**DOKUMEN PROYEK**

**12S4054 - PENAMBANGAN DATA**

**Segmentasi Wilayah Peserta BPJS Berdasarkan Jumlah dan Pertumbuhan Peserta Menggunakan K-Means Clustering**

**Disusun Oleh:**

| **12S22007** | **Tamara Y Sianipar** |
| --- | --- |
| **12S22018** | **Jesica A Siburian** |
| **12S22021** | **Krisnia Calysta Siahaan** |
| **12S22042** | **Ruth Septiana Simanullang** |

****

**PROGRAM STUDI SARJANA SISTEM INFORMASI**

**FAKULTAS INNFORMATIKA DAN TEKNIK ELEKTRO (FITE)**

**INSTITUT TEKNOLOGI DEL**

**(2024/2025)**

**BAB I**

**Pendahuluan**

### **1.1 Latar Belakang**

BPJS Kesehatan merupakan lembaga yang menyelenggarakan program jaminan sosial kesehatan bagi seluruh masyarakat Indonesia. Seiring dengan meningkatnya jumlah peserta, BPJS menghadapi tantangan dalam memastikan distribusi layanan yang merata dan efisien di seluruh wilayah. Salah satu beban terbesar yang ditanggung oleh sistem kesehatan nasional adalah penyakit tidak menular seperti **Diabetes Mellitus (DM)**, yang memerlukan penanganan jangka panjang dan berkelanjutan.

Untuk mendukung efisiensi dan pemerataan layanan, sangat penting bagi BPJS untuk mengetahui wilayah-wilayah yang memiliki **jumlah peserta tinggi** serta **pertumbuhan signifikan** dari tahun ke tahun. Dengan begitu, BPJS dapat mengalokasikan sumber daya secara lebih tepat sasaran, terutama untuk penyakit kronis seperti DM.

Salah satu pendekatan yang dapat digunakan untuk memahami pola distribusi peserta adalah melalui **segmentasi wilayah** berdasarkan karakteristik jumlah peserta dan tren pertumbuhan. Metode **K-Means Clustering** dapat digunakan untuk mengelompokkan wilayah ke dalam kategori seperti pertumbuhan rendah, sedang, dan tinggi, sehingga dapat mendukung pengambilan keputusan strategis secara berbasis data.

### **1.2 Rumusan Masalah**

Berdasarkan latar belakang di atas, rumusan masalah dalam proyek ini adalah:

1. Bagaimana mengelompokkan wilayah berdasarkan jumlah peserta dan pertumbuhan peserta BPJS dengan penyakit Diabetes Mellitus?
2. Wilayah mana saja yang masuk dalam kategori pertumbuhan peserta tinggi, sedang, atau rendah?
3. Bagaimana hasil segmentasi ini dapat dimanfaatkan dalam perencanaan layanan dan alokasi sumber daya?

### **1.3 Tujuan**

Adapun tujuan dari proyek ini dibagi menjadi dua, yaitu:

* **Tujuan Bisnis**:  
  + Menganalisis pola pemanfaatan layanan kesehatan oleh peserta penderita Diabetes Mellitus dan Tuberkulosis.
  + Menilai efektivitas pembiayaan dan kualitas distribusi layanan kesehatan berdasarkan jenis fasilitas dan wilayah.
* **Tujuan Teknis**:  
  + Menerapkan algoritma **K-Means Clustering** untuk mengelompokkan wilayah berdasarkan data peserta.
  + Menyajikan hasil segmentasi dalam bentuk visualisasi yang informatif dan mudah dipahami.

### **1.4 Manfaat**

Proyek ini diharapkan dapat memberikan manfaat sebagai berikut:

* Memberikan wawasan berbasis data kepada BPJS Kesehatan dalam merumuskan kebijakan alokasi layanan kesehatan untuk penyakit DM.
* Membantu mengidentifikasi wilayah prioritas untuk intervensi dan penambahan fasilitas kesehatan.
* Mendorong efisiensi pembiayaan dan peningkatan mutu layanan berdasarkan distribusi peserta.

### **1.5 Ruang Lingkup**

Analisis dalam proyek ini akan difokuskan pada data kepesertaan BPJS Kesehatan tahun 2021 yang terkait dengan penyakit Diabetes Mellitus. Variabel utama yang digunakan dalam segmentasi meliputi jumlah peserta per wilayah dan indikasi pertumbuhan (jika tersedia dalam data historis atau diturunkan dari data terkait). Segmentasi dilakukan berdasarkan wilayah administratif seperti kabupaten/kota atau provinsi, tergantung pada cakupan data.

**BAB II**

**Business Understanding**

### **2.1 Objektif Bisnis**

BPJS Kesehatan memiliki peran vital dalam menyediakan layanan kesehatan yang merata dan terjangkau bagi seluruh penduduk Indonesia. Untuk mencapai efisiensi dan efektivitas layanan, dibutuhkan analisis mendalam terhadap data pemanfaatan layanan kesehatan, terutama untuk penyakit dengan beban pembiayaan tinggi seperti Diabetes Mellitus (DM) dan Tuberkulosis (TB).

Objektif utama dari proyek ini adalah untuk menganalisis pola layanan dan kepesertaan BPJS Kesehatan yang berkaitan dengan penyakit DM dan TB. Fokus utama diarahkan pada:

* Pola pemanfaatan layanan kesehatan oleh peserta penderita DM dan TB.
* Jumlah peserta aktif dan jenis fasilitas kesehatan yang digunakan (Fasilitas Kesehatan Tingkat Pertama/FKTP dan Fasilitas Kesehatan Rujukan Tingkat Lanjutan/FKRTL).
* Efektivitas pembiayaan berdasarkan jenis diagnosis (kapitasi atau non-kapitasi).
* Segmentasi peserta berdasarkan wilayah, status kepesertaan, dan jenis layanan yang diterima.

Dengan memahami aspek-aspek tersebut, pengambil kebijakan dapat menyusun strategi pelayanan dan pembiayaan yang lebih tepat sasaran.

### **2.2 Tujuan Teknis**

Dalam mendukung tercapainya tujuan bisnis, proyek ini memiliki tujuan teknis sebagai berikut:

* Menyusun analisis deskriptif dan prediktif berdasarkan data pemanfaatan layanan peserta BPJS yang menderita DM dan TB.
* Mengelompokkan pasien berdasarkan jenis layanan, diagnosis, dan status kepesertaan.
* Menentukan segmentasi wilayah atau klasifikasi peserta berdasarkan karakteristik tertentu.
* Merancang visualisasi data yang dapat memberikan insight mendalam bagi pengambil keputusan.
* Mengintegrasikan berbagai sumber data (misalnya kepesertaan dan pembiayaan) untuk memberikan gambaran yang lebih utuh dan strategis.

### **2.3 Rencana Proyek**

Agar tujuan dapat tercapai secara sistematis, proyek ini dirancang dengan tahapan-tahapan sebagai berikut:

1. Pemilihan Dataset  
    Memilih salah satu kategori dataset terlebih dahulu, misalnya dataset kontekstual terkait penyakit Diabetes Mellitus (DM), untuk dianalisis lebih lanjut.
2. Eksplorasi Struktur Data (.dta)  
    Membaca dan memahami struktur dataset berformat .dta, termasuk kolom-kolom utama seperti ID peserta, jenis kelamin, wilayah, status aktif, jenis faskes, dan sebagainya.
3. Data Understanding Lanjutan  
    Melakukan eksplorasi data lebih mendalam untuk memahami karakteristik data, distribusi variabel, serta kualitas data yang tersedia.
4. Analisis dan Visualisasi  
    Melakukan analisis statistik deskriptif dan prediktif, serta menyusun visualisasi data (grafik, peta, dashboard) untuk mendukung interpretasi hasil.
5. Penyusunan Insight dan Rekomendasi  
    Menyusun insight dari hasil analisis yang dapat digunakan untuk pengambilan keputusan, baik dalam hal pengelolaan layanan, kebijakan pembiayaan, maupun strategi segmentasi peserta.

**BAB III**

**DATA UNDERSTANDING**

**3.1 Mengumpulkan Data**

**3.2 Menelaah Data**

**3.3 Memvalidasi Data**