klaster digabung. *Dendogram* ini menunjukkan bagaimana proses suatu klaster akan terbentuk. Diawali dengan setiap objek yang direpresentasikan sebagai klaster berpisah hingga saat bagaimana semuanya berada dalam suatu klaster yang sama (Hair et al., 2010).

Menurut Johnson dan Wichern (2014), metode hierarki ini mempunyai beberapa teknik klasterisasi diantaranya yaitu:

a. *Single Linkage*

Teknik klasterisasi ini terjadi saat klaster-klaster yang terbentuk berdasarkan jarak antara anggota yang terdekat (*minimum distance*) di antara kedua klaster. Teknik ini diawali dengan cara mencari suatu jarak minimum  $D = \{d_{ij}\}$  kemudian menggabungkan objek-objek yang sesuai, dimisalkan objek  $U$  dan  $V$  tergabung menjadi suatu klaster ( $UV$ ). Kemudian melakukan perhitungan jarak antara klaster ( $UV$ ) yang terbentuk terhadap objek lainnya. Adapun jarak di antara klaster ( $UV$ ) terhadap klaster  $W$  lainnya dapat dihitung dengan perhitungan rumus sebagai berikut:

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (1)$$

Sehingga jarak yang berada antara klaster ( $ij$ ) dengan klaster  $k$  lainnya secara umum dapat dihitung menggunakan perhitungan rumus sebagai berikut:

$$d_{(ij)k} = \min\{d_{ik}, d_{jk}\} \quad (2)$$

b. *Complete Linkage*

Teknik klasterisasi ini terjadi saat klaster-klaster yang terbentuk berdasarkan jarak antara anggota yang terjauh (*maximum distance*) di antara kedua klaster. Teknik ini dimulai dengan cara menemukan suatu jarak minimum  $D = \{d_{ij}\}$  kemudian menggabungkan objek-objek yang sesuai, dimisalkan objek  $U$  dan  $V$  tergabung menjadi suatu klaster ( $UV$ ). Kemudian melakukan perhitungan jarak antara klaster ( $UV$ ) yang terbentuk terhadap objek lainnya. Adapun jarak di antara klaster ( $UV$ ) terhadap klaster  $W$  lainnya dapat dihitung dengan perhitungan rumus sebagai berikut:

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\} \quad (3)$$

Sehingga jarak yang berada antara klaster ( $ij$ ) dengan klaster  $k$  lainnya secara umum dapat dihitung dengan menggunakan perhitungan rumus sebagai berikut:

$$d_{(ij)k} = \max\{d_{ik}, d_{jk}\} \quad (4)$$

### c. Average Linkage

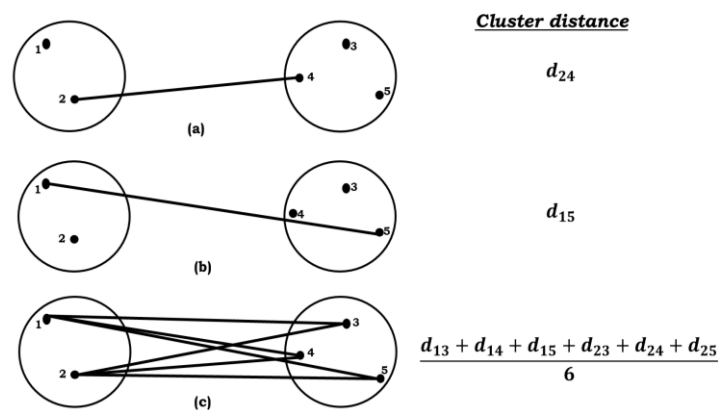
Teknik klusterisasi ini terjadi saat kluster-kluster yang terbentuk berdasarkan jarak rata-rata (*average distance*) di antara kedua kluster. Teknik ini diawali dengan cara menemukan jarak minimum  $D = \{d_{ij}\}$  kemudian menggabungkan objek-objek yang sesuai, dimisalkan objek  $U$  dan  $V$  tergabung menjadi suatu kluster ( $UV$ ). Kemudian melakukan perhitungan jarak antara kluster ( $UV$ ) yang terbentuk terhadap dengan objek lainnya. Adapun jarak di antara kluster ( $UV$ ) terhadap kluster  $W$  dapat dihitung dengan menggunakan perhitungan rumus sebagai berikut:

$$d_{(UV)W} = \text{average}\{d_{UW}, d_{VW}\} \quad (5)$$

Sehingga jarak yang berada antara kluster ( $ij$ ) dengan kluster  $k$  lainnya secara umum dapat dihitung dengan menggunakan perhitungan rumus sebagai berikut:

$$d_{(ij)k} = \text{average}\{d_{ik}, d_{jk}\} \quad (6)$$

Ketiga teknik yang diuraikan diatas dalam metode hierarki dapat diilustrasikan sebagai gambar di bawah ini:



Sumber: Johnson dan Wichern (2014)

**Gambar 3.** Jenis Teknik dari Metode Hierarki

## 2. Metode Non-Hierarki

Metode non-hierarki merupakan metode klusterisasi yang didasari pada *partitioning* dengan mendapatkan partisi dari data sehingga objek lainnya dalam suatu kluster akan memiliki kesamaan satu sama lain daripada objek yang berada dalam kluster lain (Triyanto, 2015). Metode partisi (*partitioning methods*) merupakan suatu metode dengan cara mempartisi data sebanyak kelompok ( $k$ ) dimana setiap partisi tersebut mewakili kluster dengan ( $k$ ) lebih kecil bila dibanding terhadap jumlah data ( $n$ ) yang dimiliki. Metode non-hierarki memiliki iterasi, mulai dari awal menentukan jumlah *cluster* hingga dilakukannya perhitungan untuk memperoleh suatu hasil *cluster*. Adapun metode partisi yaitu terdiri atas metode *k-means* dan *k-medoid* (Han et al., 2012).

a. *K-Means Cluster*

*K-Means Cluster* ialah salah satu dari metode pengklasteran data non-hierarki dimana ( $k$ ) akan diartikan sebagai konstanta banyaknya jumlah kluster yang akan dibentuk. Berdasarkan metode non-hierarki, metode *k-means* akan berupaya untuk membagi data yang ada ke dalam suatu bentuk dua atau lebih kluster dengan demikian metode ini akan membutuhkan jumlah kluster awal yang ditentukan diawal perhitungan sebagai ( $k$ ), serta kemudian akan membagi ( $n$ ) objek ke dalam ( $k$ ) kluster. Kemudian sesuai dengan ciri dari *cluster* yang baik, pada metode *k-means* kesamaan antar anggota dalam suatu kluster akan sangat tinggi sedangkan kesamaan antar anggota terhadap kluster lain yang terbentuk akan sangat rendah. Adapun kesamaan dalam anggota pada kluster ini dihitung berdasarkan dengan nilai kedekatan objek terhadap *mean* pada suatu kluster atau *centroid* kluster (Muzakir, 2014).

Menurut Ediyanto, Mara dan Satyahadewi (2013), adapun tahapan dalam melakukan perhitungan pengelompokan menggunakan algoritma *k-means* ini diantaranya sebagai berikut:

1. Menentukan ( $k$ ) sebagai jumlah kluster yang akan di bentuk.

Pada tahapan ini, menentukan banyaknya suatu kluster ( $k$ ) dilakukan dengan beberapa upaya pertimbangan teoritis atau menggunakan metode elbow dalam menentukan seberapa banyak kluster yang terbentuk.

2. Membangkitkan ( $k$ ) titik pusat kluster (*centroid*) awal secara acak

Pada tahapan ini, penentuan titik pusat awal dilakukan dengan cara acak dari objek-objek yang ada sebanyak ( $k$ ) *cluster*, selanjutnya untuk menghitung titik pusat *cluster* ke- $i$  menggunakan perhitungan rumus sebagai berikut:

$$v = \frac{\sum_{i=1}^n x_i}{n} ; i = 1,2,3, \dots, n \quad (7)$$

keterangan:

$v$  : Titik pusat (*centroid*) pada kluster

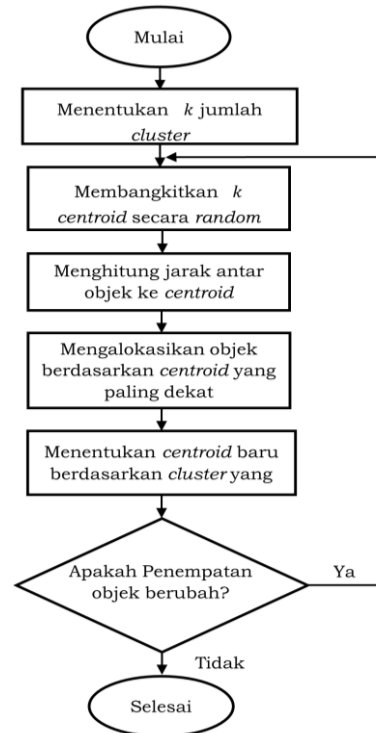
$x_i$  : Objek ke- $i$

$n$  : Banyaknya objek yang menjadi anggota kluster

3. Menghitung jarak setiap objek ke masing-masing *centroid* dari *cluster* yang dibentuk menggunakan rumus *Euclidean* pada persamaan (18).
4. Melakukan alokasi terhadap masing-masing objek ke suatu titik pusat yang paling dekat. Selanjutnya, menentukan titik pusat yang baru berdasarkan kluster yang terbentuk.
5. Melakukan menentukan titik pusat baru berdasarkan *cluster* yang terbentuk dengan menggunakan persamaan (7).

6. Ulangi langkah ke-3. Apabila alokasi objek telah sama dengan iterasi sebelumnya, maka iterasi dihentikan. Jika tidak, maka kembali ke cara yang ke-2.

Berdasarkan tahapan dari algoritma *k-means* tersebut, berikut adalah diagram alur yang terbentuk yaitu:



**Gambar 4.** Diagram Alur Algoritma *K-Means*

#### b. *K-Medoid Cluster*

*K-Medoid Cluster* merupakan teknik dari metode non-hierarki, sesuai dengan metode non-hierarki *k-medoid* berupaya untuk membagi data yang ada ke dalam suatu bentuk dua atau lebih klaster dengan demikian metode tersebut membutuhkan jumlah klaster awal yang ditentukan terlebih dahulu sebagai  $k$ , serta kemudian membagi  $n$  objek ke dalam  $k$  klaster (Han et al., 2012). Menurut Setyawati (2017), *k-medoid* ialah suatu metode yang menggunakan metode pengklasteran partisi untuk mengklasterkan sekumpulan  $n$  objek data menjadi sejumlah  $k$  klaster. Objek yang akan mewakili sebuah klaster akan disebut sebagai *medoid*. Klaster akan dibangun dengan cara menghitung kedekatan yang dimiliki antara *medoid* terhadap objek *non-medoid*.

Menurut Susanti (2021), pada proses perhitungan menggunakan algoritma *k-medoid* dapat terjadi perbedaan hasil yang diperoleh baik dari perhitungan secara manual dan *output* dari *software*. Hal ini dapat terjadi diakibatkan karena pada saat penentuan *medoid* di awal perhitungan akan berdampak pada hasil anggota *cluster* yang diperoleh. Menurut Han et al (2012), algoritma *k-medoid* ini



menggunakan *medoid* sebagai pusat klasternya sehingga *medoid* adalah suatu objek yang letaknya berada terpusat di dalam suatu klaster. Berikut adalah algoritma dari algoritma *k-medoid*, diantaranya yaitu:

1. Menentukan jumlah *cluster* ( $k$ ) sebagai banyaknya klaster yang dibentuk.
2. Membangkitkan ( $k$ ) *centroid* klaster secara acak.
3. Menghitung jarak objek *non-medoid* terhadap *medoid* di setiap *cluster*. Perhitungan jarak menggunakan persamaan (18), mengelompokkan setiap objek *non-medoid* ke dalam *medoid* yang terdekat. Selanjutnya, menghitung semua total jarak yang ada.
4. Lakukan perulangan *medoid* berdasarkan objek *non-medoid* dari *cluster* yang terbentuk sebagai kandidat untuk *medoid* yang baru.
5. Menghitung jarak di setiap objek *non-medoid* terhadap *medoid* yang baru serta mengalokasikan setiap objek *non-medoid* ke kandidat *medoid* yang terdekat, dilanjutkan dengan menghitung total jarak yang ada.
6. Menghitung selisih total jarak yang diperoleh ( $d_{total}$ ) dengan menggunakan perhitungan rumus sebagai berikut:

$$d_{total} = Total\ Jarak\ Baru - Total\ Jarak\ Lama \quad (8)$$

keterangan:

*Total Jarak Baru* : Nilai dari total jarak pada kandidat *medoid* yang baru

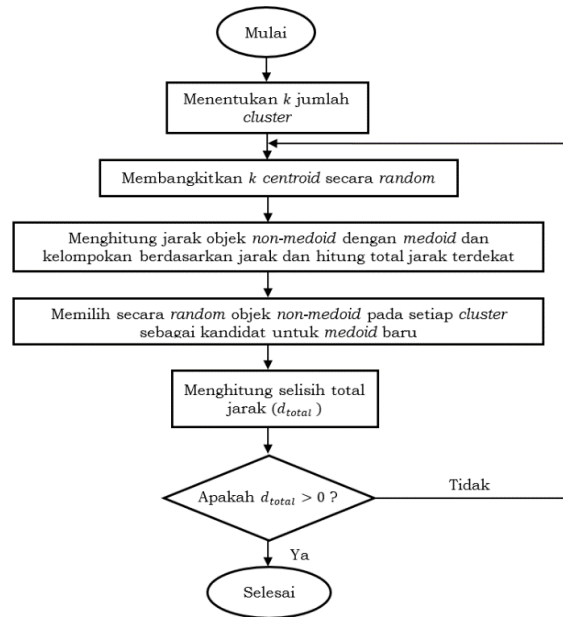
*Total Jarak Lama* : Nilai dari total jarak pada *medoid* yang lama

Nilai total jarak didapatkan dengan persamaan:

$$Total\ jarak = \sum_{i=1}^n d_{(i,j)} \quad (9)$$

7. Jika didapatkan nilai  $d_{total} < 0$ , maka kandidat *medoid* yang baru tersebut menjadi *medoid* yang baru dan jika diperoleh  $d_{total} > 0$  maka iterasi akan dihentikan. Jika tidak, ulangi langkah ke-4 hingga ke-7 sampai tidak ada perubahan *medoid* atau  $d_{total} > 0$ .

Berdasarkan tahapan dari algoritma *k-medoid* tersebut, berikut adalah diagram alur yang terbentuk yaitu:



**Gambar 5.** Diagram Alur Algoritma *K-Medoid*

## 2.5 Asumsi Analisis Cluster

Analisis *cluster* mempunyai asumsi atau syarat yang harus terpenuhi sebelum data dapat dilakukan perhitungan lebih lanjut. Adapun asumsi analisis klaster, yaitu:

1. Sampel mewakili populasi atau sampel *representative*

Menurut Putriana, Setyawan dan Noeryanti (2016) sampel yang telah diambil dari populasi harus dapat mewakili dari keseluruhan populasi dimana untuk menguji apakah sampel sudah layak untuk dianalisis maka dapat menggunakan uji *Kaiser-Meyer-Olkin* (KMO). Jika nilai hasil uji KMO berkisar diantara 0,5 hingga 1 maka dapat dikatakan bahwa sampel yang digunakan telah mewakili populasi yang digunakan atau sampel *representative*. Adapun hipotesis dan kriteria keputusan dalam uji *Kaiser-Meyer-Olkin* (KMO) adalah sebagai berikut:

$H_0$  : Sampel tidak mewakili populasi

$H_1$  : Sampel telah mewakili populasi

kriteria keputusan  $H_0$  ditolak jika  $KMO \geq 0,5$ , artinya sampel telah mewakili populasi atau  $H_0$  diterima jika  $KMO < 0,5$ , artinya sampel tidak mewakili populasi.

Menurut Nugroho (2008), menghitung nilai KMO dapat dihitung dengan rumus sebagai berikut:

$$KMO = \frac{\sum_i^n \sum_{j \neq i}^n r_{ij}^2}{\sum_i^n \sum_{j \neq i}^n r_{ij}^2 + \sum_i^n \sum_{j \neq i}^n a_{ij}^2}, \text{ dengan } a_{ij} = -\frac{v_{ij}}{\sqrt{v_{ii}v_{jj}}} \quad (10)$$

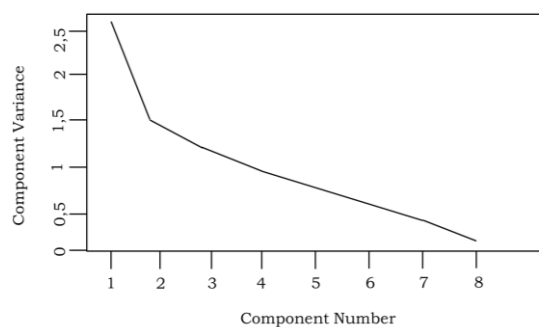
keterangan:

$a_{ij}$  : Koefisien korelasi parsial dari variabel  $i$  dan  $j$

- $r_{ij}$  : Koefisien korelasi sederhana dari variabel  $i$  dan  $j$   
 $v_{ij}$  : Kovarians dari variabel  $i$  dan  $j$   
 $i$  : 1,2,3, ...  $n$   
 $j$  : 1,2,3, ...  $n$   
 $n$  : Jumlah variabel

2. Tidak memiliki korelasi yang kuat antar variabel bebas

Mendeteksi korelasi yang kuat antar variabel bebas dapat dilihat apabila nilai korelasi antar variabel bebas ( $r$ ) besar dari 0,5 jika hal tersebut terjadi maka harus diatasi atau dapat menghilangkan variabel yang mempunyai nilai korelasi melebihi kriteria. Selain menghilangkan variabel, dapat juga di atasi menggunakan *Principal Component Analysis* (PCA) atau yang dikenal sebagai analisis komponen utama adalah analisis statistika multivariabel yang bisa digunakan untuk mentransformasi sejumlah variabel menjadi beberapa variabel baru yang tetap mempertahankan total keragaman dari variabel aslinya. Terdapat tiga teknik yang dapat digunakan dalam menentukan banyaknya komponen utama. Teknik pertama melihat pada kumulatif proporsi keragaman total yang mampu dijelaskan, minimum persentasi keragaman yang mampu dijelaskan terlebih dahulu dan banyaknya komponen yang paling kecil hingga batas itu terpenuhi dijadikan sebagai banyaknya komponen utama yang digunakan. Tidak ada patokan baku berapa batasan minimum, sebagian menyebutkan 70%, 80%, bahkan ada yang 90%. Teknik kedua melihat akar ciri yang terbentuk dari dengan akar ciri kurang dari satu maka tidak digunakan dan teknik ketiga merupakan penggunaan grafik *scree plot*. Banyaknya komponen utama yang digunakan dalam metode ini yaitu sebanyak  $k$ , yang mana  $k$  tersebut merupakan titik pada *scree plot* yang *plotnya* curam ke kiri namun tidak curam ke kanan. Untuk mengetahui variabel awal mana saja yang masuk ke dalam variabel baru dapat dilihat melalui rotasi faktor. Beberapa ahli menyarankan rotasi varimax, karena rotasi ini lebih mendekati kenyataan (Nugroho, 2008). Menurut Everitt dan Hothorn (2011), banyaknya komponen utama yang dapat terbentuk melalui *scree plot* tersebut ditunjukkan pada gambar berikut:



**Gambar 6.** *Scree Plot*

Menurut Nugroho (2008) hipotesis dan kriteria keputusan dalam uji ini adalah sebagai berikut:

$H_0$  : Terdapat korelasi yang kuat antar variabel bebas

$H_1$  : Tidak terdapat korelasi yang kuat antar variabel bebas

kriteria keputusan  $H_0$  ditolak jika  $r < 0,5$ , artinya tidak terdapat korelasi yang kuat antar variabel bebas atau  $H_0$  diterima jika  $r > 0,5$ , artinya terdapat korelasi yang kuat antar variabel bebas.

Menurut Johnson dan Wichern (2014), perhitungan nilai koefisien korelasi dapat dilakukan dengan menggunakan perhitungan koefisien korelasi pearson. Adapun nilai koefisien korelasi pearson untuk dua variabel kuantitatif dihitung dengan rumusan sebagai berikut:

$$r_{x_i, x_j} = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \sqrt{\sum_{i=1}^n (x_j - \bar{x}_j)^2}} \quad (11)$$

keterangan:

$r_{x_i, x_j}$  : Korelasi antara variabel ke- $i$  dan variabel ke- $j$

$x_i, x_j$  : Variabel ke- $i$  dan variabel ke- $j$

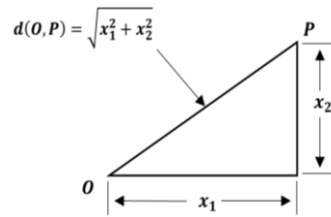
$\bar{x}_i, \bar{y}_i$  : Nilai rata-rata variabel ke- $i$  dan variabel ke- $j$

## 2.6 Ukuran Kesamaan Jarak

Ukuran kedekatan diperlukan untuk mengukur kesamaan objek yang akan dikelompokkan ke dalam suatu *cluster*. Pendekatan yang paling umum adalah mengukur kesamaan, yang dinyatakan sebagai jarak (*distance*) antara dua objek. Objek yang memiliki jarak lebih pendek akan lebih mirip satu sama lain daripada pasangan dengan jarak antara dua objek. Ukuran jarak yang digunakan dalam analisis *cluster* adalah jarak *Euclidean*. Jarak *Euclidean* adalah cara yang paling umum untuk mengukur jarak dalam analisis *cluster*, dimana ukuran kesamaan jarak dengan menggunakan perhitungan jarak *Euclidean* mempunyai tingkat akurasi yang lebih baik bila dibandingkan dengan metode yang lainnya (Nishom, 2019). Menurut Johnson dan Wichern (2014), jarak *Euclidean* adalah perhitungan dari jarak antara dua titik yang diturunkan berdasarkan teorema *pythagoras*. Jarak ini adalah jarak geometris di antara kedua objek. Apabila memperhitungkan titik  $P(x_1, x_2)$ , kemudian ditarik garis lurus dengan jarak  $d(O, P)$  dari titik  $P$  menuju titik asal  $O(0,0)$ . Berdasarkan teorema tersebut jarak  $d(O, P)$  dapat dituliskan dengan persamaan yaitu:

$$d(O, P) = \sqrt{x_1^2 + x_2^2} \quad (12)$$

Keadaan tersebut dapat diilustrasikan ke dalam gambar seperti berikut:



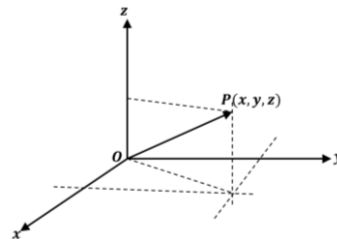
**Gambar 7.** Jarak pada Vektor Dimensi Dua

Perhitungan jarak pada ruang yang memiliki dimensi tiga dengan titik  $P(x,y,z)$  kemudian titik asalnya yaitu  $O(0,0,0)$  ditunjukkan pada **Gambar 8** dimisalkan bahwa  $x = x_1$ ,  $y = x_2$ , dan  $z = x_3$  maka akan diperoleh jarak sebagai berikut:

$$d^2(O,P) = (OX)^2 + (OY)^2 + (OZ)^2 \quad (13)$$

$$d^2(O,P) = (x)^2 + (y)^2 + (z)^2 \quad (14)$$

$$d(O,P) = \sqrt{(x_1)^2 + (x_2)^2 + (x_3)^2} \quad (15)$$



**Gambar 8.** Jarak pada Vektor Dimensi Tiga

Apabila titik  $P$  memiliki koordinat sebanyak  $p$  sehingga  $P(x_1, x_2, \dots, x_p)$ , kemudian ditarik garis lurus dari titik  $P$  terhadap titik asal yaitu  $O(0,0, \dots, 0)$ . Sehingga akan diperoleh jarak pada titik  $d(O,P)$  yang dapat dituliskan sebagai berikut:

$$d(O,P) = \sqrt{x_1^2 + x_2^2 + \dots + x_p^2} \quad (16)$$

Berdasarkan teorema *pythagoras* dapat dihitung jarak di antara kedua objek  $P$  dan  $Q$  terhadap koordinat titik  $P(x_1, x_2, \dots, x_p)$  kemudian koordinat titik  $Q(y_1, y_2, \dots, y_p)$  yaitu sebagai berikut:

$$d(O,P) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \quad (17)$$

Perhitungan jarak *Euclidean* pada analisis kluster di antara dua objek data sebanyak  $p$  variabel bisa dihitung menggunakan perhitungan sebagai berikut yaitu:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (18)$$

keterangan:

$d_{ij}$  : Jarak antara objek ke- $i$  dan objek ke- $j$

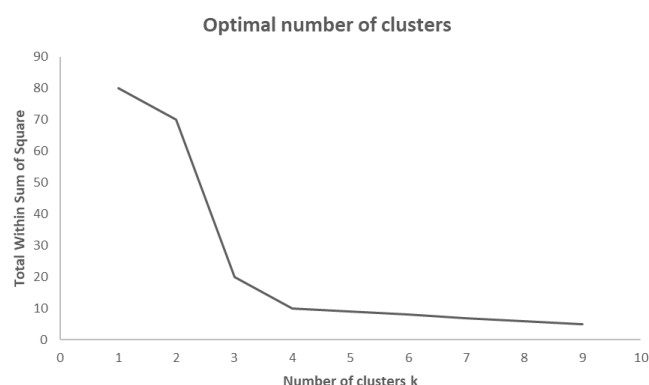
$x_{ik}$  : Nilai objek ke- $i$  pada variabel ke- $k$

$x_{jk}$  : Nilai objek ke- $j$  pada variabel ke- $k$

## 2.7 Metode *Elbow* untuk Penentuan Jumlah *Cluster* Optimal

Masalah utama dalam teknik analisis *k-means* dan *k-medoid* adalah menentukan seberapa banyak jumlah *cluster* yang digunakan. Tidak ada aturan yang khusus dalam menentukan seberapa banyak *cluster* yang digunakan dalam metode non-hierarki, namun ada suatu metode yang dapat digunakan untuk menentukan jumlah *cluster* optimal yaitu metode *Elbow*. Metode *Elbow* adalah suatu metode yang dapat digunakan dalam menghasilkan suatu informasi untuk menentukan seberapa banyak jumlah *cluster* terbaik dengan cara melihat nilai persentase perbandingan antara jumlah *cluster* yang akan membentuk siku atau *elbow* pada suatu titik dari nilai total *Within cluster Sum of Square* atau WSS (Han et al., 2012).

Menurut Merliana et al (2015) banyaknya *cluster* yang dapat terbentuk melalui metode *elbow*, dilihat dari nilai  $k$  yang mengalami penurunan paling drastis dimana akan membentuk *elbow* atau siku. Terlihat penurunan paling drastis yang membentuk siku ada pada  $k = 3$  yang ditunjukkan oleh gambar berikut:



**Gambar 9.** Grafik Metode *Elbow*

Menurut Irwanto dan Soelaiman (2012), Nilai yang digunakan dalam metode *elbow* ini merupakan perhitungan *Within cluster Sum of Square* (WSS). Adapun perhitungan WSS adalah sebagai berikut:

$$WSS = \sum_{k=1}^K \sum_{x_i \in S_k} (x_i - c_k)^2 \quad (19)$$

keterangan:

$K$  : Banyaknya *cluster*

$x_i$  : Data ke- $i$

$S_k$  : Himpunan dari elemen *cluster*  $k$

$c_k$  : Rata-rata (pusat) dalam *cluster*  $k$

## 2.8 Calinski Harabasz Index

*Calinski Harabasz Index* (CHI) adalah metode evaluasi *cluster* yang digunakan untuk mengevaluasi kelayakan atau kebaikan dari suatu hasil algoritma *cluster* yang digunakan. Perhitungan CHI menggunakan nilai *Sum of Squared Within cluster* (SSW) sebagai *compactness*, yaitu mengukur seberapa dekat objek-objek dalam *cluster* yang sama dan *Sum of Squared Between cluster* (SSB) sebagai *separation*, yaitu mengukur seberapa baik suatu *cluster* yang terpisah dengan *cluster* lainnya. Hal ini menjadi kelebihan dari *Calinski Harabasz Index* (CHI) bila dibandingkan dengan metode lainnya, CHI meyakinkan bahwa sesama anggota dalam suatu *cluster* semakin mirip sedangkan antar *cluster* yang berbeda semakin berbeda. Dengan kata lain, CHI mampu mengetahui *cluster* yang memiliki kriteria atau kondisi paling optimal sehingga terpilih *cluster* yang terbaik (Kassambara, 2017).

Menurut Khairati, Adlina, Hertono dan Handari (2019) *Calinski Harabasz Index* (CHI) digunakan untuk mengukur kebaikan objek data yang dialokasikan ke dalam suatu *cluster* yang ditunjukkan apabila nilai CHI yang dihasilkan lebih besar bila dibandingkan dengan nilai CHI lainnya maka *cluster* dengan hasil nilai CHI terbesar merupakan hasil *cluster* terbaik. Adapun perhitungan dengan menggunakan CHI dihitung menggunakan persamaan sebagai berikut:

$$CH_K = \frac{SSB}{SSW} \times \frac{N-K}{K-1} \quad (20)$$

keterangan:

$CH_K$  : *Calinski harabasz index* pada  $K$  *cluster*

$N$  : Jumlah objek yang digunakan

$K$  : Jumlah *cluster*

Dimana  $SSW$  dan  $SSB$  diperoleh dari persamaan berikut:

$$SSW = \sum_{i=1}^K \sum_{x \in C_i} d^2(m_i, x) \quad (21)$$

$$SSB = \sum_{i=1}^N n_i (d^2(m_i, M)) \quad (22)$$

$$M = \frac{\sum_{i=1}^n x_i}{n} \quad (23)$$

keterangan:

$x_i$  : Objek ke- $i$

$c_i$  : *Cluster* ke- $i$

$m_i$  : Pusat *cluster* ke- $i$

$n_i$  : Banyaknya anggota *cluster* ke- $i$

$M$  : *Centroid Global*

### III. METODOLOGI PENELITIAN

#### 3.1 Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Bersumber dari website Badan Pusat Statistik Provinsi Jambi <https://jambi.bps.go.id> berdasarkan publikasi Provinsi Jambi dalam angka tahun 2021.

#### 3.2 Objek dan Variabel Penelitian

##### I. Objek

Objek yang digunakan dalam penelitian ini adalah seluruh kabupaten/kota di Provinsi Jambi sebanyak 11 kabupaten/kota.

##### II. Variabel

Variabel yang digunakan dalam penelitian ini merupakan variabel yang menggambarkan bahwa penyakit menular tertinggi di setiap kabupaten/kota di Provinsi Jambi. Adapun variabel penelitian yang digunakan yaitu didefinisikan pada tabel berikut:

**Tabel 1.** Variabel Penelitian

Variabel	Definisi	Satuan
$X_1$	Jumlah kasus diare	Jiwa
$X_2$	Jumlah kasus malaria	Jiwa
$X_3$	Jumlah kasus tuberkulosis paru	Jiwa
$X_4$	Jumlah kasus pneumonia	Jiwa
$X_5$	Jumlah kasus HIV/AIDS	Jiwa
$X_6$	Jumlah kasus DBD	Jiwa
$X_7$	Jumlah kasus kusta	Jiwa
$X_8$	Jumlah kasus IMS	Jiwa
$X_9$	Jumlah kasus campak	Jiwa

#### 3.3 Struktur Data Penelitian

Penelitian ini memiliki struktur data sebagai berikut:

**Tabel 2.** Struktur Data Penelitian

Kabupaten/Kota di Provinsi Jambi	$X_1$	$X_2$	$X_3$	...	$X_9$
$i = 1$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	...	$X_{1,9}$
$i = 2$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	...	$X_{2,9}$
$i = 3$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	...	$X_{3,9}$
$i = 4$	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	...	$X_{4,9}$
$i = 5$	$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	...	$X_{5,9}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i = 11$	$X_{11,1}$	$X_{11,2}$	$X_{11,3}$	...	$X_{11,9}$

dengan  $i = 1,2,3, \dots, 11$  merupakan kabupaten/kota di Provinsi Jambi yang terdiri dari:

- 1 : Kerinci
- 2 : Merangin



- 3 : Sarolangun
- 4 : Batanghari
- 5 : Muaro Jambi
- 6 : Tanjung Jabung Timur
- 7 : Tanjung Jabung Barat
- 8 : Tebo
- 9 : Bungo
- 10 : Kota Jambi
- 11 : Kota Sungai Penuh

### 3.4 Teknik Analisis Data

Metode analisis yang digunakan pada penelitian ini merupakan metode *K-Means* dan *K-Medoid* dengan ukuran jarak *Euclidean*. Adapun langkah-langkah yang dilakukan pada penelitian ini adalah:

1. Merumuskan Masalah

Perumusan masalah adalah suatu proses yang sangat mendasar dalam analisis *cluster*. Kemudian melakukan identifikasi objek dan variabel yang digunakan dalam penelitian harus jelas serta relevan terhadap permasalahan yang ada. Kemudian, dilanjutkan dengan pengumpulan dan penginputan data.

2. Melakukan uji asumsi

Dalam melakukan pengolahan data menggunakan analisis *cluster* terdapat beberapa asumsi yang harus dipenuhi pada analisis *cluster* tersebut di antaranya sebagai berikut:

- a. Sampel mewakili populasi atau sampel *representative*

Menguji apakah sampel sudah layak untuk dianalisis maka dapat menggunakan uji *Kaiser-Mayer-Olkin* (KMO). Jika nilai hasil uji KMO berkisar 0,5 sampai 1 maka dapat dikatakan bahwa sampel yang digunakan telah mewakili populasi yang digunakan atau sampel *representative*.

- b. Tidak memiliki korelasi yang kuat antar variabel bebas

Apabila nilai korelasi antar variabel bebas ( $r$ ) besar dari 0,5 maka harus di atasi dengan menghilangkan variabel yang mempunyai nilai korelasi melebihi kriteria atau dapat juga di atasi menggunakan *Principal Component Analysis* (PCA).

3. Menentukan ukuran jarak kesamaan antar objek

Objek dengan jarak yang lebih pendek akan lebih mirip satu sama lain daripada pasangan dengan jarak antara dua objek. Pada penelitian ini

menentukan ukuran jarak kesamaan antar objek yang digunakan adalah perhitungan jarak *Euclidean*.

4. Melakukan pengelompokan

Proses pengelompokan menggunakan metode *k-means* dan *k-medoid*. Dengan jumlah *cluster* ditentukan menggunakan metode *Elbow*. Setelah proses pengelompokan dilakukan secara manual akan dibandingkan dengan pengelompokan menggunakan *software R*.

5. Evaluasi

Melakukan evaluasi dengan cara perbandingan hasil algoritma *k-means* dan *k-medoid* menggunakan *Calinski Harabasz Index (CHI)*.

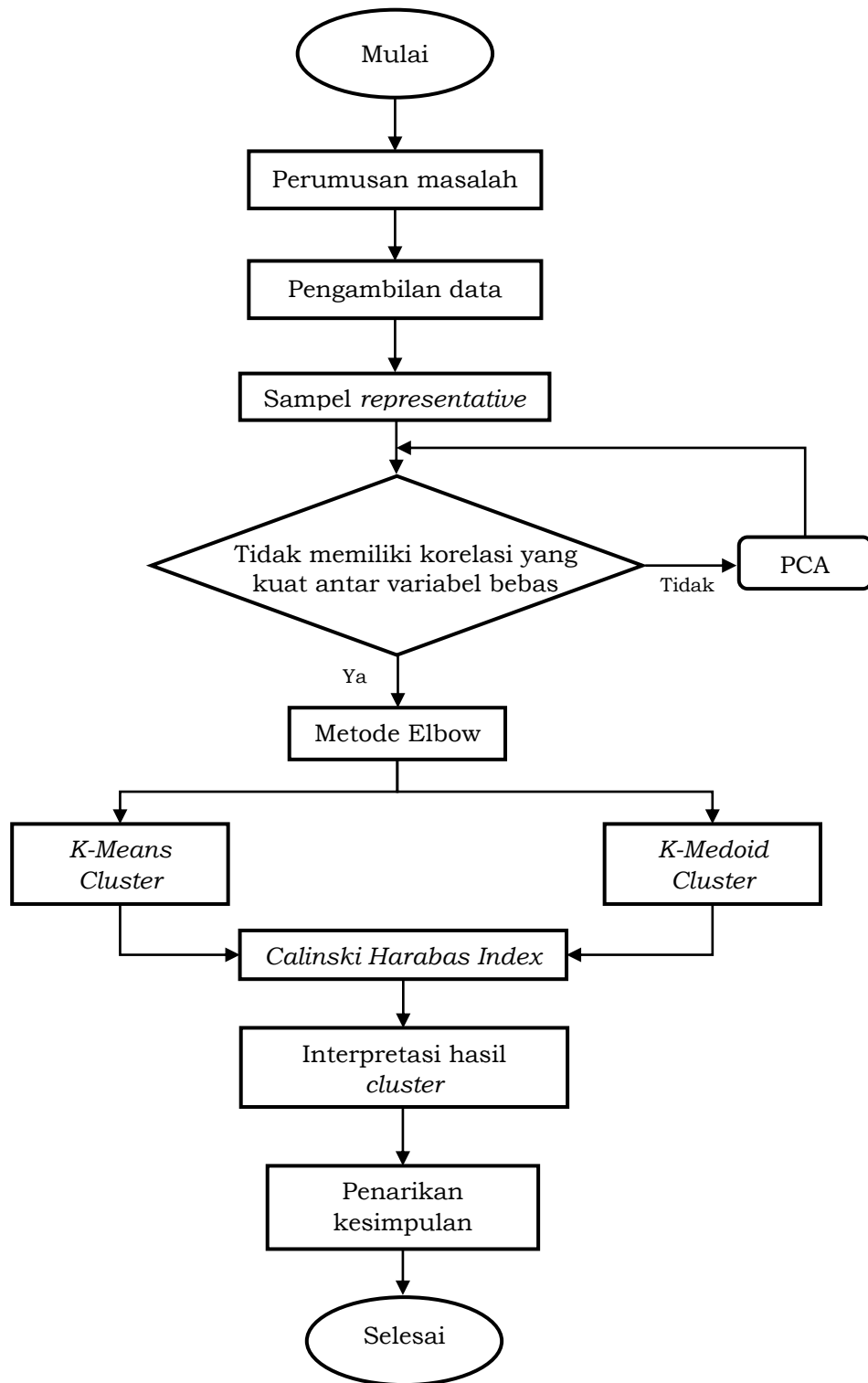
6. Melakukan interpretasi *cluster*

Memberikan informasi terkait *cluster* yang terbentuk setelah mendapatkan hasil dari perhitungan evaluasi menggunakan *calinski harabasz index*.

7. Penarikan kesimpulan

Pada akhir dari metode penelitian ini dilakukan penarikan kesimpulan sebagai jawaban dari permasalahan.

### 3.5 Diagram Alur Penelitian



**Gambar 10.** Diagram Alur Penelitian

## IV. HASIL DAN PEMBAHASAN

### 4.1 Deskripsi Data

Data diperoleh melalui *website* Badan Pusat Statistik Provinsi Jambi berdasarkan publikasi Provinsi Jambi dalam angka tahun 2021 yang memuat data penyakit menular tahun 2020, data yang digunakan pada penelitian ini terdapat pada **Lampiran 1**. Adapun deskripsi data penelitian adalah sebagai berikut:

**Tabel 3.** Deskripsi Data

Variabel	N	Minimum	Maximum	Mean	Std. Deviation
$X_1$	11	732	6653	4216,2727	1798,9073
$X_2$	11	200	6619	2735,2727	2373,2199
$X_3$	11	52	1021	439,8182	275,3193
$X_4$	11	16	536	164,7273	165,3028
$X_5$	11	0	683	69,5455	203,5423
$X_6$	11	44	717	186,2727	188,6961
$X_7$	11	0	30	5,5455	8,6529
$X_8$	11	0	254	25	76,0591
$X_9$	11	0	40	5,6364	11,8260

**Tabel 3** diatas menyajikan data yang digunakan dalam penelitian dengan jumlah data yaitu sebanyak 11 data berdasarkan jumlah kabupaten/kota yang ada di Provinsi Jambi. Variabel  $X_1$  memiliki nilai minimum sebesar 732 jiwa yang berada di Kota Sungai Penuh, nilai maksimumnya sebesar 6653 jiwa berada di Kabupaten Merangin dengan nilai meannya sebesar 4216,2727 jiwa dan std. Deviasinya sebesar 1798,9073 jiwa. Variabel  $X_2$  memiliki nilai minimum sebesar 200 jiwa yang berada di Kabupaten Tanjung Jabung Barat, nilai maksimumnya sebesar 6619 jiwa berada di Kota Jambi dengan nilai meannya sebesar 2735,2727 jiwa dan std. Deviasinya sebesar 2373,2199 jiwa. Variabel  $X_3$  memiliki nilai minimum sebesar 52 jiwa yang berada di Kota Sungai Penuh, nilai maksimumnya sebesar 1021 jiwa berada di Kota Jambi dengan nilai meannya sebesar 439,8182 jiwa dan std. Deviasinya sebesar 275,3193 jiwa. Variabel  $X_4$  memiliki nilai minimum sebesar 16 jiwa yang berada di Kabupaten Tanjung Jabung Barat, nilai maksimumnya sebesar 536 jiwa berada di Kabupaten Muaro Jambi dengan nilai meannya sebesar 164,7273 jiwa dan std. Deviasinya sebesar 165,3028 jiwa. Variabel  $X_5$  memiliki nilai minimum sebesar 0 jiwa yang berada di Kota Sungai Penuh, nilai maksimumnya sebesar 683 jiwa berada di Kota Jambi dengan nilai meannya sebesar 69,5455 jiwa dan std. Deviasinya sebesar 203,5423 jiwa. Variabel  $X_6$  memiliki nilai minimum sebesar 44 jiwa yang berada di Kabupaten Sarolangun, nilai maksimumnya sebesar 717 jiwa berada di Kota Jambi dengan nilai meannya sebesar 186,2727 jiwa dan std. Deviasinya sebesar 188,6961 jiwa. Variabel  $X_7$  memiliki nilai minimum sebesar 0 jiwa yang berada di Kabupaten Kerinci, Tanjung Jabung Barat, dan Kota Sungai Penuh, nilai maksimumnya

sebesar 30 jiwa berada di Kabupaten Tanjung Jabung Timur dengan nilai meannya sebesar 5,5455 jiwa dan std. Deviasinya sebesar 8,6529 jiwa. Variabel  $X_8$  memiliki nilai minimum sebesar 0 jiwa yang berada di Kabupaten Kerinci, Sarolangun, Batanghari, Muaro Jambi, Tanjung Jabung Timur, Tebo, dan Kota Sungai Penuh, nilai maksimumnya sebesar 254 jiwa berada di Kota Jambi dengan nilai meannya sebesar 25 jiwa dan std. Deviasinya sebesar 76,0591 jiwa. Variabel  $X_9$  memiliki nilai minimum sebesar 0 jiwa yang berada di Kabupaten Kerinci, Sarolangun, Muaro Jambi, Tanjung Jabung Barat, dan Bungo, nilai maksimumnya sebesar 40 jiwa berada di Kota Jambi dengan nilai meannya sebesar 5,6364 jiwa dan std. Deviasinya sebesar 11,8260 jiwa.

#### 4.2 Asumsi Analisis Cluster

Sebelum melakukan pengelompokan terlebih dahulu akan dilakukan pemeriksaan asumsi. Adapun asumsi-asumsi yang diperiksa adalah sebagai berikut:

a. Sampel mewakili populasi atau sampel *representative*

Dalam menguji apakah sampel sudah layak untuk dianalisis maka dapat menggunakan uji *Kaiser-Mayer-Olkin* (KMO) dengan hipotesis dan kriteria keputusan sebagai berikut:

$H_0$  : Sampel tidak mewakili populasi

$H_1$  : Sampel telah mewakili populasi

Kriteria keputusan  $H_0$  ditolak jika  $KMO \geq 0,5$ . Berikut adalah hasil pengujian KMO menggunakan *software R*:

<b>Tabel 4.</b> Hasil Uji KMO	
<i>Kaiser-Meyer-Olkin Measure of Sampling Adequacy.</i>	0,545

Berdasarkan **Tabel 4** diperoleh nilai KMO sebesar 0,545 sehingga  $H_0$  ditolak yang berarti bahwa sampel telah mewakili populasi.

b. Tidak memiliki korelasi yang kuat antar variabel bebas

Dalam asumsi kedua ini, karena korelasi Pearson digunakan untuk data berdistribusi normal maka akan diujikan hal tersebut terlebih dahulu. Normalitas data dapat dilihat melalui uji normalitas Kolmogorov-Smirnov. Adapun hipotesis dan kriteria keputusan dalam uji ini adalah sebagai berikut:

$H_0$  : Data tidak berdistribusi normal

$H_1$  : Data telah berdistribusi normal

Kriteria keputusan  $H_0$  ditolak jika nilai signifikan  $> 0,05$ . Adapun hasil pengujiannya adalah sebagai berikut:

**Tabel 5.** Hasil Uji Normalitas Tiap Variabel

<b>Variabel</b>	<b>Sig.</b>
$X_1$	0,949
$X_2$	0,922
$X_3$	0,920
$X_4$	0,638
$X_5$	0,006
$X_6$	0,386
$X_7$	0,282
$X_8$	0,013
$X_9$	0,139

Berdasarkan **Tabel 5** diperoleh bahwa nilai signifikan pada variabel  $X_5$  dan  $X_8 < 0,05$  sehingga  $H_0$  diterima yang berarti bahwa data tidak berdistribusi normal. Selanjutnya nilai korelasi Pearson dari seluruh variabel dapat dilihat pada **Lampiran 2**. Adapun hipotesis dan kriteria keputusan dalam uji ini adalah sebagai berikut:

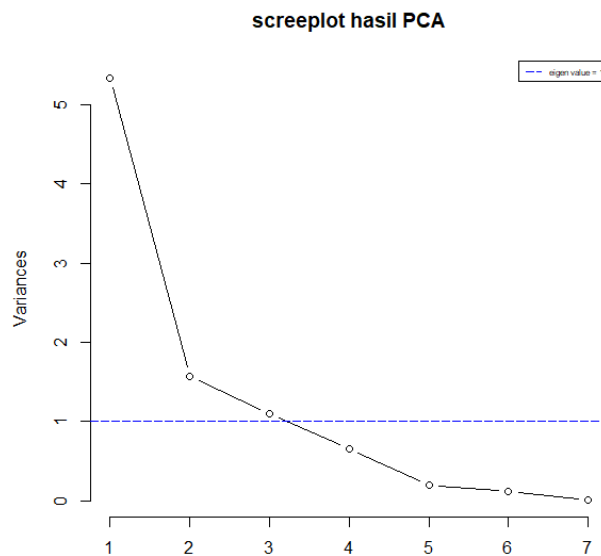
$H_0$  : Terdapat korelasi yang kuat antar variabel bebas

$H_1$  : Tidak terdapat korelasi yang kuat antar variabel bebas

Kriteria keputusan  $H_0$  ditolak jika  $r < 0,5$ , artinya tidak terdapat korelasi yang tinggi atau  $H_0$  diterima jika  $r > 0,5$ , artinya terdapat korelasi yang kuat antar variabel bebas.

Berdasarkan hasil tersebut terdapat nilai korelasi antar variabel yang nilainya melebihi 0,5 ( $r > 0,5$ ). Hal tersebut mengindikasikan bahwa terdapat korelasi antar variabel yang kuat antar variabel bebas, karena data tidak berdistribusi normal dan terdapat korelasi yang kuat antar variabel bebas maka akan di atasi menggunakan PCA.

Analisis faktor menggunakan metode ekstraksi komponen utama (PCA) ditentukan berdasarkan metode *scree plot*. *Scree plot* adalah suatu *plot* nilai eigen terhadap jumlah faktor yang diekstraksi. Titik pada tempat dimana *scree* mulai terjadi menunjukkan banyaknya faktor yang tepat. Titik ini terjadi ketika *scree* mulai terlihat mendatar. Pada **Gambar 10** terlihat bahwa *scree plot* mulai mendatar pada ekstraksi variabel-variabel awal menjadi 3 faktor.



**Gambar 11.** Scree Plot

Berdasarkan hasil *scree plot* di atas dapat disimpulkan bahwa ekstraksi faktor yang paling tepat adalah 3 faktor, sehingga data penyakit menular terbagi menjadi 3 faktor dan 3 faktor inilah yang akan digunakan sebagai data untuk melakukan analisis selanjutnya. Untuk menentukan variabel mana saja yang masuk ke dalam suatu faktor dapat dilihat dari hasil *rotated component matrix* yang terlihat pada **Lampiran 4**. Adapun data hasil dari analisis faktor ini dapat terlihat pada tabel berikut ini:

**Tabel 6.** Data Hasil Analisis Faktor

<b>n</b>	<b>V1</b>	<b>V2</b>	<b>V3</b>
1	-1,7497	0,6903	-0,271
2	0,8688	-1,8646	-1,0113
3	-0,6410	0,0651	-0,9872
4	-0,6762	0,6830	-0,2963
5	0,4202	-1,9528	0,3759
6	-0,9418	-0,0045	2,8732
7	-0,8616	1,0946	-0,4705
8	-0,1681	-1,1179	0,2117
9	-0,5664	-0,5952	-0,3762
10	6,4689	1,2792	0,1661
11	-2,1533	1,7227	-0,2144

Adapun variabel hasil analisis faktor yang memuat keseluruhan informasi dari 9 variabel yang ada dapat ditunjukkan pada tabel berikut:

**Tabel 7.** Variabel Hasil Analisis Faktor

<b>Variabel</b>	<b>Komponen</b>
V1	Jumlah kasus malaria
	Jumlah kasus tuberkulosis paru
	Jumlah kasus HIV/AIDS
	Jumlah kasus DBD

	Jumlah kasus IMS
	Jumlah kasus campak
V2	Jumlah kasus diare
	Jumlah kasus pneumonia
V3	Jumlah kasus kusta

Selanjutnya berdasarkan data baru hasil analisis faktor akan dilakukan pengecekan asumsi kembali yaitu sebagai berikut:

- a. Sampel mewakili populasi atau sampel *representative*

Berikut adalah hasil pengujian KMO menggunakan *software R*:

<b>Tabel 8.</b> Hasil Uji KMO Data Baru	
<i>Kaiser-Meyer-Olkin Measure of Sampling Adequacy.</i>	0,50

Berdasarkan **Tabel 8** diperoleh nilai KMO sebesar 0,50 yang berarti bahwa sampel telah mewakili populasi.

- b. Tidak memiliki korelasi yang kuat antar variabel bebas

Dalam asumsi kedua ini, karena korelasi Pearson digunakan untuk data berdistribusi normal maka akan diujikan hal tersebut terlebih dahulu. Adapun hasil pengujiannya adalah sebagai berikut:

**Tabel 9.** Hasil Uji Normalitas Tiap Variabel Data Baru

Variabel	Sig
V1	0,435
V2	0,935
V3	0,402

Berdasarkan **Tabel 9** diperoleh nilai signifikan tiap variabel besar dari 0,05 yang berarti bahwa data baru telah berdistribusi normal. Selanjutnya nilai korelasi Pearson dari seluruh variabel dapat dilihat pada tabel berikut:

**Tabel 10.** Hasil Korelasi Data Baru

Variabel	V1	V2	V3
V1	1	0,000	0,000
V2	0,000	1	0,000
V3	0,000	0,000	1

Berdasarkan **Tabel 10** diperoleh bahwa setiap variabel bebas tidak memiliki korelasi yang lebih dari 0,5 ( $r < 0,5$ ) maka dapat disimpulkan bahwa setiap variabel bebas tidak memiliki korelasi yang kuat. Sehingga setiap asumsi telah terpenuhi.

### 4.3 Menentukan Jumlah Cluster Optimal

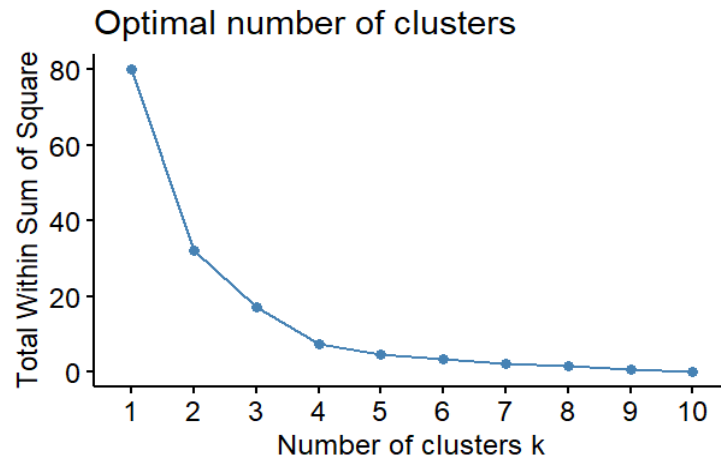
Pada pengelompokan *k-means* dan *k-medoid*, sebelum dilakukannya perhitungan terlebih dahulu jumlah *cluster* ditentukan. Dalam hal ini, penentuan jumlah *cluster* menggunakan metode *elbow* untuk mendapatkan jumlah *cluster* yang optimal. Metode *elbow* ini menggunakan nilai total *Within Sum of Square* (WSS) sebagai penentu *k* jumlah *cluster* optimalnya atau dapat dilihat



berdasarkan hasil grafik metode *elbow* dimana jumlah *cluster* yang akan membentuk siku pada suatu titik merupakan jumlah *cluster* optimalnya.

Adapun jumlah *cluster* optimal pada pengelompokan *k-means* dan *k-medoid* adalah sebagai berikut:

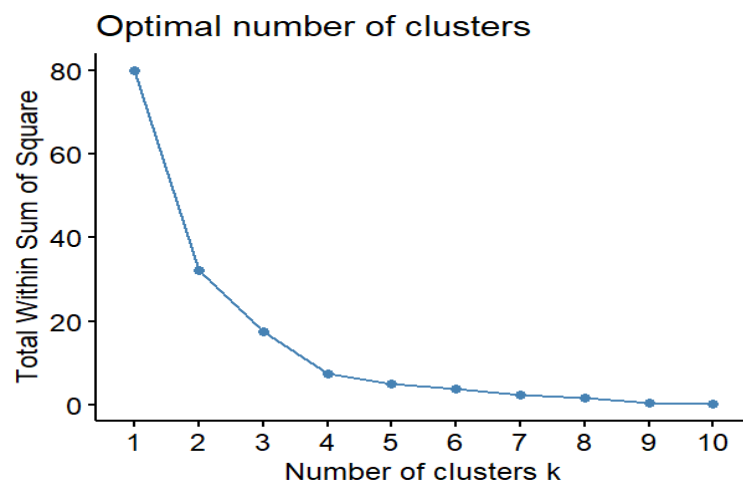
a. Jumlah *cluster* optimal pada pengelompokan *k-means*



**Gambar 12.** Grafik *Elbow K-Means*

Pada pengelompokan *k-means* dengan penentuan jumlah *cluster* optimal menggunakan metode *elbow* pada  $k = 2$  terlihat garis mengalami penurunan drastis yang membentuk *elbow* atau siku serta nilai *Within cluster Sum of Square* sebesar 59,8%, yang artinya bahwa dengan menggunakan metode ini diperoleh  $k$  optimal pada saat berada di  $k = 2$ .

b. Jumlah *cluster* optimal pada pengelompokan *k-medoid*



**Gambar 13.** Grafik *Elbow K-Medoid*

Pada pengelompokan *k-medoid* dengan penentuan jumlah *cluster* optimal menggunakan metode *elbow* pada  $k = 2$  terlihat garis mengalami penurunan drastis yang membentuk *elbow* atau siku serta nilai *Within cluster Sum of Square*

sebesar 59,8%, yang artinya bahwa dengan menggunakan metode ini diperoleh  $k$  optimal pada saat berada di  $k = 2$ .

#### 4.4 Pengelompokan dengan Algoritma K-Means

a. Titik pusat *cluster* (*centroid*) awal

Penentuan *centroid* awal sebagai titik pusat *cluster* awal ditentukan sesuai dengan jumlah  $k$  yang diambil. Jumlah  $k$  *cluster* optimal berdasarkan metode *elbow* sebanyak 2 *cluster* optimal maka jumlah *centroid* awal yang akan dipilih sebanyak 2. Tabel berikut merupakan tabel *centroid* awal pada perhitungan secara manual yaitu sebagai berikut:

<b>Centroid</b>	<b>V1</b>	<b>V2</b>	<b>V3</b>
C1	6,4689	1,2792	0,1661
C2	-0,641	0,0651	-0,9872

Berdasarkan **Tabel 11** C1 adalah *centroid* pertama dengan mengambil objek ke-10 sebagai pusat *clusternya*. Sedangkan C2 adalah *centroid* kedua dengan mengambil objek ke-3 sebagai pusat *clusternya*.

b. Perhitungan jarak antara *centroid* dan objek data

Ukuran jarak antara *centroid* awal yang disimbolkan sebagai C1 dan C2 terhadap objek data menggunakan ukuran jarak *euclidean*. Perhitungan manual jarak antara *centroid* pertama (C1) terhadap objek data adalah sebagai berikut: (selengkapnya pada **Lampiran 9.**)

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

$$\begin{aligned} d_{1,c1} &= \sqrt{\sum_{k=1}^3 (x_{1k} - c_{1k})^2} \\ &= \sqrt{(x_{11} - c_{11})^2 + (x_{12} - c_{12})^2 + (x_{13} - c_{13})^2} \\ &= \sqrt{(-1,7497 - (6,4689))^2 + (0,6903 - (1,2792))^2 + ((-0,271) - (0,1661))^2} \\ &= \sqrt{(-8,2186)^2 + (-0,589)^2 + (-0,4371)^2} \\ &= 8,2512 \end{aligned}$$

$$\begin{aligned} d_{2,c1} &= \sqrt{\sum_{k=1}^3 (x_{2k} - c_{1k})^2} \\ &= \sqrt{(x_{21} - c_{11})^2 + (x_{22} - c_{12})^2 + (x_{23} - c_{13})^2} \\ &= \sqrt{(0,8688 - (6,4689))^2 + ((-1,8646) - (1,2792))^2 + ((-1,0113) - (0,1661))^2} \\ &= \sqrt{(-5,6001)^2 + (-3,1438)^2 + (-1,1775)^2} \\ &= 6,5292 \end{aligned}$$

⋮

$$\begin{aligned} d_{11,c1} &= \sqrt{\sum_{k=1}^3 (x_{11k} - c_{1k})^2} \\ &= \sqrt{(x_{111} - c_{11})^2 + (x_{112} - c_{12})^2 + (x_{113} - c_{13})^2} \\ &= \sqrt{(-2,1533) - (6,4689))^2 + (1,7227 - (1,2792))^2 + ((-0,2145) - (0,1661))^2} \end{aligned}$$

$$= \sqrt{(-8,6222)^2 + (0,4434)^2 + (-0,3806)^2}$$

$$= 8,6419$$

Perhitungan manual jarak antara *centroid* kedua (C2) terhadap objek data adalah sebagai berikut: (selengkapnya pada **Lampiran 9.**)

$$d_{1,c2} = \sqrt{\sum_{k=1}^3 (x_{1k} - c_{2k})^2}$$

$$= \sqrt{(x_{11} - c_{21})^2 + (x_{12} - c_{22})^2 + (x_{13} - c_{23})^2}$$

$$= \sqrt{(-1,7497 - (-0,641))^2 + (0,6903 - (0,0651))^2 + ((-0,271) - (-0,9872))^2}$$

$$= \sqrt{(-1,1087)^2 + (0,6251)^2 + (0,7161)^2}$$

$$= 1,4604$$

$$d_{2,c2} = \sqrt{\sum_{k=1}^3 (x_{2k} - c_{2k})^2}$$

$$= \sqrt{(x_{21} - c_{21})^2 + (x_{22} - c_{22})^2 + (x_{23} - c_{23})^2}$$

$$= \sqrt{((0,8688) - (-0,641))^2 + ((-1,8646) - (0,0651))^2 + ((-1,0113) - (-0,9872))^2}$$

$$= \sqrt{(1,5098)^2 + (-1,9297)^2 + (-0,0241)^2}$$

$$= 2,4502$$

⋮

$$d_{11,c2} = \sqrt{\sum_{k=1}^3 (x_{11k} - c_{2k})^2}$$

$$= \sqrt{(x_{111} - c_{21})^2 + (x_{112} - c_{22})^2 + (x_{113} - c_{23})^2}$$

$$= \sqrt{(-2,1533 - (-0,641))^2 + (1,7227 - (0,0651))^2 + ((-0,2145) - (-0,9872))^2}$$

$$= \sqrt{(-1,5123)^2 + (1,65756)^2 + (0,7727)^2}$$

$$= 2,3731$$

c. Mengelompokkan objek berdasarkan *centroid* yang paling dekat

Setelah melakukan perhitungan jarak antara *centroid* dan objek data, dilakukan pengelompokan objek berdasarkan *centroid* atau titik pusat yang paling dekat. Tabel berikut merupakan tabel pengelompokan objek berdasarkan *centroid* yang paling dekat adalah sebagai berikut:

**Tabel 12.** Alokasi Objek Terhadap *Centroid* Terdekat

No.	C1	C2	Kedekatan	Hasil
1	8,2512	1,4605	1,4605	2
2	6,5292	2,4503	2,4503	2
3	7,3044	0,0000	0,0000	2
4	7,1848	0,9275	0,9275	2
5	6,8613	2,6564	2,6564	2
6	7,9934	3,8727	3,8727	2
7	7,3604	1,1728	1,1728	2
8	7,0568	1,7495	1,7495	2
9	7,3009	0,9027	0,9027	2
10	0,0000	7,3044	0,0000	1
11	8,6420	2,3731	2,3731	2

Berdasarkan **Tabel 12** diperoleh hasil pengelompokan objek berdasarkan *centroid* yang paling dekat ke dalam *cluster* tersebut. Selanjutnya, menentukan titik pusat (*centroid*) baru berdasarkan *cluster* yang terbentuk. Adapun perhitungan manual titik pusat (*centroid*) baru adalah sebagai berikut:

$$\begin{aligned}
 v_{1,2} &= \frac{\sum_{i=1}^{10} x_n}{10} \\
 &= \frac{(-1,7496 + 0,8688 - 0,641 - 0,06762 + 0,4202 - 0,9418 - 0,8616 - 0,1681 - 0,5664 - 2,1533)}{10} \\
 &= \frac{-6,469}{10} \\
 &= -0,6469 \\
 v_{2,2} &= \frac{\sum_{i=1}^{10} x_n}{10} \\
 &= \frac{(0,6903 - 1,8646 + 0,0651 + 0,6830 - 1,9528 - 0,0045 + 1,0946 - 1,1179 - 0,5952 + 1,7227)}{10} \\
 &= \frac{-1,279}{10} \\
 &= -0,1279 \\
 v_{3,2} &= \frac{\sum_{i=1}^{10} x_n}{10} \\
 &= \frac{(-0,271 - 1,0113 - 0,9872 - 0,2963 + 0,3759 + 2,8732 - 0,4705 + 0,2117 - 0,3762 - 0,2145)}{10} \\
 &= \frac{-0,1661}{10} \\
 &= -0,0166
 \end{aligned}$$

Untuk titik pusat baru pada *cluster* 1, karena terdapat satu anggota yaitu objek ke-10 maka objek tersebut akan menjadi *centroid* yang baru. Sehingga diperoleh *centroid* atau titik pusat *cluster* yang baru pada tabel berikut:

**Tabel 13.** *Centroid* Baru

<b>Centroid</b>	<b>V1</b>	<b>V2</b>	<b>V3</b>
C1	6,4689	1,2792	0,1661
C2	-0,6469	-0,1279	-0,0166

Berdasarkan **Tabel 13** diperoleh *centroid* yang baru, kemudian dilakukan perhitungan jarak antara *centroid* baru terhadap objek data yang dapat dilihat pada **Lampiran 9**. Serta dilakukan pengelompokan objek berdasarkan *centroid* baru yang paling dekat. Sehingga diperoleh hasil alokasi objek terhadap *centroid* terdekat pada iterasi kedua adalah sebagai berikut:

**Tabel 14.** Hasil Iterasi Kedua *K-Means*

<b>No.</b>	<b>C1</b>	<b>C2</b>	<b>Kedekatan</b>	<b>Hasil</b>	
				<b>Iterasi 2</b>	<b>Iterasi 1</b>
1	8,251232	1,396528	1,396528	2	2
2	6,529202	2,510539	2,510539	2	2
3	7,304418	0,989595	0,989595	2	2

4	7,184837	0,858364	0,858364	2	2
5	6,861299	2,150134	2,150134	2	2
6	7,99343	2,907455	2,907455	2	2
7	7,360443	1,321633	1,321633	2	2
8	7,056767	1,123177	1,123177	2	2
9	7,300912	0,595113	0,595113	2	2
10	0	7,255909	0	1	1
11	8,641984	2,394409	2,394409	2	2

Berdasarkan **Tabel 14** terlihat bahwa hasil iterasi kedua telah sama dengan iterasi sebelumnya, maka perhitungan telah berhenti. Baik perhitungan manual maupun dengan menggunakan *software R* diperoleh dari 11 objek penelitian dengan 2 *cluster* diperoleh hasil sebagai berikut:

**Tabel 15.** Anggota *Cluster* Algoritma *K-Means*

<b>Cluster</b>	<b>Jumlah Anggota</b>	<b>Wilayah Kabupaten/Kota</b>
1	1	Kota Jambi
2	10	Kerinci, Merangin, Sarolangun, Batanghari, Muaro Jambi, Tanjung Jabung Barat, Tanjung Jabung Timur, Tebo, Bungo, dan Kota Sungai Penuh

Setelah diperoleh hasil pengelompokan maka diperlukan sebuah interpretasi dan memberikan karakteristik di setiap *cluster* yang terbentuk dengan tujuan untuk memberikan gambaran isi dari masing-masing *cluster* yang diperoleh. Adapun untuk menggambarkan isi dari masing-masing *cluster* tersebut dilakukan perhitungan rata-rata di setiap variabel pada *cluster* yang terbentuk. Dalam hal ini, nilai rata-rata di setiap variabel pada *cluster* yang terbentuk sama halnya dengan nilai *centroid* yang terbentuk sehingga nilai rata-rata untuk masing-masing variabel di setiap *cluster* ditunjukkan oleh tabel berikut:

**Tabel 16.** Nilai Rata-Rata Tiap Variabel *Cluster K-Means*

<b>Variabel</b>	<b>Cluster 1</b>	<b>Cluster 2</b>
V1	6,4689	-0,6469
V2	1,2793	-0,1279
V3	0,1661	-0,0166

Berdasarkan **Tabel 16** diperoleh karakteristik *cluster* 1 memiliki 3 variabel yang nilainya di atas rata-rata dan karakteristik *cluster* 2 berbanding terbalik dengan *cluster* 1, nilai setiap variabel pada *cluster* 2 memiliki nilai di bawah rata-rata. Maka berdasarkan analisis karakteristik dan pertimbangan pada masing-masing *cluster* yang telah diperoleh maka dapat disimpulkan bahwa:

- a) *Cluster* 1 merupakan *cluster* dengan penyakit menular yang dikategorikan “tinggi”. Sehingga wilayah kabupaten/kota yang termasuk ke dalam *cluster* 1 merupakan wilayah kabupaten/kota dengan jumlah kasus penyakit menular yang tinggi.

b) *Cluster 2* merupakan *cluster* dengan penyakit menular yang dikategorikan “rendah”. Sehingga wilayah kabupaten/kota yang termasuk ke dalam *cluster 2* merupakan wilayah kabupaten/kota dengan jumlah kasus penyakit menular yang rendah.

Hasil pengelompokan algoritma *k-means* dengan menggunakan *software R* diperoleh hasil sebagai berikut:

*Clustering vector*:

[1] 2 2 2 2 2 2 2 2 2 1 2

Hasil *output* untuk pengelompokan menggunakan *software R* dapat dilihat pada **Lampiran 8**.

#### 4.5 Pengelompokan dengan Algoritma *K-Medoid*

a. *Medoid* awal

Penentuan *centroid* awal sebagai *medoid* awal ditentukan sesuai dengan jumlah *k* yang diambil. Jumlah *k cluster* optimal berdasarkan metode *elbow* sebanyak 2 *cluster* optimal maka jumlah *medoid* awal yang akan dipilih sebanyak 2. **Tabel 15** merupakan tabel *medoid* awal pada perhitungan secara manual yaitu sebagai berikut:

**Tabel 17. Medoid Awal**

<i>Medoid</i>	<b>V1</b>	<b>V2</b>	<b>V3</b>
C1	0,8688	-1,8646	-1,0113
C2	-0,1681	-1,1179	0,2117

Berdasarkan **Tabel 17** C1 adalah *medoid* pertama dengan mengambil objek ke-2 sebagai pusat *clusternya*. Sedangkan C2 adalah *medoid* kedua dengan mengambil objek ke-8 sebagai pusat *clusternya*.

b. Perhitungan jarak antara *medoid* dan objek data

Ukuran jarak antara *medoid* awal yang disimbolkan sebagai C1 dan C2 terhadap objek data menggunakan ukuran jarak *euclidean*. Perhitungan manual jarak antara *medoid* pertama (C1) terhadap objek data adalah sebagai berikut: (selengkapnya pada **Lampiran 10**.)

$$\begin{aligned}
 d_{ij} &= \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \\
 d_{1,c1} &= \sqrt{\sum_{k=1}^3 (x_{1k} - c_{1k})^2} \\
 &= \sqrt{(x_{11} - c_{11})^2 + (x_{12} - c_{12})^2 + (x_{13} - c_{13})^2} \\
 &= \sqrt{(-1,7497 - (0,8688))^2 + (0,6903 - (-1,8646))^2 + ((-0,271) - (-1,0113))^2} \\
 &= \sqrt{(-2,6185)^2 + (2,5548)^2 + (0,7403)^2} \\
 &= 3,7325 \\
 d_{2,c1} &= \sqrt{\sum_{k=1}^3 (x_{2k} - c_{1k})^2} \\
 &= \sqrt{(x_{21} - c_{11})^2 + (x_{22} - c_{12})^2 + (x_{23} - c_{13})^2}
 \end{aligned}$$

$$\begin{aligned}
&= \sqrt{((0,8688) - (0,8688))^2 + ((-1,8646) - (-1,8646))^2 + ((-1,0113) - (-1,0113))^2} \\
&= \sqrt{(0)^2 + (0)^2 + (0)^2} \\
&= 0
\end{aligned}$$

⋮

$$\begin{aligned}
d_{11,c1} &= \sqrt{\sum_{k=1}^3 (x_{11k} - c_{1k})^2} \\
&= \sqrt{(x_{111} - c_{11})^2 + (x_{112} - c_{12})^2 + (x_{113} - c_{13})^2} \\
&= \sqrt{(-2,1533 - (0,8688))^2 + (1,7227 - (-1,8646))^2 + ((-0,2145) - (-1,0113))^2} \\
&= \sqrt{(-3,0221)^2 + (3,5872)^2 + (0,7968)^2} \\
&= 4,7578
\end{aligned}$$

Perhitungan manual jarak antara *medoid* kedua (C2) terhadap objek data adalah sebagai berikut: (selengkapnya pada **Lampiran 10.**)

$$\begin{aligned}
d_{1,c2} &= \sqrt{\sum_{k=1}^3 (x_{1k} - c_{2k})^2} \\
&= \sqrt{(x_{11} - c_{21})^2 + (x_{12} - c_{22})^2 + (x_{13} - c_{23})^2} \\
&= \sqrt{(-1,7497 - (-0,1681))^2 + (0,6903 - (-1,1179))^2 + ((-0,271) - (0,2117))^2} \\
&= \sqrt{(-1,5816)^2 + (1,8082)^2 + (-0,4828)^2} \\
&= 2,4503
\end{aligned}$$

$$\begin{aligned}
d_{2,c2} &= \sqrt{\sum_{k=1}^3 (x_{2k} - c_{2k})^2} \\
&= \sqrt{(x_{21} - c_{21})^2 + (x_{22} - c_{22})^2 + (x_{23} - c_{23})^2} \\
&= \sqrt{((0,8688) - (-0,1681))^2 + ((-1,8646) - (-1,1179))^2 + ((-1,0113) - (0,2117))^2} \\
&= \sqrt{(1,0369)^2 + (-0,7466)^2 + (-1,2231)^2} \\
&= 1,7688
\end{aligned}$$

⋮

$$\begin{aligned}
d_{11,c2} &= \sqrt{\sum_{k=1}^3 (x_{11k} - c_{2k})^2} \\
&= \sqrt{(x_{111} - c_{21})^2 + (x_{112} - c_{22})^2 + (x_{113} - c_{23})^2} \\
&= \sqrt{(-2,1533 - (-0,1681))^2 + (1,7227 - (-1,1179))^2 + ((-0,2145) - (0,2117))^2} \\
&= \sqrt{(-1,9852)^2 + (2,8406)^2 + (-0,4262)^2} \\
&= 3,4916
\end{aligned}$$

c. Mengelompokkan objek *non-medoid* berdasarkan *medoid* yang paling dekat

Setelah melakukan perhitungan jarak antara *medoid* dan objek *non-medoid*, dilakukan pengelompokan objek *non-medoid* berdasarkan *medoid* yang paling dekat serta dihitung total jarak terdekatnya. Tabel berikut merupakan tabel pengelompokan objek *non-medoid* berdasarkan *medoid* yang paling dekat adalah sebagai berikut:

**Tabel 18.** Alokasi Objek Terhadap *Medoid* Terdekat

No.	C1	C2	Kedekatan	Hasil
1	3,7325	2,4503	2,4503	2
2	0	1,7688	0	1
3	2,4502	1,7495	1,7495	2
4	3,0640	1,9390	1,9390	2
5	1,4606	1,0344	1,0344	2
6	4,6720	2,9868	2,9868	2
7	3,4703	2,4169	2,4169	2
8	1,7688	0	0	2
9	2,0185	0,8818	0,8818	2
10	6,5292	7,0567	6,5292	1
11	4,7578	3,4916	3,4916	2
Total			23,4799	

Berdasarkan **Tabel 18** diperoleh hasil pengelompokan objek berdasarkan *medoid* yang paling dekat ke dalam suatu *cluster* dan total jaraknya. Selanjutnya, memilih objek *non-medoid* baru berdasarkan *cluster* yang terbentuk. Adapun objek yang dipilih sebagai *non-medoid* adalah sebagai berikut:

**Tabel 19.** Objek *Non-Medoid*

<i>Non-Medoid</i>	V1	V2	V3
C1	6,4689	1,2792	0,1661
C2	-2,1533	1,7227	-0,2145

Perhitungan dilakukan kembali untuk mengukur jarak antara objek ke setiap *non-medoid* dapat dilihat pada **Lampiran 10**. Sehingga diperoleh hasil iterasi kedua *k-medoid* adalah sebagai berikut:

**Tabel 20.** Hasil Iterasi Kedua *K-Medoid*

No.	C1	C2	Kedekatan	Hasil
1	8,2512	1,1099	1,1099	2
2	6,5292	4,7578	4,7578	2
3	7,3044	2,3731	2,3731	2
4	7,1848	1,8081	1,8081	2
5	6,8612	4,5255	4,5255	2
6	7,9934	3,7396	3,7396	2
7	7,3604	1,4589	1,4589	2
8	7,0567	3,4916	3,4916	2
9	7,3009	2,8137	2,8137	2
10	0	8,6419	0	1
11	8,6419	0	0	2
Total			26,0786	

Berdasarkan **Tabel 20** diperoleh hasil nilai total jarak baru sebesar 26,0786. Selanjutnya menghitung selisih total jarak dari kedua iterasi adalah sebagai berikut:

$$d_{total} = \text{Total Jarak Baru} - \text{Total Jarak Lama}$$



$$= 26,0786 - 23,4799$$

$$= 2,5987$$

Karena  $d_{total} > 0$ , maka iterasi berhenti dan hasil *cluster* mengikuti hasil pada iterasi pertama. Sehingga dapat disimpulkan bahwa dari 11 objek penelitian terbentuk 2 *cluster* dengan anggota untuk masing-masing *cluster* disajikan dalam tabel berikut:

**Tabel 21.** Anggota *Cluster* Algoritma *K-Medoid*

<b>Cluster</b>	<b>Jumlah Anggota</b>	<b>Wilayah Kabupaten/Kota</b>
1	2	Merangin dan Kota Jambi
2	9	Kerinci, Sarolangun, Batanghari, Muaro Jambi, Tanjung Jabung Barat, Tanjung Jabung Timur, Tebo, Bungo, dan Kota Sungai Penuh

Setelah diperoleh hasil pengelompokan maka diperlukan sebuah interpretasi dan memberikan karakteristik di setiap *cluster* yang terbentuk dengan tujuan untuk memberikan gambaran isi dari masing-masing *cluster* yang diperoleh. Adapun untuk menggambarkan isi dari masing-masing *cluster* tersebut dilakukan perhitungan rata-rata disetiap variabel pada *cluster* yang terbentuk. Adapun perhitungan nilai rata-rata untuk hasil *cluster k-medoid* dapat dilihat pada **Lampiran 14**. Berikut adalah tabel nilai rata-rata masing-masing *cluster* yaitu sebagai berikut:

**Tabel 22.** Nilai Rata-Rata Tiap Variabel *Cluster K-Medoid*

<b>Variabel</b>	<b>Cluster 1</b>	<b>Cluster 2</b>
V1	3,6689	-0,8153
V2	-0,2927	0,0650
V3	-0,4226	0,0939

Berdasarkan **Tabel 22** diperoleh karakteristik pada *cluster 1* memiliki 1 variabel yang nilainya di atas rata-rata yaitu variabel 1 dan memiliki 2 variabel di bawah rata-rata yaitu variabel 2 dan variabel 3. Sedangkan, karakteristik pada *cluster 2* memiliki 2 variabel yang nilainya di atas rata-rata yaitu variabel 2 dan variabel 3 serta memiliki 1 variabel yang nilainya di bawah rata-rata yaitu variabel 1. Maka berdasarkan analisis karakteristik dan pertimbangan pada masing-masing *cluster* yang telah diperoleh maka dapat disimpulkan bahwa:

- Cluster 1* merupakan *cluster* dengan penyakit menular yang dikategorikan “rendah”. Sehingga wilayah kabupaten/kota yang termasuk ke dalam *cluster 1* merupakan wilayah kabupaten/kota dengan jumlah kasus penyakit menular yang rendah.
- Cluster 2* merupakan *cluster* dengan penyakit menular yang dikategorikan “tinggi”. Sehingga wilayah kabupaten/kota yang termasuk ke dalam *cluster 2*

merupakan wilayah kabupaten/kota dengan jumlah kasus penyakit menular yang tinggi.

Adapun hasil pengelompokan algoritma *k-medoid* dengan menggunakan *software R* diperoleh hasil sebagai berikut:

*Clustering vector*:

[1] 1 1 1 1 1 1 1 1 1 2 1

Hasil *output* untuk pengelompokan menggunakan *software R* dapat terlihat pada **lampiran 8**. Berdasarkan hasil pengelompokan di atas maka dapat disimpulkan bahwa dari 11 objek penelitian terbentuk 2 *cluster* dengan masing-masing anggota *cluster* disajikan pada tabel berikut:

**Tabel 23.** Anggota Cluster Output R K-Medoid

Cluster	Jumlah Anggota	Wilayah Kabupaten/Kota
1	10	Kerinci, Merangin, Sarolangun, Batanghari, Muaro Jambi, Tanjung Jabung Barat, Tanjung Jabung Timur, Tebo, Bungo, dan Kota Sungai Penuh
2	1	Kota Jambi

Pada **Tabel 23** dapat diketahui bahwa pengelompokan menggunakan algoritma *k-means* dengan jarak *Euclidean* menghasilkan *cluster* 1 yang berjumlah 10 anggota dan *cluster* 2 berjumlah 1 anggota.

#### 4.6 Perbandingan Hasil Pengelompokan Secara Manual dan Output R

Berdasarkan hasil analisis *cluster k-means* dan *k-medoid* yang telah diperoleh secara manual dan *output software R* dapat terlihat pada tabel berikut:

**Tabel 24.** Perbandingan Hasil Perhitungan Manual dan Output R

n	Wilayah Kabupaten / kota	Manual		Output R	
		K-Means	K-Medoid	K-Means	K-Medoid
1	Kerinci	2	2	2	1
2	Merangin	2	1	2	1
3	Sarolangun	2	2	2	1
4	Batang Hari	2	2	2	1
5	Muaro Jambi	2	2	2	1
6	Tanjung Jabung Timur	2	2	2	1
7	Tanjung Jabung Barat	2	2	2	1
8	Tebo	2	2	2	1
9	Bungo	2	2	2	1
10	Kota Jambi	1	1	1	2
11	Kota Sungai Penuh	2	2	2	1

Berdasarkan **Tabel 24** terlihat bahwa hasil pengelompokan dengan menggunakan algoritma *k-means* secara manual dan menggunakan *software R* memiliki hasil yang sama. Sedangkan, hasil pengelompokan dengan menggunakan algoritma *k-medoid* secara manual dan menggunakan *software R*

memiliki perbedaan. Perbedaan hasil yang diperoleh menggunakan algoritma *k-medoid* dikarenakan pada pengambilan *medoid* saat proses pengelompokan yang diambil secara acak/*random*. Sehingga memungkinkan terjadinya perbedaan hasil yang diperoleh antara perhitungan secara manual dan menggunakan *software R*. Pengelompokan menggunakan algoritma *k-medoid* secara manual, *cluster 1* memiliki 2 anggota sedangkan *cluster 1* pada *output software R* memiliki 9 anggota. Anggota yang memiliki perbedaan *cluster* setelah dilakukan perhitungan manual dan *software R* adalah Kerinci, Sarolangun, Batanghari, Muaro Jambi, Tanjung Jabung Barat, Tanjung Jabung Timur, Tebo, Bungo, dan Kota Sungai Penuh.

#### 4.7 Evaluasi Hasil Cluster

Setelah diperoleh hasil pengelompokan maka selanjutnya adalah menghitung nilai *calinski harabasz index* untuk mengevaluasi hasil *cluster* dari algoritma *k-means* dan *k-medoid*. Dengan jumlah objek sebanyak 11 wilayah kabupaten/kota diperoleh jumlah *cluster* optimal menggunakan metode *elbow* untuk algoritma *k-means* dan *k-medoid* masing-masing sebanyak 2 *cluster*. Maka perhitungan *calinski harabasz index* diperoleh sebagai berikut:

$$\begin{aligned} CH_K &= \frac{SSB}{SSW} \times \frac{N - K}{K - 1} \\ CH_2 &= \frac{SSB}{SSW} \times \frac{11 - 2}{2 - 1} \\ &= \frac{SSB}{SSW} \times \frac{9}{1} \\ &= \frac{SSB}{SSW} \times 9 \end{aligned}$$

Adapun perhitungan nilai *calinski harabasz index* pada setiap hasil algoritma *k-means* dan *k-medoid* adalah sebagai berikut:

a. Nilai *calinski harabasz index k-means*

Pada perhitungan manual dan *software R* diperoleh hasil *cluster* yang sama, sehingga akan menggunakan *centroid* atau pusat *cluster* yang sama untuk mendapatkan nilai CHI. Adapun *centroid* perhitungan manual dapat dilihat pada **Tabel 11** dan *centroid k-means* dari *output R* dapat dilihat pada **Lampiran 8**. Selanjutnya menghitung nilai SSW dan SSB dari hasil pengelompokan menggunakan algoritma *k-means*.

Sebelum menghitung SSW diperlukan terlebih dahulu untuk menghitung kuadrat jarak antar pusat *cluster* ke anggota-anggotanya, Perhitungan manual kuadrat jarak antara *centroid* pertama terhadap objek data adalah sebagai berikut: (selengkapnya pada **Lampiran 10**.)

$$d^2(m_i, x) = \left( \sqrt{\sum_{k=1}^k (m_{ik} - x_{ik})^2} \right)^2$$

$$\begin{aligned} d^2(m_1, x_1) &= \left( \sqrt{(m_{11} - x_{11})^2 + (m_{12} - x_{12})^2 + (m_{13} - x_{13})^2} \right)^2 \\ &= \left( \sqrt{(6,4689 - (-1,7496))^2 + (1,2792 - 0,6903)^2 + (0,1661 - (-0,2709))^2} \right)^2 \\ &= \left( \sqrt{(8,2185)^2 + (0,5889)^2 + (0,4371)^2} \right)^2 \\ &= 68,0828 \end{aligned}$$

$$\begin{aligned} d^2(m_1, x_2) &= \left( \sqrt{(m_{11} - x_{21})^2 + (m_{12} - x_{22})^2 + (m_{13} - x_{23})^2} \right)^2 \\ &= \left( \sqrt{(6,4689 - 0,8688)^2 + (1,2792 - (-1,8645))^2 + (0,1661 - (-1,0113))^2} \right)^2 \\ &= \left( \sqrt{(5,6000)^2 + (3,1438)^2 + (1,1774)^2} \right)^2 \\ &= 42,6305 \end{aligned}$$

⋮

$$\begin{aligned} d^2(m_1, x_{11}) &= \left( \sqrt{(m_{11} - x_{111})^2 + (m_{12} - x_{112})^2 + (m_{13} - x_{113})^2} \right)^2 \\ &= \left( \sqrt{(6,4689 - (-2,1532))^2 + (1,2792 - 1,7227)^2 + (0,1661 - (-0,2144))^2} \right)^2 \\ &= \left( \sqrt{(8,6222)^2 + (-0,4434)^2 + (0,3806)^2} \right)^2 \\ &= 74,6839 \end{aligned}$$

Perhitungan manual kuadrat jarak antara *centroid* kedua terhadap objek data adalah sebagai berikut: (selengkapnya pada **Lampiran 11.**)

$$\begin{aligned} d^2(m_2, x_1) &= \left( \sqrt{(m_{21} - x_{11})^2 + (m_{22} - x_{12})^2 + (m_{23} - x_{13})^2} \right)^2 \\ &= \left( \sqrt{(-0,6468 - (-1,7496))^2 + (-0,1279 - 0,6903)^2 + (-0,0166 - (-0,2709))^2} \right)^2 \\ &= \left( \sqrt{(1,1027)^2 + (-0,8182)^2 + (0,2543)^2} \right)^2 \\ &= 1,9502 \end{aligned}$$

$$\begin{aligned} d^2(m_2, x_2) &= \left( \sqrt{(m_{21} - x_{21})^2 + (m_{22} - x_{22})^2 + (m_{23} - x_{23})^2} \right)^2 \\ &= \left( \sqrt{(-0,6468 - 0,8688)^2 + (-0,1279 - (-1,8645))^2 + (-0,0166 - (-1,0113))^2} \right)^2 \\ &= \left( \sqrt{(-1,5157)^2 + (1,7366)^2 + (0,9947)^2} \right)^2 \\ &= 6,3028 \end{aligned}$$

⋮

$$\begin{aligned} d^2(m_2, x_{11}) &= \left( \sqrt{(m_{21} - x_{111})^2 + (m_{22} - x_{112})^2 + (m_{23} - x_{113})^2} \right)^2 \\ &= \left( \sqrt{(-0,6468 - (-2,1532))^2 + (-0,1279 - 1,7227)^2 + (0,0166 - (-0,2144))^2} \right)^2 \\ &= \left( \sqrt{(1,5063)^2 + (-1,8506)^2 + (0,1978)^2} \right)^2 \\ &= 5,7332 \end{aligned}$$

Setelah perhitungan kuadrat jarak antar *centroid* ke anggota-anggotanya, maka diperoleh total SSWnya dapat dilihat pada tabel berikut: (selengkapnya pada **Lampiran 11.**)

**Tabel 25.** Hasil Perhitungan Kuadrat Jarak CHI *K-Means*

No.	$d^2(m_1, x)$	$d^2(m_2, x)$
1	68,0828	1,9503
2	42,6305	6,3028
3	53,3545	0,9793
4	51,6219	0,7368
5	47,0774	4,6231
6	63,8949	8,4533
7	54,1761	1,7467
8	49,7980	1,2615
9	53,3033	0,3542
10	0,0000	52,6482
11	74,6839	5,7332
Total	558,6233	84,7893

Berdasarkan **Tabel 25** maka dapat diperoleh nilai SSWnya sebagai berikut:

$$\begin{aligned}
 SSW &= \sum_{i=1}^K \sum_{x \in C_i} d^2(m_i, x) \\
 &= 558,6233 + 84,7893 \\
 &= 643,4126
 \end{aligned}$$

Selanjutnya menghitung nilai SSB, diperlukan terlebih dahulu menghitung kuadrat jarak antar *centroid* ke anggota-anggotanya dan  $M$  merupakan *centroid global* atau rata-rata keseluruhan objek data pada suatu variabel hasil perhitungan nilai  $M$  setiap variabel dapat dilihat pada **Lampiran 12.**

Perhitungan manual kuadrat jarak antara *centroid* pertama (C1) terhadap objek data adalah sebagai berikut:

$$\begin{aligned}
 d^2(m_1, M) &= \left( \sqrt{(m_{11} - M_1)^2 + (m_{12} - M_2)^2 + (m_{13} - M_3)^2} \right)^2 \\
 &= \left( \sqrt{(6,4689 - 0,000036)^2 + (1,2792 - (-0,000018))^2 + (0,1661 - 0,000018)^2} \right)^2 \\
 &= \left( \sqrt{(6,468864)^2 + (1,279218)^2 + (0,166082)^2} \right)^2 \\
 &= 43,5105 \\
 d^2(m_2, M) &= \left( \sqrt{(m_{21} - M_1)^2 + (m_{22} - M_2)^2 + (m_{23} - M_3)^2} \right)^2 \\
 &= \left( \sqrt{(-0,6468 - 0,000036)^2 + (-0,12792 - (-0,000018))^2 + (-0,0166 - 0,000018)^2} \right)^2 \\
 &= \left( \sqrt{(-0,646836)^2 + (-0,127902)^2 + (-0,016618)^2} \right)^2 \\
 &= 0,4351
 \end{aligned}$$

$$n_1 = 1 \text{ dan } n_2 = 10$$

$$\begin{aligned} SSB &= n_1 d^2(m_1, M) + n_2 d^2(m_2, M) \\ &= 1 * (43,5105) + 10 * (0,4351) \\ &= 47,8615 \end{aligned}$$

$$\begin{aligned} CH_2 &= \frac{SSB}{SSW} \times 9 \\ &= \frac{47,8615}{643,4126} \times 9 \\ &= 0,6694 \end{aligned}$$

b. Nilai *calinski harabasz index k-medoid*

Pada perhitungan manual dan *software R* diperoleh hasil *cluster* yang berbeda, sehingga akan menggunakan *medoid* yang berbeda untuk mendapatkan nilai CHI. Adapun *medoid* perhitungan manual dapat dilihat pada **Tabel 17** dan *medoid* dari *output R* dapat dilihat pada **Lampiran 8**. Selanjutnya menghitung nilai SSW dan SSB dari hasil pengelompokan menggunakan algoritma *k-medoid*.

Menghitung kuadrat jarak antar *medoid* ke anggota-anggotanya, Perhitungan manual jarak antara *medoid* pertama terhadap objek data adalah sebagai berikut: (selengkapnya pada **Lampiran 13**.)

$$d^2(m_i, x) = \left( \sqrt{\sum_{k=1}^k (m_{ik} - x_{ik})^2} \right)^2$$

$$\begin{aligned} d^2(m_1, x_1) &= \left( \sqrt{(m_{11} - x_{11})^2 + (m_{12} - x_{12})^2 + (m_{13} - x_{13})^2} \right)^2 \\ &= \left( \sqrt{(0,8688 - (-1,7496))^2 + (-1,8645 - 0,6903)^2 + (-1,0113 - (-0,2709))^2} \right)^2 \\ &= \left( \sqrt{(2,6185)^2 + (-2,5548)^2 + (-0,7403)^2} \right)^2 \\ &= 13,9319 \end{aligned}$$

$$\begin{aligned} d^2(m_1, x_2) &= \left( \sqrt{(m_{11} - x_{21})^2 + (m_{12} - x_{22})^2 + (m_{13} - x_{23})^2} \right)^2 \\ &= \left( \sqrt{(0,8688 - 0,8688)^2 + (-1,8645 - (-1,8645))^2 + ((-1,0113) - (-1,0113))^2} \right)^2 \\ &= \left( \sqrt{(0)^2 + (0)^2 + (0)^2} \right)^2 \\ &= 0 \end{aligned}$$

⋮

$$\begin{aligned} d^2(m_1, x_{11}) &= \left( \sqrt{(m_{11} - x_{111})^2 + (m_{12} - x_{112})^2 + (m_{13} - x_{113})^2} \right)^2 \\ &= \left( \sqrt{(0,8688 - (-2,1532))^2 + (-1,8645 - 1,7227)^2 + (-0,2144 - (-1,0113))^2} \right)^2 \\ &= \left( \sqrt{(3,0221)^2 + (-3,5872)^2 + (0,7968)^2} \right)^2 \\ &= 22,6367 \end{aligned}$$

Perhitungan manual kuadrat jarak antara *medoid* kedua terhadap objek data adalah sebagai berikut: (selengkapnya pada **Lampiran 13**.)

$$d^2(m_2, x_1) = \left( \sqrt{(m_{21} - x_{11})^2 + (m_{22} - x_{12})^2 + (m_{23} - x_{13})^2} \right)^2$$

$$\begin{aligned}
&= (\sqrt{(-0,1680 - (-1,74960))^2 + (-1,1179 - 0,6903)^2 + (0,2117 - (-0,2709))^2})^2 \\
&= (\sqrt{(1,5815)^2 + (-1,8082)^2 + (0,4827)^2})^2 \\
&= 3,5940
\end{aligned}$$

$$\begin{aligned}
d^2(m_2, x_2) &= (\sqrt{(m_{21} - x_{21})^2 + (m_{22} - x_{22})^2 + (m_{23} - x_{23})^2})^2 \\
&= (\sqrt{(-0,1680 - 0,8688)^2 + (-1,1179 - (-1,8645))^2 + (0,2117 - (-1,0113))^2})^2 \\
&= (\sqrt{(-1,0369)^2 + (0,7466)^2 + (1,2231)^2})^2 \\
&= 3,1286
\end{aligned}$$

⋮

$$\begin{aligned}
d^2(m_2, x_{11}) &= (\sqrt{(m_{21} - x_{111})^2 + (m_{22} - x_{112})^2 + (m_{23} - x_{113})^2})^2 \\
&= (\sqrt{(-0,1680 - (-2,1532))^2 + (-1,1179 - 1,7227)^2 + (0,2117 - (-0,2144))^2})^2 \\
&= (\sqrt{(1,9852)^2 + (-2,8406)^2 + (0,4262)^2})^2 \\
&= 12,1919
\end{aligned}$$

Setelah perhitungan kuadrat jarak antar *medoid* ke anggota-anggotanya, maka diperoleh total SSWnya dapat dilihat pada tabel berikut: (selengkapnya pada **Lampiran 13.**)

**Tabel 26.** Hasil Perhitungan Kuadrat Jarak CHI *K-Medoid*

No.	$d^2(m_1, x)$	$d^2(m_2, x)$
1	13,9319	3,5940
2	0	3,1286
3	6,0038	3,0607
4	9,3886	3,7598
5	2,1334	1,0700
6	21,8277	8,9214
7	12,0435	5,8418
8	3,1286	0
9	4,0745	0,7776
10	42,6304	49,7979
11	22,6367	12,1919
Total	137,7997	92,1442

Berdasarkan **Tabel 26** maka dapat diperoleh nilai SSWnya sebagai berikut:

$$\begin{aligned}
SSW &= \sum_{i=1}^K \sum_{x \in c_i} d^2(m_i, x) \\
&= 137,7997 + 92,1442 \\
&= 229,939
\end{aligned}$$

Selanjutnya menghitung nilai SSB, diperlukan terlebih dahulu menghitung kuadrat jarak antar *centroid* ke anggota-anggotanya dan  $M$

merupakan *centroid global* atau rata-rata keseluruhan objek data pada suatu variabel hasil perhitungan nilai  $M$  setiap variabel dapat dilihat pada **Lampiran 12**.

Perhitungan manual kuadrat jarak antara *centroid* pertama (C1) terhadap objek data adalah sebagai berikut:

$$\begin{aligned} d^2(m_1, M) &= (\sqrt{(m_{11} - M_1)^2 + (m_{12} - M_2)^2 + (m_{13} - M_3)^2})^2 \\ &= (\sqrt{(0,8688 - 0,000036)^2 + (-1,8645 - (-0,000018))^2 + (-1,0113 - 0,000018)^2})^2 \\ &= (\sqrt{(0,868764)^2 + (-1,864482)^2 + (-1,011318)^2})^2 \\ &= 5,2542 \end{aligned}$$

$$\begin{aligned} d^2(m_2, M) &= (\sqrt{(m_{21} - M_1)^2 + (m_{22} - M_2)^2 + (m_{23} - M_3)^2})^2 \\ &= (\sqrt{(-0,1680 - 0,000036)^2 + (-1,1179 - (-0,000018))^2 + (0,2117 - 0,000018)^2})^2 \\ &= (\sqrt{(-0,168036)^2 + (-1,117882)^2 + (0,211682)^2})^2 \\ &= 1,3228 \end{aligned}$$

$$n_1 = 2 \text{ dan } n_2 = 9$$

$$\begin{aligned} SSB &= n_1 d^2(m_1, M) + n_2 d^2(m_2, M) \\ &= 2 * (5,2542) + 9 * (1,3228) \\ &= 22,4139 \end{aligned}$$

$$\begin{aligned} CH_2 &= \frac{SSB}{SSW} \times 9 \\ &= \frac{22,4139}{229,9440} \times 9 \\ &= 0,8772 \end{aligned}$$

Selanjutnya perhitungan nilai *calinski harabsz index k-medoid* dengan menggunakan hasil *medoid* dari *software R* dengan perhitungan dapat dilihat pada **Lampiran 13**. Diperoleh nilai CHI sebesar 0,7181. Berdasarkan perhitungan diperoleh nilai *calinski harabsz index* pada masing-masing metode dari hasil *cluster* manual dan hasil *output R* dapat dilihat pada tabel berikut:

**Tabel 27.** Hasil Evaluasi Cluster

<b>Calinski Harabsz Index</b>			
No.	Algoritma	Cluster Manual	Cluster Output R
1	<i>K-Means</i>	0,6694	0,6694
2	<i>K-Medoid</i>	0,8772	0,7181

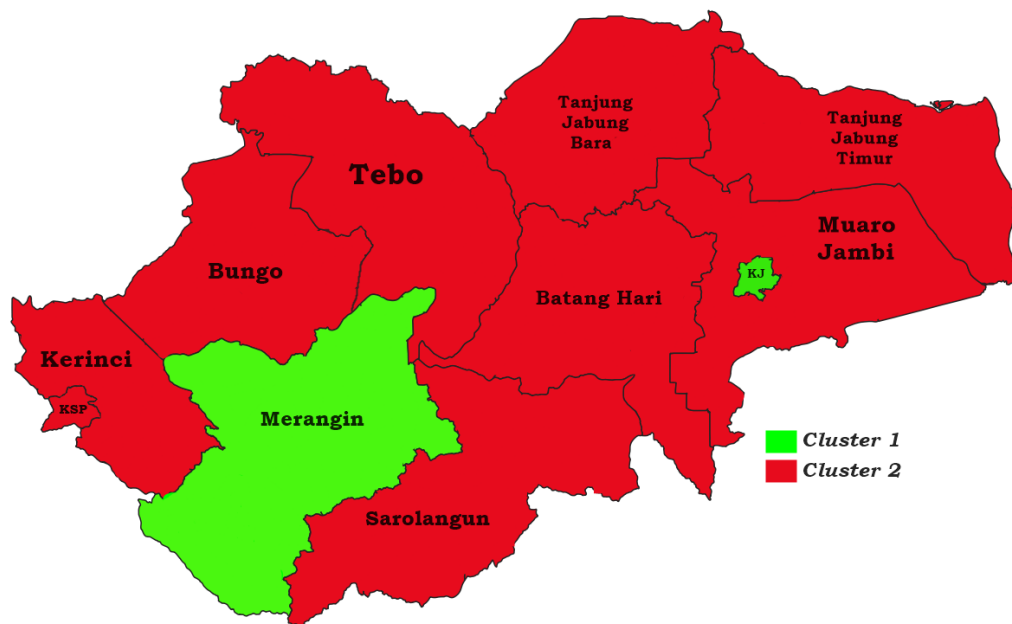
Berdasarkan **Tabel 27** terlihat bahwa hasil evaluasi *cluster* menggunakan *calinski harabsz index* diperoleh nilai tertinggi pada hasil *cluster* manual dengan algoritma *k-medoid* yaitu sebesar 0,8772. Berdasarkan dari itu maka dapat



disimpulkan bahwa hasil pengelompokan terbaik adalah hasil *cluster* manual dengan menggunakan algoritma *k-medoid*.

#### 4.8 Peta Penyebaran *Cluster*

Berdasarkan hasil evaluasi *cluster* menggunakan *calinski harabasz index* diperoleh hasil terbaik adalah hasil algoritma *k-medoid* dengan perhitungan manual. Maka daripada itu peta penyebaran yang terbentuk adalah sebagai berikut:



**Gambar 14.** Peta Penyebaran *Cluster K-Medoid*

Berdasarkan peta penyebaran tersebut dapat terlihat bahwa kabupaten/kota yang termasuk ke dalam *cluster* 1 ditunjukkan oleh warna hijau yang menunjukkan jumlah kasus penyakit menular yang rendah dan kabupaten/kota yang termasuk ke dalam *cluster* 2 ditunjukkan oleh warna merah yang menunjukkan jumlah kasus penyakit menular yang tinggi. Adapun anggota *cluster* 1 yaitu Kota Jambi dan Merangin. Kemudian, anggota *cluster* 2 yaitu Kerinci, Sarolangun, Batanghari, Muaro Jambi, Tanjung Jabung Barat, Tanjung Jabung Timur, Tebo, Bungo, dan Kota Sungai Penuh.