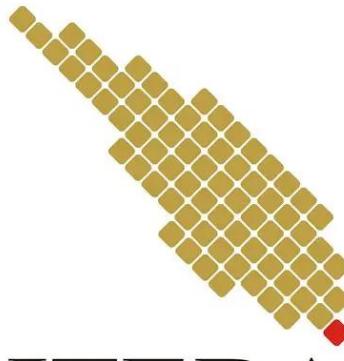


**Implementasi K-Means S4 untuk Evaluasi  
Kapasitas Kesehatan dan Risiko Penyakit di Sumbagsel**



**ITERA**

DISUSUN OLEH KELOMPOK 2:

Disusun Oleh:

Kelompok 2 Kelas RC

Donna Maya Puspita (123450028)

Labo Napitupulu (123450037)

Aprilia Dewi Hutapea (123450040)

**PROGRAM STUDI SAINS DATA  
FAKULTAS SAINS  
INSTITUT TEKNOLOGI SUMATERA  
2025**

## **Abstrak**

Penelitian ini bertujuan mengembangkan fungsi algoritma K-Means berbasis sistem objek S4 dalam bahasa R serta menerapkannya pada pemetaan profil risiko penyakit menular dan evaluasi kapasitas fasilitas kesehatan di Provinsi Lampung dan Sumatera Selatan. Pengembangan sistem S4 dilakukan untuk menghasilkan struktur fungsi yang lebih formal, konsisten, dan mudah diperluas, sehingga proses klasterisasi dapat dijalankan secara terstandarisasi dan terintegrasi dengan data epidemiologi dan infrastruktur kesehatan. Penelitian menggunakan data sekunder kabupaten/kota yang mencakup indikator beban penyakit TBC, HIV, Kusta, Malaria, dan DBD serta data fasilitas kesehatan seperti rumah sakit, puskesmas, klinik, dan posyandu. Setelah melalui proses preprocessing, standardisasi, dan eksekusi algoritma, terbentuk tiga klaster wilayah dengan karakteristik risiko yang berbeda. Klaster 2 muncul sebagai zona krisis DBD dengan insidensi tertinggi namun fasilitas primer yang terbatas; Klaster 3 menjadi episentrum TBC dan HIV dengan kapasitas layanan kesehatan melimpah; sedangkan Klaster 1 menunjukkan wilayah dengan risiko moderat dan fasilitas relatif memadai. Temuan ini mengungkap adanya ketidakseimbangan antara beban penyakit dan kapasitas layanan kesehatan di beberapa wilayah, sehingga analisis klaster dapat menjadi dasar penting bagi perencanaan intervensi kesehatan masyarakat yang lebih tepat sasaran. Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan S4 tidak hanya meningkatkan kualitas pengembangan fungsi analitik, tetapi juga memberikan nilai praktis dalam pemetaan risiko epidemiologis lintas wilayah.

## **BAB I**

### **PENDAHULUAN**

#### 1.1 Latar Belakang

Penyakit menular hingga saat ini masih menjadi salah satu isu strategis dalam pembangunan kesehatan nasional. Penyebarannya tidak hanya dipengaruhi oleh faktor biologis, tetapi juga oleh kondisi sosial, demografis, ekonomi, dan lingkungan yang berbeda-beda di setiap daerah. Provinsi Lampung dan Sumatera Selatan merupakan dua wilayah dengan karakteristik geografis dan mobilitas penduduk yang beragam, sehingga menyebabkan variasi tingkat kerentanan terhadap penyakit menular. Perbedaan kondisi ini menuntut adanya analisis risiko yang dapat menggambarkan pola persebaran penyakit secara lebih komprehensif agar intervensi kesehatan dapat direncanakan dengan efektif dan tepat sasaran.

Dalam beberapa tahun terakhir, pendekatan berbasis data menjadi semakin penting dalam mendukung kebijakan kesehatan masyarakat. Metode *clustering* seperti algoritma K-Means merupakan salah satu teknik yang sering digunakan dalam kajian epidemiologi untuk mengelompokkan wilayah berdasarkan karakteristik tertentu, termasuk tingkat risiko penyakit menular. Namun, penerapan metode ini sering kali menghadapi kendala dalam hal struktur kode, fleksibilitas, dan kemampuan untuk dikembangkan lebih lanjut. Bahasa pemrograman R menyediakan sistem objek S4 yang memungkinkan pengembangan fungsi dengan struktur yang lebih formal dan ketat, sehingga proses analisis dapat dilakukan secara konsisten, terstandarisasi, dan mudah direproduksi.

Pengembangan fungsi K-Means berbasis sistem objek S4 memiliki potensi besar untuk meningkatkan kualitas analisis epidemiologi. Dengan memanfaatkan struktur objek yang lebih terorganisasi, fungsi tersebut dapat dirancang untuk memiliki validasi input yang lebih baik, dokumentasi internal yang jelas, serta kemampuan untuk diintegrasikan dengan analisis spasial atau indikator kesehatan lainnya. Pendekatan ini sangat relevan ketika digunakan untuk memetakan tingkat risiko penyakit menular di berbagai wilayah dan menghubungkannya dengan kapasitas fasilitas kesehatan yang tersedia.

Evaluasi kapasitas fasilitas kesehatan merupakan aspek penting dalam manajemen pelayanan kesehatan, terutama ketika dikaitkan dengan tingkat risiko suatu wilayah. Ketidaksesuaian antara tingkat kebutuhan dan ketersediaan fasilitas dapat berdampak pada lambatnya penanganan kasus, meningkatnya risiko penularan, serta rendahnya kualitas layanan kesehatan. Oleh karena itu, penelitian yang menggabungkan pengembangan metode analitik berbasis pemrograman dengan aplikasi nyata di bidang kesehatan masyarakat menjadi sangat dibutuhkan. Kajian ini diharapkan dapat memberikan kontribusi baik dari sisi metodologis maupun dari sisi penerapan kebijakan kesehatan daerah.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, penelitian ini disusun untuk menjawab beberapa pertanyaan utama sebagai berikut:

1. Bagaimana merancang dan mengembangkan fungsi algoritma K-Means dengan menggunakan sistem objek S4 pada bahasa Rc supaya memiliki struktur yang lebih mudah dikembangkan?
2. Bagaimana fungsi yang dikembangkan tersebut dapat diterapkan untuk melakukan pemetaan profil risiko penyakit menular di Provinsi Lampung dan Sumatera Selatan?
3. Bagaimana hasil pemetaan tingkat risiko tersebut dapat digunakan untuk menilai kesesuaian antara risiko wilayah dan kapasitas fasilitas kesehatan yang tersedia?

## 1.3 Tujuan Peneltian

Penelitian ini memiliki beberapa tujuan yang ingin dicapai, yaitu:

1. Mengembangkan fungsi algoritma K-Means dalam bahasa R dengan memanfaatkan sistem objek S4 untuk menghasilkan fungsi yang terstruktur dan mudah diintegrasikan dengan analisis lanjutan.

2. Menghasilkan pemetaan profil risiko penyakit menular di Provinsi Lampung dan Sumatera Selatan berdasarkan hasil pengelompokan wilayah menggunakan fungsi K-Means yang dikembangkan
3. Melakukan evaluasi terhadap kecukupan kapasitas fasilitas kesehatan pada setiap klaster risiko untuk mengetahui apakah ketersediaan layanan sudah sejalan dengan kebutuhan wilayah.

#### 1.4 Manfaat Penelitian

##### a. Manfaat Praktis

1. Memberikan alat analisis yang dapat dimanfaatkan instansi kesehatan, pemerintah daerah, dan lembaga penelitian untuk memetakan risiko penyakit menular secara lebih sistematis.
2. Menyediakan informasi yang membantu proses perencanaan dan pengambilan keputusan, terutama dalam menilai kecukupan kapasitas fasilitas kesehatan dibandingkan tingkat risiko wilayah.
3. Mendukung penyusunan kebijakan kesehatan daerah, termasuk penentuan prioritas wilayah intervensi serta perencanaan kebutuhan sumber daya dan layanan kesehatan.

##### b. Manfaat Akademis

1. Memberikan kontribusi pada pengembangan metode analisis data, khususnya terkait penerapan sistem objek S4 di R untuk algoritma *clustering*.
2. Menjadi referensi bagi peneliti atau akademisi yang mengkaji hubungan antara *clustering*, analisis spasial, dan perencanaan pelayanan kesehatan.
3. Menjadi dasar bagi studi lanjutan yang membutuhkan fungsi analitik yang lebih terstruktur dan fleksibel untuk berbagai aplikasi di bidang epidemiologi dan kesehatan masyarakat.

## **BAB II**

### **LANDASAN TEORI**

#### 2.1 Algoritma K-Means

Clustering merupakan salah satu teknik analisis data yang digunakan untuk mengelompokkan objek berdasarkan kemiripan karakteristik di antara mereka. Dalam penelitian modern, teknik ini memegang peranan penting karena mampu membantu peneliti memahami struktur alami yang tersembunyi dalam data. Melalui proses pengelompokan, objek-objek yang memiliki sifat serupa dapat ditempatkan dalam kelompok yang sama, sehingga hubungan antar data menjadi lebih mudah dipahami. Langkah ini juga sering digunakan sebagai tahap awal sebelum dilakukan analisis lanjutan, karena dapat menyederhanakan kompleksitas data dan memberikan gambaran awal mengenai pola-pola yang mungkin muncul dalam suatu dataset (Wahyudi dkk., 2020).

Metode K-Means merupakan salah satu pendekatan clustering non-hierarki yang paling populer dan banyak dimanfaatkan dalam berbagai bidang. Metode ini mengelompokkan data ke dalam sejumlah cluster yang telah ditentukan sebelumnya. Setiap objek hanya akan menjadi anggota satu cluster, sehingga pengelompokan bersifat tegas dan tidak tumpang tindih. K-Means dikenal sebagai metode yang efisien karena proses perhitungannya relatif sederhana, namun tetap mampu menghasilkan pembagian kelompok yang cukup akurat. Keunggulan ini menjadikan K-Means pilihan yang banyak digunakan dalam analisis perilaku, segmentasi pengguna, hingga eksplorasi data dalam jumlah besar (Nur Khomarudin, 2003).

Tahapan awal dalam algoritma K-Means adalah menentukan jumlah cluster yang ingin dibentuk. Setelah jumlah tersebut ditetapkan, algoritma memilih sejumlah titik awal sebagai pusat cluster atau centroid secara acak. Selanjutnya, setiap objek dihitung jaraknya terhadap seluruh pusat cluster. Objek kemudian dikelompokkan ke dalam cluster yang memiliki pusat paling dekat. Salah satu metode yang paling umum digunakan untuk menghitung jarak antar objek adalah Euclidean Distance, yang

mengukur jarak geometris antar dua titik dalam ruang multidimensi. Rumus Euclidean Distance dinyatakan sebagai berikut:

$$d(x, y) = \sqrt{\sum (x_i - y_i)^2}$$

$d(x,y)$  : jarak atau ukuran ketidakmiripan antara dua objek

$x_i$  : nilai variable ke-i pada objek x

$y_i$  : nilai variable ke-i pada objek y

Setelah semua objek dialokasikan, pusat cluster diperbarui dengan menghitung rata-rata nilai atribut dari seluruh anggota cluster. Pembaruan ini dilakukan untuk memastikan bahwa pusat cluster dapat mewakili kondisi terbaru dari data yang berada di dalamnya. Proses penghitungan ulang jarak dan pembaruan pusat cluster dilakukan secara berulang hingga mencapai kondisi stabil, yaitu ketika pusat cluster tidak lagi mengalami perubahan yang signifikan. Dengan proses iteratif ini, algoritma K-Means mampu menghasilkan kelompok yang memiliki tingkat kemiripan tinggi di dalamnya, serta perbedaan yang jelas antar cluster.

Secara keseluruhan, K-Means menawarkan pendekatan yang terstruktur dan efektif dalam memahami pola pengelompokan dalam data. Sifatnya yang fleksibel, efisien, dan mudah diterapkan membuat metode ini menjadi alat analisis yang relevan dalam berbagai kebutuhan penelitian. Hasil yang diperoleh dari proses clustering ini dapat digunakan untuk mendukung pengambilan keputusan, memperkuat analisis lanjutan, maupun mengidentifikasi kecenderungan tertentu dalam dataset yang diteliti.

## 2.2 Sistem Objek S4 dalam R

Sistem objek S4 dalam bahasa R merupakan salah satu pendekatan pemrograman berorientasi objek yang dirancang untuk memberikan kerangka kerja yang lebih formal, ketat, dan terstandarisasi dibandingkan sistem objek R lainnya seperti S3. Keberadaan sistem S4 sangat penting terutama dalam penelitian ilmiah, pengembangan perangkat analitik, serta implementasi algoritma yang membutuhkan ketelitian struktural dan konsistensi definisi objek. Dalam konteks pengembangan fungsi analitis, khususnya pada algoritma K-Means untuk pemetaan risiko penyakit

menular, S4 menjadi fondasi yang memungkinkan peneliti membangun model yang lebih terorganisir, dapat diuji, dan mudah dipelihara (Mccaffrey, 2017).

Pemrograman berorientasi objek (Object-Oriented Programming/OOP) pada R melalui sistem S4 merupakan pendekatan yang dirancang untuk memberikan struktur yang lebih ketat, eksplisit, dan terkontrol dalam pengembangan program. Berbeda dari sistem S3 yang bersifat lebih fleksibel namun informal, S4 menawarkan mekanisme formal untuk mendefinisikan kelas, menetapkan atribut (slot), serta mengatur validasi struktur objek. Pendekatan ini memungkinkan pengembang untuk membangun model data yang lebih konsisten dan mudah dipelihara, terutama pada proyek-proyek berskala besar atau penelitian yang menuntut integritas data yang tinggi(Mccaffrey, 2017).

Dalam sistem S4, sebuah kelas didefinisikan menggunakan fungsi `setClass()`. Di dalamnya, setiap komponen objek ditetapkan sebagai *slot* dengan tipe data tertentu. Penetapan tipe ini menjadi salah satu keunggulan utama S4, karena setiap objek yang dibuat harus mematuhi struktur yang telah dibakukan. Dengan demikian, potensi kesalahan akibat ketidaksesuaian tipe dapat diminimalkan sejak tahap awal. Selain itu, sistem S4 menyediakan mekanisme pemeriksaan validitas objek melalui fungsi `validity`, yang dapat digunakan untuk memastikan bahwa nilai setiap slot memenuhi aturan tertentu sebelum objek dianggap sah (Faisal, 2016).

Proses pembuatan objek pada sistem S4 dilakukan melalui konstruktor generik, biasanya menggunakan fungsi `new()`. Fungsi ini memanggil definisi kelas yang telah ditetapkan sebelumnya, kemudian menghasilkan objek yang memenuhi seluruh persyaratan strukturalnya. Pada tahap ini, validasi otomatis akan dijalankan sehingga setiap objek yang terbentuk benar-benar berada dalam kondisi yang sesuai dengan desain kelas.

S4 juga mendukung konsep pewarisan (inheritance), di mana suatu kelas dapat dibangun berdasarkan kelas lain. Hal ini memberikan fleksibilitas dalam memperluas fungsionalitas tanpa harus menulis ulang seluruh struktur data. Selain itu, sistem S4 menerapkan metode generik melalui `setGeneric()` dan `setMethod()`, yang memungkinkan pengembang mendefinisikan fungsi dengan perilaku berbeda bergantung pada kelas objek yang diterimanya. Dengan cara ini, OOP pada S4 tetap

mempertahankan prinsip *polymorphism* sebagaimana yang terdapat pada paradigma OOP secara umum.

Secara keseluruhan, penggunaan OOP S4 dalam R memberikan kerangka kerja yang lebih terstruktur, terutama untuk aplikasi yang memerlukan ketelitian tinggi dalam pengelolaan data. Dengan adanya definisi kelas yang eksplisit, tipe data yang terkontrol, validasi objek, dan pewarisan antar kelas, S4 menjadi pilihan yang tepat bagi penelitian dan pengembangan perangkat analitis yang membutuhkan konsistensi dan kejelasan dalam representasi objek.

Sistem S4 didasarkan pada prinsip-prinsip fundamental dalam pemrograman berorientasi objek, seperti enkapsulasi, abstraksi, pewarisan, dan polimorfisme. Namun, S4 memberikan formalitas tambahan dalam proses definisi class, struktur data, serta pemetaan fungsi generik terhadap objek tertentu.

a. Class dan Slot

Sistem S4 didasarkan pada prinsip-prinsip fundamental dalam pemrograman berorientasi objek, seperti enkapsulasi, abstraksi, pewarisan, dan polimorfisme. Namun, S4 memberikan formalitas tambahan dalam proses definisi class, struktur data, serta pemetaan fungsi generik terhadap objek tertentu.

b. Method dan Generic Function

Sistem S4 didasarkan pada prinsip-prinsip fundamental dalam pemrograman berorientasi objek, seperti enkapsulasi, abstraksi, pewarisan, dan polimorfisme. Namun, S4 memberikan formalitas tambahan dalam proses definisi class, struktur data, serta pemetaan fungsi generik terhadap objek tertentu.

c. Enkapsulasi

Sistem S4 didasarkan pada prinsip-prinsip fundamental dalam pemrograman berorientasi objek, seperti enkapsulasi, abstraksi, pewarisan, dan polimorfisme. Namun, S4 memberikan formalitas tambahan dalam proses definisi class, struktur data, serta pemetaan fungsi generik terhadap objek tertentu.

d. Inheritance

Sistem S4 didasarkan pada prinsip-prinsip fundamental dalam pemrograman berorientasi objek, seperti enkapsulasi, abstraksi, pewarisan, dan polimorfisme. Namun, S4 memberikan formalitas tambahan dalam proses definisi class, struktur data, serta pemetaan fungsi generik terhadap objek tertentu.

e. Polimorfisme

Sistem S4 didasarkan pada prinsip-prinsip fundamental dalam pemrograman berorientasi objek, seperti enkapsulasi, abstraksi, pewarisan, dan polimorfisme. Namun, S4 memberikan formalitas tambahan dalam proses definisi class, struktur data, serta pemetaan fungsi generik terhadap objek tertentu.

### 2.3 Epidemiologi Penyakit Menular

Epidemiologi penyakit menular merupakan cabang ilmu yang mempelajari distribusi, pola penyebaran, dan faktor-faktor yang memengaruhi terjadinya penyakit yang dapat ditransmisikan dari satu individu ke individu lainnya. Dalam kajian ini, penyakit dipandang sebagai suatu fenomena yang tidak terjadi secara acak, melainkan mengikuti pola tertentu yang dipengaruhi oleh interaksi antara agen penyebab, penjamu (host), dan lingkungan. Pendekatan epidemiologi menekankan pada upaya untuk mengidentifikasi bagaimana suatu penyakit muncul, menyebar, serta apa saja determinan yang berperan dalam meningkatkan maupun menurunkan risiko penyebarannya(Amalina dkk., 2025).

Konsep dasar epidemiologi yang banyak digunakan dalam analisis penyakit menular meliputi beberapa ukuran frekuensi penyakit seperti *incidence rate* dan *prevalence rate*. Insiden merujuk pada jumlah kasus baru yang muncul dalam periode tertentu, sehingga sangat berguna untuk menilai dinamika penularan dan efektivitas intervensi kesehatan masyarakat. Sementara itu, prevalensi menggambarkan seluruh kasus yang ada pada suatu waktu tertentu dan memberikan gambaran umum tentang beban penyakit dalam suatu populasi. Selain kedua konsep tersebut, ukuran lain yang sering digunakan adalah *case fatality rate* yang menunjukkan proporsi kematian di

antara kasus yang terdiagnosis, serta *attack rate* yang sering digunakan pada kondisi KLB untuk menilai kecepatan penularan (Amalina dkk., 2025).

Dalam epidemiologi penyakit menular, pemahaman mengenai rantai penularan merupakan aspek fundamental. Penularan dapat terjadi melalui berbagai mekanisme, antara lain kontak langsung, droplet, airborne, atau melalui vektor tertentu. Setiap mekanisme memiliki karakteristik yang berbeda sehingga memerlukan pendekatan pengendalian yang spesifik. Selain itu, faktor risiko seperti kepadatan penduduk, mobilitas, kondisi lingkungan, sanitasi, cakupan imunisasi, serta karakteristik sosial ekonomi turut menentukan tingkat kerentanan suatu wilayah terhadap penyakit menular.

Penggunaan pendekatan analitik seperti klasterisasi menjadi relevan dalam epidemiologi modern, terutama ketika data yang digunakan bersifat multivariat dan mencakup berbagai indikator risiko. Metode seperti K-Means dapat membantu mengelompokkan wilayah berdasarkan pola epidemiologis yang serupa, sehingga daerah dengan tingkat risiko yang sama dapat diidentifikasi secara lebih sistematis. Proses klasterisasi ini sangat bermanfaat untuk pemetaan profil risiko, karena memungkinkan pengambilan keputusan berbasis bukti untuk menentukan area prioritas, alokasi sumber daya, maupun intervensi kesehatan masyarakat yang tepat Sasaran (Amalina dkk., 2025).

Dalam konteks Provinsi Lampung dan Sumatera Selatan, pendekatan epidemiologi menjadi semakin penting mengingat kedua wilayah memiliki kondisi demografis, mobilitas penduduk, dan heterogenitas lingkungan yang dapat memengaruhi penyebaran penyakit menular. Variasi tingkat kepadatan penduduk, kapasitas fasilitas kesehatan, serta perbedaan faktor risiko antar kabupaten/kota membuat analisis berbasis klaster menjadi relevan untuk memahami perbedaan pola epidemiologi antarwilayah. Dengan demikian, pemahaman terhadap epidemiologi penyakit menular tidak hanya membantu mengenali karakteristik penyakit, tetapi juga memberikan dasar ilmiah dalam menyusun kebijakan kesehatan yang lebih efektif dan responsif terhadap kondisi lokal.

## 2.4 Pemetaan Profil Risiko

Pemetaan profil risiko merupakan pendekatan analitis yang bertujuan menggambarkan tingkat kerentanan suatu wilayah terhadap penyakit menular berdasarkan kombinasi indikator epidemiologis, demografis, lingkungan, dan kapasitas pelayanan kesehatan. Dalam epidemiologi penyakit menular, pemetaan risiko menjadi dasar penting dalam memahami bagaimana penyakit tersebar secara spasial serta wilayah mana yang memiliki potensi penyebaran lebih tinggi dibandingkan wilayah lainnya.

Secara konseptual, profil risiko dibangun melalui serangkaian indikator yang merepresentasikan determinan penyakit, seperti jumlah kasus, kepadatan penduduk, tingkat mobilitas, kondisi sanitasi, cakupan imunisasi, dan indikator lainnya. Setiap indikator memiliki kontribusi yang berbeda dalam membentuk tingkat risiko suatu wilayah.

Untuk membentuk profil risiko, setiap indikator terlebih dahulu diseragamkan melalui proses normalisasi atau standardisasi agar dapat digabungkan secara setara. Setelah itu, nilai indikator dapat diolah menjadi skor komposit menggunakan teknik penjumlahan, pembobotan, atau metode analitis lain seperti analisis komponen utama. Skor tersebut kemudian diklasifikasikan ke dalam kategori risiko, seperti rendah, sedang, atau tinggi.

Dalam pemetaan, konsep dasar representasi spasial seperti choropleth map, batas administrasi wilayah, skema warna, dan legenda berperan penting dalam menyampaikan informasi risiko. Teknik analisis spasial juga dapat digunakan untuk mendeteksi pola seperti klaster kasus atau area berisiko tinggi.

Hubungan antara pemetaan profil risiko dan klasterisasi, seperti metode K-Means, sangat erat. K-Means mengelompokkan wilayah berdasarkan kemiripan indikator sehingga wilayah yang memiliki karakteristik epidemiologis serupa berada dalam satu klaster. Pendekatan ini sangat penting ketika indikator bersifat multivariat.

Dalam konteks penyakit menular seperti kusta dan DBD, pemetaan profil risiko membantu mengidentifikasi wilayah dengan beban penyakit tinggi atau kondisi

lingkungan kurang mendukung, sehingga analisis ini dapat digunakan sebagai dasar kebijakan dan intervensi kesehatan yang lebih tepat sasaran.

## 2.5 Pemodelan Data

Pemodelan data pada penelitian ini berfungsi sebagai fondasi analitis untuk memastikan bahwa struktur data, transformasi data, pemilihan variabel, serta teknik klasterisasi yang digunakan mampu mencerminkan kondisi epidemiologis serta kapasitas fasilitas kesehatan secara akurat. Tahapan ini mencakup proses preprocessing, penentuan variabel yang relevan, dan pengaturan struktur data dalam sistem objek S4 pada perangkat lunak R. Pendekatan ini penting karena kualitas pemodelan data secara langsung mempengaruhi kinerja algoritma K-Means maupun variasinya seperti Minkowski metric K-Means (Hadi Prakoso dkk., 2023).

### A. Konsep Preprocessing Data

Preprocessing merupakan tahap esensial untuk mempersiapkan data agar memenuhi syarat matematis dan statistika sebelum digunakan dalam klasterisasi. Langkah-langkah penting dalam preprocessing adalah sebagai berikut.

### B. Normalisasi

Normalisasi bertujuan menyetarakan rentang nilai antarvariabel sehingga variabel dengan skala besar tidak mendominasi variabel yang skalanya kecil. Untuk penelitian kesehatan masyarakat, variabel seperti angka kematian ibu, angka kematian bayi, atau prevalensi penyakit sering kali memiliki rentang nilai yang berbeda. Salah satu teknik normalisasi yang umum digunakan adalah normalisasi Min–Max:

$$x_{\text{norm}} = (x - \min(x)) / (\max(x) - \min(x))$$

Teknik ini memetakan nilai  $x$  ke dalam rentang 0 sampai 1 sehingga seluruh variabel berperan seimbang dalam perhitungan jarak K-Means.

### C. Standarisasi

Standarisasi digunakan apabila distribusi variabel berbeda jauh, atau ketika terdapat variabel yang berskala sangat besar. Teknik ini mengubah nilai setiap variabel agar memiliki mean = 0 dan standar deviasi = 1 dengan formula:

$$z = (x - \mu) / \sigma$$

Metode ini relevan ketika variabel epidemiologis memiliki variasi yang lebar sehingga diperlukan penyetaraan agar algoritma tidak berat sebelah pada variabel tertentu.

### D. Scaling

Scaling adalah proses penyetaraan skala data secara umum. Pada klasterisasi berbasis jarak seperti K-Means dan Minkowski metric K-Means, scaling menjadi sangat penting karena jarak antar observasi dihitung berdasarkan selisih nilai tiap fitur. Perbedaan skala yang ekstrim akan membuat variabel tertentu mendominasi jarak secara tidak proporsional (Mirkin, 2012).

### E. Pemilihan Variabel

Pemilihan variabel adalah langkah fundamental dalam penelitian kuantitatif yang menghubungkan konsep teoretis dengan data empiris. Secara metodologis, variabel harus dipilih berdasarkan landasan konseptual yang kuat, bukan secara sembarangan. Sebuah variabel yang baik harus memenuhi tiga kriteria utama: relevansi teoritis (sesuai dengan konsep yang diwakili), validitas dan reliabilitas (dapat diukur secara konsisten), serta memiliki variasi yang memadai untuk menggambarkan perbedaan fenomena. Proses ini melibatkan operasionalisasi konsep, yaitu menerjemahkan ide abstrak menjadi indikator terukur. Selain variabel utama, variabel pendukung juga dapat digunakan untuk memperkuat analisis, asalkan tetap relevan dan tidak bias.

## F. Teori Minkowski Metric K-Means dan Feature Weighting

K-Means klasik menggunakan jarak Euclidean, namun untuk data yang kompleks atau memiliki karakteristik berbeda, Minkowski metric K-Means dapat digunakan. Jarak Minkowski untuk dua vektor  $x$  dan  $y$  didefinisikan sebagai:

$$d(x, y) = (\sum |x_v - y_v|^p)^{(1/p)}$$

Jika  $p = 2$ , maka rumus ini menjadi jarak Euclidean. Jika  $p = 1$ , jarak tersebut menjadi Manhattan distance. Versi umum dari fungsi objektif Minkowski K-Means adalah:

$$W_p(S, c) = \sum \sum |y_iv - c_kv|^p$$

Feature weighting atau pembobotan fitur digunakan untuk menyesuaikan pengaruh masing-masing variabel terhadap pembentukan klaster. Dalam Minkowski Weighted K-Means (MWK-Means), bobot fitur dihitung berdasarkan variasi dalam klaster. Semakin besar variasi suatu fitur, semakin kecil bobotnya agar tidak mendistorsi proses klasterisasi (Mirkin, 2012).

Bobot fitur dihitung dengan:

$$w_v = 1 / \sum v' (D_v / D_v')^{(1/(p-1))}$$

Bobot ini kemudian digunakan untuk menghasilkan jarak baru:

$$d_w(x, y) = (\sum w_v |x_v - y_v|^p)^{(1/p)}$$

Pendekatan ini relevan untuk data kesehatan karena variabel epidemiologis jarang berada dalam skala yang seragam, sehingga pembobotan dapat memperbaiki ketepatan klaster.

## G. Struktur Data dalam Sistem Objek S4 pada R

Sistem objek S4 dalam R digunakan untuk membuat kelas dengan struktur formal, validasi ketat, dan definisi slot yang jelas. Dalam pengembangan fungsi K-Means berbasis S4, struktur data dapat mencakup:

- slot data: menyimpan data mentah atau data hasil preprocessing,
- slot scaled\_data: menyimpan data yang telah dinormalisasi atau distandarisasi,
- slot k: menyimpan jumlah klaster,
- slot centroids: menyimpan titik pusat *cluster*,
- slot result: menyimpan hasil klasterisasi, iterasi, dan metrik evaluasi.

Struktur formal ini memastikan bahwa proses analisis terorganisir dengan baik dan dapat diperluas. Data epidemiologis yang disusun dalam bentuk matriks  $n \times p$ , di mana n adalah jumlah wilayah dan p adalah jumlah variabel, sangat sesuai untuk dimodelkan dalam kerangka S4.

Keseluruhan tahapan pemodelan data ini memberikan landasan kuat bagi pelaksanaan analisis K-Means dan variannya. Dengan preprocessing yang tepat, pemilihan variabel yang relevan, teori matematis yang sesuai, serta penerapan struktur data berbasis S4, hasil klasterisasi diharapkan mampu menggambarkan pola risiko penyakit menular dan kapasitas fasilitas kesehatan secara akurat dan dapat dipertanggungjawabkan secara akademik.

## 2.6 Evaluasi Kapasitas Fasilitas Kesehatan

Evaluasi kapasitas fasilitas kesehatan merupakan komponen kunci dalam analisis kerentanan suatu wilayah terhadap penyakit menular. Kapasitas ini tidak hanya dimaknai sebagai ketersediaan sarana fisik, tetapi juga mencakup kecukupan tenaga, kemampuan respons diagnostik, dan keluasan jangkauan layanan kesehatan. Dalam konteks pengendalian penyakit menular, kapasitas fasilitas kesehatan menentukan seberapa cepat suatu wilayah mampu mendekripsi, merespons, dan menanggulangi peningkatan kasus, terutama ketika terjadi potensi Kejadian Luar Biasa (KLB) (Santoso dkk., 2011).

Salah satu indikator fundamental dalam evaluasi kapasitas adalah jumlah dan distribusi tenaga kesehatan (nakes). Tenaga kesehatan, khususnya dokter, perawat, epidemiolog, dan petugas surveilans, memiliki peran sentral dalam upaya deteksi dini, penegakan diagnosis, serta pelaksanaan tindakan penanggulangan. Ketimpangan distribusi tenaga kesehatan sering kali menyebabkan keterlambatan respons, terutama

pada daerah dengan akses terbatas. Dalam konteks KLB, pedoman penanggulangan KLB menekankan pentingnya kesiapan petugas untuk melakukan investigasi epidemiologi secara cepat dan sistematis demi mencegah perluasan transmisi.

Selain tenaga kesehatan, ketersediaan tempat tidur (TT), termasuk tempat tidur isolasi, menjadi indikator kapasitas pelayanan kuratif. Lonjakan kasus pada penyakit menular akut dapat menyebabkan tekanan besar terhadap rumah sakit maupun puskesmas rawat inap. Ketidakmampuan fasilitas kesehatan menyediakan TT yang memadai sering kali berdampak pada meningkatnya mortalitas, terutama bila kasus yang membutuhkan perawatan intensif tidak tertangani secara optimal. Oleh karena itu, evaluasi kapasitas TT tidak hanya memperhitungkan jumlah fisik, namun juga kesesuaianya dengan kebutuhan standar pelayanan untuk penyakit menular.

Laboratorium kesehatan turut menjadi bagian yang tidak terpisahkan dari kapasitas sistem kesehatan. Laboratorium berfungsi memastikan ketepatan diagnosis, mempercepat konfirmasi etiologi, dan memungkinkan deteksi pola penyebaran penyakit secara lebih akurat. Dalam kerangka surveilans dan respon KLB, kemampuan laboratorium untuk melakukan pemeriksaan spesifik (misalnya kultur, PCR, serologi) menentukan kecepatan verifikasi kasus serta penyusunan langkah penanggulangan. Wilayah tanpa fasilitas laboratorium yang memadai cenderung mengalami keterlambatan dalam verifikasi kasus, sehingga risiko pelebaran transmisi menjadi lebih tinggi (Santoso dkk., 2011).

Di tingkat layanan primer, puskesmas memegang peran strategis sebagai garda terdepan sistem kesehatan. Puskesmas bertanggung jawab terhadap kegiatan surveilans, deteksi kasus di masyarakat, pelacakan kontak, serta edukasi kesehatan. Kapasitas puskesmas, baik dari sisi SDM, sarana penunjang, maupun jangkauan wilayah kerja, sangat menentukan kemampuan daerah dalam melakukan respons dini. Dalam berbagai pedoman penanggulangan KLB, puskesmas sering menjadi unit pertama yang diinstruksikan untuk melakukan investigasi lapangan, pengumpulan data, hingga penyuluhan kepada masyarakat.

Indikator berikutnya adalah cakupan pelayanan kesehatan, termasuk cakupan imunisasi, cakupan deteksi dini, serta akses masyarakat terhadap fasilitas kesehatan.

Cakupan pelayanan yang rendah sering dikaitkan dengan meningkatnya kerentanan populasi, terutama pada penyakit menular yang dapat dicegah dengan imunisasi (PD3I). Kawasan dengan cakupan imunisasi tidak merata, misalnya, memiliki risiko lebih besar mengalami peningkatan kasus dan perluasan klaster.

Secara konseptual, kapasitas fasilitas kesehatan memiliki hubungan erat dengan tingkat risiko penyakit menular. Risiko tidak hanya ditentukan oleh besarnya paparan dan kerentanan populasi, tetapi juga oleh kemampuan sistem kesehatan merespons. Dua wilayah dengan tingkat risiko penularan yang sama dapat menunjukkan dampak epidemiologis berbeda ketika kapasitas responnya berbeda. Karena itu, analisis risiko penyakit menular idealnya disandingkan dengan evaluasi kapasitas fasilitas kesehatan agar interpretasi risiko menjadi lebih komprehensif.

Integrasi antara penilaian risiko (risk assessment) dan penguatan kapasitas (capacity building) merupakan pendekatan penting dalam manajemen kesehatan masyarakat modern. Hasil penilaian risiko harus diterjemahkan ke dalam prioritas peningkatan kapasitas, seperti penguatan tenaga surveilans, peningkatan sarana laboratorium, penambahan TT isolasi, atau peningkatan cakupan pelayanan primer. Dengan demikian, penguatan kapasitas bukan hanya reaktif terhadap KLB yang sedang berlangsung, tetapi bersifat preventif untuk meminimalkan dampak epidemi di masa mendatang (Santoso dkk., 2011).

Secara keseluruhan, evaluasi kapasitas fasilitas kesehatan menjadi fondasi penting dalam membangun sistem kewaspadaan dini dan respons. Kapasitas yang kuat memungkinkan wilayah tidak hanya mendeteksi peningkatan kasus secara cepat, tetapi juga menahan laju penularan, meminimalkan fatalitas, dan memastikan KLB dapat dikendalikan dalam waktu singkat sesuai dengan amanat pedoman penanggulangan penyakit menular.

## **BAB III**

### **Metode Penelitian**

#### 3.1 Jenis Data

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan memanfaatkan data sekunder. Data sekunder dipilih sebagai basis penelitian karena efisiensi dalam memperoleh cakupan wilayah yang luas (lintas provinsi) dengan standarisasi data yang terjamin oleh instansi pemerintah.

Data yang dikumpulkan merupakan data *cross-sectional* (satu titik waktu) yang bersumber dari publikasi resmi digital Badan Pusat Statistik (BPS) Provinsi Lampung dan Provinsi Sumatera Selatan (Tahun 2024). Struktur data yang digunakan dalam penelitian ini dikategorikan ke dalam dua dimensi analisis:

1. Dimensi Beban Penyakit (*Demand Side*): Data ini merepresentasikan tingkat kebutuhan layanan kesehatan masyarakat. Data mencakup statistik epidemiologi dari lima penyakit menular prioritas, yaitu Tuberkulosis (TBC), HIV/AIDS, Kusta, Malaria, dan Demam Berdarah Dengue (DBD). Data ini bersifat numerik (rasio) yang menggambarkan jumlah kasus absolut maupun tingkat insidensi (*incidence rate*) di setiap wilayah.
2. Dimensi Kapasitas Fasilitas (*Supply Side*): Data ini merepresentasikan ketersediaan infrastruktur fisik kesehatan. Data mencakup jumlah unit layanan kesehatan dari berbagai tingkatan, mulai dari fasilitas kesehatan tingkat pertama (FKTP) seperti Puskesmas dan Klinik, hingga fasilitas kesehatan rujukan tingkat lanjut (FKRTL) seperti Rumah Sakit Umum dan Khusus.

Populasi atau unit analisis dalam penelitian ini adalah seluruh wilayah administrasi tingkat II di kedua provinsi, yang terdiri dari 32 Kabupaten/Kota (15 Kabupaten/Kota di Lampung dan 17 Kabupaten/Kota di Sumatera Selatan).

### 3.2 Teknik Pengumpulan dan Pra-pemrosesan Data

Proses pengumpulan dan pengolahan data dilakukan secara komputasi menggunakan bahasa pemrograman R. Teknik ini dipilih untuk menjamin reproduksibilitas (*reproducibility*) dan akurasi pengolahan data yang kompleks. Tahapan pelaksanaan pengumpulan data adalah sebagai berikut:

1. Akuisisi Data Digital (*Digital Data Ingestion*): Data mentah (*raw data*) diakuisisi dari repositori data terbuka (*open data repository*) yang di-hosting di GitHub. Pengambilan data dilakukan secara terprogram menggunakan fungsi `read.csv()` untuk membaca file berekstensi *Comma Separated Values* langsung dari URL sumber. Metode ini meminimalisir risiko *human error* yang sering terjadi pada proses penyalinan data manual.
2. Integrasi Data Vertikal (*Vertical Data Integration*): Mengingat data bersumber dari dua provinsi yang berbeda, dilakukan teknik penggabungan baris (*row binding*) menggunakan fungsi `bind_rows()` dari paket pustaka dplyr. Proses ini menyatukan dataset Lampung dan Sumatera Selatan ke dalam satu struktur data tunggal (*dataframe*) yang seragam, sehingga memungkinkan analisis komparatif lintas provinsi.
3. Transformasi dan Standardisasi Variabel:
  - a. Renaming: Melakukan penyesuaian nama variabel dari format asli BPS yang panjang dan mengandung karakter khusus (spasi, tanda baca) menjadi format pemrograman yang standar (misalnya: Jumlah Kasus... menjadi TBC).
  - b. Normalisasi Satuan: Untuk variabel penyakit tertentu (Kusta, Malaria, DBD), data dikonversi menjadi rasio per 100.000 penduduk. Langkah ini krusial untuk menghilangkan bias populasi, sehingga wilayah dengan penduduk sedikit dapat diperbandingkan secara adil dengan wilayah padat penduduk.

4. Penanganan Data Hilang (*Missing Value Imputation*): Ditemukan adanya nilai kosong (NA) pada beberapa sel data penyakit. Teknik imputasi yang digunakan adalah *Zero Imputation* (mengganti NA dengan 0). Pendekatan ini didasarkan pada asumsi bahwa dalam laporan statistik resmi BPS, tidak adanya angka pelaporan pada kolom penyakit spesifik di suatu wilayah mengindikasikan tidak ditemukannya kasus penyakit tersebut pada periode pelaporan.
5. Integrasi Data Relasional (*Relational Joining*): Tahap akhir pengumpulan data adalah menggabungkan dataset "Profil Penyakit" dan dataset "Fasilitas Kesehatan". Teknik yang digunakan adalah *Left Join* dengan variabel kunci (*primary key*) nama "Kabupaten/Kota". Teknik ini memastikan bahwa data fasilitas dipetakan secara presisi ke wilayah yang sesuai, menghasilkan satu dataset master yang komprehensif untuk analisis.

### 3.3 Variabel yang Diamati

Variabel penelitian diklasifikasikan secara ketat berdasarkan fungsinya dalam arsitektur sistem komputasi S4 yang dikembangkan. Pemisahan ini bertujuan untuk menjaga objektivitas algoritma klasterisasi.

#### 3.3.1 Variabel Input Klasterisasi (*Clustering Variables*)

Variabel berikut berfungsi sebagai indikator risiko dan menjadi satu-satunya input bagi algoritma K-Means untuk menentukan pola pengelompokan wilayah.

*Tabel 3.1 Variabel Input*

Nama Variabel	Definisi Operasional	Skala Data	Satuan
TBC	Jumlah total kasus baru Tuberkulosis yang terkonfirmasi secara klinis/bakteriologis.	Rasio	Kasus
Pengobatan TBC	Jumlah pasien TBC yang telah menyelesaikan masa pengobatan dan dinyatakan sembuh.	Rasio	Kasus
HIV	Jumlah kasus baru infeksi HIV yang dilaporkan oleh layanan VCT.	Rasio	Kasus
Kusta_per_100000	Tingkat penemuan kasus baru Kusta yang dinormalisasi terhadap jumlah penduduk.	Rasio	Per100rb Penduduk
Malaria_per_100000	Angka kesakitan ( <i>morbidity rate</i> ) Malaria yang dinormalisasi.	Rasio	Per100rb Penduduk
DBD_per_100000	Angka kesakitan Demam Berdarah Dengue yang dinormalisasi.	Rasio	Per100rb Penduduk

### 3.3.2 Variabel Evaluasi Kapasitas (*Evaluation Variables*)

Variabel ini tidak diikutsertakan dalam perhitungan jarak Euclidean pada algoritma K-Means untuk mencegah bias (agar pengelompokan murni berdasarkan beban penyakit, bukan jumlah fasilitas). Variabel ini digunakan pada tahap *post-analysis* untuk mengevaluasi kesiapan infrastruktur (*Supply*) pada setiap klaster yang terbentuk.

*Tabel 3.2 Variabel Evaluasi*

Nama Variabel	Definisi Operasional	Skala Data	Satuan
Total_RS	Akumulasi jumlah Rumah Sakit Umum (RSU) dan Rumah Sakit Khusus (RSK) baik milik pemerintah maupun swasta.	Rasio	Unit
Total_Puskesmas	Akumulasi jumlah Puskesmas dengan fasilitas Rawat Inap dan Non-Rawat Inap.	Rasio	Unit
Klinik	Jumlah Klinik Pratama yang beroperasi di wilayah tersebut.	Rasio	Unit
Posyandu	Jumlah Pos Pelayanan Terpadu aktif yang melayani kesehatan ibu dan anak.	Rasio	Unit

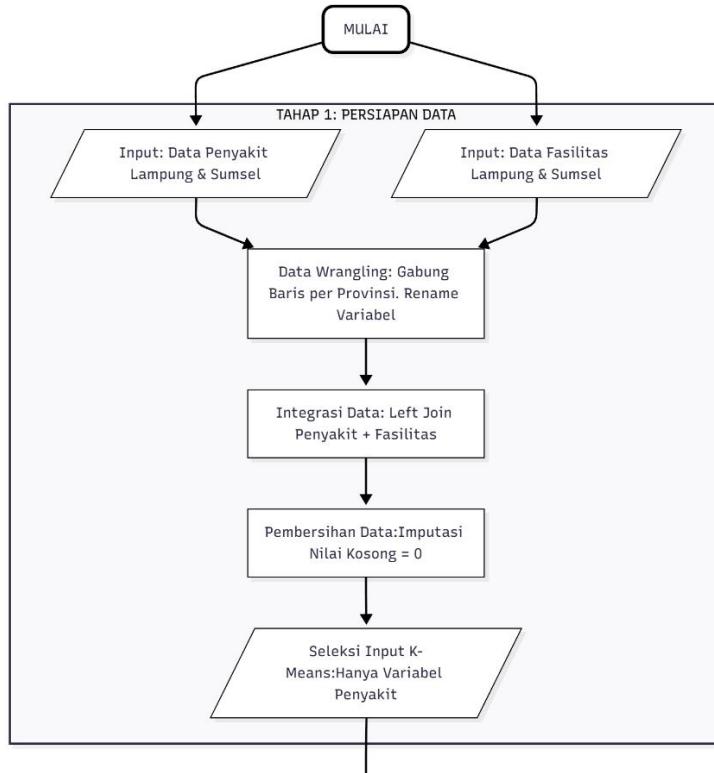
### 3.4 Diagram Alir

Alur logika pemrograman yang dikembangkan dalam penelitian ini cukup kompleks. Untuk memudahkan pemahaman, diagram alir disajikan dalam tiga segmen utama yang berurutan:

- (1) Persiapan Data, (2) Pengembangan Sistem S4, dan (3) Analisis Evaluasi.

### 3.4.1 Tahap 1: Persiapan dan Integrasi Data

Bagian pertama menggambarkan proses akuisisi data mentah (*raw data*) hingga menjadi *input* yang siap diolah.



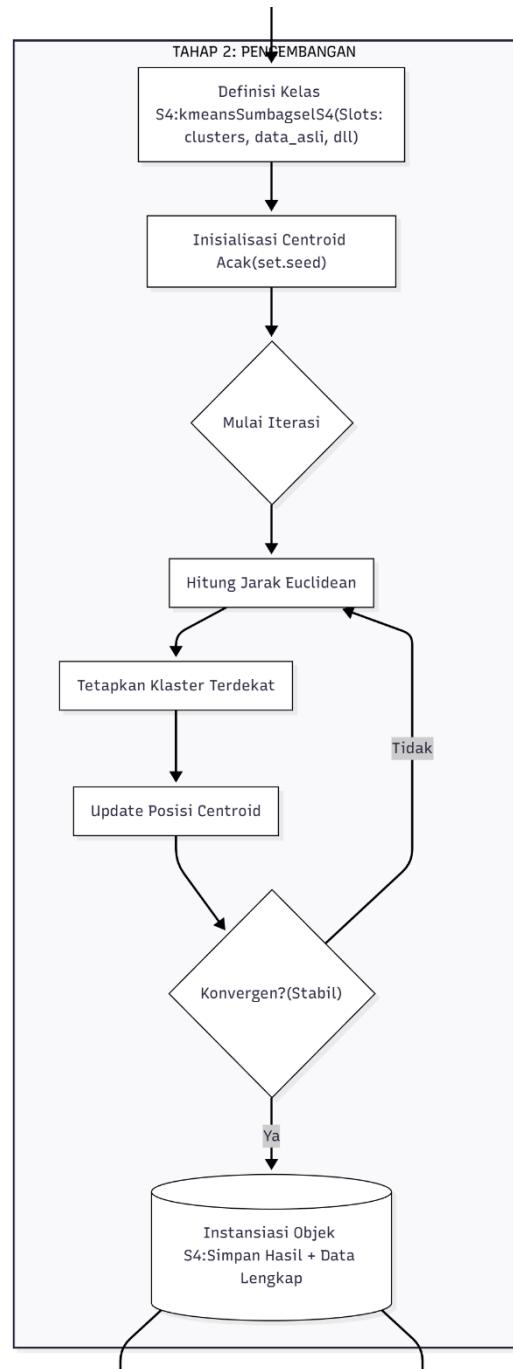
Gambar 3.1 Diagram Tahap 1 : Persiapan Data

Penjelasan Alur:

1. Akuisisi: Sistem membaca data profil penyakit dan fasilitas kesehatan dari repositori digital.
2. Wrangling: Data dari Provinsi Lampung dan Sumsel digabungkan (*bind rows*), kemudian nama variabel distandarisasi.
3. Integrasi: Data fasilitas ditempelkan ke data penyakit menggunakan *Left Join* untuk membentuk dataset lengkap.
4. Seleksi: Sebelum masuk ke tahap algoritma, variabel fasilitas dipisahkan agar tidak membiaskan perhitungan klasterisasi.

### 3.4.2 Tahap 2: Pengembangan Sistem S4 dan Algoritma

Bagian kedua merupakan inti dari penelitian, di mana struktur objek didefinisikan dan algoritma K-Means dijalankan.



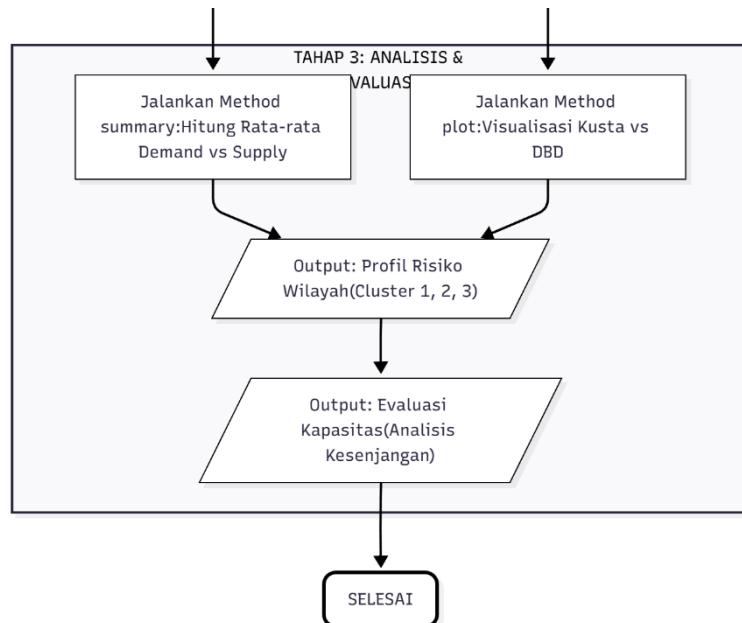
Gambar 3.1 Diagram Alir Tahap 2 : Pengembangan

Penjelasan Alur:

- 4.1 Definisi Kelas: Cetak biru objek kmeansSumbagselS4 dibuat untuk menyiapkan wadah (*slots*) bagi data.
- 4.2 Inisialisasi: Pusat klaster awal (*centroid*) ditentukan secara acak dengan kontrol set.seed.
- 4.3 Iterasi (Looping): Sistem memasuki siklus perulangan: menghitung jarak Euclidean dan memperbarui posisi centroid.
- 4.4 Konvergensi: Jika posisi centroid stabil, iterasi berhenti. Hasil akhir (label klaster) dan data asli dibungkus (*encapsulated*) menjadi satu objek S4 baru.

#### 3.4.3 Tahap 3: Analisis dan Evaluasi Kapasitas

Bagian terakhir menggambarkan bagaimana hasil olahan objek S4 diterjemahkan menjadi wawasan kebijakan (*insight*).



Gambar 3.3 Diagram Alir Tahap 3 : Analisis dan Evaluasi

Penjelasan Alur:

- 4.1 Eksekusi Metode: Metode generik summary dan plot dijalankan pada objek hasil.

- 4.2 Output Profil Risiko: Menghasilkan pengelompokan wilayah berdasarkan beban penyakit (Cluster 1, 2, 3).
- 4.3 Output Evaluasi Kapasitas: Menghasilkan tabel perbandingan antara beban penyakit (*Demand*) dengan ketersediaan fasilitas (*Supply*) untuk mendeteksi kesenjangan pelayanan (*Gap Analysis*)

## **BAB IV**

### **Hasil dan Pembahasan**

#### 4.1 Deskripsi Data

Data yang digunakan dalam penelitian ini mencakup rekam medis penyakit menular dan data infrastruktur kesehatan dari 32 Kabupaten/Kota di Provinsi Lampung dan Sumatera Selatan.

##### 4.1.1 Struktur Data Sampel

Tabel \_ menampilkan cuplikan (*head*) dari dataset final setelah melalui proses pra-pemrosesan. Data ini terdiri dari variabel penyakit yang telah dinormalisasi (Kusta dan DBD per 100.000 penduduk) serta variabel total fasilitas kesehatan.

*Tabel 4.1 Struktur Data Sampel*

Kabupaten Kota	TBC	HIV	Kusta_per_ 100000	DBD_per_ 100000	Total_ RS	Total_ Puskesmas
Lampung Barat	53.3	4	0.98	208.94	2	15
Tanggamus	37.0	5	3.98	65.19	2	26
Lampung Selatan	78.6	31	0.69	25.70	7	28
Lampung Timur	39.8	22	1.52	71.76	5	36
Lampung Tengah	51.0	5	1.55	77.96	10	39

Berdasarkan hasil pra-pemrosesan data, terbentuk *dataframe* akhir yang memuat informasi terpadu mengenai profil kesehatan dan infrastruktur di 32 kabupaten/kota. Tabel sampel data (5 baris pertama) menunjukkan struktur variabel

yang digunakan dalam analisis. Pada sisi beban penyakit (*Demand Side*), terdapat variabel TBC dan HIV yang diukur dalam satuan jumlah kasus absolut, merepresentasikan beban penyakit menular kronis di setiap wilayah. Selain itu, terdapat variabel Kusta\_per\_100000 dan DBD\_per\_100000 yang telah dinormalisasi menjadi angka insidensi per 100.000 penduduk; langkah ini krusial untuk membandingkan tingkat risiko penyakit berbasis lingkungan secara adil antarwilayah dengan populasi berbeda. Pada sisi ketersediaan fasilitas (*Supply Side*), tabel menyajikan variabel agregat Total\_RS (penjumlahan Rumah Sakit Umum dan Khusus) serta Total\_Puskesmas (penjumlahan Puskesmas Rawat Inap dan Non-Rawat Inap).

Data fasilitas ini disandingkan dalam baris yang sama dengan data penyakit untuk setiap kabupaten/kota (seperti Lampung Barat, Tanggamus, dst.), yang memungkinkan analisis kesenjangan (*gap analysis*) antara kebutuhan penanganan penyakit dan kapasitas layanan kesehatan yang tersedia secara langsung. Struktur data yang terintegrasi ini menjadi landasan bagi algoritma K-Means untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik risiko kesehatannya.

#### 4.1.2 Ringkasan Statistik

Analisis statistik deskriptif dilakukan untuk memahami karakteristik pemerataan dan penyebaran data sebelum dilakukan pengelompokan. Tabel 6.2 menyajikan ringkasan statistik yang meliputi nilai rata-rata (*mean*), nilai tengah (*median*), nilai maksimum, dan simpangan baku (*standar deviasi*) untuk variabel-variabel kunci.

*Tabel 4.2 Ringkasan Statistik*

	Mean	Median	Maximum	Std_Dev
TBC	779.74	374.00	7096.00	1307.69
HIV	15.94	6.50	143.00	28.50
Kusta_per_100000	2.39	1.54	7.80	2.03
DBD_per_100000	106.66	74.10	411.82	85.90
Total_RS	5.31	3.50	32.00	6.14
Total_Puskesmas	21.12	19.50	42.00	10.01

Berdasarkan tabel di atas, terlihat adanya fenomena *positive skewness* atau kemencengan data ke kanan pada hampir seluruh variabel penyakit, terutama DBD, TBC, dan HIV. Hal ini ditandai dengan nilai *mean* yang konsisten lebih besar dibandingkan nilai *median*. Kondisi ini mengindikasikan bahwa mayoritas wilayah di Lampung dan Sumatera Selatan sebenarnya memiliki beban penyakit yang relatif rendah, namun rata-rata keseluruhan tertarik naik oleh keberadaan sejumlah wilayah dengan kasus ekstrem (*outliers*).

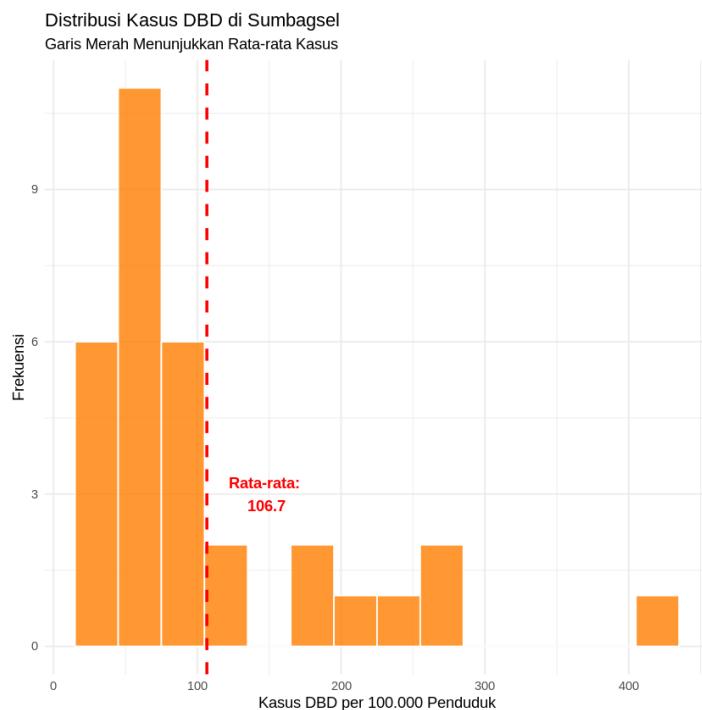
Secara spesifik, variabel Demam Berdarah Dengue (DBD) menunjukkan variabilitas yang sangat tinggi, dengan simpangan baku (85,90) yang hampir menyamai nilai rata-ratanya. Perbedaan yang mencolok antara nilai maksimum (411,82) dan median (74,10) mengonfirmasi bahwa insidensi DBD tidak tersebar merata, melainkan meledak secara lokal di wilayah tertentu. Pola serupa juga terlihat pada data fasilitas kesehatan, di mana jumlah Rumah Sakit memiliki variasi yang jauh lebih tinggi

dibandingkan Puskesmas, menandakan adanya sentralisasi fasilitas rujukan di wilayah perkotaan tertentu.

#### 4.1.3 Visualisasi Distribusi Data

Visualisasi distribusi frekuensi dilakukan menggunakan histogram untuk memvalidasi temuan statistik di atas dan melihat pola sebaran data secara lebih intuitif. Analisis visual ini dibagi menjadi dua dimensi, yaitu dimensi beban penyakit (*Demand*) dan dimensi ketersediaan fasilitas (*Supply*).

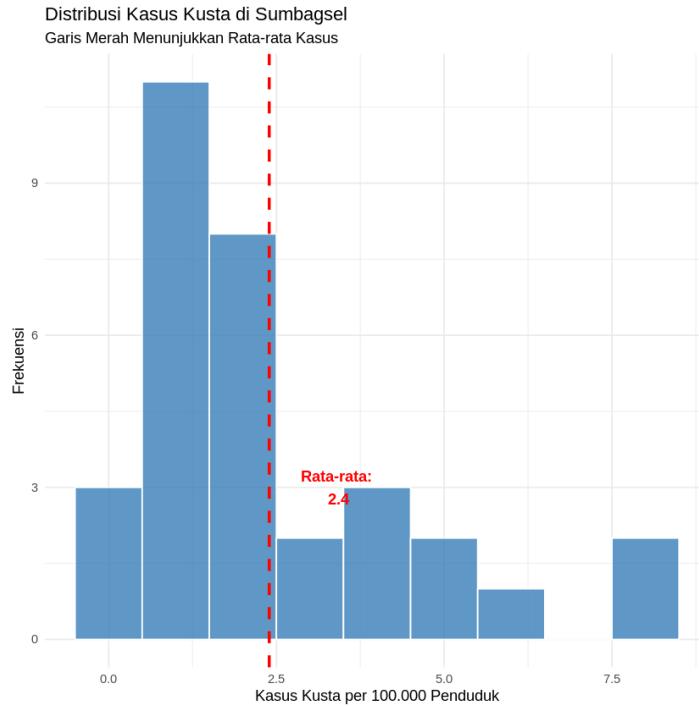
##### A. Distribusi Profil Penyakit (Demand Side)



*Gambar 4.1 Histogram Kasus DBD per Wilayah*

Histogram pada Gambar \_ memperlihatkan bentuk distribusi yang menjulur panjang ke kanan (*long-tail*). Batang-batang grafik yang tinggi terkonsentrasi di sisi kiri, menunjukkan bahwa sebagian besar kabupaten/kota berada pada zona kasus DBD rendah hingga sedang. Namun, keberadaan titik-titik data yang terpisah jauh di ujung kanan sumbu X memvisualisasikan adanya wilayah *hotspot* wabah. Pola distribusi ini

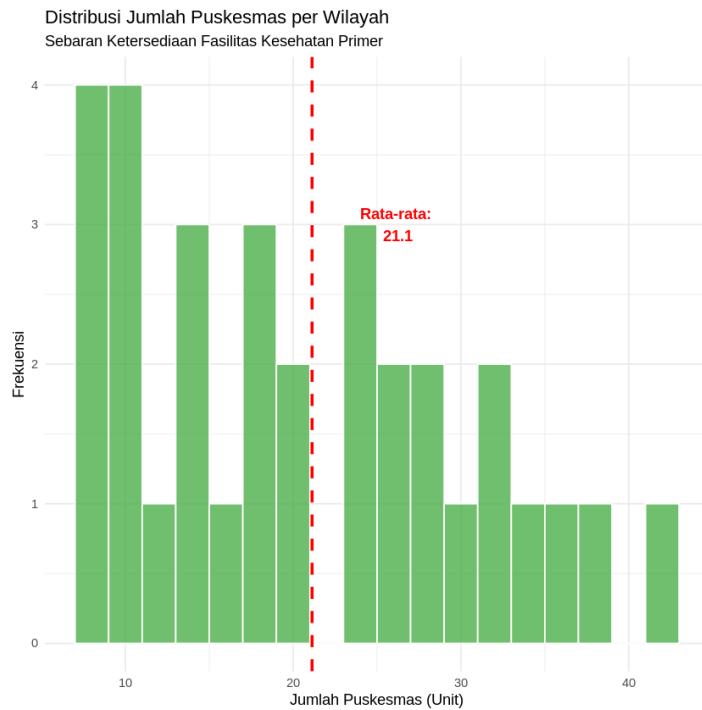
menegaskan bahwa pendekatan rata-rata tidak efektif untuk kebijakan DBD, sehingga diperlukan pendekatan berbasis klaster untuk menangani wilayah-wilayah ekstrem tersebut secara intensif.



Gambar 4.2 Histogram Kasus Kusta per Wilayah

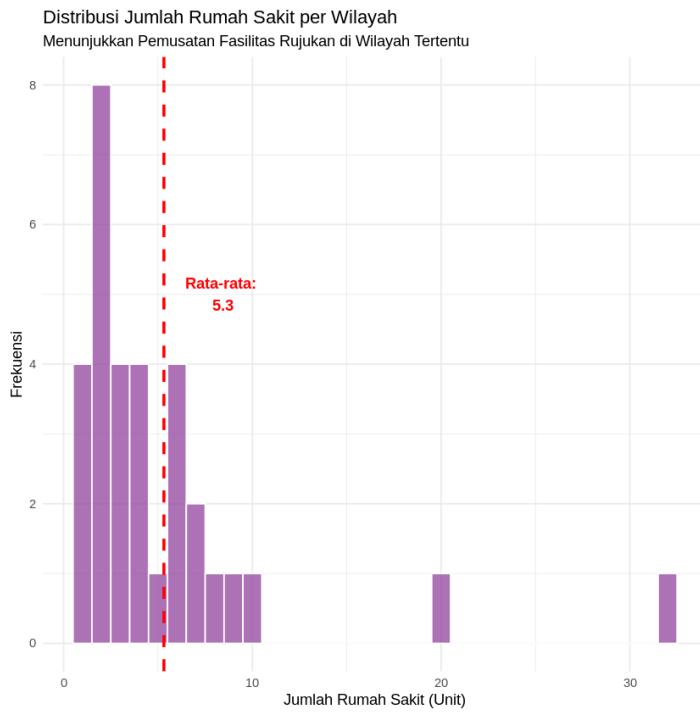
Pola distribusi yang tidak jauh berbeda juga terlihat pada kasus Kusta sebagaimana ditampilkan pada Gambar \_\_\_. Meskipun prevalensi Kusta secara umum rendah, grafik menunjukkan adanya diskontinuitas data di mana terdapat beberapa wilayah yang terpisah di sebelah kanan dengan angka kasus yang signifikan. Variabilitas ini menjadi sinyal kuat bagi algoritma K-Means untuk memisahkan wilayah tersebut ke dalam kelompok risiko tinggi tersendiri.

## B. Distribusi Fasilitas Kesehatan (Supply Side)



Gambar 4.3 Histogram Jumlah Puskesmas per Wilayah

Berbeda dengan pola data penyakit yang sangat timpang, distribusi jumlah Puskesmas pada Gambar \_ cenderung lebih menyebar (*dispersed*) di sekitar nilai rata-ratanya. Meskipun terlihat lebih merata, perhatian khusus perlu diberikan pada wilayah yang berada di sisi paling kiri grafik, yang mengindikasikan jumlah fasilitas di bawah standar rata-rata regional. Analisis kesenjangan nantinya akan menyoroti apakah wilayah dengan fasilitas minim ini beririsan dengan wilayah yang memiliki beban penyakit tinggi.



*Gambar 4.4 Histogram Jumlah Rumah Sakit per Wilayah*

Distribusi Rumah Sakit pada Gambar \_ menunjukkan ketimpangan spasial yang paling tajam dibandingkan variabel lainnya. Grafik didominasi mutlak oleh frekuensi tinggi di angka rendah (0-5 unit), sementara ekor grafik memanjang ke kanan hingga angka 32 unit. Hal ini mengonfirmasi bahwa layanan kesehatan rujukan di Sumbagsel sangat tersentralisasi di beberapa titik pusat pertumbuhan utama, yang berpotensi menimbulkan hambatan akses bagi wilayah penyanga dengan beban penyakit menular tinggi.

#### 4.2 Hasil Implementasi Klasterisasi

Berdasarkan simulasi dengan parameter  $k=3$  dan inisialisasi acak terkontrol (`set.seed(123)`), terbentuk tiga profil wilayah yang berbeda secara signifikan. Berikut adalah ringkasan statistik hasil metode `summary()`:

A. Profil Rata-rata Beban Penyakit (*Demand Side*)

*Tabel 4.3 Rata – rata Beban Penyakit*

Cluster	Anggota	TBC (Kasus)	HIV (Kasus)	Kusta (Kasus)	DBD( Kasus)
1	22	571,32	13,91	2,13	68,62
2	7	176,67	5,14	2,22	246,20
3	3	3.715,33	56,00	4,73	60,00

Tabel ini menunjukkan karakteristik risiko kesehatan yang mendominasi setiap kelompok.

B. Profil Rata-rata Ketersediaan Fasilitas (*Supply Side*)

*Tabel 4.4 Rata-rata Ketersediaan Fasilitas*

Cluster	Total RS (Unit)	Total Puskesmas (Unit)	Klinik (Unit)	Posyandu (Unit)
1	2,86	20,73	28,41	483,91
2	2,86	11,14	18,86	284,86
3	10,67	22,00	110,67	834,67

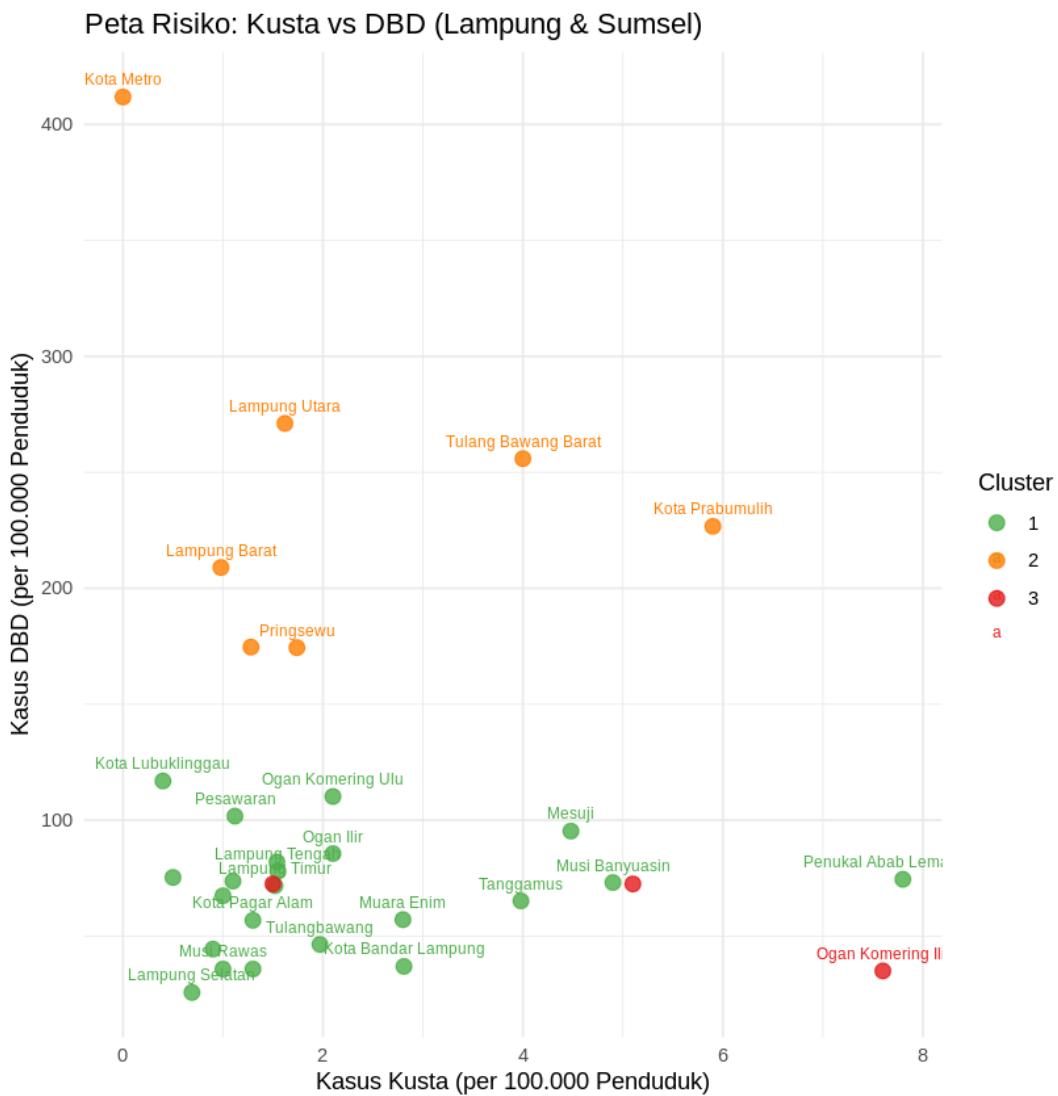
Tabel ini menunjukkan kapasitas infrastruktur kesehatan rata-rata yang dimiliki oleh wilayah dalam setiap kelompok cluster.

### C. Daftar Anggota Klaster (Wilayah)

*Tabel 4.5 Daftar Anggota Klaster*

Klaster	Profil Risiko	Anggota Wilayah (Kabupaten/Kota)
1	Wilayah Penyanga	Empat Lawang, Kota Bandar Lampung, Kota Lubuklinggau, Kota Pagar Alam, Lahat, Lampung Selatan, Lampung Tengah, Lampung Timur, Mesuji, Muara Enim, Musi Banyuasin, Musi Rawas, Musi Rawas Utara, Ogan Ilir, Ogan Komering Ulu, Ogan Komering Ulu Selatan, Ogan Komering Ulu Timur, Penukal Abab Lematang Ilir, Pesawaran, Tanggamus, Tulangbawang, Way Kanan.
2	Zona Wabah DBD	Kota Metro, Kota Prabumulih, Lampung Barat, Lampung Utara, Pesisir Barat, Pringsewu, Tulang Bawang Barat.
3	Hotspot TBC/HIV	Banyu Asin, Kota Palembang, Ogan Komering Ilir.

#### D. Peta Risiko: Kusta vs DBD di Provinsi Lampung dan Sumsel



Gambar 4.5 Peta Risiko Kutsa vs DBD di Sumbagsel

Sumbu X merepresentasikan kasus Kusta, Sumbu Y merepresentasikan kasus DBD. Warna titik menunjukkan keanggotaan klaster.

### 4.3 Interpretasi

Analisis lanjutan dilakukan dengan membandingkan profil beban penyakit (*Demand*) terhadap ketersediaan fasilitas kesehatan (*Supply*). Interpretasi ini mengintegrasikan karakteristik anggota wilayah dalam setiap klaster untuk memberikan konteks spasial yang akurat.

#### 4.3.1 Interpretasi Klaster 2: Wilayah Krisis DBD dengan Kesenjangan Fasilitas

Klaster 2 beranggotakan 7 wilayah campuran antara kota administratif kecil dan kabupaten, yaitu: Kota Metro, Kota Prabumulih, Lampung Barat, Lampung Utara, Pesisir Barat, Pringsewu, dan Tulang Bawang Barat.

- Analisis Risiko (*Demand*): Klaster ini merupakan zona merah untuk penyakit berbasis lingkungan. Rata-rata kejadian Demam Berdarah Dengue (DBD) mencapai 246,20 kasus per 100.000 penduduk, angka tertinggi di seluruh wilayah observasi. Tingginya kasus di wilayah seperti Kota Metro dan Pringsewu kemungkinan dipengaruhi oleh kepadatan pemukiman semi-urban yang mendukung perkembangbiakan vektor nyamuk.
- Evaluasi Kapasitas (*Supply*): Temuan paling kritis ada pada klaster ini. Meskipun menghadapi wabah DBD tertinggi, rata-rata jumlah Puskesmas di wilayah ini hanya 11,14 unit (terendah).
- Implikasi Kebijakan: Terjadi kesenjangan infrastruktur primer yang serius. Wilayah-wilayah ini memerlukan prioritas alokasi anggaran untuk penguatan fungsi kesehatan lingkungan di Puskesmas yang ada, atau penambahan unit baru untuk meningkatkan rasio layanan terhadap penduduk berisiko.

#### 4.3.2 Interpretasi Klaster 3: Episentrum Penyakit Menular (Wilayah Palembang & Penyangga)

Klaster 3 memiliki anggota yang sangat spesifik, yaitu Kota Palembang beserta dua wilayah penyangganya di Sumatera Selatan: Banyuasin dan Ogan Komering Ilir (OKI).

- Analisis Risiko (*Demand*): Klaster ini adalah "hotspot" utama beban penyakit kronis di Sumbagsel. Rata-rata kasus TBC sangat ekstrem (3.715 kasus) dan

HIV tertinggi (56 kasus). Konsentrasi ini mengindikasikan bahwa Kota Palembang sebagai pusat metropolitan memiliki laju transmisi penyakit udara dan kontak seksual yang sangat tinggi, yang turut memengaruhi wilayah penyangganya (aglomerasi urban).

- Evaluasi Kapasitas (*Supply*): Kapasitas fasilitas di klaster ini sangat melimpah, dengan rata-rata 10,67 Rumah Sakit dan 110,67 Klinik. Hal ini wajar mengingat posisi Palembang sebagai pusat rujukan regional.
- Implikasi Kebijakan: Ketersediaan fasilitas fisik sudah memadai (*over-supply* relatif terhadap wilayah lain). Masalah utama di klaster ini adalah pengendalian transmisi. Diperlukan strategi *active case finding* yang agresif di wilayah padat penduduk Palembang, Banyuasin, dan OKI untuk memutus rantai penularan TBC dan HIV, bukan sekadar mengandalkan pengobatan kuratif di Rumah Sakit.

#### 4.3.3 Interpretasi Klaster 1: Wilayah Stabil dengan Fasilitas Memadai

Klaster 1 mencakup 22 wilayah mayoritas, termasuk Kota Bandar Lampung, Lampung Selatan, Lahat, dan Muara Enim.

- Analisis Risiko (*Demand*): Masuknya Kota Bandar Lampung ke dalam klaster ini (bukan Cluster 3) menunjukkan bahwa beban penyakit di Ibu Kota Provinsi Lampung relatif lebih terkendali dibandingkan Palembang, atau setara dengan rata-rata kabupaten lainnya dalam model ini. Secara umum, wilayah ini memiliki profil penyakit moderat (TBC 571, DBD 68)
- Evaluasi Kapasitas (*Supply*): Distribusi fasilitas di wilayah ini tergolong sangat baik. Rata-rata jumlah Puskesmas mencapai 20,73 unit, hampir dua kali lipat lebih banyak dibandingkan Klaster 2. Keberadaan fasilitas primer yang kuat ini kemungkinan berkontribusi pada rendahnya angka kejadian luar biasa (KLB) DBD di wilayah-wilayah ini.
- Implikasi Kebijakan: Fokus kebijakan di 22 wilayah ini adalah pemeliharaan standar pelayanan minimal (SPM). Sistem kesehatan yang ada sudah berjalan

pada jalur yang tepat dalam menyeimbangkan beban penyakit dengan ketersediaan layanan.

#### 4.4 Diskusi

Bagian ini mendiskusikan relevansi temuan hasil klasterisasi dengan teori kesehatan masyarakat serta membandingkannya dengan pola epidemiologi dan metodologi yang ditemukan pada penelitian-penelitian terdahulu.

##### 4.4.1 Validasi Metode K-Means untuk Pemetaan Penyakit

Penelitian ini membuktikan bahwa algoritma K-Means efektif dalam mengidentifikasi pola penyebaran penyakit menular yang bersifat multivariat. Hasil pengelompokan yang memisahkan wilayah wabah DBD (Cluster 2) dan wilayah penyakit kronis (Cluster 3) sejalan dengan penelitian Bastian, Sujadi, & Febrianto (2018). Dalam studi mereka di Kabupaten Majalengka, algoritma K-Means juga terbukti berhasil mengelompokkan kecamatan berdasarkan tingkat kerawanan penyakit menular (TBC, DBD, Diare) menjadi tiga klaster prioritas.

Konsistensi hasil ini mengonfirmasi bahwa metode *unsupervised learning* seperti K-Means merupakan alat yang reliabel (*robust*) untuk memetakan zonasi kesehatan, tidak hanya pada skala kabupaten tetapi juga pada skala lintas provinsi seperti yang dilakukan dalam penelitian ini. Selain itu, penelitian Faridah dkk. (2024) di DKI Jakarta juga mendukung penggunaan K-Means untuk mendeteksi *hotspot* penyebaran penyakit menular di wilayah padat penduduk.

##### 4.4.2 Fenomena *Inverse Care Law* pada Klaster DBD

Temuan paling kritis dalam penelitian ini adalah kondisi paradoks pada Cluster 2, di mana wilayah dengan beban wabah DBD tertinggi (rata-rata 246,2 kasus/100.000 penduduk) justru memiliki rata-rata jumlah Puskesmas terendah (11,14 unit). Kondisi ketimpangan ini mendukung teori klasik "Inverse Care Law" yang dicetuskan oleh Hart (1971). Teori ini menyatakan bahwa "*ketersediaan pelayanan kesehatan yang baik cenderung bervariasi secara terbalik dengan kebutuhan populasi yang dilayani*".

Artinya, daerah yang paling sakit seringkali mendapatkan sumber daya yang paling sedikit.

Relevansi temuan ini diperkuat oleh studi Wijayanti dkk. (2016) mengenai epidemiologi Dengue di Jawa, yang menemukan bahwa faktor ketersediaan fasilitas kesehatan masyarakat memiliki korelasi signifikan dalam manajemen pengendalian vektor DBD. Rendahnya jumlah Puskesmas di Cluster 2 menjadi hambatan struktural dalam pelaksanaan program preventif (seperti *fogging* dan pemantauan jentik), yang menyebabkan kasus DBD di wilayah tersebut tetap tinggi secara persisten. Hal ini juga sejalan dengan Noor et al. (2021) yang menyoroti bahwa akses terbatas ke pelayanan primer berdampak langsung pada tingginya angka kesakitan di masyarakat.

#### 4.4.3 Dinamika Penyakit Urban dan Kepadatan Penduduk

Terbentuknya Cluster 3 (Palembang, dan wilayah penyangga) sebagai *hotspot* TBC dan HIV menegaskan pengaruh faktor demografi urban terhadap transmisi penyakit. Meskipun hasil evaluasi fasilitas menunjukkan bahwa *Supply* di klaster ini sangat tinggi (rata-rata 10 RS dan 110 Klinik), beban penyakit tetap ekstrem.

Hal ini sejalan dengan temuan Fuady dkk. (2021) yang menganalisis spasial TBC di daerah urban. Penelitian tersebut menyimpulkan bahwa kepadatan hunian dan mobilitas penduduk di kota besar merupakan faktor risiko dominan yang mempercepat transmisi *airborne disease* seperti TBC, terlepas dari ketersediaan akses ke fasilitas kesehatan. Penelitian Putri dkk. (2023) juga menemukan pola serupa di mana penyakit menular cenderung terkonsentrasi di pusat-pusat pertumbuhan ekonomi. Temuan ini mengimplikasikan bahwa strategi penambahan fasilitas fisik di Cluster 3 tidak akan efektif menurunkan kasus jika tidak disertai dengan intervensi kesehatan lingkungan dan deteksi dini yang agresif.

#### 4.4.4 Keunggulan Metodologis Sistem S4

Secara teknis, penelitian ini menawarkan kebaruan (*novelty*) dalam penerapan sistem objek S4 untuk analisis kebijakan kesehatan. Berbeda dengan pendekatan prosedural standar yang digunakan pada penelitian sebelumnya, sistem S4 yang

dikembangkan mampu melakukan enkapsulasi data fasilitas (@data\_asli) di dalam objek hasil klasterisasi.

Pendekatan ini memitigasi risiko kesalahan pemetaan data (*data mismatch*) yang sering terjadi saat menggabungkan data dari sumber berbeda (Penyakit vs Fasilitas). Dengan demikian, fungsi yang dikembangkan dalam penelitian ini tidak hanya akurat secara matematis, tetapi juga memiliki integritas data yang lebih tinggi untuk mendukung pengambilan keputusan berbasis bukti (*evidence-based policy*).

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan hasil penelitian mengenai pengembangan fungsi K-Means berbasis sistem objek S4 serta penerapannya pada pemetaan profil risiko penyakit menular dan evaluasi kapasitas fasilitas kesehatan di Provinsi Lampung dan Sumatera Selatan, diperoleh beberapa poin kesimpulan utama.

1. Pengembangan fungsi K-Means dengan pendekatan sistem objek S4 berhasil menghasilkan struktur fungsi yang lebih terorganisasi, konsisten, dan mudah diperluas. Pendekatan ini memungkinkan proses analisis berjalan secara sistematis serta memudahkan integrasi antara data beban penyakit dan kapasitas fasilitas kesehatan.
2. Hasil klasterisasi menunjukkan adanya tiga kelompok wilayah dengan karakteristik risiko yang berbeda. Klaster 2 merupakan zona krisis DBD, Klaster 3 menjadi episentrum penyakit kronis TBC dan HIV, sedangkan Klaster 1 mencerminkan kondisi stabil dengan risiko moderat dan fasilitas memadai.
3. Integrasi hasil klasterisasi dan data fasilitas kesehatan mengungkapkan adanya ketidakseimbangan antara risiko penyakit dan kapasitas layanan di beberapa wilayah. Analisis ini menunjukkan pentingnya pemetaan risiko berbasis klaster sebagai dasar perencanaan kesehatan yang lebih tepat sasaran.

## 5.2 Saran

Berdasarkan temuan penelitian, terdapat beberapa rekomendasi yang dapat dipertimbangkan untuk penelitian lanjutan maupun implementasi kebijakan kesehatan daerah.

1. Penguatan kapasitas fasilitas kesehatan pada wilayah berisiko tinggi, khususnya klaster dengan insidensi DBD tinggi.
2. Penerapan strategi active case finding pada wilayah dengan beban penyakit kronis seperti TBC dan HIV.
3. Penggunaan metode klasterisasi alternatif dalam penelitian lanjutan untuk meningkatkan akurasi analisis.
4. Integrasi analisis spasial dalam sistem S4 untuk menghasilkan peta risiko yang lebih komprehensif.
5. Penambahan variabel determinan epidemiologi untuk memperkaya pemetaan profil risiko.
6. Pengembangan lanjutan sistem S4 sebagai alat analitik untuk berbagai penyakit menular.

## DAFTAR PUSTAKA

- Amalina, A., Perdana, A. A., Rayendra, M. F., Rosmawati, E., Rahmahdhani, A., Maksuk, D., Ashari, A. E., Yuniati, F., & Aini, R. (2025). *Epidemiologi Penyakit Menular*.
- Bastian, A., Sujadi, H., & Febrianto, G. (2018). "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)". *Jurnal Sistem Informasi (JSI)*, 14(1), 26-32.
- Faisal, M. R. (2016). *Seri Pemograman Pengenalan Bahasa Pemograman R*.
- Faridah, R. N., Wulandari, I. R., & Fauziah, Y. (2024). "Regional Clustering Based on Infectious Disease Spread in DKI Jakarta Using the K-Means Algorithm". *Jurnal Teknologi Informasi dan Komunikasi*, 19(3), 12-20.
- Fuady, A., Arifin, B., & Yunita, J. (2021). "Analisis Spasial Pengaruh Kepadatan Hunian Terhadap Persebaran Kasus Tuberkulosis Paru". *Cita Husada Journal of Community Health*, 3(1), 25-34.
- Hadi Prakoso, B., Rachmawati, E., Rachmatta Putro Mudiono, D., Vestine, V., Eko Julianto Suyoso, G., Studi Manajemen Informasi Kesehatan, P., Kesehatan, J., Negeri Jember, P., & Timur, J. (2023). Klasterisasi Puskesmas dengan K-Means Berdasarkan Data Kualitas Kesehatan Keluarga dan Gizi Masyarakat. Dalam *Jurnal Buana Informatika* (Vol. 14, Nomor 1).
- Hart, J. T. (1971). "The Inverse Care Law". *The Lancet*, 297(7696), 405-412.
- Mccaffrey, J. (2017). *R Programming Succinctly*. www.syncfusion.com.
- Mirkin, B. (2012). *Clustering: A Data Recovery Approach*.
- Noor, M., Arisanti, N., & Puspitasari, I. (2021). "Pengaruh Akses Pelayanan Kesehatan Primer Terhadap Tingkat Kesehatan Masyarakat di Indonesia". *Jurnal Kebijakan Kesehatan Indonesia*, 10(1), 35-42.
- Nur Khomarudin, A. (2003). *Teknik Data Mining : Algoritma K-Means Clustering*.  
<https://agusnkhom.wordpress.com>

- Putri, R. A., Wibowo, A., & Nugraha, A. L. (2023). "Spatial Analysis of Dengue Disease and Health Facilities in Jakarta Province". *E-Journal Kesehatan Masyarakat*, 11(2), 88-95.
- Santoso, H., Rosliany, Hapsari, R. B., Nasir, A. M., Purwanto, E., Jaya, I., Delianna, J., Indriani, N., Rosmaniar, Soitawati, Muhiriyah, E., & Nugroho, G. W. (2011). *Pedoman Penyelidikan dan Penanggulangan Kejadian Luar Biasa Penyakit Menular dan Keracunan Pangan (Pedoman Epidemiologi Penyakit) Edisi Revisi Tahun 2011*.
- Tanjung, R., Nurmaini, S., & Zulkarnain, Z. (2021). "The Spatial Distribution of Pulmonary Tuberculosis in Kabanjahe District, Karo Regency, North Sumatera, Indonesia". *Open Access Macedonian Journal of Medical Sciences*, 9(E), 1-6.
- Wahyudi, M., Masitha, Saragih, R., & Solikhun. (2020). *Data Mining : Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*.
- Wijayanti, S. P., Porphyre, T., Chase-Topping, M., Rainey, S. M., McFarlane, M., Schnettler, E., ... & Kohl, A. (2016). "Dengue in Java, Indonesia: Relevance of Climate, Land Use, and Public Health Facilities". *International Journal of Environmental Research and Public Health*, 13(3), 305.

## LAMPIRAN

### 1. Code Definisi Kelas S-4

```
# Definisi Kelas S4 (Blueprint Objek)
setClass("kmeansSumbagsels4",
  slots = list(
    # Menyimpan vektor label hasil pengelompokan (1, 2, atau 3)
    clusters = "numeric",

    # Menyimpan data frame lengkap (gabungan penyakit & fasilitas)
    # Slot ini krusial untuk analisis Gap Analysis (Supply vs Demand)
    data_asli = "data.frame",
    # Menyimpan label nama wilayah (Kabupaten/Kota)
    kab_kota = "character",
    # Menyimpan matriks pusat klaster yang telah dinormalisasi
    centroids_norm = "matrix"
  )
)
```

### 2. Code Method Summary

```
setMethod("summary", "kmeansSumbagsels4",
  function(object) {
    cat("ANALISIS PROFIL RISIKO & KAPASITAS(S4)\n")

    cat("\n[1] Ukuran Cluster (Demand):\n")
    print(table(Cluster = object@clusters))

    #Ambil data lengkap dari slot S4
    data_full <- object@data_asli
    data_full$Cluster <- object@clusters

    cat("\n[2] Rata-rata Beban Penyakit (DEMAND):\n")
    #Pilih kolom penyakit utama untuk ditampilkan
    cols_penyakit <- c("TBC", "HIV", "Kusta_per_10000", "DBD_per_100000", "Cluster")
    profil_penyakit <- aggregate(. ~ Cluster, data = data_full[, cols_penyakit], FUN = mean)
    print(round(profil_penyakit, 2))

    cat("\n[3] Rata-rata Ketersediaan Fasilitas (SUPPLY):\n")
    #pilih kolom fasilitas untuk evaluasi kapasitas
    cols_faskes <- c("Total_RS", "Total_Puskesmas", "Klinik", "Posyandu", "Cluster")
    profil_faskes <- aggregate(. ~ Cluster, data = data_full[, cols_faskes], FUN = mean)
    print(round(profil_faskes, 2))
  }
)
```

### 3. Code Method Plot

```
setMethod("plot", signature(x = "kmeansSumbagselS4", y = "missing"),
  function(x) {

    data_plot <- x@data_asli
    data_plot$Cluster <- as.factor(x@clusters)
    data_plot$KabupatenKota <- x@kab_kota

    warna_kustom <- c("1" = "#4DAF4A", "2" = "#FF7F00", "3" = "#E41A1C")

    p <- ggplot(data_plot, aes(x = Kusta_per_100000, y = DBD_per_100000, color = Cluster)) +
      geom_point(size = 3, alpha = 0.8) +
      geom_text(aes(label = KabupatenKota), vjust = -1, size = 2.5, check_overlap = TRUE) +
      ggtitle("Peta Risiko: Kusta vs DBD (Lampung & Sumsel)") +
      xlab("Kasus Kusta (per 100.000 Penduduk)") +
      ylab("Kasus DBD (per 100.000 Penduduk)") +
      theme_minimal() +
      scale_color_manual(values = warna_kustom)

    print(p)
  }
)
```