



**KERJA PRAKTIK - IF184801**

## **Deteksi Anomali pada Dataset BPJS Tahun 2020**

Departemen Informatika Institut Teknologi Sepuluh Nopember  
Surabaya

Jl. Teknik Kimia, Institut Teknologi Sepuluh Nopember Surabaya,  
Sukolilo, Surabaya 60111, Indonesia

Periode: 14 Februari 2023 – 7 Oktober 2023

### **Oleh:**

Dewangga Dharmawan      05111940000029

### **Pembimbing Departemen**

Dini Adni Navastara, S.Kom., M.Sc.

### **Pembimbing Lapangan**

Dr. Rarasmaya Indraswari, S.Kom, M.Kom.

**DEPARTEMEN TEKNIK INFORMATIKA**

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya 2023



**KERJA PRAKTIK - IF184801**

**Deteksi Anomali pada Dataset BPJS Tahun 2020**

Departemen Informatika Institut Teknologi Sepuluh Nopember  
Surabaya

Jl. Teknik Kimia, Institut Teknologi Sepuluh Nopember Surabaya,  
Sukolilo, Surabaya 60111, Indonesia

Periode: 14 Februari 2023 – 7 Oktober 2023

**Oleh:**

Dewangga Dharmawan 05111940000029

**Pembimbing Departemen**

Dini Adni Navastara, S.Kom., M.Sc.

**Pembimbing Lapangan**

Dr. Rarasmaya Indraswari, S.Kom, M.Kom.

DEPARTEMEN TEKNIK INFORMATIKA

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya 2023

## **DAFTAR ISI**

<b>DAFTAR GAMBAR</b> .....	7
<b>DAFTAR TABEL</b> .....	9
<b>LEMBAR PENGESAHAN KERJA PRAKTIK</b> .....	11
<b>ABSTRAK</b> .....	13
<b>KATA PENGANTAR</b> .....	15
<b>BAB I PENDAHULUAN</b> .....	17
<b>1.1 Latar Belakang</b> .....	17
<b>1.2 Tujuan</b> .....	18
<b>1.3 Manfaat</b> .....	18
<b>1.4 Rumusan Masalah</b> .....	18
<b>1.5 Lokasi dan Waktu Kerja Praktik</b> .....	19
<b>1.6 Metodologi Kerja Praktik</b> .....	19
<b>1.6.1 Perumusan Masalah</b> .....	19
<b>1.6.2 Studi Literatur dan Teori</b> .....	19
<b>1.6.3 Analisis dan Perancangan Sistem</b> .....	19
<b>1.6.4 Implementasi Program</b> .....	20
<b>1.6.5 Pengujian dan Evaluasi</b> .....	20
<b>1.6.6 Kesimpulan dan Saran</b> .....	20
<b>1.7 Sistematika Laporan</b> .....	20
<b>1.7.1 Bab I Pendahuluan</b> .....	20
<b>1.7.2 Bab II Profil Perusahaan</b> .....	21
<b>1.7.3 Bab III Tinjauan Pustaka</b> .....	21

1.7.4	<b>Bab IV Analisis dan Perancangan Program</b>	
		21
1.7.5	<b>Bab V Implementasi Sistem .....</b>	21
1.7.6	<b>Bab VI Pengujian dan Evaluasi.....</b>	21
1.7.7	<b>Bab VII Kesimpulan dan Saran.....</b>	21
<b>BAB II PROFIL PERUSAHAAN.....</b>		23
2.1	<b>Profil Departemen Informatika Institut Teknologi Sepuluh Nopember Surabaya.....</b>	23
2.2	<b>Lokasi (Perlu lebih banyak informasi, visi misi, lab-lab, dsb. Dari web (Masukan ke sitasi)) .....</b>	23
2.3	<b>Sejarah .....</b>	23
2.4	<b>Visi.....</b>	24
2.5	<b>Misi.....</b>	24
2.6	<b>Tujuan .....</b>	24
2.7	<b>Sasaran.....</b>	25
2.8	<b>Fasilitas .....</b>	25
2.9	<b>Akreditasi.....</b>	27
2.10	<b>Sumber Informasi.....</b>	27
<b>BAB III TINJAUAN PUSTAKA .....</b>		29
3.1	<b>BPJS .....</b>	29
3.2	<b>Jenis Fasilitas Kesehatan.....</b>	29
3.3	<b>Pembagian Peserta BPJS.....</b>	30
3.4	<b>ICD .....</b>	31
3.4.1	<b>ICD-10.....</b>	32

3.4.2	ICD-9-CM Prosedur .....	36
3.5	INA-CBG .....	37
3.5.1	Case-Mix Main Group.....	39
3.5.2	Kelompok Tipe Kasus.....	42
3.5.3	Tingkat Keparahan.....	43
3.6	Tarif INA-CBG.....	44
3.7	Machine Learning.....	45
3.8	Anomali dan Jenis-jenisnya .....	46
3.9	Boxplot .....	49
3.10	PCA .....	50
3.11	Algoritma Deteksi Anomali.....	52
3.12	Pengukuran Evaluasi Deteksi Anomali.....	55
<b>BAB IV ANALISIS DAN PERANCANGAN PROGRAM</b>		
.....		59
4.1	Analisis Dataset .....	59
4.1.1	Dataset kepesertaan .....	60
4.1.2	Dataset Kunjungan FKRTL.....	73
4.1.3	Dataset Kunjungan FKTP Non-Kapitasi.....	93
4.2	Perancangan Sistem.....	103
<b>BAB V IMPLEMENTASI SISTEM</b> .....		107
5.1	Pengelompokan Kelas-Kelas Minoritas .....	107
5.2	Boxplot Modifikasi.....	108
5.3	PCA Dataset.....	109

5.4	DBSCAN .....	110
5.5	Isolation Forest .....	111
5.6	LOF .....	111
5.1.5	Train-Test Split.....	112
5.1.4	Pengukuran Metrik.....	113
<b>BAB VI PENGUJIAN DAN EVALUASI .....</b>		<b>115</b>
6.1	Dataset Peserta .....	115
6.2	Dataset Kunjungan FKRTL.....	117
6.3	Dataset Kunjungan FKTP Non-Kapitasi.....	123
<b>BAB VII KESIMPULAN DAN SARAN .....</b>		<b>131</b>
7.1	Kesimpulan.....	131
7.2	Saran .....	133
<b>DAFTAR PUSTAKA .....</b>		<b>135</b>
<b>BIODATA PENULIS .....</b>		<b>141</b>

## DAFTAR GAMBAR

Gambar 1 Contoh Grafik ROC-AUC.....	57
Gambar 2 Distribusi Usia Peserta BPJS .....	64
Gambar 3 Distribusi Bulan dan Tanggal Lahir Peserta BPJS	65
Gambar 4 Heatmap Bulan dan Tanggal Lahir Peserta .....	65
Gambar 5 Distribusi Bobot Peserta BPJS .....	71
Gambar 6 Scatterplot Bobot dan Usia Peserta BPJS .....	72
Gambar 7 Distribusi Banyak Kunjungan FKRTL.....	74
Gambar 8 Grafik Distribusi Tahun Kunjungan FKRTL dan Pulang dari FKRTL.....	76
Gambar 9 Distribusi Hari Kunjungan FKRTL.....	76
Gambar 10 Distribusi Lama Kunjungan FKRTL.....	77
Gambar 11 Distribusi Banyak Prosedur.....	91
Gambar 12 Distribusi Tarif Kunjungan FKRTL Setelah Verifikasi.....	93
Gambar 13 Distribusi Banyak Kunjungan FKTP Non-Kapitasi .....	94
Gambar 14 Distribusi Banyak Kunjungan, Tindakan, dan Pulang dari FKTP Non-Kapitasi Tiap Tahun .....	96
Gambar 15 Flowchart Perancangan Sistem .....	103
Gambar 16 Grafik ROC-AUC dari Masing-Masing Algoritma Deteksi Anomali Numerik Multivariat Dataset Kunjungan FKRTL .....	119
Gambar 17 Grafik ROC-AUC dari Masing-Masing Algoritma Deteksi Anomali Multidimensional pada Dataset Kunjungan FKRTL .....	121
Gambar 18 Grafik ROC-AUC dari Masing-Masing Algoritma Deteksi Anomali Numerik Multivariat pada Dataset Kunjungan FKTP Non-Kapitasi .....	126
Gambar 19 Grafik ROC-AUC dari Masing-Masing Algoritma Deteksi Anomali Multidimensional pada Dataset Kunjungan FKTP Non-Kapitasi .....	128

*[Halaman ini sengaja tidak berisi kecuali kalimat ini]*

## DAFTAR TABEL

Tabel 1 Bab ICD-10 Edisi 2010 .....	34
Tabel 2 Bab ICD-9-CM Prosedur .....	37
Tabel 3 Kode INA-CBG .....	39
Tabel 4 Kode CMG Biasa .....	40
Tabel 5 Kode CMG Spesial .....	41
Tabel 6 Kode Tipe Kasus INA-CBG.....	43
Tabel 7 Tingkat Keparahan pada INA-CBG .....	44
Tabel 8 Jenis Data Anomali .....	47
Tabel 9 Jenis Hubungan Keluarga Peserta BPJS .....	61
Tabel 10 Banyak Keluarga Peserta BPJS yang Memiliki Dua Pasang dan Hubungan Keluarganya.....	62
Tabel 11 Jenis Kelamin Peserta BPJS .....	62
Tabel 12 Status Hubungan Perkawinan Peserta BPJS .....	63
Tabel 13 Daerah Asal Peserta BPJS.....	67
Tabel 14 Daerah Asal Faskes Pendaftaran Peserta BPJS .....	67
Tabel 15 Kepemilikan Faskes Pendaftaran Peserta BPJS ....	68
Tabel 16 Jenis Faskes Pendaftaran Peserta BPJS.....	69
Tabel 17 Kepemilikan dan Jenis Faskes Pendaftaran Peserta BPJS.....	69
Tabel 18 Tingkat Kelas Perawatan Peserta BPJS.....	70
Tabel 19 Segmentasi Peserta BPJS .....	70
Tabel 20 Status Peserta BPJS.....	72
Tabel 21 Kunjungan ke FKTP Kapitasi .....	74
Tabel 22 Daerah Asal FKRTL.....	78
Tabel 23 Kepemilikan FKRTL.....	79
Tabel 24 Jenis FKRTL .....	79
Tabel 25 Tipe FKRTL .....	80
Tabel 26 Jenis Pelayanan FKRTL .....	80
Tabel 27 Jenis Poli .....	81
Tabel 28 Kelas Perawatan FKRTL.....	82
Tabel 29 Segmentasi Peserta pada Kunjungan FKRTL .....	82
Tabel 30 Status Pulang Kunjungan FKRTL.....	83

Tabel 31 Bab Diagnosis Masuk Kunjungan FKRTL .....	85
Tabel 32 Bab Diagnosis Primer Kunjungan FKRTL .....	85
Tabel 33 Digit Pertama Kode INA-CBG .....	86
Tabel 34 Digit Kedua INA-CBG .....	86
Tabel 35 Digit Keempat INA-CBG.....	87
Tabel 36 Daerah Asal Faskes Perujuk.....	88
Tabel 37 Kepemilikan Faskes Perujuk.....	88
Tabel 38 Jenis Faskes Perujuk .....	89
Tabel 39 Tipe Faskes Perujuk .....	90
Tabel 40 Kelompok Tarif INA-CBG.....	92
Tabel 41 Daerah Asal FKTP Non-Kapitasi .....	97
Tabel 42 Kepemilikan FKTP Non-Kapitasi .....	97
Tabel 43 Jenis FKTP Non-Kapitasi.....	98
Tabel 44 Tipe FKTP Non-Kapitasi .....	98
Tabel 45 Jenis Layanan FKTP Non-Kapitasi.....	99
Tabel 46 Segmentasi Peserta pada Kunjungan FKTP Non-Kapitasi .....	99
Tabel 47 Bab Diagnosis Kunjungan FKTP Non-Kapitasi..	100
Tabel 48 Jenis Tindakan pada Kunjungan FKTP Non-Kapitasi .....	101
Tabel 49 Distribusi Tarif Kunjungan FKTP Non-Kapitasi Setelah Verifikasi .....	102
Tabel 50 Anomali Kategorikal dan Campuran pada Dataset Kunjungan FKRTL Berdasarkan Jenis Layanan dan Lama Kunjungan.....	117
Tabel 51 Hasil dari Tes Anomali Numerik Multivariat Dataset Kunjungan FKRTL .....	118
Tabel 52 Hasil dari Tes Anomali Multidimensional Dataset Kunjungan FKRTL .....	122
Tabel 53 Hasil dari Tes Anomali Numerik Multivariat Dataset Kunjungan FKTP Non-Kapitasi.....	125
Tabel 54 Hasil dari Tes Anomali Multidimensional Dataset Kunjungan FKTP Non-Kapitasi.....	127

**LEMBAR PENGESAHAN  
KERJA PRAKTIK  
Deteksi Anomali pada Dataset BPJS Tahun 2020**

Oleh:

Dewangga Dharmawan

05111940000029

Disetujui oleh Pembimbing Kerja Praktik:

1. Dini Adni Navastara, S.Kom,  
M.Sc.  
NIP. 198510172015042001

  
(Pembimbing Departemen)

2. Dr. Rarasmaya Indraswari,  
S.Kom.  
NIP. 1995202012057

  
(Pembimbing Lapangan)

*[Halaman ini sengaja tidak berisi kecuali kalimat ini]*

## **Deteksi Anomali pada Dataset BPJS Tahun 2020**

Nama Mahasiswa : Dewangga Dharmawan  
NRP : 05111940000029  
Departemen : Informatika FTEIC-ITS  
Pembimbing Departemen : Dini Adni Navastara,  
S.Kom, M.Sc.  
Pembimbing Lapangan : Dr. Rarasmaya Indraswari,  
S.Kom, M.Kom.

### **ABSTRAK**

*BPJS Kesehatan adalah badan yang bertanggungjawab atas pemberian bantuan terhadap kesehatan penduduk Indonesia. Dalam BPJS, terdapat berbagai peserta yang menggunakan jasa BPJS untuk mengurangi beban biaya pembayaran rumah sakit. Beberapa pasien memiliki data yang berada di luar jangkauan biasanya. Sehingga, penelitian ini mengusulkan cara-cara untuk mencari anomali dalam dataset-dataset BPJS dengan menggunakan Boxplot modifikasi, DBSCAN, Isolation Forest, dan manual untuk menemukan anomali numerik dan kategorikal, univariat dan multivariat, dan multidimensional. Ditemukan maksimal 1.547 peserta yang memiliki ketidakcocokan informasi kategorikal, maksimal 26,85% kunjungan FKRTL memiliki nilai numerik diluar batas bawah atau atas, maksimal 61 data dimana kunjungan FKRTL yang seharusnya rawat jalan memiliki lama kunjung lebih dari satu hari, 0,07% dari dataset kunjungan FKRTL merupakan anomali numerik multivariat menurut Isolation Forest AND LOF, 1,79% dari data kunjungan FKRTL merupakan anomali multidimensional menurut model hybrid Isolation Forest OR LOF, 236*

*kunjungan FKTP non-kapitasi rawat jalan yang mana lama kunjungannya lebih dari satu hari, maksimal 12.89% data kunjungan FKTP non-kapitasi yang memiliki nilai numerik yang berada di luar batas bawah dan atas, 0,55 persen dari data FKTP non-kapitasi memiliki anomali numerik multivariat, dan 0,72 persen dari data FKTP non-kapitasi menurut DBSCAN atau Isolation Forest AND LOF merupakan data anomali multidimensional.*

***Kata Kunci : BPJS, Deteksi Anomali, Machine Learning, Analisis Data, Anomali***

## **KATA PENGANTAR**

Puji syukur kepada Allah SWT bahwasannya penulis memiliki semangat untuk menulis catatan Kerja Praktik berjudul Deteksi Anomali pada Dataset BPJS Tahun 2023. Penulis juga memiliki kekurangan mengenai waktu dan fokus pada pengerjaan kerja praktik. Terima kasih kepada beberapa orang yang membantu menyusun dan memberikan semangat dalam penulisan kerja praktik ini, diantaranya adalah kedua orang tua penulis, Ibu Dini Adni Navastara S.Kom, M.Sc. sebagai dosen pembimbing departemen, Ibu Rarasmaya Indraswari S.Kom, M.Kom. sebagai pembimbing lapangan, dan teman-teman penulis yang memberikan saran dan pengalaman dalam melakukan kerja praktik pada bidang lainnya.

Surabaya, 7 Oktober 2023  
Dewangga Dharmawan

*[Halaman ini sengaja tidak berisi kecuali kalimat ini]*

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Semua hal yang ada di alam semesta dapat menjadi sebuah data yang dapat dikumpulkan menjadi kumpulan data, atau dataset, yang disimpan dalam berbagai media. Seperti alam semesta, tidak semua data memiliki nilai yang sama dan sempurna. Dalam suatu dataset, terdapat data yang tidak sesuai dengan data-data lainnya. Data tersebut dapat disebut sebagai anomali. Anomali dalam data memberikan informasi yang dapat digunakan bagi pembaca dan pengguna data. Sehingga, deteksi anomali menjadi salah satu bidang yang dipelajari dan selalu dikembangkan selama ratusan tahun (Nassif et al., 2021). Contoh dari anomali adalah pola yang tidak biasa dalam grafik berbasis waktu untuk mencegah kerugian finansial dan sosial (Ren et al., 2019), pola dalam suatu gambar produk dalam pekerjaan Quality Assurance (Roth et al., 2022), kejadian yang mencurigakan atau berbahaya dalam rekaman lalu lintas (Fadhlurohman, 2021).

Machine Learning adalah salah satu teknik yang digunakan dalam deteksi anomali dan cukup populer dalam memecahkan masalah-masalah dalam ilmu komputer (Elmrabit et al., 2020) (Nassif et al., 2021). Machine Learning memecahkan masalah-masalah tersebut dengan mempelajari riwayat sebelumnya (Nassif et al., 2021) (Wang et al., 2021). Contoh lainnya dari deteksi anomali menggunakan Machine Learning adalah deteksi kesehatan pasien berdasarkan Internet of Things (IoT) menggunakan SVM untuk mengetahui ketidakwajaran dalam catatan kondisi kesehatan pasien dan serangan terhadap IoT oleh (Said et al., 2021), deteksi anomali dalam data layanan kesehatan bagi penduduk di daerah terpencil menggunakan K-means dan variasinya oleh (Kavitha

et al., 2021), deteksi ketidaksesuaian dari data berbasis waktu dalam IOT oleh (Cook et al., 2019), deteksi anomali dalam dataset KDD99 yang memiliki data berupa serangan jaringan menggunakan Sub-Space Clustering dan One Class-SVM tanpa pengetahuan sebelumnya oleh (Pu et al., 2020), dan mendeteksi ketidakwajaran dalam tingkah seorang dalam lingkungan smart house menggunakan model Markov tersembunyi oleh (Sanchez et al., 2020).

Dalam kerja praktik ini, dataset yang akan dicari anomalnya adalah dataset-dataset dari BPJS berupa dataset kunjungan FKRTL dan FKTP non-kapitasi dan dataset peserta BPJS. Seperti dataset-dataset lainnya, dataset BPJS memiliki data anomali, seperti suatu kunjungan yang lebih mahal tetapi lebih lama daripada kunjungan lainnya serta ketidaksesuaian antara dua atribut dalam suatu data. Deteksi anomali dilakukan dengan menggunakan algoritma deteksi anomali boxplot, DBSCAN, Isolation Forest, dan LOF dan deteksi anomali manual dengan mengfilter dataset berdasarkan jenis atribut.

## **1.2 Tujuan**

Tujuan kerja praktik ini adalah memenuhi kewajiban mata kuliah kerja praktik, menemukan data-data BPJS yang tergolong sebagai anomali, dan menjelaskan kenapa data-data tersebut merupakan anomali.

## **1.3 Manfaat**

Manfaat penelitian ini adalah mendapatkan informasi mengenai data-data yang tidak biasa dalam pola data BPJS.

## **1.4 Rumusan Masalah**

Ada dua rumusan masalah dalam kerja praktik ini:

1. Bagaimana persiapan data agar anomali dalam data dapat dideteksi?
2. Bagaimana mendeteksi anomali dalam dataset BPJS?
3. Bagaimana mengevaluasi deteksi anomali dari dataset BPJS?

## **1.5 Lokasi dan Waktu Kerja Praktik**

Pengerjaan Kerja Praktik dilakukan di Departemen Informatika Institut Teknologi Sepuluh Nopember dan dimulai pada tanggal 14 Februari 2023 sampai 7 Oktober 2023.

## **1.6 Metodologi Kerja Praktik**

### **1.6.1 Perumusan Masalah**

Dataset yang didapatkan dari BPJS memiliki data-data yang memiliki perbedaan dengan data-data normal didalamnya. Data-data tersebut disebut sebagai data-data anomali. Data-data anomali dibagi menurut jenis anomali, seperti univariat dan multivariat.

### **1.6.2 Studi Literatur dan Teori**

Untuk melakukan deteksi anomali, diperlukan pengetahuan mengenai atribut-atribut apa saja yang ada di dalam dataset BPJS. Penjelasan atribut-atribut tersebut dapat diketahui melalui berbagai landasan hukum BPJS, sumber-sumber lainnya, dan pengasumsian. Selain itu, berbagai algoritma yang digunakan dalam deteksi anomali dirujuk pada berbagai penelitian sebelumnya untuk mencari algoritma yang tepat untuk mencari anomali.

### **1.6.3 Analisis dan Perancangan Sistem**

Setelah mendapatkan dataset, dataset akan diproses. Setelah diproses, data-data anomali akan

dikurangi dimensionalitasnya dengan PCA sebelum dicari menggunakan algoritma boxplot, DBSCAN, Isolation Forest, dan manual.

#### **1.6.4 Implementasi Program**

Program dijalankan setelah menerima dataset. Informasi-informasi yang ada di dalam dataset akan diproses untuk mengetahui data anomali dan relasinya dengan data non-anomali.

#### **1.6.5 Pengujian dan Evaluasi**

Hasil dari deteksi anomali akan dievaluasi menggunakan algoritma ROC-AUC dan nilai-nilai dari Confusion Matrix (True Positive, False Positive, False Negative, True Negative). Model dan algoritma yang terbaik menjadi model utama untuk menentukan anomali yang ada di dalam dataset BPJS.

#### **1.6.6 Kesimpulan dan Saran**

Dataset-dataset BPJS dilakukan pre-processing, EDA, standarisasi, dan encoding sebelum dan bersamaan dengan deteksi anomali. Algoritma deteksi anomali, selain boxplot, akan dievaluasi menggunakan algoritma ROC-AUC dan Confusion Matriks untuk ditemukan nilai-nilai metrik yang terbaik untuk menemukan data-data anomali. Diharapkan bahwa penelitian selanjutnya dapat lebih baik daripada penelitian ini.

### **1.7 Sistematika Laporan**

#### **1.7.1 Bab I Pendahuluan**

Bab ini berisi latar belakang, tujuan, manfaat, rumusan masalah, lokasi dan waktu kerja praktik, metodologi, dan sistematika laporan.

### **1.7.2 Bab II Profil Perusahaan**

Bab ini berisi profil Departemen Informatika Institut Teknologi Sepuluh Nopember.

### **1.7.3 Bab III Tinjauan Pustaka**

Bab ini berisi dasar teori dan aplikasi yang digunakan dalam kerja praktik yang ditulis.

### **1.7.4 Bab IV Analisis dan Perancangan Program**

Bab ini berisi tahapan analisis dataset dan pembuatan program.

### **1.7.5 Bab V Implementasi Sistem**

Bab ini berisi proses berjalannya program

### **1.7.6 Bab VI Pengujian dan Evaluasi**

Bab ini berisi hasil program dan evaluasi penilaian algoritma deteksi anomali yang digunakan.

### **1.7.7 Bab VII Kesimpulan dan Saran**

Bab ini berisi kesimpulan dan saran yang ada dari kerja praktik.

*[Halaman ini sengaja tidak berisi kecuali kalimat ini]*

## **BAB II**

### **PROFIL PERUSAHAAN**

#### **2.1 Profil Departemen Informatika Institut Teknologi Sepuluh Nopember Surabaya**

Institut Teknologi Sepuluh Nopember Surabaya, atau ITS, adalah universitas yang terletak di Surabaya, Indonesia. ITS memiliki fokus pendidikan tinggi berbasis teknologi informasi dan komunikasi. ITS memiliki sembilan fakultas, diantaranya adalah Fakultas Teknologi Elektro dan Informatika Cerdas, atau FTEIC. FTEIC berfokus pada bidang kelistrikan, elektronika, sistem dan teknologi informasi, dan komputasi. FTEIC memiliki enam departemen, diantaranya adalah Departemen Informatika. Departemen Informatika memiliki fokus terhadap ilmu komputer dan teknologi informasi dan komunikasi.

#### **2.2 Lokasi (Perlu lebih banyak informasi, visi misi, lab-lab, dsb. Dari web (Masukan ke sitasi))**

Jl. Teknik Kimia, Institut Teknologi Sepuluh Nopember Surabaya, Sukolilo, Surabaya 60111, Indonesia

#### **2.3 Sejarah**

Departemen Informatika Institut Teknologi Sepuluh Nopember Surabaya pertama kali didirikan pada 1985 dengan nama Program Studi Teknik Komputer melalui perintah Direktorat Jendral Pendidikan Tinggi. Pada tahun 1993, Program Studi Teknik Komputer berubah nama menjadi Jurusan Teknik Komputer. Pada tahun 1996, berdasarkan Surat Keputusan Direktur Jendral Pendidikan Tinggi Nomor 224/DIKTI/Kep/1996 tanggal 11 Juli 1996, Jurusan Teknik Komputer berubah nama menjadi Jurusan Teknik Informatika, atau Departemen Informatika. Departemen Informatika

menyediakan program Sarjana (S1) sejak 1985, Pasca Sarjana (S2) sejak 1994, dan Doktor (S3) pada 2011.

## **2.4 Visi**

Departemen Informatika, sebagai bagian dari Institut Teknologi Sepuluh Nopember Surabaya (ITS), memiliki visi yang sejalan, yaitu menjadikan ITS sebagai perguruan tinggi dengan reputasi internasional dalam ilmu pengetahuan, teknologi, dan seni, terutama yang menunjang industri dan kelautan yang berawawasan lingkungan, menjadikan Departemen Informatika sebagai inovator bidang informatika yang unggul di tingkat nasional dengan reputasi internasional serta berperan aktif dalam upaya memajukan dan mengsejahterakan bangsa.

## **2.5 Misi**

Departemen Informatika ITS memiliki misi menyelenggarakan proses pembelajaran yang berkualitas dan memenuhi standar nasional maupun internasional, melaksanakan penelitian yang inovatif, bermutu, dan bermanfaat, meningkatkan pemanfaatan teknologi informasi dan komunikasi untuk masyarakat, dan menjalin kemitraan dengan berbagai lembaga, baik di dalam maupun di luar negeri.

## **2.6 Tujuan**

Departemen Informatika ITS memiliki tujuan menghasilkan lulusan yang kompeten di bidang Informatika serta memiliki daya saing dan kemandirian untuk berkompetisi di tingkat nasional dan internasional, melakukan perbaikan proses pembelajaran secara berkesinambungan, menghasilkan karya penelitian yang inovatif dan bermanfaat bagi masyarakat, serta publikasi di jurnal nasional ataupun internasional bereputasi, memberikan kontribusi bagi peningkatan mutu kehidupan masyarakat, dan mengambil peran aktif dalam

kegiatan bidang informatika pada tingkat nasional dan internasional.

## **2.7 Sasaran**

Departemen Informatika ITS memiliki sasaran berupa peningkatan kualitas PBM, peningkatan kualitas lulusan yang unggul pada tingkat nasional, pertumbuhan dan keberlanjutan yang merupakan pengembangan unggulan dan dikenalnya departemen Informatika di tingkat internasional, dan peningkatan citra departemen Informatika sebagai unggulan di tingkat nasional dan semakin dikenalnya departemen Informatika di tingkat internasional.

## **2.8 Fasilitas**

Departemen Informatika ITS memiliki dua belas laboratorium, yaitu Laboratorium Rekayasa Perangkat Lunak (RPL), Laboratorium Komputasi Berbasis Jaringan (KBJ), Laboratorium Komputasi Cerdas dan Visi (KCV), Laboratorium Arsitektur dan Jaringan Komputer (AJK), Laboratorium Grafika, Interaksi, dan Game (GIGa), Laboratorium Algoritma dan Pemrograman (AlPro), Laboratorium Manajemen Cerdas Informasi (MCI), Laboratorium Pemodelan dan Komputasi Terapan (PKT), dua Laboratorium Pemrograman, Laboratorium Pascasarjana S2, dan Laboratorium Pascasarjana S3.

Laboratorium RPL memiliki fokus pada keahlian pengujian perangkat lunak, mengelola proyek perangkat lunak, mengurangi resiko kesalahan pada perangkat lunak, dan membuat perangkat lunak. Laboratorium KBJ menekankan pada pembangunan infrastruktur jaringan, sistem grid, aplikasi jaringan sesuai standar, dan aplikasi multimedia berbasis jaringan. Laboratorium KCV memiliki fokus terhadap manipulasi dan analisis data citra pada berbagai bidang

aplikasi, menerapkan metode sistem cerdas dalam berbagai bidang aplikasi, dan memodelkan dan mengoptimasikan sistem nyata. Laboratorium AJK menekankan pada pembangunan berbagai jenis arsitektur jaringan sesuai standar teknologi terbaru dan menerapkan keamanan jaringan. Laboratorium GIGa menekankan pada desain, pengembangan, dan dokumentasi proses pembuatan game sesuai dengan standar, membuat model tiga dimensi, pemrograman di dalam realitas virtual, dan aplikasi realitas virtual tiga dimensi dengan menggunakan game engine. Laboratorium AIPro menekankan pada perancangan dan analisa algoritma dalam menyelesaikan permasalahan secara efektif dan efisien berdasarkan kaidah pemrograman yang kuat, aplikasi model pemrograman yang mendasari berbagai bahasa pemrograman yang ada, dan dapat memilih bahasa pemrograman untuk menghasilkan aplikasi yang sesuai. Laboratorium MCI memiliki fokus terhadap analisis, sintesa, dan evaluasi proses bisnis dan sistem informasi pada sistem Enterprise, implementasi rekayasa pengetahuan ke dalam suatu aplikasi, investigasi, pengujian, evaluasi kematangan, dan kepatutan terhadap prosedur standar dan tata kelola teknologi informasi, melakukan tata kelola proyek dan sumber daya manusia, dan merancang dan mengimplementasikan solusi basis data terdistribusi dan teknologi Big Data. Laboratorium PKT memiliki fokus pada riset dan kerjasama industri pada bidang pemodelan dan simulasi, peramalan sains, optimasi, dan komputasional ilmiah.

Dua Laboratorium Pemrograman digunakan sebagai tempat perkuliahan dan praktikum, seminar dan workshop, dan ujian online. Dua laboratorium pascasarjana digunakan untuk mahasiswa yang menempuh ilmu S2 dan S3.

## **2.9 Akreditasi**

Pada masa penulisan buku kerja praktik ini, Departemen Informatika ITS dinyatakan sebagai program studi berakreditasi A oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) Indonesia pada tahun 2021 sampai 2026. Departemen Informatika ITS pertama kali dinyatakan sebagai program studi berakreditasi A oleh BAN-PT Indonesia pada 2011. Departemen ITS juga diakui oleh ASEAN University Network (AUN-QA) pada tahun 2015 sampai 2019.

## **2.10 Sumber Informasi**

Sumber referensi informasi mengenai profil perusahaan didapatkan dari situs utama Departemen Informatika Institut Teknologi Sepuluh Nopember Surabaya (<https://www.its.ac.id/informatika> ).

*[Halaman ini sengaja tidak diisi kecuali kalimat ini]*

## **BAB III**

### **TINJAUAN PUSTAKA**

#### **3.1 BPJS**

BPJS adalah badan hukum yang bertujuan untuk memberikan bantuan dana terhadap seluruh rakyat Indonesia untuk memenuhi kebutuhan hidup dasar rakyat Indonesia. (Indonesia, 2011). Menurut Pasal 5 dan 6 dalam Undang-Undang yang sama, BPJS terbagi menjadi dua berdasarkan jenis bantuan yang dapat diberikan, yaitu BPJS Ketenagakerjaan dan BPJS Kesehatan.

BPJS Ketenagakerjaan memberikan bantuan pada bidang ketenagakerjaan yang terdiri dari kecelakaan kerja, hari tua, pensiun, dan kematian. BPJS Ketenagakerjaan diperluas untuk mencakup kehilangan pekerjaan (Indonesia, 2022). Sementara itu, BPJS Kesehatan, yang merupakan salah satu fokus utama dalam kerja praktik ini, memberikan bantuan pada bidang kesehatan.

#### **3.2 Jenis Fasilitas Kesehatan**

Terdapat dua jenis fasilitas kesehatan yang ada di Indonesia, Fasilitas Kesehatan Tingkat Pertama (FKTP) dan Fasilitas Kesehatan Rujukan Tingkat Lanjutan (FKRTL) (Indonesia 2016) (Indonesia 2018). FKTP adalah fasilitas kesehatan yang menyediakan pelayanan bersifat non-spesialistik sedangkan FKRTL adalah fasilitas kesehatan yang menyediakan pelayanan bersifat spesialistik. Jenis FKTP adalah klinik pratama, puskesmas, dan dokter umum sedangkan FKRTL berupa rumah sakit secara umum. Pasien dari FKTP dapat dirujuk ke FKRTL untuk melakukan prosedur tingkat lanjut.

### **3.3 Pembagian Peserta BPJS**

Peserta BPJS terbagi menjadi beberapa jenis berdasarkan pekerjaan mereka. Pembagian tersebut adalah Pekerja Bukan Penerima Upah (PBPU), Pekerja Penerima Upah (PPU), Bukan Pekerja, dan Pekerja Bantuan Iuran (PBI) (Indonesia, 2018).

Peserta PBPU adalah peserta yang bekerja atas risiko mereka sendiri. Peserta PBPU terdiri dari pekerja mandiri dan pekerja tidak mandiri dan tidak menerima gaji. Peserta PPU terdiri atas pejabat negara, anggota Dewan Perwakilan Rakyat Daerah (DPRD), pegawai negeri sipil (PNS), prajurit, polisi, kepala dan perangkat desa, pegawai swasta, dan pekerja lainnya yang menerima gaji. Peserta bukan pekerja adalah peserta yang tidak termasuk dalam pembagian peserta BPJS sebelumnya atau semua orang yang tidak bekerja dengan menerima gaji, upah, dan sebagainya. Peserta bukan pekerja terdiri dari investor, pemberi pekerjaan, penerima dana pensiun, veteran, perintis kemerdekaan dan para kerabatnya yang merupakan janda, duda, atau anak yatim dan/atau piatu, dan orang-orang lainnya yang dapat membayar iuran BPJS. Peserta PBPU, PPU, dan bukan pekerja dapat memberikan kepesertaan BPJS kepada anggota kerabat peserta. Kerabat peserta yang dimaksud adalah pasangan pernikahan peserta, anak-anak mereka yang belum menikah, berpenghasilan, dan mencapai umur 21 atau 25 jika mereka masih menempuh pendidikan formal, orang tua peserta, dan mertua peserta. Peserta PBI, atau penerima bantuan iuran, adalah peserta yang didanai oleh Anggaran Pendapatan Belanja Negara (APBN) atau Daerah (APBD) dalam bidang kesehatan. Peserta PBI ditentukan oleh menteri yang mengurus bidang sosial. Peserta PBI pada umumnya merupakan fakir miskin dan tidak mampu.

### 3.4 ICD

ICD, atau International Classification of Diseases (Klasifikasi Penyakit Internasional), adalah sistem klasifikasi diagnosis penyakit internasional yang ditanggung oleh Organisasi Kesehatan Dunia, atau dalam bahasa Inggris sebagai World Health Organization, atau disingkat sebagai WHO. WHO adalah divisi dari Persatuan Bangsa-Bangsa, atau PBB, yang bertanggung jawab atas kesehatan umum internasional. ICD berguna dalam pencatatan dan statistika yang berhubungan dengan perawatan dalam fasilitas kesehatan. (WHO, 2023)

ICD memiliki sejarah yang panjang, menurut publikasi di WHO (WHO, 2021). Pencetusan klasifikasi penyakit yang akan menjadi ICD dimulai dari seorang statistikawan Prancis bernama Jacques Bertillon di sebuah pertemuan Institut Statistik Internasional di Chicago, Amerika Serikat, pada 1893. Ia membuat daftar klasifikasi penyebab kematian berdasarkan klasifikasi-klasifikasi penyakit yang terdapat di Prancis, Inggris, Jerman, dan Swiss. Daftar klasifikasi tersebut hanya diadopsi oleh beberapa negara di Eropa, yang juga memiliki klasifikasi penyakit lokal. Daftar klasifikasi ini direvisi tiga kali sebelum Bertillon meninggal pada 1922. Oleh karena itu, revisi klasifikasi tersebut, yang pada waktu tersebut bernama Daftar Penyebab Kematian Internasional, dipegang oleh Liga Bangsa-Bangsa. Daftar klasifikasi tersebut direvisi dua kali sebelum Liga Bangsa-Bangsa berubah menjadi Persatuan Bangsa-Bangsa, atau disingkat menjadi PBB, setelah Perang Dunia Kedua berakhir pada 1945. Daftar klasifikasi tersebut menjadi tanggung jawab bagi divisi kesehatan umum PBB, WHO. Setelah revisi keenam dilakukan pada 1946, Daftar Penyebab Kematian Internasional berubah menjadi ICD, dan hasil revisi keenam ICD dinamakan menjadi ICD-6, berdasarkan banyak

revisi yang dilakukan sejak Bertillon mulai mencetuskan klasifikasi penyakitnya. ICD direvisi lima kali dengan revisi terbaru, ICD-11, mulai digunakan pada 1 Januari 2022.

ICD digunakan oleh berbagai fasilitas kesehatan di dunia, termasuk Indonesia. ICD juga digunakan dalam dataset BPJS. Tetapi, menurut isi dataset yang digunakan dan Peraturan Menteri Kesehatan Nomor 76 Tahun 2016 tentang Pedoman Indonesian Case Base Groups (INA-CBG) dalam Pelaksanaan Jaminan Kesehatan Nasional, ICD yang digunakan adalah ICD-10, yang selesai direvisi pada 1989, untuk diagnosis dan ICD-9-CM untuk prosedur penanganannya. Perlu diketahui bahwa ICD-9-CM, dengan CM merupakan kependekan dari Clinical Modification, adalah modifikasi dari ICD-9, yang mana direvisi pada 1975, yang dilakukan oleh Pusat Kesehatan Statistik Kesehatan Amerika Serikat pada 1979. Walaupun kedua ICD merupakan revisi ke-9 dan 10, kedua ICD yang digunakan adalah edisi 2010.

### **3.4.1 ICD-10**

ICD-10 adalah revisi ke-sepuluh ICD yang ditetapkan pada 1989 (WHO, 2004). ICD-10 terdiri dari tiga volume: volume pertama yang terdiri dari klasifikasi penyakit utama, volume kedua yang merupakan panduan penggunaan ICD dan merupakan sumber utama dalam definisi ICD-10 dalam kerja praktik ini, dan volume ketiga yang berisi semua klasifikasi penyakit dari volume pertama yang diurutkan secara alfabetis.

Penyakit dalam ICD-10 terbagi menjadi 22 bab yang dideskripsikan dalam tabel 1 berikutnya. Masing-masing bab terbagi menjadi beberapa blok yang terdiri dari beberapa kategori yang mirip.

Masing-masing kategori terdiri dari subkategori. Penyakit dalam ICD-10 memiliki kode yang terdiri dari satu huruf alfabet yang merupakan kode dari masing-masing bab, dengan beberapa pengecualian, dua angka yang merupakan kode kategori, dan satu angka terakhir yang merupakan kode subkategori. Dalam ICD pada umumnya, angka terakhir merupakan angka desimal dari tiga karakter pertama. Tetapi, dalam dataset dalam kerja praktik ini, angka terakhir bukan merupakan angka desimal.

Contoh penyakit yang terdaftar di ICD-10 adalah E110 yang merupakan kode bagi penyakit diabetes mellitus tanpa ketergantungan insulin dengan koma. E110 berada dalam bab keempat yang memiliki kode E dan merupakan kelompok penyakit endokrin, nutrisi, dan metabolik, kategori diabetes mellitus tanpa ketergantungan insulin yang memiliki kode 11, dan subkategori koma yang memiliki kode 0.

*Tabel 1 Bab ICD-10 Edisi 2010*

Nomor Bab	Deskripsi Bab	Karakter pertama dalam kode
I	Penyakit menular dan parasit	A (Bakteri dan Virus) dan B (Virus, parasit, dan lain-lain)
II	Neoplasma/Tumor	C (Tumor ganas) dan D jika kode D diikuti dengan kategori antara 00 sampai 48 (Tumor jinak)
III	Penyakit darah dan organ pembentuk darah dan penyakit yang berhubungan dengan sistem imun	D dengan kategori selain pada bab II
IV	Penyakit endokrin, nutrisi, dan metabolisme	E
V	Penyakit jiwa	F
VI	Penyakit saraf	G
VII	Penyakit mata dan sekitarnya	H dengan kategori antara 00 sampai 59
VIII	Penyakit telinga dan sekitarnya	H dengan kategori selain pada bab VII
IX	Penyakit pada sistem perdarahan	I
X	Penyakit pada sistem pernafasan	J
XI	Penyakit pada sistem pencernaan	K
XII	Penyakit kulit dan jaringan bawah kulit	L

Nomor Bab	Deskripsi Bab	Karakter pertama dalam kode
XIII	Penyakit pada jaringan otot, tulang, dan penghubung	M
XIV	Penyakit pada saluran kemih	N
XV	Kehamilan, kelahiran, dan masa puerperium (Masa setelah kelahiran bagi seorang ibu)	O
XVI	Gangguan pada masa kelahiran	P
XVII	Masalah malformasi keturunan, deformasi, dan kromosomal	Q
XVIII	Penemuan, gejala, dan pertanda klinis yang tidak dijelaskan dimanapun	R
XIX	Luka, racun, dan konsekuensi penyebab eksternal lainnya	S (Penyakit yang berada di anggota tubuh pasien) dan T (Penyakit yang tidak terikat pada lokasi anggota tubuh)
XX	Penyebab kematian dan penyakit eksternal	V (Kecelakaan dengan kendaraan), W (Kecelakaan lainnya), X (Kecelakaan lainnya (00-59), luka yang disengaja (60-84), dan luka yang dilakukan oleh

Nomor Bab	Deskripsi Bab	Karakter pertama dalam kode
		orang lain (85-99)), Y (Luka yang dilakukan oleh orang lain)
XXI	Faktor yang mempengaruhi status kesehatan dan kontak dengan layanan kesehatan	Z
XXII	Kode dengan tujuan khusus	U

### 3.4.2 ICD-9-CM Prosedur

ICD-9-CM adalah modifikasi dari ICD-9 yang dibuat oleh Pusat Statistik Kesehatan Nasional Amerika Serikat (NCHS) dan Pusat Layanan Perawatan dan Pelayanan Medis Amerika Serikat (CMMS) (CDC, 2021). Dalam perihal BPJS dan kerja praktik ini, prosedur penanganan diagnosis menggunakan prosedur yang ada dalam ICD-9-CM edisi 2010. Prosedur dalam ICD-9-CM terbagi menjadi 17 bab yang diperjelas dalam tabel 2 berikutnya. Masing-masing bab terbagi menjadi beberapa kategori dan subkategori. Kode prosedur ICD-9-CM terdiri dari dua angka pertama sebagai kode bagian, satu angka kategori, dan satu angka subkategori. Contoh kode prosedur ICD-9-M adalah 2311 yang merupakan kode prosedur pencabutan sisa dari akar gigi. Dua angka pertama adalah bab pencabutan dan pengembalian gigi. Satu angka berikutnya berupa kategori operasi pencabutan gigi. Satu angka terakhir adalah subkategori pencabutan sisa dari akar gigi.

*Tabel 2 Bab ICD-9-CM Prosedur*

Bab	Deskripsi Bab
00	Prosedur dan intervensi yang tidak dijelaskan dimanapun
01-05	Operasi jaringan saraf
06 dan 07	Operasi jaringan endokrin
08-16	Operasi mata
17	Prosedur diagnosis dan terapi lainnya
18-20	Operasi telinga
21-29	Operasi hidung, mulut, dan tenggorokan
30-34	Operasi jaringan pernafasan
35-39	Operasi peredaran darah
40-41	Operasi jaringan darah dan limpa
42-54	Operasi jaringan pencernaan
55-59	Operasi jaringan kemih
60-64	Operasi kemaluan pria
65-71	Operasi kemaluan wanita
72-75	Prosedur kelahiran
76-84	Operasi jaringan tulang dan otot
85-86	Operasi kulit
87-99	Prosedur diagnosis dan terapi lainnya

### **3.5 INA-CBG**

Indonesia's Case Based Groups (INA-CBG) adalah sistem pembayaran case-based yang diterapkan di FKRTL Indonesia (Indonesia, 2017). Sistem INA-CBG mulai diterapkan di seluruh FKRTL Indonesia pada 1 Januari 2014. Sistem pembayaran case-based, atau casemix, adalah salah satu jenis dari pembayaran prospektif, dimana biaya total layanan kesehatan sudah diketahui sebelum layanan diberikan. Sistem casemix adalah sistem dimana pelayanan rumah sakit dikelompokkan berdasarkan kemiripan dan sumber daya yang digunakan.

INA-CBG mengelompokan layanan berdasarkan pada ICD-10 revisi tahun 2010 untuk diagnosis yang diberikan dan ICD-9-CM revisi tahun 2010 untuk prosedur yang diberikan. Masing-masing layanan dibentuk dalam suatu kode yang terdiri dari empat bagian: kode Case-Mix Main Groups (CMG), tipe kelompok kasus, spesifikasi kelompok kasus, dan tingkat keparahan kelompok kasus. Informasi untuk masing-masing bagian terdapat pada tabel 3. Informasi lebih lanjut mengenai INA-CBG dan tarifnya terdapat pada Peraturan Menteri Kesehatan Republik Indonesia Nomor 52 Tahun 2016.

Salah satu contoh kode INA-CBG adalah Q-5-44-0. Dalam data FKRTL, deskripsi Q-5-44-0 adalah Penyakit Kronis Lain-Lain. Q berarti pasien menerima layanan rawat jalan. 5 berarti rawat jalan tersebut tidak memiliki prosedur. 44 adalah spesifikasi dari Q. 0 berarti bahwa pasien menerima layanan rawat jalan.

*Tabel 3 Kode INA-CBG*

Posisi Bagian Kode INA-CBG	Deskripsi Bagian Kode INA-BCG	Rentang
1	Kode Casemix Main Groups (CMG)	Kode Biasa: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, S, T, U, V, W, X, Z Kode Spesial: SF, YY, DD, II, RR, CG
2	Kelompok tipe kasus	0-9
3	Spesifikasi CMG	01-99
4	Tingkat keparahan kelompok kasus	0, I, II, III

### **3.5.1 Case-Mix Main Group**

Kode CMG adalah bagian pertama dari kode INA-CBG. Kode CMG terbagi menjadi dua, kode biasa dan kode spesial. Kode biasa terdiri dari satu huruf dari hampir seluruh alfabet latin (A-Z, kecuali Q, R, dan Y). Kode biasa CMG disesuaikan dengan ICD-10. Sementara kode spesial adalah kode yang digunakan sebagai biaya tambahan untuk mengurangi resiko finansial bagi rumah sakit. Kode spesial memiliki dua huruf. Kode CMG dispesifikasikan oleh bagian ketiga dari kode INA-CBG. Kode CMG biasa dan spesial diperjelas pada masing-masing tabel 4 dan 5 berikutnya.

*Tabel 4 Kode CMG Biasa*

Kode CMG	Deskripsi Kode CMG dalam Kelompok
A	Penyakit menular dan parasitik
B	Sistem hepatobiliari dan pankreatik (Hati, limpa, dan pankreas)
C	Myeloproliferative dan Neoplasm (Kanker dan tumor)
D	Sistem penghasil darah dan sistem imun
E	Sistem endokrin, nutrisi, dan metabolisme
F	Kesehatan mental dan tingkah laku
G	Sistem saraf pusat
H	Mata dan sekitarnya
I	Sistem peredaran darah
J	Sistem pernafasan
K	Sistem pencernaan
L	Kulit, jaringan sub-kulit, dan payudara
M	Sistem muskuloskeletal (Otot dan tulang) dan jaringan penghubung (Sendi)
N	Sistem nephro-urineri (Ginjal dan saluran kemih)
O	Persalinan
P	Bayi
Q	Rawat Jalan
S	Luka, keracunan, dan efek buruk dari narkoba
T	Penyalahgunaan dan ketergantungan pada narkoba
U	Telinga, hidung, mulut, dan tenggorokan
V	Sistem reproduksi pria
W	Sistem reproduksi wanita
X	Kode Error
Z	Faktor yang mempengaruhi status kesehatan dan hubungan lainnya dengan layanan kesehatan.

*Tabel 5 Kode CMG Spesial*

Kode CMG	Deskripsi Kode CMG	Keterangan Kode CMG
SF	Kelompok Sub-Akut	Diperuntukan bagi pasien yang mendapatkan layanan psikiatri dan kusta dan memiliki waktu rawat antara 43 sampai 103 hari. Tarif INA-CBG ditambahkan dengan tarif sub-akut.
CF	Kelompok Kronis	Diperuntukan bagi pasien yang mendapatkan layanan psikiatri dan kusta dan memiliki waktu rawat antara 104 sampai 180 hari. Tarif INA-CBG ditambahkan dengan tarif sub-akut dan tarif kronis.
YY	Prosedur Spesial	Penggantian pinggang dan lutut, PCI, Keratoplasti, Pankreatektomi, perbaikan septal prostesis jantung, operasi stereotaktik dan radioterapi, torakotomi, lobektomi atau bilobektomi, vitrektomi, Phacoemulsifikasi, Cholangiograf

Kode CMG	Deskripsi Kode CMG	Keterangan Kode CMG
DD	Obat-obatan Spesial	Streptokinase, Deferiprone, Deferoksamin, Deferasirox, Albumin untuk Septikaemia, Faktor anti-Hemofilia
II	Investigasi Spesial	Obat nuklir, MRI, dan diagnostik dan prosedur penggambaran mata
RR	Protesis Spesial	Elektroda Subdural Grid, Cote Graft, Protesis TMJ, implan pinggang dan lutut, prostesis Evar dan Tevar

### 3.5.2 Kelompok Tipe Kasus

Kelompok Tipe Kasus adalah bagian kedua dari kode INA-CBG yang menunjukkan apakah pelayanan tersebut merupakan rawat inap atau rawat jalan beserta spesifikasinya. Pembagian tipe kasus diperjelas pada tabel 6.

*Tabel 6 Kode Tipe Kasus INA-CBG*

Jenis Tipe Kasus	Kode Tipe Kasus	Deskripsi
Jalan	2	Prosedur Besar Rawat Jalan
	3	Prosedur Signifikan Rawat Jalan
	5	Rawat Jalan Bukan Prosedur
	7	Rawat Jalan Kebidanan
	9	Rawat Jalan Neonatal
Inap	1	Prosedur Rawat Inap
	4	Rawat Inap Bukan Prosedur
	6	Rawat Inap Kebidanan
	8	Rawat Inap Neonatal
Error	0	Kode Error

### **3.5.3 Tingkat Keparahan**

Tingkat Keparahan adalah bagian keempat dari kode INA-CBG. Bagian ini tergantung apakah layanan berupa rawat jalan atau rawat inap. Jika layanan yang diterima merupakan rawat jalan, maka tingkat keparahan tidak dimasukkan. Tingkat Keparahan tertulis sebagai angka Romawi, kecuali 0. Penjelasan masing-masing tingkat keparahan kode INA-CBG diperjelas pada tabel 7.

*Tabel 7 Tingkat Keparahan pada INA-CBG*

Tingkat Keparahan	Deskripsi
0	Rawat Jalan
I	Rawat Inap Ringan (Tanpa komplikasi)
II	Rawat Inap Sedang (Komplikasi tingkat rendah)
III	Rawat Inap Berat (Komplikasi tingkat tinggi)

### **3.6 Tarif INA-CBG**

Tarif INA-CBG adalah tarif paket layanan rumah sakit yang memperhitungkan harga dari beberapa rumah sakit dari suatu kelompok dan data dari Jaminan Kesehatan Nasional, JKN, yang merupakan bagian dari BPJS. Rumah sakit-rumah sakit tersebut dikelompokkan menjadi enam kelompok: rumah sakit umum pusat nasional Dr. Cipto Mangunkusumo, satu kelompok tambahan (RS Jantung dan Pembuluh Darah Harapan Kita (Pusat Jantung Nasional), RS Anak dan Bunda Harapan Kita (Pusat Kesehatan Ibu dan Anak Nasional), dan RS Kanker Dharmais (Pusat Kanker Nasional). Kelompok ini berlaku tarif INA-CBG non-khusus jika layanan yang mereka berikan merupakan non-khusus.), RS Kelas A, B, C, dan D (Baik milik pemerintah maupun swasta) (Khusus RS yang belum memiliki penetapan kelas, maka RS tersebut tergolong RS kelas D). (Indonesia, 2023).

Beberapa rumah sakit yang dikelompokkan merupakan rumah sakit khusus. Jika rumah sakit khusus tersebut memberikan pelayanan selain kekhususan mereka, maka tarif INA-CBG yang diterapkan berada dalam satu kelas dibawah kelas rumah sakit khusus tersebut. Belum ditentukan bagaimana jika aturan tersebut berlaku rumah sakit kelas D yang menyediakan layanan di luar kekhususan mereka. Daftar

jenis bidang rumah sakit khusus yang tertulis dalam peraturan menteri adalah kanker, jantung dan pembuluh darah, jiwa, paru, kusta, ortopedi, mata, THT (Telinga, Hidung, Tenggorokan), dan gigi dan mulut. Jika ada RS Khusus yang mana bidangnya tidak tercatat di atas, maka tarif INA-CBG sesuai dengan kelas yang ditetapkan.

Tarif INA-CBG merupakan biaya pertama pasien sebelum ditambahkan dengan biaya tambahan INA-CBG, jika ada. Setelah itu, biaya total INA-CBG diverifikasi oleh BPJS sehingga biaya total pasien berkurang.

### **3.7 Machine Learning**

Machine Learning (ML) adalah keahlian komputer untuk melakukan tugas lebih efisien tanpa diprogram dengan mempelajari dan membaca pengalaman sebelumnya sendiri agar meningkatkan otomatisasi (Mahesh, 2020) (Janiesch et al., 2021) (Alzubi, 2018). ML dapat terbagi menjadi berbagai bidang diantaranya regresi, clustering, klasifikasi, reinforcement (Memainkan permainan misalnya), dan deteksi anomali (Alzubi 2018). Jenis dari ML terdiri dari supervised learning (Decision Tree, Naive Bayes, SVM, Regression), unsupervised learning (K-Means), semi-supervised learning (Campuran dari supervised dan unsupervised), reinforcement (AI memainkan catur), evolusioner (Perkembangan dan adaptasi), ensemble (Random Forest), Neural Network, pengurangan dimensi (PCA), dan campuran (Alzubi et al. 2018) (Janiesch et al. 2021.) (Mahesh, 2020). Machine Learning digunakan dalam berbagai industri, salah satunya adalah aplikasi transportasi, industri, agrikultur, dan sebagainya.

### **3.8 Anomali dan Jenis-jenisnya**

Dalam berbagai referensi, data anomali adalah sebuah data yang terletak dalam sebuah kelompok data yang tidak biasa dan tidak sesuai dengan pola yang ada di dalam kelompok data tersebut. Tetapi, data anomali masih memiliki definisi yang luas dan tidak jelas. Paper utama yang dapat menjelaskan klasifikasi dari anomali adalah dua paper milik Ralph Foorthuis (2019) (2023) dimana paper diterbitkan pada 2019 menjelaskan efek metode diskretisasi terhadap enam jenis anomali. Pada tahun 2023, enam jenis anomali tersebut dikembangkan menjadi sembilan bagian.

Menurut penelitian Foorthuis, data anomali dibagi menjadi enam bagian berdasarkan jenis data dan hubungan antar-data. Dua tahun kemudian, Foorthuis menambahkan dimensi tingkatan anomali. Sehingga, data anomali dapat diklasifikasikan menjadi sembilan bagian seperti yang ditampilkan di tabel 8. Menurut tabel tersebut, data anomali dapat diklasifikasikan berdasarkan jenis data anomali, jenis hubungan antar data, tingkatan anomali dalam sebuah data.

Tabel 8 Jenis Data Anomali

		Jenis Data Anomali				
		Numerik	Kategori	Campuran		
Jenis Hubungan	Univariat	Anomali Numerik Tidak Biasa	Anomali Kategori Tidak Biasa	Anomali Campuran Sederhana	Atomik	Tingkat Anomali
	Multivariat	Anomali Numerik Multidimensi	Anomali Kategori Multidimensi	Anomali Campuran Multidimensi		
		Anomali Numerik Agregat	Anomali Kategori Agregat	Anomali Campuran Agregat	Agregat	

Data anomali diklasifikasikan menjadi data numerik, data kategori, dan data campuran. Data numerik merupakan data kontinu yang berupa angka, seperti panjang lengan, lebar kepala, dan jumlah sepatu yang dimiliki. Data kategori merupakan data yang hanya dapat didefinisikan dalam kata, seperti nama orang, nama lokasi, dan nama buku. Jenis data ketiga, data campuran, merupakan campuran dari data numerik dan data kategori, seperti umur seorang dari masing-masing kota.

Data anomali juga diklasifikasikan berdasarkan hubungan antar data; univariat dan multivariat. Data univariat adalah sebuah data yang tidak bergantung pada data-data lainnya, contohnya adalah sebuah himpunan yang memiliki nama-nama kota di Jawa Timur. Data multivariat adalah sebuah data yang saling bergantung sehingga dapat membentuk sebuah

pola, contohnya adalah hubungan antara data umur orang dan tinggi orang.

Data anomali juga dapat diklasifikasikan berdasarkan besar anomali; atomik dan agregat. Data atomik, seperti namanya, adalah data satuan yang tidak membentuk suatu kelompok. Contoh situasi data anomali atomik adalah diantara mahasiswa yang memakai baju berwarna biru yang sedang pergi ke Kebun Binatang, terdapat satu orang yang memakai baju kuning. Data agregat adalah data yang membentuk suatu kelompok yang memiliki sifat yang sama. Dalam hal ini, kesamaan data-data tersebut adalah sifatnya yang berbeda daripada data-data lainnya. Contoh situasi data tersebut adalah suatu kelompok mahasiswa dibagi menjadi beberapa kelompok dengan memakai baju yang memiliki warna yang berbeda-beda. Masing-masing kelompok ditugaskan untuk membuat patung dari kertas untuk suatu acara. Tetapi, kelompok mahasiswa yang memakai baju warna hijau membuat patung dari besi bekas.

Menurut penelitian Samariya dan Thakkar (2023), data anomali dibagi menjadi empat jenis; anomali titik (Satu data yang berbeda dengan data-data lainnya), anomali kelompok (Kelompok data yang berbeda dengan data-data lainnya), anomali lokal (Data yang berbeda berdasarkan data-data lainnya yang dekat dengan data tersebut), dan anomali kolektif (Suatu kelompok data yang berbeda dibandingkan dengan seluruh data di dalam dataset). Menurut penelitian Nassif dkk. (2021), data anomali dapat dibagi menjadi anomali titik, anomali konteks (Data anomali yang anomali dalam suatu keadaan), dan anomali kolektif. Walaupun begitu, klasifikasi data-data tersebut dapat diaplikasikan pada klasifikasi milik

Foorthuis dikarenakan klasifikasi tersebut lebih mendalam dan konkrit.

### 3.9 Boxplot

Boxplot adalah suatu algoritma yang dapat menampilkan distribusi dari kelompok data beserta data yang dianggap sebagai anomali (Walker et al., 2018). Boxplot hanya dapat digunakan pada data univariat numerik. Boxplot terdiri dari tiga bagian, kotak, kumis, dan outlier. Bagian kotak boxplot merupakan data-data yang dapat dianggap sebagai data normal. Bagian kumis adalah perpanjangan dari bagian kotak yang menentukan batas atas dan batas bawah dari ke-normalan dari suatu data, sehingga data-data anomali dapat diketahui, yang juga disebut sebagai outlier. IQR, atau rentang antar kuartil, digunakan untuk mencari mencari batas atas dan batas bawah yang ditulis sebagai berikut, dengan  $q_1$  dan  $q_3$  merupakan kuartil awal dan kuartil akhir dari suatu data yang memiliki panjang  $N$ .

$$IQR = q_3 - q_1$$

$$L_{bawah} = q_1 - (1.5 * IQR)$$

$$L_{atas} = q_3 + (1.5 * IQR)$$

Rumus Boxplot demikian merupakan rumus Boxplot tradisional yang ditemukan oleh Spear pada 1952 dan dikembangkan oleh Tukey pada 1972. Nilai 1.5 pada persamaan tersebut ditemukan oleh Tukey setelah melakukan beberapa percobaan dengan menggunakan nilai antara 1 dan 2. Rumus Boxplot sering digunakan oleh berbagai perangkat lunak statistika, seperti seaborn, untuk menemukan nilai anomali.

Tetapi, rumus Boxplot tradisional memiliki kelemahan ketika rumus Boxplot diterapkan pada kumpulan data yang memiliki distribusi data yang asimetris (Babura et al., 2018). Sehingga, kondisi tersebut memunculkan situasi dimana data yang seharusnya outlier merupakan bagian dari data yang normal dan data yang terdeteksi sebagai outlier merupakan data normal. Oleh karena itu, berbagai penelitian memodifikasi rumus Boxplot agar dapat mengakomodasikan data terdistribusi asimetris, salah satunya adalah yang digunakan oleh Walker dkk. (2018).

$$\text{Lower fence} = Q1 - 1.5(Q3 - Q1)\left(\frac{Q2 - Q1}{Q3 - Q2}\right)$$

$$\text{Upper fence} = Q3 + 1.5(Q3 - Q1)\left(\frac{Q3 - Q2}{Q2 - Q1}\right)$$

Rumus Boxplot modifikasi ini akan digunakan dalam data univariat numerik, seperti klasifikasi anomali sebelumnya. Tetapi, jika nilai selisih penyebut adalah 0, maka nilai tersebut diabaikan.

### 3.10 PCA

PCA adalah salah satu teknik pengurangan dimensi yang paling populer digunakan dalam bidang statistik. Fungsi ini menampilkan berbagai data dalam suatu dataset sebagai grafik Kartesian dua dimensi untuk menampilkan informasi baru mengenai dataset tersebut, terutama untuk mencari kemiripan antara berbagai jenis data (Labrin and Urdinez, 2020) dan menjadikan data yang memiliki berbagai atribut menjadi lebih sederhana.

Sementara nilai dimensi data menjadi lebih sederhana, begitu pula penjelasan mengenai proses PCA yang dirujuk dari Jayson (2018). PCA mengurangi dimensionalitas data dengan menggunakan nilai variansi dari masing-masing atribut data. Sebagai contoh adalah dimisalkan sebuah dataset yang memiliki beberapa atribut dan dipilih dua atribut untuk menjadi atribut awal dalam proses PCA. Dua atribut tersebut, dimisalkan  $x$  dan  $y$ , memiliki nilai yang dapat dijadikan sebagai titik dalam sebuah grafik baru, yang dinamakan grafik A. Hal ini menjadi alasan mengapa PCA hanya menerima dataset numerik. Grafik A memiliki titik-titik data dari kedua atribut  $x$  dan  $y$ . Masing-masing atribut memiliki nilai rata-rata yang menjadi titik pusat yang baru bagi grafik A, yang disebut sebagai titik O, dengan posisi titik data dipindahkan secara relatif terhadap titik O. Proses berikutnya ditulis dalam bentuk daftar :

1. Buat sebuah garis, dinamakan garis PC1 (Principal Component 1), dalam grafik A sehingga garis PC1 melewati titik 0 dan jumlah kuadrat jarak antara posisi proyeksi tegak lurus beberapa titik data terhadap garis PC1 dan titik 0 memiliki nilai sebesar mungkin.
2. Lalu, tambah garis PC2 yang tegak lurus terhadap PC1 karena terdapat dua atribut yang digunakan dalam grafik A. Dua garis tersebut membuat grafik baru bernama PC dengan posisi titik data dari garis A direlatifkan berdasarkan garis PC1 dan PC2.
3. Masing-masing nilai garis grafik PC memiliki nilai variansi. Variansi didapatkan dengan membagi nilai eigen masing-masing garis dengan banyak data dikurangi satu. Nilai eigen didapatkan dari jumlah kuadrat jarak antara posisi proyeksi tegak lurus beberapa titik data terhadap masing-masing dan titik 0.

4. Nilai variansi PC1 cenderung lebih besar daripada PC2 dan garis PC berikutnya ketika atribut data lainnya dimasukkan dalam grafik A.
5. Total nilai variansi dari berbagai garis PC adalah 1 yang mana berarti bahwa garis-garis PC tersebut mencakup dan mewakili seluruh data. Tetapi, agar banyak atribut berkurang tetapi seluruh data dapat direpresentasikan, nilai variansi yang dapat digunakan antara 0.8 sampai 0.9 atau tidak mendekati nilai 1 tetapi masih cukup besar.

### 3.11 Algoritma Deteksi Anomali

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) adalah algoritma clustering berdasarkan kepadatan yang paling sering digunakan dalam komunitas ilmiah pada masa ini (Hahsler, 2019). DBSCAN menentukan kelompok dalam sebuah data menggunakan nilai radius terkecil ( $\epsilon$ ) dan nilai batas kepadatan terkecil ( $\rho_{\min}$ ) yang terdiri dari titik-titik data dalam sebuah kelompok.

DBSCAN membaca data dan menentukan posisi data dalam sebuah grafik (Deng, 2020). DBSCAN memilih salah satu titik data dalam grafik dan mencari jarak antara titik-titik data lainnya. Jika sebuah titik memiliki jarak dengan titik yang dipilih lebih kecil atau sama dengan nilai  $\epsilon$ , maka titik tersebut digabungkan dengan kelompok dari titik tersebut. Jika tidak, maka pindah ke titik berikutnya. Jika kelompok titik yang dipilih memiliki kepadatan data yang sama dengan atau lebih besar daripada  $\rho_{\min}$ , maka titik-titik dalam kelompok tersebut ditandai sebagai sudah dikelompokkan. DBSCAN mengulangi lagi proses sebelumnya dengan titik lainnya yang belum ditandai sampai seluruh data dilewati sekali. Jika sebuah titik dikelompokkan terhadap titik-titik lainnya, titik data tersebut

dimasukan dalam sebuah kelompok baru. Setelah semua data dilewati sekali, titik-titik data dari berbagai kelompok dibandingkan. Jika sebuah kelompok memiliki titik-titik data yang sama dengan kelompok lainnya, maka kelompok tersebut digabungkan. Proses ini berlangsung sampai tidak ada kelompok baru yang dapat dibuat. Hasil dari DBSCAN adalah kelompok-kelompok dalam sebuah data dan titik-titik data yang tidak memiliki kelompok, yang disebut sebagai noise, outlier, atau anomali.

Nilai  $\epsilon$  dapat dicari menggunakan algoritma Nearest Neighbors (Pendregosa, 2011). Algoritma Nearest Neighbors adalah algoritma untuk mencari titik-titik data terdekat dalam sebuah titik data. Jarak antara suatu titik dengan titik-titik data lainnya, termasuk titik yang sama, dapat didapatkan dari algoritma tersebut. Untuk mencari nilai  $\epsilon$ , nilai terkecil dari kelompok jarak antara satu titik data dan satu titik data terdekat selain titik data yang sama.

Isolation Forest adalah algoritma deteksi tidak terarah (Fang et al. 2022). Prinsip Isolation Forest adalah membagi titik-titik data menjadi jenis data tree dengan memilih suatu titik data secara acak yang akan menjadi root dari tree. Titik-titik data yang berada di kedua sisi sebuah titik data akan dipisahkan menjadi node kiri dan kanan dari sebuah titik root. Proses ini berulang pada kedua node sampai pada sebuah nilai atau tidak ada titik data yang dapat dibagi lagi. Titik data anomali dapat ditentukan berdasarkan rata-rata tinggi masing-masing node dalam tree dan batas banyak titik data anomali yang dapat ditentukan. Algoritma ini berdasarkan pada sebuah ide bahwa sebuah data anomali terisolasi terhadap titik-titik data lainnya

Local Outlier Factor, atau LOF, adalah algoritma untuk menentukan sebuah anomali (Budiarto et al., 2019). LOF memiliki dua langkah utama. Langkah pertama adalah menentukan local reachability density suatu titik data, dimisalkan  $a$ , data terhadap titik-titik lainnya. Local reachability adalah hasil dari banyak titik data yang berada di dalam lingkaran yang memiliki jari-jari sepanjang jarak antara titik data  $a$  dan titik terdekat ke- $k$  terhadap  $a$  menurut  $k$ -nearest neighbor dibagi dengan jumlah reachability distance antara titik  $a$  dan titik-titik data di dalam lingkaran  $k$ -nearest neighbor-nya  $a$ , dianggap sebagai  $N(a)$ . Reachability distance adalah nilai terbesar antara jarak antara titik  $a$  terhadap titik terdekat ke- $k$  terhadap  $a$  dan jarak antara titik  $a$  dan titik terdekat terhadap  $a$ .

$$rd(a, b) = \max \{knn(b), d(a, b)\}$$

$$lrd(a) = \frac{n}{\sum_{i=0}^n rd(a, b_i)}$$

$$b \in N(a)$$

$$n = |N(a)|$$

Langkah kedua adalah menentukan nilai Local Outlier Factor. Local Outlier Factor dapat ditemukan untuk masing-masing titik data dengan cara pembagian antara jumlah local reachability density semua titik data dalam  $N(a)$  dan perkalian antara banyak titik data dari  $N(a)$  dan local reachability density titik  $a$ .

$$LOF(a) = \frac{\sum_{i=0}^n lrd(b_i)}{n \cdot lrd(a)}$$

$$b \in N(a)$$

### 3.12 Pengukuran Evaluasi Deteksi Anomali

Evaluasi model deteksi anomali dapat dilakukan menggunakan Confusion Matrix dan algoritma ROC-AUC. Confusion Matrix adalah sebuah matrix yang membandingkan nilai dua jenis kelompok data (Bowers & Zhou, 2019). Salah satu jenis Confusion Matrix adalah Binary Confusion Matrix, dimana hanya ada dua jenis data dalam dua kelompok data, yaitu benar atau salah, 1 atau 0, dan sebagainya. Jenis Confusion Matrix tersebut biasanya digunakan untuk membandingkan kumpulan data tes dan data prediksi. Confusion Matrix tersebut terdiri dari empat jenis nilai, yaitu True Positive (TP, banyak data yang benar di dalam kedua kumpulan data), False Positive (FP, banyak data yang salah dalam kumpulan data tes tetapi benar dalam kumpulan data prediksi), False Negative (FN, banyak data yang benar dalam kumpulan data tes tetapi salah dalam kumpulan data prediksi), dan True Negative (TN, banyak data yang salah di dalam kedua kumpulan data). Keempat jenis nilai tersebut membentuk metrik-metrik penilaian bagi sebuah model prediksi, yaitu nilai akurasi, presisi, recall, spesifiktas positif dan negatif, kappa (K), dan F1 yang mana memiliki rumus sebagai berikut.

$$N = TP + FP + TN + FN$$

$$R = ((TP + FN). (TP + FP) + (FP + TN). (FN + TN))/N^2$$

$$Accuracy = \frac{TP + TN}{N}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Specificity_{pos} = \frac{TN}{FP + TN}$$

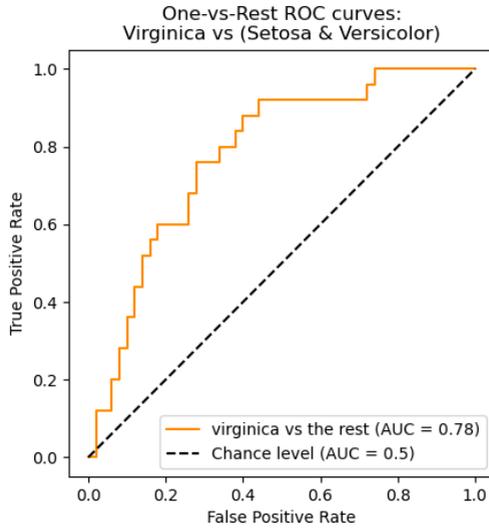
$$Specificity_{neg} = \frac{FP}{FP + TN}$$

$$Specificity_{neg} = 1 - Specificity_{pos}$$

$$K = \frac{Accuracy - R}{1 - R}$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

Diantara metrik-metrik pengukuran tersebut, metrik recall dan spesifiktas positif saling berhubungan, dimana nilai recall menjadi lebih besar dan spesifiktas negatif lebih kecil dan sebaliknya. Kedua metrik tersebut dapat membentuk menjadi sebuah kurva bernama ROC (Receiver Operating Characteristics). Dalam ROC, nilai recall dan spesifiktas negatif juga disebut sebagai nilai tingkat positif asli (True Positive Rate, TPR) dan nilai tingkat positif palsu (False Positive Rate, FPR). Daerah di bawah kurva ROC bernama AUC (Area Under the Curve). Sehingga, metode ini bernama ROC-AUC. Lebih besar nilai ROC-AUC menjadi lebih baik. Contoh dari grafik ROC-AUC terdapat pada gambar 1.



*Gambar 1 Contoh Grafik ROC-AUC  
(Pedregosa et al., 2011)*

*[Halaman ini sengaja tidak berisi kecuali kalimat ini]*

## **BAB IV**

### **ANALISIS DAN PERANCANGAN PROGRAM**

#### **4.1 Analisis Dataset**

Dataset BPJS diambil dari BPJS. Dataset BPJS yang didapatkan terdiri tiga dataset; dataset kepesertaan, dataset kunjungan FKRTL, dan dataset kunjungan FKTP non-kapitasi. Dataset kepesertaan terdiri dari 143.496 peserta, dataset kunjungan FKRTL terdiri dari 5.096.036 kunjungan, dan dataset kunjungan FKTP non-kapitasi memiliki 8.020. Dataset kunjungan FKRTL dan FKTP non-kapitasi memiliki data yang merujuk pada dataset kepesertaan. Dataset kunjungan FKRTL memiliki 129.350 peserta unik dan data kunjungan FKTP non-kapitasi memiliki 1.803 peserta unik. Dari 1.803 peserta tersebut, 1700 diantaranya juga terdaftar dalam dataset FKRTL.

Masing-masing dataset dilakukan Exploratory Data Analysis (DEA) untuk mengetahui karakteristik pada masing-masing data, diantaranya adalah data kosong dan konteksnya, distribusi data, nilai rata-rata, median, dan modus, outlier (Sahoo et al., 2019), dan isi dari data tersebut. Selain itu, masing-masing atribut dalam dataset akan dilakukan korelasi terhadap atribut lainnya yang memiliki potensi hubungan dengan atribut tersebut agar dapat ditemukan informasi baru di dalam data tersebut. Diketahui bahwa analisis ketiga dataset tersebut dilakukan sebelum dan setelah melakukan pre-processing.

### **4.1.1 Dataset kepesertaan**

Dataset kepesertaan memiliki delapan belas kolom yang menjelaskan identitas peserta, fasilitas kesehatan tempat peserta mendaftar, dan status peserta menurut BPJS. Masing-masing peserta memiliki nomor peserta (PSTV01) yang unik. Identitas peserta terdiri dari nomor keluarga (PSTV02), tanggal lahir (PSTV03), hubungan keluarga (PSTV04), jenis kelamin (PSTV05), status perkawinan (PSTV06), daerah tempat tinggal (PSTV09 dan PSTV10), dan tahun meninggal (PSTV18). Masing-masing peserta mendaftar BPJS di faskes. Data faskes tersebut memiliki informasi berupa kepemilikan faskes (PSTV11), jenis faskes (PSTV12), dan lokasi faskes (PSTV13 dan PSTV14). Masing-masing peserta diidentifikasi oleh BPJS berupa kelas rawat (PSTV07), segmentasi (PSTV08), bobot (PSTV15), tahun pengambilan data (PSTV16), dan status (PSTV17). Informasi mengenai dataset peserta BPJS terdapat pada lampiran 1.

#### ***4.1.1.1 Data Keluarga Peserta***

Masing-masing peserta memiliki nomor keluarga pada PSTV02. Tetapi, tidak semua peserta memiliki nomor keluarga yang unik. Hal ini dikarenakan beberapa peserta memiliki anggota keluarga yang juga mendaftar menjadi peserta BPJS. Sehingga, beberapa peserta memiliki nomor keluarga yang sama dengan peserta lainnya yang berarti bahwa peserta tersebut memiliki hubungan keluarga dengan peserta, seperti yang tertulis pada PSTV04. Dari PSTV02, terdapat 567 nomor keluarga yang dimiliki oleh lebih dari satu orang. Diantara 567 keluarga tersebut, hanya ada satu keluarga yang terdiri dari tiga anggota. Keluarga lainnya terdiri dari dua orang.

PSTV04 adalah jenis hubungan keluarga seorang peserta terhadap peserta lainnya, jika memiliki hubungan. Dalam PSTV04, hubungan keluarga peserta terdiri dari peserta, suami, istri, anak, dan tambahan (hubungan keluarga selain suami, istri, dan anak). Peserta yang memiliki hubungan keluarga peserta merupakan peserta yang memiliki anggota keluarga lainnya yang mendaftar ke BPJS terlebih dahulu. Dari tabel 9 yang menjelaskan banyak data PSTV04, 59,3 peserta memiliki jenis hubungan keluarga berupa peserta pada PSTV04. 40,28 persen dari seluruh peserta BPJS memiliki status hubungan keluarga selain peserta pada PSTV04. Ini dapat diasumsikan bahwa para peserta tersebut memiliki anggota keluarga yang mendaftar BPJS sebelumnya, tetapi tidak tertulis di dalam dataset kepesertaan yang digunakan. Diantara 567 keluarga, terdapat 32 keluarga yang tidak memiliki anggota keluarga peserta. Informasi lebih lanjut terdapat pada tabel 10

*Tabel 9 Jenis Hubungan Keluarga Peserta BPJS*

Jenis Hubungan Keluarga	Banyak Peserta	Persentase
Peserta	85.100	59,30
Istri	47.079	32,81
Suami	5.331	3,71
Tambahan	4.686	3,27
Anak	1.300	0,91

*Tabel 10 Banyak Keluarga Peserta BPJS yang Memiliki Dua Pasangan dan Hubungan keluarganya*

	Peserta	Istri	Suami	Anak	Tambahan
Peserta	0	424	87	10	12
Istri		3	0	5	2
Suami			0	0	1
Anak				2	0
Tambahan					19

Menurut PSTV05, yang berisi data jenis kelamin peserta, terdapat 59,81 persen peserta perempuan. Hal ini membentuk rasio seks sebesar 0,67 yang didapatkan dari membagi antara banyak peserta laki-laki dan perempuan. Visualisasi dari nilai tersebut terdapat pada tabel 11.

*Tabel 11 Jenis Kelamin Peserta BPJS*

Jenis Kelamin	Banyak Peserta	Persentase
Perempuan	85.829	59,81
Laki-Laki	57.667	40,19

Masing-masing peserta memiliki status perkawinan yang tertulis dalam kolom PSTV06. 83.72% dari semua peserta telah melakukan pernikahan (Kawin), sementara 9.48% dari semua peserta telah melakukan proses perceraian (Cerai) dan

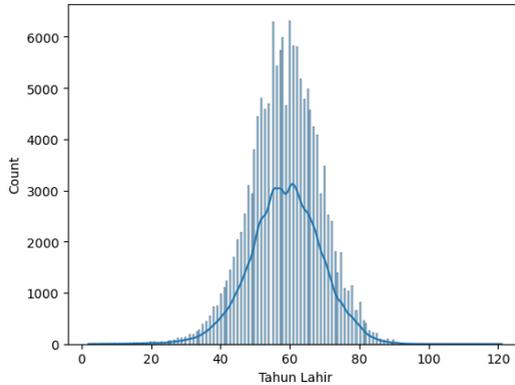
6.8% dari semua peserta belum melakukan pernikahan (Belum Kawin). Visualisasi ini terdapat pada tabel 12.

*Tabel 12 Status Hubungan Perkawinan Peserta BPJS*

Status Hubungan	Banyak Peserta	Persentase
Kawin	120.142	83,72
Cerai	13.602	9,48
Belum Kawin	9.752	6,80

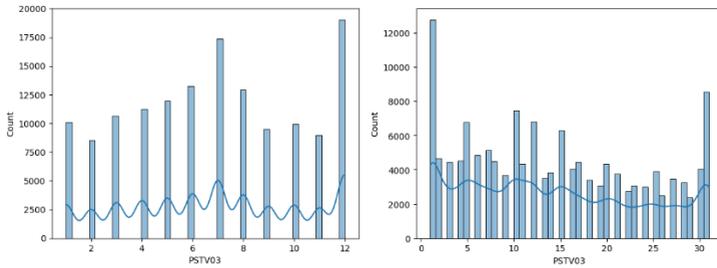
#### ***4.1.1.2 Tanggal Lahir dan Umur Peserta***

Tanggal lahir peserta pada PSTV03 memiliki format tahun-bulan-hari (YYYY-MM-DD). Menurut PSTV16, umur peserta dapat ditarik dengan mengurangi 2020 dari PSTV16 dengan tahun lahir peserta pada PSTV03. Dari hasil tersebut, umur peserta memiliki rentang dari 2 tahun sampai 121 tahun, dengan mayoritas peserta berasal dari rentang umur 50 sampai 70 tahun, umur peserta terbanyak adalah 60 tahun, dan rata-rata umur peserta adalah 59 tahun, yang mana merupakan peserta yang tergolong dalam umur senior. Distribusi umur peserta tergolong sebagai persebaran distribusi normal, seperti yang tergambar pada gambar 2.

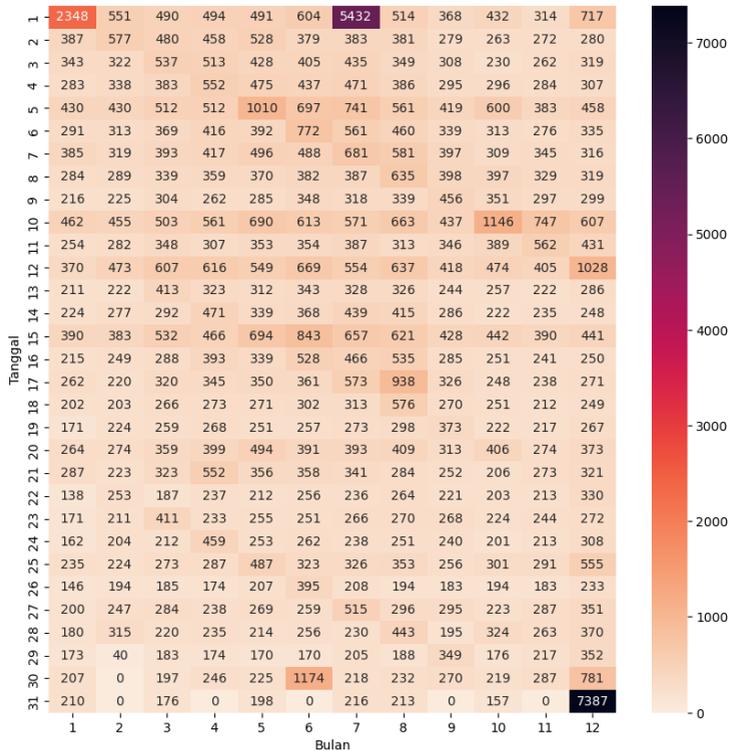


*Gambar 2 Distribusi Usia Peserta BPJS*

Menurut tanggal lahir peserta, bulan Juli dan Desember menjadi bulan lahir yang paling banyak dan tanggal 1 dan 31 menjadi tanggal lahir yang paling banyak di dataset kepesertaan. Jika masing-masing peserta dibagi berdasarkan penanggalan, tanggal lahir yang memiliki jumlah yang lebih menonjol adalah tanggal 1 Januari, 31 Desember, dan 1 Juli, yang mana merupakan awal, akhir, dan pertengahan tahun. Selain itu, terdapat juga sebuah pola tanggal lahir dimana nilai tanggal lahir memiliki nilai yang sama dengan nilai bulan lahir, seperti tanggal 5 bulan 5, tanggal 6 bulan 6, dan sebagainya. Grafik dan heatmap yang menjelaskan distribusi tersebut terdapat pada gambar 4.



Gambar 3 Distribusi Bulan dan Tanggal Lahir Peserta BPJS



Gambar 4 Heatmap Bulan dan Tanggal Lahir Peserta

#### ***4.1.1.3 Daerah Asal Peserta dan Faskes Tempat Pendaftaran Peserta***

Daerah asal peserta terbagi menjadi dua, provinsi (PSTV09) dan kabupaten/kota (PSTV10). Provinsi dan kabupaten/kota asal peserta, sebagaimana provinsi dan kabupaten/kota yang tertulis di dataset lainnya, merujuk pada 34 provinsi dan kabupaten/kota yang berada di dalamnya di Indonesia pada saat dataset dibuat. Dalam kerja praktik ini, 34 provinsi dikelompokkan berdasarkan pulau-pulau besar di Indonesia untuk visualisasi data. 80 peserta yang berada di dalam dataset memiliki daerah asal yang tidak terdefiniskan. Menurut PSTV10, peserta-peserta tersebut berasal dari luar negeri atau tidak diketahui asalnya. Distribusi peserta BPJS tertinggi cenderung berada di pulau Jawa, dengan persentase sebesar 54,42%, diikuti oleh peserta dari pulau Sumatra dengan persentase 22.57%. Visualisasi data ini terdapat pada tabel 13.

Masing-masing faskes tempat peserta terdaftar juga memiliki provinsi (PSTV13) dan kabupaten/kota (PSTV14). Distribusi faskes tempat terdaftar memiliki kemiripan dengan daerah asal peserta. Walaupun begitu, terdapat 1.257 peserta yang terdaftar di luar daerah asal mereka. Peserta tersebut termasuk 80 peserta yang memiliki daerah asal tidak terdefiniskan, dengan 76 orang mendaftar di faskes yang terletak di pulau Jawa. Visualisasi data ini terdapat pada gambar 14.

*Tabel 13 Daerah Asal Peserta BPJS*

Nama Kelompok Daerah	Jumlah	Persentase
Jawa	78.804	54,42
Sumatra	32.389	22,57
Sulawesi	12.367	8,62
Bali dan Nusa Tenggara	7.366	5,13
Kalimantan	10.191	7,10
Maluku dan Papua	3.019	2,10
Tidak Diketahui	80	0,06

*Tabel 14 Daerah Asal Faskes Pendaftaran Peserta BPJS*

Nama Daerah	Jumlah	Persentase
Jawa	78.406	54,64
Sumatra	32.148	22,40
Sulawesi	12.335	8,60
Bali dan Nusa Tenggara	10.258	7,15
Kalimantan	7.412	5,17
Maluku dan Papua	2.937	2,05
Tidak Diketahui	0	0,00

#### **4.1.1.4 Data Deskripsi Faskes Pendaftaran**

Masing-masing peserta mendaftar BPJS di faskes. Masing-masing faskes yang ada di dalam dataset memiliki data berupa kepemilikan (PSTV11), jenis (PSTV12), dan lokasi faskes (PSTV13 dan PSTV14). Kepemilikan faskes dapat dibagi menjadi dua, pemerintahan dan swasta. Faskes milik pemerintahan terbagi menjadi pemerintahan tiap sub-daerah (Pusat, provinsi, dan kabupaten/kota), BUMN, TNI (Angkatan Darat, Laut, dan Udara), dan

Kepolisian Republik Indonesia. Faskes milik pemerintahan dalam dataset mencakup 64.22 persen dari seluruh dataset, dengan faskes milik pemerintahan kabupaten/kota mencakup 58.82 persen dari seluruh dataset. Visualisasi ini terdapat pada tabel 15.

*Tabel 15 Kepemilikan Faskes Pendaftaran Peserta BPJS*

Pemerintah atau Swasta	Kepemilikan	Banyak Peserta	Persentase
Pemerintahan	Pemerintahan Kabupaten/Kota	84.405	58,82
	BUMN	2.196	1,53
	Pemerintahan Provinsi	1.923	1,34
	TNI AD	1.508	1,05
	POLRI	1.398	0,97
	TNI AL	383	0,27
	TNI AU	341	0,24
	Vertikal/Pusat	10	0,01
	Total	92.164	64,23
Swasta	Swasta	51.332	35,77

Jenis faskes tempat peserta mendaftar terbagi menjadi tiga, yaitu puskesmas, klinik pratama, dan dokter umum. Puskesmas menjadi faskes dengan jumlah terbanyak di dataset kepesertaan dengan besar 60.01 persen dari seluruh dataset. Visualisasi ini terdapat pada tabel 16. Ketika PSTV12 dikorelasikan dengan PSTV11, semua faskes puskesmas dimiliki oleh pemerintahan provinsi dan kabupaten/kota sementara semua dokter umum dimiliki oleh BUMN dan Swasta, seperti pada tabel 17.

*Tabel 16 Jenis Faskes Pendaftaran Peserta BPJS*

Jenis Faskes	Banyak Peserta	Persentase
Puskesmas	86105	60.01
Klinik Pratama	36173	25.21
Dokter Umum	21218	14.79

*Tabel 17 Kepemilikan dan Jenis Faskes Pendaftaran Peserta BPJS*

	Puskesmas	Klinik pratama	Dokter umum
Vertikal/Pusat	0	10	0
Pemerintah Provinsi	1860	63	0
Pemerintah kabupaten/kota	84245	160	0
POLRI	0	1398	0
TNI AD	0	1508	0
TNI AL	0	383	0
TNI AU	0	341	0
BUMN	0	2154	42
Swasta	0	30156	21176

#### **4.1.1.5 Status Peserta Menurut BPJS**

Kelas perawatan peserta (PSTV07) terdiri dari tiga tingkat, yaitu kelas I, II, dan III. Kelas rawat yang paling banyak adalah kelas I dengan besar 38.71 persen dari seluruh dataset diikuti oleh kelas III dengan besar 38.44 persen dari seluruh dataset. Visualisasi ini terdapat pada tabel 18.

*Tabel 18 Tingkat Kelas Perawatan Peserta BPJS*

Tingkat Kelas Perawatan	Banyak Peserta	Persentase
Pertama	55.547	38.71
Kedua	32.791	22.85
Ketiga	55.158	38.44

Peserta dibagi berdasarkan jenis pekerjaan mereka pada kolom PSTV08, yaitu pekerja bukan penerima upah (PBPU), pekerja penerima upah (PPU), bukan pekerja, dan pekerja bantuan iuran (PBI) yang didanai oleh anggaran daerah (APBD) atau anggaran nasional (APBN). Dua jenis pembagian terbanyak adalah PBPU dan PPU dengan besar masing-masing sebesar 29.78 dan 29.23 persen dari seluruh data. Visualisasi ini terdapat pada tabel 19. Semua peserta yang tergolong PBI hanya memiliki kelas rawat tingkat III.

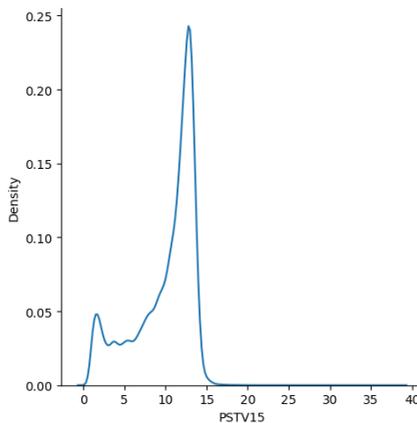
*Tabel 19 Segmentasi Peserta BPJS*

Segmentasi Peserta	Banyak Peserta	Persentase
PBPU	42.738	29,78
PPU	41.942	29,23
PBI APBN	15.116	10,53
PBI APBD	18.500	12,89
Bukan Pekerja	25.200	17,56

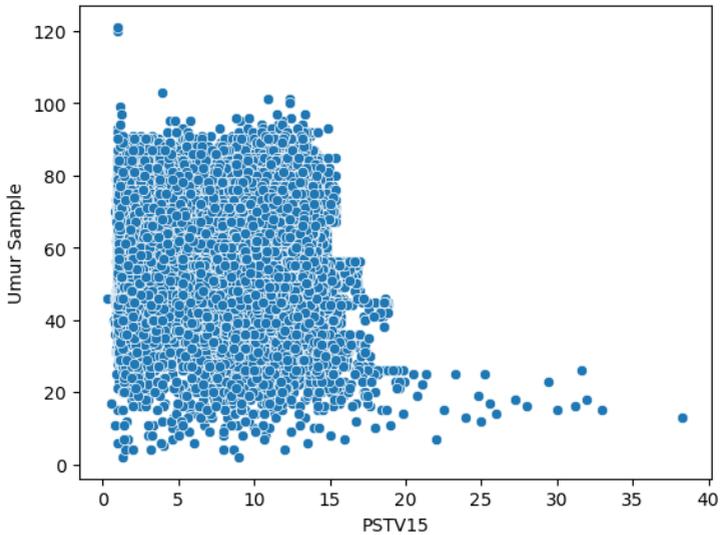
Masing-masing peserta di peserta BPJS memiliki bobot dalam PSTV15. Tetapi, tidak ada informasi lebih lanjut mengenai deskripsi PSTV15. Walaupun begitu, menurut rentang PSTV15, yang memiliki rentang antara 0,35 sampai 38,3, PSTV15 dapat diasumsikan bukan merupakan berat badan.

Salah satu asumsi penjelasan PSTV15 adalah tingkat kepentingan dari masing-masing peserta dalam menggunakan jasa BPJS. Bobot peserta memiliki rata-rata 10 dan nilai yang paling banyak berkisar antara 10 sampai 15. Distribusi bobot peserta tidak biasa, dengan bobot peserta dapat dibagi menjadi tiga berdasarkan besar distribusi dalam visualisasi persebaran bobot peserta, yaitu antara 0 dan 10, 10 dan 15, dan lebih dari 15. Distribusi ini divisualisasikan pada gambar 5.

Diantara semua atribut yang ada dalam dataset peserta BPJS, hanya atribut umur yang memiliki korelasi dengan bobot peserta, dengan bobot peserta diatas 20 dimiliki oleh peserta yang berumur di bawah 27 tahun dan peserta yang berumur di atas 100 tahun memiliki bobot dibawah 4. Penyebaran ini divisualisasikan pada gambar 6.



*Gambar 5 Distribusi Bobot Peserta BPJS*



*Gambar 6 Scatterplot Bobot dan Usia Peserta BPJS*

Dalam dataset kepesertaan, terdapat atribut bernama PSTV16 yang berisi tahun dimana data masing-masing peserta diambil. Tetapi, atribut ini hanya memiliki satu data, yaitu 2020.

Peserta memiliki tiga jenis status pada PSTV17 berdasarkan keaktifan mereka, yaitu aktif, tidak aktif dan meninggal. 94.59 persen dari seluruh peserta merupakan peserta aktif. Visualisasi ini terdapat pada tabel 20.

*Tabel 20 Status Peserta BPJS*

Status Kepesertaan	Banyak Peserta	Persentase
Aktif	135.730	94,59
Tidak Aktif	3.430	2,39
Meninggal	4.336	3,02

PSTV18 berisi tahun meninggal peserta BPJS. PSTV18 memiliki 92,84 data kosong, yang mengasumsikan bahwa peserta yang tidak memiliki tahun meninggal belum meninggal pada waktu penulisan dataset peserta BPJS. Tahun meninggal peserta yang tertulis adalah 2019 dan 2020.

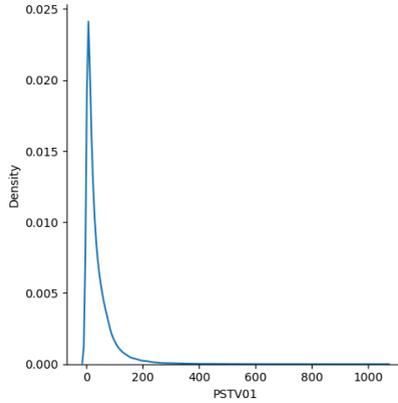
#### **4.1.2 Dataset Kunjungan FKRTL**

Dataset Kunjungan FKRTL dibagi menjadi sembilan bagian, yaitu data dari luar dataset FKRTL, data deskripsi kunjungan FKRTL, diagnosis ICD-10, diagnosis INA-CBG, deskripsi faskes perujuk, prosedur, dan tarif. Informasi lebih lanjut mengenai dataset kunjungan FKRTL terdapat pada lampiran 2.

##### **4.1.2.1 Data dari luar dataset**

Dataset kunjungan FKRTL memiliki tiga atribut yang berasal dari data kepesertaan, yaitu ID peserta (PSTV01), ID Keluarga (PSTV02), dan bobot peserta (PSTV15), dan satu data yang berasal dari dataset kunjungan FKTP Kapitasi (FKP02) yang tidak ada di dalam penelitian ini. PSTV01 dan PSTV02 di dataset kunjungan FKRTL memiliki data yang sama, sementara PSTV15 dari dataset kunjungan FKRTL memiliki perbedaan tetapi sangat kecil, dengan perbedaan terbesar memiliki besar  $7 \cdot 10^{-15}$ . Sehingga, PSTV15 secara relatif sama dengan PSTV15 di dataset kepesertaan.

Menurut PSTV01, terdapat 129.350 peserta yang mengunjungi FKRTL, rata-rata seorang pasien mengunjungi FKRTL sebanyak 39,4 kali, dan pasien dengan kunjungan terbanyak memiliki 1.058 kunjungan. Visualisasi ini terdapat pada gambar 7.



*Gambar 7 Distribusi Banyak Kunjungan FKRTL*

FKP02 terdiri dari kode kunjungan FKTP Kapitasi dan data kosong. Masing-masing peserta yang pernah mengunjungi FKRTL dapat mengunjungi FKTP Kapitasi sebelumnya atau tidak. Dari seluruh kunjungan FKRTL, terdapat 55,67 persen kunjungan tidak mengunjungi FKTP Kapitasi sebelumnya. Sehingga, data FKP02 diubah menjadi data peserta yang mengunjungi FKTP Kapitasi sebelumnya. Visualisasi ini terdapat pada tabel 21.

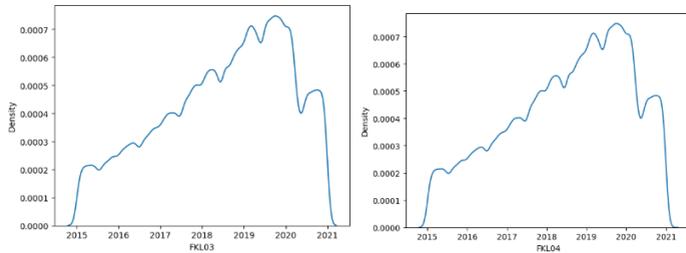
*Tabel 21 Kunjungan ke FKTP Kapitasi*

Pernah Mengunjungi FKTP Kapitasi Sebelumnya	Banyak Kunjungan	Persentase
Ya	2258932	44.33
Tidak	2837104	55.67
Total	5096036	100

#### **4.1.2.2 Data deskripsi kunjungan FKRTL**

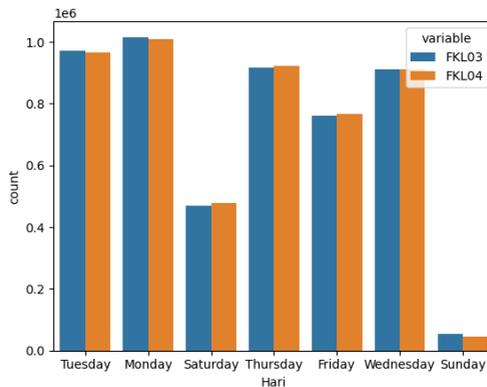
Data deskripsi kunjungan FKRTL adalah kelompok kolom yang terdiri dari ID kunjungan (FKL02) yang unik, tanggal kedatangan pasien (FKL03), tanggal pulang pasien (FKL04), provinsi dan kabupaten/kota lokasi FKRTL (FKL05 dan FKL06), kepemilikan, jenis, dan tipe FKRTL (FKL07, FKL08, FKL09), tingkat pelayanan (FKL10), jenis poli yang digunakan (FKL11), segmentasi peserta pada saat akses FKRTL (FKL12), kelas iuran premi peserta (FKL13), dan status kepulangan FKRTL (FKL14)

Menurut tanggal kunjungan dan kepulangan pasien, rentang tahun kunjungan FKRTL adalah antara 2015 dan 2021 dengan jumlah peserta yang memasuki FKRTL memiliki pola meningkat mulai dari 2015 sampai 2019. Tetapi, pada 2020, jumlah peserta yang mengunjungi FKRTL menurun dengan drastis pada bulan April. Salah satu kemungkinan penyebab penurunan jumlah kunjungan adalah munculnya wabah COVID-19 pada 2020. Lalu, ada 356 peserta yang keluar, atau memiliki rencana keluar, dari FKRTL pada awal 2021, pada bulan Januari dan Februari. Setiap tahun, selalu ada penurunan kunjungan pada tengah tahun. Puncak kunjungan peserta ke FKRTL adalah pada tahun 2019. Visualisasi ini terdapat pada gambar 8.



*Gambar 8 Grafik Distribusi Tahun Kunjungan FKRTL dan Pulang dari FKRTL*

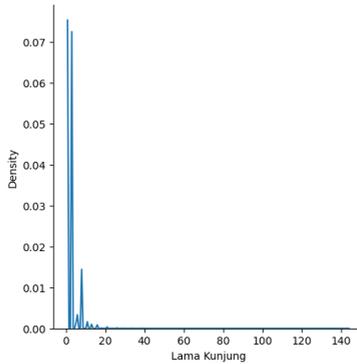
Dalam hari kedatangan ke FKRTL dan kepulangan dari FKRTL, hari Sabtu dan Minggu, hari yang disebut sebagai hari akhir pekan, merupakan hari dimana pengunjung FKRTL menurun, seperti pada gambar 9.



*Gambar 9 Distribusi Hari Kunjungan FKRTL*

Lama kunjungan seorang pasien dapat diambil dengan mencari selisih dari tanggal kedatangan dan

tanggal kunjungan ditambah satu hari karena seorang pasien yang datang dan pulang pada hari yang sama terhitung sebagai satu hari, bukan nol hari. 94,74% peserta berada di FKRTL dalam satu hari saja. Rata-rata lama waktu pasien mengunjungi FKRTL adalah 1,22 hari, atau sekitar 5,28 jam. Lama waktu kunjungan terlama di dataset kunjungan FKRTL adalah 144 hari. Distribusi lama kunjungan FKRTL cenderung lebih menonjol ketika lama kunjungan di bawah 10 hari. Visualisasi ini terdapat pada gambar 10.



*Gambar 10 Distribusi Lama Kunjungan FKRTL*

Lokasi provinsi dan kabupaten/kota FKRTL yang dikunjungi tertulis pada kolom FKL05 dan FKL06. Sebagaimana pada lokasi asal peserta, catatan kunjungan FKRTL di pulau Jawa lebih sering muncul daripada daerah-daerah lainnya dengan banyak 45.72 persen. Visualisasi ini terdapat pada tabel 22.

*Tabel 22 Daerah Asal FKRTL*

<b>Nama Daerah</b>	<b>Banyak Kunjungan</b>	<b>Persentase</b>
Jawa	3.232.108	63,42
Sumatera	1.019.900	20,01
Sulawesi	291.790	5,73
Kalimantan	267.479	5,25
Nusa Tenggara	235.027	4,61
Maluku-Papua	49.729	0,98

Seperti faskes yang berada di data peserta, FKRTL kunjung peserta BPJS memiliki data kepemilikan (FKL07) dan jenis (FKL08). Kepemilikan FKRTL terbagi menjadi dua, yaitu pemerintah dan swasta, yang memiliki rasio banyak kunjungan dari masing-masing jenis kepemilikan mendekati 1:1. Visualisasi ini terdapat pada tabel 23. Jenis FKRTL di dalam data ini terdiri dari dua jenis, yaitu rumah sakit biasa dan klinik, yang terdiri dari klinik utama dan klinik pratama. FKRTL berjenis rumah sakit menjadi jenis FKRTL yang paling banyak dikunjungi dengan mencakup 98,41 persen dari seluruh dataset. Visualisasi ini terdapat pada tabel 24.

FKRTL kunjung peserta BPJS memiliki tipe FKRTL pada FKL09 yang terdiri dari rumah sakit bertipe yang dibagi menjadi rumah sakit pemerintahan dan swasta, terdiri dari empat tingkat (I, II, III, dan IV) dan juga merupakan tipe FKRTL yang paling banyak di dalam data dengan besar 44% untuk masing-masing dari pemerintahan dan swasta, rumah sakit TNI dan POLRI yang juga memiliki empat tingkat, dan rumah sakit khusus. Terdapat 4.972 kunjungan yang tidak memiliki tipe FKRTL dan 1.410 yang tidak tergolong

dalam tipe-tipe sebelumnya, yaitu Non Provider Gawat Darurat dan Klinik Non Rawat Inap, yang hanya dimiliki oleh FKRTL berjenis klinik pratama. Terdapat 37.974 RS milik pemerintah yang bertipe RS Swasta dan 78.808 RS milik swasta yang bertipe RS Pemerintah. Visualisasi ini terdapat pada tabel 25.

*Tabel 23 Kepemilikan FKRTL*

Pemerintahan atau Swasta	Kepemilikan FKRTL	Banyak	Persentase
Swasta	Swasta	2.432.732	47,74
Pemerintahan	Kabupaten/Kota	1.503.107	29,50
	Provinsi	424.081	8,32
	Vertikal	222.169	4,36
	TNI (AD, AL, AU)	277.432	5,44
	BUMN	150.265	2,95
	Polri	86.250	1,69
	Total	2.663.304	52,26

*Tabel 24 Jenis FKRTL*

Jenis FKRTL	Banyak Kunjungan	Persentase
Rumah Sakit	5.014.911	98,41
Klinik Utama	80.817	1,59
Klinik Pratama	308	0,01

*Tabel 25 Tipe FKRTL*

Nama Tipe RS	Jumlah Kunjungan RS	Persentase
RS Swasta Bertipe	2262267	44.39
RS Pemerintah Bertipe	2265985	44.47
RS TNI dan RS Polri	359958	7.06
RS Khusus (Jantung, Jiwa, dsb.)	201444	3.95
RS Lain (Non Provider Gawat Darurat, Klinik Non Rawat Inap)	1410	0.03
No Data	4972	0.10

FKL10 terdiri dari dua jenis, Rawat Jalan Tingkat Lanjut (RJTL) dan Rawat Inap Tingkat Lanjut (RITL). Pelayanan RITL merupakan pelayanan yang menyediakan layanan yang memerlukan penginapan di rumah sakit, sedangkan pelayanan RJTL tidak. Ketika dikaitkan dengan lama kunjung pasien ke FKRTL, pasien yang mendapatkan pelayanan RJTL seharusnya memiliki lama kunjung hanya 1 hari sedangkan pasien yang mendapatkan pelayanan RITL dapat memiliki lama kunjung lebih dari 1 hari. Layanan FKRTL RJTL mencakup 94,68 persen dari seluruh dataset. Visualisasi ini terdapat pada tabel 26.

*Tabel 26 Jenis Pelayanan FKRTL*

Jenis Pelayanan FKRTL	Jumlah Kunjungan	Persentase
RJTL (Rawat Jalan)	4.825.035	94,68
RITL (Rawat Inap)	271.001	5,32

PSTV11 berisikan poli yang digunakan pada tiap kunjungan FKRTL. Poli yang dikunjungi pasien cenderung berbeda-beda untuk setiap FKRTL yang dikunjungi dan belum ditemukan aturan pusat mengenai nama dan singkatan dari 127 poli di rumah sakit. Poli yang memiliki nilai yang menonjol adalah poli INT, yang mungkin berarti poli penyakit dalam (Internal). Poli INT mencakup 40,86 persen dari seluruh dataset. Poli kedua terbanyak adalah SAR yang mencakup 8,36 persen. Visualisasi ini terdapat pada tabel 27. Di dalam atribut poli FKRTL, ada 39 data kosong. Selain itu, ada data di atribut poli bernama Missing yang merupakan poli yang tidak ternama atau bukan sebuah poli. Data kosong di dalam atribut poli dapat diubah menjadi Missing. Sehingga, data Missing dalam PSTV11 mencakup sebesar 5,34 persen dari seluruh dataset.

*Tabel 27 Jenis Poli*

Jenis Poli	Banyak Kunjungan	Persentase
INT	2.082.185	40,86
Bukan INT	3.013.851	59,14

Setiap kunjungan FKRTL, pasien mendapatkan pembagian berdasarkan pekerjaan mereka (FKL12) dan tingkat perawatan yang mereka dapatkan (FKL13). Perlu diketahui bahwa pembagian berdasarkan pekerjaan dan tingkat perawatan masing-masing peserta tidak sama dengan yang ada di dataset kepesertaan. Jika sama, maka nama kolom tersebut bukan FKL12 dan FKL13, melainkan PSTV07 dan

PSTV08, seperti data dari kepesertaan pada PSTV01, PSTV02, dan PSTV15. Pembagian peserta berdasarkan pekerjaan terdiri dari PBPU, PPU, bukan pekerja, dan PBI, baik dari anggaran daerah maupun nasional dengan pembagian terbesar terletak pada PBPU yang mencakup 35,96 persen dari seluruh dataset. Visualisasi ini terdapat pada tabel 29. Sementara, kelas perawatan yang paling besar adalah kelas III dengan besar 96,61 persen dari seluruh dataset. Visualisasi ini terdapat pada tabel 28.

*Tabel 28 Kelas Perawatan FKRTL*

Kelas Perawatan	Banyak Kunjungan	Persentase
Kelas 1	109.707	2,15
Kelas 2	63.282	1,24
Kelas 3	4.923.047	96,61

*Tabel 29 Segmentasi Peserta pada Kunjungan FKRTL*

Segmentasi Peserta	Banyak Kunjungan	Persentase
PBPU	1.832.299	35,96
PPU	1.339.397	26,28
Bukan Peserta	1.211.260	23,77
PBI APBD	415.343	8,15
PBI APBN	297.737	5,84

PSTV14 menuliskan status terakhir peserta setelah mengunjungi FKRTL. Peserta yang pulang dari FKRTL memiliki lima jenis status: sehat, meninggal, merujuk ke tempat lain, dipulangkan paksa, dan tidak diketahui. Visualisasi ini terdapat pada tabel 30.

Dari semua pasien yang mengunjungi FKRTL, 94,60% pasien selalu pulang dengan sehat, 5,39% pasien memiliki satu status pulang meninggal, 0,0085% memiliki dua status pulang meninggal, dan ada 1 pasien yang memiliki 9 status pulang meninggal. Satu pasien tersebut mengunjungi FKRTL sebanyak 332 kali. Belum diketahui apakah status BPJS dapat diberikan kepada kerabat lainnya.

*Tabel 30 Status Pulang Kunjungan FKRTL*

Status Pulang	Banyak Kunjungan	Persentase
Sehat	4.991.307	97,94
Meninggal	7.008	0,14
Merujuk ke Tempat Lain	29.709	0,58
Pulang Paksa	3.247	0,06
Tidak Diketahui	64.765	1,27

#### **4.1.2.3 Data diagnosis peserta**

Diagnosis peserta dalam dataset kunjungan FKRTL dibagi menjadi dua, diagnosis masuk dan diagnosis primer. Diagnosis masuk adalah alasan seorang peserta BPJS mengunjungi fasilitas kesehatan sedangkan diagnosis primer adalah diagnosis yang dilakukan di dalam fasilitas kesehatan dan yang paling penting. Diagnosis peserta terdiri dari delapan kolom yang memiliki kode diagnosis yang merujuk dari ICD-10 edisi 2010. Empat kolom pertama (FKL15, FKL15A, FKL16, FKL16A) menjelaskan diagnosis masuk dan empat kolom berikutnya (FKL17, FKL17A, FKL18, FKL18A) menjelaskan diagnosis primer. Empat kolom dari masing-masing diagnosis

terbagi lagi menjadi dua berdasarkan susunan kode ICD-10. Dua kolom pertama (FKL15, FKL15A, FKL17, 17A) memiliki tiga digit pertama dari kode ICD-10 dan dua kolom berikutnya (FKL16, FKL16A, FKL18, FKL18A) memiliki satu digit tambahan dari tiga digit sebelumnya untuk mengspesifikasikan diagnosis yang diberikan. Masing-masing dari dua kolom tersebut terbagi lagi menjadi dua, satu kolom (FKL15, FKL16A, FKL17, FKL18A) mendeskripsikan diagnosis yang diberikan dengan kata sementara satu kolom lainnya tidak (FKL15A, FKL16, FKL17A, FKL18). Di dalam data diagnosis, terdapat data yang memiliki nama '9999' dimana diagnosis tidak diberikan dan juga adanya data kosong di dalam data diagnosis tersebut. Sehingga, data kosong di dalam data diagnosis akan diganti dengan nilai '9999' yang menandakan bahwa tidak ada diagnosis yang diberikan pada kunjungan tersebut.

Diagnosis masuk dan primer pada tiap kunjungan FKRTL dikelompokkan berdasarkan hierarki pertama, atau bab, dari diagnosis ICD-10. Bab-bab yang paling banyak pada diagnosis masuk kunjungan FKRTL adalah bab IV (Penyakit endokrin, nutrisi, dan metabolisme), XXI (Faktor status kesehatan dan kontak layanan kesehatan), dan IX (Penyakit sistem peredaran darah), yang memiliki kode utama berupa huruf E, Z, dan I. Bab IV, XXI, dan IX masing-masing memiliki besar 26.91%, 17.66%, dan 14.46%. Visualisasi ini terdapat pada tabel 31. Dalam diagnosis masuk, terdapat 6,71 persen dari seluruh kunjungan tidak memiliki diagnosis masuk. Sementara, bab pada diagnosis primer yang paling sering adalah bab XXI

dengan kode Z dengan besar 81,49 persen dari seluruh dataset. Visualisasi ini terdapat pada tabel 32. Terdapat 464 kunjungan yang tidak memiliki diagnosis primer. Menggabungkan dua jenis diagnosis tersebut, terdapat 12 kunjungan FKRTL yang tidak memiliki diagnosis masuk dan primer.

*Tabel 31 Bab Diagnosis Masuk Kunjungan FKRTL*

Bab Diagnosis Masuk	Banyak Kunjungan	Persentase
E	1.371.198	26,91
Z	900.078	17,66
I	736.732	14,46
Lain-Lain	2.088.028	40,97

*Tabel 32 Bab Diagnosis Primer Kunjungan FKRTL*

Bab Diagnosis Primer	Banyak Kunjungan	Persentase
Z	4.152.624	81,49
Lain-lain	943.412	18,51

#### **4.1.2.4 Data INA-CBG**

Data INA-CBG di dalam dataset kunjungan FKRTL terdiri dari enam kolom; kode INA-CBG (FKL19), deskripsi kode INA-CBG (FKL19A), dan empat digit dari FKL19 (FKL20, FKL21, FKL22, dan FKL23). Dalam digit pertama INA-CBG (FKL20), yang merupakan kode Casemix kelompok utama INA-CBG, kode Q (Rawat Jalan) muncul lebih sering dengan besar 69.32%. Visualisasi ini terdapat pada tabel 33. Digit pertama ini dispesifikasikan pada digit ketiga (FKL22). Digit kedua INA-CBG (FKL21)

menjelaskan tipe kelompok kasus yang didapatkan. Tipe kelompok kasus ini terbagi menjadi rawat jalan (2, 3, 5, 7, dan 9) dan rawat inap (1, 4, 6, dan 8) dengan rawat jalan mencakup 94,68 persen dalam dataset kunjungan FKRTL. Visualisasi ini terdapat pada tabel 34. Digit terakhir INA-CBG (FKL23) menjelaskan tingkat keparahan diagnosis yang didapatkan. Tingkat keparahan INA-CBG terdiri dari tiga tingkat (I, II, dan III yang berkorelasi dengan tingkat keparahan ringan, sedang, dan berat) dan hanya berlaku bagi peserta yang mendapatkan layanan rawat inap. Kode INA-CBG memiliki kode error, tetapi kode tersebut tidak ditemukan di dalam dataset kunjungan FKRTL. Visualisasi ini terdapat pada tabel 35.

*Tabel 33 Digit Pertama Kode INA-CBG*

Digit Pertama INA-CBG	Banyak Kunjungan	Persentase
Q	3.532.583	69,32
Bukan Q	1.563.453	30,68

*Tabel 34 Digit Kedua INA-CBG*

Digit Kedua INA-CBG	Banyak Kunjungan	Persentase
Rawat Jalan	3.532.583	69,32
Rawat Inap	1.563.453	30,68

*Tabel 35 Digit Keempat INA-CBG*

Digit Keempat INA-CBG	Banyak Kunjungan	Persentase
Rawat Jalan (0)	4.825.065	94,68
Rawat Inap Ringan (I)	173.649	3,41
Rawat Inap Sedang (II)	74.166	1,46
Rawat Inap Berat (III)	23.156	0,45

#### **4.1.2.5 Data Faskes Perujuk**

Data faskes perujuk peserta terdiri dari lokasi provinsi (FKL25) dan kabupaten/kota faskes perujuk (FKL26), dan kepemilikan (FKL27), jenis (FKL28), dan tipe faskes perujuk (FKL29). Sebagaimana lokasi FKRTL, kunjungan FKRTL memiliki faskes perujuk terbanyak dari pulau Jawa, yang mencakup 62,73 persen dari seluruh dataset. Visualisasi ini terdapat pada tabel 36. Terdapat 36.425 kunjungan yang tidak melalui faskes perujuk dan 53 kunjungan tidak memiliki faskes perujuk yang terdefinisi. Kunjungan yang tidak melalui faskes perujuk memiliki data pada FKL27, FKL28, dan FKL29 bertuliskan 'Bukan Perujuk'.

*Tabel 36 Daerah Asal Faskes Perujuk*

Daerah FKRTL	Jumlah Kunjungan	Persentase
Jawa	3.196.998	62,73
Sumatera	1.023.488	20,08
Sulawesi	288.137	5,65
Kalimantan	264.661	5,19
Nusa Tenggara	234.911	4,61
Maluku-Papua	51.363	1,01
Tidak dirujuk	36.425	0,71
Tidak terdefinisi	53	0,00

Kepemilikan faskes perujuk juga terbagi menjadi dua, yaitu milik pemerintah dan swasta. Kunjungan dimana faskes perujuk merupakan milik pemerintah mencakup 60,61 persen. Visualisasi ini terdapat pada tabel 37.

*Tabel 37 Kepemilikan Faskes Perujuk*

Pemerintahan atau Swasta	Kepemilikan FKRTL	Banyak	Persentase
Swasta	Swasta	1.970.807	38,67
Pemerintahan	Kabupaten/Kota	2.603.328	51,09
	Provinsi	163.724	3,21
	Vertikal	56.076	1,1
	TNI (AD, AL, AU)	135.205	2,65
	BUMN	70.321	1,38
	Polri	60.150	1,18
	Total	3.088.804	60,61
Bukan Rujukan	Bukan Rujukan	36.425	0,71

Jenis faskes perujuk memiliki jenis yang lebih banyak daripada jenis FKRTL, yaitu terdiri dari puskesmas, rumah sakit, klinik (pratama dan utama), dokter umum, apotik, pemerintah (Atau jenis faskes milik pemerintah yang tidak terdefinisi), tidak tertera, dan lain-lain. Jenis faskes lain-lain terdiri dari dokter gigi, laboratorium, PMI, dan optik. 81 kunjungan yang mana faskes perujuknya tidak memiliki jenis faskes semuanya adalah milik swasta. Terdapat 46,21 persen dari seluruh kunjungan dimana faskes perujuk memiliki jenis puskesmas. Visualisasi ini terdapat pada tabel 38.

*Tabel 38 Jenis Faskes Perujuk*

Jenis Faskes Perujuk	Banyak Kunjungan	Persentase
Puskesmas	2.354.857	46,21
Rumah Sakit	1.084.377	21,28
Klinik Pratama	1.055.100	20,70
Dokter Umum	544.600	10,69
Bukan Rujukan	36.425	0,71
Klinik Utama	16.060	0,32
Lain-lain	3.367	0,07
Apotik	993	0,02
Pemerintah (Tidak dispesifikasi)	176	0,00
Tidak tertera	81	0,00

Faskes perujuk memiliki tipe faskes yang lebih banyak. Tipe faskes perujuk terdiri dari rawat inap dan non-rawat inap (Termasuk klinik), RS bertipe (Baik swasta maupun pemerintah), RS TNI dan Polri, RS Khusus, PPK (Pemberi Pelayanan Kesehatan, yang

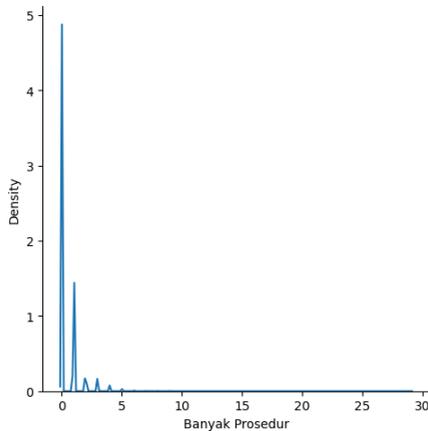
terbagi menjadi PPK Luar Wilayah, PPK Tidak Ditunjuk, dan PPK Lain-lain), apotik (Terbagi menjadi apotik kronis, PRB (Program Rujuk Balik), dan PRB dan kronis), bukan rujukan, tidak tertera, dan lain-lain. Tipe faskes lain-lain mencakup dokter praktek perorangan, dokter gigi, non-provider gawat darurat, laboratorium, PMI (Palang Merah Indonesia), optik, RS CoB (Coordination of Benefit, dimana peserta mendapatkan asuransi dari dua atau lebih perusahaan asuransi), klaim khusus, IVA/Pap Smear (Operasi pada Serviks), dan RS Kelas D Pratama (Rumah sakit yang terletak di daerah terpencil). 51 persen dari seluruh kunjungan memiliki tipe faskes rujukan non-inap. 153 faskes memiliki banyak kunjungan yang tidak tertera. Visualisasi ini terdapat pada tabel 39.

*Tabel 39 Tipe Faskes Perujuk*

Tipe Faskes Perujuk	Banyak Kunjungan	Persentase
Non-Rawat Inap	2.598.881	51,00
Rawat Inap	810.712	15,91
Lain-lain	547.527	10,74
RS Swasta Bertipe	518.519	10,17
RS Pemerintah Bertipe	470.101	9,22
RS TNI Polri	63.565	1,25
RS Khusus	47.947	0,94
Bukan Rujukan	36.425	0,71
PPK	1.213	0,02
Apotik	993	0,02
Tidak tertera	153	0,00

#### 4.1.2.6 Prosedur

Prosedur yang dilakukan pada kunjungan FKRTL ditulis pada FKL30. Satu kunjungan ke FKRTL dapat tidak menggunakan prosedur atau menggunakan prosedur lebih dari satu. Rata-rata pasien mendapatkan 0,59 prosedur dan pasien dengan prosedur terbanyak memiliki banyak prosedur sebanyak 29 prosedur dalam satu kunjungan. Visualisasi ini terdapat pada gambar 11.



*Gambar 11 Distribusi Banyak Prosedur*

#### 4.1.2.7 Biaya FKRTL

Ada tiga jenis tarif di dataset FKRTL; tarif INA-CBG (Tarif awal FKRTL), biaya tambahan, dan tarif verifikasi BPJS. Tarif INA-CBG (FKL32) adalah biaya awal dari rumah sakit yang diberikan kepada peserta. Tarif INA-CBG dikelompokkan berdasarkan lima jenis daerah di Indonesia (FKL31). Visualisasi ini terdapat pada tabel 40.

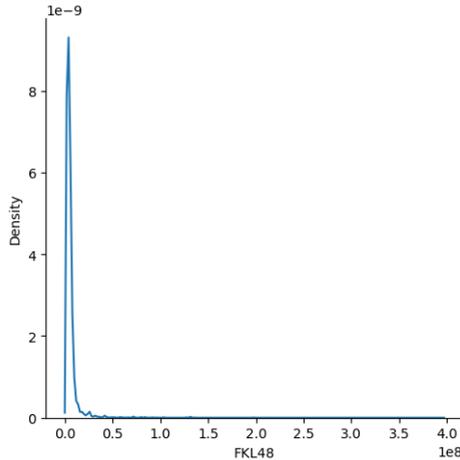
*Tabel 40 Kelompok Tarif INA-CBG*

Kelompok Tarif INA-CBG	Letak Daerah	Banyak Kunjungan	Persentase
Kelas 1	Pulau Jawa	3.232.111	63,42
Kelas 2	Lampung, Riau, Sumatera Barat, Sumatera Selatan, Bali, NTB	660.984	12,97
Kelas 3	Pulau Sumatera selain yang termasuk pada Kelas 2, Kalimantan Barat, Pulau Sulawesi	882.186	17,31
Kelas 4	Pulau Kalimantan selain Kalimantan Barat	218.477	4,29
Kelas 5	Kepulauan Maluku, Pulau Papua, dan NTT	102.278	2,01

Biaya tambahan adalah biaya opsional yang ditambahkan pada tarif awal INA-CBG. Biaya tambahan tersebut terbagi berdasarkan lima kode spesial INA-CBG, yaitu kelompok sub-akut, prosedur spesial, prostesis spesial, investigasi spesial, dan obat-obatan spesial. Biaya tambahan tertulis dari FKL33 sampai FKL46 yang menjelaskan kode, deskripsi kode, dan biaya tambahan yang tertulis. Akan tetapi, setidaknya 99,6 persen dari seluruh kunjungan tidak menggunakan biaya tambahan, dan biaya tambahan

tersebut ditambahkan pada tarif awal INA-CBG sehingga menghasilkan tarif akhir pada FKL47.

Tarif akhir peserta kunjungan FKRTL akan diverifikasi oleh BPJS sehingga tarif akhir peserta dapat berkurang dan nilainya akan dituliskan pada FKL48. Rata-rata biaya verifikasi BPJS adalah 58.597,8 dan biaya verifikasi BPJS yang paling besar adalah 396.933.400. Menurut FKL48, terdapat 14 kunjungan yang tidak memiliki biaya. Visualisasi ini terdapat pada gambar 12.



*Gambar 12 Distribusi Tarif Kunjungan FKRTL Setelah Verifikasi*

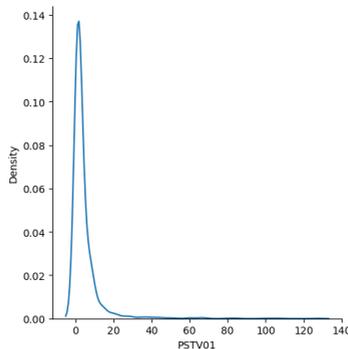
#### **4.1.3 Dataset Kunjungan FKTP Non-Kapitasi**

Dataset kunjungan FKTP non-kapitasi terdiri dari 20 kolom yang terbagi menjadi data dari data peserta BPJS (PSTV01, PSTV02, dan PSTV15), data deskripsi kunjungan (PNK02, PNK03, PNK04, PNK05, PNK06, PNK07, PNK08, PNK09, PNK10, PNK11, dan PNK12), diagnosis (PNK13,

PNK13A, PNK14, PNK15), tindakan (PNK16), dan tarif (PNK17 dan PNK18). Hanya terdapat tiga data kosong pada PNK13. Sehingga, data kosong pada kolom tersebut akan digantikan dengan '9999' seperti diagnosis pada dataset kunjungan FKRTL. Informasi mengenai dataset kunjungan FKTP non-kapitasi terdapat pada lampiran 3.

#### ***4.1.3.1 Data dari Data Peserta BPJS***

PSTV01, PSTV02, dan PSTV15 pada dataset kunjungan FKTP non-kapitasi memiliki data yang sama dengan dataset peserta BPJS, tanpa ada perbedaan nilai pada PSTV15. Terdapat 1.803 peserta yang berbeda dalam dataset kunjungan FKTP non-kapitasi, dengan rata-rata kunjungan pada masing-masing peserta sebanyak 4,45 kali dan kunjungan terbanyak sebanyak 127 kali. Visualisasi ini terdapat pada gambar 13.



*Gambar 13 Distribusi Banyak Kunjungan FKTP Non-Kapitasi*

#### ***4.1.3.2 Deskripsi Kunjungan FKTP Non-Kapitasi***

Deskripsi kunjungan FKTP non-kapitasi terdiri dari ID kunjungan (PNK02), tanggal kunjungan, tindakan, dan pulang (PNK03, PNK04, dan PNK05), lokasi provinsi dan kabupaten/kota FKTP (PNK06 dan PNK07), kepemilikan,

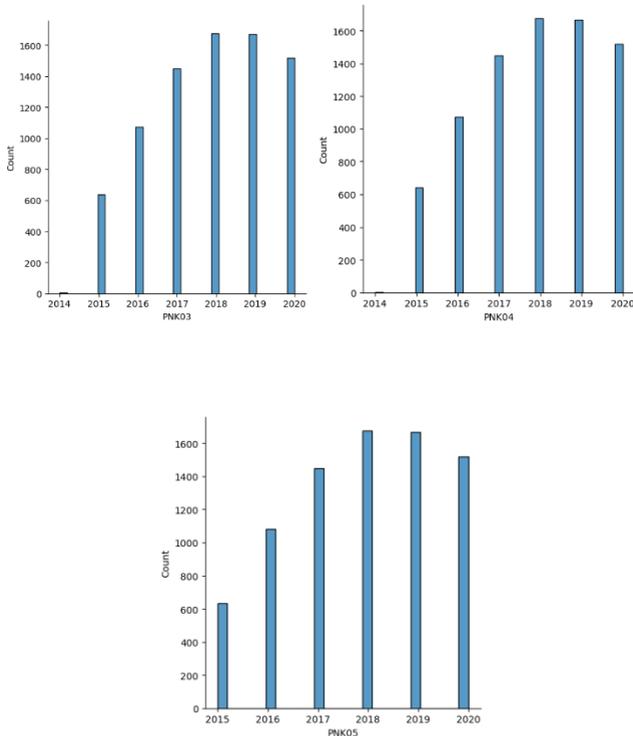
jenis, dan tipe FKTP (PNK08, PNK09, dan PNK10), jenis layanan FKTP (PNK11), dan pembagian peserta berdasarkan pekerjaan (PNK12).

Tidak seperti dua dataset sebelumnya, dataset kunjungan FKTP non-kapitasi tidak memiliki kolom yang memiliki nilai unik dan dimana ID primer-nya, ID kunjungan (PSTV02), memiliki data yang tidak semuanya berupa nilai unik. Sehingga, dataset kunjungan FKTP non-kapitasi memiliki data duplikat sebanyak 234 data. Terdapat 4.939 ID kunjungan yang unik dan ada 8.020 banyak kunjungan ke FKTP non-kapitasi. Terdapat 895 ID kunjungan yang berulang lebih dari satu kali dengan maksimal sembilan kali, dengan enam ID kunjungan berulang sebanyak sembilan kali.

Dataset kunjungan FKTP non-kapitasi memiliki tiga jenis tanggal, yaitu tanggal datang (PNK03), tanggal tindakan (PNK04), dan tanggal pulang (PNK05). Ketiga data tanggal tersebut dapat ditarik tiga data baru berupa lama tunggu (Antara PNK03 dan PNK04), lama tindakan (Antara PNK04 dan PNK05), dan lama kunjungan (Antara PNK03 dan PNK05). Sama seperti pada dataset kunjungan FKRTL, tiga data baru tersebut ditambahkan satu hari karena ketika tidak ada selisih antara ketiga data tanggal tersebut, maka lama waktu terhitung sebagai satu hari. Rata-rata lama tunggu, tindakan, dan kunjungan masing-masing adalah 1,15 hari (3,6 jam); 1,66 (15,84 jam); dan 1,8 hari (19,2 jam) dengan lama tunggu, tindakan, dan kunjungan terlama masing-masing selama 224, 223, dan 224 hari. Semua jenis data waktu lebih mendekati nilai 0.

Menurut data tanggal FKTP non-kapitasi, rentang tahun kunjungan FKTP non-kapitasi adalah 2014 sampai 2020, dengan enam kunjungan dan dua tindakan pada 2014. Pola

distribusi kunjungan FKTP non-kapitasi tiap tahun tidak jauh beda dengan pola distribusi kunjungan FKRTL, dengan tahun 2019 mengalami penurunan setelah terjadi peningkatan antara 2015 dan 2018. Visualisasi ini terdapat pada gambar 14.



*Gambar 14 Distribusi Banyak Kunjungan, Tindakan, dan Pulang dari FKTP Non-Kapitasi Tiap Tahun*

Distribusi daerah lokasi FKTP non-kapitasi (PNK06 dan PNK07) sama seperti pada dataset-dataset sebelumnya, dimana daerah pulau Jawa memiliki distribusi yang lebih besar, dengan mencakup 66,93 persen dari seluruh dataset

kunjungan FKTP non-kapitasi. Visualisasi ini terdapat pada tabel 41.

*Tabel 41 Daerah Asal FKTP Non-Kapitasi*

Nama Daerah	Jumlah	Persentase
Jawa	5.368	66,93
Sumatra	1.278	15,94
Sulawesi	701	8,74
Bali dan Nusa Tenggara	321	4,00
Kalimantan	214	2,67
Maluku dan Papua	138	1,72

Pada PNK08, FKTP non-kapitasi terbagi menjadi dua jenis kepemilikan, yaitu milik pemerintah dan swasta. FKTP non-kapitasi milik pemerintah terdiri dari pemerintah provinsi, pemerintah kabupaten/kota, TNI (Angkatan Darat dan Udara), Polri, dan BUMN. Visualisasi ini terdapat pada tabel 42.

*Tabel 42 Kepemilikan FKTP Non-Kapitasi*

Pemerintahan atau Swasta	Kepemilikan FKRTL	Banyak	Persentase
Swasta	Swasta	4.493	56,02
Pemerintahan	Kabupaten/Kota	3.342	41,67
	BUMN	131	1,63
	TNI AD	35	0,44
	Provinsi	11	0,14
	Polri	6	0,07
	TNI AU	2	0,02
	Total		3.527

Jenis FKTP non-kapitasi (PNK09) terbagi menjadi puskesmas, klinik pratama, laboratorium, dan lain-lain, yang

diantaranya adalah dokter umum dan jejaring. Kunjungan FKTP non-kapitasi dengan jenis puskesmas mencakup 41,3 persen dari seluruh data, diikuti oleh jenis laboratorium yang mencakup 30,1 persen. Visualisasi ini terdapat pada tabel 43.

*Tabel 43 Jenis FKTP Non-Kapitasi*

Jenis FKTP Non-Kapitasi	Banyak Kunjungan	Persentase
Puskesmas	3.312	41,30
Laboratorium	2.414	30,10
Klinik Pratama	1.545	19,26
Lain-lain	749	9,34

Tipe FKTP non-kapitasi (PNK10) terbagi menjadi rawat inap dan non-rawat inap (Termasuk klinik), laboratorium, dan lain-lain, yang terdiri dari dokter praktek perorangan, PPK lain-lain, IVA/Pap Smear, dan RS Kelas D Pratama. Visualisasi ini terdapat pada tabel 44.

*Tabel 44 Tipe FKTP Non-Kapitasi*

Tipe FKTP Non-Kapitasi	Banyak Kunjungan	Persentase
Non-Rawat Inap	2.634	32,84
Rawat Inap	2.414	30,10
Laboratorium	2.222	27,71
Lain-Lain	750	9,35

Kunjungan FKTP Non-Kapitasi (PNK11) memiliki tiga jenis layanan, RJTP (Rawat Jalan Tingkat Pertama), RITP (Rawat Inap Tingkat Pertama), dan Promotif. RJTP dan Promotif dikelompokkan sebagai non-rawat inap dikarenakan lama kunjung kedua jenis layanan tersebut seharusnya hanya dalam satu hari. Visualisasi ini terdapat pada tabel 45.

*Tabel 45 Jenis Layanan FKTP Non-Kapitasi*

Jenis Layanan	Banyak Kunjungan	Persentase
RJTP	3.278	40,87
RITP	1.605	20,01
Promotif	3.137	39,11

Jenis pembagian peserta pada dataset kunjungan FKTP non-kapitasi (PNK12) terbagi menjadi PBI (Baik dari APBN maupun APBD), PPU, bukan pekerja, dan PBPU. Kunjungan peserta kelompok PBI menjadi kunjungan yang paling banyak yang mencakup 45,65 persen. Visualisasi ini terdapat pada tabel 46.

*Tabel 46 Segmentasi Peserta pada Kunjungan FKTP Non-Kapitasi*

Segmentasi Peserta	Banyak Kunjungan	Persentase
PBI	3.661	45,65
PPU	2.036	25,39
Bukan Pekerja	1.197	14,93
PBPU	1.126	14,04

#### **4.1.3.3 Diagnosis**

Diagnosis pada dataset kunjungan FKTP non-kapitasi menggunakan kode ICD-10 edisi 2010. Menggunakan cara yang sama dengan dataset kunjungan FKRTL, terdapat empat bab yang mencakup lebih dari 10 persen dari seluruh dataset, yaitu bab XXI (Faktor status kesehatan dan kontak layanan kesehatan dengan kode Z ), IV (Penyakit endokrin, nutrisi, dan metabolisme dengan kode E), XV (Penanganan kelahiran dengan kode O), dan IX (Penyakit sistem peredaran dengan kode I). Visualisasi ini terdapat pada tabel 47.

*Tabel 47 Bab Diagnosis Kunjungan FKTP Non-Kaptiasi*

Bab Diagnosis	Banyak Kunjungan	Persentase
Z	2.543	31,71
E	2.512	31,32
O	1.434	17,88
I	1.152	14,36
Lain-Lain	379	4,73

#### **4.1.3.4 Tindakan**

PNK16 memiliki data berupa tindakan-tindakan dalam bentuk kata yang dapat dikelompokkan menjadi kelompok KB, Gula, Kolestrol, Ginjal, dan Kehamilan. Visualisasi ini terdapat pada tabel 48.

1. Kelompok tindakan KB adalah kelompok tindakan yang berhubungan dengan KB, diantaranya adalah suntik KB dan operasi IUD (Intrauterine Device).
2. Kelompok tindakan gula adalah kelompok tindakan yang berhubungan dengan gula darah. Kelompok ini terdiri dari gula darah puasa, gula darah post-prandial, gula darah sewaktu, dan HbA1c (Tes tingkat gula darah).
3. Kelompok tindakan kolestrol adalah kelompok tindakan yang memiliki hubungan terhadap kolestrol, diantaranya adalah trigliserida, total, LDL (Low-density lipoprotein), dan HDL (High-density lipoprotein).
4. Kelompok tindakan ginjal adalah kelompok tindakan yang berhubungan dengan ginjal. Kelompok ini terdiri dari Ureum, Kreatinin, dan Microalbuminaria
5. Kelompok tindakan kehamilan adalah kelompok tindakan yang berhubungan dengan persalinan,

kehamilan, dan hal-hal lainnya yang berhubungan dengan natalitas. Kelompok tindakan ini terdiri dari PNC (Postnatal care/pasca-lahir), ANC (Antenatal care/pre-lahir), komplikasi kebidanan dan neo-natal, paket persalinan per vaginam, tindakan pasca persalinan, dan penanganan keguguran.

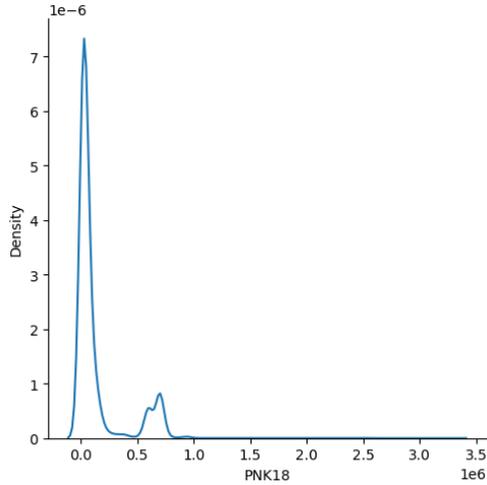
6. Kelompok tindakan teknis adalah kelompok tindakan yang tidak berhubungan kesehatan secara langsung. Kelompok tindakan ini terdiri dari evakuasi medis/ambulans dan rawat inap.

*Tabel 48 Jenis Tindakan pada Kunjungan FKTP Non-Kapitasi*

Kelompok Tindakan	Banyak Kunjungan	Persentase
Kehamilan	2.754	34,34
Gula	1.425	17,77
Kolestrol	1.256	15,66
KB	1.221	15,22
Ginjal	923	11,51
Teknis	441	5,50

#### **4.1.3.5 Tarif**

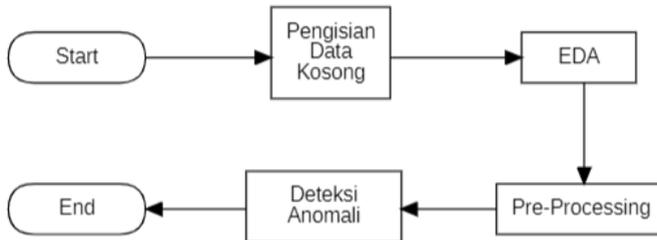
Tarif dalam dataset kunjungan FKTP non-kapitasi terbagi menjadi dua, yaitu tarif awal (PNK17) dan tarif verifikasi BPJS (PNK18). Tetapi, kedua kolom tersebut memiliki nilai yang sama, atau tidak terjadi pengurangan biaya dalam dataset kunjungan FKTP. Sehingga, hanya PNK18 yang digunakan. PNK18 memiliki rata-rata sebesar 133.854,70 dengan nilai terbesar adalah 3.300.000. Visualisasi ini terdapat pada gambar 49.



*Tabel 49 Distribusi Tarif Kunjungan FKTP Non-Kapitasi Setelah Verifikasi*

Tetapi, jika tarif kunjungan dibagi dengan lama kunjungan, maka rata-rata tarif sebesar 103.610,8 dengan tarif terbesar masih memiliki besar 3.300.000. Tarif yang akan digunakan adalah tarif harian kunjungan FKTP non-kapitasi.

## 4.2 Perancangan Sistem



*Gambar 15 Flowchart Perancangan Sistem*

Seperti flowchart pada gambar 15, langkah pertama dalam flowchart adalah mengambil dataset dari BPJS yang sudah disediakan, yaitu dataset peserta BPJS, kunjungan FKRTL, dan kunjungan FKTP non-kapitasi. Jika terdapat data kosong dalam dataset, maka data kosong tersebut diisi oleh sebuah nilai atau dihapuskan jika cukup banyak dan tergantung dari konteks nilai kosong tersebut. Masing-masing dataset akan dilakukan Exploratory Dataset Analysis, dimana dataset akan dilakukan analisis pada masing-masing atribut untuk mencari data kosong dan konteksnya, distribusi data, statistik data, konteks atribut dan isinya terhadap keseluruhan data, dan informasi dan konteks baru yang didapatkan dari dataset tersebut.

Langkah kedua adalah pre-processing. Dataset peserta BPJS digabungkan dengan dataset-dataset kunjungan berdasarkan tiap kunjunga, bukan tiap peserta. Masing-masing atribut dalam data akan dicari apakah terdapat kelas-kelas yang memiliki jumlah yang sedikit. Jika hanya terdapat dua kelas dengan perbedaan jumlah yang signifikan, maka atribut

tersebut akan dihapuskan. Jika terdapat banyak kelas dengan perbedaan jumlah yang signifikan dibandingkan dengan kelas mayoritas, maka kelas-kelas tersebut akan digabungkan. Penggabungan ini juga tergantung dari konteks masing-masing atribut. Atribut-atribut ordinal akan diubah menjadi numerik. Beberapa atribut akan membentuk atribut lainnya jika memungkinkan, seperti mencari lama kunjungan, tarif per hari, atau umur peserta.

Sebelum dihapus atribut-atribut yang tidak penting, seperti ID peserta, dan atribut-atribut yang masih tidak seimbang setelah penggabungan kelas, masing-masing dataset akan dicari anomali kategorikal dan campuran multivariat secara manual berdasarkan fakta yang ditemukan, seperti ketidaksesuaian data peserta dan ketidaksesuaian jenis layanan terhadap lama kunjungan. Setelah data duplikat dari dataset dihapus, maka dataset akan dicari anomali numerik univariat menggunakan boxplot.

Langkah berikutnya adalah mencari anomali numerik multivariat dan anomali multidimensional. Khusus untuk anomali multidimensional, atribut-atribut kategorikal akan dijadikan sebagai data dummy dengan One Hot Encoder. Untuk masing-masing anomali, kedua dataset lalu dilakukan standarisasi menggunakan Standard Scaler. Hasil dari standarisasi lalu diproses menggunakan PCA agar banyak dimensi dalam kedua data berkurang sampai nilai jumlah variansi dalam data mendekati 90%. Hasil dari PCA lalu diujikan terhadap berbagai algoritma deteksi anomali, yaitu DBSCAN, Isolation Forest, dan LOF. Tetapi, dikarenakan Isolation Forest dan LOF memiliki parameter perkiraan banyak anomali, maka ditambahkan algoritma deteksi anomali berupa data campuran hasil dari Isolation Forest dan LOF berupa AND

dan OR agar dapat mengetahui data mana saja yang menurut kedua algoritma setuju merupakan anomali. Hasil uji dari masing-masing algoritma diukur dengan menggunakan berbagai metrik pengukuran, yang paling utama adalah algoritma ROC-AUC. Metrik-metrik lainnya adalah Confusion Matriks dan nilai-nilai turunannya berupa akurasi, presisi, recall, f1, spesifisitas, dan kappa. Algoritma dengan hasil uji terbaik akan menjadi algoritma utama dalam menentukan data anomali dalam kedua dataset kunjungan. Khusus dataset kunjungan FKRTL, dataset tersebut akan disampel sebesar 1% dikarenakan lama eksekusi program dan performa hardware yang digunakan.

*[Halaman ini sengaja tidak berisi kecuali kalimat ini]*

## **BAB V**

### **IMPLEMENTASI SISTEM**

#### **5.1 Pengelompokan Kelas-Kelas Minoritas**

Pseudocode 1 menjelaskan cara-cara mengelompokkan kelas-kelas minoritas di dalam sebuah atribut dalam sebuah dataset yang memiliki input berupa dataset (*df*), nama atribut dalam dataset (*column\_name*), kelompok dari kelompok kelas-kelas minoritas (Contohnya adalah kelompok pulau Jawa, pulau Sumatera, dan sebagainya) (*group*), kelompok kelas pengganti (Contohnya adalah Jawa, Sumatera, dan sebagainya) (*label*), dan nilai boolean yang menentukan apakah nilai dalam atribut dataset diubah atau hanya dilihat hasilnya saja (*replace*).

Baris 2 menyalin atribut dalam dataset. Baris 3 memulai proses pengulangan dengan panjang sepanjang banyak kelompok kelas-kelas minoritas. Dalam masing-masing pengulangan, terdapat pengulangan pada baris 5 yang mengulangi isi dari kelompok dari kelompok kelas-kelas minoritas. Kelas-kelas minoritas dalam suatu kelompok digantikan dengan nilai label dalam kelompok label pada posisi nilai yang sama pada baris 7. Baris 10 menampilkan besar masing-masing kelas pada atribut salinan setelah kelas minoritas digantikan, atau dikelompokkan. Pada baris 11, nilai *replace* dicek apakah bernilai *True* atau *False*. Pada baris 12, jika *True*, maka atribut dataset asli digantikan dengan atribut dataset salinan. Dataset dikembalikan pada baris 14.

### Fungsi pengelompokan kelas-kelas minoritas

---

```
1 Replacing_atr(df, column_name, group, label, replace) :  
  
   Input : Dataset (df), nama atribut (column_name), array dua  
   dimensi berisi array kelompok kelas (group), array kelas pengganti  
   (label), nilai boolean (replace)  
   Output : Dataset (df)  
  
2 temp <- df[column_name]  
3 for each number in lenght of group  
4 start  
5   for each member in group[number]  
6   start  
7     temp -> replace(member, label[number])  
8   end  
9 end  
10 shows value counts of temp  
11 if replace is True then  
12   df[column_name] <- temp  
13 end if  
14 return df
```

*Pseudocode 1 Fungsi Pengelompokan Kelas-Kelas Minoritas*

## 5.2 Boxplot Modifikasi

Pseudocode 2 menjelaskan fungsi boxplot yang memiliki input berupa satu array berisikan data numerik dan output berupa nilai batas bawah dan batas atas. Baris 2 menentukan nilai rentang inter-kuartil dengan mencari selisih antara kuartil akhir dan kuartil awal array data. Baris 3 mencari selisih antara kuartil tengah dan kuartil awal array data. Jika nilai selisih tersebut adalah 0 pada baris 5, maka nilai selisih tersebut dianggap sebagai 1 pada baris 6. Baris 4 mencari selisih antara kuartil akhir dan kuartil tengah array data. Jika nilai selisih tersebut adalah 0 pada baris 8, maka nilai selisih tersebut diantap 1 pada baris 9. Batas bawah dan batas atas dicari pada baris 11 dan 12. Batas bawah dan batas atas dikembalikan pada baris 13.

## Fungsi Boxplot modifikasi

---

```
1 | Boxplot(data) :  
   |  
   | Input : Satu array numerik (data)  
   | Output : Batas bawah (low_thres), dan batas atas (high_thres)  
  
2 | IQR <- Q3(data) – Q1(data)  
3 | Q2Q1 <- Q2(data) – Q1(data)  
4 | Q3Q2 <- Q3(data) – Q2(data)  
5 | If Q2Q1 is 0 then  
6 |   Q2Q1 <- 1  
7 | End if  
8 | If Q3Q2 is 0 then  
9 |   Q3Q2 <- 1  
10 | End if  
11 | Low_Thres <- Q1(data) – (1.5 * IQR * ((Q2(data) – Q1(data)) /  
   | Q3Q2)  
12 | High_Thres <- Q3(data) + (1.5 * IQR * ((Q3(data) – Q2(data)) /  
   | Q2Q1)  
13 | Return low_thres, high_thres
```

*Pseudocode 2 Fungsi Boxplot Modifikasi*

### 5.3 PCA Dataset

Pseudocode 3 menjelaskan tentang pengurangan dimensi dataset (data) menjadi nilai dimensi yang ditentukan (dimension) sehingga nilai total variansi menjadi sekitar 90 persen. Pada baris 2, fungsi PCA dengan parameter dimensi dimasukkan dan mengubah nilai dataset. Pada baris 3, nilai variansi PCA akan ditampilkan. Data yang sudah dikurangi dimensinya akan dikembalikan pada baris 4.

#### Fungsi PCA dataset

---

```
1 | PCA_data(dimension, data) :  
   |  
   | Input : Target dimensi (dimension), dataset (data)  
   | Output : Data setelah reduksi dimensi (new_data)  
2 | new_data <- PCA(dimension) <- fit_transform(data)  
3 | Show sum of PCA variance  
4 | Return new_data
```

*Pseudocode 3 Fungsi PCA Terhadap Dataset*

## 5.4 DBSCAN

Pseudocode 4 menjelaskan proses deteksi anomali pada sebuah data dengan input berupa data latihan (train) untuk memprediksi data tes (test), nilai dimensi dari PCA (dimension) untuk mengukur titik data maksimal dalam sebuah kelompok DBSCAN, dan skala epsilon (eps\_scale) untuk mengukur besar jarak antar titik data dalam DBSCAN.

Baris 2 mencari titik terdekat masing-masing titik data selain titik data yang sama menggunakan Nearest Neighbor dengan memasukan data training dan mengetes data tes. Baris 3 menghasilkan hasil dari fungsi DBSCAN dengan input nilai epsilon berupa nilai terbesar dari jarak terdekat suatu titik data terhadap titik lainnya dari Nearest Neighbor dikalikan dengan skala epsilon dan banyak titik data dalam suatu kelompok yang memiliki besar dimensi data ditambah dua. DBSCAN tersebut dimasukan data tes sebelum dikembalikan sebagai hasil fungsi pada baris 4.

Fungsi proses DBSCAN	
1	DBSCAN_data(train, test, dimension, eps_scale):  Input : data training (train), data tes (test), besar dimensi setelah PCA (dimension), besar jarak epsilon (eps_scale) Output : hasil DBSCAN data (DB_result)
2	distance, index <- NearestNeighbors() <- fit(train) <- kneighbors(test)
3	DB_result <- DBSCAN(max(sort(distance[[1]])) * eps_scale, dimension + 2) <- fit(test)
4	return DB_result

*Pseudocode 4 Fungsi Proses DBSCAN*

## 5.5 Isolation Forest

Pseudocode 5 menjelaskan proses Isolation Forest dengan input berupa data pelatihan (train), data tes (test), dan persentase perkiraan banyak anomali dalam sebuah data (contamination). Baris 2 mengembalikan nilai Isolation Forest dengan parameter contamination yang memasukan nilai data train dan memprediksi nilai data tes.

Fungsi proses Isolation Forest	
1	Isofor(train, test, contamination):  Input : data training (train), data tes (test), prediksi persentase data anomali (contamination) Output : hasil prediksi dengan Isolation Forest
2	Return IsolationForest(contamination) <- fit(train) <- predict(test)

*Pseudocode 5 Fungsi Proses Isolation Forest*

## 5.6 LOF

Pseudocode 6 menjelaskan proses LOF yang memiliki input data pelatihan (train), data tes (test), persentase perkiraan banyak anomali dalam sebuah data (contamination), dan nilai boolean yang menentukan apakah data yang akan dites merupakan data yang baru atau tidak (novelty). Pada baris 2,

fungsi LOF dengan parameter containment dan novelty diinisialisasi. Pada baris 3, jika nilai novelty adalah False, maka fungsi mengembalikan nilai dari hasil memasukan data train pada LOF untuk memprediksi data train pada baris 4. Jika nilai novelty adalah True, maka fungsi mengembalikan nilai dari hasil memasukan data train pada LOF untuk memprediksi data tes pada baris 6.

Fungsi proses LOF	
1	LOF_data(train, test, contamination, novelty):  Input : data training (train), data tes (test), prediksi persentase data anomali (contamination), nilai bool (novelty) Output : hasil prediksi dengan LOF
2	LOF_start <- LOF(containment, novelty)
3	If novelty is False then
4	Return LOF_start <- fit(train) <- predict(train)
5	else
6	Return LOF_start <- fit(train) <- predict(test)

*Pseudocode 6 Fungsi Proses LOF*

### 5.1.5 Train-Test Split

Pseudocode 7 menjelaskan fungsi pembagian dataset menjadi dataset pelatihan dan tes dengan input berupa dataset (data), hasil deteksi anomali dari dataset (result), dan besar persentase data tes (test\_size). Baris 2 memasukan input data, result, dan test\_size pada fungsi train\_test\_split yang mana hasilnya berupa dataset latihan (X\_Train dan y\_train) dan dataset tes (X\_test dan y\_test). Nilai-nilai tersebut dikembalikan pada baris 3.

Fungsi pembagian data dengan train-test split	
1	split(data, result, test_size) :
	Input : Dataset (data), hasil deteksi anomali (result), besar data tes (test_size)
	Output : dataset training (X_train), dataset tes (X_test), hasil deteksi anomali train (y_train), hasil deteksi anomali tes (y_test)
2	X_train, X_test, y_train, y_test = train_test_split(data, result, test_size)
3	Return X_train, X_test, y_train, y_test

*Pseudocode 7 Fungsi Pembagian Data Menjadi Data Train dan Data Test*

#### 5.1.4 Pengukuran Metrik

Pseudocode 8 menjelaskan tampilan nilai ROC-AUC dan metrik-metrik evaluasi model deteksi anomali dengan input nilai hasil tes (y\_test) dan nilai hasil yang seharusnya (y\_pred). Baris 2 menampilkan nilai dan grafik ROC-AUC antara y\_test dan y\_pred. Baris 3 menghasilkan nilai TP, FP, FN, dan TN dari confusion matrix antara y\_test dan y\_pred yang diitampilkan pada baris 4. Keempat nilai tersebut digunakan untuk mendapatkan nilai metrik-metrik penilaian model deteksi anomali berupa akurasi, presisi, recall, f1, spesifitas positif dan negatif, dan nilai Kappa pada baris 5 sampai 12. Baris 13 menampilkan seluruh nilai metrik-metrik tersebut.

### Fungsi pengukuran metrik

---

```
1 | metrics(y_test, y_pred) :  
   |  
   | Input : Hasil dari tes (y_test), hasil prediksi (y_pred)  
   | Output : Tidak ada  
  
2 | Show ROC-AUC(y_test, y_pred)  
3 | TP, FP, FN, TN <- Confusion_matrix(y_test, y_pred)  
4 | Show TP, FP, FN, TN  
5 | Accuracy <- (TP + TN) / length of y_pred  
6 | Precision <- TP / (TP + FP)  
7 | Recall <- TP / (TP + FN)  
8 | F1 <- (2 * Precision * Recall) / (Precision + Recall)  
9 | Specificity_pos <- TN / (FP + TN)  
10 | Specificity_neg <- 1 - Specificity_pos  
11 | R <- ((TP + FN) * (TP + FP) + (FP + TN) * (FN + TN)) / length of  
   | y_pred  
12 | Kappa <- (Accuracy - R) / (1 - R)  
13 | Shows Accuracy, Precision, Recall, F1, Specificity_pos,  
   | Specificity_neg, and Kappa
```

*Pseudocode 8 Fungsi Penilaian Model Deteksi Anomali*

## **BAB VI**

### **PENGUJIAN DAN EVALUASI**

#### **6.1 Dataset Peserta**

Dataset pertama yang akan dilakukan deteksi anomali ini adalah dataset peserta BPJS sebelum diberikan kepada dataset-dataset kunjungan. Dalam dataset peserta, terdapat dua jenis tahun meninggal dalam PSTV18 jika seorang peserta meninggal, yaitu 2019 dan 2020. Alasan yang memungkinkan sebab tidak ada tahun lain selain 2019 dan 2020 adalah dataset peserta BPJS diambil pada tahun 2020 dan sebagian peserta pada saat tersebut sudah meninggal ketika dilakukan pengambilan sampel. Tetapi, diantara peserta-peserta BPJS dalam dataset tersebut, terdapat 5.947 peserta yang memiliki tahun meninggal pada PSTV18 tidak memiliki status peserta pada PSTV17 tertulis meninggal; 164 peserta tidak aktif dan sisanya masih aktif pada waktu pengisian dataset.

Anomali lainnya dalam dataset peserta BPJS adalah status pribadi peserta, terutama hal yang mengenai keluarga (PSTV02), jenis hubungan keluarga (PSTV04), jenis kelamin (PSTV05), dan status kawin (PSTV06). Menurut PSTV02, terdapat 567 keluarga, dimana seorang peserta mengajak anggota keluarga lainnya untuk mendaftar menjadi peserta BPJS. Tetapi, terdapat 57.797 peserta BPJS yang tidak memiliki anggota keluarga lainnya dalam dataset peserta BPJS tetapi tidak memiliki status keluarga peserta pada PSTV04. Ada kemungkinan bahwa anggota keluarga para peserta tersebut yang memiliki status keluarga peserta yang mengajak mereka untuk mendaftar ke BPJS berada pada dataset peserta BPJS lainnya pada tahun yang berbeda. Diantara 567 keluarga, terdapat satu keluarga yang mengajak dua orang anggota keluarganya, yaitu anak dan suaminya, sedangkan sisanya hanya mengajak satu orang. Tetapi, diantara 566 keluarga yang

mengajak satu anggota keluarganya, terdapat 32 keluarga yang tidak memiliki anggota keluarga peserta, seperti pada tabel 10. Dalam PSTV04 dan PSTV05, terdapat 15 peserta yang memiliki status keluarga dan jenis kelamin yang tidak sesuai. 9 peserta yang memiliki status keluarga istri memiliki jenis kelamin laki-laki dan 6 peserta yang memiliki status keluarga suami memiliki jenis kelamin perempuan. Mengenai status keluarga dan status kawin, terdapat 1.423 peserta yang memiliki status keluarga pada PSTV04 yang tertulis suami atau istri memiliki status perkawinan pada PSTV06 bukan kawin (Belum kawin atau cerai) dan terdapat 17 keluarga yang mana anggota keluarganya merupakan pasangan suami dan istri dimana salah satu atau kedua peserta dari anggota keluarga peserta memiliki data status perkawinan yang tidak sesuai, dimana salah satu anggota keluarga mengaku bahwa mereka sudah menikah sementara anggota keluarga lainnya, yang seharusnya sudah menikah dengan anggota keluarga tersebut, belum menikah atau sudah cerai dan sebaliknya atau kedua anggota keluarga tersebut, yang salah satunya memiliki status hubungan suami atau istri, belum kawin atau sudah cerai.

Selain itu, setelah dilakukan deteksi anomali campuran multivariat antara umur peserta pada saat penulisan dataset dan status perkawinan peserta, terdapat 13 peserta yang memiliki status kawin tetapi memiliki umur pada saat penulisan dataset berumur di bawah 18 tahun, yang merupakan batas umur seseorang yang dapat diakui sebagai sudah dewasa. Diantara 13 peserta tersebut, 7 peserta merupakan anak seorang peserta, 5 peserta merupakan anggota keluarga lainnya dari peserta, dan satu orang merupakan seorang peserta. 13 peserta tersebut memiliki rentang umur antara 8 sampai 17 tahun.

## 6.2 Dataset Kunjungan FKRTL

Anomali kategorikal dan campuran multivariat yang ditemukan berfokus pada jenis layanan (Rawat Inap dan Jalan) dan lama kunjungan. Terdapat lima atribut dan kelas yang memiliki atribut berhubungan dengan jenis layanan dan lama kunjungan, yaitu lama kunjungan, jenis layanan (FKL10), kelas Klinik Non-Rawat Inap pada FKL09, kelas Q pada FKL20, FKL21, dan FKL23. Visualisasi ini terdapat pada tabel 50.

*Tabel 50 Anomali Kategorikal dan Campuran pada Dataset Kunjungan FKRTL Berdasarkan Jenis Layanan dan Lama Kunjungan*

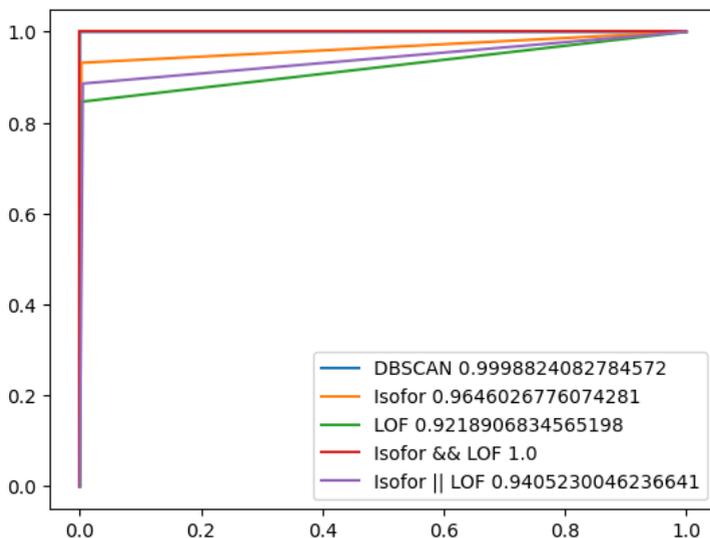
	RITL (FKL10)	Lama Kunjung > 1
RJTL (FKL10)		61
Q (FKL20)	29	50
Rawat Jalan (FKL21)	30	61
0 (FKL23)	30	61

Mengenai anomali numerik univariat, setelah dilakukan penghapusan atribut imbalance dan duplikasi sebanyak 2.969.440 data sehingga banyak data FKRTL menjadi 2.126.596 data, terdapat 452.020 data (21,26%) dimana tarif harian setelah verifikasi BPJS (FKL48) berada di luar rentang batas bawah dan batas atas Rp 170.998,35 dan Rp 2.100.422,3. Diantara data yang menggunakan layanan rawat inap, terdapat 4.254 data (0,2 %) yang berada di luar rentang batas bawah dan batas atas 0,75 dan 15 hari. Terdapat 60.005 data (2,82 %) yang mana jumlah prosedur harian lebih dari batas atas 2.5 prosedur per hari. Terdapat 11.496 data (0,54 %)

dimana peserta memiliki bobot di atas batas atas 14,86. Terdapat 43.095 data (2,03 %) dimana umur masuk peserta di luar rentang batas bawah dan batas atas 35,29 dan 87,75 tahun.

*Tabel 51 Hasil dari Tes Anomali Numerik Multivariat Dataset Kunjungan FKRTL*

Nama Algoritma	ROC-AUC	Akurasi	Presisi	Recall	F1	Spesifisitas Positif	Kap pa
DBSCAN	0,9999	0,9998	0,9998	1	0,9999	0,6667	0,8
IsoFor	0,9646	0,9967	0,9974	0,9993	0,9983	0,7885	0,8525
LOF	0,9219	0,9962	0,9976	0,9986	0,9981	0,7674	0,803
IsoFor && LOF	1	1	1	1	1	1	1
IsoFor    LOF	0,9405	0,9929	0,995	0,9978	0,9964	0,7692	0,8199



*Gambar 16 Grafik ROC-AUC dari Masing-Masing Algoritma Deteksi Anomali Numerik Multivariat Dataset Kunjungan FKRTL*

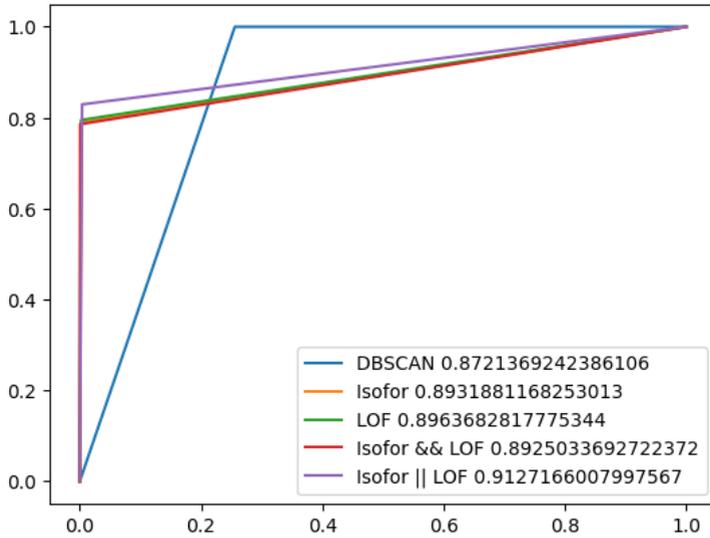
Anomali numerik multivariat, menurut algoritma Isolation Forest-LOF AND yang memiliki nilai terbaik, memiliki beberapa jenis deskripsi anomali. Dari data sample sebesar 21.266 data, terdapat 15 data (0,07 %) yang merupakan data anomali. Semua data anomali numerik multivariat tidak mengunjungi FKTP Kapitasi sebelumnya. :

- Banyak prosedur antara 10 dan 11 prosedur, tarif harian antara 140.000 dan 160.000, banyak prosedur kurang dari 11, rawat jalan, dan kelompok daerah 3 (38)
- Lama kunjung lebih dari 33 hari dan umur peserta lebih dari 75 tahun (12)
- Bobot dibawah 5, umur peserta lebih dari 67 tahun, lama kunjung lebih dari 4, tarif harian kurang dari

21.000.000, kelompok daerah 3, dan rawat inap tingkat parah (187)

- Umur peserta kurang dari 20, rawat jalan, tarif harian dibawah 10.000.000, kelompok daerah 1 (3.822)
- Banyak prosedur harian lebih dari 7, rawat inap ringan dengan lama kunjung satu hari saja, tarif harian kurang dari 30.000.000, bobot peserta di atas 14, kelompok daerah 1 (3)
- Umur peserta di bawah 50, rawat inap ringan dengan lama kunjung selama dua hari, kelompok daerah 1, bobot peserta antara 12,5 dan 13,2, dan banyak prosedur harian kurang dari 0.5 (256)
- Tarif harian kurang dari 30.000.000, rawat inap ringan dengan lama kunjung selama dua hari, kelompok daerah 1, bobot peserta lebih dari 11, dan umur peserta lebih dari 58 (5.220)
- Tarif harian lebih dari 40.000.000, bobot antara 12 dan 13, dan umur peserta lebih dari 58, rawat inap ringan (419)

Visualisasi evaluasi tes anomali numerik multivariat dataset kunjungan FKRTL terdapat pada tabel 51 dan grafik ROC-AUC anomali numerik multivariat dataset kunjungan FKRTL divisualisasikan pada gambar 16.



*Gambar 17 Grafik ROC-AUC dari Masing-Masing Algoritma Deteksi Anomali Multidimensional pada Dataset Kunjungan FKRTL*

*Tabel 52 Hasil dari Tes Anomali Multidimensional Dataset  
Kunjungan FKRTL*

Nama Algoritma	RO C-AUC	Akurasi	Presisi	Recall	F1	Spesifitas Positif	Kap pa
DBSCAN	0,8721	0,7612	0,7443	1	0,8534	0,2167	0,2777
IsoFor	0,8932	0,9941	0,9969	0,9971	0,997	0,7759	0,7796
LOF	0,8964	0,996	0,9979	0,9981	0,998	0,775	0,7828
IsoFor && LOF	0,8925	0,9986	0,9993	0,9993	0,9993	0,7857	0,785
IsoFor    LOF	0,9127	0,9929	0,9962	0,9966	0,9964	0,8095	0,8157

Menurut IsoFor OR LOF, data anomali multidimensional pada dataset kunjungan FKRTL memiliki banyak 380 data (1.79) dan terdiri dari beberapa bagian.

- Tarif harian lebih dari 40 juta Rupiah, kelompok daerah 1, banyak prosedur kurang dari 1, faskes FKRTL terletak di Jawa bagian barat, rawat inap, poli selain INT, diagnosis masuk dalam kelompok EZI dan diagnosis primer selain Z, diagnosis INA-CBG kelompok selain Q, dirujuk dari faskes dari Jawa bagian barat, peserta merupakan peserta laki-laki sudah menikah dengan bobot kurang dari 12, tinggal di dan mendaftar di faskes pemerintah di provinsi Jawa bagian Barat. Kunjungan ini tidak mengunjungi FKTP kapitasi sebelumnya. (124)

- Tarif harian antara 10 juta dan 40 juta rupiah, lama kunjung kurang dari 10 hari, rawat inap, poli lain-lain, diagnosis primer selain Z, dan diagnosis INA-CBG selain Q (15.116)
- Tarif harian antara 1 juta dan 10 juta, lama kunjung di atas 10 hari, faskes milik pemerintahan, rawat inap parah, poli lain-lain, diagnosis primer selain Z, diagnosis INA-CBG selain Q, kelompok daerah 1, tidak mengunjungi FKTP kapitasi sebelumnya (665)
- Tarif antara 1 juta dan lebih dari 10 juta, lama kunjungan kurang dari 10, rawat inap, poli lain-lain, diagnosis primer lain-lain, dan INA-CBG selain Q (197.933)
- Tarif kurang dari 1 juta dan lama kunjungan di atas 10 (45)
- Tarif kurang dari 1 juta, lama kunjungan kurang dari 10 hari, dan rawat inap (141)
- Tarif kurang dari 1 juta, lama kunjungan selama 1 hari, rawat jalan, mengunjungi FKTP kapitasi sebelumnya, dan banyak prosedur antara 10 dan 1 (170.841)
- Data lain yang cenderung lebih umum. Data yang tergolong anomali disimpan ke dalam file .csv

Visualisasi evaluasi tes anomali multivariat dataset kunjungan FKRTL terdapat pada tabel 52 dan grafik ROC-AUC tes tersebut divisualisasikan pada gambar 17.

### **6.3 Dataset Kunjungan FKTP Non-Kapitasi**

Anomali yang pertama ditemukan dalam dataset kunjungan FKTP non-kapitasi adalah anomali berupa hubungan tipe faskes kunjungan (PNK10) terhadap lama

kunjung. Tipe faskes kunjungan terbagi menjadi rawat inap, non-rawat inap, laboratorium, dan lain-lain. Kunjungan dengan tipe faskes non-rawat inap, yang mana waktu kunjungan seharusnya tidak perlu lebih dari satu hari, tetapi memiliki lama kunjungan lebih dari satu hari adalah sebanyak 236 kunjungan, dengan lama kunjungan terlama selama 24 hari.

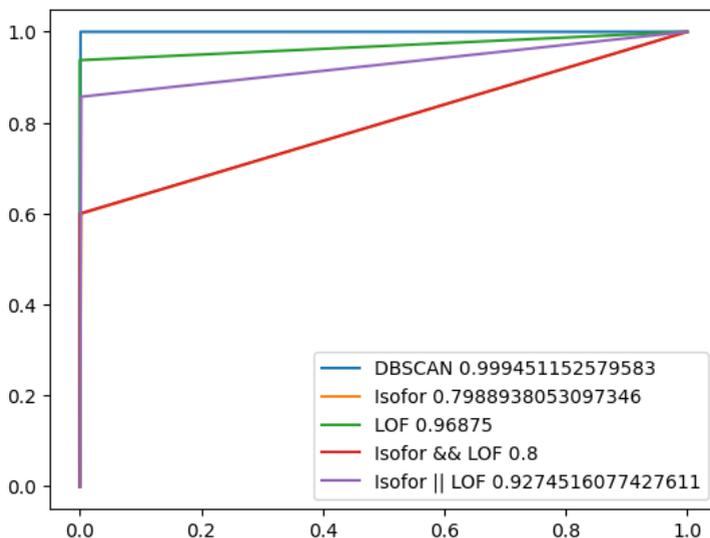
Setelah atribut-atribut imbalance dan data-data duplikat sebanyak 3.454 data dihapuskan sehingga besar data dataset kunjungan FKTP non-kapitasi sebanyak 4.566 data, anomali berikutnya berupa numerik. Menggunakan boxplot modifikasi, anomali numerik univariat ditemukan pada masing-masing atribut numerik. Pada atribut tarif harian (PNK18 dibagi lama kunjungan), terdapat 17 data (0,37 %) yang mana tarif hariannya diatas batas atas data sebesar Rp 755.275,00. Mengenai lama kunjung, tunggu, dan tindakan, data yang akan dicari anomalnya berupa data kunjungan yang memiliki jenis layanan rawat inap karena lama layanan yang digunakan dalam rawat inap tidak terbatas pada 1 hari saja, tidak seperti rawat jalan. Terdapat 220 kunjungan (4,82% dari seluruh data) dimana lama kunjungnya lebih dari batas atas data sebesar 2 hari. Terdapat 150 kunjungan (3,29%) yang memiliki lama tunggu bukan satu hari, yang merupakan batas bawah dan atas lama tunggu. Terdapat 40 kunjungan (0,88%) yang mana lama tindakannya tidak berada pada rentang antara 1 sampai 3,5 hari. Hanya ada satu kunjungan dimana seorang peserta memiliki bobot diatas 15,03. Terdapat 161 data (3,53%) dimana umur peserta ketika mengunjungi FKTP non-kapitasi tidak berada pada rentang antara 41,14 dan 91,2 tahun.

Anomali berikutnya adalah anomali numerik multivariat dengan menggunakan atribut bobot peserta, tarif harian, lama kunjung, tunggu, dan tindakan, dan umur masuk

peserta. Anomali tersebut dicari menggunakan PCA untuk mengurangi dimensi dan algoritma-algoritma deteksi anomali berupa DBSCAN, Isolation Forest, dan LOF. Hasil dari Isolation Forest dan LOF dapat digabungkan secara AND atau OR. Hasil dari masing-masing algoritma diukur dengan membagi data menjadi data training dan test dengan rasio 4:1. Hasil dari tes algoritma deteksi anomali diukur menggunakan nilai ROC-AUC, lalu diikuti oleh metrik confusion matrix, akurasi, presisi, recall, f1, Spesifisitas, dan Kappa. Visualisasi evaluasi tes anomali numerik multivariat dataset kunjungan FKTP terdapat pada tabel 53 dan grafik ROC-AUC data tes anomali numerik multivariat dataset kunjungan FKTP divisualisasikan pada gambar 18.

*Tabel 53 Hasil dari Tes Anomali Numerik Multivariat Dataset Kunjungan FKTP Non-Kapitasi*

Nama Algoritma	ROC-AUC	Akurasi	Presisi	Recall	F1	Spesifisitas Positif	Kappa
DBSCAN	0,9995	0,9989	0,9989	1	0,9995	0,75	0,8566
IsoFor	0,7989	0,9934	0,9979	0,9956	0,9967	0,75	0,6634
LOF	0,9688	0,9989	1	0,9989	0,9994	1	0,9672
IsoFor && LOF	0,8	0,9978	1	0,9978	0,9989	1	0,749
IsoFor    LOF	0,9275	0,9945	0,9978	0,9966	0,9972	0,9	0,8753



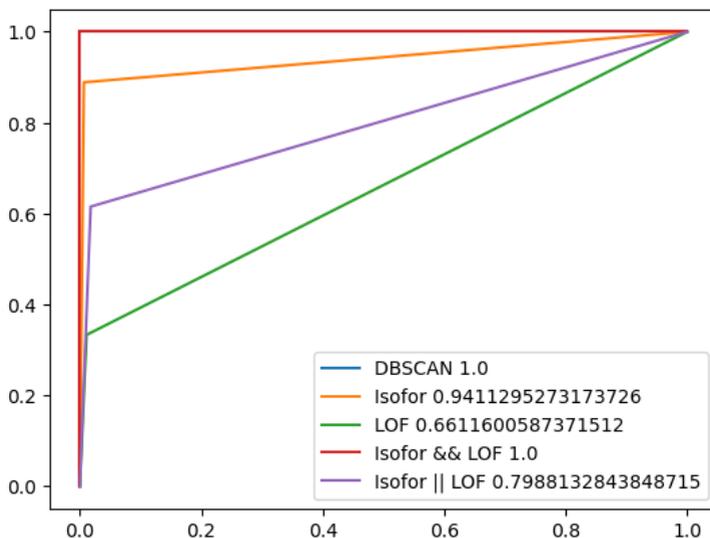
*Gambar 18 Grafik ROC-AUC dari Masing-Masing Algoritma Deteksi Anomali Numerik Multivariat pada Dataset Kunjungan FKTP Non-Kapitasi*

Algoritma dengan nilai ROC-AUC terbaik adalah DBSCAN. Tetapi, algoritma dengan metrik terbaik secara mayoritas adalah LOF. Menurut DBSCAN dan LOF, data yang merupakan anomali merupakan data yang memiliki lama kunjung lebih dari 200 hari dan data yang memiliki tarif harian lebih dari Rp 3.300.000,00. Terdapat 25 data yang memiliki nilai tersebut, yang mencakup 0,55 persen dari seluruh data.

Anomali terakhir adalah anomali multidimensional, dimana atribut non-numerik menjadi data dummy dan menjadi numerik dan digabungkan dengan atribut numerik. Visualisasi evaluasi tes anomali multidimensional terdapat pada tabel 54 dan grafik ROC-AUC tes ini divisualisasikan pada gambar 19.

*Tabel 54 Hasil dari Tes Anomali Multidimensional Dataset  
Kunjungan FKTP Non-Kapitasi*

Nama Algoritma	RO C-AUC	Akurasi	Persisi	Recall	F1	Spesifisitas Positif	Kap pa
DBSCAN	1	1	1	1	1	1	1
IsoFor	0,9411	0,9923	0,9934	0,9989	0,9961	0,5714	0,692
LOF	0,6612	0,9847	0,989	0,9956	0,9923	0,1667	0,2153
IsoFor && LOF	1	1	1	1	1	1	1
IsoFor    LOF	0,7988	0,977	0,9822	0,9944	0,9883	0,3333	0,4218



*Gambar 19 Grafik ROC-AUC dari Masing-Masing Algoritma Deteksi Anomali Multidimensional pada Dataset Kunjungan FKTP Non-Kapitasi*

Dua algoritma terbaik untuk mencari anomali multidimensional adalah DBSCAN dan IsoFor && LOF. Kedua algoritma menemukan anomali multidimensional yang mana lama lama kunjung suatu kunjungan di atas 200 hari, seperti anomali numerik multivariat, yang terdiri dari 24 data. Data yang menurut IsoFor && LOF saja merupakan anomali adalah kunjungan dimana seorang peserta laki-laki yang sudah menikah dengan umur antara 60 sampai 61 tahun dimana peserta tersebut berasal dari daerah Jawa bagian tengah dan timur, tergolong PPU menurut faskes tempat mendaftar, yang merupakan bukan puskesmas dan milik swasta di daerah yang sama, tetapi PBI di faskes FKTP non-kapitasi milik swasta lain-lain di daerah Jawa bagian Barat, mengunjungi fakseks FKTP non-kapitasi dengan layanan rawat jalan, mendapatkan

diagnosis bab E, mendapatkan tindakan berhubungan dengan gula darah, dan mendapatkan tarif harian sebesar Rp 20.000,00. Terdapat 8 data yang memiliki data tersebut. Sehingga, banyak data yang merupakan anomali dalam dataset kunjungan FKTP non-kapitasi sebanyak 33 data, yang mencakup 0,72 persen dari seluruh data.

*[Halaman ini sengaja tidak berisi kecuali kalimat ini]*

## **BAB VII**

### **KESIMPULAN DAN SARAN**

#### **7.1 Kesimpulan**

Sebelum melakukan deteksi anomali, dataset-dataset BPJS harus dipersiapkan terlebih dahulu. Dataset-dataset kunjungan digabungkan dengan dataset peserta BPJS berdasarkan ID peserta sehingga masing-masing data dalam dataset-dataset kunjungan memiliki informasi mengenai peserta. Jenis atribut yang harus ada berupa atribut numerik dan kategorikal. Atribut yang tidak berjenis numerik dan kategorikal, seperti tanggal, diubah menjadi atribut numerik atau kategorikal. Data-data kosong dalam sebuah atribut diubah dan diisi dengan kelas yang menjelaskan bahwa data merupakan data kosong, tergantung dari masing-masing atribut. Kedua dataset dilakukan EDA untuk menemukan informasi mengenai atribut-atribut dan data-data dalam kedua dataset. Atribut numerik yang dapat dibagi dengan lama kunjungan dibagikan dengan lama kunjungan, seperti tarif kunjungan dan banyak prosedur.

Dalam kedua dataset, terdapat tiga jenis dalam anomali kategorikal dan campuran multivariat yang dapat ditemukan, yaitu informasi peserta BPJS (ID keluarga, jenis kelamin, hubungan keluarga, dan jenis perkawinan), hubungan antara tahun meninggal peserta, status peserta, dan status peserta ketika pulang dari FKRTL, dan hubungan antara lama kunjung faskes dengan jenis layanan, terutama terhadap jenis layanan rawat jalan yang hanya memiliki lama kunjung satu hari. Deteksi anomali ini dilakukan secara manual dengan bantuan Python. Dalam masing-masing atribut kategorikal, terdapat kelas-kelas minoritas, atau anomali kategorikal univariat, yang membuat atribut tersebut tidak seimbang dibandingkan dengan kelas mayoritas. Sehingga, kelas-kelas minoritas tersebut dalam sebuah atribut digabungkan agar atribut-atribut tersebut seimbang. Tetapi, jika sebuah atribut masih tidak seimbang

setelah penggabungan kelas-kelas minoritas, maka atribut tersebut dihapuskan. Atribut-atribut lainnya yang dihapus merupakan atribut yang tidak digunakan secara konteks seperti tahun meninggal dan status peserta pada saat penulisan dataset peserta BPJS, atribut yang redundan seperti atribut yang menjelaskan suatu atribut lainnya dan atribut yang hanya memiliki satu jenis kelas, dan atribut yang menjelaskan ID peserta, keluarga, dan kunjungan. Setelah beberapa atribut dihapuskan, data duplikat di dalam kedua dataset dihapuskan untuk mencari anomali numerik univariat menggunakan boxplot yang dimodifikasi untuk mengatasi atribut yang memiliki data yang tidak seimbang. Nilai dalam atribut yang lebih besar daripada batas atas boxplot dan lebih kecil daripada batas kecil boxplot merupakan anomali numerik univariat.

Kedua dataset dilakukan standarisasi bagi atribut-atribut numerik dan diubah menjadi data dummy dengan One Hot Encoder bagi atribut-atribut kategorikal. Anomali numerik multivariat dan anomali multidimensional dicari dengan cara yang hampir sama, dengan anomali multidimensional menggunakan atribut numerik dan kategorikal sedangkan anomali numerik multivariat menggunakan atribut numerik saja. Khusus anomali multidimensional, atribut-atribut kategorikal diubah menjadi data dummy sebelum dilakukan standarisasi. Khusus dataset kunjungan FKRTL, dikarenakan performa perangkat keras dan besar dataset yang digunakn, dataset tersebut akan di-sample sebesar 1 persen dari banyak data. Kedua anomali tersebut dicari dengan melakukan reduksi dimensi menggunakan PCA sehingga nilai variansi masing-masing dataset memiliki nilai mendekati 90 persen. Terdapat tiga algoritma yang digunakan untuk mencari kedua anomali tersebut : DBSCAN, Isolation Forest (Isofor), dan LOF, dan terdapat lima cara menggunakan anomali tersebut : DBSCAN, Isofor, LOF, Isofor AND LOF, dan Isofor OR LOF. Masing-masing dataset menghasilkan data prediksi dari masing-masing algoritma. Lalu, dataset dan data prediksi tersebut dibagi

menjadi data training dan data test, serta data prediksi training dan data prediksi test, dengan rasio 4:1. Masing-masing algoritma dilatih (Training) dengan data training sebelum melakukan prediksi terhadap data test untuk menghasilkan data hasil tes. Data tersebut dibandingkan dengan data prediksi test untuk mencari nilai TP, FP, FN, dan TN yang menghasilkan nilai ROC-AUC, confusion matrix, dan nilai akurasi, presisi, recall, f1, spesifisitas, dan kappa. Hasil prediksi algoritma dengan nilai-nilai terbaik akan menjadi data anomali numerik multivariat dan multidimensional.

## **7.2 Saran**

Diharapkan bahwa data-data dalam BPJS dapat diperbaiki sehingga tidak adanya suatu pertentangan fakta antara berbagai jenis atribut dataset. Selain itu, deteksi anomali berikutnya diharapkan lebih baik daripada deteksi anomali yang dilakukan dalam penulisan ini.

*[Halaman ini sengaja tidak berisi kecuali kalimat ini]*

## DAFTAR PUSTAKA

- Alzubi, J., Nayyar, A., & Kumar, A. (2018, November). Machine Learning from Theory to Algorithms: An Overview. *In Journal of Physics: Conference Series*, 1142, 12012.
- Babura, B. I., Adam, M. B., Samad, A. R. A., Fitrianto, A., & Yusif, B. (2018, November). Analysis and Assessment of Boxplot Characters for Extreme Data. *In Journal of Physics: Conference Series*. 1132(1), 12078. IOP Publishing.
- Bowers, A. J., & Zhou, X. (2019). Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes. *JESPAR*, 24(1), 20-46.
- Budiarto, E. H., Permanasari, A. E., & Fauziati, S. (2019, Juli). Unsupervised Anomaly Detection Using K-means, Local Outlier Factor and One Class SVM. In *2019 5th ICST*, 1, 1-5. IEEE.
- Centers for Disease Control and Prevention. (2021, 3 November). ICD-9-CM. <https://www.cdc.gov/nchs/icd/icd9cm.htm>
- Cook, A. A., Mısırlı, G., & Fan, Z. (2019). Anomaly Detection for IoT Time-Series Data: A Survey. *IEEE Internet of Things Journal*, 7(7), 6481-6494.
- Deng, D. (2020, September). DBSCAN clustering algorithm based on density. *In 2020 7th IFEEA*. 949-953. IEEE.
- Elmrabit, N., Zhou, F., Li, F., & Zhou, H. (2020, June). Evaluation of Machine Learning Algorithms for Anomaly Detection. *In 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 1-8.
- Fadhlorrohman, M., Muliawati, A., Hananto, B. (2021). Analisis Kinerja Intrusion Detection System pada Deteksi

Anomali dengan Metode Decision Tree Terhadap Serangan Siber. *Jurnal Ilmu Komputer dan Agri-Informatika*, 8(2), 90-94.

Fang, N., Fang, X., & Lu, K. (2022). Anomalous Behavior Detection Based on the Isolation Forest Model with Multiple Perspective Business Processes. *Electronics*, 11(21), 3640.

Foorhuis, R. (2021). On the Nature and Types of Anomalies: A Review of Deviations in Data. *Int J Data Sci Anal*, 12, 297–331. <https://doi.org/10.1007/s41060-021-00265-1>

Foorhuis, R. (2019). All or in-cloud: how the identification of six types of anomalies is affected by the discretization method. *In Artificial Intelligence: 30th Benelux Conference, BNAIC 2018, Hertogenbosch, The Netherlands, 25-42*. Springer International Publishing.

Hahsler, M., Piekenbrock, M., & Doran, D. (2019). DBSCAN: Fast density-based clustering with R. *Journal of Statistical Software*, 91, 1-30

Indonesia. (2011). Undang-Undang Republik Indonesia Nomor 24 Tahun 2011 tentang Badan Penyelenggara Jaminan Sosial. *Lembaran Negara RI Tahun 2011 Nomor 116, Tambahan Lembaran RI Nomor 5256*.

Indonesia. (2016). Peraturan Menteri Kesehatan Republik Indonesia Nomor 52 Tahun 2016 tentang Standar Tarif Pelayanan Kesehatan dalam Penyelenggaraan Program Jaminan Kesehatan. *Lembaran Negara RI Tahun 2016 Nomor 1601*.

Indonesia. (2017). Peraturan Menteri Kesehatan Republik Indonesia Nomor 76 Tahun 2017 Tentang Pedoman Indonesian Case Base Groups (INA-CBG) dalam Pelaksanaan Jaminan Kesehatan Nasional. *Lembaran Negara RI Tahun 2017 Nomor 92*.

Indonesia. (2018). Peraturan Menteri Kesehatan Republik Indonesia Nomor 82 Tahun 2018 tentang Jaminan Kesehatan. *Lembaran Negara RI Tahun 2018 Nomor 165*.

Indonesia. (2022). Peraturan Pemerintah Pengganti Undang-Undang Republik Indonesia Nomor 2 Tahun 2022 tentang Cipta Kerja. *Lembaran Negara RI Tahun 2022 Nomor 238, Tambahan Lembaran RI Nomor 6841*.

Indonesia. (2023). Peraturan Menteri Kesehatan Republik Indonesia Nomor 3 Tahun 2023 tentang Standar Tarif Pelayanan Kesehatan dalam Penyelenggaraan Program Jaminan Kesehatan. *Lembaran Negara RI Tahun 2023 Nomor 35*.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine Learning and Deep Learning. *Electronic Markets*, 31(3), 685-695.

Kavitha, M., Srinivas, P. V. V. S., Kalyampudi, P. L., & Srinivasulu, S. (2021, September). Machine Learning Techniques for Anomaly Detection in Smart Healthcare. *In 3rd ICIRCA*, 1350-1356.

Labrín, C., & Urdinez, F. (2020). Principal Component Analysis. *In R for Political Data Science*, 375-393. Chapman and Hall/CRC.

Mahesh, B. (2020). Machine Learning Algorithms-A Review. *IJSR*. 9(1), 381-386.

Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M. (2021). Machine Learning for Anomaly Detection: A Systematic Review. *IEEE Access*, 9, 78658-78700. <https://doi.org/10.1109/ACCESS.2021.3083060>

Payla, Jayson. (2018). Principal Component Analysis. ResearchGate.

[https://www.researchgate.net/publication/326587453\\_Principals\\_Component\\_Analysis](https://www.researchgate.net/publication/326587453_Principals_Component_Analysis)

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *JMLR*, 12, 2825 – 2830.

Pu, G., Wang, L., Shen, J., & Dong, F. (2020). A Hybrid Unsupervised Clustering-Based Anomaly Detection Method. *Tsinghua Science and Technology*, 26(2), 146-153.

Ren, H., Xu, B., Wang, Y., Yi, C., Huang, C., Kou, X., Xing, T., Yang, M., Tong, J., & Zhang, Q. (2019, July). Time-Series Anomaly Detection Service at Microsoft. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3009-3017. <https://doi.org/10.1145/3292500.3330680>

Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., & Gehler, P. (2022). Towards Total Recall in Industrial Anomaly Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14318-14328.

Said, A. M., Yahyaoui, A., & Abdellatif, T. (2021). Efficient Anomaly Detection for Smart Hospital IoT Systems. *Sensors*, 21(4), 1026. <https://doi.org/10.3390/s21041026>

Sahoo, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory data analysis using Python. *International Journal of Innovative Technology and Exploring Engineering*, 8(12), 4727-4735.

Samariya, D., Thakkar, A. (2023). A Comprehensive Survey of Anomaly Detection Algorithms. *Ann. Data. Sci.* 10, 829–850. <https://doi.org/10.1007/s40745-021-00362-9>

Sánchez, V. G., Lysaker, O. M., & Skeie, N. O. (2020). Human Behaviour Modelling for Welfare Technology Using Hidden Markov Models. *Pattern Recognition Letters*, 137, 71-79. <https://doi.org/10.1016/j.patrec.2019.09.022>

Walker, M. L., Dovoedo, Y. H., Chakraborti, S., Hilton, C. W., (2018). An Improved Boxplot for Univariate Data. *JOUR.* 72(4), 348-353. <https://doi.org/10.1080/00031305.2018.1448891>

Wang, S., Balarezo, J. F., Kandeepan, S., Al-Hourani, A., Chavez, K. G., & Rubinstein, B. (2021). Machine Learning in Network Anomaly Detection: A Survey. *IEEE Access*, 9, 152379-152396.

World Health Organization. (2004). ICD-10 : International Statistical Classification of Diseases and Related Health Problems : Tenth Revision, 2nd Ed. World Health Organization. <https://iris.who.int/handle/10665/42980>

World Health Organization. (2021, 1 January). History of the Development of the ICD. <https://www.who.int/publications/m/item/history-of-the-development-of-the-icd>

World Health Organization. (2023). International Statistical Classification of Diseases and Related Health Problems (ICD). <https://www.who.int/standards/classifications/classification-of-diseases>

*[Halaman ini sengaja tidak berisi kecuali kalimat ini]*

## **BIODATA PENULIS**

Nama : Dewangga Dharmawan  
Tempat, Tanggal Lahir : Madiun, 20 Desember 2000  
Jenis Kelamin : Laki-laki  
Telepon : +6281252165248  
Email : [dewangga.dharmawan@gmail.com](mailto:dewangga.dharmawan@gmail.com)

### **AKADEMIS**

Kuliah : Departemen Teknik Informatika –  
FTEIC, ITS  
Angkatan : 2019  
Semester : 9 (Sembilan)