

# Klasterisasi Penyakit pada Data Klaim Rujukan Tingkat Lanjut BPJS Kesehatan Menggunakan Algoritma *Density-Based Spatial Clustering of Application with Noise*

Muklas Rivai<sup>1</sup>, Misbahul Huda<sup>2</sup>, Rosni\*<sup>3</sup>, Karina Sylfia Dewi<sup>4</sup>

<sup>1,2,3,4</sup>Program Studi Sains Aktuaria, Fakultas Sains, Institut Teknologi Sumatera, Jl. Terusan Ryacudu, Way Huwi,  
Kec. Jati Agung, Kabupaten Lampung Selatan, Lampung

Email: [muklas.rivai@at.itera.ac.id](mailto:muklas.rivai@at.itera.ac.id), [hmisbahul08@gmail.com](mailto:hmisbahul08@gmail.com), [rosni@at.itera.ac.id](mailto:rosni@at.itera.ac.id), [karina.dewi@at.itera.ac.id](mailto:karina.dewi@at.itera.ac.id)

**Abstract**— Over time and with the advancement of technology, an increasing number of disease-claim submissions have been received by Badan Penyelenggara Jaminan Sosial (BPJS) for Health, causing data accumulation to the point that the dataset can now be categorized as Big Data. One of the challenges of Big Data is that it cannot be processed using conventional methods, thus requiring specialized approaches such as data clustering. The purpose of this study is to determine the optimal number of clusters and to analyze the characteristics of the cluster groups. The type of data used is secondary data obtained from the BPJS Health database. The data used consists of claim data from Fasilitas Kesehatan Rujukan Tingkat Lanjutan (FKRTL) under BPJS Health from January 2019 to December 2020. The variables used include childbirth, accidents, catastrophic diseases, and other diseases. The stages of the clustering process include data normalization, parameter determination, application of the *Density-Based Spatial Clustering of Application with Noise (DBSCAN)* algorithm, and evaluation of cluster results using the silhouette index. The results of the clustering analysis on FKRTL claim data based on the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)*, show that there are three clusters and one noise cluster, with an average silhouette index of 0.6595942, indicating that the model has a medium structure. Cluster 1 consists of two members with dominant claim categories being accidents and other diseases, cluster 2 consists of 27 members with childbirth as the dominant claim category, cluster 3 consists of four members with catastrophic diseases and other diseases as the dominant claim categories, and the noise cluster consists of one member with childbirth as the dominant claim category.

**Key word**—Clustering; DBSCAN; FKRTL; ICD-10; Silhouette Index

**Abstrak**—Seiring waktu dan perkembangan zaman, semakin banyak pengajuan klaim penyakit yang masuk ke Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan dan menyebabkan penumpukan data sehingga kumpulan data tersebut sudah tergolong ke dalam Big Data. Salah satu hambatan dari Big Data ialah pengolahannya yang tidak bisa menggunakan cara biasa sehingga diperlukan cara khusus seperti *clustering* data. Tujuan dari penelitian ini ialah menentukan jumlah kluster optimal dan menganalisis karakteristik kelompok kluster. Jenis data yang digunakan adalah data sekunder yang diperoleh melalui database BPJS Kesehatan. Data yang digunakan adalah data klaim Fasilitas Kesehatan Rujukan Tingkat Lanjut (FKRTL) BPJS Kesehatan pada Januari 2019 hingga Desember 2020. Variabel yang digunakan adalah persalinan, kecelakaan, penyakit katastrofik, dan penyakit lainnya. Tahapan pelaksanaan clustering meliputi normalisasi data, penentuan parameter, penerapan algoritma *Density-Based Spatial Clustering of Application with Noise (DBSCAN)* dan evaluasi hasil kluster menggunakan indeks *silhouette*. Hasil analisis *clustering* pada data klaim FKRTL berdasarkan *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)* menunjukkan terdapat tiga kluster dan satu kluster *noise* dengan nilai rata-rata indeks *silhouette* sebesar 0,6595942 yang berarti model memiliki struktur *medium*. Kluster 1 terdiri dari dua anggota dengan kategori klaim dominan ialah kecelakaan dan penyakit lainnya, kluster 2 terdiri dari 27 anggota dengan kategori klaim dominan adalah persalinan, kluster 3 terdiri dari empat anggota dengan kategori klaim dominan adalah katastrofik dan penyakit lainnya, dan kluster *noise* terdiri dari satu anggota dengan kategori klaim dominan adalah persalinan.

**Kata kunci**— DBSCAN; FKRTL; ICD-10; Indeks *Silhouette*; Klasterisasi

## I. PENDAHULUAN

BPJS Kesehatan atau Badan Penyelenggara Jaminan Sosial Kesehatan merupakan suatu badan hukum dalam menyelenggarakan program jaminan sosial dibidang kesehatan yang berlandaskan Undang Undang Nomor 24 Tahun 2011 tentang jaminan sosial [1]. BPJS Kesehatan memiliki satu program utama yang biasa dikenal dengan JKN atau Jaminan Kesehatan Nasional. JKN merupakan program yang dibuat dengan tujuan memberikan kepastian jaminan kesehatan yang menyeluruh bagi seluruh rakyat Indonesia untuk dapat hidup sehat, produktif dan sejahtera. Program JKN mulai berlaku sejak 1 Januari 2014 dan bersifat wajib untuk seluruh rakyat Indonesia yang nantinya seluruh rakyat Indonesia wajib untuk menyisihkan sebagian uangnya untuk jaminan kesehatan di masa depan [2]. Pelayanan kesehatan yang diberikan melalui JKN berupa bentuk pengobatan dan perawatan. Sejak peluncurannya pada tahun 2014, Program JKN telah mengalami perkembangan yang signifikan dalam memberikan layanan kesehatan kepada seluruh penduduk Indonesia. JKN BPJS Kesehatan berupaya untuk menyediakan akses kesehatan yang lebih merata dan berkualitas bagi masyarakat, dengan memastikan bahwa setiap warga negara mendapatkan perlindungan finansial dan pelayanan medis yang layak. Pada Program JKN peserta mendapat tiga tingkat fasilitas kesehatan, salah satunya adalah Fasilitas Kesehatan Rujukan Tingkat Lanjutan (FKRTL). FKRTL adalah fasilitas Kesehatan yang melakukan pelayanan Kesehatan per orang yang bersifat spesialis atau subspecialis yang meliputi rawat jalan Tingkat lanjut, rawat inap Tingkat lanjut, dan rawat inap di ruang perawatan khusus [3][4]. Pasien dirujuk ke FKRTL jika fasilitas Kesehatan Tingkat pertama dan kedua tidak sanggup menangani penyakit pasien.

Permasalahan terkait jaminan kesehatan yang terjadi pada setiap daerah di Indonesia tidak sama. Perbedaan tersebut akan berdampak pada kebijakan yang akan dilaksanakan disetiap daerah. Kemudian perkembangan teknologi dimana semakin banyak data yang dihasilkan, BPJS Kesehatan mulai memanfaatkan potensi Big Data pada data kesehatan yang terkumpul guna meningkatkan efisiensi dan kualitas layanan. Data yang didapat bersumber dari berbagai aspek, termasuk klaim pengobatan, riwayat medis, dan profil peserta. Himpunan data yang banyak ini nantinya akan diproses dan diolah menjadi pedoman dalam mengidentifikasi pola kesehatan, pemahaman risiko, serta kebutuhan layanan kesehatan yang spesifik untuk setiap kelompok. Pengelompokan provinsi untuk mengelompokkan karakteristik yang sama dan pengolahan Big Data salah satunya yang dapat digunakan adalah *clustering* [5][6][7][8].

Metode *clustering* memiliki berbagai variasi yang masing-masing memiliki kelebihan dan kekurangan tergantung pada sifat data yang dihadapi. Salah satunya ialah *Density Based Spatial Clustering of Application with Noise (DBSCAN)*. Klaster atau kelompok yang diperoleh biasanya didasarkan pada suatu faktor, contohnya daerah atau wilayah seperti penelitian yang dilakukan oleh Siti Sundari dkk [9]. Penelitian yang digarap berupa pengelompokan provinsi di Indonesia berdasarkan data imunisasi penyakit campak balita pada tahun 2004 hingga 2017 menggunakan metode K-Medoids dengan luaran yang diperoleh tiga kelompok (klaster) yang terdiri dari klaster rendah (2 provinsi), klaster sedang (30 provinsi) dan klaster tinggi (2 provinsi). Dengan adanya penelitian seperti ini, pemerintah Indonesia dapat mengambil langkah efektif dalam memberantas penyakit campak seperti melakukan sosialisasi dan pemerataan dalam pemberian imunisasi campak balita terutama terhadap kelompok klaster rendah yang berisi Provinsi Kalimantan Utara dan Provinsi Papua.

Klasterisasi provinsi di bidang kesehatan memiliki manfaat yang besar sebagai pedoman dalam menentukan kebijakan dan langkah efektif yang akan diambil. Pemerintah dan instansi yang bergerak di bidang kesehatan seperti BPJS Kesehatan dapat meningkatkan efisiensi, kualitas layanan kesehatan, serta tindakan promotif dan preventif secara nasional. Penelitian yang dilakukan oleh Nayuni Dwitri dkk [10] membahas tentang pengelompokan provinsi berdasarkan penyebaran kasus Covid-19 pada tahun 2020 menggunakan metode K-Means. Luaran yang didapat dari penelitian tersebut terdapat tiga klaster dengan rincian klaster satu sebagai penyebaran dalam jangka besar klaster dua sebagai penyebaran dalam jangka sedang dan klaster tiga sebagai penyebaran dalam jangka kecil. Berdasarkan

hasil penelitian tersebut, diperlukan penanganan khusus terhadap klaster satu yang berisi Provinsi DKI Jakarta guna menekan angka penyebaran virus Covid-19 di Indonesia.

Program yang dijalankan BPJS Kesehatan memiliki banyak pertimbangan baik dalam kebijakan berdasarkan penyakit maupun tindakan yang menunjang program itu sendiri. Penentuan keputusan dalam kebijakan berdasarkan penyakit tentunya diperlukan banyak pertimbangan. Salah satunya adalah pengelompokan daerah berdasarkan banyaknya penyakit yang pastinya sangat berbeda keadaan dan penanganannya. Berdasarkan uraian tersebut, penelitian ini berfokus pada klasterisasi ICD-10 data klaim FKRTL BPJS Kesehatan dengan menerapkan metode DBSCAN. Tujuan penelitian adalah menentukan jumlah klaster yang optimal berdasarkan data klaim FKRTL BPJS Kesehatan serta menganalisis karakteristik kelompok klaster provinsi di Indonesia. Hasil penelitian diharapkan dapat memberikan kontribusi dalam perumusan kebijakan promotif dan preventif di bidang kesehatan.

## II. METODE PENELITIAN

### A. Sumber Data

Jenis data yang digunakan penulis ialah data sekunder yang diperoleh melalui *database* Badan Penyelenggara Jaminan Sosial Kesehatan. Data yang digunakan adalah data Klaim FKRTL BPJS Kesehatan pada Januari 2019 hingga Desember 2020. Penulis menggunakan satu teknik pengumpulan data yaitu Dokumentasi. Teknik dokumentasi diartikan sebagai cara pengumpulan data, dengan mencatat atau mengambil data yang sudah ada dalam dokumen atau arsip [11]. Data yang diperoleh berasal dari arsip data Badan Penyelenggara Jaminan Sosial Kesehatan. Pada penelitian ini populasi yang digunakan ialah klaim FKRTL BPJS Kesehatan.

### B. Data Penelitian

Peneliti mengklasifikasi ulang 22 kategori penyakit pada ICD-10 menjadi empat variabel klasifikasi klaim guna memperkecil jumlah variabel. Empat variabel tersebut adalah Persalinan, Kecelakaan, Penyakit Katastropik, dan Penyakit Lainnya. Variabel persalinan merupakan variabel yang berisikan klaim yang berkaitan dengan persalinan pada ibu hamil. Biasanya kategori persalinan berada pada kode O60 hingga O80 dalam ICD-10. Variabel kecelakaan merupakan variabel yang berisikan klaim yang berkaitan dengan kecelakaan transportasi darat, laut dan udara. Kategori kecelakaan berada pada kode V01 hingga V99 dalam ICD-10. Variabel penyakit katastropik berisikan klaim yang berkaitan dengan penyakit-penyakit katastropik seperti jantung, gagal ginjal, diabetes dan lainnya. Variabel ini rata-rata berada pada kode C00-C97, D37-D76, dan I00-I79 pada kode ICD-10. Untuk variabel penyakit lainnya berisikan klaim penyakit yang tidak termasuk kedalam ketiga variabel sebelumnya.

### C. Prosedur Penelitian

Pada analisis penelitian ini, terdapat tahapan-tahapan yang harus dilakukan, yaitu:

#### 1. *Pre-processing* Data.

Sebelum melakukan pengolahan data, peneliti akan memilah variabel yang digunakan dari data sampel keseluruhan menggunakan perangkat lunak R Studio. Data sampel yang digunakan pada penelitian ini ialah data klaim FKRTL BPJS Kesehatan. Dalam data klaim FKRTL BPJS Kesehatan, terdapat 55 variabel yang dimulai dari PSTV01 tentang “Nomor Peserta” hingga FKL43 tentang “Tarif *special investigation*”. Variabel yang diambil dari data sampel tersebut ialah variabel FKL03 tentang Tanggal datang kunjungan FKRTL, variabel FKL05 tentang Provinsi peserta, dan variabel FKL15A tentang kode diagnosis masuk ICD-10. Tahapan selanjutnya ialah *grouping* data menjadi empat variabel klaim berdasarkan FKL15A atau kode diagnosis masuk ICD-10. Selain pemilahan variabel dan *grouping* data, peneliti akan melakukan transformasi data dengan membagi jumlah data klaim dengan data sampel kepesertaan BPJS Kesehatan tahun 2019 dan 2020. Hal ini dilakukan agar meminimalisir perbedaan skala yang besar antar variabel yang digunakan. Hasil pembagian tersebut kemudian dikalikan dengan 100% hingga satuan akhir yang didapat berupa persentase. Lalu hasil *pre-processing* data akan dilanjutkan ke proses *clustering*.

## 2. Proses clustering.

Pada tahapan ini, data yang sudah didapat dari data *pre-processing* akan diolah menggunakan metode DBSCAN dengan bantuan R Studio. Proses *clustering* dimulai dari menentukan beberapa nilai *Minimum Points* (MinPts) dan *Epsilon* (Eps). Penentuan nilai MinPts dan Eps dapat dilakukan dengan fungsi *kNNdisplot*. Nilai *k* ditentukan oleh peneliti yang nantinya akan digunakan sebagai MinPts pada proses *clustering*. Nilai yang terdapat di posisi *knee* menjadi acuan utama pemilihan nilai Eps yang *optimum*. Posisi *knee* merupakan suatu keadaan dimana terdapat perubahan yang signifikan pada nilai Eps di suatu sampel atau titik data. Setelah mendapat beberapa nilai MinPts dan Eps, dilakukan pembuatan kluster menggunakan metode DBSCAN.

DBSCAN merupakan salah satu metode clustering yang mengelompokkan data berdasarkan kepadatan titik data dalam set data, dengan mengidentifikasi wilayah yang padat dan titik data yang terisolasi [12]. DBSCAN biasanya diterapkan pada data yang memiliki banyak *noise* atau *outlier*. Hal ini dilakukan karena data *noise* yang diproses model ini tidak dianggap ke kluster manapun. Untuk melakukan proses *clustering* pada metode DBSCAN, diperlukan dua parameter input yakni Eps dan MinPts. Eps merupakan jarak maksimal yang diizinkan antar dua data dalam satu kluster. Sedangkan MinPts merupakan jumlah minimal data dalam jarak radius *epsilon* agar terbentuk kluster [13][14].

$$N_{Eps}(x) = \{y \in D \mid d(x, y) \leq Eps\} \quad (1)$$

$N_{Eps}(x)$  : titik sekitar dari  $x$  dalam radius  $Eps$   
 $D$  : gugus data  
 $dist(x, y)$  : jarak *Euclidean* dari objek  $x$  dan  $y$   
 $Eps$  : radius atau ambang batas.

Untuk metode jarak yang digunakan biasanya menggunakan jarak *Euclidean* yang didefinisikan sebagai berikut [15].

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

Beberapa istilah dalam metode DBSCAN antara lain [16]:

1. *Core*: merupakan titik pusat dalam suatu kluster yang didasari oleh densitas dimana ada sejumlah titik yang harus berada dalam Eps dan MinPts yang telah ditentukan.
2. *Border*: merupakan titik yang menjadi Batasan dalam kawasan titik pusat (*core*).
3. *Noise*: merupakan titik yang tidak terjangkau oleh *core* dan bukan merupakan *border*.

$$Noise = \{x \in D \mid \forall i : x \notin C_i\} \quad (3)$$

dimana  $D$  merupakan gugus data, dan  $C_i$  merupakan kluster ke- $i$

4. Densitas Terjangkau Langsung (*Directly density-reachable*): merupakan sebuah titik yang terhubung secara langsung dengan titik pusat (*core*)

$$x \in N_{Eps}(y) \wedge |N_{Eps}(y)| \geq MinPts \quad (4)$$

$N_{Eps}(y)$  : titik sekitar  $y$  dalam radius  $Eps$

$MinPts$  : minimal titik dalam kluster

5. Densitas terjangkau (*density-reachable*): merupakan titik yang terhubung secara tidak langsung dengan titik pusat (*core*)
6. Densitas Terhubung: merupakan titik yang saling terhubung satu sama lain oleh titik lain.

Saat membuat kluster dengan DBSCAN, data akan dikelompokkan dengan tetangganya. Sepasang amatan dikatakan bertetangga jika jarak antara kedua pengamatan lebih kecil dari sama dengan nilai epsilon. Secara umum, cara kerja dari algoritma DBSCAN adalah sebagai berikut:

1. Menentukan nilai MinPts dan Eps yang akan digunakan.
2. Memilih data *random*  $p$  sebagai pusat *core* (kluster awal).
3. Menghitung jarak setiap data input terhadap data " $p$ " menggunakan rumus jarak *euclidean*.

4. Mengelompokkan data yang *density-reachable* dengan amatan “p”.
  5. Jika amatan “p” adalah *border points* dan tidak ada amatan yang *density-reachable* dengan amatan “p”, maka lanjutkan pada amatan lainnya.
  6. Melakukan perulangan dari langkah 3 hingga 6, sampai semua amatan diproses.
3. *Evaluasi model klaster*

Untuk mendapatkan satu model klaster yang optimal, peneliti akan melakukan evaluasi pada model klaster yang ada. Evaluasi model akan dilakukan menggunakan Indeks *silhouette*. Indeks *silhouette* digunakan untuk mengevaluasi jumlah klaster yang optimal. Untuk menghitung nilai indeks *silhouette* diperlukan jarak antar objek dengan menggunakan rumus *euclidean distance* [17]. Terdapat beberapa tahapan yang harus dilakukan untuk menghitung nilai indeks *silhouette*, diantaranya adalah:

1. Menghitung rata-rata jarak dari objek  $i$  dengan seluruh objek yang ada dalam satu kelompok dimana nilai rata-ratanya akan disebut  $a_i$ .
2. Menghitung rata-rata jarak dari objek  $i$  dengan objek yang berada di kelompok lainnya. Lalu dari semua jarak rata-rata, diambil nilai terkecil yang akan disebut  $b_i$ .
3. Menentukan nilai indeks *silhouette* pada objek  $i$  dengan persamaan berikut:

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (5)$$

Hasil pengelompokan dikatakan baik jika nilai indeks *silhouette* bernilai positif ( $a_i < b_i$ ) dan  $a_i$  mendekati nilai 0. Namun terkadang nilai  $S_i$  dapat bernilai *negative*. Hal ini dapat diartikan bahwa struktur kelompok yang dihasilkan *overlapping* sehingga objek  $i$  lebih tepat dimasukkan ke dalam kelompok yang lain [17]. Berikut adalah nilai indeks *silhouette* berdasarkan Kaufman dan Rousseeuw [18]:

Tabel 1. Nilai *Coefficient Silhouette*

Nilai SC	Keterangan
$0.7 < SC \leq 1$	Struktur Kelompok Kuat
$0.5 < SC \leq 0.7$	Struktur Kelompok Medium
$0.25 < SC \leq 0.5$	Struktur Kelompok Lemah
$SC \leq 0.25$	Tidak ada Struktur

Ket: SC = *Coefficient Silhouette* yang didapat dari rata-rata nilai indeks *silhouette*.

#### 4. Analisis klaster

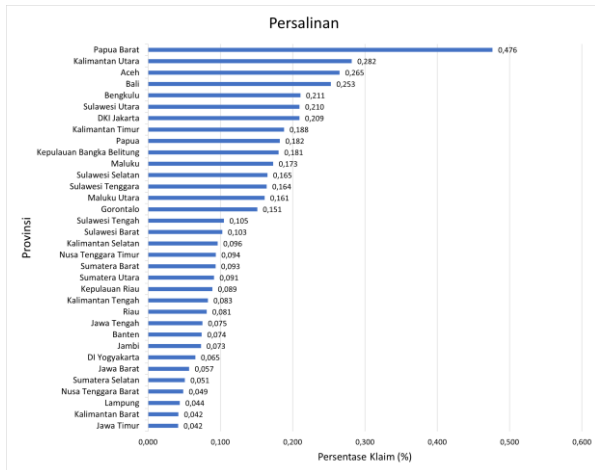
Setelah mendapat model klaster yang paling optimal, peneliti akan melakukan pembuatan klaster menggunakan metode DBSCAN. Setelah mendapatkan hasil klaster, peneliti akan melakukan analisis berdasarkan karakteristik dari tiap klaster yang ada beserta karakteristik dari klaster *noise*.

### III. HASIL DAN PEMBAHASAN

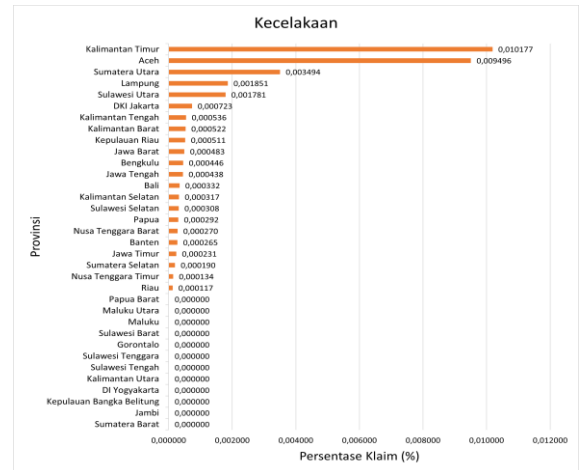
#### A. Statistika Deskriptif

Data yang digunakan pada penelitian ini adalah data klaim FKRTL BPJS Kesehatan pada periode 2019-2020. Peneliti mengambil variabel Provinsi, Tanggal Klaim dan Kode ICD-10 yang kemudian diolah hingga menghasilkan empat variabel klaim berupa variabel Persalinan, Kecelakaan, Penyakit Katastropik, dan Penyakit Lainnya. Nilai rata-rata persentase pengajuan klaim pada variabel Persalinan sebesar 0,137627%. Persentase pengajuan klaim terbesar pada variabel Persalinan berada di Provinsi Papua Barat dengan nilai 0,476463%. Salah satu faktor yang menyebabkan banyaknya persalinan ialah minimnya informasi terkait program Keluarga Berencana (KB). Berdasarkan publikasi yang diterbitkan oleh Kementerian Kesehatan, cakupan peserta KB aktif di Provinsi Papua Barat tergolong cukup rendah dengan nilai 29%. Hal ini menandakan masih banyaknya masyarakat Provinsi Papua Barat yang belum menyadari pentingnya perencanaan keluarga tersebut [19]. Persentase pengajuan klaim terkecil pada variabel Persalinan berada di Provinsi Jawa Timur dengan nilai 0,041623%. Salah satu penyebab sedikitnya jumlah klaim variabel persalinan di Provinsi Jawa Timur ialah kecilnya laju pertumbuhan

penduduk per tahun. Menurut Badan Pusat Statistik (BPS), laju pertumbuhan penduduk Provinsi Jawa Timur Tahun 2020 sebesar 0,92%. Grafik persentase klaim variabel persalinan per provinsi dapat dilihat pada Gambar 1.

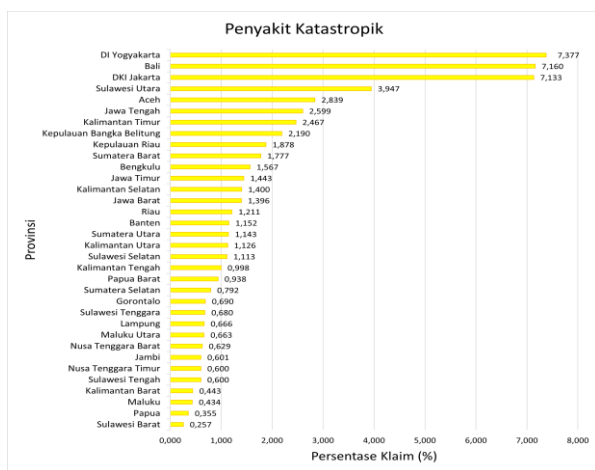


Gambar 1. Grafik Persentase Klaim Variabel Persalinan Per Provinsi

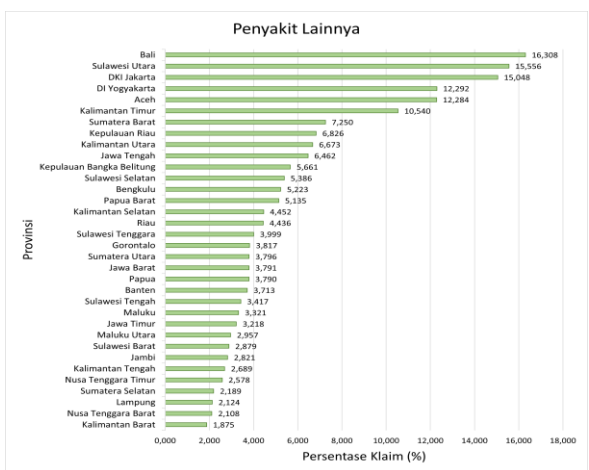


Gambar 2. Grafik Persentase Klaim Variabel Kecelakaan Per Provinsi

Berdasarkan Gambar 2, nilai rata-rata persentase pengajuan klaim pada variabel Kecelakaan sebesar 0,000968%. Persentase pengajuan klaim terbesar tercatat di Provinsi Kalimantan Timur dengan nilai 0,010177%, yang diduga dipengaruhi oleh kondisi jalan di wilayah tersebut. Sebaliknya, persentase pengajuan klaim terkecil sebesar 0% ditemukan di Provinsi Sumatera Barat, Jambi, Kepulauan Bangka Belitung, DI Yogyakarta, Kalimantan Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, dan Papua Barat, yang dapat dikaitkan dengan kondisi infrastruktur jalan yang relatif baik. Selanjutnya Visualisasi persentase klaim variabel Penyakit Katastropik per provinsi ditunjukkan pada Gambar 3. Pada variabel Penyakit Katastropik, nilai rata-rata persentase pengajuan klaim mencapai 1,772367%. Persentase tertinggi tercatat di Provinsi di Yogyakarta dengan nilai 7,376753%, didominasi oleh klaim penyakit Tipe 2 Diabetes Mellitus, Hipertensi Jantung, Gagal Jantung, dan Penyakit Ginjal Akut. Penyakit-penyakit tersebut termasuk kategori katastropik karena berpotensi menurunkan sistem imun, memicu komplikasi serius, serta memerlukan biaya pengobatan yang tinggi. Persentase terendah pada variabel ini ditemukan di Provinsi Sulawesi Barat dengan nilai 0,256538%.



Gambar 3. Grafik Persentase Klaim Variabel Penyakit Katastropik Per Provinsi



Gambar 4. Grafik Persentase Klaim Variabel Penyakit Lainnya Per Provinsi

Penyakit Katastropik Per Provinsi

Penyakit Lainnya Per Provinsi

Grafik persentase klaim variabel Penyakit Lainnya per provinsi ditunjukkan pada Gambar 4. Pada variabel Penyakit Lainnya, nilai rata-rata persentase pengajuan klaim tercatat sebesar 5,724027%. Klaim terbesar terdapat di Provinsi Bali dengan nilai 16,30812%. Beberapa penyakit dengan klaim tinggi pada tahun 2019–2020 di provinsi ini meliputi schizoprenia (628 klaim), epilepsi (718 klaim), katarak senilis (655 klaim), asma (598 klaim), nyeri panggul (710 klaim), demam dan penyakit lainnya (1606 klaim), serta tindakan cuci darah (1650 klaim). Sebaliknya, persentase klaim terendah pada variabel ini terdapat di Provinsi Kalimantan Barat dengan nilai 1,875224%. Kondisi tersebut dipengaruhi oleh rendahnya persentase jumlah peserta BPJS Kesehatan di provinsi tersebut, yaitu sekitar 49,98% dari total populasi.

B. Pre-processing Data

Pre-processing data adalah tahapan dari data mining yang bertujuan untuk mengolah data mentah menjadi data berkualitas yang siap untuk dilanjutkan ke proses selanjutnya. Tahapan dari pre-processing data yaitu seleksi dan pembersihan data serta transformasi data. Transformasi data dilakukan jika terdapat beberapa kondisi. Salah satu kondisi utamanya adalah terdapat perbedaan skala pada setiap variabel. Perbedaan skala yang besar pada variabel data yang ada dalam data penelitian ada pada variabel kecelakaan dan variabel penyakit lainnya. Hasil dari transformasi data yang dilakukan berupa persentase jumlah klaim pada data yang digunakan peneliti. Data klaim yang ada dibagi dengan jumlah sampel peserta tiap provinsi pada tahun 2019 dan 2020 yang didapat dari Data Kepesertaan BPJS Kesehatan tahun 2020 dan kemudian dikali dengan 100 persen.

Tabel 2. Data klaim FKRTL Setelah Transformasi

Provinsi	Persalinan (%)	Kecelakaan (%)	Katastropik (%)	Penyakit Lainnya (%)
Aceh	0,264952233	0,009496496	2,838977417	12,283717310
Sumatera Utara	0,091276340	0,003493831	1,143106702	3,796172383
Sumatera Barat	0,093413613	0	1,777039515	7,250459344
Riau	0,081177920	0,000116635	1,210903967	4,435860113
Jambi	0,073212965	0	0,600876147	2,820866658
Sumatera Selatan	0,050997557	0,000190053	0,792394332	2,189220574
Bengkulu	0,210806860	0,000445680	1,567458195	5,223375049
Lampung	0,044009973	0,001851280	0,665871738	2,124175444
Kepulauan Bangka Belitung	0,180658182	0	2,189870121	5,661436802
Kepulauan Riau	0,088953417	0,000511227	1,878246288	6,826407918
...	...	...	...	...
Papua	0,182199970	0,000292456	0,354748897	3,790227297

C. Analisis Clustering

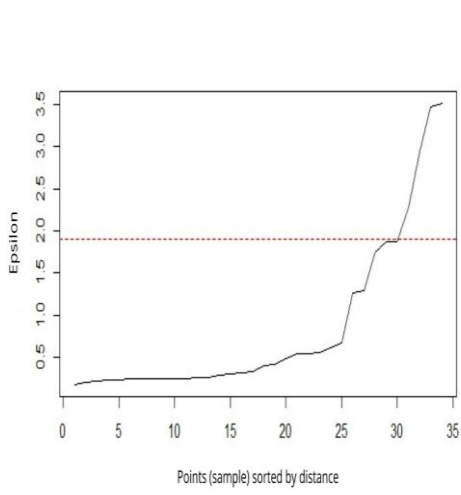
Analisis clustering dilakukan dengan bantuan software R Studio. Tahap pertama yang dilakukan ialah pemanggilan data excel yang telah di proses. Agar variabel data yang digunakan seimbang secara skala dalam analisis clustering, dilakukan standardisasi pada data yang digunakan dengan bantuan fungsi “scale” di software R Studio. Berikut merupakan tabel sebagian data klaim yang telah diseimbangkan secara skala.

Tabel 3. Data Klaim Setelah Standardisasi Data

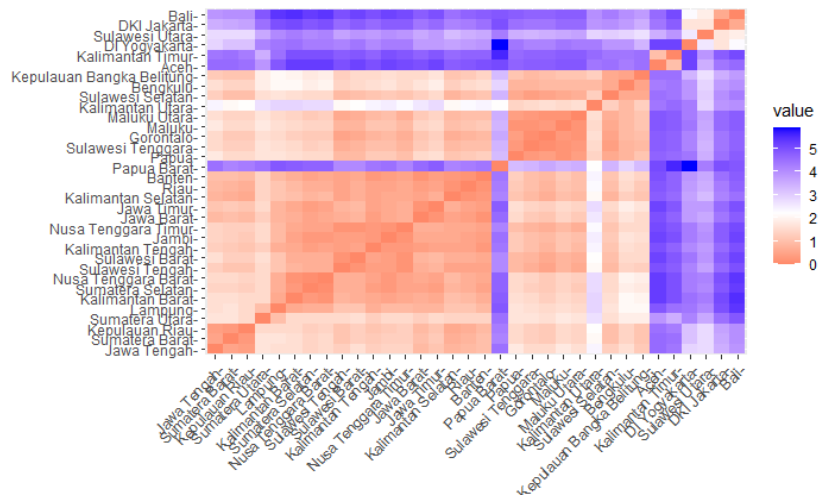
Provinsi	Persalinan	Kecelakaan	Katastropik	Penyakit Lainnya
Aceh	1,393594734	3,618343400	0,561326737	1,607010775
Sumatera Utara	-0,507312288	1,071592672	-0,331162154	-0,472291164
Sumatera Barat	-0,483919534	-0,410735223	0,002458793	0,373949491

Riau	-0,617840919	-0,361250512	-0,295482367	-0,315578717
Jambi	-0,705018461	-0,410735223	-0,616522790	-0,711224224
Sumatera Selatan	-0,948169229	-0,330101427	-0,515732175	-0,865966607
Bengkulu	0,800966029	-0,221646356	-0,107837940	-0,122651156
Lampung	-1,024649308	0,374707390	-0,582317445	-0,881901541
Kepulauan Bangka	0,470984514	-0,410735223	0,219719889	-0,015333592
Belitung	0,470984514	-0,410735223	0,219719889	-0,015333592
Kepulauan Riau	-0,532737008	-0,193837122	0,055721061	0,270064214
...	...	...	...	...
Papua	0,487859599	-0,286654986	-0,746052615	-0,473747607

Penetapan parameter DBSCAN berupa MinPts dan menentukan Eps menggunakan kNNdistplot. kNNdistplot digunakan untuk memilih Eps *optimum* dari jarak rata-rata setiap data ke-k tetangga terdekatnya yang disajikan pada Gambar 5. Nilai k yang dimaksud di kNNdistplot ialah nilai MinPts pada metode DBSCAN. Nilai yang terdapat di posisi *knee* menjadi acuan utama pemilihan nilai Eps yang optimum. Posisi *knee* merupakan suatu keadaan dimana terdapat perubahan yang signifikan pada nilai Eps di suatu sampel atau titik data. Berdasarkan kNNdistplot di bawah ini, MinPts dan Eps yang digunakan peneliti ialah MinPts = k = 2 dan Eps = 1,9. Selanjutnya, menghitung jarak antar provinsi berdasarkan empat variabel klaim menggunakan metode *euclidean*. Visualisasi jarak hasil perhitungan dapat dilihat pada Gambar 6. Jarak antar provinsi ditunjukkan dengan indikator *value* yang ada pada sebelah kanan grafik. Nilai 0 disimbolkan sebagai warna merah muda menandakan bahwa tidak ada jarak antar provinsi. Sebaliknya, nilai 6 disimbolkan sebagai warna biru menandakan bahwa terdapat jarak antar provinsi. Semakin dekat jarak antar provinsi, maka nilai indikatornya akan mendekati 0 dan berwarna merah muda. Semakin jauh jarak antar provinsi, maka nilai indikatornya akan mendekati 6 dan berwarna biru. Hal ini cukup membantu dalam pembuatan kelompok atau *clustering*. Provinsi yang berdekatan akan dianggap sebagai satu kelompok atau klaster yang sama dan provinsi yang berjauhan akan dianggap sebagai kelompok atau klaster yang berbeda.



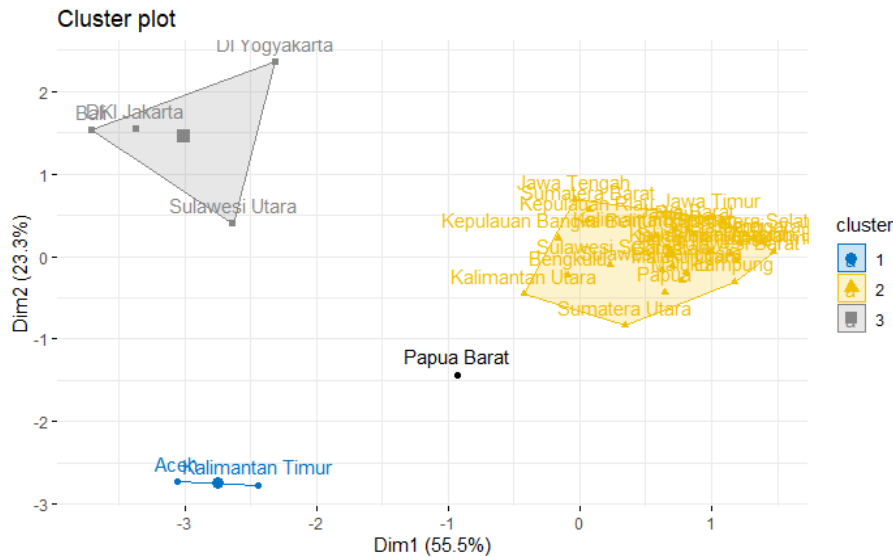
Gambar 5. kNNdistplot



Gambar 6. Visualisasi Jarak Antar Provinsi

Model klaster yang didapat pada penentuan kelompok atau klaster dengan metode DBSCAN divisualisasikan menjadi plot berdasarkan kelompok atau klaster yang terbentuk pada Gambar 7. Untuk mengevaluasi kualitas model *clustering* yang dihasilkan dengan menghitung rata-rata indeks *silhouette*. Rata-rata indeks *silhouette* mengevaluasi klaster berdasarkan jarak antar titiknya. Semakin dekat jarak titik dengan anggota klasternya, semakin baik struktur pada klaster tersebut. Semakin jauh jarak suatu titik dengan anggota klaster yang berbeda, semakin baik pula struktur model klasternya. Rata-rata

indeks *silhouette* yang dihasilkan dari model ini bernilai 0,6595942 yang berarti model *clustering* yang dihasilkan memiliki struktur *medium* atau standar. Hasil *clustering* dapat disajikan pada Gambar 7.



Gambar 7. Visualisasi Klaster Provinsi di Indonesia

Hasil *clustering* menunjukkan terdapat tiga klaster dan satu klaster *noise*. Terdapat dua anggota di dalam Klaster 1 yang berisikan Provinsi Aceh dan Provinsi Kalimantan Timur. Klaster 2 terdiri dari 27 anggota provinsi yang berisikan Provinsi Sumatera Barat, Provinsi Sumatera Utara, Provinsi Riau, Provinsi Sumatera Selatan, Provinsi Lampung, Provinsi Jambi, Provinsi Bengkulu, Provinsi Kepulauan Bangka Belitung, Provinsi Kepulauan Riau, Provinsi Banten, Provinsi Jawa Barat, Provinsi Jawa Tengah, Provinsi Jawa Timur, Provinsi Nusa Tenggara Timur, Provinsi Nusa Tenggara Barat, Provinsi Kalimantan Barat, Provinsi Kalimantan Tengah, Provinsi Kalimantan Utara, Provinsi Kalimantan Selatan, Provinsi Sulawesi Tengah, Provinsi Sulawesi Tenggara, Provinsi Gorontalo, Provinsi Sulawesi Barat, Provinsi Sulawesi Utara, Provinsi Maluku, Provinsi Maluku Utara, dan Provinsi Papua. Klaster 3 terdiri dari empat anggota provinsi yang berisikan Provinsi DKI Jakarta, Provinsi DI Yogyakarta, Provinsi Bali, dan Provinsi Sulawesi Utara. Sedangkan klaster *noise* hanya terdapat satu anggota provinsi yang berisikan Provinsi Papua Barat. Analisis dari hasil *clustering* ditampilkan pada tabel berikut:

Tabel 3. Analisis Hasil DBSCAN *Clustering*

Klaster	Jumlah Anggota Provinsi	Kategori Penyakit yang Banyak di Klaim
Klaster 1	2	Kecelakaan, Penyakit Lainnya
Klaster 2	27	Persalinan
Klaster 3	4	Katastropik, Penyakit Lainnya
Noise	1	Persalinan

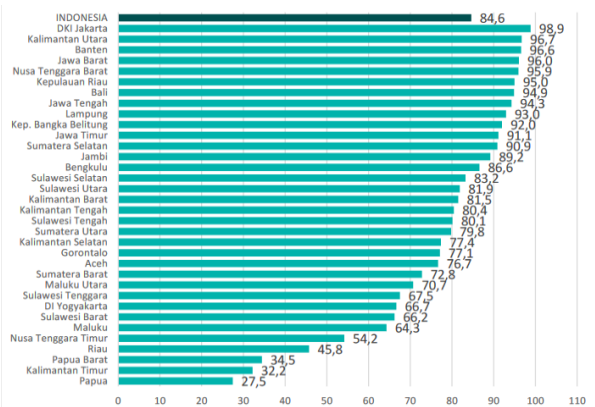
Berdasarkan Tabel 3 terlihat bahwa pada klaster 1, kategori penyakit yang banyak di klaim ialah kecelakaan dan penyakit lainnya. Seperti yang sudah dijelaskan di statistika deskriptif bahwasanya salah satu penyebab terjadinya kecelakaan di Provinsi Kalimantan Timur ialah kondisi dan permukaan jalan yang kurang baik. Berikut merupakan grafik kondisi jalan dan bentuk permukaan jalan di Provinsi Kalimantan Timur yang dilansir dari publikasi Statistik Transportasi Darat BPS [20].

Tabel 4. Panjang Jalan Kab./Kota di Kalimantan Timur Menurut Kondisi Jalan/Jenis Permukaan Jalan Tahun 2020

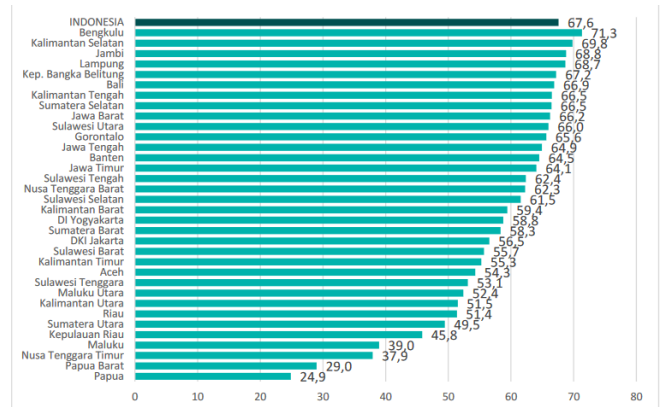
Provinsi	Baik (Km)	Sedang (Km)	Rusak (Km)	Rusak Berat (Km)	Total (Km)
Kalimantan Timur	3 162	3 024	2 464	2 216	10 866
Provinsi	Aspal (Km)	Kerikil (Km)	Tanah (Km)	Lainnya (Km)	Total (Km)
Kalimantan Timur	2 384	3 841	2 063	2 578	10 866

Selain itu, jumlah pelanggaran lalu lintas juga berpengaruh terhadap jumlah kasus kecelakaan. Dilansir dari laman resmi data Pemerintahan Provinsi Kalimantan Timur, didapati pada tahun 2020 pelanggaran lalu lintas terjadi sebanyak 37.323 kasus dengan tindakan penilangan dan peneguran sebanyak 101.416 kali [21]. Hal ini mencerminkan bahwa masih banyak pengguna jalan di Provinsi Kalimantan Timur yang kurang mematuhi lalu lintas ditambah dengan kondisi jalan yang kurang baik membuat persentase klaim FKRTL BPJS Kesehatan di variabel kecelakaan pada tahun 2019 dan 2020 menduduki peringkat tertinggi di Indonesia. Sedangkan di Provinsi Aceh, salah satu penyebab terjadinya kecelakaan ialah pelanggaran lalu lintas yang dilakukan oleh pengguna jalan. Untuk kategori penyakit lainnya, nilai rata-rata persentase klaim penyakit lainnya sebesar 11,41199%. Nilai persentase klaim penyakit lainnya di Provinsi Aceh sebesar 12,28372% dengan klaim penyakit terbanyaknya ialah dispepsia sebanyak 596 klaim, dorsalgia sebanyak 698 dan demam dan lainnya sebanyak 1419 klaim. Sedangkan nilai persentase klaim penyakit lainnya di Provinsi Kalimantan Timur sebesar 10,54026% dengan klaim penyakit terbanyaknya ialah penyakit pulpitis sebanyak 823 klaim, dispepsia sebanyak 950 dan cuci darah sebanyak 1815.

Selanjutnya ialah Klaster 2 dengan jumlah anggota 27 provinsi. Kategori penyakit yang banyak diklaim pada klaster 2 ialah kategori persalinan dan penyakit lainnya. Nilai rata-rata dari persentase klaim kategori persalinan sebesar 0,111557%. Untuk nilai terbesar dari persentase klaim kategori persalinan ialah Provinsi Kalimantan Utara dengan nilai 0,281784%. Salah satu penyebabnya ialah adanya faktor Cakupan Pelayanan Kesehatan Ibu Hamil K4 yang cukup besar di Kalimantan Utara, yakni sebesar 96.7%. Grafik cakupan pelayanan kesehatan ibu hamil k4 menurut provinsi tahun 2020 disajikan pada Gambar 8. Nilai terkecil dari persentase klaim kategori persalinan ialah Provinsi Jawa Timur dengan nilai 0,041623%. Salah satu faktor yang menyebabkan kecilnya nilai klaim kategori persalinan di Provinsi Jawa Timur ialah laju pertumbuhan penduduk yang kecil, yakni sebesar 0,92% pada tahun 2020. Selain itu cakupan peserta KB aktif di Provinsi Jawa Timur cenderung menengah dengan persentase 64,1% [19]. Hal ini dapat menekan laju pertumbuhan penduduk dan jumlah persalinan di Provinsi Jawa Timur. Grafik cakupan peserta kb aktif tahun 2020 disajikan pada Gambar 9.



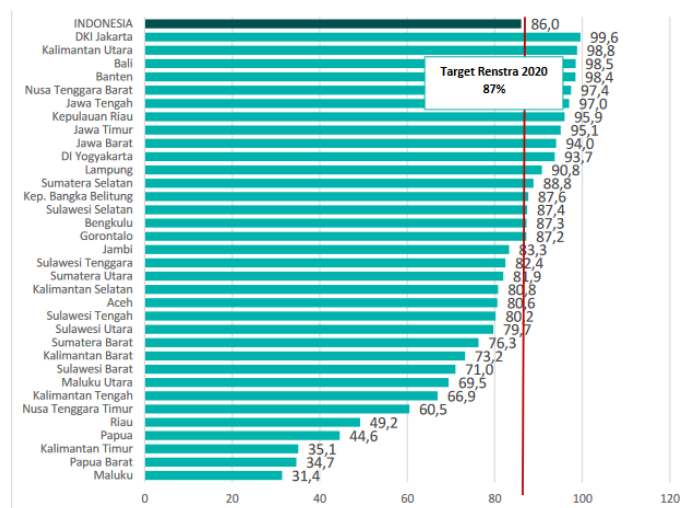
Gambar 8. Grafik Persentase Cakupan Pelayanan Kesehatan Ibu Hamil K4 Menurut Provinsi Tahun 2020



Gambar 9. Grafik Persentase Cakupan Peserta KB Aktif di Indonesia Tahun 2020

Kategori penyakit yang banyak diklaim pada klaster 3 ialah penyakit katastrofik. Rata-rata persentase pengajuan klaim penyakit katastrofik sebesar 6,404267%. Dari keempat anggota provinsi di Klaster 3, persentase pengajuan klaim penyakit katastrofik tertinggi adalah Provinsi DI Yogyakarta dengan besar nilai 7,376753%. Sedangkan persentase pengajuan klaim penyakit katastrofik terendah ada pada Provinsi Sulawesi Utara dengan besar nilai 3,946548%. Dari keempat anggota provinsi di Klaster 3, jumlah klaim penyakit katastrofik terbesar ialah Penyakit Ginjal Kronis dengan jumlah klaim 8950 di Provinsi DKI Jakarta, 6684 di Provinsi Bali, 4022 di Provinsi DI Yogyakarta, dan 2421 di Provinsi Sulawesi Utara. Kategori penyakit terbanyak kedua di Klaster 3 adalah penyakit lainnya dengan rata-rata persentase pengajuan klaim sebesar 14,80125%. Nilai persentase klaim penyakit lainnya di Provinsi DKI Jakarta sebesar 15,04825% dengan klaim penyakit terbanyaknya ialah dorsalgia sebanyak 1887 klaim, pulpitis sebanyak 1500 klaim dan demam dan lainnya sebanyak 1341 klaim.

Terakhir adalah klaster *noise* yang hanya berisikan satu anggota yakni Provinsi Papua Barat. Berdasarkan Tabel 3, klaim variabel penyakit yang menonjol ialah klaim persalinan. Berbeda dengan Klaster 2, persalinan di klaster *noise* memiliki persentase jumlah klaim yang sangat berbeda. Persentase jumlah klaim persalinan di Provinsi Papua Barat sebesar 7,28%. Karena perbedaan yang signifikan antara Provinsi Papua Barat dan anggota pada klaster 2 membuat Provinsi Papua Barat termasuk dalam pencilan atau *noise*. Seperti halnya yang sudah dijelaskan pada Gambar 1, salah satu penyebab banyaknya jumlah klaim persalinan yang ada di Provinsi Papua Barat ialah minimnya pengetahuan masyarakat Provinsi Papua Barat mengenai program Keluarga Berencana (KB). Peserta aktif KB di Provinsi Papua Barat pada tahun 2020 sebanyak 29%. Hal ini menyebabkan banyaknya persalinan yang tidak diinginkan terjadi sehingga jumlah klaim persalinan membengkak. Selain itu, klaim yang diajukan oleh peserta BPJS di Provinsi Papua Barat mencakup penyakit yang terjadi pada ibu hamil dan pasca persalinan. Penyakit yang terjadi pada ibu hamil disebabkan karena banyak faktor, salah satunya ialah imunitas atau kekebalan tubuh ibu hamil. Seringkali ibu hamil mengalami kekurangan zat besi yang disebabkan oleh banyaknya darah yang digunakan oleh tubuh ibu hamil. Oleh karena itu diperlukan tambahan suplemen untuk ibu hamil seperti tablet penambah darah. Pemberian tablet tambah darah di Provinsi Papua Barat tergolong sangat rendah. Pada tahun 2020 pemberian tablet tambah darah di Provinsi Papua Barat sebanyak 29,8%. Oleh karena itu banyak ibu hamil di Provinsi Papua Barat yang mengalami penyakit sehingga diperlukan penanganan khusus di fasilitas kesehatan seperti rumah sakit dan lainnya. Selain itu, cakupan persalinan di fasilitas pelayanan kesehatan Provinsi Papua Barat tahun 2020 tergolong sangat rendah, yakni sebanyak 34,7%. Grafik pendukung terkait persalinan di fasilitas kesehatan tahun 2020 disajikan pada Gambar 10.



Gambar 10. Grafik Persentase Cakupan Persalinan di Fasilitas Pelayanan Kesehatan Menurut Provinsi Tahun 2020

Hal ini disebabkan banyak faktor, salah satunya ialah budaya *blood taboo* atau tindakan isolasi ibu hamil. Tindakan ini merupakan budaya setempat yang didasari anggapan bahwa darah yang dikeluarkan oleh perempuan melalui alat kelaminnya ialah darah pembawa sial. Hal ini menyebabkan banyak ibu hamil menjelang persalinan diisolasi dan ditempatkan di rumah yang berbeda. Aktivitas seperti makan, memasak dan tidur dilakukan sendiri oleh ibu hamil selama kurang-lebih 3 minggu menjelang proses persalinan dan berdampak pada kesehatan ibu hamil.

#### IV. SIMPULAN

Berdasarkan penelitian yang telah dilakukan, didapati kesimpulan sebagai berikut:

1. Berdasarkan metode *Density Based Spatial Clustering of Application with Noise* (DBSCAN), didapati jumlah kluster optimal sebanyak 3 kluster dengan kluster 1 yang terdiri dari dua anggota provinsi, kluster 2 yang terdiri dari 27 anggota provinsi, kluster 3 yang terdiri dari empat anggota provinsi, dan kluster *noise* yang terdiri dari satu anggota dengan nilai rata-rata indeks *silhouette* 0,6595942 yang berarti model *clustering* yang dihasilkan memiliki struktur *medium* atau standar.
2. Hasil analisis Kluster menunjukkan bahwa kategori penyakit yang banyak diklaim pada Kluster 1 ialah kategori kecelakaan dan penyakit lainnya. Nilai tertinggi kategori kecelakaan dan penyakit lainnya adalah Provinsi Aceh. Kluster 2 dengan kategori penyakit yang banyak diklaim ialah Persalinan. Nilai tertinggi klaim kategori persalinan ialah Provinsi Kalimantan Utara. Kluster 3 dengan kategori penyakit yang banyak diklaim ialah Katastropik dan Penyakit Lainnya. Nilai tertinggi klaim penyakit katastropik di Provinsi DI Yogyakarta. Nilai tertinggi klaim penyakit lainnya di Provinsi Bali. Kluster *noise* dengan kategori penyakit yang banyak diklaim ialah persalinan dengan jumlah klaim yang termasuk kedalam *noise* atau pencilan..

#### DAFTAR PUSTAKA

- [1] Y. S. Sy., “Klasterisasi Pasien Rawat Inap Peserta BPJS Berdasarkan Jenis Penyakit Menggunakan Algoritma K-Means,” *Jurnal Sistem Informasi dan Teknologi*, vol. 5, no. 2, pp. 33-37, 2023.
- [2] P. M. Putri dan P. B. Murdi, “Pelayanan Kesehatan di Era Jaminan Kesehatan Nasional Sebagai Program Badan Penyelenggara Jaminan Sosial Kesehatan,” *Wacana Hukum: Jurnal Fakultas Hukum Universitas Slamet Riyadi*, vol. 25, no. 1, pp. 80-97, 2019.
- [3] P. R. Indonesia, “Peraturan Presiden Republik Indonesia Nomor 82 Tahun 2018 tentang Jaminan Kesehatan,” 17 September 2023. [Online]. Available: <https://jdih.kemenkeu.go.id/fullText/2018/82TAHUN2018PERPRES.pdf>.
- [4] I. G. Andirasdini, D. Saputra, M. Rivai, and S. E. M. Putra, “Analysis of the Health Social Security Administration (BPJS Kesehatan) Claim Amount Using Random Forest Regression,” *Indonesian Actuarial Journal*, vol. 1, no. 1, pp. 1-8, 2025.
- [5] M. Rivai, R. Rosni, N. Azzanina, A. N. I. B. Ratam, B. Z. Zarrouby, F. Perangin Angin, I. H. Aqil, L. Latifah, and R. K. Laia, “Analisis Klasifikasi Curah Hujan Harian Menggunakan Metode Decision Tree (Studi Kasus: Stasiun Pengamatan Radin Inten II Lampung),” *JOSTECH Journal of Science and Technology*, vol. 5, no. 2, pp. 215-222, 2025.
- [6] K.S. Dewi, M. Rivai, A. Nurhasanah, “Analisis Clustering Menggunakan Metode K-Means Pada Data Produksi Tanaman Biofarmaka Di Indonesia.” *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*. Vol. 6, no. 2, pp. 699-708, 2025.
- [7] M. Rivai, A.K. Hutabarat, D.H. Halawa, I Satrio, M.S. Tarmizi, N. Frasiska, N. Pratiwi, Z. Alma, “Analisis Jumlah Kecelakaan Lalu Lintas, Korban, dan kerugian di Provinsi Lampung Tahun 2021 dengan Metode K-Means Cluster.” *Indonesian Journal of Applied Mathematics*. vol. 4, no. 1, pp. 19-25, 2024.

- 
- [8] M. Rivai, N. Frasiska. "Analysis of Fuzzy C-Means Clustering Method on Grouping Provinces in Indonesia Based on Economic Growth in 2023." *International Journal of Scientific Research in Science, Engineering and Technology*, vol. 11 no. 6, pp. 380-388, 2024.
- [9] S. Sundari, I. S. Damanik, A. P. Windarto, H. S. Tambunan, J. dan A. Wanto, "Analisis K-Medoids Clustering dalam Pengelompokan Data Imunisasi Campak Balita di Indonesia," *Seminar Nasional Riset Information Science (SENARIS)*, pp. 687-696, 2019.
- [10] N. Dwitri, J. A. Tampubolon, S. Prayoga, P. F. I. R. Zer dan D. Hartama, "Penerapan Algoritma K-Means dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 di Indonesia," *Jurnal Teknologi Informasi*, vol. 4, no. 1, pp. 128-132, 2020.
- [11] Djaali, *Metodologi Penelitian Kuantitatif*, Jakarta: PT Bumi Aksara, 2020.
- [12] G. Urva, Desyanti, I. Albanna, M. S. Sungkar, I. M. A. O. Gunawan, I. Adhicandra, S. Ramadhan, R. L. Rahardian, Herlawati, R. T. Handayanto, A. A. G. B. Ariana, Hartatik, P. D. Atika dan S. Junaidi, *Penerapan Data Mining di Berbagai Bidang: Konsep, Metode, dan Studi Kasus*, Jambi: PT. Sonpedia Publishing Indonesia, 2023.
- [13] A. Bastian, H. Sujadi, and G. Febrianto, "Penerapan Algoritma K-Means Clustering Analysis pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," *Jurnal Sistem Informasi*, vol. 14, no. 1, pp. 26-32, 2018.
- [14] G. Urva, Desyanti, I. Albanna, M. S. Sungkar, I. M. A. O. Gunawan, I. Adhicandra, S. Ramadhan, R. L. Rahardian, Herlawati, R. T. Handayanto, A. A. G. B. Ariana, Hartatik, P. D. Atika, and S. Junaidi, *Penerapan Data Mining di Berbagai Bidang: Konsep, Metode, dan Studi Kasus*. Jambi: PT. Sonpedia Publishing Indonesia, 2023.
- [15] A. Setiawan, "Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris," *Jurnal Sains dan Edukasi Sains*, vol. 5, no. 1, pp. 28-37, 2022.
- [16] W. P. Nurmayanti, D. J. Ratnaningsih, S. Nisrina, A. Rahim, M. Malthuf, and W. Kusuma, "Clustering of BPJS National Health Insurance Participant Using DBSCAN Algorithm," *Jurnal Varian* vol. 6, no. 1, 2022.
- [17] L. Qadrini, "Metode K-Means dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017," *Jurnal Statistika*, vol. 13, no. 2, pp. 5-11, 2020.
- [18] L. Kaufman dan P. J. Rousseuw, *Finding Groups in Data: an Introduction to Cluster Analysis*, New York: Wiley, 1990.
- [19] Kementerian Kesehatan Republik Indonesia, *Profil Kesehatan Indonesia 2020*, Jakarta: Kementerian Kesehatan Republik Indonesia, 2021.
- [20] Badan Pusat Statistik, *Statistik Transportasi Darat 2020*, Jakarta Pusat: BPS RI, 2021.
- [21] Pemerintah Provinsi Kalimantan Timur, "Data Pelanggaran Lalu Lintas di Wilayah Kalimantan Timur Tahun 2020," 2021. [Online]. Available: <https://data.kaltimprov.go.id/dataset/00de8d1f-bf1c-40db-9ecf-e900eae48573/resource/3bab59ee-37e7-4c3b-b6ea-b2faa96cf0a1/download/data-pelanggaran-lalu-lintas-di-wilayah-kalimantan-timur-tahun-2020.csv>. [Diakses 3 Juni 2025].