

# Penerapan *Data Mining* dengan Menggunakan Algoritma C4.5 pada Klasifikasi Fasilitas Kesehatan Provinsi di Indonesia

Anastasia Carolina Rumahorbo dan Kemal Ade Sekarwati

Program Studi Sistem Informasi Universitas Gunadarma  
Jl. Margonda Raya No. 100, Depok Jawa Barat, Indonesia  
E-mail : anastasiacarolinar@gmail.com, ade@staff.gunadarma.ac.id

## Abstrak

Kesehatan menjadi faktor penting dalam masyarakat dan merupakan salah satu pondasi suatu negara. Negara yang kuat memiliki masyarakat yang sehat dan sejahtera. Kesehatan dan kesejahteraan masyarakat merupakan salah satu fokus utama bagi pemerintahan. Pembangunan yang terus dilakukan untuk memperbaiki pelayanan serta fasilitas kesehatan merupakan salah satu usaha pemerintah untuk membangun kesehatan bangsa dan terwujudnya negara yang sehat. Fasilitas pelayanan kesehatan merupakan hak dasar masyarakat yang harus dipenuhi dalam pembangunan kesehatan sehingga akan terjadi pemerataan provinsi terhadap suatu indeks. Fasilitas kesehatan dapat dianalisa dengan menggunakan analisis *data mining*, yaitu dengan metode classification *decision tree*. Penelitian ini meliputi tahap pemahaman data, tahap pengolahan data, tahap pemodelan, dan tahap evaluasi. Pengambilan data berdasarkan data fasilitas kesehatan Indonesia tahun 2017. Data tersebut diubah menjadi data ratio, diolah dengan menggunakan software RStudio yang hasilnya menjadi sebuah model *decision tree* untuk mengetahui peringkat indeks pembangunan manusia, dan kemudian dievaluasi menggunakan metode *K-Fold Cross Validation*, serta dibandingkan dengan beberapa algoritma diantaranya C4.5 tanpa *cross validation* dan C4.5 dengan *cross validation*. Hasil perbandingan didapatkan nilai akurasi dari algoritma C4.5 dengan *cross validation* lebih baik, yaitu 97,5% dibandingkan algoritma C4.5 tanpa *cross validation*, yaitu 91,17%.

**Kata Kunci** : *Data mining*, Algoritma C4.5, Fasilitas Kesehatan

## Pendahuluan

Kesehatan menjadi faktor penting dalam masyarakat dan merupakan salah satu pondasi suatu negara. Negara yang kuat memiliki masyarakat yang sehat dan sejahtera. Kesehatan dan kesejahteraan masyarakat merupakan salah satu fokus utama bagi pemerintahan. Pembangunan yang terus dilakukan untuk memperbaiki pelayanan serta fasilitas kesehatan merupakan salah satu usaha pemerintah untuk membangun kesehatan bangsa dan terwujudnya negara yang sehat. Fasilitas pelayanan kesehatan merupakan hak dasar masyarakat yang harus dipenuhi dalam pembangunan kesehatan. Derajat

kesehatan masyarakat suatu negara dipengaruhi oleh keberadaan sarana atau fasilitas pelayanan kesehatan. Pasal 1 ayat 7 Undang-Undang Nomor 36 Tahun 2009 tentang kesehatan masyarakat menyatakan bahwa fasilitas pelayanan kesehatan adalah suatu alat dan/atau tempat yang digunakan untuk menyelenggarakan upaya pelayanan kesehatan, baik promotif, preventif, kuratif, maupun rehabilitatif yang dilakukan pemerintah, pemerintah daerah, dan/atau masyarakat.

Jumlah Puskesmas di Indonesia sampai dengan Desember 2017 adalah 9.825 unit, yang terdiri dari 3.454 unit Puskesmas rawat inap dan 6.371 unit Puskesmas non rawat inap. Provinsi dengan rasio Puskesmas ter-

hadap kecamatan tertinggi adalah Provinsi DKI Jakarta sebesar 7,73 Puskesmas per kecamatan, sedangkan Papua memiliki rasio terendah sebesar 0,70 Puskesmas per kecamatan. Rasio Puskesmas per kecamatan tersebut dapat menggambarkan kondisi aksesibilitas masyarakat terhadap pelayanan kesehatan primer [1].

Tahun 2017, terdapat 8.610 Klinik yang tersebar di seluruh Indonesia, yang terdiri dari 969 Klinik Utama dan 7.641 Klinik Pratama. Provinsi dengan Klinik Utama terbanyak adalah Provinsi DKI Jakarta sebesar 207 Klinik Utama, sedangkan untuk Provinsi dengan Klinik Pratama terbanyak adalah Provinsi Jawa Tengah sebesar 919 Klinik Pratama dan Provinsi dengan jumlah Klinik Pratama paling sedikit adalah Kalimantan Utara sebesar 1 Klinik Pratama. Jumlah Rumah Sakit di Indonesia sampai dengan tahun 2017 adalah 2.776, yang terdiri dari 2.198 Rumah Sakit Umum (RSU) dan 578 Rumah Sakit Khusus (RSK). Rumah Sakit juga dikelompokkan berdasarkan fasilitas dan kemampuan pelayanan menjadi Kelas A, Kelas B, Kelas C, dan Kelas D. Tahun 2017 terdapat 71 RS Kelas A, 397 RS Kelas B, 1340 RS Kelas C, 737 RS Kelas D dan Kelas D Pratama, dan 231 RS lainnya belum ditetapkan kelas.

Terpenuhi atau tidaknya kebutuhan masyarakat terhadap pelayanan kesehatan rujukan dan perorangan di suatu wilayah dapat dilihat dari rasio tempat tidur terhadap 1.000 penduduk. Standar WHO (World Health Organization) adalah 1 tempat tidur untuk 1.000 penduduk. Rasio tempat tidur di rumah sakit di Indonesia di tahun 2017 sebesar 1,16 per 1.000 penduduk. Jumlah tempat tidur di Indonesia sudah tercukupi menurut WHO. Walaupun rasio tempat tidur terhadap jumlah penduduk di Indonesia pada tahun 2017 telah mencukupi, namun ketika diuraikan per provinsi ditemukan masih ada delapan provinsi dengan rasio tempat tidur terhadap penduduknya kurang mencukupi, yaitu Riau (0,99), Lampung (0,88), Banten (0,88), Jawa Barat (0,83), Sulawesi Barat (0,83), Nusa Tenggara Timur (0,82), Kalimantan Tengah (0,79), dan Nusa Tenggara Barat (0,68), sedangkan rasio tempat tidur rumah sakit tertinggi terdapat di Provinsi DKI Jakarta sebesar 2,24, Sulawesi Utara sebesar 2,15, dan DI Yogyakarta sebesar 1,83.

Jumlah Posyandu di Indonesia tahun 2017 adalah sebanyak 294.428 Posyandu dan sebanyak 169.087 atau sekitar 57,43% Posyandu Aktif. Posyandu Aktif adalah Posyandu yang mampu melaksanakan kegiatan utamanya secara rutin setiap bulan (KIA: ibu hamil, ibu nifas, bayi, balita, KB, imunisasi, gizi, pencegahan dan penanggulangan diare) dengan cakupan masing-masing minimal 50% dan melakukan kegiatan tambahan [1].

Berdasarkan data di atas yang diambil dari dokumen yang dihasilkan oleh Kementerian Kesehatan Republik Indonesia melalui situs <http://www.depkes.go.id/>, pemerintah seharusnya melakukan pemerataan fasilitas kesehatan yang ada terutama untuk provinsi yang masuk ke dalam tingkatan yang rendah dan seharusnya memberikan perhatian lebih kepada provinsi tersebut. *Decision tree* digunakan sebagai alat bantu dalam penelitian ini dalam mencari dan membuat keputusan masalah serta *decision tree* juga memiliki tingkat ketelitian (akurasi) yang baik.

Untuk melakukan implementasi algoritma *decision tree* dapat menggunakan Bahasa R. Bahasa R sendiri efektif dalam pengelolaan data dan menggunakan standar tertinggi bagi analisis data. Penelitian ini membatasi masalah pada data kesehatan yang dikeluarkan oleh Kementerian Kesehatan Republik Indonesia tahun 2017. Data kesehatan hanya mengambil tabel yang berhubungan dengan fasilitas kesehatan dan mengimplementasikan data fasilitas kesehatan tersebut menggunakan metode classification *decision tree* dengan algoritma C4.5 menggunakan Bahasa R. Tujuan dari penelitian ini adalah mengimplementasikan data fasilitas kesehatan Indonesia tahun 2017 dengan menggunakan metode classification *decision tree* menggunakan Bahasa R, menganalisis kinerja algoritma *decision tree* pada data fasilitas kesehatan Indonesia tahun 2017, dan melakukan klasifikasi berdasarkan fasilitas sebagai pendukung keputusan pemerintah dalam pemerataan fasilitas kesehatan yang ada terutama untuk Provinsi yang masuk ke dalam tingkatan yang rendah. Untuk itu akan dibuat penulisan dengan judul "Penerapan *data mining* Dengan Menggunakan Algoritma C4.5 Pada Klasifikasi Fasilitas Kesehatan Provinsi Di Indonesia". *Decision Tree* dapat mengidentifikasi dan melihat hubungan antara faktor yang mempengaruhi suatu masalah dan dapat mencari penyelesaian

terbaik dengan memperhitungkan faktor-faktor tersebut serta dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah [2].

### **Data Mining**

Pada dasarnya *data mining* berhubungan erat dengan analisa data dan penggunaan perangkat lunak untuk mencari pola kesamaan dalam sekumpulan data. Berikut proses tahapan *data mining* dapat dilihat pada penjelasan di bawah ini [3] :

1. Pembersihan Data (*Data Cleaning*) Proses pembersihan data ini merupakan proses menghilangkan noise dan data yang tidak konsisten atau data yang tidak relevan. Data-data yang harus dibersihkan seperti data yang memiliki isian yang tidak sempurna, seperti data yang hilang atau data yang tidak valid. Pembersihan data ini juga akan mempengaruhi hasil dari proses *data mining* yang dihasilkan karena data yang akan ditangani berkurang secara kuantitas dan kompleksitas pemrosesannya.
2. Integrasi Data (*Data Integration*) Proses ini merupakan proses penggabungan data ke dalam satu database dimana terkadang proses *data mining* yang akan dilakukan memerlukan data lebih dari satu database atau memerlukan pemrosesan menggunakan database lain. Integrasi data perlu dilakukan secara cermat karena dapat menghasilkan hasil yang menyimpang jika terjadi kesalahan pada proses ini.
3. Seleksi Data (*Data Selection*) Proses penyeleksian data yang dilakukan karena tidak semua data yang ada di dalam database akan digunakan, hanya data yang sesuai untuk di analisis yang akan digunakan.
4. Transformasi Data (*Data Transformation*) Transformasi data adalah pengubahan data atau penggabungan data yang dilakukan sebelum akan dilakukan proses *data mining* karena dalam proses *data mining* membutuhkan format data yang khusus sebelum data diaplikasikan.
5. Proses *data mining* Proses utama yang dilakukan berdasarkan metode yang sudah

dipilih untuk menghasilkan suatu pengetahuan.

6. Evaluasi Pola (*Pattern Evaluation*) Proses ini diidentifikasi pola-pola yang ada ke dalam knowledge based dimana hasil dari teknik *data mining* dievaluasi untuk dilakukan proses penilaian apakah pola hipotesa yang ada benar-benar tercapai, bila hasil pemrosesan tidak sesuai maka dapat dilakukan perbaikan pada proses *data mining*, mencoba metode pemrosesan yang lain atau menerima hasil ini sebagai bahan acuan untuk penelitian *data mining* dimasa akan datang.
7. Presentasi Pengetahuan (*Knowledge Persentation*) Presentasi pengetahuan merupakan penyajian hasil dari pemrosesan yang ada. Pada tahap ini terfokus pada bagaimana penyajian atau visualisasi dari hasil pengetahuan yang sudah didapat sehingga dapat dipahami oleh pengguna atau yang memanfaatkan hasil dari pemrosesan tersebut.

### **Pengelompokan Data Mining**

*Data Mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, berikut penjelasannya dapat dilihat dibawah ini [4].

1. Deskripsi Terkadang secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola.
2. Estimasi Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada penilaian berikutnya estimasi nilai dari vari-

abel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

3. Prediksi Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali dalam prediksi nilai dari hasil akan ada di masa mendatang. Contoh prediksi dalam bisnis dan penelitian adalah prediksi harga beras dalam tiga bulan mendatang dan prediksi presentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.
4. Klasifikasi Klasifikasi terdapat target variabel kategori. Sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Kemudian untuk menentukan pendapatan seorang pegawai, dipakai cara klasifikasi dalam *data mining*.
5. Pengklusteran Pengklusteran merupakan pengelompokan record, pengamatan atau memperhatikan, dan membentuk kelas objek-objek yang mempunyai kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain. Contoh pengklusteran dalam bisnis dan penelitian, antara lain :

- (a) Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- (b) Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku finansial.
- (c) Melakukan pengklusteran terhadap ekspresi dari gen untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.

6. Asosiasi Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Asosiasi mencari kombinasi jenis barang yang akan terjual untuk bulan depan. Contoh asosiasi dalam bisnis dan penelitian, antara lain :

- (a) Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran upgrade layanan yang diberikan.
- (b) Menentukan barang dalam supermarket yang dibeli secara bersamaan dan yang tidak pernah dibeli secara bersamaan.

### Profil Kesehatan Indonesia 2017

Profil Kesehatan Indonesia 2017 ini menyajikan data dan informasi tentang Demografi, Sarana Kesehatan, Sumber Daya Manusia Kesehatan, Pembiayaan Daerah, Kesehatan Keluarga, Pengendalian Penyakit, dan Kesehatan Lingkungan. Data dan informasi yang ditampilkan pada Profil Kesehatan Indonesia dapat membantu dalam membandingkan capaian pembangunan kesehatan antara satu provinsi dengan provinsi lainnya, mengukur capaian pembangunan kesehatan di Indonesia, serta sebagai dasar untuk perencanaan program pembangunan kesehatan selanjutnya.

### Indeks Pembangunan Manusia

Menurut United Nations Development Programme (UNDP), Indeks Pembangunan Manusia (IPM) mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. IPM dibangun melalui pendekatan tiga dimensi dasar, sebagai ukuran kualitas hidup, yaitu umur panjang dan sehat, pengetahuan dan standar hidup layak. Pembangunan manusia Indonesia terus mengalami peningkatan dari tahun ke tahun baik dengan metode lama (1996-2013) dan metode baru (2010-2017). Dengan metode baru selama periode 2010-2017, nilai IPM Indonesia telah meningkat 4,28 poin, yaitu dari 66,53 tahun 2010 menjadi 70,81 pada tahun 2017. Bila dilihat berdasarkan target APBN, yang menargetkan IPM sebesar 70,1 sudah melampaui tar-

get (IPM 2017=70,81). Untuk melihat capaian IPM antar wilayah dapat dilihat melalui pengelompokan IPM ke dalam beberapa kategori, yaitu :

- $IPM < 60$  : IPM rendah
- $60 \leq IPM < 70$  : IPM sedang
- $70 \leq IPM < 80$  : IPM tinggi
- $\geq 80$  : IPM sangat tinggi

## R Programming

R programming adalah suatu fasilitas perangkat lunak terpadu untuk manipulasi data, simulasi, kalkulasi dan peragaan grafik. R memiliki kemampuan menganalisis data dengan sangat efektif dan dilengkapi dengan operator pengolahan array dan matriks. R memiliki kemampuan penampilan grafik untuk peragaan datanya. R programming dapat di unduh secara gratis di <http://cran.p-project.org> [6].

## RStudio

RStudio adalah lingkungan utama pengembangan terintegrasi untuk R. Salah satu keunggulan pada RStudio ini adalah dapat dijalankan pada browser. Sehingga dengan dijalankan di atas browser, maka pengguna tidak memerlukan lagi instalasi R, kecuali paket(package) pemrograman sesuai dengan kebutuhan pengguna [7].

## Metode Penelitian

### Tahap Pemahaman Data

Sumber data penelitian diperoleh dari data yang dikumpulkan berdasarkan dokumen-dokumen keterangan kesehatan yang dihasilkan oleh Kementerian Kesehatan Republik Indonesia melalui situs <http://www.depkes.go.id/>. Data yang digunakan dalam penelitian ini adalah data sarana atau fasilitas kesehatan menurut provinsi tahun 2017 yang terdiri dari 34 provinsi.

### Tahap Pengolahan Data

#### 1. Ekstraksi Data

Tahapan pertama yaitu mempersiapkan data fasilitas kesehatan Indonesia tahun 2017 yang diperoleh dari website Kementerian Kesehatan Republik Indonesia dan mengekstraksi data menjadi dataset

67 variabel. Data tersebut dapat dilihat pada Tabel 1.

2. Ubah Data Menjadi Bentuk .csv Tahapan kedua yaitu mengubah data menjadi bentuk .csv. Dataset yang telah dijelaskan pada Tabel 1 tersebut disimpan dalam format .csv agar dapat digunakan pada software RStudio.
3. Ubah Data Menjadi Ratio Tahapan ketiga yaitu mengubah data menjadi ratio. Berdasarkan Peraturan Pemerintah Republik Indonesia Nomor 78 Tahun 2007 pada pasal 6 ayat 1, semua variabel Tabel 1 tersebut dibagi dengan jumlah penduduk kemudian dikalikan dengan 10.000 untuk dijadikan data ratio.
4. Pilih Variabel Pemilihan variabel menjadi langkah selanjutnya pada data ratio karena data ratio tersebut tidak semua memiliki nilai. Pemilihan variabel bertujuan agar data yang bernilai 0 pada keseluruhan kolom tidak dipakai seperti pada data RatioRSKKEP dan data JPEND juga tidak dipakai karena data jumlah penduduk menjadi pembagi di setiap kolom sehingga jumlah penduduk tersebut sudah terdapat di semua kolom. Jadi dari 67 variabel berubah menjadi 65 variabel.

### Tahap Pemodelan

#### 1. Siapkan data training

Tahap pertama adalah mempersiapkan data training. Data training diambil dari 70% dari data fasilitas kesehatan Indonesia tahun 2017 atau dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.

#### 2. Tentukan akar dari pohon

Tahap kedua adalah menentukan akar dari pohon. Untuk menentukan akar dari pohon, didasarkan pada nilai information gain tertinggi dari atribut-atribut yang ada. Sebelum mendapatkan nilai information gain, terlebih dahulu mencari nilai entropy.

Tabel 1: Data Fasilitas Kesehatan

NO	Variable	Keterangan
1	PRI	Puskesmas Rawat Inap
2	PNRI	Puskesmas Non Rawat Inap
3	TP	Total Puskesmas
4	TKU	Total Klinik Utama
5	TKP	Total Klinik Pratama
6	RSUK	Rumah Sakit Umum Kemkes
7	RSKK	Rumah Sakit Khusus Kemkes
8	JRK	Jumlah RS Kemkes
9	RSUKEP	Rumah Sakit Umum Kepolisian
10	RSUKKEP	Rumah Sakit Khusus Kepolisian
11	JRKEP	Jumlah RS Kepolisian
12	RSUTNI	Rumah Sakit Umum TNI
13	RSKTNI	Rumah Sakit Khusus TNI
14	JRTNI	Jumlah RS TNI
15	RSUKL	Rumah Sakit Umum Kem Lain
16	RSKKL	Rumah Sakit Khusus Kem Lain
17	JRKL	Jumlah RS Kementerian Lain
18	RSUPP	Rumah Sakit Umum Pem Prov
19	RSKPP	Rumah Sakit Khusus Peme Prov
20	JRPP	Jumlah RS Pemerintah Provinsi
21	RSUPK	Rumah Sakit Umum Pem Kab
22	RSKPK	Rumah Sakit Khusus Pem Kab
23	JRPK	Jumlah RS Pem Kabupaten
24	RSUPemkot	Rumah Sakit Umum Pem Kota
25	RSKPenkot	Rumah Sakit Khusus Pem Kota
26	JRPemkot	Jumlah RS Pemerintah Kota
27	RSUS	Rumah Sakit Umum Swasta
28	RSKS	Rumah Sakit Khusus Swasta
29	JRS	Jumlah RS Swasta
30	TRSU	Total Rumah Sakit Umum
31	TRSK	Total Rumah Sakit Khusus
32	TRSUK	Total Semua Rumah Sakit
33	RSKA	Rumah Sakit Kelas A
34	TTKA	Tempat Tidur Kelas A
35	RSKB	Rumah Sakit Kelas B
36	TTKB	Tempat Tidur Kelas B
37	RSKC	Rumah Sakit Kelas C
38	TTKC	Tempat Tidur Kelas C
39	RSKD	Rumah Sakit Kelas D
40	TTKD	Tempat Tidur Kelas D
41	RSNK	Rumah Sakit Non Kelas
42	TTNK	Tempat Tidur Non Kelas
43	TRSK	Total Rumah Sakit Kelas
44	TTTK	Total Tempat Tidur Kelas
45	TTKVVIP	Total Tempat Tidur Kelas VVIP
46	TTKVIP	Total Tempat Tidur Kelas VIP
47	TTK1	Total Tempat Tidur Kelas 1
48	TTK2	Total Tempat Tidur Kelas 2
49	TTK3	Total Tempat Tidur Kelas 3
50	TTRL	Total Tempat Tidur Rawat Inap Lain
51	TTRNRI	Total Tempat Tidur Non Rawat Inap Lain
52	IF	Industri Farmasi
53	IT	Industri Obat Tradisional
54	UKOT	Usaha Kecil Obat Tradisional
55	PRAK	Produksi Alat Kesehatan
56	PKRT	Produksi Perbekalan Kesehatan dan Rumah Tangga
57	IK	Industri Kosmetik
58	PBF	Pedagang Besar Farmasi
59	APK	Apotek
60	TOB	Toko Obat
61	PYAK	Penyalur Alat Kesehatan
62	IFKS	Instansi farmasi kabupaten/kota standar
63	JIFK	Jumlah Instansi Farmasi kabupaten/kota
64	TPA	Total Posyandu Aktif
65	JKP	Jumlah Keseluruhan Posyandu
66	JPEND	Jumlah Penduduk
67	PIPM	Peringkat Indeks Pembangunan Manusia

Entropy merupakan suatu parameter untuk mengukur tingkat keberagaman dari kumpulan data dan digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dari entropy disajikan pada persamaan (1) [8].

$$Entropy(S) = \sum_{i=1}^n -pi * log_2 pi \quad (1)$$

Keterangan :

S adalah himpunan kasus

n adalah jumlah nilai yang ada pada atribut target (jumlah kelas klasifikasi)

pi adalah jumlah proporsi sampel (peluang) untuk kelas i

- Hitung nilai information gain Tahap ketiga adalah menghitung nilai information gain. Information gain adalah salah satu attribute selection measure yang digunakan untuk memilih test attribute tiap node pada tree [9]. Untuk menghitung nilai information gain suatu atribut digunakan rumus seperti yang tertera dalam persamaan (2) [8].

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan :

S adalah himpunan kasus

A adalah atribut penentu

n adalah jumlah partisi atribut A

|Si| adalah jumlah kasus untuk nilai i

|S| adalah jumlah seluruh kasus dalam S

Entropy(Si) adalah entropy untuk kasus yang memiliki nilai i

- Ulangi langkah kedua hingga semua record terpartisi

Tahap keempat adalah mengulang langkah kedua, menghitung nilai entropy hingga semua record terpartisi. Proses partisi pohon keputusan akan berhenti disaat :

- Semua record dalam simpul N mendapat kelas yang sama

- Tidak ada atribut dalam record yang di partisi lagi
- Tidak ada record dalam cabang yang kosong

### Tahap Evaluasi

Evaluasi dengan confusion matrix menghasilkan nilai accuracy, sensitivity, precision, dan specificity. Rumus confusion matrix yang digunakan dalam penelitian ini menggunakan rumus accuracy, sensitivity, precision, dan specificity yang dapat dilihat di bawah ini.

Rumus accuracy :

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (3)$$

Rumus sensitivity :

$$Sensitivity = \frac{TP}{FN + TP} \times 100\% \quad (4)$$

Rumus precision :

$$Precision = \frac{TP}{FP + TP} \times 100\% \quad (5)$$

Rumus specificity :

$$Specificity = \frac{TN}{FP + TN} \times 100\% \quad (6)$$

## Hasil dan Pembahasan

### Dataset Input

Tahapan pengolahan data yang telah dijelaskan pada pemilihan variabel menghasilkan dataset 67 variabel yang kemudian dataset tersebut diubah menjadi ratio serta dilakukan pemilihan variabel menjadi 65 variabel. Berikut 65 variabel yang menjadi dataset input dalam software RStudio dapat dilihat pada Tabel 2.

### Import Dataset

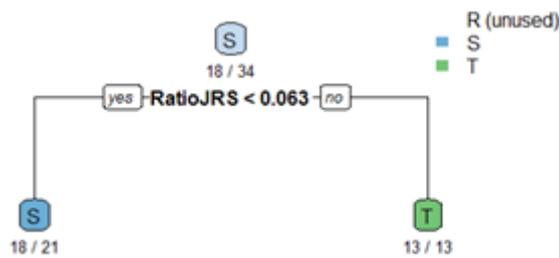
Hal pertama yang dilakukan saat import dataset ke dalam RStudio, yaitu memastikan terlebih dahulu bahwa dataset sudah disimpan ke dalam format .csv. File yang sudah dalam format .csv kemudian di impor ke dalam RStudio dengan cara pada menu Environment pilih Import Dataset – From Text(Base) lalu pilih file dataset dengan format .csv. Pilih opsi yes untuk menggunakan Heading agar mempermudah dalam penamaan setiap variabel. Perintah view dapat digunakan untuk menampilkan dataset yang sudah di impor ke dalam RStudio.

Tabel 2: Dataset input

NO	Variable	Keterangan
1	RatioPRI	Ratio Puskesmas Rawat Inap
2	RatioPNRI	Ratio Puskesmas Non Rawat Inap
3	RatioTP	Ratio Total Puskesmas
4	RatioTKU	Ratio Total Klinik Utama
5	RatioTKP	Ratio Total Klinik Pratama
6	RatioRSUK	Ratio Rumah Sakit Umum Kemkes
7	RatioRSKK	Ratio Rumah Sakit Khusus Kemkes
8	RatioJRK	Ratio Jumlah RS Kemkes
9	RatioRSUKEP	Ratio Rumah Sakit Umum Kepolisian
10	RatioJRKEP	Ratio Jumlah RS Kepolisian
11	RatioRSUTNI	Ratio Rumah Sakit Umum TNI
12	RatioRSKTNI	Ratio Rumah Sakit Khusus TNI
13	RatioJRTNI	Ratio Jumlah RS TNI
14	RatioRSUKL	Ratio Rumah Sakit Umum Kementerian Lain
15	RatioRSKKL	Ratio Rumah Sakit Khusus Kementerian Lain
16	RatioJRKL	Ratio Jumlah RS Kementerian Lain
17	RatioRSUPP	Ratio Rumah Sakit Umum Pemerintahan Provinsi
18	RatioRSKPP	Ratio Rumah Sakit Khusus Pemerintahan Provinsi
19	RatioJRPP	Ratio Jumlah RS Pemerintahan Provinsi
20	RatioRSUPK	Ratio Rumah Sakit Umum Pemerintah Kabupaten
21	RatioRSKPK	Ratio Rumah Sakit Khusus Pemerintah Kabupaten
22	RatioJRPK	Ratio Jumlah RS Pem Kab
23	RatioRSUPemKot	Ratio Rumah Sakit Umum Pemerintah Kota
24	RatioRSKPemKot	Ratio Rumah Sakit Khusus Pemerintah Kota
25	RatioJRPemKot	Ratio Jumlah RS Pem Kota
26	RatioRSUS	Ratio Rumah Sakit Umum Swasta
27	RatioRSKS	Ratio Rumah Sakit Khusus Swasta
28	RatioJRS	Ratio Jumlah RS Swasta
29	RatioTRSU	Ratio Total Rumah Sakit Umum
30	RatioTRSUK	Ratio Total Rumah Sakit Khusus
31	RatioTRSUK	Ratio Total Semua Rumah Sakit
32	RatioRSKA	Ratio Rumah Sakit Kelas A
33	RatioTTKA	Ratio Tempat Tidur Kelas A
34	RatioRSKB	Ratio Rumah Sakit Kelas B
35	RatioTTKB	Ratio Tempat Tidur Kelas B
36	RatioRSKC	Ratio Rumah Sakit Kelas C
37	RatioTTKC	Ratio Tempat Tidur Kelas C
38	RatioRSKD	Ratio Rumah Sakit Kelas D
39	RatioTTKD	Ratio Tempat Tidur Kelas D
40	RatioRSNK	Ratio Rumah Sakit Non Kelas
41	RatioTTNK	Ratio Tempat Tidur Non Kelas
42	RatioTRSK	Ratio Total Rumah Sakit Kelas
43	RatioTTTK	Ratio Total tempat Tidur Kelas
44	RatioTTKVVIP	Ratio Total Tempat Tidur Kelas VVIP
45	RatioTTKVIP	Ratio Total Tempat Tidur Kelas VIP
46	RatioTTK1	Ratio Total Tempat Tidur Kelas 1
47	RatioTTK2	Ratio Total Tempat Tidur Kelas 2
48	RatioTTK3	Ratio Total Tempat Tidur Kelas 3
49	RatioTTRIL	Ratio Total Tempat Tidur Rawat Inap Lain
50	RatioTTRNRI	Ratio Total Tempat Tidur Non Rawat Inap Lain
51	RatioIF	Ratio Industri Farmasi
52	RatioIT	Ratio Industri Obat Tradisional
53	RatioUKOT	Ratio Usaha Kecil Obat Tradisional
54	RatioPRAK	Ratio Produksi Alat Kesehatan
55	RatioPKRT	Ratio Produksi Perbekalan Kesehatan dan Rumah Tangga
56	RatioIK	Ratio Industri Kosmetik
57	RatioPBF	Ratio Pedagang Besar Farmasi
58	RatioAPK	Ratio Apotek
59	RatioTOB	Ratio Toko Obat
60	RatioPYAK	Ratio Penyalar Alat Kesehatan
61	RatioIFKS	Ratio Instansi Farmasi Kabupaten/kota Standar
62	RatioJIFK	Ratio Jumlah Instansi Farmasi Kabupaten/kota
63	RatioIPA	Ratio Total Posyandu Aktif
64	RatioJKP	Ratio Jumlah Keseluruhan Posyandu
65	PIPM	Peringkat Indeks Pembangunan Manusia

### Membuat Model Tree C4.5 tanpa Cross-Validation

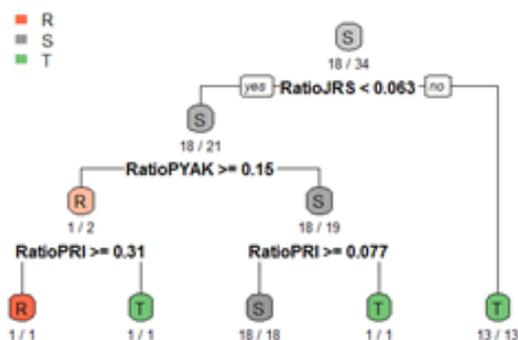
Langkah pertama yang dilakukan adalah memanggil Library J48 kemudian pilih target variabel sebagai hasil, yaitu PIPM secara default dan akan menghasilkan model seperti pada Gambar 1 di bawah ini.



Gambar 1: Tampilan Tree Default

Gambar 1 menjelaskan bahwa variabel yang diujikan pada data fasilitas kesehatan Indonesia tahun 2017, yaitu variabel RatioJRS. Variabel RatioJRS menunjukkan angka 18/34 yang artinya dari 34 provinsi dipecah menjadi 21 provinsi masuk ke dalam klasifikasi peringkat S atau Sedang dan peringkat T atau Tinggi menunjukkan angka 13/13 yang artinya dari semua 13 provinsi tersebut masuk ke dalam klasifikasi peringkat T atau Tinggi.

Decision tree pada Gambar 1 telah terbentuk, tetapi percabangan yang ditampilkan belum seluruhnya terlihat karena dalam klasifikasi peringkat S atau Sedang didefinisikan 18/21 yang artinya dari 21 provinsi hanya terdapat 18 provinsi dan tiga provinsi lainnya belum diketahui. Untuk mengetahui tiga provinsi tersebut decision tree dibuat kembali dengan beberapa pengaturan tambahan agar dapat terbentuk sepenuhnya.



Gambar 2: Tampilan Tree Full

Cara untuk menampilkan tree sepenuhnya, yaitu mengubah nilai complexity parameter

menjadi 0 dan minimal pembagian node menjadi 2. Complexity parameter digunakan untuk mengontrol ukuran pohon keputusan dan untuk memilih ukuran pohon optimal. Setelah mengubah nilai complexity parameter dan minimal pembagian node maka hasilnya dapat dilihat pada Gambar 2.

Gambar 2 menjelaskan bahwa variabel yang diujikan pada data fasilitas kesehatan Indonesia tahun 2017, yaitu RatioJRS, RatioPYAK, dan RatioPRI. Variabel tersebut yang menjadi acuan untuk data yang akan dimasukkan. Variabel RatioJRS menunjukkan angka 18/34 yang artinya 34 provinsi dipecah menjadi 21 provinsi masuk ke dalam klasifikasi peringkat S atau Sedang dan 13 provinsi masuk ke dalam klasifikasi peringkat T atau Tinggi. Jika variabel RatioJRS tidak kurang dari 0,063 maka variabel output PIPM menjadi T atau peringkat Tinggi, jika variabel RatioJRS kurang dari 0,063 maka akan dicek kembali pada variabel RatioPYAK. Variabel RatioPYAK menunjukkan angka 18/21 yang artinya 21 provinsi dipecah menjadi dua provinsi masuk ke dalam klasifikasi peringkat R atau Rendah dan 19 provinsi masuk ke dalam klasifikasi peringkat S atau Sedang. Jika variabel RatioPYAK tidak lebih besar sama dengan 0,15 maka akan dicek kembali pada variabel RatioPRI.

Variabel RatioPRI menunjukkan angka 18/19 yang artinya 19 provinsi dipecah menjadi 18 provinsi masuk ke dalam klasifikasi peringkat S atau Sedang dan satu provinsi masuk ke dalam klasifikasi peringkat T atau Tinggi. Jika variabel RatioPRI tidak lebih besar sama dengan 0,077 maka variabel output PIPM menjadi T atau peringkat Tinggi dan jika variabel RatioPRI lebih besar sama dengan 0,077 maka output PIPM menjadi S atau peringkat Sedang.

Jika variabel RatioPYAK lebih besar sama dengan 0,15 maka akan dicek kembali pada variabel RatioPRI. Variabel RatioPRI menunjukkan angka 1/2 yang artinya dua provinsi dipecah menjadi satu provinsi masuk ke dalam klasifikasi peringkat R atau Rendah dan satu provinsi masuk ke dalam klasifikasi peringkat T atau Tinggi. Jika variabel RatioPRI tidak lebih besar sama dengan 0,31 maka variabel output PIPM menjadi T atau peringkat Tinggi dan jika variabel RatioPRI lebih besar sama dengan 0,31 maka variabel output PIPM menjadi R atau peringkat Rendah.

Decision tree pada Gambar 2 dibandingkan

dengan data fasilitas kesehatan Indonesia tahun 2017 yang terdapat pada lampiran L-6 maka dihasilkan RatioJRS tidak kurang dari 0,063 sebanyak 13 provinsi yang memiliki variabel output PIPM T atau peringkat Tinggi, yaitu provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Kepulauan Riau, DKI Jakarta, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Timur, dan Sulawesi Utara. Jika RatioJRS kurang dari 0,063 maka akan dicek kembali pada variabel RatioPYAK.

Jika RatioPYAK tidak lebih besar sama dengan 0,15 maka akan dicek kembali pada variabel RatioPRI. Jika RatioPRI tidak lebih besar sama dengan 0,077 maka provinsi yang memiliki variabel output PIPM T atau peringkat Tinggi sebanyak satu provinsi, yaitu provinsi Jawa Barat. Jika variabel RatioPRI lebih besar sama dengan 0,077 maka provinsi yang memiliki variabel output PIPM S atau peringkat Sedang sebanyak 18 provinsi, yaitu provinsi Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, dan Papua Barat.

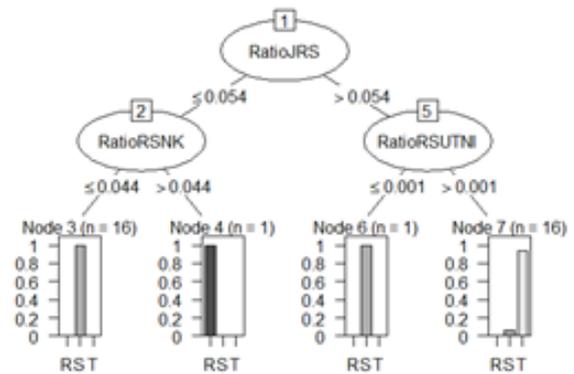
Jika variabel RatioPYAK lebih besar sama dengan 0,15 maka akan dicek kembali pada variabel RatioPRI. Jika RatioPRI tidak lebih besar sama dengan 0,31 maka provinsi yang memiliki variabel output PIPM T atau peringkat Tinggi sebanyak satu provinsi, yaitu provinsi Sulawesi Selatan dan jika variabel RatioPRI lebih besar sama dengan 0,31 maka provinsi yang memiliki variabel output PIPM R atau peringkat Rendah sebanyak satu provinsi, yaitu provinsi Papua.

### ***K-Fold Cross Validation***

*K-Fold Cross Validation* merupakan metode untuk mengevaluasi kinerja suatu algoritma. K adalah melipat data sebanyak K dan mengiterasikannya juga sebanyak K sehingga algoritma tersebut memiliki nilai keakuratan terhadap suatu data.

Tahap ini model dari algoritma *decision tree* yang diujikan adalah algoritma C4.5. Pada setiap model hal yang dilakukan adalah membuat fold dan untuk penggunaan jumlah fold terbaik untuk uji validitas dianjurkan menggunakan 10-

*fold cross validation* dalam model (Septiani, 2017). Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih percabangan optimal sampai tidak ada cabang lagi yang dihasilkan. Dalam RStudio terdapat method “J48” yang digunakan untuk menguji model tersebut dan method “cv” untuk *cross validation*. Hasil *decision tree* C4.5 menggunakan perintah plot dapat dilihat pada Gambar 3.



Gambar 3: Tampilan Algoritma C4.5 dengan *Cross Validation*

Gambar 3 menunjukkan bahwa variabel yang diujikan pada data fasilitas kesehatan Indonesia tahun 2017, yaitu RatioJRS, RatioRSNK, dan RatioRSUTNI. Variabel tersebut yang menjadi acuan untuk data yang akan dimasukkan, jika data dimasukkan dan variabel RatioJRS kurang dari sama dengan 0,054 maka akan dicek kembali pada variabel RatioRSNK.

Jika variabel RatioRSNK kurang dari sama dengan 0,044 maka variabel output PIPM menjadi S atau peringkat Sedang. Jika variabel RatioRSNK lebih besar dari 0,044 maka variabel output PIPM menjadi R atau peringkat Rendah.

Jika variabel RatioJRS lebih dari 0,054 maka akan dicek kembali pada variabel RatioRSUTNI. Jika variabel RatioRSUTNI kurang dari 0,001 maka variabel output PIPM menjadi S atau peringkat Sedang dan jika variabel RatioRSUTNI lebih dari 0,001 maka variabel output PIPM menjadi S dan T atau peringkat Sedang dan Tinggi.

*Decision tree* pada Gambar 3 dibandingkan dengan data fasilitas kesehatan Indonesia tahun 2017 yang terdapat pada lampiran L-6 maka dihasilkan RatioJRS kurang dari sama dengan

0,054 akan dicek kembali pada variabel RatioRSNK.

Jika variabel RatioRSNK kurang dari sama dengan 0,044 maka provinsi yang memiliki variabel output PIPM S atau peringkat Sedang, yaitu provinsi Jambi, Sumatera Selatan, Bengkulu, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, dan Papua Barat dan jika variabel RatioRSNK lebih dari 0,044 maka provinsi yang memiliki variabel output PIPM S atau peringkat Rendah, yaitu provinsi Papua.

Jika variabel RatioJRS lebih dari 0,054 maka akan dicek kembali pada variabel RatioRSUTNI. Jika variabel RatioRSUTNI kurang dari 0,001 maka provinsi yang memiliki variabel output PIPM menjadi S atau peringkat Sedang, yaitu provinsi Kepulauan Bangka Belitung dan jika variabel RatioRSUTNI lebih dari 0,001 didapatkan provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Lampung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Timur, Sulawesi Utara, Sulawesi Selatan.

ditambah tools untuk mengukur kinerja suatu algoritma, yaitu confusion matrix. Hasil dari confusion matrix dapat dilihat pada Tabel 3. Tabel 3 menjelaskan bahwa setiap kelas R, S, dan T memiliki nilai yang berbeda pada setiap parameternya. Pada sensitivity, prediksi kelas R didapatkan hasil NA dimana hasil tersebut memiliki kemungkinan bahwa data yang diuji terlalu sedikit dan untuk kelas S dan T didapatkan hasil sensitifitas yang sama.

Pada specificity yang menunjukkan kemampuan model untuk mengenal kelas pada S dan T memiliki tingkat spesifik yang sama sedangkan untuk R hanya memiliki hasil 1. Pada nilai pos pred value dan neg pred value sebagai hasil prediksi nilai kelas positif dan hasil prediksi nilai kelas negative pada kelas S dan T memiliki hasil yang sama sedangkan untuk kelas R didapatkan hasil NA yang memiliki kemungkinan data yang diuji terlalu sedikit.

Pada nilai prevalence, nilai detection rate, dan nilai detection prevalence kelas S lebih sering muncul pada data kesehatan Indonesia tahun 2017 yang memiliki nilai 0,5556 atau 55,56% dari data keseluruhan dan pada nilai balance accuracy kelas S dan T memiliki nilai akurasi yang seimbang.

Tabel 3: Confusion Matrix C4.5

	Class: R	Class: S	Class: T
<i>Sensitivity</i>	NA	1,0000	1,0000
<i>Specificity</i>	1	1,0000	1,0000
<i>Pos Pred Value</i>	NA	1,0000	1,0000
<i>Neg Pred Value</i>	NA	1,0000	1,0000
<i>Prevalence</i>	0	0,5556	0,4444
<i>Detection Rate</i>	0	0,5556	0,4444
<i>Detection Prevalence</i>	0	0,5556	0,4444
<i>Balanced Accuracy</i>	NA	1,0000	1,0000

Setelah mendapatkan hasil decision tree algoritma C4.5 dengan *cross validation* dan *variable important*, maka selanjutnya mengukur kinerja suatu algoritma. Dalam RStudio ter-

### Hasil Perbandingan Algoritma *Decision Tree* Terhadap Peringkat IPM

Hasil perbandingan algoritma *decision tree* C4.5 tanpa *cross validation* dan algoritma *decision tree* C4.5 dengan *cross validation* terhadap peringkat IPM pada data fasilitas kesehatan Indonesia tahun 2017 dapat dilihat pada Tabel 4.

Tabel 4 menjelaskan bahwa disetiap algoritmanya terdapat perbedaan. Perbedaan tersebut dapat dilihat dari peringkat IPM terhadap algoritma C4.5 tanpa *cross validation* dan algoritma C4.5 dengan *cross validation*.

Pada algoritma C4.5 tanpa *cross validation* terdapat 3 provinsi yang berbeda dengan peringkat IPM, yaitu provinsi Jawa Barat diprediksi dari peringkat Tinggi menjadi peringkat Sedang, provinsi Sulawesi Selatan diprediksi dari peringkat Tinggi menjadi peringkat Sedang, dan provinsi Papua diprediksi dari peringkat Rendah menjadi peringkat Sedang.

Tabel 4: Hasil Analisis

Provinsi	Peringkat IPM	C4.5 Tanpa Cross Validation	C4.5 Dengan Cross Validation
Aceh	Tinggi	Tinggi	Tinggi
Sumatera Utara	Tinggi	Tinggi	Tinggi
Sumatera Barat	Tinggi	Tinggi	Tinggi
Riau	Tinggi	Tinggi	Tinggi
Jambi	Sedang	Sedang	Sedang
Sumatera Selatan	Sedang	Sedang	Sedang
Bengkulu	Sedang	Sedang	Sedang
Lampung	Sedang	Sedang	<b>Tinggi</b>
Kep. Bangka Belitung	Sedang	Sedang	Sedang
Kep. Riau	Tinggi	Tinggi	Tinggi
DKI Jakarta	Tinggi	Tinggi	Tinggi
Jawa Barat	Tinggi	<b>Sedang</b>	Tinggi
Jawa Tengah	Tinggi	Tinggi	Tinggi
DI Yogyakarta	Tinggi	Tinggi	Tinggi
Jawa Timur	Tinggi	Tinggi	Tinggi
Banten	Tinggi	Tinggi	Tinggi
Bali	Tinggi	Tinggi	Tinggi
Nusa Tenggara Barat	Sedang	Sedang	Sedang
Nusa Tenggara Timur	Sedang	Sedang	Sedang
Kalimantan Barat	Sedang	Sedang	Sedang
Kalimantan Tengah	Sedang	Sedang	Sedang
Kalimantan Selatan	Sedang	Sedang	Sedang

Kalimantan Timur	Tinggi	Tinggi	Tinggi
Kalimantan Utara	Sedang	Sedang	Sedang
Sulawesi Utara	Tinggi	Tinggi	<b>Sedang</b>
Sulawesi Tengah	Sedang	Sedang	Sedang
Sulawesi Selatan	Tinggi	<b>Sedang</b>	Tinggi
Sulawesi Tenggara	Sedang	Sedang	Sedang
Gorontalo	Sedang	Sedang	Sedang
Sulawesi Barat	Sedang	Sedang	Sedang
Maluku	Sedang	Sedang	Sedang
Maluku Utara	Sedang	Sedang	Sedang
Papua Barat	Sedang	Sedang	Sedang
Papua	Rendah	<b>Sedang</b>	Rendah

Pada algoritma C4.5 dengan *cross validation* terdapat 2 provinsi yang berbeda dengan peringkat IPM, yaitu provinsi Lampung diprediksi dari peringkat Sedang menjadi peringkat Tinggi dan provinsi Sulawesi Utara diprediksi dari peringkat Tinggi menjadi peringkat Sedang.

## Penutup

Hasil implementasi dari data fasilitas kesehatan Indonesia tahun 2017 menggunakan metode *classification decision tree* dengan menggunakan Bahasa R mendapatkan variabel Ratio-JRS yang menjadi akar dari pohon keputusan yang artinya variabel ini yang akan diperiksa pertama kali saat data baru dimasukkan dan berperan penting pada penentuan output variabel peringkat indeks pembangunan manusia.

Hasil analisis kinerja algoritma *decision tree* C4.5 pada data fasilitas kesehatan Indonesia tahun 2017, yaitu mendapatkan hasil perbandingan nilai akurasi antara algoritma C4.5 tanpa *cross validation* dengan algoritma C4.5 dengan *cross validation*. Hasil nilai akurasi menunjukkan bahwa algoritma C4.5 dengan *cross validation* lebih baik dengan nilai 97,5%

dibandingkan algoritma C4.5 tanpa *cross validation* dengan nilai 91,17% serta mendapatkan hasil perbandingan algoritma *decision tree* C4.5 terhadap peringkat IPM. Pada algoritma C4.5 dengan *cross validation* terdapat 2 provinsi yang berbeda dengan peringkat IPM, yaitu provinsi Lampung diprediksi dari peringkat Sedang menjadi peringkat Tinggi dan provinsi Sulawesi Utara diprediksi dari peringkat Tinggi menjadi peringkat Sedang. Pada algoritma C4.5 tanpa *cross validation* terdapat 3 provinsi yang berbeda dengan peringkat IPM, yaitu provinsi Jawa Barat diprediksi dari peringkat Tinggi menjadi peringkat Sedang, provinsi Sulawesi Selatan diprediksi dari peringkat Tinggi menjadi peringkat Sedang, dan provinsi Papua diprediksi dari peringkat Rendah menjadi peringkat Sedang.

Berdasarkan hasil klasifikasi data fasilitas kesehatan Indonesia diperlukan adanya peningkatan fasilitas kesehatan terutama di provinsi dengan nilai peringkat Rendah maupun Sedang. Fasilitas Kesehatan mempengaruhi peningkatan peringkat indeks pembangunan manusia di tahun berikutnya. Hasil penelitian ini dapat dijadikan sebagai pengembangan dalam pengambilan keputusan dalam meningkatkan fasilitas kesehatan Indonesia pada Kementerian Kesehatan maupun lembaga yang memiliki keterkaitan dengan data fasilitas kesehatan Indonesia.

Penelitian ini hanya menggunakan satu kriteria dalam menguji algoritma *decision tree*, yaitu menghitung nilai akurasi. Pada penelitian selanjutnya untuk menguji algoritma *decision tree* akan lebih baik jika semua kriteria ikut diujikan seperti akurasi, kecepatan komputasi, robustness, skalabilitas, dan interpretabilitas. Penelitian ini juga menggunakan data dengan 2210 record. Pada penelitian selanjutnya untuk mengestimasi akurasi yang digunakan pada sebuah algoritma akan lebih baik jika record yang digunakan lebih banyak sehingga kemungkinan akurasi akan lebih akurat dalam sebuah algoritma. Penelitian ini masih menggunakan cara manual dalam mencocokkan provinsi yang masuk ke dalam klasifikasi Rendah, Sedang, atau Tinggi antara algoritma C4.5 yang telah dihasilkan di RStudio dengan data yang ada. Penelitian selanjutnya

akan lebih baik jika dibuatkan program dalam mencocokkan provinsi tersebut supaya meminimalisir human error.

## Daftar Pustaka

- [1] Sekretariat Jendral, "Profil Kesehatan Indonesia 2017", diakses daring <http://www.depkes.go.id/>, Maret 2019
- [2] Bambang Hermanto, S.N. Azhari, & Fajri Profesio Putra, "Analisis Kinerja *Decision Tree* C4.5 dalam Prediksi Potensi Pelunasan Kredit Calon Debitur", Vol 2, No 2, 2017.
- [3] J. Han & M. Kamber, "Data mining Concepts and Techniques", Third Edition, San Francisco: Mofgan Kaufan Publisher, 2011.
- [4] Budanis Dwi Meilani Achmad, & Fauzi Slamet, "Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode *Decision Tree*", Jurnal IPTEK ITS, Vol 16, No 1, 2012.
- [5] Sigit Abdillah, "Penerapan Algoritma *Decision Tree* C4.5 Untuk Diagnosa Penyakit Stroke Dengan Klasifikasi data mining Pada Rumah Sakit Santa Maria Pemasang", Jurnal Ilmiah Semarang, 2015.
- [6] Kelik Sussolaikah & Aslan Alwi, "Sentiment Analysis Terhadap Acara Televisi Mata Naja Berdasarkan Opini Masyarakat Pada Microblogging Twitter", KNASTIK, 2016.
- [7] Dian Anggraeni, Lutfi Ali Muharom, & Alfian Futuhul Hadi, "Rancangan Bangun Data Warehouse dan R Berbasis Web", Prosiding Seminar Nasional Matematika, 2014.
- [8] Daniel T Larose, "Discovering Knowledge in Data: An Introduction to data mining, New Jersey: John Willey & Sons, Inc, 2009.
- [9] Selvia Lorena, Wendi Zarman, & Hamidah Ida, "Analisis Dan Penerapan Algoritma C4.5 Dalam data mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik", Prosiding Seminar Nasional Aplikasi Sains & Teknologi, 2014.