



TUGAS AKHIR – SS141501

**PENGELOMPOKAN PENDERITA TUBERKULOSIS
DALAM RUMAH TANGGA DI SURABAYA
DENGAN METODE CART (*CLASSIFICATION
AND REGRESSION TREES*) *BAGGING***

**ADHEALA NATHASYA WISAKSONO
NRP 1313 100 013**

**Dosen Pembimbing
Dr. Bambang Widjanarko Otok, M.Si**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**



FINAL PROJECT – SS141501

**CLASSIFICATION OF TUBERCULOSIS PATIENTS
IN THE HOUSEHOLDS IN SURABAYA USING
CLASSIFICATION AND REGRESSION TREES
(CART) BAGGING**

**ADHEALA NATHASYA WISAKSONO
NRP 1313 100 013**

**Supervisor
Dr. Bambang Widjanarko Otok, M.Si**

**UNDERGRADUATE PROGRAM
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
SEPULUH NOPEMBER INSTITUTE OF TECHNOLOGY
SURABAYA 2017**



TUGAS AKHIR – SS141501

**PENGELOMPOKAN PENDERITA TUBERKULOSIS
DALAM RUMAH TANGGA DI SURABAYA
DENGAN METODE CART (*CLASSIFICATION
AND REGRESSION TREES*) *BAGGING***

**ADHEALA NATHASYA WISAKSONO
NRP 1313 100 013**

**Dosen Pembimbing
Dr. Bambang Widjanarko Otok, M.Si**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**

LEMBAR PENGESAHAN

**PENGELOMPOKAN PENDERITA TUBERKULOSIS
DALAM RUMAH TANGGA DI SURABAYA DENGAN
METODE CART (*CLASSIFICATION AND
REGRESSION TREES*) BAGGING**

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada

Program Studi Sarjana Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :

Adheala Nathasya Wisaksono
NRP. 1313 100 013

Disetujui oleh Pembimbing:

Dr. Bambang Widjanarko Otok, M.Si
NIP. 19681124 199412 1 001

()

Mengetahui,
Kepala Departemen

Dr. Suhartono
NIP. 19710929 199512 1 001



SURABAYA, JULI 2017



ABSTRAK



PENGELOMPOKAN PENDERITA TUBERKULOSIS DALAM RUMAH TANGGA DI SURABAYA DENGAN METODE CART (CLASSIFICATION AND REGRESSION TREES) BAGGING

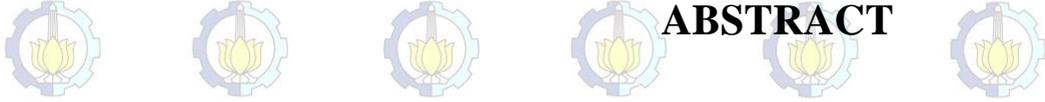
Nama Mahasiswa : Adheala Nathasya Wisaksono
NRP : 1313 100 013
Departemen : Statistika FMIPA-ITS
Dosen Pembimbing : Dr. Bambang Widjanarko Otok, M.Si.

Abstrak

Tuberkulosis di Indonesia merupakan penyebab kematian utama ketiga setelah penyakit jantung dan saluran pernafasan. Angka kejadian tuberkulosis di Jawa Timur mencapai 44.086 kasus pada tahun 2015 dengan Surabaya sebagai daerah dengan kasus tertinggi tuberkulosis di Jawa Timur. Setidaknya 10% kasus tuberkulosis terjadi di Surabaya dengan 2.330 kasus BTA Positif dan 2.409 kasus BTA Negatif. Tuberkulosis dikelompokkan menjadi dua kategori berdasarkan hasil pemeriksaan dahak yaitu BTA Positif dan BTA Negatif di mana risiko penularan BTA Positif lebih tinggi daripada risiko penularan BTA Negatif. Metode yang sesuai dalam pengelompokan kasus tuberkulosis adalah CART (Classification and Regression Trees) dan bagging karena mampu menggambarkan karakteristik penderita BTA Positif dan BTA Negatif dalam rumah tangga. Hasil penelitian ini yaitu ketepatan klasifikasi analisis CART pada data testing sebesar 76% dan penggunaan bagging tidak memberikan peningkatan ketepatan klasifikasi. Sedangkan faktor-faktor yang berpengaruh terhadap kejadian tuberkulosis adalah penghasilan KRT per bulan, tempat pembuangan air limbah, tingkat pendidikan KRT, sumber air bersih yang digunakan dan status gizi.

Kata kunci :Bagging, CART, ketepatan klasifikasi, tuberkulosis

“Halaman ini sengaja dikosongkan”



CLASSIFICATION OF TUBERCULOSIS PATIENTS IN THE HOUSEHOLDS IN SURABAYA USING CLASSIFICATION AND REGRESSION TREES (CART) BAGGING

Name : Adheala Nathasya Wisaksono
NRP : 1313 100 013
Department : Statistics
Supervisor : Dr. Bambang Widjanarko Otok, M.Si.

Abstract

Tuberculosis in Indonesia is third main cause of death after heart disease and respiratory disorder. In East Java, there were 44.086 cases of tuberculosis in 2015 while 10% of this cases happened in Surabaya which made Surabaya as the biggest area with tuberculosis in East Java. About 2.330 cases of BTA Positive and 2.409 cases of BTA Negative happened in Surabaya by 2015. Tuberculosis is classified into two different categories based on the result of mucus inspection which is BTA Positive and BTA Negative where the transmission risk of BTA Positive is bigger than BTA Negative. The use of CART (Classification and Regression Trees) and bagging is expected to be able to define the characteristics of each category. The result of this research are the accuracy of CART in data testing is 76% and the use of bagging in CART doesn't give the increase of the accuracy. While the influential factors are income of the household's head, the disposal wastage place, education of the household's head, source of clean water and nutrition status.

Keywords : Accuracy, bagging, CART, tuberculosis

“Halaman ini sengaja dikosongkan”



KATA PENGANTAR



KATA PENGANTAR

Puji syukur kehadirat Allah SWT, atas segala rahmat dan karunia-Nya yang tak pernah henti diberikan, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul

**“Pengelompokan Penderita Tuberkulosis dalam Rumah
Tangga di Surabaya dengan Metode CART (*Classification
and Regression Trees*) Bagging”**

dengan baik dan tepat waktu. Pada kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Bambang Widjanarko O, S.Si, M.Si selaku pembimbing dan Ibu Ir. Sri Pingit Wulandari, M.Si selaku pembimbing tidak resmi yang telah sabar dan memberikan waktunya untuk membimbing dan memberikan masukan dalam penyelesaian Tugas Akhir ini.
2. Ibu Santi Wulan Purnami, M.Si., Ph.D, Bapak Dr. Sutikno, M.Si dan Bapak Muhammad Mashuri, M.T selaku dosen penguji atas segala kritikan dan saran yang sangat membangun.
3. (Alm.) Papa Wisaksono, Mama Siti Nurhasanah, seluruh keluarga besar serta semua pihak yang tidak dapat disebutkan satu persatu atas semua dukungan kepada penulis.

Penulis menyadari masih banyaknya kekurangan dalam pembuatan laporan Tugas Akhir ini, besar harapan bagi penulis untuk dapat menerima saran dan kritik yang bersifat membangun guna perbaikan di masa mendatang. Semoga laporan Tugas Akhir ini dapat bermanfaat bagi penelitian selanjutnya.

Surabaya, Juli 2017

Penulis

“Halaman ini sengaja dikosongkan”



DAFTAR ISI



DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
TITLE PAGE	iii
LEMBAR PENGESAHAN	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	7
1.5 Batasan Masalah.....	7
BAB II TINJAUAN PUSTAKA	
2.1 Tuberkulosis	9
2.2 <i>Classification and Regression Trees</i> (CART)	14
2.2.1 Pembentukan Pohon Klasifikasi	17
2.2.2 Pemangkasan Pohon Klasifikasi	20
2.2.3 Penentuan Pohon Klasifikasi Optimal.....	21
2.3 Ilustrasi CART dengan Pemilihan Pemilah Indeks Gini ...	23
2.4 <i>Bootstrap Aggregating</i> (<i>Bagging</i>).....	27
2.5 Ilustrasi <i>Bagging</i> untuk Pohon Klasifikasi	28
2.6 Ketepatan Klasifikasi.....	29
BAB III METODE PENELITIAN	
3.1 Sumber Data.....	31
3.2 Kerangka Konsep Penelitian	31
3.3 Variabel Penelitian	32
3.4 Langkah Analisis Penelitian.....	35

3.5 Diagram Alir Penelitian.....	35
BAB IV ANALISIS DAN PEMBAHASAN	
4.1 Karakteristik Deskriptif Kejadian Tuberkulosis di Surabaya	37
4.2 Analisis CART untuk Klasifikasi Kejadian Tuberkulosis di Surabaya	47
4.2.1 Kemungkinan Pemilah untuk Setiap Variabel Prediktor	47
4.2.2 Pembentukan Pohon Klasifikasi Maksimal	49
4.2.3 Pemangkasan (<i>Pruning</i>) Pohon Klasifikasi Maksimal	52
4.2.4 Pemilihan Pohon Klasifikasi Optimal	54
4.2.5 Ketepatan Klasifikasi Analisis CART Penderita Tuberkulosis dalam Rumah Tangga di Surabaya 59	
4.3 Analisis CART <i>Bagging</i> untuk Klasifikasi Penderita Tuberkulosis dalam Rumah Tangga di Surabaya	61
4.4 Perbandingan Ketepatan Klasifikasi Analisis CART dan CART <i>Bagging</i>	67
BAB V KESIMPULAN DAN SARAN	
5.1 Kesimpulan.....	69
5.2 Saran.....	70
DAFTAR PUSTAKA	73
LAMPIRAN	77



DAFTAR GAMBAR



DAFTAR GAMBAR

	Halaman
Gambar 2.1	Gambaran Struktur Pohon Klasifikasi 16
Gambar 2.2	Ilustrasi Pohon Klasifikasi 27
Gambar 2.3	Ilustrasi <i>Bagging</i> pada Pohon Klasifikasi 29
Gambar 3.1	Kerangka Konsep Penelitian..... 32
Gambar 3.2	Diagram Alir Penelitian..... 36
Gambar 4.1	Jumlah Kejadian Tuberkulosis Berdasarkan Masing-masing Kategori..... 37
Gambar 4.2	Perbandingan Jumlah Anggota Keluarga Berdasarkan Masing-Masing Kategori 38
Gambar 4.3	Perbandingan Tingkat Pendidikan KRT Berdasarkan Masing-Masing Kategori 39
Gambar 4.4	Perbandingan Jenis Kelamin Berdasarkan Masing-Masing Kategori 40
Gambar 4.5	Perbandingan Penghasilan KRT per Bulan Berdasarkan Masing-Masing Kategori 41
Gambar 4.6	Perbandingan Riwayat Penyakit Berdasarkan Masing-Masing Kategori 42
Gambar 4.7	Perbandingan Kebiasaan Merokok Berdasarkan Masing-Masing Kategori 42
Gambar 4.8	Perbandingan Status Gizi Berdasarkan Masing-Masing Kategori 43
Gambar 4.9	Perbandingan Kepemilikan Toilet/ WC/ Jamban Berdasarkan Masing-Masing Kategori 44
Gambar 4.10	Perbandingan Sumber Air Bersih Berdasarkan Masing-Masing Kategori 45
Gambar 4.11	Perbandingan Tempat Pembuangan Air Limbah dari Kamar Mandi/Tempat Cuci/Dapur Berdasarkan Masing-Masing Kategori 46
Gambar 4.12	Perbandingan Kepadatan Hunian Rumah Berdasarkan Masing-Masing Kategori 46

Gambar 4.13	Pemilah Pohon Klasifikasi Maksimal	51
Gambar 4.15	Topologi Pohon Klasifikasi Optimal	55



DAFTAR TABEL



DAFTAR TABEL

	Halaman
Tabel 2.1	Variabel Penelitian Ilustrasi CART24
Tabel 2.2	Sampel Pengamatan Ilustrasi CART24
Tabel 2.3	Proporsi Pengamatan Variabel Respon25
Tabel 2.4	Perhitungan <i>Goodness of Split</i> Variabel X_125
Tabel 2.5	Perhitungan <i>Goodness of Split</i> Variabel X_226
Tabel 2.6	Perbandingan <i>Goodness of Split</i> Variabel X_1 dan X_225
Tabel 3.1	Variabel Penelitian33
Tabel 4.1	Banyaknya Kemungkinan Pemilah Variabel Prediktor48
Tabel 4.2	Variabel Terpenting dari Pohon Klasifikasi Maksimal49
Tabel 4.3	Urutan Pembentukan Pohon Klasifikasi (<i>Tree Sequence</i>)53
Tabel 4.4	Hasil Perhitungan <i>Cost Complexity Measure</i>53
Tabel 4.5	Hasil Perhitungan <i>Cost Complexity Measure (CV)</i>54
Tabel 4.6	Pelabelan Kelas Simpul Terminal57
Tabel 4.7	Hasil Klasifikasi Analisis CART untuk Data <i>Training</i>60
Tabel 4.8	Hasil Klasifikasi Analisis CART untuk Data <i>Testing</i>61
Tabel 4.9	Hasil Klasifikasi Analisis CART <i>Bagging</i> untuk Data <i>Training</i> Replikasi 10 dan 2562
Tabel 4.10	Hasil Klasifikasi Analisis CART <i>Bagging</i> untuk Data <i>Training</i> Replikasi 5063
Tabel 4.11	Hasil Klasifikasi Analisis CART <i>Bagging</i> untuk Data <i>Training</i> Replikasi 100, 125, dan 15064
Tabel 4.12	Hasil Klasifikasi Analisis CART <i>Bagging</i> untuk Data <i>Testing</i> Replikasi 10 dan 2565
Tabel 4.13	Hasil Klasifikasi Analisis CART <i>Bagging</i> untuk Data <i>Testing</i> Replikasi 5066

Tabel 4.14	Hasil Klasifikasi Analisis CART <i>Bagging</i> untuk Data <i>Testing</i> Replikasi 100, 125, dan 150.....	67
Tabel 4.15	Nilai 1-APER Analisis CART dan Analisis CART <i>Bagging</i> Penderita Tuberkulosis dalam Rumah Tangga di Surabaya.....	68



DAFTAR LAMPIRAN



DAFTAR LAMPIRAN

	Halaman
Lampiran A Data Pengamatan Kejadian Tuberkulosis	77
Lampiran B Statistik Deskriptif Variabel Prediktor.....	78
Lampiran C Output Pohon Klasifikasi <i>Cross-Validation</i> <i>10-Fold</i>	81
Lampiran D Output <i>Missclasification</i>	83
Lampiran E Akurasi Hasil Klasifikasi Analisis CART	83
Lampiran F Informasi Simpul Terminal Pohon Klasifikasi Optimal	84
Lampiran G Akurasi Hasil Klasifikasi Analisis CART <i>Bagging</i>	94
Lampiran H Surat Keterangan Penelitian.....	101

“Halaman ini sengaja dikosongkan”



BAB I
PENDAHULUAN



BAB I

PENDAHULUAN

1.1 Latar Belakang

Peningkatan kesadaran akan kesehatan oleh masyarakat saat ini sudah semakin tinggi (Goldberg, 1994). Namun meningkatnya kesadaran ini menjadi tidak berarti seiring dengan meningkatnya jumlah penderita penyakit menular dan penyakit tidak menular di Indonesia. Menurut Mohammad Ali Toha Direktur Utama Rumah Sakit Umum Pusat (RSUP) Persahabatan, jumlah penyakit tidak menular di Indonesia naik, tetapi penyakit menular juga tetap permasalahan yang besar (Bimantara, 2016). Lebih lanjut, salah satu penyakit yang saat ini menjadi fenomena besar di Indonesia adalah tuberkulosis (TB).

Tuberkulosis merupakan penyakit yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Biasanya penyakit ini menyerang organ paru manusia. Namun terdapat pula beberapa kasus lain di mana tuberkulosis menyerang organ selain paru. Berdasarkan hasil survei terbaru, Indonesia menempati negara terbanyak kedua penderita tuberkulosis di dunia setelah India (Bimantara, 2016). Kasus kejadian tuberkulosis di Indonesia pada tahun 2016 sudah mencapai 1 juta kasus per tahun atau mencapai persentase 10% dari seluruh kasus di dunia. Menurut Departemen Kesehatan RI, 95% penderita tuberkulosis paru berada di negara berkembang (Rusnoto, Rahmaatullah, & Udiono, 2008). Sekitar 75% penderita tuberkulosis adalah kelompok usia yang paling produktif secara ekonomi (15-50 tahun) dengan tingkat sosial ekonomi yang rendah (Kementerian Kesehatan RI, 2014). Tuberkulosis di Indonesia merupakan penyebab kematian utama ketiga setelah penyakit jantung dan saluran pernafasan (Rusnoto *et al*, 2008). Risiko penularan tuberkulosis di Indonesia dianggap cukup tinggi dan bervariasi pada angka 1-2% setiap tahunnya.

Angka kejadian tuberkulosis di Jawa Timur pada tahun 2014 mencapai 43.725 kasus (Nurhartanto, 2014). Pada tahun 2015, kejadian tuberkulosis di Jawa Timur mencapai 44.086

kasus (Kementerian Kesehatan Republik Indonesia, 2016) dengan 4.739 kasus tuberkulosis terjadi di Surabaya. Angka ini menempatkan Surabaya sebagai daerah dengan angka tuberkulosis tertinggi di Jawa Timur. Setidaknya terdapat 2.330 kasus BTA Positif dan 2.409 kasus BTA Negatif di Surabaya pada tahun 2015 (Dinas Kesehatan Kota Surabaya, 2015). Tingginya kejadian tuberkulosis di Surabaya menurut Harsono selaku kepala Dinas Kesehatan Jawa Timur disebabkan oleh padatnya kota Surabaya dan banyaknya lingkungan yang tidak sehat (Anwar, 2016). Kepadatan dan lingkungan yang tidak sehat menyebabkan penyebaran bakteri tuberkulosis menjadi lebih mudah. Dalam artikel lain dikatakan pula oleh Harsono bahwa gaya hidup dan kondisi lingkungan tidak sehat serta kurangnya ventilasi udara dan caayamatahari ke dalam rumah, ruangan, atau rumah dalam kondisi lembab, serta sanitasi yang kurang baik menjadi penyebab tuberkulosis berkembang (Hastareksa, 2016). Selain itu, tingginya kasus tuberkulosis disebabkan rendahnya pencegahan yang dilakukan masyarakat dan tidak adanya kecenderungan untuk berobat ketika sakit (Hastareksa, 2016).

Faktor lainnya yang mempengaruhi kemungkinan seseorang menderita tuberkulosis paru adalah daya tahan tubuh yang rendah, di antaranya disebabkan gizi buruk atau HIV/AIDS. Selain itu terdapat beberapa faktor lain yang mempengaruhi kemungkinan seseorang menderita tuberkulosis paru seperti adanya sumber penularan, riwayat kontak dengan penderita tuberkulosis paru, tingkat sosial ekonomi, tingkat paparan, virulensi basil, daya tahan tubuh rendah berkaitan dengan genetik, keadaan gizi, usia, nutrisi, hingga imunisasi (Amir & Alsegaf, 1989). Keadaan rumah yang meliputi suhu dalam rumah, ventilasi, pencahayaan, kelembaban, kepadatan penghuni, lingkungan sekitar rumah, dan pekerjaan juga dapat menjadi faktor yang mempengaruhi kemungkinan seseorang menderita tuberkulosis paru (Amir & Alsegaf, 1989).

Bentuk gejala yang ditunjukkan penderita tuberkulosis cukup beragam. Gejala umum yang sering ditemui yaitu batuk-batuk yang bisa menjadi batuk berdahak dan berlangsung selama 21 hari atau lebih, batuk mengeluarkan darah, dada yang terasa sakit saat bernafas atau batuk, tidak nafsu makan, penurunan berat badan, demam dan menggigil, berkeringat secara berlebihan pada malam hari, dan kelelahan. Namun, tidak semua bakteri tuberkulosis menyebabkan penderitanya menunjukkan gejala-gejala tersebut. Kasus di mana bakteri tuberkulosis bersembunyi dan aktif pada suatu hari disebut tuberkulosis laten dan sekitar sepertiga penduduk dunia mengidap tuberkulosis laten.

Pengobatan tuberkulosis bertujuan untuk menyembuhkan penderita, mencegah kematian, mencegah kekambuhan, memutuskan rantai penularan dan mencegah terjadinya resistensi bakteri tahan asam terhadap obat antibiotik. Pengobatan tuberkulosis dilakukan melalui pemberian antibiotik yang harus dihabiskan oleh penderita tuberkulosis dalam jangka waktu tertentu sesuai dengan resep dokter. Jenis antibiotik yang biasa digunakan untuk pengobatan tuberkulosis paru yaitu *isoniazid*, *rifampicin*, *pyrazinamide* dan *ethambutol* (Manalu, 2010). Pemberian jenis obat ini pada beberapa penderita tidak menyebabkan efek samping. Namun pada beberapa penderita lainnya ditemui efek samping seperti mual, muntah, penurunan nafsu makan, sakit kuning, urin berwarna gelap, serta gatal-gatal pada kulit (Kementerian Kesehatan RI, 2014).

Masa penyembuhan tuberkulosis berbeda-beda untuk setiap penderita. Tergantung kondisi kesehatan serta tingkat keparahan tuberkulosis yang diderita. Umumnya, kondisi penderita akan membaik dan tuberkulosis berhenti menular setelah penderita mengonsumsi antibiotik selama dua minggu. Untuk memastikan kesembuhan total penderita tuberkulosis perlu mengonsumsi antibiotik selama 6 bulan. Jika konsumsi obat dihentikan sebelum waktu yang dianjurkan, bakteri tuberkulosis tidak bisa hilang meski penderita sudah merasa kondisinya lebih baik.

Penelitian mengenai tuberkulosis sudah banyak dilakukan oleh peneliti kesehatan dan peneliti di bidang lainnya. Dalam penelitian berjudul Hubungan Antara Umur, Jenis Kelamin, dan Kepadatan Hunian dengan Kejadian TB Paru pada Pasien Rawat Jalan di Rumah Sakit Umum Daerah Noongan dikatakan bahwa terdapat hubungan antara jenis kelamin dan kejadian tuberkulosis paru pada pasien rawat jalan RSUD Noongan dan tidak terdapat hubungan antara umur dan kejadian tuberkulosis paru serta antara kepadatan hunian dan kejadian tuberkulosis paru (Korua, Kapantow, & Kawatu, 2015). Sedangkan dalam penelitian yang berjudul Faktor-faktor yang Mempengaruhi Kejadian TB Paru dan Upaya Penanggulangannya dikatakan bahwa berbagai faktor seperti kebiasaan tidak menutup mulut saat batuk, status gizi, jenis kelamin, usia, faktor sosial ekonomi, pengobatan, hingga pengetahuan tentang tuberkulosis paru berpengaruh terhadap kejadian tuberkulosis paru (Manalu, 2010).

Tuberkulosis mudah menular melalui percik dahak penderitanya. Risiko penularan yang diberikan penderita tuberkulosis BTA Positif lebih besar dibandingkan dengan risiko penularan yang diberikan oleh penderita BTA Negatif (Gultom & Yahya, 2013). Ukuran yang digunakan untuk mengetahui risiko penularan tuberkulosis setiap tahun adalah ARTI (*Annual Risk of Tuberculosis Infection*). ARTI merupakan proporsi penduduk berisiko terinfeksi tuberkulosis selama satu tahun. Risiko penularan tuberkulosis di Indonesia mencapai 1-2% per tahun (Rusnoto *et al*, 2008). ARTI 1% berarti 10 orang dari 1000 penduduk terinfeksi setiap tahun. Adanya perbedaan risiko penularan tuberkulosis BTA Positif dan BTA Negatif dirasa perlu dilakukan pengelompokan untuk mengetahui karakteristik penderita BTA Positif dan BTA Negatif penderita tuberkulosis dalam rumah tangga khususnya di Surabaya karena Surabaya merupakan daerah dengan kasus tuberkulosis tertinggi di Jawa Timur.

Untuk melakukan pengelompokan pada penderita tuberkulosis dalam rumah tangga di Surabaya diperlukan suatu

metode yang mampu mengatasi adanya dugaan hubungan antara faktor-faktor yang mempengaruhi kejadian tuberkulosis. Metode pengelompokan atau klasifikasi yang akan digunakan yaitu CART (*Classification and Regression Trees*) dengan menggunakan pendekatan *Bagging (Bootstrap Aggregating)*. Penggunaan pendekatan *bagging* dilakukan untuk meningkatkan akurasi klasifikasi yang dihasilkan metode CART.

Metode CART (*Classification and Regression Trees*) merupakan metode yang menggunakan teknik pohon keputusan. Metode ini dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen dan Charles J. Stone pada sekitar tahun 1980-an. CART merupakan metodologi statistik nonparametrik yang dikembangkan untuk topik klasifikasi, baik untuk variabel dengan respon kategorik maupun kontinu (Breiman, Friedman, Olshen, & Stone, 1984). Metode ini merupakan metode yang dapat diterapkan pada data dengan ukuran obyek dan variabel yang besar (Breiman *et al*, 1984). CART menghasilkan pohon regresi ketika variabel respon yang digunakan merupakan variabel kontinu. Sedangkan jika variabel respon yang digunakan merupakan variabel kategorik maka akan dihasilkan pohon klasifikasi.

Pohon klasifikasi yang dihasilkan CART cenderung tidak stabil dan mempengaruhi hasil klasifikasi. Untuk mengatasi ketidakstabilan ini digunakan pendekatan *bagging (bootstrap aggregating)* yang dikembangkan oleh Breiman. *Bagging* merupakan pendekatan yang dapat diterapkan untuk membantu meningkatkan stabilitas dan kekuatan prediksi pada metode klasifikasi dan metode regresi (Breiman, 1996). Sehingga hasil klasifikasi dan prediksi yang dihasilkan lebih baik.

Penelitian yang menggunakan metode CART (*Classification and Regression Trees*) dengan pendekatan *bagging (bootstrap aggregating)* sudah banyak dilakukan untuk meneliti berbagai fenomena yang terjadi di masyarakat. Dalam penelitian yang berjudul *Bagging Classification Trees Untuk Prediksi Risiko Preeklampsia (Studi Kasus: Ibu Hamil Kategori*

Penerima Jampersal di RSUD Dr. Moewardi Surakarta) diperoleh hasil penelitian yaitu metode *bagging classification trees* dapat diterapkan untuk mengklasifikasikan ibu hamil dengan risiko preeklampsia atau tidak dan pengklasifikasian dengan metode ini menunjukkan ketepatan klasifikasi sebesar 86% (Mukid, Wuryandari, Ratnaningrum, & Rahayu, 2015). Sedangkan dalam penelitian yang berjudul *Bagging CART* pada Klasifikasi Anak Putus Sekolah dikatakan bahwa status kelangsungan pendidikan anak usia wajib belajar dipengaruhi umur anak, status bekerja anak, pengeluaran rumah tangga, pendidikan kepala rumah tangga, jumlah anggota rumah tangga, dan jenis kelamin anak. Penerapan teknik *bagging* pada CART menghasilkan ketepatan klasifikasi yang lebih tinggi dibandingkan algoritma CART (Otok & Sumarmi, 2009). Sejauh ini banyak penelitian CART Bagging yang memberikan hasil bahwa *bagging* dapat meningkatkan akurasi pohon klasifikasi yang dihasilkan CART menggunakan beberapa jenis replikasi. Namun, dalam penelitian yang dilakukan Breiman (1996), *bagging* tidak selalu meningkatkan akurasi pohon klasifikasi dan pohon regresi yang dihasilkan CART.

Berdasarkan uraian tersebut, permasalahan yang akan dibahas dalam penelitian ini adalah pengelompokan penderita tuberkulosis dalam rumah tangga di Kota Surabaya dengan menggunakan metode CART *Bagging*.

1.2 Rumusan Masalah

Rumusan masalah yang diajukan dalam penelitian ini yaitu bagaimana karakteristik faktor-faktor penyebab terjadinya tuberkulosis serta pengelompokan penderita tuberkulosis dalam rumah tangga di Surabaya dengan menggunakan metode CART (*Classification and Regression Trees*) *Bagging* (*Bootstrap Aggregating*).

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah mengetahui karakteristik faktor-faktor penyebab terjadinya

tuberkulosis. Selain itu, ingin diketahui pengelompokan penderita tuberkulosis dalam rumah tangga di Surabaya dengan metode CART (*Classification and Regression Trees*) *Bagging* (*Bootstrap Aggregating*). Selain itu, hasil pengelompokan terhadap kejadian tuberkulosis di Surabaya diharapkan mampu membantu pemerintah dalam mengatasi kejadian tuberkulosis di Surabaya.

1.4 Manfaat Penelitian

Penelitian ini diharapkan mampu memberikan manfaat berupa tambahan pengetahuan terkait metode CART (*Classification and Regression Trees*) *Bagging* (*Bootstrap Aggregating*). Selain itu, diharapkan pula mampu membantu pemerintah Kota Surabaya dalam menangani kasus tuberkulosis di Surabaya.

1.5 Batasan Masalah

Batasan yang ditetapkan dalam penelitian ini yaitu data yang digunakan adalah data penderita tuberkulosis tahun 2016 di Surabaya di mana data hanya diambil di 4 wilayah di Surabaya. Cakupan pengambilan data pada penelitian ini yaitu wilayah Surabaya Pusat, Surabaya Timur, Surabaya Barat dan Surabaya Selatan.

“Halaman ini sengaja dikosongkan”



BAB II
TINJAUAN PUSTAKA



BAB II

TINJAUAN PUSTAKA

2.1 Tuberkulosis

Tuberkulosis di Indonesia merupakan penyebab kematian utama ketiga setelah penyakit jantung dan saluran pernafasan (Rusnoto *et al*, 2008). Risiko penularan tuberkulosis di Indonesia dianggap cukup tinggi dan bervariasi pada angka 1-2% setiap tahunnya. Berdasarkan hasil survei terbaru, Indonesia menempati negara terbanyak kedua penderita tuberkulosis di dunia setelah India (Bimantara, 2016). Kasus kejadian tuberkulosis di Indonesia pada tahun 2016 sudah mencapai 1 juta kasus per tahun atau mencapai persentase 10% dari seluruh kasus di dunia. Menurut Departemen Kesehatan RI, 95% penderita tuberkulosis paru berada di negara berkembang (Rusnoto *et al*, 2008). Sekitar 75% penderita tuberkulosis adalah kelompok usia yang paling produktif secara ekonomi (15-50 tahun) dengan tingkat sosial ekonomi yang rendah (Kementerian Kesehatan RI, 2014).

Tuberkulosis merupakan penyakit yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Bakteri *Mycobacterium tuberculosis* pertama kali diperkenalkan oleh Robert Koch di Berlin, Jerman pada 24 Maret 1882 (Muniroh, Aisah, & Mifbakhuddin, 2013). Bakteri tuberkulosis berbentuk batang dan bersifat tahan asam sehingga dikenal juga sebagai Batang Tahan Asam (BTA). Penyakit ini menyerang organ tubuh manusia seperti paru dan organ tubuh lainnya sehingga tuberkulosis diklasifikasikan menjadi tuberkulosis paru dan tuberkulosis ekstra paru (Departemen Kesehatan RI, 2009). Tuberkulosis ekstra paru menyerang organ tubuh lain seperti pleura, selaput otak, selaput jantung (*pericardium*), kelenjar lymfe, tulang, persendian kulit, usus, ginjal, saluran kencing, alat kelamin dan lain-lain.

Selain diklasifikasikan berdasarkan organ yang diserang bakteri tuberkulosis, penyakit ini juga diklasifikasikan berdasarkan hasil pemeriksaan dahak. Menurut hasil pemeriksaan dahak, tu-

berkubosis dibedakan menjadi dua kategori (Perhimpunan Dokter Paru Indonesia, 2006), yaitu:

1. Tuberkulosis BTA Positif

Ciri-ciri penderita tuberkulosis BTA Positif yaitu:

- a. Minimal 2 dari 3 spesimen dahak menunjukkan BTA Positif
- b. Hasil pemeriksaan satu spesimen dahak menunjukkan BTA Positif dan kelainan radiologi menunjukkan gambaran tuberkulosis aktif.
- c. Hasil pemeriksaan satu spesimen dahak menunjukkan BTA Positif dan terdapat perkembangbiakkan *Mycobacterium tuberculosis*.

2. Tuberkulosis BTA Negatif

Ciri-ciri penderita tuberkulosis BTA Negatif yaitu:

- a. Hasil pemeriksaan dahak tiga kali menunjukkan BTA negatif, gambaran klinis, dan kelainan radiologi menunjukkan tuberkulosis aktif.
- b. Hasil pemeriksaan dahak tiga kali menunjukkan BTA Negatif dan terdapat perkembangbiakkan *Mycobacterium tuberculosis*.

Penularan tuberkulosis seringkali terjadi melalui udara. Dalam Pedoman Nasional Pengendalian Tuberkulosis yang diterbitkan oleh Kementerian Kesehatan RI penularan tuberkulosis dibedakan menjadi empat cara (Kementerian Kesehatan RI, 2014), yaitu:

1. Sumber penularan adalah pasien tuberkulosis BTA Positif melalui percik renik dahak yang dikeluarkannya. Namun, bukan berarti bahwa pasien tuberkulosis dengan hasil pemeriksaan BTA Negatif tidak mengandung kuman dalam dahaknya.
2. Sumber penularan tuberkulosis juga dapat berasal dari percikan dahak pasien tuberkulosis BTA Negatif. Tingkat penularan pasien tuberkulosis BTA Positif adalah 65% sedangkan tingkat penularan pasien tuberkulosis negatif dengan kultur positif mencapai 26% dan tingkat pasien tu-

berkubosis negatif dengan kultur negatif dan foto toraks positif adalah 17%.

3. Infeksi akan terjadi apabila orang lain menghirup udara yang mengandung percik renik dahak yang infeksius tersebut.
4. Pada waktu batuk atau bersin, pasien menyebarkan kuman ke udara dalam bentuk percikan dahak.

Menurut Hiswani (2009) beberapa faktor yang memengaruhi kejadian tuberkulosis yaitu faktor sosial ekonomi, status gizi, usia, dan jenis kelamin. Sedangkan menurut Fatimah (2008), faktor demografi dan lingkungan juga memengaruhi kejadian tuberkulosis. Widyasari (2011) menambahkan bahwa riwayat penyakit yang pernah diderita turut mempengaruhi kejadian tuberkulosis. Pola perilaku penderita tuberkulosis menurut Notoadmojo (2005) mempengaruhi kejadian tuberkulosis dan pendidikan menurut penelitian yang dilakukan Perkumpulan Pemberantasan Tuberkulosis Indonesia (2002) juga memengaruhi kejadian tuberkulosis. Selanjutnya dalam penelitian ini digunakan gabungan faktor-faktor yang telah diteliti sebelumnya. Berikut merupakan uraian faktor-faktor yang mempengaruhi kejadian tuberkulosis.

1. Demografi

Demografi dijelaskan sebagai ilmu yang mempelajari secara statistik dan matematik tentang besar, komposisi, dan distribusi penduduk serta perubahan-perubahannya sepanjang masa melalui bekerjanya lima komponen demografi, yaitu kelahiran, kematian, perkawinan, migrasi dan mobilitas sosial (Bogue, 1969). Penelitian yang telah dilakukan sebelumnya oleh Fatimah (2008) menunjukkan bahwa faktor demografi yang mempengaruhi terjadinya tuberkulosis adalah umur, jenis kelamin, dan kondisi sosial. Berdasarkan usia, penyakit tuberkulosis dapat menyerang siapa saja terutama penduduk usia produktif (15-50 tahun) (Departemen Kesehatan RI, 2009). Menurut Kementerian Kesehatan RI (2014), 75% penderita tuberkulosis berusia 15-50 tahun. Jika ditinjau dari jenis kelaminnya, penyakit tuberkulosis lebih banyak menyerang laki-laki daripada perempuan. Hal ini

disebabkan oleh kebiasaan merokok yang lebih banyak dilakukan laki-laki daripada perempuan. Kebiasaan merokok dapat meningkatkan risiko infeksi tuberkulosis pada pria sebanyak 2,2 kali (Kesehatan Masyarakat, 2011).

2. Lingkungan

Faktor lingkungan yang berpengaruh terhadap terjadinya penyakit tuberkulosis yaitu faktor kesehatan lingkungan rumah dan sanitasi dasar seperti pencahayaan, ventilasi dan jenis dinding di rumah (Fatimah, 2008). Dinas Kesehatan Provinsi Jawa Timur mendefinisikan rumah sehat sebagai bangunan rumah yang memenuhi syarat kesehatan seperti ketersediaan jamban sehat, sarana air bersih, tempat pembuangan sampah, sarana pembuangan air limbah, ventilasi yang baik, kepadatan hunian rumah yang memenuhi syarat dan lantai rumah yang tidak terbuat dari tanah (Dinas Kesehatan Provinsi Jawa Timur, 2013). Kepadatan hunian rumah yang memenuhi syarat diatur dalam KEPMENKES RI NO. 829/MENKES/SK/VII/1999 tentang Persyaratan Kesehatan Perumahan yang berbunyi rumah dikatakan tidak padat penghuninya bila luas lantai kamar tidur $\geq 8\text{m}^2/\text{orang}$.

3. Status gizi

Keadaan malnutrisi atau kekurangan kalori, protein, vitamin, zat besi dan lain-lain akan mempengaruhi daya tahan tubuh seseorang sehingga rentan terhadap penyakit termasuk penyakit tuberkulosis (Hiswani, 2009). Keadaan ini merupakan faktor penting yang berpengaruh di negara miskin baik pada orang dewasa maupun anak-anak (Hiswani, 2009). Gizi yang seimbang berpengaruh terhadap ketahanan tubuh seseorang terutama dalam menghadapi serangan penyakit. Penyakit tuberkulosis yang disebabkan oleh bakteri semakin mudah menyerang calon penderita dengan ketahanan tubuh yang rendah.

Menurut Supriasa (2002), status gizi merupakan ekspresi dari keadaan keseimbangan dalam bentuk variabel tertentu atau perwujudan malnutrisi dalam bentuk variabel tertentu. Selanjutnya status gizi dibedakan menjadi dua yaitu status gizi normal dan malnutrisi yang mewakili keadaan kelebihan dan kekurangan

gizi. Status gizi normal merupakan keadaan tubuh yang mencerminkan keseimbangan antara konsumsi dan penggunaan gizi oleh tubuh. Sedangkan malnutrisi adalah keadaan patologis akibat kekurangan atau kelebihan secara relatif maupun absolut satu atau lebih zat gizi (Supariasa, 2002).

4. Riwayat Penyakit Penyerta Tuberkulosis

Tidak jauh berbeda dengan status gizi, riwayat penyakit yang pernah diderita turut mempengaruhi penyebab terjadinya tuberkulosis. Adanya riwayat penyakit yang pernah diderita meningkatkan risiko seseorang menderita tuberkulosis. Beberapa penyakit seperti *diabetes mellitus* (DM), HIV/AIDS, gagal ginjal, kanker, dan hepatitis akut merupakan penyakit yang dapat menyebabkan penyakit tuberkulosis. Dalam Jurnal Kesehatan Masyarakat dikatakan bahwa penderita *diabetes mellitus* memiliki risiko 5 kali lebih besar terinfeksi tuberkulosis dibandingkan dengan orang tanpa riwayat penyakit *diabetes mellitus* (Widyasari, 2011).

5. Sosial Ekonomi

Pendapatan merupakan salah satu indikator faktor sosial ekonomi seseorang. Pendapatan penting untuk menopang kebutuhan seseorang. Pendapatan yang tidak mencukupi bagi suatu rumah tangga menyebabkan kurangnya daya beli dalam memenuhi konsumsi rumah tangga khususnya konsumsi makanan sehingga secara tidak langsung berpengaruh pada kesehatan anggota rumah tangga. Umumnya penderita tuberkulosis di dunia berasal dari keluarga dengan kondisi sosial ekonomi lemah atau miskin (WHO, 2013). Rumah tangga miskin dalam indikator yang ditetapkan oleh Badan Pusat Statistik Surabaya adalah rumah tangga dengan sumber penghasilan kepala rumah tangga di bawah Rp 600.000,00 per bulan (Dinas Kesehatan Pemerintah Kota Surabaya, 2015).

Selain pendapatan, faktor sosial ekonomi lain yang juga mempengaruhi terjadinya tuberkulosis yaitu keadaan rumah, kepadatan hunian, lingkungan perumahan, lingkungan dan sanitasi tempat kerja yang buruk (Hiswani, 2009). Rumah yang terlalu

padat tidak mampu memberikan kualitas udara dan pencahayaan yang baik.

6. Pola Perilaku Penderita Tuberkulosis

Salah satu faktor yang mempengaruhi terjadinya tuberkulosis yaitu pola perilaku penderita tuberkulosis. Seseorang yang memiliki perilaku positif terhadap kesehatan seperti mengkonsumsi makanan bergizi dan mencuci tangan sebelum makan maka kemungkinan besar orang tersebut lebih sehat (Notoadmojo, 2005). Beberapa perilaku yang dapat meningkatkan risiko terjadinya tuberkulosis adalah kebiasaan merokok, mengkonsumsi narkotika dan minuman keras.

7. Pendidikan

Faktor lain yang secara tidak langsung mempengaruhi terjadinya tuberkulosis adalah pendidikan. Pendidikan merupakan usaha yang dengan sengaja dipilih untuk mempengaruhi dan membantu anak yang bertujuan untuk meningkatkan ilmu pengetahuan, jasmani dan akhlak sehingga secara perlahan bisa mengantarkan anak kepada tujuan dan cita-citanya yang paling tinggi (Yunus, 1993). Pendidikan mencerminkan pengetahuan dan perilaku seseorang. Sehingga diasumsikan orang dengan pendidikan yang lebih tinggi memiliki pengetahuan yang lebih terkait dengan pola hidup sehat dan pengobatan mengenai suatu penyakit. Semakin tinggi pendidikan seseorang akan menambah pengetahuan dan kesadaran mereka terutama di bidang kesehatan (Mazayudha & Mandakir, 2014). Namun, berdasarkan hasil penelitian, orang dengan pendidikan tinggi belum tentu mempunyai kesadaran yang lebih baik tentang penyakitnya dibandingkan orang dengan pendidikan rendah (Perkumpulan Pemberantasan Tuberkulosis Indonesia, 2012). Sehingga pengaruh faktor pendidikan terhadap terjadinya tuberkulosis perlu diteliti lebih lanjut.

2.2 Classification and Regression Trees (CART)

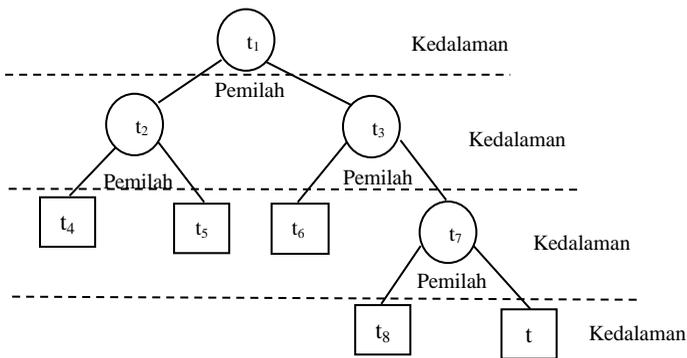
Metode CART (*Classification and Regression Trees*) merupakan metode yang menggunakan teknik pohon keputusan. Metode ini dikembangkan oleh empat orang yaitu, Leo Breiman, Jerome H. Friedman, Richard A. Olshen dan Charles J. Stone pa-

da sekitar tahun 1980-an. CART merupakan metodologi statistik nonparametrik yang dikembangkan untuk topik klasifikasi, baik untuk variabel dengan respon kategorik maupun kontinu (Breiman *et al*, 1984). Metode ini merupakan metode yang dapat diterapkan pada data dengan ukuran obyek dan variabel yang besar (Breiman *et al*, 1984). CART menghasilkan pohon regresi ketika variabel respon yang digunakan merupakan variabel kontinu. Sedangkan jika variabel respon yang digunakan merupakan variabel kategorik maka akan dihasilkan pohon klasifikasi. CART memiliki kemampuan untuk menyeleksi variabel-variabel dan interaksi-interaksi variabel yang paling penting dalam menentukan hasil atau variabel responnya. Tujuan utama CART yaitu untuk mendapatkan suatu kelompok data yang akurat sebagai ciri dari suatu klasifikasi. Kekurangan metode CART yaitu pohon klasifikasi yang dihasilkan tidak stabil artinya ketika terjadi sedikit perubahan pada data *learning* maka hasil prediksi pohon yang diperoleh dapat mengalami perubahan cukup besar sehingga mempengaruhi hasil klasifikasi (Sutton, 2005). Sedangkan kelebihan CART dibandingkan dengan metode klasifikasi lainnya yaitu hasil yang diperoleh lebih mudah diinterpretasikan, lebih akurat dan lebih cepat proses perhitungannya (Otok & Sumarmi, 2009). Kelebihan lain yang dimiliki CART yaitu dapat diterapkan pada himpunan data berjumlah besar, memiliki banyak variabel dan skala variabel campuran. Sedangkan menurut Lewis (2000), kelebihan CART adalah sebagai berikut:

1. Tidak memiliki asumsi yang harus dipenuhi karena merupakan metode nonparametrik.
2. Memudahkan dalam hal eksplorasi dan pengambilan keputusan pada struktur data yang kompleks dan multi variabel karena struktur data dapat dilihat secara visual.
3. Hasil klasifikasi akhir berbentuk sederhana dan dapat mengklasifikasikan data baru secara lebih efisien serta mudah diinterpretasikan.

Cara kerja CART dalam pembuatan sebuah pohon klasifikasi dikenal dengan istilah *binary recursive partitioning*. Istilah *binary*

digunakan karena setiap simpul induk akan selalu mengalami pemilahan ke dalam tepat dua simpul anak. Penggunaan istilah *recursive* berarti bahwa proses pemilahan tersebut diulang kembali pada setiap simpul anak sebagai hasil pemilahan sebelumnya, sehingga simpul anak tersebut menjadi simpul induk. Proses pemilahan ini dilakukan secara terus menerus hingga tidak dapat dilakukan pemilahan. Sedangkan istilah *partitioning* merujuk pada pengertian bahwa *learning sample* yang dimiliki dipilah lagi ke dalam bagian yang lebih kecil (Lewis, 2000).



Gambar 2.1 Gambaran Struktur Pohon Klasifikasi

Simpul utama pohon klasifikasi yang mengandung keseluruhan data disimbolkan dengan t_1 . Simpul anak atau *internal nodes* ditandai dengan t_2 , t_3 , dan t_7 . Selanjutnya simpul-simpul ini mengalami pemilahan dan menjadi simpul induk. Sedangkan t_4 , t_5 , t_6 , t_8 , dan t_9 merupakan simpul akhir (*terminal nodes*) atau simpul akhir yang sudah homogen. Simpul-simpul tersebut berada dalam tingkat kedalaman (*depth*) tertentu. Struktur pohon klasifikasi secara lebih jelas ditunjukkan oleh Gambar 2.1.

Pengklasifikasian dengan metode pohon klasifikasi melibatkan empat komponen berbeda, yaitu variabel respon, variabel prediktor, data *learning* dan data *testing*. Sedangkan langkah-langkah penerapan algoritma CART terdiri dari pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi, dan penentuan pohon klasifikasi optimal.

2.2.1 Pembentukan Pohon Klasifikasi

Langkah pembentukan pohon klasifikasi dibagi menjadi tiga tahapan berbeda, yaitu tahap pemilihan pemilah (*classifier*), penentuan simpul terminal, dan penandaan label kelas. Tahap pertama dalam pembentukan pohon klasifikasi yaitu pemilihan pemilah (*classifier*) yang dilakukan dengan tujuan untuk mengurangi tingkat heterogenitas pada simpul induk dan mendapatkan simpul anak dengan tingkat homogenitas tinggi. Pada tahapan pemilihan pemilah, setiap pemilihan hanya bergantung pada nilai yang berasal pada variabel prediktor. Jika variabel prediktor merupakan variabel kontinu dengan ruang sampel berukuran n dan terdapat n amatan yang berbeda, maka terbentuk $n-1$ jenis pemilahan yang berbeda. Namun, jika variabel prediktor merupakan variabel berskala kategorik ordinal bertaraf L , maka akan diperoleh $L-1$ jenis pemilahan.

Menurut Breiman *et al* (1984), metode pemilihan pemilah dilakukan secara coba-coba karena belum ada ketentuan tentang metode pemilihan pemilah terbaik. Beberapa macam metode pemilihan pemilah yang biasa digunakan diantaranya indeks Gini, indeks Informasi, indeks Twoing, dan indeks Entropy. Metode yang akan digunakan untuk tahap pemilihan pemilah dalam penelitian ini yaitu indeks Gini. Kelebihan indeks Gini dibandingkan dengan indeks lainnya yaitu proses perhitungan yang sederhana, relatif cepat, dan sesuai untuk diterapkan dalam berbagai kasus. Berikut merupakan rumus perhitungan indeks Gini.

$$i(t) = \sum_{i,j=1} p(j|t)p(i|t), i \neq j \quad (2.1)$$

di mana:

$i(t)$ = fungsi heterogenitas indeks Gini

$p(i|t)$ = proporsi kelas i pada simpul t

$p(j|t)$ = proporsi kelas j pada simpul t

Pemilihan pemilah terbaik dilakukan setelah diperoleh beberapa jenis pemilah. Dasar pemilihan pemilah dilakukan dengan menggunakan kriteria *Goodness of Split*. *Goodness of Split* meru-

pakan nilai yang digunakan sebagai acuan pemilihan pemilah oleh suatu pemilah s pada suatu simpul t . Ukuran ini menunjukkan penurunan heterogenitas suatu kelas dengan fungsi berikut.

$$\phi(s, t) = \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2.2)$$

di mana:

$\phi(s, t)$ = nilai *Goodness of Split*

$i(t)$ = fungsi heterogenitas pada simpul t

P_L = proporsi pengamatan simpul kiri

P_R = proporsi pengamatan simpul kanan

$i(t_L)$ = fungsi heterogenitas pada simpul anak kiri

$i(t_R)$ = fungsi heterogenitas pada simpul anak kanan

Nilai *Goodness of Split* tertinggi menunjukkan pemilah terbaik atau dengan kata lain pemilah tersebut mampu mereduksi heterogenitas secara lebih tinggi dibandingkan reduksi yang terjadi pada pemilah lainnya. Pemilah terbaik yang terpilih mampu menunjukkan variabel penting yang akan digunakan dalam proses pembentukan pohon klasifikasi. Penentuan skor untuk variabel penting pada pohon klasifikasi dihitung melalui rumusan berikut.

$$skor = \sum_{i=1}^n \phi(s, t_i) \quad (2.3)$$

di mana $\phi(s, t_i)$ merupakan nilai *Goodness of Split* pada setiap simpul (*improvement*). Skor untuk variabel penting dalam pohon klasifikasi diperoleh melalui penjumlahan *Goodness of Split* dari masing-masing variabel yang berperan sebagai pengganti untuk setiap simpul (California Statistical Software Inc., 2000).

Setelah tahapan pemilihan pemilah selesai, tahapan selanjutnya yaitu penentuan simpul terminal. Tahapan ini dilakukan untuk memutuskan apakah simpul t akan mengalami pemilahan kembali atau menjadi simpul terminal berdasarkan kriteria yang telah ditentukan. Kriteria yang digunakan untuk memutuskan apakah suatu simpul tidak akan dipilah lagi atau menjadi simpul terminal, yaitu:

1. Jika hanya ada satu pengamatan ($n = 1$) pada setiap simpul anak atau jika terdapat batasan minimum n . Pengembangan pohon klasifikasi akan berhenti jika pada simpul yang ada hanya terdapat kurang dari 5 pengamatan ($n_i < 5$) (Breiman *et al*, 1984).
2. Jika sudah mencapai batasan jumlah level yang telah ditentukan atau tingkat kedalaman pada pohon klasifikasi sudah maksimal.

Tahap terakhir dalam pembentukan pohon klasifikasi yaitu penandaan label kelas. Penandaan label kelas pada simpul terminal dilakukan untuk membedakan karakteristik dari klasifikasi pengamatan tiap kelas variabel respon yang terbentuk. Penandaan label kelas simpul terminal dilakukan berdasarkan aturan jumlah terbanyak. Rumusan yang digunakan untuk menghitung jumlah terbanyak, yaitu:

$$p(j_0 | t) = \max_j p(j | t) = \max_j \frac{N_j(t)}{N(t)} \quad (2.4)$$

di mana

$p(j | t)$ = proporsi kelas j pada simpul t

$N_j(t)$ = banyaknya amatan kelas j pada simpul terminal t

$N(t)$ = jumlah total pengamatan dalam simpul terminal t

Jika rumusan (2.4) terpenuhi maka label kelas yang digunakan untuk simpul terminal t adalah j_0 . Pada tahap ini kesalahan klasifikasi setiap kelas diasumsikan sama. Label j_0 digunakan untuk menandai estimasi kesalahan klasifikasi pada simpul t merupakan estimasi kesalahan klasifikasi terkecil. Untuk menghitung kesalahan klasifikasi tersebut digunakan perhitungan $r(t) = 1 - \max_j p(j | t)$.

Pada beberapa kasus khusus ditemukan simpul terminal dengan banyaknya data setiap kategori memiliki jumlah yang sama. Kasus seperti ini membuat penandaan label kelas tidak bisa dilakukan dengan aturan jumlah terbanyak. Untuk mengatasi adanya kasus seperti ini digunakan aturan berupa penandaan label

kelas dengan mengacu pada kategori variabel respon dengan jumlah yang lebih sedikit (Steinberg dan Golovnya, 2007). Misalnya pada simpul terminal ke delapan ditemui data dengan kategori 1 dan kategori 2 variabel prediktor masing-masing berjumlah dua data sehingga akurasi 50%. Sedangkan variabel respon yang digunakan memiliki dua kelas dengan kelas kedua berjumlah lebih sedikit daripada kategori satu, sehingga simpul terminal 8 diberi label sebagai kelas 2.

2.2.2 Pemangkasan Pohon Klasifikasi

Pohon klasifikasi yang terbentuk sebelumnya merupakan pohon klasifikasi maksimal atau masih berukuran besar. Terbentuknya pohon klasifikasi maksimal ini disebabkan oleh penghentian pohon klasifikasi yang didasarkan pada banyaknya amatan pada simpul terminal atau besarnya tingkat homogenitas. Pembentukan pohon klasifikasi yang besar dapat menimbulkan kasus *overfitting* atau nilai dugaan melebihi nilai sebenarnya. Sehingga, banyaknya pemilahan yang dilakukan dapat mengakibatkan semakin kecilnya tingkat kesalahan dalam prediksi. Sedangkan jika dalam tahap pemilahan diberikan batasan ketika pemilahan masih layak dilakukan maka dapat terjadi kasus *underfitting* atau dugaan yang kurang dari nilai sebenarnya (Breiman *et al*, 1984). Sehingga perlu dilakukan pemangkasan untuk memperoleh nilai dugaan yang tepat untuk pohon klasifikasi yang terbentuk.

Adanya kemungkinan *overfitting* maupun *underfitting* pada pohon klasifikasi yang terbentuk menyebabkan diperlukannya pemangkasan untuk mendapatkan pohon klasifikasi optimal. Untuk memperoleh pohon klasifikasi yang optimal dilakukan pemangkasan (*pruning*) terhadap bagian pohon yang kurang penting. Tujuan dilakukannya pemangkasan adalah mendapatkan ukuran pohon klasifikasi yang layak tanpa mengorbankan ketepatan atau kebaikannya.

Untuk mendapatkan ukuran pemangkasan sehingga diperoleh pohon klasifikasi yang layak digunakan ukuran *cost com-*

plexity (Breiman *et al*, 1984). Untuk $\alpha \geq 0$ maka ukuran *cost complexity* yang digunakan adalah:

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \quad (2.5)$$

di mana:

$R_\alpha(T)$ = *cost complexity measure*, kompleksitas suatu pohon T pada kompleksitas α

$R(T)$ = *resubstitution estimate*, kesalahan klasifikasi T

α = *complexity parameter*, *cost* bagi penambahan satu simpul akhir pada pohon T

$|\tilde{T}|$ = banyaknya simpul terminal pada pohon T

Ukuran *cost complexity* digunakan untuk menentukan $T(\alpha)$ yang meminimumkan $R_\alpha(T)$ pada seluruh bagian atau untuk setiap α . Persamaan yang digunakan untuk menentukan bagian $T(\alpha) < T_{\max}$ yang dapat meminimumkan $R_\alpha(T)$ dapat dituliskan sebagai berikut.

$$R_\alpha(T(\alpha)) = \min_{T < T_{\max}} R_\alpha(T) = \min_{T < T_{\max}} [R(T) + \alpha |\tilde{T}|] \quad (2.6)$$

Proses pemangkasan (*pruning*) dilakukan secara berulang hingga tidak mungkin dilakukan pemangkasan kembali. Sehingga diperoleh hasil berupa sub pohon yang semakin kecil dan tersarang.

2.2.3 Penentuan Pohon Klasifikasi Optimal

Pohon klasifikasi yang berukuran besar memberikan nilai dugaan pengganti yang paling kecil sehingga cenderung dipilih untuk menduga nilai variabel respon. Namun, pohon klasifikasi yang berukuran besar terlalu kompleks. Untuk menghindari kompleksitas yang diberikan pohon klasifikasi yang berukuran besar diperlukan pohon klasifikasi optimal dengan struktur yang lebih sederhana dan memberikan nilai dugaan pengganti yang cenderung kecil. Untuk menentukan pohon klasifikasi yang optimal dapat digunakan metode penduga sampel uji (*test sample estimate*) dan penduga validasi silang lipat V (*validation V-fold estimate*).

mate). Breiman dalam (Otok & Sumarmi, 2009), menyarankan penggunaan penduga validasi silang lipat V (*validation V-fold estimate*) untuk menghitung nilai dugaan pengganti pada sampel yang jumlahnya kurang dari 3000, sedangkan penduga sampel uji (*test sample estimate*) dapat digunakan untuk menghitung nilai dugaan pengganti pada jumlah sampel lebih dari 3000.

Berikut merupakan metode yang dapat digunakan untuk menentukan pohon klasifikasi optimal.

1. Penduga Sampel Uji (*Test Sample Estimate*)

Metode ini digunakan untuk data sampel yang berjumlah lebih dari 3000. Penggunaan metode ini diawali dengan membagi data menjadi dua bagian yaitu data *learning* (L_1) dan data *testing* (L_2). Data *learning* (L_1) digunakan untuk membentuk pohon T , sedangkan data *testing* (L_2) digunakan untuk menduga $R^{ts}(T_t)$.

Rumusan yang digunakan dalam metode ini ialah:

$$R^{ts}(T_t) = \frac{1}{N_2} \sum_{(x_n, j_n) \in L_2} X(d(x_n) \neq j_n) \quad (2.7)$$

di mana:

$R^{ts}(T_t)$ = total proporsi kesalahan *test sample estimate*

N_2 = jumlah pengamatan dari data *learning* (L_1)

$X(d(x_n) \neq j_n)$ merupakan pernyataan yang bernilai 0 jika dalam tanda kurung salah dan bernilai 1 jika dalam tanda kurung benar. Pohon klasifikasi optimal dipilih berdasarkan (T_t) dengan

$$R^{ts}(T_t) = \min_t R^{ts}(T_t).$$

2. Penduga Validasi Silang Lipat V (*Validation V-Fold Estimate*)

Metode ini digunakan untuk menghitung nilai dugaan pengganti pada sampel yang jumlahnya kurang dari 3000. Metode ini sering digunakan untuk menghindari *overlapping* pada data. Pembagian data pada metode ini dilakukan secara acak yaitu dengan membagi data pengamatan (L) menjadi V bagian (*fold*) yang saling independen dengan ukuran yang kurang lebih sama

besar. Proses pembentukan pohon klasifikasi dilakukan dengan salah satu bagian menjadi data *testing* dan bagian lainnya menjadi data *learning* dengan pengulangan sebanyak V kali. Selanjutnya dilakukan penghitungan rata-rata akurasi dari V model yang terbentuk (Lewis, 2000). Nilai V standar yang digunakan adalah 10 (*cross validation 10-fold*) namun untuk data yang lebih kecil dapat digunakan nilai V yang lebih kecil (Timofeev, 2004). Rumusan yang digunakan untuk menghitung penduga validasi silang lipat V adalah sebagai berikut,

$$R^{CV}(T_t) = \frac{1}{V} \sum_{v=1}^V R^{CV}(T_t^{(v)}) \quad (2.8)$$

di mana:

$R^{CV}(T_t)$ = total proporsi kesalahan penduga validasi silang lipat
 V

= jumlah *fold* yang digunakan

Nilai $R^{CV}(T_t)$ merupakan nilai yang diperoleh melalui penerapan data *learning* pada pohon klasifikasi T dari hasil penggunaan validasi silang lipat V . Selanjutnya nilai $R^{CV}(T_t)$ menggantikan nilai $R(T)$ dalam persamaan (2.5) kemudian ditentukan nilai *cost complexity minimum* untuk mendapatkan pohon klasifikasi optimal.

2.3 Ilustrasi CART dengan Pemilihan Pemilah Indeks Gini

Pembentukan pohon klasifikasi dengan menggunakan pemilihan pemilah indeks Gini dapat diilustrasikan secara sederhana dengan menggunakan satu variabel respon berskala nominal dan dua variabel prediktor yang masing-masing berskala ordinal dan nominal seperti variabel penelitian yang digunakan. Pada ilustrasi ini variabel respon yang digunakan memiliki dua kategori yaitu kelas 1 untuk BTA Positif dan kelas 2 untuk BTA Negatif. Selanjutnya variabel prediktor X_1 berupa variabel Jumlah Anggota Keluarga dan X_2 adalah variabel Jenis Kelamin. Tabel 2.1 menyajikan keterangan variabel yang digunakan dalam ilustrasi CART dengan pemilihan pemilah Indeks Gini.

Tabel 2.1 Variabel Penelitian Ilustrasi CART

Variabel	Keterangan	Kategori	Skala Data
Y	Kejadian Tuberkulosis	1 = Tuberkulosis BTA Positif 2 = Tuberkulosis BTA Negatif	Nominal
X ₁	Jumlah Anggota Keluarga	1 = > 7 orang 2 = ≤ 7 orang 2 = Tidak Bersekolah	Ordinal
X ₂	Jenis Kelamin	1 = Laki-laki 2 = Perempuan	Nominal

Sampel yang digunakan dalam ilustrasi CART berjumlah 10 sampel pengamatan. Sampel pengamatan yang digunakan dalam ilustrasi disajikan dalam Tabel 2.2.

Tabel 2.2 Sampel Pengamatan Ilustrasi CART

Sampel Pengamatan ke-	Y	X ₁	X ₂
1	1	2	2
2	1	2	1
3	2	1	2
4	1	2	1
5	1	1	2
6	2	2	1
7	2	1	1
8	1	2	2
9	2	2	1
10	1	2	1

Langkah pertama yang perlu dilakukan dalam pembentukan pohon klasifikasi adalah pemilihan pemilah dengan menggunakan indeks Gini. Pemilah yang terpilih akan menjadi simpul induk yang selanjutnya memilah menjadi dua simpul anak. Perhitungan nilai heterogenitas simpul induk dapat dilakukan dengan mengacu pada rumus (2.1). Setelah melakukan pemilihan pemilah selanjutnya dilakukan pemilihan *threshold* yang kemudian dilanjutkan dengan perhitungan *goodness of split*. Apabila nilai *goodness of split* telah ditentukan maka dilakukan pemilahan simpul induk

menjadi dua simpul anak. Berikut rangkaian perhitungan manual indeks Gini untuk variabel respon.

Tabel 2.3 Proporsi Pengamatan Variabel Respon

Y	$N_i(t)$	$p(i t)$
Kelas 1 (BTA Positif)	6	6/10=0,6
Kelas 2 (BTA Negatif)	4	4/10=0,4
Total	10	1

$$i(t) = \sum_{i,j=1} p(j|t)p(i|t) = p(1|t)p(2|t) + p(2|t)p(1|t)$$

$$= (0,6)(0,4) + (0,4)(0,6) = 0,48$$

Setelah diperoleh nilai heterogenitas indeks Gini, dilakukan penentuan *threshold* untuk variabel X_1 . Variabel X_1 merupakan variabel berskala ordinal dengan dua kategori sehingga banyaknya kemungkinan *threshold* hanya satu kemungkinan saja. Perhitungan berikut merupakan penjelasan dari banyaknya kemungkinan *threshold*.

$$L - 1 = 2 - 1 = 1$$

Sehingga, sampel amatan pada variabel X_1 dengan kategori 1 dipilah sebagai simpul kiri dan kategori 2 sebagai simpul kanan. Perhitungan nilai *goodness of split* untuk variabel X_1 disajikan dalam Tabel 2.4 berikut.

Tabel 2.4 Perhitungan *Goodness of Split* Variabel X_1

Simpul Kiri		Simpul Kanan	
$N_L=3$	$p_L=0,3$	$N_R=7$	$p_R=0,7$
$N(1 t_L)=1$	$p(1 t_L)=1/3$	$N(1 t_R)=5$	$p(1 t_R)=5/7$
$N(2 t_L)=2$	$p(2 t_L)=2/3$	$N(2 t_R)=2$	$p(2 t_R)=2/7$

Simpul kiri: $i(t_L)=[(1/3)(2/3)+(2/3)(1/3)]=8/9$

Simpul kanan: $i(t_R)=[(5/7)(2/7)+(2/7)(5/7)]=20/14$

Goodness of split untuk *threshold* X_1 :2 adalah

$$\phi(s, t) = 0,48 - 0,3(8/9) - 0,7(20/14) = 0,0609$$

Selanjutnya dilakukan rangkaian perhitungan yang sama pada variabel X_2 sehingga dapat diketahui nilai *goodness of split* tertinggi yang kemudian digunakan untuk menentukan pemilah terbaik dari pohon klasifikasi. Variabel X_2 merupakan variabel

berskala nominal dengan dua kategori sehingga banyaknya kemungkinan *threshold* pada variabel ini adalah satu kemungkinan saja. Perhitungan untuk mendapatkan banyaknya kemungkinan pemilah adalah sebagai berikut.

$$2^{L-1} - 1 = 2^{2-1} - 1 = 1$$

sampel amatan pada variabel X_2 dengan kategori 1 dipilah sebagai simpul kiri dan kategori 2 sebagai simpul kanan. Perhitungan nilai *goodness of split* untuk variabel X_2 disajikan dalam Tabel 2.5 berikut.

Tabel 2.5 Perhitungan *Goodness of Split* untuk Variabel X_2

Simpul Kiri		Simpul Kanan	
$N_L=5$	$p_L=0,5$	$N_R=5$	$p_R=0,5$
$N(1/t_L)=2$	$p(1/t_L)=2/5$	$N(1/t_R)=4$	$p(1/t_R)=4/5$
$N(2/t_L)=3$	$p(2/t_L)=3/5$	$N(2/t_R)=1$	$p(2/t_R)=1/5$

Simpul kiri: $i(t_L)=[(2/5)(3/5)+(3/5)(2/5)]=12/25$

Simpul kanan: $i(t_R)=[(4/5)(1/5)+(1/5)(4/5)]=8/25$

Goodness of split untuk *threshold* X_2 1:2 adalah

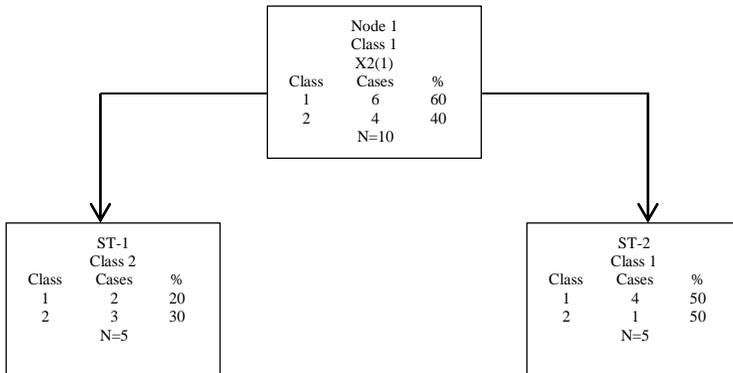
$$\phi(s, t) = 0,48 - 0,5(12/25) - 0,5(8/25) = 0,08$$

Setelah mendapatkan hasil perhitungan *goodness of split* dilakukan perbandingan nilai untuk mendapatkan nilai *goodness of split* tertinggi sehingga diperoleh pemilah terbaik untuk pohon klasifikasi. Berikut nilai *goodness of split* untuk masing-masing variabel.

Tabel 2.6 Perbandingan *Goodness of Split* Variabel X_1 dan X_2

Variabel	Threshold	Goodness of split
X_1	1:2	0,0609
X_2	1:2	0,08

Berdasarkan Tabel 2.6 dapat diketahui bahwa variabel X_2 memiliki nilai *goodness of split* tertinggi sehingga dipilih menjadi pemilah utama dalam pohon klasifikasi. Sampel amatan dalam variabel X_2 dengan kategori 1 masuk ke dalam simpul kiri sedangkan sampel amatan dengan kategori 2 masuk ke dalam simpul kanan. Berikut gambaran hasil pembentukan pohon klasifikasi.



Gambar 2.2 Ilustrasi Pohon Klasifikasi

Proses pemilahan terjadi terus menerus sesuai dengan cara kerja CART yaitu *binary recursive partitioning* atau pemilahan simpul menjadi tepat dua pemilah hingga tidak dapat dilakukan pemilahan lagi. Menurut Breiman (1984), proses pemilahan pada pembentukan pohon klasifikasi berhenti ketika pada simpul terminal hanya terdapat minimal 5 amatan saja.

2.4 *Bootstrap Aggregating (Bagging)*

Bagging adalah teknik yang diperkenalkan Breiman untuk meningkatkan stabilitas dan kekuatan prediksi pada metode regresi dan klasifikasi (Breiman, 1996). Namun, penggunaan *bagging* tidak selalu meningkatkan akurasi, hal ini dapat dipengaruhi oleh keadaan data dan pembagian *training testing* yang digunakan (Breiman, 1996). *Bagging* juga digunakan untuk mereduksi variansi estimator khususnya pada data berdimensi tinggi. *Bagging* merupakan bagian dari *bootstrap* yang dilakukan dengan membangkitkan sampel data secara berpasangan. *Bagging* sering digunakan pada metode klasifikasi karena dapat meningkatkan akurasi klasifikasi (Buhlmann & Yu, 2002). Studi yang dilakukan Breiman menunjukkan bahwa menggunakan 50 kali *bagging* untuk pohon klasifikasi menghasilkan ketepatan klasifikasi yang baik, sedangkan untuk pohon regresi dapat dilakukan sebanyak 25

kali untuk mendapatkan hasil klasifikasi yang baik (Mukid *et al*, 2015).

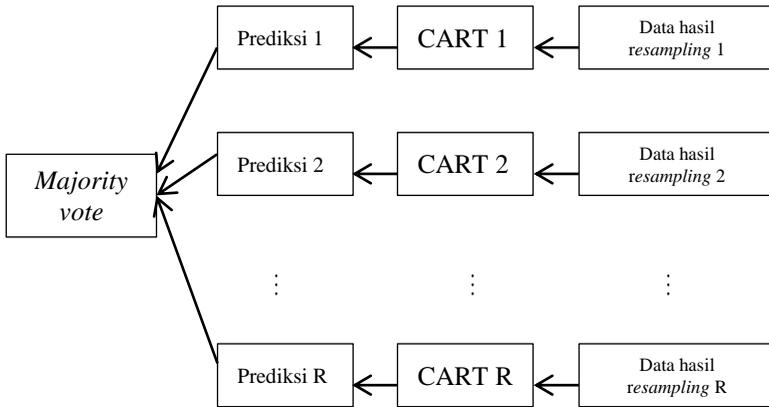
Proses pembuatan dugaan secara *bagging* menggunakan pohon klasifikasi adalah sebagai berikut (Sutton, 2005):

1. Tahapan *bootstrap*
 - a. Ambil sampel acak dengan pemilihan berukuran n dari data *learning*.
 - b. Bentuk pohon klasifikasi optimum dari sampel acak yang telah diambil.
 - c. Ulangi langkah a-b sebanyak B kali sehingga diperoleh B pohon klasifikasi. Nilai B yang digunakan yaitu 10, 25, 50, 100, 125 dan 150. Hal ini didasarkan pada pernyataan Sutton (2005) bahwa *bagging* pada umumnya memberikan hasil yang baik dari replikasi rendah. Jika saat replikasi rendah *bagging* memberikan hasil yang kurang baik, *bagging* dilakukan dengan replikasi hingga 100 atau lebih.
2. Tahapan *aggregating*

Lakukan prediksi gabungan berdasarkan B pohon klasifikasi yang terbentuk dengan menggunakan aturan *majority vote* (suara terbanyak).

2.5 Ilustrasi *Bagging* untuk Pohon Klasifikasi

Bagging merupakan teknik yang memanfaatkan proses pembangkitan data sampel secara berpasangan. Penggunaan *bagging* dalam CART diharapkan dapat meningkatkan stabilitas dan akurasi pada pohon klasifikasi dan pohon regresi yang dihasilkan. *Bagging* dapat meningkatkan stabilitas dan akurasi pada pohon klasifikasi melalui aturan *majority vote*. Sedangkan pada peningkatan akurasi dan stabilitas pohon regresi diperoleh melalui penggunaan rata-rata pada saat pengambilan sampel. Ilustrasi penerapan *bagging* pada CART dapat digambarkan sebagai berikut.



Gambar 2.3 Ilustrasi *Bagging* pada Pohon Klasifikasi

Langkah pertama dalam penggunaan *bagging* dimulai dengan pengambilan sampel acak sebanyak n dari data *learning*. Pengambilan sampel ini dilakukan dengan pengembalian sehingga tidak menutup kemungkinan bahwa suatu data terambil beberapa kali sebagai sampel acak. Selanjutnya dilakukan pembentukan pohon klasifikasi menggunakan sampel acak yang telah terambil tersebut. Setelah pohon klasifikasi optimal terbentuk tahap selanjutnya dilakukan prediksi pada setiap kelas sehingga diperoleh ketepatan klasifikasi setiap pohon. Langkah pengambilan sampel acak sebanyak n hingga melakukan prediksi pada setiap pohon klasifikasi dilakukan sebanyak replikasi *bagging* yang digunakan. Langkah terakhir dalam proses *bagging* adalah melakukan *majority vote* terhadap hasil prediksi yang diperoleh.

2.6 Ketepatan Klasifikasi

Untuk mengukur ketepatan klasifikasi dapat digunakan adalah 1-APER. Ukuran ini merupakan kebalikan dari APER sehingga menggambarkan proporsi pengamatan atau observasi yang diprediksi secara benar. Rumusan yang dapat digunakan dalam perhitungan 1-APER adalah sebagai berikut.

$$1 - \text{APER} = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}} \quad (2.10)$$

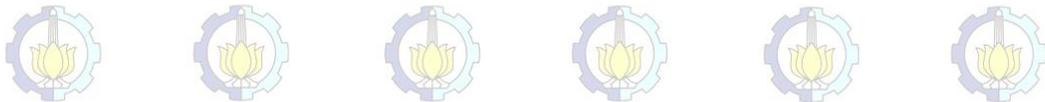
di mana:

n_{11} : kejadian gagal dari hasil amatan yang tepat diprediksi sebagai kejadian gagal

n_{12} : kejadian gagal dari hasil amatan yang tepat diprediksi sebagai kejadian sukses

n_{21} : kejadian sukses dari hasil amatan yang tepat diprediksi sebagai kejadian gagal

n_{22} : kejadian sukses dari hasil amatan yang tepat diprediksi sebagai kejadian sukses



BAB III
METODOLOGI PENELITIAN

BAB III

METODOLOGI PENELITIAN

3.1 Sumber Data

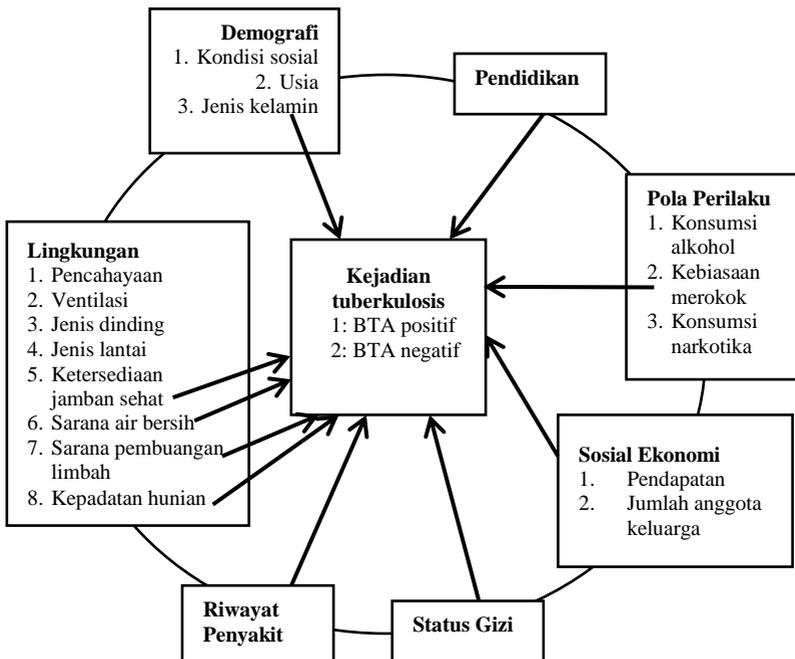
Data dalam penelitian ini merupakan data sekunder yang diambil dari penelitian berjudul “Analisis Ketahanan Pangan Rumah Tangga Terhadap Penderita Penyakit *Tuberculosis* dengan Pendekatan *Geographically Weighted Poisson Regression* (Studi Kasus Pantai Pesisir Surabaya). Penelitian ini disusun oleh Ir. Sri Pingit Wulandari, M.Si., Dra. Destri Susilaningrum, M. Si., dan Dr. I Nyoman Latra, M. Si. pada tahun 2016. Penelitian ini dilakukan untuk mendeskripsikan ketahanan pangan rumah tangga dan jumlah kasus penyakit *Tuberculosis* di tengah Kota Surabaya. Selain itu, melalui penelitian ini diperoleh model pengaruh ketahanan pangan rumah tangga terhadap jumlah kasus penyakit *Tuberculosis* di daerah tengah Kota Surabaya dengan pendekatan *Geographically Weighted Poisson Regression* (GWPR).

Jumlah sampel yang digunakan dalam penelitian ini sebanyak 124 sampel dengan metode *sampling* yang digunakan adalah *simple random sampling* (SRS) dengan taksiran parameter proporsional. Wilayah yang menjadi sampel penelitian merupakan wilayah tengah Kota Surabaya yang terdiri dari 10 kecamatan yaitu kecamatan Bubutan, Dukuh Pakis, Genteng, Gubeng, Sawahan, Simokerto, Sukomanunggal, Tambaksari, Tegalsari dan Wonokromo di mana unit sampel penelitian berupa rumah tangga dengan penderita tuberkulosis.

3.2 Kerangka Konsep Penelitian

Kejadian tuberkulosis dikategorikan menjadi dua yaitu BTA Positif dan BTA Negatif (Perhimpunan Dokter Paru Indonesia, 2006). Kejadian tuberkulosis disebabkan oleh infeksi bakteri *Mycobacterium tuberculosis* dan beberapa faktor lainnya. Beberapa faktor tersebut diantaranya demografi, lingkungan, status gizi, riwayat penyakit penyerta tuberkulosis, sosial ekonomi, pola perilaku penderita tuberkulosis, dan pendidikan.

Selanjutnya, faktor-faktor tersebut dijabarkan dalam variabel penelitian yaitu, jumlah anggota keluarga, tingkat pendidikan KRT (kepala rumah tangga), jenis kelamin, penghasilan KRT per bulan, riwayat penyakit, kebiasaan merokok, status gizi, kepemilikan toilet/WC/jamban, sumber air bersih, tempat pembuangan limbah dari kamar mandi/tempat cuci dapur, dan kepadatan hunian rumah. Kerangka konsep yang digunakan dalam penelitian ini ditunjukkan oleh Gambar 3.1 berikut.



Gambar 3.1 Kerangka Konsep Penelitian

3.3 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari satu variabel respon (Y) yaitu jenis penyakit tuberkulosis paru dan sebelas variabel prediktor (X). Berikut variabel penelitian yang digunakan.

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan	Kategori	Definisi Operasional
Y	Kejadian Tuberkulosis	1 = Tuberkulosis BTA Positif	Penderita tuberkulosis yang menunjukkan gejala aktif dan gejala pasif
		2 = Tuberkulosis BTA Negatif	
X ₁	Jumlah Anggota Keluarga	1 = > 7 orang	Jumlah orang yang terkumpul dan tinggal di suatu tempat di bawah satu atap dalam keadaan saling ketergantungan
		2 = ≤ 7 orang	
X ₂	Tingkat Pendidikan KRT	1 = Bersekolah	Status pendidikan yang diperoleh KRT
		2 = Tidak Bersekolah	
X ₃	Jenis Kelamin	1 = Laki-laki	Perbedaan bentuk, sifat dan fungsi biologi laki-laki dan perempuan
		2 = Perempuan	
X ₄	Penghasilan KRT per Bulan	1 = < Rp 600.000,00	Kemampuan ekonomi yang diterima atau diperoleh setiap bulannya
		2 = ≥ Rp 600.000,00	

Tabel 3.1 Variabel Penelitian (lanjutan)

Variabel	Keterangan	Kategori	Definisi Operasional
X ₅	Riwayat Penyakit	1 = Ada	Penyakit yang diderita sebelum terjangkit tuberkulosis
		2 = Tidak Ada	
X ₆	Kebiasaan Merokok	1 = Ya	Adanya kebiasaan merokok penderita tuberkulosis
		2 = Tidak	
X ₇	Status Gizi	1 = Gizi Kurang/Lebih 2 = Gizi Normal	Kecukupan gizi yang dikonsumsi
X ₈	Kepemilikan Toilet/WC/Jamban	1 = Tidak Punya/Umum	Kepemilikan tempat pembuangan hajat atau kotoran manusia
		2 = Sendiri	
X ₉	Sumber Air Bersih	1 = Sumur	Keberadaan sumber air bersih untuk kebutuhan sehari-hari
		2 = PDAM	
X ₁₀	Tempat Pembuangan Air Limbah dari Kamar Mandi/Tempat Cuci/Dapur	1 = Langsung ke Got/Sungai	Keberadaan tempat pembuangan limbah dari kamar mandi/ tempat cuci/ dapur
		2 = Di Tanah/Penampungan Tertutup di Pekarangan	
X ₁₁	Kepadatan Hunian Rumah	1 = Tidak Memenuhi Syarat (< 8m ² /orang) 2 = Memenuhi Syarat (≥ 8m ² /orang)	Sejumlah manusia dalam setiap unit ruangan

3.4 Langkah Analisis Penelitian

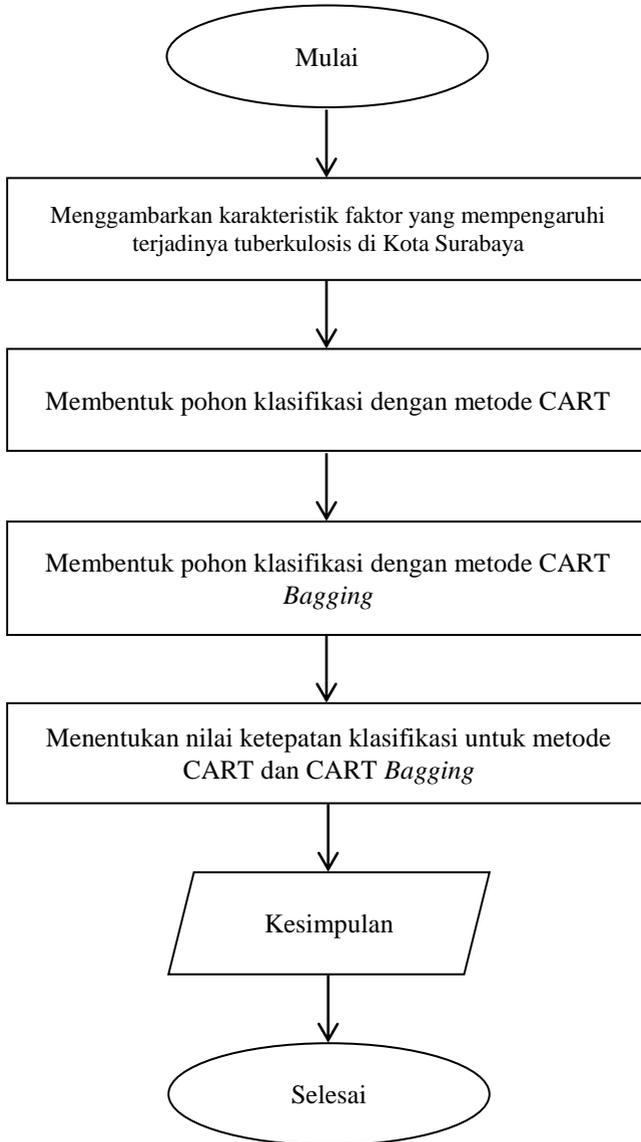
Langkah analisis yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Melakukan analisis statistika deskriptif untuk mengetahui karakteristik setiap variabel penelitian.
2. Membagi data menjadi dua bagian yaitu data *training* dan data *testing* dengan perbandingan 80:20.
3. Membangun pohon klasifikasi optimal dengan algoritma CART.
 - a. Membentuk pohon klasifikasi maksimal melalui pemilihan pemilah, penentuan simpul terminal dan penandaan label kelas.
 - b. Melakukan pemangkasan pohon klasifikasi maksimal.
 - c. Memilih pohon klasifikasi optimal dengan menggunakan penduga validasi silang lipat 10.
4. Membangun pohon agregat menggunakan prosedur *bagging*.
 - a. Mengambil sampel berukuran n dari data *training*
 - b. Melakukan langkah *a* sebanyak B kali. Nilai B yang digunakan yaitu 10, 25, 50, 100, 125, dan 150.
5. Menghitung nilai ketepatan klasifikasi untuk metode CART dan CART *bagging*.
6. Membandingkan hasil ketepatan klasifikasi.
7. Melakukan penarikan kesimpulan.

3.5 Diagram Alir Penelitian

Diagram alir dari penelitian ini dapat dilihat pada Gambar

3.1.



Gambar 3.2 Diagram Alir Penelitian



BAB IV
HASIL DAN PEMBAHASAN



BAB IV ANALISIS DAN PEMBAHASAN

4.1 Karakteristik Penderita Tuberkulosis dalam Rumah Tangga di Surabaya

Tuberkulosis merupakan penyebab kematian ketia setelah penyakit jantung dan saluran pernafasan. Penyakit ini disebabkan oleh bakteri *Mycobacterium tuberculosis* dan mudah tersebar melalui udara. Berdasarkan hasil pemeriksaan dahak, tuberkulosis dibedakan menjadi dua kategori yaitu BTA Positif dan BTA Negatif. Dalam penelitian ini dilakukan observasi terhadap 124 pengamatan dengan tuberkulosis baik yang termasuk ke dalam kategori BTA Positif maupun BTA Negatif.

Berdasarkan observasi yang telah dilakukan terdapat 63 penderita BTA Positif dan 61 penderita BTA Negatif. Beberapa variabel yang memengaruhi terjadinya tuberkulosis diantaranya jumlah anggota keluarga, tingkat pendidikan, jenis kelamin, penghasilan, riwayat penyakit kebiasaan merokok, status gizi, kepemilikan toilet/WC/jamban, sumber air bersih, tempat pembuangan air limbah dari kamar mandi/tempat cuci/ dapur dan kepadatan hunian rumah.



Gambar 4.1 Jumlah Kejadian Tuberkulosis Berdasarkan Masing-Masing Kategori

Gambar 4.1 menunjukkan jumlah kejadian tuberkulosis berdasarkan kategori BTA Positif dan BTA Negatif. Gambar 4.1 menunjukkan bahwa perbedaan jumlah antara kejadian

tuberkulosis dengan kategori BTA Positif dan BTA Negatif tidak terlalu besar. Namun jumlah kejadian tuberkulosis dengan kategori BTA Positif lebih banyak daripada jumlah kejadian tuberkulosis dengan kategori BTA Negatif.

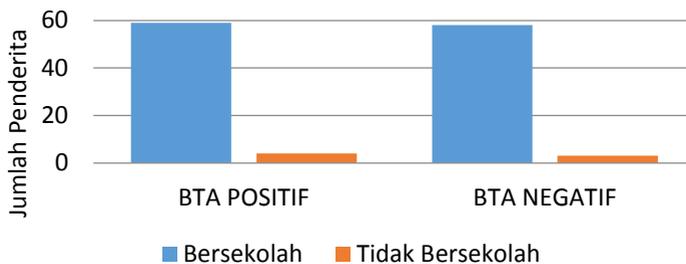
Salah satu variabel yang memengaruhi kejadian tuberkulosis di Surabaya adalah jumlah anggota keluarga. Jumlah anggota keluarga didefinisikan sebagai jumlah orang yang terkumpul dan tinggal di suatu tempat di bawah satu atap dalam keadaan saling ketergantungan. Variabel ini dapat memengaruhi terjadinya tuberkulosis dari berbagai aspek seperti ketersediaan makanan hingga aspek kesehatan. Gambar 4.2 menyajikan perbandingan jumlah anggota keluarga pada kedua kategori kejadian tuberkulosis.



Gambar 4.2 Perbandingan Jumlah Anggota Keluarga Berdasarkan Masing-masing Kategori

Baik pada kategori BTA Positif maupun BTA Negatif, lebih banyak penderita yang tinggal dengan anggota keluarga berjumlah kurang dari 7 orang. Pada kategori BTA Positif, terdapat 5 penderita yang tinggal dengan anggota keluarga yang berjumlah lebih dari 7 orang. Sedangkan 58 penderita lainnya pada kategori yang sama tinggal dengan anggota keluarga yang berjumlah kurang dari tujuh orang. Untuk kategori BTA Negatif, terdapat 6 penderita yang tinggal dengan anggota keluarga yang berjumlah lebih dari 7 orang. Sedangkan 55 penderita yang lain pada kategori yang sama tinggal dengan anggota keluarga yang berjumlah kurang dari 7 orang.

Variabel pendidikan yang diteliti adalah tingkat pendidikan kepala rumah tangga (KRT) karena KRT turut berperan serta dalam penyebaran informasi dan edukasi kepada anggota keluarga lainnya. Kategori yang digunakan dalam penelitian ini untuk membedakan jenjang pendidikan KRT adalah kategori 1 untuk bersekolah dan kategori 2 untuk tidak bersekolah. Karakteristik kejadian tuberkulosis berdasarkan pendidikan disajikan pada Gambar 4.3.

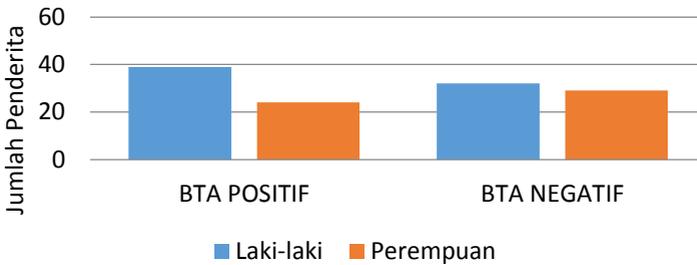


Gambar 4.3 Perbandingan Tingkat Pendidikan KRT Berdasarkan Masing-masing Kategori

Berdasarkan Gambar 4.3 diperoleh informasi bahwa baik pada kategori BTA Positif maupun BTA Negatif, lebih banyak KRT yang termasuk ke dalam kategori bersekolah daripada KRT yang termasuk ke dalam kategori tidak bersekolah. Pada kategori BTA Positif, 59 penderita tinggal dengan KRT bersekolah. Hanya 4 penderita pada kategori yang sama tinggal dengan KRT yang tidak bersekolah. Sedangkan pada kategori BTA Negatif, 58 penderita tinggal dengan KRT bersekolah dan 3 penderita sisanya tinggal dengan KRT yang tidak bersekolah.

Karakteristik kejadian tuberkulosis berdasarkan jenis kelamin dijelaskan melalui Gambar 4.4. Kejadian tuberkulosis dapat dijelaskan berdasarkan jenis kelamin terkait dengan kebiasaan merokok yang lebih banyak dilakukan laki-laki dibandingkan perempuan. Sehingga kemungkinan seorang laki-laki mengalami tuberkulosis lebih tinggi daripada seorang perempuan.

Berdasarkan Gambar 4.4 baik pada kategori BTA Positif maupun BTA Negatif, penderita berjenis kelamin laki-laki berjumlah lebih banyak dibandingkan dengan penderita berjenis kelamin perempuan. Pada kategori BTA Positif, penderita berjenis kelamin laki-laki berjumlah 39 penderita dan penderita berjenis kelamin perempuan berjumlah 24 penderita. Sedangkan pada kategori BTA Negatif, penderita berjenis kelamin laki-laki berjumlah 32 penderita dan penderita berjenis kelamin perempuan berjumlah 29 penderita.

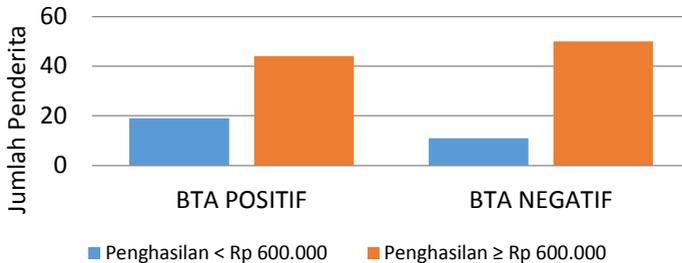


Gambar 4.4 Perbandingan Jenis Kelamin Berdasarkan Masing-masing Kategori

Kejadian tuberkulosis tidak hanya dikaitkan dengan faktor lingkungan tetapi juga dikaitkan dengan faktor sosial ekonomi seperti pendapatan. Penelitian ini menggunakan variabel penghasilan KRT per bulan dengan menggunakan batas minimum pendapatan di Surabaya yaitu sebesar Rp 600.000,00. Selanjutnya perbandingan mengenai penghasilan KRT per bulan disajikan pada Gambar 4.5.

Berdasarkan Gambar 4.5 diketahui bahwa pada kedua kategori lebih banyak KRT yang memiliki penghasilan lebih dari Rp 600.000,00 dibandingkan dengan KRT dengan penghasilan kurang dari Rp 600.000,00. Pada kategori BTA Positif, 19 penderita tinggal dengan KRT yang memiliki penghasilan kurang dari Rp 600.000,00 sedangkan 44 penderita yang lain tinggal dengan KRT yang memiliki penghasilan lebih dari Rp 600.000,00. Pada kategori BTA Negatif, jumlah penderita yang

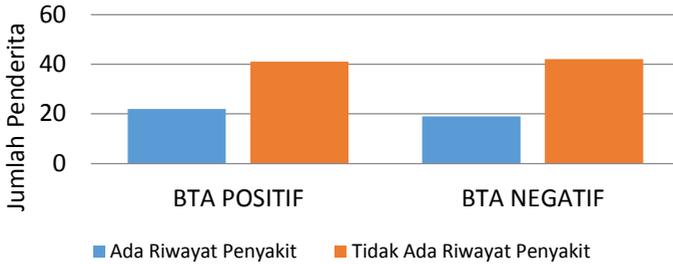
tinggal dengan KRT dengan penghasilan kurang dari Rp 600.000,00 lebih rendah jika dibandingkan dengan kategori BTA Positif yaitu sebanyak 11 penderita. Sisanya sebanyak 50 penderita tinggal dengan KRT dengan penghasilan lebih dari Rp 600.000,00 pada kategori BTA Negatif.



Gambar 4.5 Perbandingan Penghasilan KRT per Bulan Berdasarkan Masing-masing Kategori

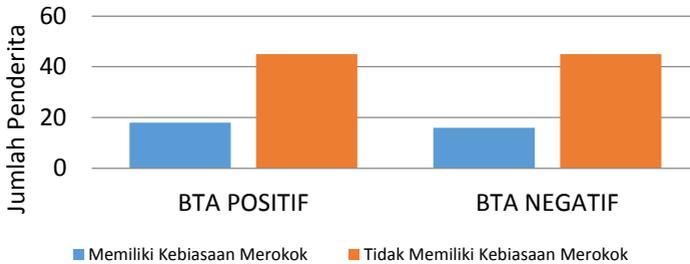
Riwayat penyakit seperti diabetes melitus (DM), HIV/AIDS, gagal ginjal, kanker, dan hepatitis akut dapat menyebabkan seseorang terjangkit penyakit tuberkulosis. Sehingga variabel riwayat penyakit turut diteliti untuk mengetahui seberapa besar pengaruh riwayat penyakit terhadap kejadian tuberkulosis khususnya di kota Surabaya. Berikut karakteristik kejadian tuberkulosis berdasarkan riwayat penyakit.

Pada Gambar 4.6 terlihat bahwa baik pada kategori BTA Positif dan BTA Negatif lebih banyak penderita yang tidak memiliki riwayat penyakit apapun sebelum menderita tuberkulosis. Pada kategori BTA Positif hanya 22 penderita dari 63 penderita yang memiliki riwayat penyakit. Sedangkan 41 penderita sisanya tidak memiliki riwayat penyakit sebelum menderita tuberkulosis. Pada kategori BTA Negatif, hanya 19 penderita yang memiliki riwayat penyakit. Jumlah ini lebih sedikit dibandingkan dengan penderita yang memiliki riwayat penyakit pada kategori BTA Positif. Sedangkan 42 penderita pada kategori BTA Negatif tidak memiliki riwayat penyakit sebelum menderita tuberkulosis.



Gambar 4.6 Perbandingan Riwayat Penyakit Berdasarkan Masing-masing Kategori

Kebiasaan merokok merupakan kebiasaan yang tidak baik bagi kesehatan manusia khususnya organ paru-paru. Berbagai jenis penyakit seperti kanker hingga komplikasi jantung disebabkan oleh kebiasaan merokok. Penyakit lain yang dapat dipicu oleh kebiasaan merokok adalah tuberkulosis. Oleh sebab itu, variabel penelitian kebiasaan merokok turut diamati dengan responden penderita tuberkulosis baik yang tergolong ke dalam kategori BTA Positif maupun BTA Negatif. Karakteristik kejadian tuberkulosis disajikan dalam Gambar 4.8 berikut.

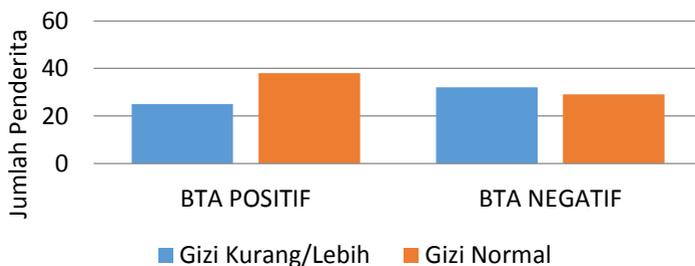


Gambar 4.7 Perbandingan Kebiasaan Merokok Berdasarkan Masing-masing Kategori

Gambar 4.7 menunjukkan bahwa lebih banyak penderita yang tidak memiliki kebiasaan merokok dibandingkan dengan

penderita yang memiliki kebiasaan merokok pada dua kategori tuberkulosis yaitu BTA Positif dan BTA Negatif. Pada kategori BTA Positif terdapat 18 penderita yang memiliki kebiasaan merokok dan 45 penderita sisanya tidak memiliki kebiasaan merokok. Sedangkan pada kategori BTA Negatif, 16 penderita memiliki kebiasaan merokok dan 45 penderita yang lain tidak memiliki kebiasaan merokok.

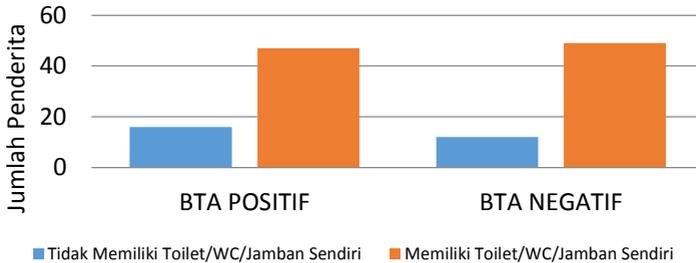
Gizi memegang peranan penting dalam menjaga kesehatan manusia. Gizi yang seimbang diperlukan untuk mencapai daya tahan tubuh yang baik dan menghindari berbagai jenis penyakit. Karakteristik kejadian tuberkulosis berdasarkan status gizi disajikan dalam Gambar 4.8. Pada gambar, terlihat bahwa 25 penderita pada kategori BTA Positif memiliki status gizi kurang/lebih sedangkan 38 penderita pada kategori yang sama memiliki status gizi normal. Pada kategori BTA Negatif, terdapat 32 penderita yang memiliki status gizi kurang/lebih dan 29 penderita sisanya memiliki status gizi normal. Berdasarkan Gambar 4.8 dapat diketahui pula bahwa pada kategori BTA Positif lebih banyak penderita yang berstatus gizi normal sedangkan pada kategori BTA Negatif lebih banyak penderita yang berstatus kurang/lebih.



Gambar 4.8 Perbandingan Status Gizi Berdasarkan Masing-masing Kategori

Masalah pembuangan kotoran manusia merupakan permasalahan yang penting karena kotoran manusia dapat menjadi sumber penyebaran berbagai penyakit. Oleh sebab itu

kepemilikan toilet/WC/Jamban sebagai tempat pembuangan kotoran manusia menjadi penting untuk diteliti. Kepemilikan toilet/WC/Jamban secara tidak langsung dapat menggambarkan bagaimana penyebaran penyakit. Karakteristik kejadian tuberkulosis berdasarkan kepemilikan toilet/WC/jamban disajikan pada Gambar 4.9.



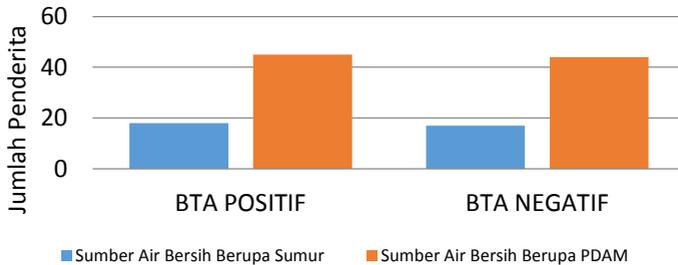
Gambar 4.9 Perbandingan Kepemilikan Toilet/WC/Jamban Berdasarkan Masing-masing Kategori

Berdasarkan Gambar 4.9, diperoleh informasi bahwa pada kategori BTA Positif terdapat 16 penderita yang tidak memiliki toilet/WC/jamban sendiri sedangkan 47 penderita sisanya sudah memiliki toilet/WC/Jamban sendiri. Pada kategori BTA Negatif diketahui bahwa 12 penderita tidak memiliki toilet/WC/Jamban sendiri dan 49 penderita sisanya sudah memiliki toilet/WC/Jamban sendiri. Melalui Gambar 4.9 dapat diketahui pula bahwa pada kedua kategori lebih banyak penderita yang sudah memiliki toilet/WC/Jamban sendiri.

Sumber air bersih merupakan hal yang penting untuk mendukung kehidupan yang baik dan layak bagi manusia. Dalam penelitian ini sumber air bersih dikaitkan dengan gambaran penyebaran penyakit. Kepemilikan sumber air bersih baik yang berasal dari sumur maupun PDAM dapat mendukung kehidupan yang baik dan layak.

Karakteristik kejadian tuberkulosis berdasarkan sumber air bersih dijelaskan melalui Gambar 4.10. Berdasarkan Gambar 4.10 dapat diketahui bahwa terdapat 18 penderita dengan sumber air

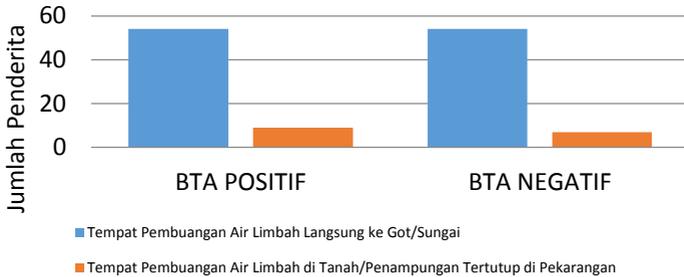
bersih berupa sumur dan 45 penderita dengan sumber air berupa PDAM pada kategori BTA Positif. Sedangkan pada kategori BTA Negatif terdapat 17 penderita yang memiliki sumber air bersih berupa sumur dan 44 penderita sisanya memiliki sumber air bersih berupa PDAM.



Gambar 4.10 Perbandingan Sumber Air Bersih Berdasarkan Masing-masing Kategori

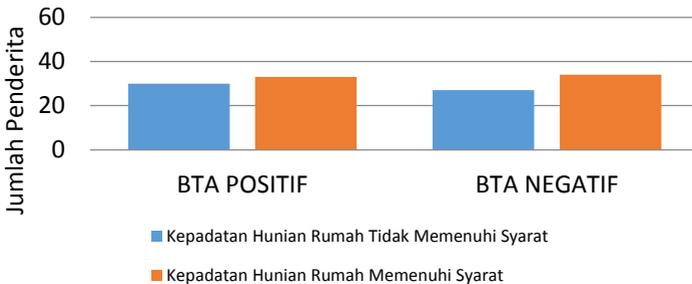
Tidak hanya kepemilikan toilet/WC/Jamban dan sumber air bersih, keberadaan tempat pembuangan air limbah dari kamar mandi/tempat cuci/dapur juga berkaitan dengan penyebaran penyakit dan kesehatan lingkungan tinggal manusia. Keberadaan tempat pembuangan air limbah dari kamar mandi/tempat cuci/dapur jika tidak tepat dapat menjadi tempat berkembangnya berbagai penyakit.

Karakteristik kejadian tuberkulosis berdasarkan tempat pembuangan air limbah dari kamar mandi/tempat cuci/dapur disajikan pada Gambar 4.11. Berdasarkan Gambar 4.11 dapat diketahui bahwa 54 penderita BTA Positif dan 54 penderita BTA Negatif membuang air limbah dari kamar mandi/tempat cuci/dapur langsung ke got atau sungai. Sedangkan 7 penderita BTA Positif dan 9 penderita BTA Negatif membuang air limbah dari kamar mandi/tempat cuci/dapur ke tanah atau penampungan tertutup di perkarangan.



Gambar 4.11 Perbandingan Tempat Pembuangan Air Limbah dari Kamar Mandi/Tempat Cuci/Dapur Berdasarkan Masing-masing Kategori

Kepadatan hunian rumah yang dapat dijelaskan melalui perbandingan jumlah anggota keluarga yang tinggal dengan luas lantai merupakan salah satu faktor yang dapat mempengaruhi kesehatan manusia. Jika terlalu padat, kesehatan lingkungan dan sanitasi di dalam rumah menjadi tidak terlalu baik sehingga dapat memengaruhi kesehatan manusia yang tinggal di dalam rumah. Karakteristik kejadian tuberkulosis berdasarkan kepadatan hunian rumah disajikan dalam Gambar 4.12.



Gambar 4.12 Perbandingan Kepadatan Hunian Rumah Berdasarkan Masing-masing Kategori

Kepadatan hunian rumah dikatakan memenuhi syarat apabila telah memenuhi KEPMENKES RI NO. 829/MENKES/SK/VII/1999 tentang Persyaratan Kesehatan

Perumahan. Aturan ini menetapkan bahwa rumah dikatakan tidak padat penghuninya bila luas lantai kamar tidur $\geq 8\text{m}^2/\text{orang}$. Berdasarkan Gambar 4.12, sebanyak 30 penderita yang termasuk ke dalam kategori BTA Positif tinggal dengan kepadatan hunian rumah yang tidak memenuhi syarat. Sedangkan 33 penderita lainnya pada kategori yang sama tinggal dengan kepadatan hunian rumah yang sudah memenuhi syarat. Pada kategori BTA Negatif, 27 penderita tinggal dengan kepadatan hunian rumah yang tidak memenuhi syarat. Sedangkan 34 penderita lainnya tinggal dengan kepadatan hunian rumah yang sudah memenuhi syarat. Gambar 4.12 menunjukkan pula bahwa pada kedua kategori lebih banyak penderita yang tinggal dengan kepadatan hunian rumah yang sudah memenuhi syarat.

4.2 Analisis CART untuk Pengelompokan Penderita Tuberkulosis dalam Rumah Tangga di Surabaya

Analisis CART dilakukan untuk mengetahui ketepatan klasifikasi kejadian tuberkulosis di Surabaya berdasarkan dua kategori yaitu BTA Positif dan BTA Negatif. Analisis ini dilakukan dengan menggunakan metode *cross validation V-folds estimate* karena data yang digunakan berjumlah kurang dari 3000 pengamatan. Pada analisis ini diterapkan 10 *folds* terhadap 99 data *training* sehingga data *training* dibagi menjadi 10 bagian untuk membentuk pohon klasifikasi. Sebelum dilakukan pembentukan pohon klasifikasi dilakukan perhitungan kemungkinan pemilah untuk setiap variabel prediktor.

4.2.1 Kemungkinan Pemilah untuk Setiap Variabel Prediktor

Untuk membentuk pohon klasifikasi diperlukan variabel-variabel yang berperan sebagai pemilah. Banyaknya kemungkinan pemilah yang akan berperan dalam pembentukan pohon klasifikasi dapat diketahui berdasarkan skala data variabel prediktor. Jika variabel prediktor berskala kontinu dengan ruang sampel berukuran n dan terdapat n amatan yang berbeda, maka kemungkinan pemilah yang terbentuk sebanyak $n-1$ jenis

pemilahan yang berbeda. Sedangkan jika variabel prediktor berskala kategorik ordinal bertaraf L maka kemungkinan pemilah yang terbentuk sebanyak $L-1$ jenis pemilahan. Banyaknya kemungkinan pemilah untuk membentuk pohon klasifikasi berdasarkan kejadian tuberkulosis di Surabaya disajikan dalam Tabel 4.1

Tabel 4.1 Banyaknya Kemungkinan Pemilah Variabel Prediktor

Variabel	Nama Variabel	Skala Data	Jumlah Kategori	Kemungkinan Pemilah
X_1	Jumlah Anggota Keluarga	Ordinal	2	$2 - 1 = 1$ pemilah
X_2	Tingkat Pendidikan KRT	Ordinal	2	$2 - 1 = 1$ pemilah
X_3	Jenis Kelamin	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_4	Penghasilan KRT per Bulan	Ordinal	2	$2 - 1 = 1$ pemilah
X_5	Riwayat Penyakit	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_6	Kebiasaan Merokok	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_7	Status Gizi	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_8	Kepemilikan Toilet/WC/Jamban	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_9	Sumber Air Bersih	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_{10}	Tempat Pembuangan Air Limbah dari Kamar Mandi/Tempat Cuci/Dapur	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_{11}	Kepadatan Hunian Rumah	Nominal	2	$2^{2-1} - 1 = 1$ pemilah

Berdasarkan Tabel 4.1, variabel jumlah anggota keluarga, tingkat pendidikan KRT, jenis kelamin, penghasilan KRT per bulan, riwayat penyakit, kebiasaan merokok, status gizi, kepemilikan toilet/WC/jamban, sumber air bersih, tempat pembuangan air limbah dari kamar mandi/tempa cuci/dapur, dan kepadatan hunian rumah memiliki satu kemungkinan pemilah saja karena variabel prediktor berskala data kategorik dengan dua taraf. Setelah dilakukan perhitungan kemungkinan pemilah untuk membentuk pohon klasifikasi, tahap selanjutnya yaitu pemilihan pemilah. Tahap ini dilakukan dengan menggunakan indeks Gini.

4.2.2 Pembentukan Pohon Klasifikasi Maksimal

Pohon klasifikasi maksimal merupakan pohon yang memiliki jumlah simpul terbanyak dengan kedalaman yang tinggi. Pada pohon maksimal diperoleh nilai dugaan pengganti yang kecil namun pohon klasifikasi berukuran terlalu besar dan kompleks. Pembentukan pohon klasifikasi maksimal diawali dengan tahapan pemilihan pemilah atau *classifier*. Pemilah yang terpilih selanjutnya akan digunakan sebagai simpul induk. Simpul induk pada pohon klasifikasi maksimal biasanya dipilih berdasarkan skor variabel prediktor tertinggi. Besarnya kontribusi masing-masing variabel prediktor terhadap pembentukan pohon klasifikasi maksimal ditampilkan dalam Tabel 4.4 berikut.

Tabel 4.2 Variabel Terpenting dari Pohon Klasifikasi Maksimal

Variabel	Nama Variabel	Skor
X ₄	Penghasilan KRT per Bulan	100,000
X ₂	Tingkat Pendidikan KRT	87,114
X ₁₀	Tempat Pembuangan Air Limbah dari Kamar Mandi/Tempat Cuci/Dapur	77,022
X ₉	Sumber Air Bersih	62,547
X ₃	Jenis Kelamin	49,227
X ₁₁	Kepadatan Hunian Rumah	45,694
X ₅	Riwayat Penyakit	36,477
X ₇	Status Gizi	36,183
X ₆	Kebiasaan Merokok	30,309
X ₈	Kepemilikan Toilet/WC/Jamban	24,644

Variabel	Nama Variabel	Skor
X ₁	Jumlah Anggota Keluarga	5,168

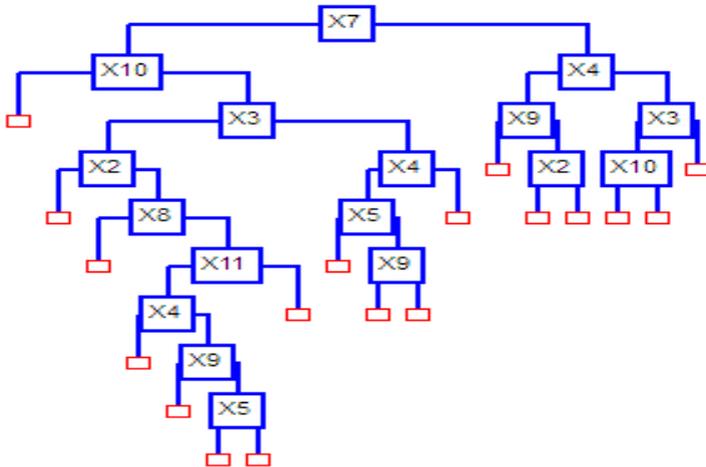
Dalam Tabel 4.2 disajikan skor variabel prediktor yang membentuk pohon klasifikasi maksimal. Berdasarkan Tabel 4.2 dapat diketahui bahwa variabel penghasilan KRT per bulan memiliki skor variabel prediktor tertinggi yaitu sebesar 100,00. Variabel tingkat pendidikan KRT memiliki skor sebesar 87,114, variabel tempat pembuangan air limbah dari kamar mandi/tempat cuci/dapur memiliki skor sebesar 77,022, variabel sumber air bersih memiliki skor sebesar 62,547, variabel jenis kelamin memberikan skor sebesar 49,227, variabel kepadatan hunian rumah memiliki skor sebesar 45,694, variabel riwayat penyakit memiliki skor sebesar 36,477, variabel status gizi memiliki skor sebesar 36,183, variabel kebiasaan merokok memiliki skor sebesar 30,309, variabel kepemilikan toilet/WC/jamban memiliki skor sebesar 26,644 dan variabel jumlah anggota keluarga memiliki skor paling rendah yaitu 5,168.

Jika ditinjau dari skor variabel prediktor, variabel prediktor dengan skor tertinggi memberikan kontribusi yang besar dalam pembentukan pohon klasifikasi maksimal. Variabel prediktor dengan skor tertinggi juga dapat dikatakan pula sebagai variabel yang mampu menurunkan heterogenitas pada pohon klasifikasi paling besar dibandingkan dengan variabel prediktor lainnya. Sedangkan variabel prediktor dengan skor terendah memberikan kontribusi kecil dalam pembentukan pohon klasifikasi maksimal.

Setelah dilakukan perhitungan skor variabel prediktor dibentuk pohon klasifikasi maksimal berdasarkan skor variabel prediktor tersebut. Selanjutnya dilakukan proses pemilahan pemilah hingga terbentuk pohon klasifikasi maksimal. Pohon klasifikasi maksimal yang terbentuk dari data kejadian tuberkulosis di Surabaya terdiri dari 10 kedalaman dan 18 simpul terminal (*terminal node*).

Simpul induk dari pohon klasifikasi maksimal yang terbentuk adalah variabel status gizi. Variabel status gizi memiliki

skor variabel prediktor sebesar 36,183 di mana skor ini lebih rendah dibandingkan dengan skor yang dimiliki oleh variabel penghasilan KRT per bulan. Biasanya variabel prediktor dengan skor tertinggi dipilih sebagai pemilah utama atau simpul induk (*parent node*). Namun pada data kejadian tuberkulosis di Surabaya pemilah utama atau simpul induk (*parent node*) bukan merupakan variabel prediktor dengan skor tertinggi. Sebenarnya perbedaan skor variabel prediktor dan pemilah utama atau simpul induk (*parent node*) pada pohon klasifikasi dapat terjadi pada data apapun. Hal ini disebabkan oleh adanya persamaan informasi antara variabel prediktor dengan skor tertinggi dan variabel yang dipilih sebagai pemilah utama atau simpul induk (*parent node*).



Gambar 4.13 Pemilah Pohon Klasifikasi Maksimal

Menurut Steinberg dan Golovnya (2007), variabel dalam analisis CART memiliki dua peran penting yaitu sebagai pemilah utama yang dipilah menjadi *node* dan sebagai pemilah pengganti (pemilah yang digunakan ketika pemilah utama mengalami *missing*). Variabel yang muncul dengan skor tertinggi tetapi tidak dipilih sebagai pemilah utama memiliki kemungkinan korelasi tinggi dan kesamaan informasi dengan variabel yang dipilih

menjadi pemilah utama. Selain itu skor variabel atau *variabel importance* lebih banyak menjelaskan tentang kontribusi suatu variabel terhadap pembentukan pohon klasifikasi. Selanjutnya pemilah pohon klasifikasi maksimal disajikan dalam Gambar 4.13.

4.2.3 Pemangkasan (*Pruning*) Pohon Klasifikasi Maksimal

Tahap selanjutnya setelah diperoleh pohon klasifikasi maksimal adalah melihat apakah pohon klasifikasi maksimal perlu dilakukan pemangkasan (*pruning*) atau tidak. Tujuan dilakukannya pemangkasan adalah untuk menghindari adanya kasus *overfitting* maupun kasus *underfitting*. Kasus *overfitting* terjadi apabila nilai dugaan yang dihasilkan pohon klasifikasi maksimal melebihi nilai sebenarnya. Sedangkan kasus *underfitting* terjadi apabila nilai dugaan yang dihasilkan pohon klasifikasi maksimal kurang dari keadaan sebenarnya. Ukuran yang digunakan untuk mengetahui apakah pohon klasifikasi maksimal perlu dilakukan pemangkasan atau tidak adalah nilai *cost complexity measure* yang dihasilkan.

Tabel 4.3. menyajikan urutan pembentukan pohon klasifikasi. Kolom *tree number* menjelaskan urutan pembentukan pohon klasifikasi. Kolom *terminal nodes* menyajikan banyaknya simpul terminal yang dihasilkan setiap urutan pembentukan pohon klasifikasi. Kolom *cross-validated relative cost* menyajikan rentang biaya kesalahan klasifikasi pohon saat penggunaan penduga validasi silang lipat V. Kolom *resubstitution relative cost* merupakan kolom yang menyajikan biaya kesalahan klasifikasi setiap pohon. Sedangkan kolom *complexity* merupakan kolom yang menyajikan biaya penambahan simpul terminal setiap pohon. Untuk mengetahui *cost complexity measure* yang bernilai minimum dapat dilakukan perhitungan menggunakan rumus (2.5).

Tabel 4.3 Urutan Pembentukan Pohon Klasifikasi (*Tree Sequence*)

<i>Tree Number</i>	<i>Terminal Nodes</i>	<i>Cross-Validated Relative Cost</i>	<i>Resubstitution Relative Cost</i>	<i>Complexity</i>
1	18	$0,722 \pm 0,097$	0,3529	0,000
2	12	$0,778 \pm 0,098$	0,389	0,003
3	10	$0,856 \pm 0,100$	0,422	0,008
4	8	$0,859 \pm 0,099$	0,478	0,014
5	5	$0,867 \pm 0,098$	0,570	0,015
6	3	$0,744 \pm 0,089$	0,667	0,024
7	2	$0,856 \pm 0,100$	0,774	0,054
8	1	$1,000 \pm 0,000$	1,000	0,113

Tabel 4.4 menyajikan hasil perhitungan nilai *cost complexity measure* untuk setiap urutan pembentukan pohon klasifikasi. Berdasarkan hasil perhitungan dalam Tabel 4.4 dapat diketahui bahwa nilai *cost complexity measure* yang paling minimum sebesar 0,3529 dengan nilai *complexity* 0,000 dan nilai *resubstitution relative cost* sebesar 0,3529. Selanjutnya nilai ini dapat disebut sebagai *cost complexity minimum*. Pada nilai ini jumlah simpul terminal pohon sebanyak 18 simpul terminal dengan kedalaman 10 dan rentang rentang biaya kesalahan klasifikasi pohon saat penggunaan penduga validasi silang lipat V sebesar 0,625 hingga 0,819. Sehingga dapat dikatakan bahwa pada pohon klasifikasi yang sudah terbentuk tidak perlu dilakukan pemangkasan.

Tabel 4.4 Hasil Perhitungan *Cost Complexity Measure*

<i>Tree Number</i>	<i>Terminal Nodes</i>	<i>Resubstitution Relative Cost</i>	<i>Complexity</i>	<i>Cost Complexity Measure</i>
1**	18	0,3529	0,000	0,3529
2	12	0,389	0,003	0,425
3	10	0,422	0,008	0,502
4	8	0,478	0,014	0,59
5	5	0,570	0,015	0,645
6	3	0,667	0,024	0,739
7	2	0,774	0,054	0,882

<i>Tree Number</i>	<i>Terminal Nodes</i>	<i>Resubstitution Relative Cost</i>	<i>Complexity</i>	<i>Cost Complexity Measure</i>
8	1	1,000	0,113	1,113

** batas pemangkasan pohon klasifikasi

4.2.4 Pemilihan Pohon Klasifikasi Optimal

Pohon klasifikasi optimal dipilih berdasarkan dengan penduga pengganti yang digunakan. Penggunaan penduga pengganti pada pemilihan pohon klasifikasi optimal didasarkan pada jumlah sampel yang digunakan untuk membangun pohon klasifikasi. Pada penelitian ini digunakan penduga validasi silang lipat V dengan *10-folds*. Pemilihan pohon klasifikasi optimal dengan menggunakan validasi silang lipat V dilakukan dengan memilih nilai *cost complexity measure* paling minimum dengan nilai $R(T)$ yang telah digantikan dengan nilai $R^{CV}(T)$. Tabel 4.5 menyajikan hasil perhitungan *cost complexity measure* untuk pemilihan pohon klasifikasi optimal.

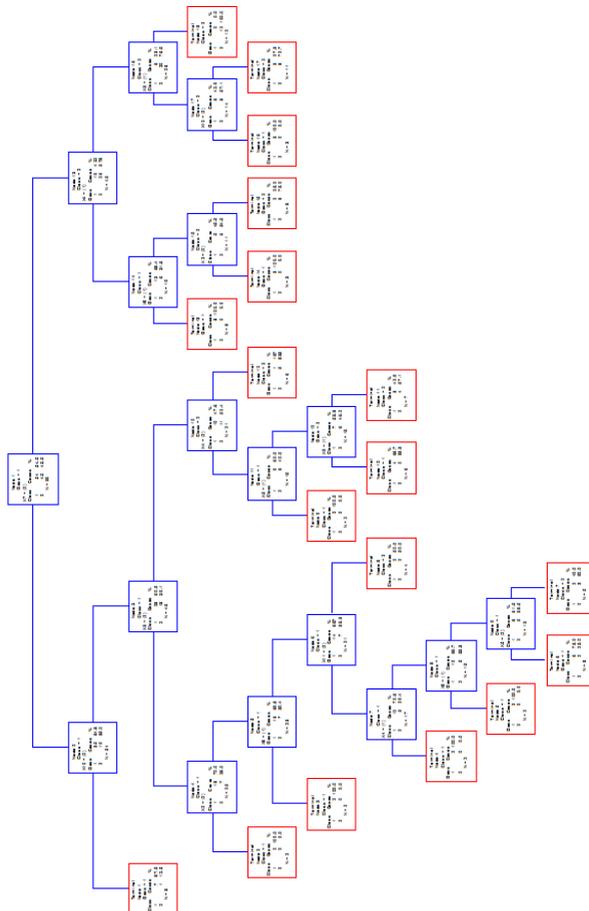
Tabel 4.5 Hasil Perhitungan *Cost Complexity Measure (CV)*

<i>Tree Number</i>	<i>Terminal Nodes</i>	<i>Cross-Validated Resubstitution Cost</i>	<i>Complexity</i>	<i>Cost Complexity Measure (CV)</i>
1**	18	-0,194	0,000	-0,194
2	12	-0,196	0,003	-0,16
3	10	-0,2	0,008	-0,12
4	8	-0,198	0,014	-0,086
5	5	-0,196	0,015	-0,121
6	3	-0,178	0,024	-0,106
7	2	-0,2	0,054	-0,092
8	1	0	0,113	0,113

** pohon klasifikasi optimal

Kolom *cross-validated resubstitution cost* diperoleh melalui selisih batas atas dan batas bawah kolom yang sama pada Tabel 4.3. Selanjutnya kolom *cost complexity measure (CV)* diperoleh melalui penerapan rumus (2.5). Berdasarkan Tabel 4.4 diketahui bahwa nilai *cost complexity minimum* berada pada pohon

klasifikasi dengan 18 simpul terminal. Sehingga pohon klasifikasi maksimal merupakan pohon klasifikasi optimal pada kasus penderita tuberkulosis dalam rumah tangga di Surabaya. Gambaran pohon klasifikasi optimal untuk klasifikasi kejadian tuberkulosis disajikan pada Gambar 4.14.



Gambar 4.14 Topologi Pohon Klasifikasi Optimal

Berdasarkan 99 pengamatan yang dilabelkan menjadi kategori atau kelas 1 (BTA Positif) dipilah menjadi dua simpul anak yaitu simpul kiri dan simpul kanan berdasarkan variabel status gizi penderita tuberkulosis. Apabila penderita tuberkulosis termasuk ke dalam kelompok status gizi normal maka akan dipilah menjadi simpul kiri dengan label kelas 1 atau BTA Positif dengan jumlah amatan sebanyak 54. Apabila penderita tuberkulosis termasuk ke dalam kelompok status gizi kurang/lebih akan dipilah menjadi simpul kanan dengan label kelas 2 atau BTA Negatif di mana jumlah amatan yang termasuk ke dalam kelas ini sebanyak 45 amatan.

Sebanyak 54 amatan pada label kelas 1 atau BTA Positif dipilah lagi menjadi dua simpul yaitu simpul kiri dan simpul kanan. Jika tempat pembuangan air limbah dari kamar mandi/tempat cuci/ dapur penderita tuberkulosis di tanah/penampungan tertutup di pekarangan maka akan dipilah menjadi simpul kiri. Sedangkan jika tempat pembuangan air limbah dari kamar mandi/tempat cuci/dapur penderita tuberkulosis langsung ke got/sungai maka akan dipilah menjadi simpul kanan. Pada pemilahan ini, simpul kiri yang memiliki 8 amatan menjadi simpul terminal pertama karena tidak dapat dipilah lagi. Sedangkan simpul kanan yang memiliki 46 amatan mengalami proses pemilahan selanjutnya. Proses pemilahan ini berlangsung secara terus menerus hingga simpul yang terbentuk tidak dapat dipilah lagi dan menghasilkan simpul terminal.

Pada data kejadian tuberkulosis di Surabaya diperoleh pohon klasifikasi maksimal dan pohon klasifikasi optimal yang sama dengan 18 simpul terminal dan 10 kedalaman. Setiap simpul terminal ini selanjutnya diberi label untuk mengetahui masing-masing kategori atau kelas pada variabel respon. Simpul terminal yang diklasifikasikan sebagai kategori 1 atau BTA Positif adalah simpul terminal 1, 2, 3, 4, 5, 6, 9, 10, 13, 14, dan 16. Sedangkan simpul terminal yang diklasifikasikan sebagai kategori 2 atau BTA Negatif adalah simpul terminal 7, 8, 11, 15, 17 dan 18.

Tabel 4.6 Pelabelan Kelas Simpul Terminal

Kelas	Simpul Terminal	Persentase (%)
BTA Positif (11 Simpul Terminal)	1	87,5
	2	100
	3	100
	4	100
	5	100
	6	75,0
	9	100
	10	66,7
	13	100
	14	100
	16	100
BTA Negatif (7 Simpul Terminal)	7	60,0
	8	50,00
	11	57,1
	12	83,3
	15	75,00
	17	72,7
	18	100

Setiap simpul terminal yang dihasilkan pohon klasifikasi optimal memberikan karakteristik berbeda berdasarkan kelas variabel respon yang telah diprediksi. Untuk mengetahui bagaimana karakteristik yang dihasilkan dilakukan pelabelan label kelas. Selanjutnya pelabelan kelas untuk setiap simpul terminal yang dihasilkan pohon klasifikasi optimal disajikan dalam Tabel 4.6. Jika ditinjau dari Tabel 4.6 dapat diketahui bahwa jumlah simpul terminal untuk kelas 1 atau kategori BTA Positif sebanyak 11 simpul terminal sedangkan jumlah simpul terminal untuk kelas 2 atau kategori BTA Negatif berjumlah 7 simpul terminal.

Untuk mengetahui karakteristik masing-masing kelas pada variabel respon dilakukan penelusuran pohon klasifikasi optimal yang telah terbentuk seperti pada Gambar 4.14. Karakteristik yang kuat dapat diketahui dengan melihat simpul terminal dengan persentase kelas terbesar pada pelabelan label kelas. Karakteristik

kelas 1 atau BTA Positif dapat diketahui dengan menelusuri simpul terminal 2, sedangkan karakteristik kelas 2 atau BTA Negatif dapat diketahui dengan menelusuri simpul terminal 18.

Jika ditinjau berdasarkan simpul terminal dengan persentase tertinggi diperoleh beberapa karakteristik untuk setiap kategori penderita tuberkulosis. Karakteristik penderita BTA Positif berdasarkan simpul terminal 2 digambarkan sebagai penderita dengan status gizi normal, memiliki tempat pembuangan air limbah dari kamar mandi/ tempat cuci/ dapur langsung ke got/sungai, berjenis kelamin perempuan dan tinggal dengan KRT yang tidak bersekolah. Berdasarkan simpul terminal 3, penderita BTA Positif memiliki kecenderungan berstatus gizi normal, memiliki tempat pembuangan air limbah dari kamar mandi/ tempat cuci/ dapur langsung ke got/sungai, berjenis kelamin perempuan, tinggal dengan KRT yang bersekolah, dan status kepemilikan toilet/ WC/ jambannya umum. Kecenderungan penderita BTA Positif menurut simpul terminal 4 berstatus gizi normal, memiliki tempat pembuangan air limbah dari kamar mandi/ tempat cuci/ dapur langsung ke got/sungai, berjenis kelamin perempuan, tinggal dengan KRT yang bersekolah, status kepemilikan toilet/ WC/ jamban pribadi, tinggal dengan kepadatan hunian yang memenuhi syarat, dan tinggal dengan KRT berpenghasilan kurang dari Rp 600.000 per bulan. Karakteristik penderita BTA Positif berdasarkan simpul terminal 5 digambarkan sebagai penderita dengan status gizi normal, memiliki tempat pembuangan air limbah dari kamar mandi/ tempat cuci/ dapur langsung ke got/sungai, berjenis kelamin perempuan, tinggal dengan KRT yang bersekolah, status kepemilikan toilet/ WC/ jamban pribadi, tinggal dengan kepadatan hunian yang memenuhi syarat, tinggal dengan KRT berpenghasilan lebih dari Rp 600.000 per bulan, dan menggunakan sumber air bersih berupa sumur. Selanjutnya menurut simpul terminal 9, penderita BTA Positif memiliki kecenderungan berupa penderita dengan status gizi normal, memiliki riwayat penyakit, tinggal dengan KRT berpenghasilan

lebih dari Rp 600.000 per bulan, berjenis kelamin laki-laki, dan memiliki tempat pembuangan air limbah dari kamar mandi/ tempat cuci/ dapur langsung ke got/sungai. Sedangkan menurut simpul terminal 13, penderita BTA Positif memiliki kecenderungan berstatus gizi kurang/lebih, tinggal dengan KRT berpenghasilan kurang dari Rp 600.000 per bulan, dan menggunakan sumber air bersih berupa sumur. Karakteristik penderita BTA Positif berdasarkan simpul terminal 14 digambarkan sebagai penderita dengan status gizi kurang/lebih, tinggal dengan KRT berpenghasilan kurang dari Rp 600.000 per bulan, menggunakan sumber air bersih berupa PDAM, dan tinggal dengan KRT yang tidak bersekolah. Sedangkan menurut simpul terminal 16, karakteristik penderita BTA Positif adalah penderita dengan status gizi kurang/lebih, tinggal dengan KRT berpenghasilan lebih dari Rp 600.000 per bulan, berjenis kelamin laki-laki dan memiliki tempat pembuangan air limbah dari kamar mandi/ tempat cuci/ dapur di tanah/ penampungan tertutup di pekarangan.

Selanjutnya karakteristik penderita BTA Negatif menurut simpul terminal 18 memiliki kecenderungan berupa penderita berstatus gizi kurang/lebih, tinggal dengan KRT berpenghasilan lebih dari Rp 600.000 per bulan dan berjenis kelamin perempuan.

4.2.5 Ketepatan Klasifikasi Analisis CART Penderita Tuberkulosis dalam Rumah Tangga di Surabaya

Ketepatan klasifikasi diperlukan untuk mengetahui kebaikan prediksi klasifikasi terutama kebaikan prediksi pohon klasifikasi optimal. Sebelumnya diperoleh pohon klasifikasi optimal kejadian tuberkulosis di Surabaya berdasarkan dua kategori yaitu BTA Positif dan BTA Negatif. Selanjutnya perlu dilakukan perhitungan ketepatan klasifikasi terhadap data *training* maupun data *testing* sehingga diketahui kebaikan prediksi klasifikasi serta kesalahan klasifikasi.

Pengukuran ketepatan klasifikasi dalam penelitian ini menggunakan nilai 1-APER yang menjelaskan proporsi pengamatan atau observasi yang diprediksikan secara benar.

Sebelum dilakukan pengukuran ketepatan klasifikasi dengan 1-APER, disajikan hasil klasifikasi kejadian tuberkulosis di Surabaya pada data *training* yang berjumlah 99 pengamatan dalam Tabel 4.7.

Tabel 4.7 Hasil Klasifikasi Analisis CART untuk Data *Training*

Kelas Observasi Y	Kelas Prediksi Y		Total
	1	2	
1	33	21	54
2	15	30	45
Total	48	51	99

Berdasarkan Tabel 4.7 dapat diketahui bahwa 33 pengamatan yang termasuk ke dalam kelas 1 atau BTA Positif diprediksikan secara tepat ke dalam kelas 1 dan pada kelas 2 atau BTA Negatif terdapat 30 pengamatan yang diprediksikan secara tepat. Namun terdapat 21 pengamatan yang termasuk ke dalam kelas 1 atau BTA Positif yang diprediksikan secara tidak tepat ke dalam kelas 2 atau BTA Negatif dan 15 pengamatan yang termasuk ke dalam kelas 2 atau BTA Negatif yang diprediksikan secara tidak tepat ke dalam kelas 1 atau BTA Positif.

Selanjutnya untuk mengetahui ketepatan klasifikasi dengan menggunakan nilai APER dilakukan perhitungan berikut.

$$1 - APER = \frac{33 + 30}{99} = 0,636$$

Berdasarkan hasil perhitungan yang telah dilakukan dapat dikatakan bahwa ketepatan klasifikasi pohon klasifikasi optimal penderita tuberkulosis dalam rumah tangga di Surabaya sebesar 0,636 atau 63,6%.

Setelah dilakukan klasifikasi dan pengukuran ketepatan klasifikasi terhadap data *training* dilakukan pula klasifikasi dan pengukuran ketepatan klasifikasi terhadap data *testing*. Jumlah data *testing* yang digunakan sebanyak 25 pengamatan.

Tabel 4.8 Hasil Klasifikasi Analisis CART untuk Data *Testing*

Kelas Observasi Y	Kelas Prediksi Y		Total
	1	2	
1	9	2	11
2	4	10	14
Total	13	12	25

Jika ditinjau dari Tabel 4.8 dapat diketahui bahwa terdapat 9 pengamatan yang diprediksikan secara tepat ke dalam kelas 1 atau BTA Positif dan 10 pengamatan yang diprediksikan secara tepat ke dalam kelas 2 atau BTA Negatif. Namun terdapat 2 pengamatan yang diprediksikan secara tidak tepat ke dalam kelas 2 atau BTA Negatif dan 4 pengamatan yang diprediksikan secara tidak tepat ke dalam kelas 1 atau BTA Positif.

Ukuran ketepatan klasifikasi untuk data *testing* kejadian tuberkulosis di Surabaya diperoleh melalui perhitungan berikut.

$$1 - APER = \frac{9 + 10}{25} = 0,76$$

Berdasarkan perhitungan yang telah dilakukan diperoleh tingkat ketepatan klasifikasi pohon klasifikasi optimal sebesar 0,76 atau 76%. Dari perhitungan ini diketahui pula bahwa ketepatan klasifikasi antara data *training* dan data *testing* memiliki perbedaan sebesar 0,124 atau 12,4%.

4.3 Ketepatan Klasifikasi CART *Bagging* untuk Klasifikasi Penderita Tuberkulosis dalam Rumah Tangga di Surabaya

Penerapan *bagging* pada CART diharapkan dapat memperbaiki stabilitas dan kekuatan prediksi pada pohon klasifikasi maupun pohon regresi yang dihasilkan dengan cara mereduksi variansi variabel prediktor. Metode *bagging* menggunakan konsep replikasi *bootstrap sampling*. Berbagai variasi replikasi disarankan untuk mengetahui peningkatan akurasi ketepatan klasifikasi. Untuk penelitian ini digunakan variasi replikasi diantaranya 10, 25, 50, 100, 125, dan 150.

Replikasi yang digunakan diterapkan pada data *training* untuk memperoleh pohon klasifikasi optimal. Kemudian pohon klasifikasi optimal yang sudah diperoleh diterapkan pada data *testing* untuk mengetahui ketepatan klasifikasi yang dihasilkan metode *bagging*. Berdasarkan perhitungan yang telah dilakukan, pada replikasi 10 dan 25 diperoleh hasil klasifikasi yang serupa yang disajikan dalam Tabel 4.9 berikut.

Tabel 4.9 Hasil Klasifikasi Analisis CART *Bagging* untuk Data *Training* Replikasi 10 dan 25

Kelas Observasi Y	Kelas Prediksi Y		Total
	1	2	
1	3	1	4
2	1	1	2
Total	4	2	6

Berdasarkan Tabel 4.9 dapat diketahui bahwa masing-masing satu pengamatan yang diprediksikan secara tidak tepat ke dalam kelas 1 atau BTA Positif dan kelas 2 atau BTA Negatif. Selain itu, dapat diketahui pula bahwa terdapat 3 pengamatan yang diprediksikan secara tepat ke dalam kelas 1 atau BTA Positif dan 1 pengamatan yang diprediksikan secara tepat ke dalam kelas 2 atau BTA Negatif.

Pengukuran ketepatan klasifikasi untuk analisis CART *bagging* dengan replikasi 10 dan 25 yang diterapkan pada data *training* dilakukan dengan perhitungan berikut.

$$1 - APER = \frac{3 + 1}{6} = 0,667$$

Dari hasil perhitungan di atas, dapat diketahui bahwa tingkat ketepatan klasifikasi sebesar 0,667 atau 66,7%.

Setelah dilakukan klasifikasi dan pengukuran ketepatan klasifikasi terhadap data *training* dengan replikasi *bagging* 10 dan 25 selanjutnya dilakukan klasifikasi dengan replikasi 50. Berikut merupakan hasil klasifikasi analisis CART *bagging* dengan replikasi 50.

Tabel 4.10 Hasil Klasifikasi Analisis CART *Bagging* untuk Data *Training* Replikasi 50

Kelas Observasi Y	Kelas Prediksi Y		Total
	1	2	
1	2	2	4
2	1	1	2
Total	4	2	6

Berdasarkan Tabel 4.10 dapat diketahui bahwa terdapat 2 pengamatan yang tepat diklasifikasikan pada kelas 1 atau BTA Positif dan 1 pengamatan tepat diklasifikasikan pada kelas 2 atau BTA Negatif. Pada hasil klasifikasi dengan replikasi 50 pada data *training* ini diketahui pula bahwa terdapat 2 pengamatan yang diprediksikan secara tidak tepat ke dalam kelas 2 atau BTA Negatif dan 1 pengamatan yang diprediksikan secara tidak tepat ke dalam kelas 1 atau BTA Positif.

Pengukuran ketepatan klasifikasi untuk analisis CART *bagging* dengan replikasi 50 yang diterapkan pada data *training* dilakukan dengan perhitungan berikut.

$$1 - APER = \frac{2 + 1}{6} = 0,5$$

Dari hasil perhitungan di atas, dapat diketahui tingkat ketepatan klasifikasi sebesar 0,5 atau 50%. Berdasarkan perhitungan dapat dikatakan pula bahwa ketepatan klasifikasi dengan replikasi *bagging* 50 mengalami penurunan.

Setelah dilakukan klasifikasi dan pengukuran ketepatan klasifikasi terhadap data *training* dengan replikasi *bagging* 10, 25, dan 50, selanjutnya dilakukan klasifikasi dan pengukuran ketepatan klasifikasi terhadap data *training* dengan replikasi *bagging* 100, 125, dan 150. Pada replikasi 100, 125, dan 150 diperoleh hasil klasifikasi yang sama. Berikut merupakan hasil klasifikasi analisis CART *bagging* dengan replikasi 100, 125, dan 150.

Tabel 4.11 Hasil Klasifikasi Analisis CART *Bagging* untuk Data *Training* Replikasi 100, 125, dan 150

Kelas Observasi Y	Kelas Prediksi Y		Total
	1	2	
1	4	0	4
2	1	1	2
Total	5	1	6

Berdasarkan Tabel 4.11 dapat diketahui bahwa terdapat 4 pengamatan yang tepat diklasifikasikan pada kelas 1 atau BTA Positif dan 1 pengamatan tepat diklasifikasikan pada kelas 2 atau BTA Negatif. Pada hasil klasifikasi dengan replikasi 100, 125, dan 150 pada data *training* ini diketahui pula bahwa tidak ada pengamatan yang diprediksikan secara tidak tepat ke dalam kelas 2 atau BTA Negatif. Namun masih terdapat satu pengamatan yang diprediksikan secara tidak tepat ke dalam kelas 1 atau BTA Positif.

Selanjutnya, pengukuran ketepatan klasifikasi untuk replikasi 100, 125, dan 150 menggunakan APER dan 1-APER disajikan dalam perhitungan berikut.

$$1 - APER = \frac{4 + 1}{6} = 0,833$$

Berdasarkan hasil perhitungan yang telah dilakukan diperoleh tingkat ketepatan klasifikasi pohon klasifikasi optimal dengan replikasi *bagging* 100, 125, dan 150 0,833 atau 83,3%. Nilai ini menunjukkan pula bahwa terjadi peningkatan akurasi dengan menggunakan replikasi *bagging* 100 dan nilai akurasinya konstan pada replikasi selanjutnya.

Setelah menerapkan replikasi *bagging* pada data *training* selanjutnya replikasi ini diterapkan pada data *testing*. Berdasarkan perhitungan yang telah dilakukan sebelumnya, hasil klasifikasi dan ketepatan klasifikasi untuk data *training* dengan replikasi *bagging* 10, dan 25 menunjukkan hasil yang sama. Sehingga hasil yang sama juga diperoleh pada saat dilakukan klasifikasi dan

pengukuran ketepatan klasifikasi pada data *testing*. Berikut hasil klasifikasi untuk data *testing* dengan replikasi 10 dan 25.

Tabel 4.12 Hasil Klasifikasi Analisis CART *Bagging* untuk Data *Testing* Replikasi 10 dan 25

Kelas Observasi Y	Kelas Prediksi Y		Total
	1	2	
1	9	2	11
2	4	10	14
Total	13	12	25

Berdasarkan hasil klasifikasi yang ditunjukkan pada Tabel 4.11 dapat diketahui bahwa terdapat 9 pengamatan yang diprediksikan secara tepat ke dalam kelas 1 atau BTA Positif dan 10 pengamatan yang diprediksikan secara tepat ke dalam kelas 2 atau BTA Negatif. Namun terdapat 2 pengamatan yang tidak diprediksikan secara tepat ke dalam kelas 2 atau BTA Negatif dan 4 pengamatan yang diprediksikan secara tidak tepat ke dalam kelas 1 atau BTA Positif.

Selanjutnya, pengukuran ketepatan klasifikasi menggunakan *APER* dan *1-APER* untuk data *testing* dengan replikasi 10 dan 25 disajikan dalam perhitungan berikut.

$$1 - APER = \frac{9 + 10}{25} = 0,76$$

Dari hasil perhitungan di atas, dapat dikatakan bahwa tingkat tingkat ketepatan klasifikasinya sebesar 0,76 atau 76%.

Selanjutnya dilakukan klasifikasi dan pengukuran ketepatan klasifikasi pada data *testing* dengan menggunakan replikasi *bagging* 50. Hasil klasifikasi data *testing* dengan replikasi *bagging* 50 disajikan dalam Tabel 4.13 berikut.

Tabel 4.13 Hasil Klasifikasi Analisis CART *Bagging* untuk Data *Testing* Replikasi 50

Kelas Observasi Y	Kelas Prediksi Y		Total
	1	2	
1	10	1	11
2	0	14	14
Total	10	15	25

Tabel 4.13 menunjukkan bahwa terdapat 10 pengamatan yang diprediksikan secara tepat ke dalam kelas 1 atau BTA Positif. Pada kelas 2 atau BTA Negatif terdapat 14 pengamatan yang diprediksikan secara tepat. Meski tidak terdapat pengamatan yang diprediksikan secara tidak tepat ke dalam kelas 1 atau BTA Positif namun masih terdapat 2 pengamatan yang diprediksikan secara tidak tepat ke dalam kelas 2 atau BTA Negatif.

Untuk mengetahui pengukuran ketepatan klasifikasi menggunakan *APER* dan $1 - \text{APER}$ pada data *testing* dengan replikasi *bagging* 50, dilakukan perhitungan berikut.

$$1 - \text{APER} = \frac{10 + 14}{25} = 0,96$$

Informasi yang dapat diperoleh dari hasil perhitungan di atas yaitu bahwa tingkat ketepatan klasifikasi pada data *testing* dengan menggunakan replikasi *bagging* 50 sebesar 0,96 atau 96%. Berdasarkan perhitungan yang telah dilakukan, peningkatan ketepatan klasifikasi dengan replikasi *bagging* 50 20% lebih besar daripada replikasi *bagging* sebelumnya.

Selanjutnya dilakukan klasifikasi dan pengukuran ketepatan klasifikasi pada data *testing* dengan menggunakan replikasi *bagging* 100, 125, dan 150. Tidak berbeda dengan hasil klasifikasi yang diberikan data *training*, hasil klasifikasi data *testing* untuk replikasi 100, 125, dan 150 memberikan hasil yang sama. Hasil klasifikasi data *testing* dengan replikasi *bagging* 100, 125, dan 150 disajikan dalam Tabel 4.14 berikut.

Tabel 4.14 Hasil Klasifikasi Analisis CART *Bagging* untuk Data *Testing* Replikasi 100, 125, dan 150

Kelas Observasi Y	Kelas Prediksi Y		Total
	1	2	
1	9	2	11
2	4	10	14
Total	13	12	25

Berdasarkan hasil klasifikasi yang ditunjukkan pada Tabel 4.14 dapat diketahui bahwa terdapat 9 pengamatan yang diprediksikan secara tepat ke dalam kelas 1 atau BTA Positif dan 10 pengamatan yang diprediksikan secara tepat ke dalam kelas 2 atau BTA Negatif. Namun terdapat 2 pengamatan yang tidak diprediksikan secara tepat ke dalam kelas 2 atau BTA Negatif dan 4 pengamatan yang diprediksikan secara tidak tepat ke dalam kelas 1 atau BTA Positif.

Selanjutnya, pengukuran ketepatan klasifikasi menggunakan APER dan 1-APER untuk data *testing* dengan replikasi 100, 125, dan 150 disajikan dalam perhitungan berikut.

$$1 - APER = \frac{9 + 10}{25} = 0,76$$

Dari hasil perhitungan di atas, dapat dikatakan bahwa tingkat ketepatan klasifikasi pohon klasifikasi optimal pada data *testing* dengan replikasi 100, 125, dan 150 sebesar 0,76 atau 76%. Ketepatan klasifikasi pada replikasi *bagging* 100 mengalami penurunan namun mencapai nilai konstan pada replikasi selanjutnya.

4.4 Perbandingan Ketepatan Klasifikasi Analisis CART dan CART *Bagging*

Setelah dilakukan pengukuran ketepatan klasifikasi pada data *training* maupun data *testing* baik menggunakan analisis CART maupun analisis CART *bagging* dengan replikasi diperoleh nilai ketepatan klasifikasi pada Tabel 4.15.

Tabel 4.15 Nilai 1-APER Analisis CART dan Analisis CART *Bagging* Penderita Tuberkulosis dalam Rumah Tangga di Surabaya

	CART	CART BAGGING					
		10	25	50	100	125	150
Data <i>training</i>	0,636	0,667	0,667	0,50	0,833	0,833	0,833
Data <i>testing</i>	0,76	0,76	0,76	0,96	0,76	0,76	0,76

Berdasarkan Tabel 4.15 dapat diketahui bahwa ketepatan klasifikasi analisis CART penderita tuberkulosis dalam rumah tangga di Surabaya sebesar 63,6% untuk data *training* dan 76% untuk data *testing*. Selanjutnya penggunaan replikasi *bagging* pada data penderita tuberkulosis dalam rumah tangga di Surabaya menghasilkan ketepatan klasifikasi yang bervariasi. Ketepatan klasifikasi konvergen diperoleh ketika replikasi 100 diterapkan pada data *training* dan *testing*. Pada data *training*, penggunaan replikasi 100 menghasilkan ketepatan klasifikasi sebesar 83,3%. Sedangkan pada data *testing*, ketepatan klasifikasi yang dihasilkan sebesar 76%.



BAB V
KESIMPULAN DAN SARAN



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan untuk pengelompokan penderita tuberkulosis dalam rumah tangga di Surabaya diperoleh kesimpulan sebagai berikut.

1. Jumlah kejadian tuberkulosis dengan kategori atau kelas 1 yaitu BTA Positif berjumlah 63 penderita sedangkan penderita dengan kategori atau kelas 2 yaitu BTA Negatif berjumlah 61 penderita. Berdasarkan variabel jumlah anggota keluarga, baik pada kelas BTA Positif maupun BTA Negatif lebih banyak penderita yang tinggal dengan anggota keluarga yang kurang dari 7 orang. Apabila ditinjau dari variabel tingkat pendidikan KRT, baik pada kelas BTA Positif maupun BTA Negatif lebih banyak penderita yang tinggal dengan KRT yang bersekolah. Jika dilihat berdasarkan variabel jenis kelamin, baik pada kelas BTA Positif maupun BTA Negatif lebih banyak penderita yang berjenis kelamin laki-laki dibandingkan dengan penderita berjenis kelamin perempuan. Berdasarkan variabel penghasilan KRT per bulan, baik pada kelas BTA Positif maupun BTA Negatif lebih banyak penderita yang tinggal dengan KRT dengan penghasilan lebih dari Rp 600.000. Pada variabel kebiasaan merokok, baik pada kelas BTA Positif maupun BTA Negatif lebih banyak penderita yang tidak memiliki kebiasaan tidak merokok. Apabila dilihat dari variabel status gizi, pada kelas BTA Positif lebih banyak penderita yang berstatus gizi normal, sedangkan pada kelas BTA Negatif lebih banyak penderita yang berstatus gizi kurang/lebih. Jika ditinjau dari variabel kepemilikan toilet/WC/jamban, baik pada kelas BTA Positif maupun BTA Negatif lebih banyak penderita yang memiliki toilet/WC/jamban sendiri. Berdasarkan variabel sumber air bersih, baik pada kelas BTA Positif maupun BTA Negatif lebih banyak penderita yang menggunakan sumber air bersih

yang berasal dari PDAM dibandingkan dengan sumur. Pada variabel tempat pembuangan air limbah dari kamar mandi/tempat cuci/dapur, baik pada kelas BTA Positif maupun BTA Negatif lebih banyak penderita yang membuang air limbah langsung ke got/sungai. Berdasarkan variabel kepadatan hunian rumah, baik pada kelas BTA Positif maupun BTA Negatif lebih banyak penderita yang tinggal dengan kepadatan hunian rumah yang memenuhi syarat.

2. Berdasarkan analisis CART yang telah dilakukan pada data kejadian tuberkulosis di Surabaya diperoleh pohon klasifikasi optimal dengan 18 simpul terminal dan 10 kedalaman. Berdasarkan pohon klasifikasi optimal yang telah dihasilkan tersebut diperoleh ketepatan klasifikasi pada data *training* sebesar 63,3% dan ketepatan klasifikasi sebesar 76% pada data *testing*.
3. Penerapan metode *bagging* pada analisis CART penderita tuberkulosis dalam rumah tangga di Surabaya tidak selalu meningkatkan ketepatan klasifikasi pada pohon klasifikasi. Hal ini dapat dibuktikan dengan ketepatan klasifikasi analisis CART yang bernilai sama dengan ketepatan klasifikasi analisis CART *bagging* yang sudah konvergen yaitu sebesar 76%.

5.2 Saran

Bagi penelitian selanjutnya dapat diterapkan pembagian data *training:testing* serta replikasi *bagging* yang lain untuk memperoleh hasil pengelompokan yang berbeda dari penelitian yang telah dilakukan sebelumnya. Selain itu dapat pula ditambahkan variabel penelitian lainnya untuk mengetahui faktor-faktor lain yang berpengaruh dalam pengelompokan tuberkulosis berdasarkan hasil pemeriksaan dahaknya.

Penggunaan *bagging* pada kasus ini tidak menghasilkan peningkatan ketepatan klasifikasi. Sehingga lebih disarankan penggunaan CART pada pengelompokan penderita tuberkulosis dalam rumah tangga di Surabaya.

Berdasarkan hasil penelitian yang telah dilakukan, ditemukan variabel terpenting dalam pembentukan pohon yaitu variabel penghasilan KRT per bulan, tempat pembuangan air limbah, tingkat pendidikan KRT, sumber air bersih yang digunakan dan status gizi. Perubahan variabel-variabel tersebut diharapkan mampu membantu pemerintah Kota Surabaya dalam menanggulangi tingginya kasus tuberkulosis dan mengurangi risiko penularan tuberkulosis.

“Halaman ini sengaja dikosongkan”



DAFTAR PUSTAKA



DAFTAR PUSTAKA

- Amir, & Alsegaf, H. 1989. *Pengantar Penyakit Paru*. Surabaya: Airlangga University Press.
- Anwar, M. Maret, 2016. *Penderita TB di Surabaya Tertinggi di Jatim, Kedua Jember, dan Ketiga Sidoarjo, Ini Data Lengkapnya*, <URL:<http://surabaya.tribunnews.com/2016/03/23/penderita-tb-di-surabaya-tertinggi-di-jatim-kedua-jember-dan-ketiga-sidoarjo-ini-data-lengkapnya>>
- Badan Pusat Statistik Kota Surabaya. 2014. *Surabaya Dalam Angka*. Surabaya: Badan Pusat Statistik Kota Surabaya.
- Bimantara, J. G. Maret, 2016. *Tuberkulosis di Indonesia Terbanyak Kedua di Dunia*. <URL:[http:// www.print.kompas.com](http://www.print.kompas.com)>
- Bogue, D. J. 1969. *Principles of Demography*. New York: John Wiley and Sons.
- Breiman , L., Friedman, J., Olshen , R. A., & Stone, C. 1984. *Classification and Regression Trees*. New York: Chapman and Hall (Wadsworth, Inc).
- Breiman, L. 1996. *Bagging Predictors*. *Machine Learning*, 123-140.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. 1984. *Classification and Regression Tree*. New York - London: Chapman Hall.
- Buhlmann, P., & Yu, B. 2002. *Analyzing Bagging*. *The Annals of Statistics Vol. 30 No. 4*, 927-961.
- California Statistical Software Inc. 2000. *CART References Guide*. San Diego, California: Salford System.
- Departemen Kesehatan RI. 2009. *Buku Saku Kader Program Penanggulangan TB*. Tangerang: Direktorat Jenderal Pengendalian Penyakit dan Penyehatan Lingkungan Departemen Kesehatan RI.
- Dinas Kesehatan Kota Surabaya. 2015. *Profil Kesehatan Tahun 2015*. Surabaya. Dinas Kesehatan Kota Surabaya

- Dinas Kesehatan Pemerintah Kota Surabaya. 2015. *Profil Kesehatan Kota Surabaya Tahun 2014*. Surabaya: Dinas Kesehatan Pemerintah Kota Surabaya.
- Dinas Kesehatan Provinsi Jawa Timur. 2013. *Profil Kesehatan Provinsi Jawa Timur Tahun 2012*. Surabaya: Dinas Kesehatan Provinsi Jawa Timur.
- Fatimah , S. 2008. *Faktor Kesehatan Lingkungan Rumah yang Berhubungan dengan Kejadian TB Paru di Kabupaten Cilacap*. Semarang: Universitas Diponegoro.
- Hastareksa. Januari, 2016. *Tuberkulosis Paru BTA Positif Jatim Tembus 15.371 Kasus*. <URL:<http://www.surabayaonline.co>>
- Hiswani. 2009. *Tuberkulosis Merupakan Penyakit Infeksi yang Masih Menjadi Masalah Kesehatan*. Medan: Universitas Sumatera Utara.
- Johnson, R. A., & Dean, W. W. 2007. *Applied Multivariate Statistical Analysis 6th Ed*. New Jersey: Prentice Hall Internatioonal Inc.
- Kementerian Kesehatan RI. 1999. KEPMENKES RI NO. 829/MENKES/SK/VII/1999 tentang Persyaratan Kesehatan Perumahan
- Kementerian Kesehatan RI. 2014. *Pedoman Nasional Pengendalian Tuberkulosis*. Jakarta: Kementerian Kesehatan Republik Indonesia Direktorat Jenderal Pengendalian Penyakit dan Penyehatan Lingkungan.
- Kesehatan Masyarakat. 2011. *Faktor-faktor Risiko Tuberkulosis (TB-Paru-TBC)* . <URL:<http://www.kesmas.tk>>
- Kementerian Kesehatan Republik Indonesia. 2016. *Profil Kesehatan Indonesia 2015*. Jakarta. Kementerian Kesehatan Republik Indonesia.
- Korua, E. S., Kapantow, H. N., & Kawatu, P. A. 2015. Hubungan Antara Umur, Jenis Kelamin, dan Kepadatan Hunian dengan Kejadian TB Paru pada Pasien Rawat Jalan di Rumah Sakit Umum Daerah Noongan.

- Lewis, R. J. 2000. An Introduction to *Classification and Regression Trees (CART) Analysis*. *Annual Meeting of the Society for Academic Emergency Medicine*.
- Manalu, H. S. 2010. Faktor-faktor yang Mempengaruhi Kejadian TB Paru dan Upaya Penanggulangannya. *Jurnal Ekologi Kesehatan*, 1340-1346.
- Mazayudha, & Mandakir. 2014. Perilaku Penderita TBC dengan Penularan pada Anggota Keluarga. *The Sun*, 37-43.
- Mukid, M. A., Wuryandari, T., Ratnaningrum, D., & Rahayu, R. S. 2015. *Bagging Classification Trees* Untuk Prediksi Risiko Preeklampsia (Studi Kasus: Ibu Hamil Kategori Penerima Jampersal di RSUD Dr. Moewardi Surakarta). *Media Statistika*, 111-120.
- Muniroh, N., Aisah, S., & Mifbakhuddin. 2013. Faktor-faktor yang Berhubungan dengan Penyembuhan Penyakit Tuberkulosis (TBC) Paru di Wilayah Kerja Puskesmas Mangkang Semarang Barat. *Jurnal Keperawatan Komunitas Vol. 1 No.1*, 33-42.
- Notoadmojo, S. 2005. *Metodologi Penelitian Kesehatan*. Jakarta: Rineka Cipta.
- Nurhartanto, S. Maret, 2015. *Jatim Peringkat Dua Tuberculosis se-Indonesia*. <URL:<http://www.encycity.co/jatim-peringkat-dua-tuberculosis-se-indonesia/>>
- Otok, B. W., & Sumarmi. 2009. *Bagging CART* pada Klasifikasi Anak Putus Sekolah. *Seminar Nasional Statistika IX*, 1-9.
- Otok, B. W., Guritno, S., & Haryatmi, S. 2006. *Bootstrap* dalam MARS untuk Klasifikasi Perbankan. *Inferensi Jurnal Statistika Volume 2 No. 1*.
- Perhimpunan Dokter Paru Indonesia. 2006. *Tuberkulosis, Pedoman Diagnosis dan Penatalaksanaan di Indonesia*. <URL:<http://www.klikdpi.com>>
- Perkumpulan Pemberantasan Tuberkulosis Indonesia. 2012. *Jurnal Tuberkulosis Indonesia Vol.8*.
- Rusnoto, Rahmaatullah, P., & Udiono, A. 2008. Faktor-Faktor yang Berhubungan dengan Kejadian TB Paru pada Usia

- Dewasa (Studi Kasus di Balai Pencegahan dan Pengobatan Penyakit Paru Pati).
- Steinberg, Dan dan Golovnya, Mikhail. 2007. *CART 6.0 User's Guide*. San Diego. Salford Systems.
- Supariasa. 2002. *Penilaian Status Gizi*. Jakarta. Penerbit Kedokteran EGC
- Sutton, C. D. 2005. *Classification and Regression Trees, Bagging, and Boosting. Handbook of Statistics*.
- Timofeev, R. 2004. *Classification and Regression Trees (CART) Theory and Application*. Berlin: Humboldt University.
- Walid, W. I. Maret, 2016. *Indonesia Negara Kedua Terbanyak Penderita TB*. <URL:<http://news.fajarnews.com/read/2016/03/26/9908/indonesia.negara.kedua.terbanyak.penderita.tb>>
- WHO. 2013. *Global Tuberculosis Report 2013*. France: WHO Press.
- Widyasari, R. M. 2011. Hubungan Antara Jenis Kepribadian, Riwayat Diabetes Melitus dan Riwayat Paparan Merokok dengan Kejadian Tuberkulosis Paru Dewasa di Wilayah Kecamatan Semarang Utara Tahun 2011. *Jurnal Kesehatan Masyarakat*, 446-453.
- Yunus, M. 1993. *Sejarah Pendidikan Islam di Indonesia*. Jakarta: Hidakarya Agung.



LAMPIRAN



LAMPIRAN

Lampiran A. Data Pengamatan Kejadian Tuberkulosis di Surabaya Tahun 2016.

No.	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁
1	1	2	1	1	1	2	2	2	1	1	1	2
2	2	2	1	2	1	2	2	1	1	2	1	2
3	2	2	1	1	1	2	2	2	1	1	1	1
4	1	2	1	1	2	2	2	1	2	1	1	1
5	1	2	1	1	2	2	2	1	2	1	1	1
6	2	2	1	2	2	2	2	1	2	2	1	2
7	1	2	1	1	1	2	2	1	2	1	1	1
8	2	1	1	1	2	2	1	1	2	1	1	1
9	2	2	1	1	2	2	1	1	2	1	1	2
10	1	2	1	2	2	2	2	2	2	2	1	2
...
60	2	1	1	1	2	1	2	1	2	1	1	2
61	1	1	1	2	2	1	2	2	2	2	1	2
62	1	2	1	1	2	2	2	1	2	2	1	2
63	2	2	1	2	2	2	2	1	1	2	1	1
64	1	2	1	2	1	2	2	1	2	1	1	2
65	2	2	1	2	1	2	2	2	1	2	1	2
66	2	2	1	2	2	2	2	1	2	2	1	1
...
119	2	2	1	1	2	1	2	1	2	2	1	2
120	2	2	1	1	2	2	1	2	1	1	1	1
121	2	2	1	2	2	2	2	1	2	2	2	2
122	2	1	1	1	2	2	2	2	1	2	2	1
123	1	2	1	1	1	2	2	2	2	2	2	2
124	1	2	1	1	2	2	2	1	1	1	2	1

Keterangan :

Y(1) : BTA Positif

Y(2) : BTA Negatif

X₁ : Jumlah anggota keluargaX₂ : Tingkat pendidikan KRTX₃ : Jenis kelaminX₄ : Penghasilan KRT per bulanX₅ : Riwayat penyakitX₆ : Kebiasaan merokokX₇ : Status giziX₈ : Sumber air bersihX₉ : Kepemilikan toilet/WC/jambanX₁₀ : Tempat pembuangan air limbah dari kamar mandi/tempat cuci/dapurX₁₁ : Kepadatan hunian rumah**Lampiran B. Statistik Deskriptif Variabel Prediktor****Jumlah Anggota Keluarga*Kejadian Tuberkulosis Crosstabulation**

Count	Kejadian Tuberkulosis		Total
	1	2	
Jumlah Anggota Keluarga	1	5	6
	2	58	55
Total	63	61	124

Tingkat Pendidikan KRT *Kejadian Tuberkulosis Crosstabulation

Count	Kejadian Tuberkulosis		Total
	1	2	
Tingkat Pendidikan KRT	1	59	58
	2	4	3
Total	63	61	124

Jenis Kelamin *Kejadian Tuberkulosis Crosstabulation

Count	Kejadian Tuberkulosis		Total
	1	2	

Jenis Kelamin	1	39	32	71
	2	24	29	53
Total		63	61	124

Penghasilan KRT per Bulan *Kejadian Tuberkulosis Crosstabulation

Count	Kejadian Tuberkulosis		Total	
	1	2		
Penghasilan KRT per Bulan	1	19	11	30
	2	44	50	94
Total		63	61	124

Riwayat Penyakit *Kejadian Tuberkulosis Crosstabulation

Count	Kejadian Tuberkulosis		Total	
	1	2		
Riwayat Penyakit	1	22	19	41
	2	41	42	83
Total		63	61	124

Kebiasaan Merokok *Kejadian Tuberkulosis Crosstabulation

Count	Kejadian Tuberkulosis		Total	
	1	2		
Kebiasaan Merokok	1	18	16	34
	2	45	45	90
Total		63	61	124

Status Gizi *Kejadian Tuberkulosis Crosstabulation

Count	Kejadian Tuberkulosis		Total	
	1	2		
Status Gizi	1	25	32	57
	2	38	29	67
Total		63	61	124

**Kepemilikan Toilet/WC/Jamban *Kejadian Tuberkulosis
Crosstabulation**

Count	Kejadian Tuberkulosis		Total	
	1	2		
Kepemilikan Toilet/WC/Jamban	1	16	12	28
	2	47	49	96
Total	63	61	124	

Sumber Air Bersih *Kejadian Tuberkulosis Crosstabulation

Count	Kejadian Tuberkulosis		Total	
	1	2		
Sumber Air Bersih	1	18	17	35
	2	45	44	89
Total	63	61	124	

**Tempat Pembuangan Air Limbah dari Kamar Mandi/Tempat
Cuci/Dapur *Kejadian Tuberkulosis Crosstabulation**

Count	Kejadian Tuberkulosis		Total	
	1	2		
Tempat Pembuangan Air Limbah dari Kamar Mandi/Tempat Cuci/Dapur	1	54	54	108
	2	9	7	16
Total	63	61	124	

Kepadatan Hunian Rumah *Kejadian Tuberkulosis Crosstabulation

Count	Kejadian Tuberkulosis		Total	
	1	2		
Kepadatan Hunian Rumah	1	30	27	57
	2	33	34	67
Total	63	61	124	

LEARNING SAMPLE VARIABLE STATISTICS				
=====				
VARIABLE		CLASS		OVERALL
		1	2	
Y	MEAN	1.000	2.000	1.455
	SD	0.000	0.000	0.500
	N	54.000	45.000	99.000
	SUM	54.000	90.000	144.000
X1	MEAN	1.963	1.933	1.949
	SD	0.191	0.252	0.220
	N	54.000	45.000	99.000
	SUM	106.000	87.000	193.000
X2	MEAN	1.093	1.044	1.071
	SD	0.293	0.208	0.258
	N	54.000	45.000	99.000
	SUM	59.000	47.000	106.000
X3	MEAN	1.444	1.489	1.465
	SD	0.502	0.506	0.501
	N	54.000	45.000	99.000
	SUM	78.000	67.000	145.000
X4	MEAN	1.648	1.733	1.687
	SD	0.482	0.447	0.466
	N	54.000	45.000	99.000
	SUM	89.000	78.000	167.000
X5	MEAN	1.593	1.600	1.596
	SD	0.496	0.495	0.493
	N	54.000	45.000	99.000
	SUM	86.000	72.000	158.000
X6	MEAN	1.815	1.689	1.758
	SD	0.392	0.468	0.431
	N	54.000	45.000	99.000
	SUM	98.000	76.000	174.000
X7	MEAN	1.648	1.422	1.545
	SD	0.482	0.499	0.500
	N	54.000	45.000	99.000
	SUM	89.000	64.000	153.000
X8	MEAN	1.796	1.800	1.798
	SD	0.407	0.405	0.404
	N	54.000	45.000	99.000
	SUM	97.000	81.000	178.000
X9	MEAN	1.685	1.733	1.707
	SD	0.469	0.447	0.457
	N	54.000	45.000	99.000
	SUM	91.000	78.000	169.000
X10	MEAN	1.278	1.156	1.222
	SD	0.452	0.367	0.418
	N	54.000	45.000	99.000
	SUM	69.000	52.000	121.000
X11	MEAN	1.537	1.667	1.596
	SD	0.503	0.477	0.493
	N	54.000	45.000	99.000
	SUM	83.000	75.000	158.000

Lampiran C. Output Pohon Klasifikasi *Cross-Validation 10-Fold*

```

=====
TERMINAL NODE INFORMATION
=====

[Breiman adjusted cost, lambda = 0.022]

```

Node	Class	Wgt	Count	N	Prob	Cost	Parent Complexity
1	1		8.00	8	0.076	0.146	0.015 [0.466]
	1		7.00	7	0.854		
	2		1.00	1	0.146		
2	1		2.00	2	0.019	0.000	0.003 [0.777]
	1		2.00	2	1.000		
	2		0.00	0	0.000		
3	1		2.00	2	0.019	0.000	0.004 [0.777]
	1		2.00	2	1.000		
	2		0.00	0	0.000		
4	1		2.00	2	0.019	0.000	0.005 [0.777]
	1		2.00	2	1.000		
	2		0.00	0	0.000		
5	1		2.00	2	0.019	0.000	0.007 [0.777]
	1		2.00	2	1.000		
	2		0.00	0	0.000		
6	1		8.00	8	0.078	0.286	0.015 [0.599]
	1		6.00	6	0.714		
	2		2.00	2	0.286		
7	2		5.00	5	0.052	0.357	0.015 [0.781]
	1		2.00	2	0.357		
	2		3.00	3	0.643		
8	2		4.00	4	0.041	0.455	0.005 [0.955]
	1		2.00	2	0.455		
	2		2.00	2	0.545		
9	1		2.00	2	0.019	0.000	0.008 [0.777]
	1		2.00	2	1.000		
	2		0.00	0	0.000		
10	1		6.00	6	0.059	0.375	0.015 [0.760]
	1		4.00	4	0.625		
	2		2.00	2	0.375		
11	2		7.00	7	0.072	0.385	0.015 [0.717]
	1		3.00	3	0.385		
	2		4.00	4	0.615		
12	2		6.00	6	0.065	0.143	0.017 [0.504]
	1		1.00	1	0.143		
	2		5.00	5	0.857		
13	1		8.00	8	0.074	0.000	0.024 [0.326]
	1		8.00	8	1.000		
	2		0.00	0	0.000		
14	1		3.00	3	0.028	0.000	0.028 [0.631]

1	3.00	3	1.000			
2	0.00	0	0.000			
15 2	8.00	8	0.085	0.217	0.028	[0.509]
1	2.00	2	0.217			
2	6.00	6	0.783			
16 1	3.00	3	0.028	0.000	0.028	[0.631]
1	3.00	3	1.000			
2	0.00	0	0.000			
17 2	11.00	11	0.117	0.238	0.028	[0.463]
1	3.00	3	0.238			
2	8.00	8	0.762			
18 2	12.00	12	0.133	0.000	0.014	[0.201]
1	0.00	0	0.000			
2	12.00	12	1.000			
Node	Learn					
1	8.000	7.000	1.000			
2	2.000	2.000	0.000			
3	2.000	2.000	0.000			
4	2.000	2.000	0.000			
5	2.000	2.000	0.000			
6	8.000	6.000	2.000			
7	5.000	2.000	3.000			
8	4.000	2.000	2.000			
9	2.000	2.000	0.000			
10	6.000	4.000	2.000			
11	7.000	3.000	4.000			
12	6.000	1.000	5.000			
13	8.000	8.000	0.000			
14	3.000	3.000	0.000			
15	8.000	2.000	6.000			
16	3.000	3.000	0.000			
17	11.000	3.000	8.000			
18	12.000	0.000	12.000			

Lampiran D. Output *Missclassification*

=====							
MISCLASSIFICATION BY CLASS							
=====							
(Cross Validation)							
Class	Prior Prob	Wgt Count	Count	Wgt Misclass	Misclass	Cost	

1	0.500	54.00	54	13.00	13	0.241	
		(54.00)	54	21.00	21	0.389)	
2	0.500	45.00	45	5.00	5	0.111	
		(45.00)	45	15.00	15	0.333)	

Total	1.000	99.00	99	18.00	18		
		(99.00)	99	36.00	36)		

Lampiran E. Akurasi Hasil Klasifikasi Analisis CART Akurasi Data *Training*

=====							
CROSS VALIDATION CLASSIFICATION TABLE							
=====							

Actual Class	Predicted Class		Actual Total
	1	2	
1	33.00	21.00	54.00
2	15.00	30.00	45.00
PRED. TOT.	48.00	51.00	99.00
CORRECT	0.611	0.667	
SUCCESS IND.	0.066	0.212	
TOT. CORRECT	0.636		
SENSITIVITY: 0.611 SPECIFICITY: 0.667 FALSE REFERENCE: 0.313 FALSE RESPONSE: 0.412 REFERENCE = "1", RESPONSE = "2"			

Akurasi Data Testing

=====

CASE-BY-CASE SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	9.00	2.00	11.00
2	4.00	10.00	14.00
PRED. TOT.	13.00	12.00	25.00
CORRECT	0.818	0.714	
SUCCESS IND.	0.378	0.154	
TOT. CORRECT	0.760		
SENSITIVITY: 0.818 SPECIFICITY: 0.714 FALSE REFERENCE: 0.308 FALSE RESPONSE: 0.167 REFERENCE = "1", RESPONSE = "2"			

=====

Lampiran F. Informasi Simpul Terminal Pohon Klasifikasi Optimal

=====

NODE INFORMATION

=====

```

*****
*                Node 1: X7                *
*                N: 99                      *
*****
*****
*                Node 2                    *   *                Node 13                *
*                N: 54                      *   *                N: 45                *
*****

```

Node 1 was split on X7
 A case goes left if X7 = (2)
 Improvement = 0.026 Complexity Threshold = 0.113

Node	Cases	Wgt	Counts	Cost	Class
1	99		99.00	0.500	1
2	54		54.00	0.394	1
13	45		45.00	0.378	2

Weighted Counts

Class	Top	Left	Right
-------	-----	------	-------

1	54.00	35.00	19.00
2	45.00	19.00	26.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.500	0.606	0.378
2	0.500	0.394	0.622

Surrogate		Split	Assoc.	Improve.
1	X4	s 2	0.139	0.004
2	X10	s 1	0.139	0.011
3	X8	s 2	0.044	.106431E-04
4	X9	s 2	0.020	0.001
5	X6	s 1	0.016	0.011

Competitor		Split	Improve.
1	X10	2	0.011
2	X6	2	0.011
3	X11	1	0.009
4	X2	2	0.005
5	X4	1	0.004

```

*****
*                   Node 2: X10                   *
*                   N: 54                          *
*****

```

```

=====
=      Terminal Node 1      = *      Node 3      *
=      N: 8                 = *      N: 46         *
=====

```

Node 2 was split on X10
A case goes left if X10 = (2)
Improvement = 0.011 Complexity Threshold = 0.015

Node	Cases	Wgt	Counts	Cost	Class
2	54		54.00	0.394	1
-1	8		8.00	0.146	1
3	46		46.00	0.435	1

Weighted Counts			
Class	Top	Left	Right
1	35.00	7.00	28.00
2	19.00	1.00	18.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.606	0.854	0.565
2	0.394	0.146	0.435

Competitor		Split	Improve.
1	X6	2	0.011
2	X4	2	0.007
3	X3	2	0.004
4	X9	2	0.002
5	X2	1	0.002

```

*****
*                   Node 3: X3                   *
*                   N: 46                          *
*****

```

```

*****
*           Node 4           *           Node 10           *
*           N: 25           *           N: 21           *
*****

```

Node 3 was split on X3
A case goes left if X3 = (2)
Improvement = 0.014 Complexity Threshold = 0.023

Node	Cases	Wgt	Counts	Cost	Class
3	46		46.00	0.435	1
4	25		25.00	0.318	1
10	21		21.00	0.431	2

Weighted Counts

Class	Top	Left	Right
1	28.00	18.00	10.00
2	18.00	7.00	11.00

Within Node Probabilities

Class	Top	Left	Right
1	0.565	0.682	0.431
2	0.435	0.318	0.569

Surrogate Split Assoc. Improve.

1 X6	s 2		0.759	0.006
2 X5	s 1		0.302	0.001
3 X8	s 2		0.250	.667150E-03
4 X9	s 2		0.241	.172595E-04
5 X11	s 2		0.112	0.001

Competitor Split Improve.

1 X6	2		0.006
2 X4	2		0.004
3 X5	1		0.001
4 X11	2		0.001
5 X2	1		0.001

```

*****
*           Node 4: X2           *
*           N: 25           *
*****

```

```

=====
=           Terminal Node 2           =           *           Node 5           *
=           N: 2           =           *           N: 23           *
=====

```

Node 4 was split on X2
A case goes left if X2 = (2)
Improvement = 0.004 Complexity Threshold = 0.003

Node	Cases	Wgt	Counts	Cost	Class
4	25		25.00	0.318	1
-2	2		2.00	0.000	1
5	23		23.00	0.344	1

Weighted Counts

Class	Top	Left	Right
1	18.00	2.00	16.00
2	7.00	0.00	7.00

Within Node Probabilities

Class	Top	Left	Right
1	0.682	1.000	0.656
2	0.318	0.000	0.344

Competitor	Split	Improve.
1 X8	1	0.004
2 X4	1	0.002
3 X1	2	.217694E-03
4 X5	1	.150779E-03
5 X9	1	.975001E-04


```

*****
*           Node 5: X8           *
*           N: 23                *
*****
=====
= Terminal Node 3 = *           Node 6           *
= N: 2           = *           N: 21          *
=====

```

Node 5 was split on X8
A case goes left if X8 = (1)
Improvement = 0.005 Complexity Threshold = 0.004

Node	Cases	Wgt	Counts	Cost	Class
5	23		23.00	0.344	1
-3	2		2.00	0.000	1
6	21		21.00	0.375	1

Weighted Counts

Class	Top	Left	Right
1	16.00	2.00	14.00
2	7.00	0.00	7.00

Within Node Probabilities

Class	Top	Left	Right
1	0.656	1.000	0.625
2	0.344	0.000	0.375

Competitor	Split	Improve.
1 X4	1	0.003
2 X11	2	0.001
3 X9	1	.322046E-03
4 X1	2	.644396E-04
5 X5	2	.747126E-05


```

*****
*           Node 6: X11          *
*           N: 21                *
*****
=====
*           Node 7           * = Terminal Node 8 =
*           N: 17           * = N: 4 =
*****
=====

```

Node 6 was split on X11
A case goes left if X11 = (2)
Improvement = 0.003 Complexity Threshold = 0.005

Node	Cases	Wgt	Counts	Cost	Class
6	21		21.00	0.375	1
7	17		17.00	0.333	1
-8	4		4.00	0.455	2

Weighted Counts

Class	Top	Left	Right
1	14.00	12.00	2.00
2	7.00	5.00	2.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.625	0.667	0.455
2	0.375	0.333	0.545

Surrogate	Split	Assoc.	Improve.
1 X1 s 2		0.273	0.000
2 X4 s 2		0.273	0.003
3 X9 s 2		0.045	.763126E-03

Competitor	Split	Improve.
1 X4 1		0.003
2 X9 1		.763126E-03
3 X5 2		.466368E-03

 * Node 7: X4 *
 * N: 17 *

```

=====
= Terminal Node 4 = * Node 8 *
= N: 2 = * N: 15 *
=====
*****
  
```

Node 7 was split on X4
 A case goes left if X4 = (1)
 Improvement = 0.005 Complexity Threshold = 0.005

Node	Cases	Wgt	Counts	Cost	Class
7	17		17.00	0.333	1
-4	2		2.00	0.000	1
8	15		15.00	0.375	1

Weighted Counts			
Class	Top	Left	Right
1	12.00	2.00	10.00
2	5.00	0.00	5.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.667	1.000	0.625
2	0.333	0.000	0.375

Competitor	Split	Improve.
1 X9 1		0.005
2 X1 1		0.002
3 X5 2		.661376E-03

 * Node 8: X9 *
 * N: 15 *

```

=====
= Terminal Node 5 = * Node 9 *
= N: 2 = * N: 13 *
=====
*****
  
```

Node 8 was split on X9

```

A case goes left if X9 = (1)
Improvement = 0.006          Complexity Threshold = 0.007

Node   Cases   Wgt Counts      Cost Class
   8     15     15.00      0.375 1
  -5     2       2.00      0.000 1
   9     13     13.00      0.429 1

                Weighted Counts
Class      Top      Left      Right
  1         10.00   2.00     8.00
  2          5.00   0.00     5.00

                Within Node Probabilities
Class      Top      Left      Right
  1         0.625   1.000   0.571
  2         0.375   0.000   0.429

Competitor      Split                               Improve.
  1 X1           1                               0.003
  2 X5           2                               0.003

*****
*                Node 9: X5                *
*                N: 13                    *
*****

=====
=      Terminal Node 6      =      Terminal Node 7      =
=                N: 8      =                N: 5      =
=====

Node 9 was split on X5
A case goes left if X5 = (2)
Improvement = 0.008          Complexity Threshold = 0.015

Node   Cases   Wgt Counts      Cost Class
   9     13     13.00      0.429 1
  -6     8       8.00      0.286 1
  -7     5       5.00      0.357 2

                Weighted Counts
Class      Top      Left      Right
  1         8.00   6.00     2.00
  2          5.00   2.00     3.00

                Within Node Probabilities
Class      Top      Left      Right
  1         0.571   0.714   0.357
  2         0.429   0.286   0.643

Surrogate      Split                               Assoc.   Improve.
  1 X1          s 2                               0.179     0.004

Competitor      Split                               Improve.
  1 X1           1                               0.004

*****
*                Node 10: X4               *
*                N: 21                    *
*****

*****
=====

```

```

*           Node 11           *           =           Terminal Node 12           =
*           N: 15           *           =           N: 6           =
*****

```

Node 10 was split on X4
A case goes left if X4 = (2)
Improvement = 0.015 Complexity Threshold = 0.017

Node	Cases	Wgt	Counts	Cost	Class
10	21		21.00	0.431	2
11	15		15.00	0.444	1
-12	6		6.00	0.143	2

Class	Weighted Counts		
	Top	Left	Right
1	10.00	9.00	1.00
2	11.00	6.00	5.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.431	0.556	0.143
2	0.569	0.444	0.857

Surrogate	Split	Assoc.	Improve.
1 X6 s 1		0.514	.746380E-03
2 X2 s 1		0.343	0.009
3 X5 s 2		0.229	.746380E-03

Competitor	Split	Improve.
1 X2 1		0.009
2 X9 1		0.002
3 X5 2		.746380E-03
4 X6 1		.746380E-03
5 X8 2		.468880E-03

```

*****
*           Node 11: X5           *
*           N: 15           *
*****

```

```

=====
=           Terminal Node 9           =           *           Node 12           *
=           N: 2           =           *           N: 13           *
=====

```

Node 11 was split on X5
A case goes left if X5 = (1)
Improvement = 0.008 Complexity Threshold = 0.008

Node	Cases	Wgt	Counts	Cost	Class
11	15		15.00	0.444	1
-9	2		2.00	0.000	1
12	13		13.00	0.493	2

Class	Weighted Counts		
	Top	Left	Right
1	9.00	2.00	7.00
2	6.00	0.00	6.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.556	1.000	0.493
2	0.444	0.000	0.507

Surrogate	Split	Assoc.	Improve.
1 X6 s 2		0.500	0.004

Competitor	Split	Improve.
1 X6	2	0.004
2 X8	2	0.001
3 X9	1	.944822E-03
4 X11	2	.226665E-03


```

*****
*           Node 12: X9           *
*           N: 13                 *
*****
=====
= Terminal Node 10 = = Terminal Node 11 =
= N: 6             = = N: 7             =
=====

```

Node 12 was split on X9
A case goes left if X9 = (1)
Improvement = 0.004 Complexity Threshold = 0.015

Node	Cases	Wgt	Counts	Cost	Class
12	13		13.00	0.493	2
-10	6		6.00	0.375	1
-11	7		7.00	0.385	2

Weighted Counts			
Class	Top	Left	Right
1	7.00	4.00	3.00
2	6.00	2.00	4.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.493	0.625	0.385
2	0.507	0.375	0.615

Surrogate	Split	Assoc.	Improve.
1 X11	s 1	0.500	.336950E-03
2 X8	s 1	0.063	.174205E-03

Competitor	Split	Improve.
1 X11	1	.336950E-03
2 X8	2	.174205E-03


```

*****
*           Node 13: X4           *
*           N: 45                 *
*****
*****
*           Node 14             *
*           N: 19               *
*****
*****
*           Node 16             *
*           N: 26               *
*****

```

Node 13 was split on X4
A case goes left if X4 = (1)
Improvement = 0.044 Complexity Threshold = 0.054

Node	Cases	Wgt	Counts	Cost	Class
13	45		45.00	0.378	2
14	19		19.00	0.356	1
16	26		26.00	0.200	2

Weighted Counts

Class	Top	Left	Right
1	19.00	13.00	6.00
2	26.00	6.00	20.00

Within Node Probabilities

Class	Top	Left	Right
1	0.378	0.644	0.200
2	0.622	0.356	0.800

Surrogate	Split	Assoc.	Improve.
1 X11 s 1		0.317	0.028
2 X2 s 2		0.149	0.023
3 X8 s 1		0.139	0.004
4 X5 s 1		0.030	0.002

Competitor	Split	Improve.
1 X11 1		0.028
2 X2 2		0.023
3 X3 1		0.016
4 X9 1		0.014
5 X10 2		0.009

 * Node 14: X9 *
 * N: 19 *

```

=====
= Terminal Node 13 = * Node 15 *
= N: 8 = * N: 11 *
=====
*****
  
```

Node 14 was split on X9
 A case goes left if X9 = (1)
 Improvement = 0.031 Complexity Threshold = 0.024

Node	Cases	Wgt	Counts	Cost	Class
14	19		19.00	0.356	1
-13	8		8.00	0.000	1
15	11		11.00	0.410	2

Weighted Counts

Class	Top	Left	Right
1	13.00	8.00	5.00
2	6.00	0.00	6.00

Within Node Probabilities

Class	Top	Left	Right
1	0.644	1.000	0.410
2	0.356	0.000	0.590

Surrogate	Split	Assoc.	Improve.
1 X5 s 2		0.075	0.003

Competitor	Split	Improve.
1 X11 1		0.023
2 X6 2		0.009
3 X2 2		0.008
4 X5 2		0.003
5 X8 1		.222995E-03

 * Node 15: X2 *
 * N: 11 *

```

*****
=====
=      Terminal Node 14      =      Terminal Node 15      =
=      N: 3                  =      N: 8                  =
=====

Node 15 was split on X2
A case goes left if X2 = (2)
Improvement = 0.026          Complexity Threshold = 0.028

Node      Cases   Wgt Counts      Cost Class
  15         11      11.00      0.410 2
 -14         3       3.00      0.000 1
 -15         8       8.00      0.217 2

                          Weighted Counts
Class      Top      Left      Right
  1         5.00     3.00     2.00
  2         6.00     0.00     6.00

                          Within Node Probabilities
Class      Top      Left      Right
  1         0.410   1.000   0.217
  2         0.590   0.000   0.783

Competitor      Split      Improve.
  1 X11          1          0.025
  2 X6           2          0.001
  3 X3           2          .254754E-03
  4 X5           2          .254754E-03
  5 X8           1          .254754E-03

*****
*      Node 16: X3      *
*      N: 26           *
*****

*****
*      Node 17      *      Terminal Node 18      =
*      N: 14       *      N: 12                  =
*****

Node 16 was split on X3
A case goes left if X3 = (1)
Improvement = 0.021          Complexity Threshold = 0.014

Node      Cases   Wgt Counts      Cost Class
  16         26      26.00      0.200 2
  17         14      14.00      0.385 2
 -18         12      12.00      0.000 2

                          Weighted Counts
Class      Top      Left      Right
  1         6.00     6.00     0.00
  2        20.00     8.00    12.00

                          Within Node Probabilities
Class      Top      Left      Right
  1         0.200   0.385   0.000
  2         0.800   0.615   1.000

Surrogate      Split      Assoc.      Improve.
  1 X6          s 1          0.319      .101327E-03
  2 X8          s 2          0.181      .304310E-04
  3 X10         s 1          0.125      0.007

```

```

4 X5      s 2      0.125      0.003
5 X9      s 2      0.097      0.001

Competitor      Split      Improve.
1 X10          2      0.007
2 X5           1      0.003
3 X1           2      0.002
4 X9           2      0.001
5 X11          2      .552105E-03

*****
*              Node 17: X10          *
*              N: 14                 *
*****

=====
=      Terminal Node 16      =      Terminal Node 17      =
=      N: 3                  =      N: 11                  =
=====

Node 17 was split on X10
A case goes left if X10 = (2)
Improvement = 0.026      Complexity Threshold = 0.028

Node      Cases      Wgt Counts      Cost      Class
17        14        14.00      0.385 2
-16       3         3.00      0.000 1
-17       11        11.00      0.238 2

Weighted Counts
Class      Top      Left      Right
1          6.00      3.00      3.00
2          8.00      0.00      8.00

Within Node Probabilities
Class      Top      Left      Right
1          0.385      1.000      0.238
2          0.615      0.000      0.762

Surrogate      Split      Assoc.      Improve.
1 X8          s 1      0.333      0.007

Competitor      Split      Improve.
1 X5           1      0.012
2 X1           2      0.008
3 X6           2      0.008
4 X8           1      0.007
5 X9           2      .659364E-03

```

Lampiran G. Akurasi Hasil Klasifikasi Analisis CART *Bagging* Akurasi Data *Training* untuk Replikasi 10

```

=====
COMMITTEE SUMMARY
=====

% CORRECT
FIRST TREE ONLY      0.667
BOOTSTRAP SET        0.667

=====
BOOTSTRAP SET CLASSIFICATION TABLE
=====

```

```

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

Actual      Predicted Class
Class              1              2              Actual
-----
I 1              3.00              1.00              4.00
B                  3.00              1.00
I 2              1.00              1.00              2.00
B                  1.00              1.00
-----
INITIAL TREE
PRED. TOT.              4.00              2.00              6.00
CORRECT                0.750              0.500
SUCCESS IND.           0.083              0.167
Tot. Correct           0.667
-----
BOOTSTRAP SET OF TREES
PRED. TOT.              4.00              2.00
CORRECT                0.750              0.500
SUCCESS IND.           0.083              0.167
Tot. Correct           0.667
-----
INITIAL TREE
  SENSITIVITY:          0.750      SPECIFICITY:          0.500
FALSE REFERENCE:       0.250      FALSE RESPONSE:       0.500
REFERENCE = "1", RESPONSE = "2"
-----
BOOTSTRAP SET OF TREES
  SENSITIVITY:          0.750      SPECIFICITY:          0.500
FALSE REFERENCE:       0.250      FALSE RESPONSE:       0.500
-----

```

Akurasi Data *Training* untuk Replikasi 25

```

=====
COMMITTEE SUMMARY
=====
                                % CORRECT
FIRST TREE ONLY                0.667
BOOTSTRAP SET                   0.667

=====
BOOTSTRAP SET CLASSIFICATION TABLE
=====

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

Actual      Predicted Class
Class              1              2              Actual
-----
I 1              3.00              1.00              4.00
B                  3.00              1.00
I 2              1.00              1.00              2.00
B                  1.00              1.00
-----
INITIAL TREE
PRED. TOT.              4.00              2.00              6.00
CORRECT                0.750              0.500
SUCCESS IND.           0.083              0.167
Tot. Correct           0.667
-----
BOOTSTRAP SET OF TREES
PRED. TOT.              4.00              2.00
CORRECT                0.750              0.500
SUCCESS IND.           0.083              0.167
Tot. Correct           0.667
-----

```

```

INITIAL TREE
  SENSITIVITY:      0.750    SPECIFICITY:      0.500
  FALSE REFERENCE:  0.250    FALSE RESPONSE:   0.500
  REFERENCE = "1", RESPONSE = "2"
-----
BOOTSTRAP SET OF TREES
  SENSITIVITY:      0.750    SPECIFICITY:      0.500
  FALSE REFERENCE:  0.250    FALSE RESPONSE:   0.500
-----
  
```

Akurasi Data Training untuk Replikasi 50

```

=====
COMMITTEE SUMMARY
=====

                                % CORRECT
FIRST TREE ONLY                  0.667
BOOTSTRAP SET                    0.500

=====
BOOTSTRAP SET CLASSIFICATION TABLE
=====

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

Actual      Predicted Class
Class              1              2          Actual
-----
I 1              3.00            1.00          4.00
B                  2.00            2.00
I 2              1.00            1.00          2.00
B                  1.00            1.00
-----
INITIAL TREE
PRED. TOT.              4.00            2.00          6.00
CORRECT                 0.750            0.500
SUCCESS IND.            0.083            0.167
Tot. Correct           0.667
-----
BOOTSTRAP SET OF TREES
PRED. TOT.              3.00            3.00
CORRECT                 0.500            0.500
SUCCESS IND.           -0.167            0.167
Tot. Correct           0.500
-----
INITIAL TREE
  SENSITIVITY:      0.750    SPECIFICITY:      0.500
  FALSE REFERENCE:  0.250    FALSE RESPONSE:   0.500
  REFERENCE = "1", RESPONSE = "2"
-----
BOOTSTRAP SET OF TREES
  SENSITIVITY:      0.500    SPECIFICITY:      0.500
  FALSE REFERENCE:  0.333    FALSE RESPONSE:   0.667
-----
  
```

Akurasi Data Training untuk Replikasi 100

```

=====
COMMITTEE SUMMARY
=====

                                % CORRECT
FIRST TREE ONLY                  0.667
BOOTSTRAP SET                    0.833

=====
BOOTSTRAP SET CLASSIFICATION TABLE
=====
  
```

```

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

```

Actual Class	Predicted Class		Actual Total
	1	2	
I 1	3.00	1.00	4.00
B	4.00	0.00	
I 2	1.00	1.00	2.00
B	1.00	1.00	

INITIAL TREE			
PRED. TOT.	4.00	2.00	6.00
CORRECT	0.750	0.500	
SUCCESS IND.	0.083	0.167	
Tot. Correct	0.667		

BOOTSTRAP SET OF TREES			
PRED. TOT.	5.00	1.00	
CORRECT	1.000	0.500	
SUCCESS IND.	0.333	0.167	
Tot. Correct	0.833		

INITIAL TREE			
SENSITIVITY:	0.750	SPECIFICITY:	0.500
FALSE REFERENCE:	0.250	FALSE RESPONSE:	0.500
REFERENCE = "1", RESPONSE = "2"			

BOOTSTRAP SET OF TREES			
SENSITIVITY:	1.000	SPECIFICITY:	0.500
FALSE REFERENCE:	0.200	FALSE RESPONSE:	0.000

Akurasi Data *Training* untuk Replikasi 125

```

=====
COMMITTEE SUMMARY
=====

```

	% CORRECT
FIRST TREE ONLY	0.667
BOOTSTRAP SET	0.833

```

=====
BOOTSTRAP SET CLASSIFICATION TABLE
=====

```

```

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

```

Actual Class	Predicted Class		Actual Total
	1	2	
I 1	3.00	1.00	4.00
B	4.00	0.00	
I 2	1.00	1.00	2.00
B	1.00	1.00	

INITIAL TREE			
PRED. TOT.	4.00	2.00	6.00
CORRECT	0.750	0.500	
SUCCESS IND.	0.083	0.167	
Tot. Correct	0.667		

BOOTSTRAP SET OF TREES			
PRED. TOT.	5.00	1.00	
CORRECT	1.000	0.500	
SUCCESS IND.	0.333	0.167	
Tot. Correct	0.833		

```

-----
INITIAL TREE
  SENSITIVITY:      0.750      SPECIFICITY:      0.500
  FALSE REFERENCE:  0.250      FALSE RESPONSE:   0.500
  REFERENCE = "1", RESPONSE = "2"
-----
BOOTSTRAP SET OF TREES
  SENSITIVITY:      1.000      SPECIFICITY:      0.500
  FALSE REFERENCE:  0.200      FALSE RESPONSE:   0.000
-----

```

Akurasi Data *Training* untuk Replikasi 150

```

=====
COMMITTEE SUMMARY
=====
                                % CORRECT
FIRST TREE ONLY                 0.667
BOOTSTRAP SET                   0.833

=====
BOOTSTRAP SET CLASSIFICATION TABLE
=====

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

Actual      Predicted Class      Actual
Class              1          2          Total
-----
I 1              3.00         1.00         4.00
B                  4.00         0.00
I 2              1.00         1.00         2.00
B                  1.00         1.00

-----
INITIAL TREE
PRED. TOT.         4.00         2.00         6.00
CORRECT            0.750         0.500
SUCCESS IND.       0.083         0.167
Tot. Correct       0.667

-----
BOOTSTRAP SET OF TREES
PRED. TOT.         5.00         1.00
CORRECT            1.000         0.500
SUCCESS IND.       0.333         0.167
Tot. Correct       0.833

-----
INITIAL TREE
  SENSITIVITY:      0.750      SPECIFICITY:      0.500
  FALSE REFERENCE:  0.250      FALSE RESPONSE:   0.500
  REFERENCE = "1", RESPONSE = "2"
-----
BOOTSTRAP SET OF TREES
  SENSITIVITY:      1.000      SPECIFICITY:      0.500
  FALSE REFERENCE:  0.200      FALSE RESPONSE:   0.000
-----

```

Akurasi Data *Testing* untuk Replikasi 10

```

=====
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE
=====

Actual      Predicted Class      Actual
Class              1          2          Total
-----
1              9.00         2.00         11.00
2              4.00        10.00         14.00
-----
PRED. TOT.     13.00        12.00         25.00

```

CORRECT	0.818	0.714	
SUCCESS IND.	0.378	0.154	
TOT. CORRECT	0.760		
SENSITIVITY:	0.818	SPECIFICITY:	0.714
FALSE REFERENCE:	0.308	FALSE RESPONSE:	0.167
REFERENCE = "1", RESPONSE = "2"			

Akurasi Data *Testing* untuk Replikasi 25

=====				
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class		Actual Total	
	1	2		

1	9.00	2.00	11.00	
2	4.00	10.00	14.00	

PRED. TOT.	13.00	12.00	25.00	
CORRECT	0.818	0.714		
SUCCESS IND.	0.378	0.154		
TOT. CORRECT	0.760			
SENSITIVITY:	0.818	SPECIFICITY:	0.714	
FALSE REFERENCE:	0.308	FALSE RESPONSE:	0.167	
REFERENCE = "1", RESPONSE = "2"				

Akurasi Data *Testing* untuk Replikasi 50

=====				
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class		Actual Total	
	1	2		

1	10.00	1.00	11.00	
2	0.00	14.00	14.00	

PRED. TOT.	10.00	15.00	25.00	
CORRECT	0.909	1.000		
SUCCESS IND.	0.469	0.440		
TOT. CORRECT	0.960			
SENSITIVITY:	0.909	SPECIFICITY:	1.000	
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.067	
REFERENCE = "1", RESPONSE = "2"				

Akurasi Data *Testing* untuk Replikasi 100

=====				
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class		Actual Total	
	1	2		

1	9.00	2.00	11.00	
2	4.00	10.00	14.00	

PRED. TOT.	13.00	12.00	25.00	
CORRECT	0.818	0.714		
SUCCESS IND.	0.378	0.154		
TOT. CORRECT	0.760			

SENSITIVITY:	0.818	SPECIFICITY:	0.714
FALSE REFERENCE:	0.308	FALSE RESPONSE:	0.167
REFERENCE = "1", RESPONSE = "2"			

Akurasi Data *Testing* untuk Replikasi 125

CASE-BY-CASE SAMPLE CLASSIFICATION TABLE

Actual Class	Predicted Class		Actual Total
	1	2	
1	9.00	2.00	11.00
2	4.00	10.00	14.00
PRED. TOT.	13.00	12.00	25.00
CORRECT	0.818	0.714	
SUCCESS IND.	0.378	0.154	
TOT. CORRECT	0.760		

SENSITIVITY:	0.818	SPECIFICITY:	0.714
FALSE REFERENCE:	0.308	FALSE RESPONSE:	0.167
REFERENCE = "1", RESPONSE = "2"			

Akurasi Data *Testing* untuk Replikasi 150

CASE-BY-CASE SAMPLE CLASSIFICATION TABLE

Actual Class	Predicted Class		Actual Total
	1	2	
1	9.00	2.00	11.00
2	4.00	10.00	14.00
PRED. TOT.	13.00	12.00	25.00
CORRECT	0.818	0.714	
SUCCESS IND.	0.378	0.154	
TOT. CORRECT	0.760		

SENSITIVITY:	0.818	SPECIFICITY:	0.714
FALSE REFERENCE:	0.308	FALSE RESPONSE:	0.167
REFERENCE = "1", RESPONSE = "2"			

Lampiran H. Surat Keterangan Penelitian

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini menerangkan bahwa:

1. Mahasiswa Statistika FMIPA-ITS dengan identitas berikut:

Nama : Adheala Nathasya Wisaksono
NRP : 1313 100 013

menyatakan bahwa data yang digunakan dalam Tugas Akhir/Thesis ini merupakan data sekunder yang diambil dari penelitian / buku / Tugas Akhir / Thesis / publikasi lainnya yaitu:

Sumber : Analisis Ketahanan Pangan Rumah Tangga Terhadap Penderita Penyakit *Tuberculosis* dengan Pendekatan *Geographically Weighted Poisson Regression* (Studi Kasus Pantai Pesisir Surabaya)

Keterangan : Penelitian diketuai oleh Ir. Sri Pingit Wulandari, M.Si sejak tanggal 01 Mei 2016 sampai dengan 12 Mei 2016 untuk keperluan Tugas Akhir/Thesis Semester Gasal/ Genap* 2016/2017

2. Tidak Keberatan/Keberatan* nama perusahaan dicantumkan dalam Tugas Akhir/Thesis mahasiswa Statistika yang akan di dimpan di Perpustakaan ITS dan dibaca di lingkungan ITS.
3. Tidak Keberatan/Keberatan* bahwa hasil analisis data dari perusahaan dipublikasikan dalam E journal ITS yaitu Jurnal Sains dan Seni ITS.

Mengetahui
Dosen Pembimbing Tugas Akhir



Dr. Bambang Widjanarko Otok, M.Si
NIP: 19681124 199412 1 001

Surabaya, Mei 2017
Menyetujui
Ketua Peneliti



Ir. Sri Pingit Wulandari, M.Si
NIP: 19620603 198701 2 001

*(coret yang tidak perlu)

“Halaman ini sengaja dikosongkan”

BIODATA PENULIS



Penulis dilahirkan di Banyuwangi, 25 Februari 1995 dengan nama lengkap Adheala Nathasya Wisaksono, biasa dipanggil Dhea atau Adheala. Penulis merupakan anak pertama dari pasangan (Alm.) Wisaksono dan Siti Nurhasanah. Pendidikan formal yang ditempuh penulis adalah SDN Karangrejo II, SMP Negeri 3 Jember, dan SMA Negeri 1 Jember. Pada tahun 2013, penulis diterima di Jurusan Statistika ITS melalui jalur SNMPTN. Semasa kuliah penulis aktif dalam beberapa organisasi diantaranya, Staf

Divisi PERS HIMASTA-ITS (2014-2015) dan Ketua Divisi PERS HIMASTA-ITS (2015-2016). Kepanitiaan yang dijalani penulis semasa kuliah diantaranya Staf Acara INTERVAL, Staf Konsumsi POSEIDON, Staf Acara CERITA PRS, Ketua Panitia Creative Writing Talkshow dan lain-lain. Penulis menerima segala kritikan, masukan, dan saran yang bersifat membangun demi meningkatkan manfaat Tugas Akhir ini. Oleh karena itu, penulis siap membantu terkait dengan metode dan ilmu statistik, jika dibutuhkan pembaca dapat menghubungi penulis melalui email: adhealanathasyal@gmail.com.