



TUGAS AKHIR – SS141501

**KLASIFIKASI KETAHANAN PANGAN RUMAH
TANGGA DI SURABAYA DENGAN PENDEKATAN
CART ARCING**

**YUSNADA ASA NURANI
NRP 1313 100 016**

**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**



TUGAS AKHIR – SS141501

**KLASIFIKASI KETAHANAN PANGAN RUMAH
TANGGA DI SURABAYA DENGAN PENDEKATAN
CART ARCING**

**YUSNADA ASA NURANI
NRP 1313 100 016**

**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 HOUSEHOLDS FOOD
SECURITY IN SURABAYA USING CART ARCING**

**YUSNADA ASA NURANI
NRP 1313 100 016**

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

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

LEMBAR PENGESAHAN

KLASIFIKASI KETAHANAN PANGAN RUMAH TANGGA DI SURABAYA DENGAN PENDEKATAN CART ARCING

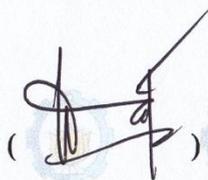
TUGAS AKHIR

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 :
YUSNADA ASA NURANI
NRP. 1313 100 016

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



KLASIFIKASI KETAHANAN PANGAN RUMAH TANGGA DI SURABAYA DENGAN PENDEKATAN CART ARCING

Nama Mahasiswa : Yusnada Asa Nurani
NRP : 1313 100 016
Departemen : Statistika
Dosen Pembimbing : Dr. Bambang Widjanarko O., M. Si.

Abstrak

Sumber daya pangan Indonesia secara umum dapat dikatakan cukup lengkap dan beragam. Dengan keragaman dan potensi pangan yang lengkap, Indonesia seharusnya mampu memenuhi kebutuhan pangan penduduk. Namun kenyataannya pengelolaan potensi pangan di Indonesia masih belum maksimal. Hal tersebut menyebabkan kondisi ketahanan pangan nasional masih jauh dari yang diharapkan. Selain itu, rumah tangga dengan penderita tuberkulosis di Surabaya cenderung memiliki tingkat ketahanan pangan rumah tangga yang rendah. Oleh karena itu, pada penelitian ini dilakukan klasifikasi ketahanan pangan rumah tangga penderita tuberkulosis di Surabaya dengan metode Classification and Regression Tree (CART) dan pendekatan Adaptive Resampling and Combining (ARCING) untuk mengetahui ketepatan klasifikasi dan faktor yang berpengaruh dalam proses klasifikasi. Hasil penelitian menunjukkan nilai ketepatan klasifikasi yang dihasilkan dengan metode CART adalah sebesar 59,6% untuk data training dan 52,0% untuk data testing. Angka tersebut naik 7,1% untuk data training dan 4,0% untuk data testing menjadi 66,7% dan 56,0% ketika pendekatan ARCING digunakan. Hasil penelitian juga menunjukkan bahwa cara suatu rumah tangga memperoleh makanan merupakan variabel yang paling berpengaruh dalam proses klasifikasi ketahanan pangan rumah tangga di Surabaya.

Kata kunci: CART, CART ARCING, Ketahanan Pangan

(halaman ini sengaja dikosongkan)



ABSTRACT



CLASSIFICATION OF HOUSEHOLD FOOD SECURITY IN SURABAYA USING CART ARCING

Name of Student : Yusnada Asa Nurani
Student Number : 1313 100 016
Department : Statistics
Supervisor : Dr. Bambang Widjanarko O., M. Si.

Abstract

Food resource in Indonesia generally is heterogeneous and complete. By this heterogeneity and big resource of food, Indonesia shall fulfill the residents need of food. But the fact is maintenance of food resource in Indonesia has not been maximized yet. This condition causes Indonesian food security is still far from expectation. Meanwhile, households with tuberculosis patients in Surabaya tend to classify as food-insecure. This study conduct a classification of households with tuberculosis patients food security in Surabaya by using Classification and Regression Tree (CART) and Adaptive Resampling and Combining (ARCING) to know the accuracy and influential factors of households food security. As for the result of this study, data training yields accuracy about 59,6% while data testing yields accuracy about 52,0%. While using ARCING, data training yields accuracy about 66,7% and 56,0% data testing yields about 56,0% in accuracy. It can be said that the accuracy is increasing by 7,1% for data training and 4,0% for data testing. As for the influential factors of this study is the way households obtain foods.

Keywords: *CART, CART ARCING, Food Security*

(halaman ini sengaja dikosongkan)



KATA PENGANTAR



KATA PENGANTAR

Assalamu'alaikum Warahmatullah Wabarakatuh.

Puji syukur senantiasa penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat, hidayah dan karunia-Nya sehingga penulis dapat Pada kesempatan ini, dengan segala kerendahan hati penulis mengucapkan terima kasih yang sebesar-besarnya kepada pihak-pihak yang telah membantu dan mendukung keberhasilan penyelesaian laporan ini, yakni:

1. Ibu, Mbak Ninis, Mas Rodli, Falih, Farras dan keluarga besar penulis atas do'a, semangat, dan motivasi dalam menjalani proses perkuliahan dan menyelesaikan Tugas Akhir ini.
2. Bapak Dr. Bambang Widjanarko Otok, M.Si selaku dosen pembimbing yang senantiasa memberikan ilmu, perhatian, bimbingan, dan pengarahan selama penulis menyelesaikan Tugas Akhir ini.
3. Ibu Dr. Irhamah, S.Si, M.Si, Bapak Dr. Sutikno, S.Si., M.Si, dan Ibu Dr. Santi Wulan Purnami, S.Si, M.Si selaku tim penguji yang telah memberikan ilmu, kritik, dan saran membangun untuk kesempurnaan Tugas Akhir ini.
4. Ibu Sri Pingit Wulandari, M.Si yang meluangkan waktunya untuk membantu dan memberikan masukan dalam penyelesaian Tugas Akhir ini.
5. Serta semua pihak yang telah memberikan dukungan baik moril maupun materiil yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari masih banyaknya kekurangan pada pembuatan laporan Tugas Akhir ini, oleh karena itu kritik dan saran membangun sangat diharapkan. Semoga Tugas Akhir ini dapat memberikan manfaat baik bagi penulis, pembaca, dan semua pihak.

Wassalamu'alaikum Warahmatullah Wabarakatuh.

Surabaya, Juli 2017

Yusnada Asa Nurani

(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 GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	
2.1 <i>Classification and Regression Trees (CART)</i>	7
2.1.1 Pembentukan Pohon Klasifikasi.....	9
2.1.2 Pemangkasan Pohon Klasifikasi.....	12
2.1.3 Penentuan Pohon Klasifikasi Optimal.....	13
2.2 Ilustrasi CART.....	15
2.3 <i>Adaptive and Resampling Combining (ARCING)</i>	20
2.4 Ilustrasi ARCING	21
2.5 Ukuran Ketepatan Klasifikasi.....	22
2.6 Ketahanan Pangan	23
BAB III METODOLOGI PENELITIAN	
3.1 Sumber Data	29

3.2	Kerangka Konsep Penelitian.....	29
3.3	Variabel Penelitian.....	31
3.4	Langkah Analisis	32
3.5	Diagram Alir.....	34

BAB IV ANALISIS DAN PEMBAHASAN

4.1	Karakteristik Faktor-Faktor Ketahanan Pangan Rumah Tangga di Surabaya	37
4.2	Klasifikasi Ketahanan Pangan Rumah Tangga di Surabaya dengan <i>Classification Trees</i>	40
4.2.1	Kemungkinan Pemilahan untuk Setiap Variabel Prediktor	41
4.2.2	Pembentukan Pohon Klasifikasi Maksimal	42
4.2.3	Pemangkasan Pohon Klasifikasi Maksimal	44
4.2.4	Penentuan Pohon Klasifikasi Optimal	46
4.2.5	Perhitungan Hasil Ketepatan Klasifikasi <i>Classification Trees</i>	50
4.3	Klasifikasi Ketahanan Pangan Rumah Tangga di Surabaya dengan CART ARCING	51
4.4	Perbandingan Hasil Klasifikasi CART dan CART ARCING	52

BAB V KESIMPULAN DAN SARAN

5.1	Kesimpulan	55
5.2	Saran	56

DAFTAR PUSTAKA LAMPIRAN



DAFTAR GAMBAR



DAFTAR GAMBAR

	Halaman
Gambar 2.1 Stuktur Pohon Klasifikasi	9
Gambar 2.2 Ilustrasi CART	20
Gambar 3.1 Kerangka Konsep Penelitian	30
Gambar 3.2 Diagram Alir Penelitian	34
Gambar 4.1 Persediaan Makanan Pokok Rumah Tangga di Surabaya	38
Gambar 4.2 Akses Fisik, Sosial, dan Ekonomi Rumah Tang- ga di Surabaya	39
Gambar 4.3 Jenis Protein yang Dikonsumsi Rumah Tangga di Surabaya	40
Gambar 4.4 Pohon Klasifikasi Maksimal	43
Gambar 4.5 Plot <i>Relative Cost</i>	45
Gambar 4.6 Pohon Klasifikasi Optimal	46

(halaman ini sengaja dikosongkan)



DAFTAR TABEL



DAFTAR TABEL

	Halaman
Tabel 2.1 Data Ilustrasi CART	16
Tabel 2.2 Jenis Pemilah Ilustrasi CART	16
Tabel 2.3 Proporsi Pengamatan pada Variabel Respon Y	17
Tabel 2.4 Pemilahan Variabel Prediktor X_2	17
Tabel 2.5 Pemilahan Variabel Prediktor X_6	18
Tabel 2.6 <i>Goodness of Split</i> Setiap Pemilah	19
Tabel 2.7 Ketersediaan Pangan Rumah Tangga	25
Tabel 2.8 Stabilitas Ketersediaan Pangan.....	25
Tabel 2.9 Aksesibilitas atau Keterjangkauan Pangan	25
Tabel 2.10 Kontinuitas Ketersediaan Pangan	26
Tabel 3.1 Stuktur Data	29
Tabel 3.2 Variabel Penelitian	31
Tabel 4.1 Kemungkinan Pemilah	41
Tabel 4.2 Skor Kontribusi Variabel Prediktor	42
Tabel 4.3 Urutan Pembentukan Pohon Klasifikasi	45
Tabel 4.4 Label Kelas Simpul Terminal	48
Tabel 4.5 Ketepatan Klasifikasi CART	51
Tabel 4.6 Ketepatan Klasifikasi CART ARCING	52
Tabel 4.7 Perbandingan Ketepatan Klasifikasi	52

(halaman ini sengaja dikosongkan)



DAFTAR LAMPIRAN



DAFTAR LAMPIRAN

	Halaman
Lampiran A Data Ketahanan Pangan Rumah Tangga di Surabaya.....	59
Lampiran B <i>Output Variable Importance</i>	60
Lampiran C <i>Output Cross Validation Classification Table</i> ...	60
Lampiran D <i>Output Tree Sequence</i>	61
Lampiran E <i>Output Missclasification</i>	61
Lampiran F Informasi Simpul Terminal.....	62
Lampiran G <i>Output</i> Ketepatan Klasifikasi CART untuk Data <i>Training</i>	63
Lampiran H <i>Output</i> Ketepatan Klasifikasi CART untuk Data <i>Testing</i>	64
Lampiran I <i>Output</i> Ketepatan Klasifikasi CART 25 Replikasi Data <i>Training</i>	65
Lampiran J <i>Output</i> Ketepatan Klasifikasi CART 50 Replikasi Data <i>Training</i>	66
Lampiran K <i>Output</i> Ketepatan Klasifikasi CART 75 Replikasi Data <i>Training</i>	67
Lampiran L <i>Output</i> Ketepatan Klasifikasi CART 100 Replikasi Data <i>Training</i>	68
Lampiran M <i>Output</i> Ketepatan Klasifikasi CART 125 Replikasi Data <i>Training</i>	69
Lampiran N <i>Output</i> Ketepatan Klasifikasi CART 150 Replikasi Data <i>Training</i>	70
Lampiran O <i>Output</i> Ketepatan Klasifikasi CART ARCING 25 Replikasi Data <i>Testing</i>	71
Lampiran P <i>Output</i> Ketepatan Klasifikasi CART ARCING 50 Replikasi Data <i>Testing</i>	71
Lampiran Q <i>Output</i> Ketepatan Klasifikasi CART ARCING 75 Replikasi Data <i>Testing</i>	72
Lampiran R <i>Output</i> Ketepatan Klasifikasi CART ARCING 100 Replikasi Data <i>Testing</i>	72
Lampiran S <i>Output</i> Ketepatan Klasifikasi CART ARCING	

	125 Replikasi Data <i>Testing</i>	73
Lampiran T	<i>Output</i> Ketepatan Klasifikasi CART ARCING	
	150 Replikasi Data <i>Testing</i>	73
Lampiran U	Surat Keterangan Data Sekunder	74



BAB I
PENDAHULUAN



BAB I

PENDAHULUAN

1.1 Latar Belakang

Meningkatnya jumlah penduduk dunia turut meningkatkan kebutuhan atas makanan, tempat tinggal, energi, dan kebutuhan-kebutuhan lainnya. Indonesia sebagai negara berkembang kini menempati peringkat keempat penduduk terbanyak dunia (detikFinance, 2014). Jumlah penduduk Indonesia hingga tahun 2015 mencapai 255 juta jiwa (Informasipedia, 2015). Seiring dengan peningkatan jumlah penduduk tentunya diperlukan pengelolaan sumber daya pangan yang memadai sehingga kebutuhan pangan penduduk terpenuhi.

Sejauh ini pengelolaan potensi pangan yang belum maksimal menyebabkan kondisi ketahanan pangan nasional masih jauh dari yang diharapkan. Menurut Menteri Pertanian Andi Amran dalam *Global Food Security Nilai Ketahanan Pangan Indonesia Tertinggi di Dunia*, Indonesia telah menunjukkan peringkat ketahanan pangan yang cukup baik dengan berada di posisi ke-71 dari 113 negara yang diobservasi pada 2016 berdasarkan data *Global Food Security Index (GFSI)* yang dirilis *The Economist Intelligence Unit* (Julianto, 2016). Namun, kondisi baik yang dinyatakan Menteri Pertanian ini belum sesuai dengan keadaan di lapangan. Setidaknya terdapat 19,4 juta penduduk Indonesia yang masih mengalami kelaparan (PoskotaNews, 2016). Seharusnya keadaan ini tidak terjadi karena Indonesia memiliki sumber daya pangan yang beraneka ragam. Terdapat setidaknya 77 jenis sumber karbohidrat, 26 jenis kacang-kacangan, 389 jenis buah-buahan, 228 jenis sayur-sayuran, 110 jenis rempah-rempahan dan bumbu-bumbuan, 40 jenis bahan minuman serta 1.260 jenis tanaman obat (Nugrayasa, 2015).

Konsep pangan di Indonesia secara jelas diatur dalam Undang-undang Nomor 7 tahun 1996 tentang Pangan, Undang-undang Nomor 18 tahun 2012 tentang Ketahanan Pangan dan Peraturan Pemerintah Nomor 28 tahun 2004 tentang Keamanan,

Mutu dan Gizi Pangan. Ketahanan pangan dalam Undang-undang Nomor 7 tahun 1996 dijelaskan sebagai kondisi terpenuhinya pangan bagi rumah tangga yang tercermin dari tersedianya pangan yang cukup, baik jumlah maupun mutunya, aman, merata, dan terjangkau. Sedangkan dalam Undang-undang Nomor 18 tahun 2012 dijelaskan sebagai kondisi terpenuhinya pangan bagi negara sampai dengan perseorangan yang tercermin dari tersedianya pangan yang cukup, baik jumlah maupun mutunya, aman, beragam, bergizi, merata, dan terjangkau serta tidak bertentangan dengan agama, keyakinan, dan budaya masyarakat untuk dapat hidup sehat, aktif, dan produktif secara berkelanjutan. Sehingga secara tidak langsung dapat disimpulkan bahwa ketahanan pangan merupakan kondisi yang perlu dicapai untuk menjamin kelangsungan hidup penduduk Indonesia.

Food and Agriculture Organization (FAO) menggambarkan ketahanan pangan melalui empat aspek utama yakni ketersediaan pangan, stabilitas ketersediaan atau pemerataan pangan, aksesibilitas atau keterjangkauan pangan, dan kualitas atau keamanan pangan. Ketersediaan pangan dalam rumah tangga yang dipakai dalam pengukuran mengacu pada pangan yang cukup dan tersedia dalam jumlah yang dapat memenuhi kebutuhan konsumsi rumah tangga dalam waktu satu bulan atau satu tahun. Stabilitas ketersediaan pangan di tingkat rumah tangga diukur berdasarkan kecukupan ketersediaan pangan dan frekuensi makan anggota rumah tangga dalam sehari. Suatu rumah tangga dikatakan memiliki stabilitas ketersediaan pangan yang baik jika mempunyai persediaan pangan lebih dari *cutting point* serta anggota rumah tangga dapat makan tiga kali atau lebih dalam sehari sesuai dengan kebiasaan makan penduduk di daerah tersebut. Indikator aksesibilitas atau keterjangkauan terhadap pangan dapat diukur dengan cara melihat kemudahan rumah tangga dalam memperoleh pangan yang diukur berdasarkan akses fisik, akses sosial, dan akses ekonomi. Berdasarkan pengukuran indikator stabilitas ketersediaan pangan dan aksesibilitas pangan dapat diukur kontinuitas ketersediaan pangan rumah tangga.

Pengukuran indikator yang terakhir adalah kualitas jenis pangan yang diukur dengan cara melihat data pengeluaran untuk konsumsi makanan (lauk pauk) sehari-hari yang mengandung protein. Berdasarkan keempat aspek utama tersebut, tingkat ketahanan pangan suatu rumah tangga dapat dibedakan menjadi dua kategori, yakni rumah tangga tahan pangan dan rumah tangga rawan pangan (Pramita, 2016).

Wulandari, Susilaningrum, & Latra (2016) melakukan penelitian tentang analisis ketahanan pangan terhadap penderita tuberkulosis di Surabaya. Hasilnya menunjukkan terdapat 56% rumah tangga yang dikategorikan sebagai rumah tangga rawan pangan dan 44% sisanya dikategorikan sebagai rumah tangga tahan pangan. Hal ini menunjukkan bahwa rumah tangga dengan penderita tuberkulosis di Surabaya cenderung tergolong dalam rumah tangga rawan pangan. Padahal, tingkat ketahanan pangan rumah tangga bagi suatu daerah dianggap menjadi dasar yang cukup penting untuk melakukan pengembangan pada daerah tersebut dalam hal kesejahteraan masyarakat.

Ketahanan pangan memiliki faktor-faktor yang diduga saling berhubungan, sehingga diperlukan suatu metode klasifikasi yang bersifat nonparametrik. Salah satu metode klasifikasi yang bersifat nonparameterik adalah *Classification and Regression Trees* (CART). CART merupakan salah satu metode untuk pengklasifikasian yang dilakukan dengan teknik pohon keputusan. Metode ini memiliki beberapa kelebihan diantaranya adalah tidak memiliki asumsi yang harus dipenuhi (Lewis, 2000). Selain itu, terdapat algoritma *Adaptive Resampling and Combining* (ARCING) yang merupakan suatu metode yang dapat mengecilkan kasus kesalahan klasifikasi. Metode ini dilakukan dengan cara melakukan *resampling* pada data *training* dengan peluang pengambilan tertentu (Breiman, 1998). Widyandoro (2011) melakukan penelitian mengenai klasifikasi kesejahteraan rumah tangga di Provinsi Jawa Timur dengan pendekatan CART ARCING. Hasil klasifikasi dengan menggunakan CART diperoleh 34 kelompok rumah tangga miskin dengan ketepatan kla-

sifikasi 70,8% dan klasifikasi yang dilakukan dengan metode CART ARCING menghasilkan ketepatan klasifikasi yang lebih tinggi yakni 78,4%.

Pada penelitian ini digunakan metode *Classification and Regression Trees* (CART) dengan pendekatan *Adaptive Resampling and Combining* (ARCING) untuk mengetahui klasifikasi ketahanan pangan rumah tangga di kota Surabaya berdasarkan faktor-faktor pembentuk indikatornya.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang tersebut, diajukan rumusan masalah yakni bagaimana karakteristik dan hasil klasifikasi ketahanan pangan rumah tangga di Surabaya, serta faktor-faktor apa saja yang berpengaruh terhadap ketahanan pangan rumah tangga di Surabaya dengan menggunakan metode CART ARCING.

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Mengetahui karakteristik faktor-faktor ketahanan pangan rumah tangga di Surabaya.
2. Memperoleh hasil klasifikasi ketahanan pangan rumah tangga di Surabaya.
3. Mengetahui faktor-faktor yang berpengaruh terhadap ketahanan pangan rumah tangga di Surabaya dengan menggunakan metode CART ARCING.

1.4 Manfaat Penelitian

Manfaat bagi pemerintah dan bidang pendidikan yang diharapkan dalam pelaksanaan penelitian ini adalah sebagai berikut.

1. Bagi pemerintah, diharapkan penelitian ini dapat membantu pemerintah daerah untuk melakukan peningkatan ketahanan pangan rumah tangga di Surabaya.

2. Bagi bidang pendidikan, diharapkan penelitian ini dapat menambah wawasan keilmuan dalam penerapan metode CART ARCING.

1.5 Batasan Masalah

Penelitian ini menggunakan batasan masalah yaitu data yang digunakan merupakan data ketahanan pangan rumah tangga penderita tuberkulosis di Surabaya tahun 2016 dengan cakupan wilayah yang digunakan adalah kecamatan Bubutan, Dukuh Pakis, Genteng, Gubeng, Sawahan, Simokerto, Sukomanunggal, Tambaksari, Tegalsari, dan Wonokromo.

(halaman ini sengaja dikosongkan)



BAB II
TINJAUAN PUSTAKA



BAB II TINJAUAN PUSTAKA

2.1 *Classification and Regression Trees (CART)*

Classification and Regression Trees (CART) adalah salah satu metode pengklasifikasian yang dilakukan dengan teknik pohon keputusan (*decision trees*). Menurut Breiman, Friedman, Stone, & Olshen (1984) model yang dihasilkan berdasarkan pada skala variabel respon, jika respon data memiliki skala interval dan rasio maka model pohon yang dihasilkan adalah pohon regresi (*regression trees*), sedangkan jika respon data memiliki skala nominal dan ordinal maka pohon yang dihasilkan adalah pohon klasifikasi (*classification trees*). CART memiliki sifat “*binary recursive partitioning*”.

Istilah “*binary*” mendefinisikan suatu simpul induk (*parent node*) yang merupakan suatu variabel prediktor terpenting yang digunakan untuk memilah dan akan menghasilkan dua simpul anak baru (*child nodes*). Istilah “*recursive*” menjelaskan proses yang terjadi dalam prosedur pembagian secara biner yang akan dilakukan secara berulang-ulang, artinya dari simpul anak yang sudah dihasilkan tersebut akan menjadi dua simpul induk baru dan akan dilakukan pemilahan lagi dan menghasilkan dua simpul anak yang baru yang lebih homogen. Begitu seterusnya sampai memenuhi kriteria tertentu sehingga sudah tidak dapat dibagi lagi. Sedangkan istilah “*partitioning*” menjelaskan bahwa proses klasifikasi dilakukan dengan cara memilah suatu himpunan data menjadi bagian-bagian atau partisi-partisi (Lewis, 2000).

Menurut Lewis (2000), metode klasifikasi CART memiliki beberapa kelebihan, diantaranya adalah sebagai berikut.

1. Tidak memiliki asumsi yang harus dipenuhi.
2. Memudahkan dalam hal eksplorasi dan pengambilan keputusan pada struktur data yang kompleks dan multivariabel karena struktur data dapat dilihat secara visual.

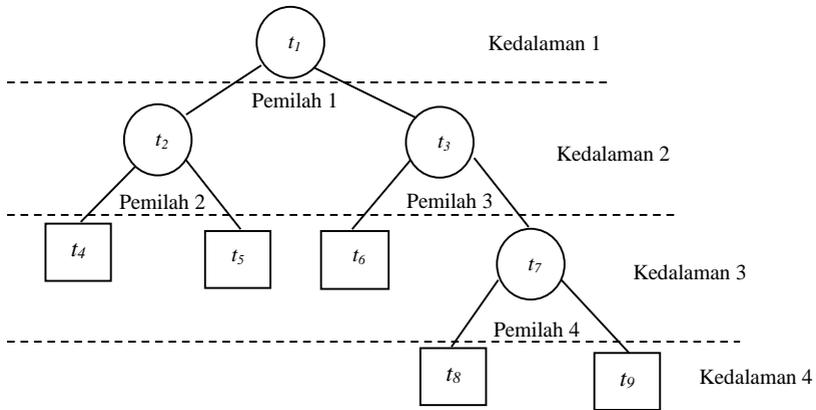
3. Hasil klasifikasi akhir berbentuk sederhana dan dapat mengklasifikasikan data baru secara lebih efisien serta mudah diinterpretasikan.

Namun, Sutton (2005) menyebutkan bahwa metode CART memiliki kekurangan yakni pohon yang terbentuk tidak stabil atau dengan kata lain ketika terjadi sedikit perubahan pada data *training* maka hasil prediksi pohon yang diperoleh dapat mengalami perbedaan yang cukup besar. Sehingga, hasil klasifikasi yang terbentuk bukan merupakan hasil klasifikasi yang paling optimal.

Lewis (2000) mengungkapkan bahwa struktur CART diawali dengan pembentukan simpul utama (*root node*) sebagai pemilah terpenting dalam menentukan klasifikasi pengamatan yang menghimpun seluruh pengamatan dari data *training* yang digunakan. Simpul utama tersebut juga menjadi simpul induk (*parent node*) pertama yang akan dipilah mejadi dua simpul baru atau disebut simpul anak (*child node*) kiri dan kanan. Selanjutnya kedua simpul anak tersebut berubah menjadi dua simpul induk dan dipilah kembali menjadi dua simpul anak baru. Langkah tersebut dilakukan berulang-ulang sampai terbentuk simpul akhir (*terminal node*) yang sudah tidak bisa dipilah lagi, hal ini menunjukkan bahwa simpul anak yang dihasilkan sudah homogen.

Gambar 2.1 merupakan suatu struktur pohon yang menggambarkan hasil proses pemilahan yang dilakukan. t_1 merupakan simpul utama yang mengandung semua data. t_2 , t_3 , dan t_7 merupakan simpul dalam (*internal nodes*) atau simpul anak yang kemudian menjadi simpul induk. t_4 , t_5 , t_6 , t_8 , dan t_9 merupakan simpul akhir (*terminal nodes*) yang sudah homogen. Simpul-simpul tersebut berada dalam tingkat kedalaman (*depth*) tertentu.

Secara umum, terdapat tiga tahapan dalam algoritma CART, yakni pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi, dan penentuan pohon klasifikasi optimal. Pada pembentukan pohon klasifikasi dilakukan pemilihan pemilah, penentuan simpul terminal, dan pemberian label kelas. Kemudian dilanjutkan dengan pemangkasan pohon klasifikasi untuk menentukan pohon klasifikasi yang optimal.



Gambar 2.1 Struktur Pohon Klasifikasi

2.1.1 Pembentukan Pohon Klasifikasi

Terdapat 3 tahap dalam pembentukan pohon klasifikasi. Tahap pertama adalah pemilihan pemilah. Setiap pemilahan hanya bergantung pada nilai yang berasal dari satu variabel independen. Jika variabel independen yang digunakan (X_j) memiliki jenis data kontinyu dengan ruang sampel berukuran n dan terdapat n nilai amatan sampel yang berbeda, maka akan terbentuk $n - 1$ jenis pemilahan yang berbeda. Jika variabel independen yang digunakan memiliki jenis data kategori nominal bertaraf L , maka akan terbentuk $2^{L-1} - 1$ pemilahan yang berbeda. Namun jika variabel independen yang digunakan memiliki jenis data kategori ordinal bertaraf L , maka akan diperoleh $L - 1$ jenis pemilahan. Metode pemilihan pemilah masih bersifat coba-coba, karena belum ada ketentuan tentang metode pemilihan pemilah yang terbaik (Breiman, *et al*, 1984). Terdapat beberapa macam metode pemilahan yang digunakan dalam CART, diantaranya indeks Gini, indeks Informasi, indeks Twoing, dan indeks Entropy. Metode pemilahan yang sering digunakan adalah indeks Gini dengan fungsi sebagai berikut.

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

dengan

$I(t)$ = fungsi heterogenitas simpul t

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

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

Dilakukan pemilihan pemilah terbaik setelah didapatkan beberapa macam jenis pemilahan. Pemilihan pemilah tersebut dilakukan berdasarkan kriteria *goodness of split*, yakni suatu nilai yang dijadikan patokan pemilahan oleh suatu pemilah s pada suatu simpul t . *Goodness of split* menunjukkan ukuran penurunan heterogenitas suatu kelas dengan fungsi sebagai berikut.

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

dengan

$\phi(s, t)$ = 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

Pemilah yang menghasilkan nilai *goodness of split* tertinggi merupakan pemilah terbaik karena mampu mereduksi heterogenitas lebih tinggi. Pemilah terbaik tersebut dapat menunjukkan variabel penting yang akan menghasilkan skor variabel untuk menunjukkan seberapa besar kontribusi variabel tersebut dalam proses pembentukan pohon. Berikut merupakan persamaan yang digunakan untuk menentukan skor pada variabel penting.

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

dengan $\phi(s, t_i)$ merupakan nilai *goodness of split* pada setiap simpul (*improvement*). Nilai skor tersebut diperoleh dengan menjumlahkan *goodness of split* dari masing-masing variabel yang berperan sebagai pengganti (*surrogate*) untuk setiap simpul (q).

Tahap kedua dari pembentukan pohon klasifikasi adalah penentuan simpul terminal. Dalam tahap ini akan diputuskan apakah simpul t akan dipilah lagi atau dapat dijadikan simpul terminal berdasarkan kriteria tertentu. Berikut merupakan kriteria yang digunakan untuk memutuskan suatu simpul tidak akan dipilah lagi atau menjadi simpul terminal.

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

Tahap terakhir adalah pemberian label kelas pada tiap simpul terminal. Penandaan tersebut perlu dilakukan agar nantinya dapat diketahui karakteristik dari klasifikasi pengamatan untuk setiap kelas variabel respon yang terbentuk. Pemberian label kelas pada simpul terminal ditentukan berdasarkan aturan jumlah terbanyak, yakni bila

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

dengan

$p(j | t)$ = 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

maka label kelas untuk simpul terminal t adalah j_0 . Asumsi yang digunakan adalah nilai kesalahan klasifikasi untuk setiap kelas besarnya sama.

Selain menggunakan aturan jumlah terbanyak, pelabelan kelas pada suatu simpul dalam CART dilakukan dengan mengacu pada kategori variabel respon yang memiliki jumlah paling sedikit. Misalnya, ketika pada suatu simpul terminal terdapat sampel amatan dalam jumlah yang sama untuk seluruh kategori, maka simpul

terminal tersebut dilabelkan sebagai simpul terminal dengan kategori variabel respon paling sedikit (Stenberg & Golovnya, 2007).

2.1.2 Pemangkasan Pohon Klasifikasi

Pohon klasifikasi yang terbentuk pertama kali merupakan pohon klasifikasi maksimal atau pohon klasifikasi yang berukuran sangat besar. Hal ini dikarenakan pohon yang dibentuk dengan aturan pemilah dan kriteria *goodness of split* penghentiannya didasarkan oleh banyak amatan pada simpul terminal yang biasanya ditetapkan sebanyak satu pengamatan saja. Pembentukan pohon klasifikasi yang besar dapat menghasilkan nilai dugaan yang melebihi nilai sebenarnya (*overfitting*). Sedangkan Breiman *et al.* (1984) berpendapat jika dalam proses pemilahan tersebut diberikan batasan padahal pemilahan masih layak untuk dilakukan maka akan menghasilkan nilai dugaan yang kurang dari nilai sebenarnya (*underfitting*). Oleh karena itu, untuk mendapatkan pohon yang layak perlu dilakukan pemangkasan pohon (*pruning*). Pemangkasan pohon tersebut dilakukan dengan cara mengurangi simpul pohon klasifikasi tanpa menyebabkan pengurangan nilai akurasi yang berarti sehingga didapatkan pohon yang layak dan tidak terlalu melebar. Penentuan ukuran pohon yang layak dapat dilakukan dengan metode *cost complexity*. Metode ini mengandalkan parameter *complexity* yang dinotasikan dengan C , yang secara perlahan nilainya meningkat selama proses pemangkasan.

Suatu simpul anak akan dipangkas jika hasil perubahan nilai kesalahan klasifikasi prediksi lebih kecil daripada C kali perubahan kompleksitas pohon. Persamaan fungsi *cost complexity* pada sub pohon (T) dari pohon klasifikasi maksimal (T_{\max}) dengan nilai $C \geq 0$ adalah sebagai berikut.

$$R_C(T) = R(T) + C |\tilde{T}| \quad (2.5)$$

dengan

$R_C(T)$ = *cost complexity measure*, kompleksitas suatu pohon T pada biaya kesalahan C

$R(T)$ = *resubstitution estimate*, kesalahan klasifikasi pohon T
 C = *complexity parameter, cost* bagi penambahan satu simpul
 akhir pada pohon T
 $|\tilde{T}|$ = banyaknya simpul terminal pada pohon T

Cost complexity pruning digunakan untuk menentukan suatu pohon bagian $T(C)$ yang meminimumkan $R_C(T)$ pada seluruh pohon bagian atau untuk setiap C . Berikut merupakan persamaan yang digunakan untuk mencari pohon bagian $T(C) < T_{\max}$ yang dapat meminimumkan $R_C(T)$.

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

Proses tersebut dilakukan berulang sampai tidak terdapat lagi pemangkasan yang mungkin dilakukan. Kemudian akan didapatkan hasil berupa deretan sub pohon yang semakin kecil.

2.1.3 Penentuan Pohon Klasifikasi Optimal

Penentuan pohon klasifikasi yang optimal dengan ukuran pohon yang terlalu besar dapat mengakibatkan tingginya nilai *cost complexity*. Hal ini dikarenakan struktur data yang digambarkan cenderung kompleks sehingga perlu dipilih pohon optimal yang berukuran sederhana namun tetap memberikan nilai estimasi pengganti yang memiliki selisih cukup kecil dibanding dengan nilai sebenarnya. Jika $R(T)$ dipilih sebagai nilai estimasi terbaik, maka akan cenderung menghasilkan pohon klasifikasi yang besar, karena pohon klasifikasi yang besar dapat menghasilkan nilai $R(T)$ yang kecil. Terdapat dua macam metode estimasi yang dapat digunakan untuk menentukan pohon klasifikasi yang optimal, yakni *test sample estimate* dan *V-fold cross validation estimate*.

1. *Test Sample Estimate*

Metode ini digunakan ketika data berukuran besar. Prosedur metode ini diawali dengan membagi data menjadi dua bagian, yakni data *training* (L_1) dan data *testing* (L_2). Pengamatan dalam L_1 digunakan untuk membentuk pohon T , sedangkan pengamatan da-

lam L_2 digunakan untuk menduga $R^{ts}(T_t)$. Berikut merupakan persamaan *test sample estimate*.

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

dengan

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

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

y_n = variabel respon ke- n dari data asli

j_n = hasil klasifikasi variabel respon ke- n

Sedangkan $X(d(y_n) \neq j_n)$ merupakan pernyataan yang akan bernilai 0 jika pernyataan dalam tanda kurung salah dan bernilai 1 jika pernyataan dalam tanda kurung benar. Untuk mengestimasi proporsi kesalahan yang dihasilkan dari proses pembentukan pohon klasifikasi tersebut, dipilih pohon klasifikasi optimal (T_t) dengan $R^{ts}(T_t) = \min_t R^{ts}(T_t)$.

2. *V-fold Cross Validation Estimate*

Metode ini digunakan ketika data berukuran kecil. Prosedur dalam metode ini adalah dengan membagi jumlah pengamatan (L) secara random menjadi V bagian (*fold*) yang saling independen dengan ukuran yang kurang lebih sama besar. Metode ini sering digunakan untuk menghindari *overlapping* pada data. Dalam pembentukan pohon, salah satu bagian akan berperan menjadi data *testing* sedangkan bagian yang lain berperan menjadi data *training*. Pembentukan pohon dilakukan dengan cara membentuk model sebanyak V kali, dengan bagian yang berbeda dari data setiap kali pembentukan yang dilakukan (Lewis, 2000). Misalkan $d^{(v)}(x)$ merupakan hasil pengklasifikasian, maka persamaan untuk *V-fold cross validation estimate* adalah sebagai berikut.

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

dengan

$R^{cv}(T_t)$ = total proporsi kesalahan *V-fold cross validation estimate*
 V = jumlah *fold* yang digunakan

Untuk mengestimasi proporsi kesalahan klasifikasi yang di hasilkan dari proses pembentukan pohon klasifikasi tersebut, dipilih terlebih dahulu pohon klasifikasi optimal (T_t) dengan

$$R^{cv}(T_t^{(v)}) = \min_t R^{cv}(T_t^{(v)}) \quad (\text{Lewis, 2000}).$$

Jumlah bagian (V) yang sering digunakan adalah 10. Hal ini dikarenakan nilai 10 merupakan nilai terbaik untuk mendapatkan estimasi kesalahan dari proses klasifikasi yang sudah dilakukan. Selain itu, dengan membagi data menjadi 10 bagian akan menghasilkan proporsi data yang seimbang (*balance*). Nilai akurasi yang dihasilkan dengan proporsi data yang seimbang akan lebih akurat (Witten, 2011). Sehingga *10-fold cross validation estimate* merupakan metode yang dijadikan sering digunakan dan dijadikan standar.

2.2 Ilustrasi CART

Berikut merupakan ilustrasi pembentukan pohon klasifikasi dengan metode pemilahan indeks gini dan pemilihan pemilah *goodness of split*. Pada ilustrasi ini digunakan 2 variabel prediktor yakni jarak rumah ke pasar terdekat (X_2) yang berskala ordinal dan jenis protein yang dikonsumsi (X_6) yang berskala nominal. Selain itu, digunakan variabel respon ketahanan pangan yang terdiri dari dua kategori, yakni kategori 1 untuk rumah tangga yang dikelompokkan ke dalam tahan pangan dan kategori 2 untuk rumah tangga yang dikelompokkan ke dalam kelompok rawan pangan. Data yang digunakan dalam ilustrasi ini dijelaskan dalam Tabel 2.1.

Tabel 2.1 Data Ilustrasi CART

Data ke-	Y	X_2	X_6
1	1	1	2
2	2	1	1
3	2	2	2
4	1	1	1
5	1	1	2
6	1	2	2
7	2	1	1
8	1	2	2
9	2	2	1
10	2	2	2

Langkah pertama dalam pembentukan pohon klasifikasi adalah penentuan jenis pemilah untuk setiap variabel. Berikut merupakan perhitungan jenis pemilah untuk variabel prediktor yang digunakan dalam ilustrasi ini.

Tabel 2.2 Jenis Pemilah Ilustrasi CART

Variabel	Nama Variabel	Skala Data	Jenis Pemilah
X_2	Jarak Rumah-Pasar	Ordinal	$2 - 1 = 1$ Pemilah
X_6	Jenis Protein yang Dikonsumsi	Nominal	$2^2 - 1 = 1$ Pemilah

Berdasarkan Tabel 2.2 dapat diketahui bahwa baik variabel X_2 maupun X_6 keduanya memiliki satu jenis pemilah. Berikutnya dihitung *goodness of split* untuk setiap jenis pemilahan dari kedua variabel prediktor yang digunakan. Perhitungan *goodness of split* tersebut bertujuan untuk mengetahui variabel prediktor yang akan menjadi pemilah utama. Tabel 2.3 menerangkan proporsi pengamatan pada variabel respon Y . Setelah diketahui proporsi pengamatan pada variabel respon Y , dilakukan perhitungan indeks gini untuk variabel respon Y berdasarkan proporsi yang didapatkan. Indeks gini tersebut berguna untuk menghitung nilai *goodness of split* baik dari variabel X_2 maupun variabel X_6 . Berikut merupakan proporsi pengamatan dan perhitungan indeks gini dari variabel respon Y .

Tabel 2.3 Proporsi Pengamatan pada Variabel Respon Y

Variabel respon (Y)	$N_i(t)$	$p(i t)$
Kelas 1	5	$\frac{5}{10} = 0,5$
Kelas 2	5	$\frac{5}{10} = 0,5$
Total	10	1,0

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

$$I(t) = p(1|t)p(2|t)$$

$$I(t) = (0,5)(0,5)$$

$$I(t) = 0,25$$

Berdasarkan hasil perhitungan didapatkan nilai indeks gini dari variabel respon Y pada ilustrasi CART ini adalah sebesar 0,25. Selanjutnya, dihitung *goodness of split* dari variabel prediktor X_2 . Perhitungan ini diawali dengan memilah data pada variabel prediktor X_2 . Data pada variabel tersebut yang memiliki kategori 1 dipilah sebagai simpul kiri, sedangkan data yang memiliki kategori 2 dipilah sebagai simpul kanan. Berikut merupakan perhitungan indeks gini dan *goodness of split* untuk variabel prediktor X_2 .

Tabel 2.4 Pemilahan Variabel Prediktor X_2

Simpul Kiri		Simpul Kanan	
$N_L = 5$	$p_L = \frac{5}{10} = 0,50$	$N_R = 5$	$p_L = \frac{5}{10} = 0,50$
$N(1/t_L) = 3$	$p(1 t_L) = \frac{3}{5} = 0,6$	$N(1/t_R) = 2$	$p(1 t_L) = \frac{2}{5} = 0,4$
$N(2/t_L) = 2$	$p(2 t_L) = \frac{2}{5} = 0,4$	$N(2/t_R) = 3$	$p(2 t_L) = \frac{3}{5} = 0,6$

Indeks gini untuk simpul kiri adalah sebesar

$$I(t_L) = \sum_{i,j=1} p(j|t_L)p(i|t_L), i \neq j$$

$$I(t_L) = p(1|t_L)p(2|t_L)$$

$$I(t_L) = (0,6)(0,4)$$

$$I(t_L) = 0,24$$

Sedangkan, indeks gini untuk simpul kanan adalah sebesar

$$I(t_R) = \sum_{i,j=1} p(j|t_R)p(i|t_R), i \neq j$$

$$I(t_R) = p(1|t_R)p(2|t_R)$$

$$I(t_R) = (0,4)(0,6)$$

$$I(t_R) = 0,24$$

Berdasarkan kedua indeks gini tersebut, dapat dihitung nilai *goodness of split* dari variabel prediktor X_2 , yakni sebesar

$$\phi(X_2, t) = I(t) - p_L I(t_L) - p_R I(t_R)$$

$$\phi(X_2, t) = (0,25) - (0,5)(0,24) - (0,5)(0,24)$$

$$\phi(X_2, t) = 0,010$$

Berikutnya, dilakukan perhitungan nilai *goodness of split* dari variabel prediktor X_6 . Perhitungan ini diawali dengan cara yang sama yakni memilah data pada variabel prediktor X_6 berdasarkan kategori data. Berikut merupakan perhitungan indeks gini dan *goodness of split* untuk variabel prediktor X_6 .

Tabel 2.5 Pemilahan Variabel Prediktor X_6

Simpul Kiri		Simpul Kanan	
$N_L = 4$	$p_L = \frac{4}{10} = 0,40$	$N_R = 6$	$p_L = \frac{6}{10} = 0,60$
$N(1/t_L)=1$	$p(1 t_L) = \frac{1}{4} = 0,25$	$N(1/t_R)=4$	$p(1 t_L) = \frac{4}{6} = 0,7$
$N(2/t_L)=3$	$p(2 t_L) = \frac{3}{5} = 0,75$	$N(2/t_R)=2$	$p(2 t_L) = \frac{2}{6} = 0,3$

Indeks gini untuk simpul kiri adalah sebesar

$$I(t_L) = \sum_{i,j=1} p(j|t_L)p(i|t_L), i \neq j$$

$$I(t_L) = p(1|t_L)p(2|t_L)$$

$$I(t_L) = (0,25)(0,75)$$

$$I(t_L) = 0,19$$

Sedangkan, indeks gini untuk simpul kanan adalah sebesar

$$I(t_R) = \sum_{i,j=1} p(j|t_R)p(i|t_R), i \neq j$$

$$I(t_R) = p(1|t_R)p(2|t_R)$$

$$I(t_R) = (0,7)(0,3)$$

$$I(t_R) = 0,21$$

Berdasarkan kedua indeks gini tersebut, dapat dihitung nilai *goodness of split* dari variabel prediktor X_6 , yakni sebesar

$$\phi(X_6, t) = I(t) - p_L I(t_L) - p_R I(t_R)$$

$$\phi(X_6, t) = (0,25) - (0,4)(0,19) - (0,6)(0,21)$$

$$\phi(X_6, t) = 0,049$$

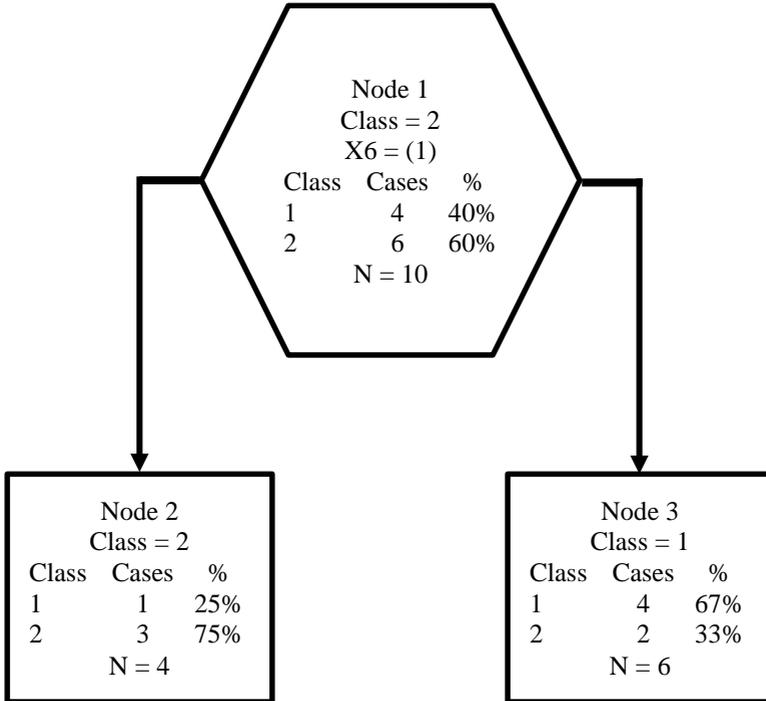
Setelah didapatkan nilai *goodness of split* dari seluruh kemungkinan pemilah, kemudian dilakukan perbandingan dari nilai tersebut. Pemilah yang memiliki nilai *goodness of split* tertinggi dipilih menjadi pemilah pertama. Berikut merupakan nilai *goodness of split* dari kedua pemilah.

Tabel 2.6 *Goodness of Split* Setiap Pemilah

Variabel	Pemilah	<i>Goodness of Split</i>
X_2	1:2	0,010
X_6	1:2	0,049

Berdasarkan Tabel 2.6 dapat diketahui bahwa variabel X_6 dengan pemilah 1:2 memiliki nilai *goodness of split* tertinggi, sehingga variabel tersebut digunakan sebagai pemilah pertama. Langkah selanjutnya yakni membangun pohon klasifikasi berda-

sarkan variabel yang terpilih menjadi pemilah pertama tersebut. Berikut merupakan pohon klasifikasi yang terbentuk.



Gambar 2.2 Ilustrasi CART

Setelah dibentuk pemilahan pertama, selanjutnya dilakukan perhitungan indeks gini dan *goodness of split* pada Node 2 dan Node 3. Jika pada simpul tersebut menunjukkan bahwa data belum homogen, maka dilakukan pemilahan kembali dengan cara yang sama. Pemilahan tersebut dilakukan secara *recursive* hingga didapatkan simpul yang homogen.

2.3 Adaptive Resampling and Combining (ARCING)

Adaptive Resampling and Combining (ARCING) merupakan suatu metode yang dapat mengecilkan kasus kesalahan klasifikasi dengan cara menekan nilai varian dalam kelompok klasi-

fikasinya. Metode ini merupakan hasil adaptasi dari metode *Bootstrap Agregating* (BAGGING). Metode ARCING lebih mampu mengecilkan nilai varian dalam kelompok dibandingkan dengan metode BAGGING. Pada metode ini dilakukan *resampling* pada data *learning* dengan rumusan peluang pengambilan tertentu. Berikut merupakan algoritma metode ARCING (Breiman, 1998).

1. Mengambil sampel dengan pengembalian dari data *training* awal (T) untuk mendapatkan data *training* baru $T^{(k)}$ dengan peluang pengambilan $p^{(k)}(l) = \frac{1}{N}$ dengan N merupakan jumlah data *training* awal.
2. Membentuk *classifier* atau pemilah (C_k) dari data $T^{(k)}$.
3. Menjalankan data *training* (T) pada C_k , sehingga dihasilkan jumlah misklasifikasi $m(l)$.
4. Memperbarui langkah ke- $k+1$ dengan rumusan peluang untuk data yang misklasifikasi sebagai berikut.

$$p^{k+1}(l) = \frac{1 + m(l)^4}{\sum (1 + m(l)^4)} \quad (2.9)$$

5. Mengulang langkah (1) sampai dengan (4) sebanyak K , dengan K merupakan jumlah replikasi yang dilakukan.
6. Mengkombinasikan ke- K pemilah (C_1, C_2, \dots, C_K) dengan *simple voting* untuk menghasilkan pohon klasifikasi yang optimal.

2.4 Ilustrasi ARCING

Berikut merupakan ilustrasi dari pendekatan ARCING dengan replikasi yang digunakan yakni sebanyak $K=5$ dan pada masing-masing replikasi diulang sebanyak $l=5$ kali dengan data *training* awal yang digunakan yakni sebanyak $N=100$ data.

1. Mengambil data *training* baru sebanyak 100 data dengan pengembalian dari data *training* awal dengan peluang untuk masing-masing data adalah $p = \frac{1}{N} = \frac{1}{100} = 0,01$. Pengambilan data tersebut diulang sebanyak 5 kali sehingga dihasilkan 5 kelompok data yang berbeda yang masing-masing terdiri dari 100 data.
2. Membentuk pohon klasifikasi dari 5 kelompok data yang telah terambil sebelumnya.
3. Menjalankan data *training* awal ke dalam 5 pohon klasifikasi pada langkah (2) sehingga didapatkan kesalahan klasifikasi pada masing-masing pohon.
4. Melakukan pembaruan peluang pengambilan pada data *training* awal yang mengalami kesalahan klasifikasi berdasarkan Persamaan (2.9) dengan $m(l)$ adalah jumlah kesalahan klasifikasi dan $l = 1, 2, 3, 4, 5$.
5. Mengulangi pengambilan data *training* baru sebanyak 100 data dengan pengembalian dari data *training* awal dengan peluang pengambilan baru yang didapatkan dari langkah (4).
6. Mengulangi langkah (2) sampai dengan (5) dengan $k = 1, 2, 3, 4, 5$ hingga didapatkan jumlah pohon yakni sebanyak $(k)(l) = (5)(5) = 25$ pohon klasifikasi yang berbeda.
7. Membandingkan kemiripan ke-25 pohon yang terbentuk dengan pohon yang dibentuk berdasarkan data *training* awal.
8. Pohon yang memiliki kemiripan paling banyak dengan pohon *training* awal dipilih menjadi pohon klasifikasi optimal dalam algoritma CART ARCING.

2.5 Ukuran Ketepatan Klasifikasi

Lewis (2000) berpendapat bahwa masalah klasifikasi mencakup empat hal utama. Empat hal utama tersebut adalah variabel respon yang bersifat kategorik, adanya variabel prediktor yang berpengaruh terhadap variabel respon, adanya data *training* yang

digunakan untuk menentukan model klasifikasi, dan adanya data *testing* yang digunakan untuk memprediksi ketepatan klasifikasi berdasarkan model yang dihasilkan oleh data *training*.

Terdapat beberapa jenis ukuran yang dapat digunakan untuk mengukur ketepatan hasil suatu klasifikasi, diantaranya adalah *Total Accuracy Rate* (1-APER). *Total Accuracy Rate* (1-APER) merupakan proporsi observasi yang diprediksi secara benar atau ukuran ketepatan klasifikasi total. Fungsi yang digunakan untuk menghitung nilai ukuran ketepatan klasifikasi total ini adalah sebagai berikut.

$$1 - APER = \frac{\text{Jumlah prediksi benar}}{\text{Jumlah total prediksi}} \quad (2.10)$$

2.6 Ketahanan Pangan

Konsep pangan di Indonesia secara jelas diatur dalam Undang-undang Nomor 7 tahun 1996 tentang Pangan, Undang-undang Nomor 18 tahun 2012 tentang Ketahanan Pangan dan Peraturan Pemerintah Nomor 28 tahun 2004 tentang Keamanan, Mutu dan Gizi Pangan. Ketahanan pangan dalam Undang-undang Nomor 18 tahun 2012 dijelaskan sebagai kondisi terpenuhinya pangan bagi negara sampai dengan perseorangan yang tercermin dari tersedianya pangan yang cukup, baik jumlah maupun mutunya, aman, beragam, bergizi, merata, dan terjangkau serta tidak bertentangan dengan agama, keyakinan, dan budaya masyarakat untuk dapat hidup sehat, aktif, dan produktif secara berkelanjutan. Sedangkan dalam Undang-undang Nomor 7 tahun 1996, ketahanan pangan dijelaskan sebagai kondisi terpenuhinya pangan bagi rumah tangga yang tercermin dari tersedianya pangan yang cukup, baik jumlah maupun mutunya, aman, merata, dan terjangkau.

Food and Agriculture Organization (FAO) menjelaskan ketahanan pangan sebagai keadaan di mana semua rumah tangga mempunyai akses baik fisik maupun ekonomi untuk mendapatkan pangan bagi seluruh anggota keluarganya dan setiap rumah tangga tidak beresiko kehilangan kedua akses tersebut (Maxwell & Frankenberger, 1992). Terdapat lima unsur yang secara tidak lang-

sung harus dipenuhi, diantaranya adalah berorientasi pada rumah tangga dan individu, dimensi waktu setiap saat pangan tersedia dan dapat diakses, menekankan pada akses pangan rumah tangga dan individu baik secara fisik, ekonomi, maupun sosial, dan berorientasi pada pemenuhan gizi, serta ditujukan untuk hidup sehat dan produktif. FAO merangkum kelima unsur tersebut ke dalam indikator-indikator yang kemudian dapat digunakan untuk menentukan ketahanan suatu rumah tangga. Menurut FAO terdapat empat indikator utama ketahanan pangan, yakni.

1. Ketersediaan pangan (*food sufficiency*); kecukupan jumlah pangan.
2. Stabilitas ketersediaan pangan; sistem distribusi pangan yang mendukung tersedianya pangan setiap saat dan merata.
3. Aksesibilitas atau keterjangkauan pangan; kemudahan rumah tangga memperoleh pangan dengan harga yang terjangkau.
4. Keamanan atau kualitas jenis pangan (*food safety*); pangan yang bebas dari kemungkinan cemaran biologis, kimia, dan benda lain yang dapat mengganggu, merugikan, dan membahayakan keadaan manusia, serta terjamin mutunya (*food quality*) yaitu memenuhi kandungan gizi dan standar perdagangan terhadap bahan makanan dan minuman.

Indikator-indikator tersebut kemudian digambarkan melalui beberapa faktor yang dapat menjelaskan ketahanan pangan suatu rumah tangga. Pusat Penelitian LIPI dalam Pemodelan Penderita Penyakit Tuberkulosis Paru Menurut Stratifikasi Ketahanan Pangan Rumah Tangga di Wilayah Tengah Kota Surabaya dengan Metode Regresi Logistik Biner (2016) menjelaskan bahwa ketersediaan pangan dalam rumah tangga yang dipakai dalam pengukuran mengacu pada pangan yang cukup dan tersedia dalam jumlah yang dapat memenuhi kebutuhan konsumsi rumah tangga dalam waktu tertentu. Hal ini dapat digambarkan dengan ketersediaan makanan pokok pada suatu rumah tangga dalam kurun satu bulan atau satu tahun. Kondisi ketersediaan pangan dapat dilihat pada Tabel 2.7.

Tabel 2.7 Ketersediaan Pangan Rumah Tangga

Makanan Pokok	Persediaan Pangan Per Bulan	Kondisi
Beras	≥ 20 hari	Cukup
	< 20 hari	Tidak Cukup
Jagung	≥ 30 hari	Cukup
	< 30 hari	Tidak Cukup

Sumber: Puslit Kependudukan-LIPI, 2009

Stabilitas ketersediaan pangan di tingkat rumah tangga diukur berdasarkan kecukupan ketersediaan pangan dan frekuensi makan anggota rumah tangga dalam sehari. Tabel 2.8 merupakan kombinasi antara ketersediaan makanan pokok dengan frekuensi makan sebagai indikator stabilitas ketersediaan pangan.

Tabel 2.8 Stabilitas Ketersediaan Pangan

Kecukupan Ketersediaan Pangan	Frekuensi Makan Anggota Rumah Tangga	
	≥ 3 kali dalam sehari	< 3 kali dalam sehari
Cukup	Stabil	Tidak Stabil
Tidak Cukup	Tidak Stabil	Tidak Stabil

Sumber: Puslit Kependudukan-LIPI, 2009

Indikator aksesibilitas atau keterjangkauan terhadap pangan dapat diukur dengan cara melihat kemudahan rumah tangga dalam memperoleh pangan yang diukur berdasarkan akses fisik, akses sosial, dan akses ekonomi yang ditentukan oleh Badan Pusat Statistika pada tahun 2007. Berikut merupakan faktor aksesibilitas atau keterjangkauan pangan.

Tabel 2.9 Aksesibilitas atau Keterjangkauan Pangan

Aksesibilitas	Baik	Buruk
Akses Fisik:		
Lokasi Pasar	Dalam kecamatan (≤ 2 km)	Luar kecamatan (> 2 km)
Akses Sosial:		
a. Jumlah anggota rumah tangga	< 7 orang	≥ 7 orang
b. Pendidikan terakhir kepala keluarga/ibu	Minimal SD	Tidak Sekolah

Tabel 2.9 Aksesibilitas atau Keterjangkauan Pangan (Lanjutan)

Aksesibilitas	Baik	Buruk
Akses Ekonomi: Cara memperoleh makanan pokok	Tidak berhutang	Berhutang

Berdasarkan pengukuran indikator stabilitas ketersediaan pangan dan aksesibilitas pangan dapat diukur kontinyuitas ketersediaan pangan rumah tangga. Tabel 2.10 merupakan penjelasan mengenai ukuran kontinyuitas ketersediaan pangan rumah tangga.

Tabel 2.10 Kontinyuitas Ketersediaan Pangan

Aksesibilitas atau Keterjangkauan Pangan	Stabilitas Ketersediaan Pangan	
	Stabil	Tidak Stabil
Baik	Kontinyu	Tidak Kontinyu
Buruk	Tidak Kontinyu	Tidak Kontinyu

Sumber: Puslit Kependudukan-LIPI, 2009

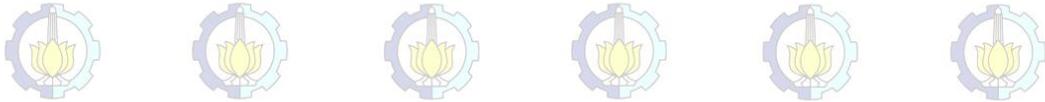
Pengukuran indikator yang terakhir adalah kualitas jenis pangan yang dapat diukur dengan cara melihat data pengeluaran untuk konsumsi makanan (lauk pauk) sehari-hari yang mengandung protein hewani dan/atau nabati. Terdapat dua kategori rumah tangga berdasarkan kriteria ini, yakni

1. Rumah tangga dengan kualitas pangan baik adalah rumah tangga yang memiliki pengeluaran untuk lauk pauk berupa protein hewani dan nabati atau protein hewani saja.
2. Rumah tangga dengan kualitas pangan tidak baik adalah rumah tangga yang memiliki pengeluaran untuk lauk pauk berupa protein nabati saja atau tidak sama sekali.

Berdasarkan aspek-aspek tersebut, tingkat ketahanan pangan suatu rumah tangga dapat dibedakan menjadi dua kategori, yakni rumah tangga tahan pangan dan rumah tangga rawan pangan. Suatu rumah tangga dikatakan tahan pangan ketika mampu memenuhi kebutuhan pangan anggotanya, baik jumlah maupun mutunya, aman, merata, dan terjangkau. Sedangkan dikatakan rawan pangan

ketika tidak mampu memenuhi kebutuhan pangan pada waktu tertentu untuk memenuhi standar kebutuhan fisiologis bagi pertumbuhan dan kesehatan (Purwatini, 2014).

(halaman ini sengaja dikosongkan)



BAB III
METODOLOGI PENELITIAN

BAB III METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari hasil penelitian yang berjudul “Analisis Ketahanan Pangan Rumah Tangga Terhadap Kasus Penderita Penyakit Tuberculosis dengan Pendekatan *Geographically Weighted Poisson Regression* (Studi Kasus Pantai Pesisir Surabaya)” yang disusun oleh Wulandari, S. P., Susilaningrum, D., dan Latra, I. N. pada tahun 2016. Data tersebut terdiri dari 124 buah yang terdapat pada Lampiran A.

Tabel 3.1 Struktur Data

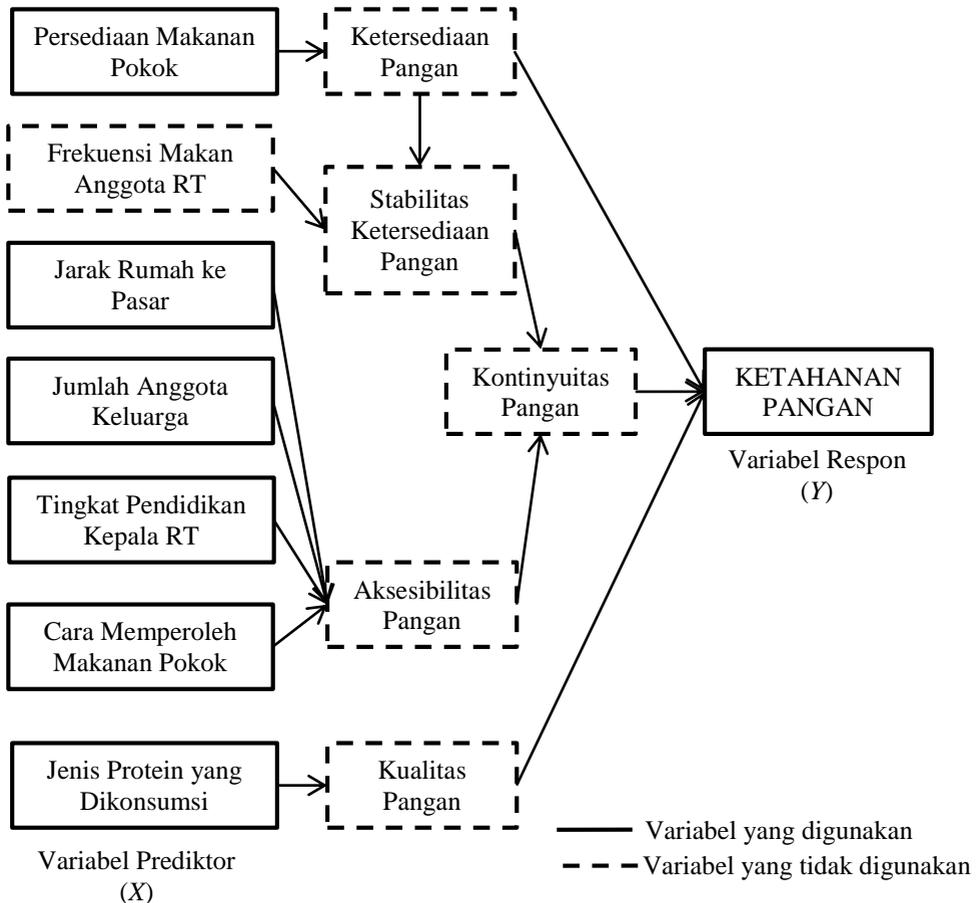
Data Ke-	Y	X_1	X_2	...	X_6
1	y_1	$x_{1,1}$	$x_{1,2}$...	$x_{1,6}$
2	y_2	$x_{2,1}$	$x_{2,2}$...	$x_{2,6}$
...
124	y_{124}	$x_{124,1}$	$x_{124,2}$...	$x_{124,6}$

Tabel 3.1 merupakan stuktur data yang digunakan dalam penelitian ini, dimana variabel respon yang digunakan memiliki dua kategori, yakni rumah tangga tahan pangan dan rumah tangga tidak tahan pangan atau rawan pangan.

3.2 Kerangka Konsep Penelitian

Berdasarkan *Food and Agriculture Organization* (FAO) ketahanan pangan digambarkan melalui empat aspek utama, yakni ketersediaan pangan, stabilitas ketersediaan atau pemerataan pangan, aksesibilitas atau keterjangkauan pangan, dan kualitas atau keamanan pangan. Pusat Penelitian Kependudukan LIPI (2009) menjelaskan bahwa ketersediaan pangan dapat diukur melalui ketersediaan makanan pokok dalam sebulan atau setahun. Stabilitas ketersediaan pangan dapat diukur melalui kecukupan ketersediaan pangan dan frekuensi makan anggota rumah tangga dalam sehari. Berdasarkan stabilitas ketersediaan dan aksesibilitas pangan dapat diketahui kontinuitas pangan dari suatu rumah

tangga. Kualitas pangan dapat diukur berdasarkan jenis protein yang dikonsumsi. Badan Pusat Statistika menjelaskan aksesibilitas atau keterjangkauan pangan dapat diukur berdasarkan 3 akses utama, yakni akses fisik, akses sosial, dan akses ekonomi. Akses fisik dapat digambarkan dengan jarak antara rumah ke pasar, akses sosial dapat digambarkan dengan jumlah anggota rumah tangga dan tingkat pendidikan kepala rumah tangga, akses ekonomi dapat digambarkan dengan cara memperoleh makanan pokok (Wulandari, Susilaningrum, & Latra, 2016). Gambar 3.1 menunjukkan kerangka konsep dari penelitian ini.



Gambar 3.1 Kerangka Konsep Penelitian

3.3 Variabel Penelitian

Terdapat dua jenis variabel yang digunakan dalam penelitian ini, yakni variabel respon (Y) dan variabel prediktor (X). Terdapat satu variabel respon dan enam variabel prediktor yang digunakan dalam penelitian ini. Variabel tersebut diharapkan dapat mewakili indikator-indikator ketahanan pangan yang ditetapkan oleh FAO. Berikut merupakan penjelasan lebih lanjut mengenai variabel penelitian yang digunakan.

Tabel 3.2 Variabel Penelitian

Keterangan	Skala Data	Kategori	Definisi Operasional
Ketahanan Pangan (Y)	Nominal	1 = Tahan Pangan 2 = Rawan Pangan	Suatu rumah tangga dikatakan tahan pangan ketika mampu memenuhi kebutuhan pangan anggotanya, baik jumlah maupun mutunya, aman, merata, dan terjangkau. Suatu rumah tangga dikatakan rawan pangan ketika tidak mampu memenuhi kebutuhan pangan pada waktu tertentu untuk memenuhi standar kebutuhan fisiologis bagi pertumbuhan dan kesehatan.
Persediaan Makanan Pokok (Beras) (X_1)	Ordinal	1 = Tersedia \geq 20 hari per bulan 2 = Tersedia $<$ 20 hari per bulan	Jumlah hari dalam sebulan dimana tersedia beras sebagai makanan pokok.
Jarak Rumah-Pasar (X_2)	Ordinal	1 = \leq 2 km 2 = $>$ 2 km	Jarak dari rumah ke pasar terdekat.
Jumlah Anggota Keluarga (X_3)	Ordinal	1 = $<$ 7 orang 2 = \geq 7 orang	Jumlah orang yang terkumpul dan tinggal di suatu tempat di bawah suatu atap dalam keadaan saling ketergantungan.

Tabel 3.2 Variabel Penelitian (Lanjutan)

Keterangan	Skala Data	Kategori	Definisi Operasional
Pendidikan Kepala Rumah Tangga (X_4)	Ordinal	1 = Minimal SD 2 = Tidak Bersekolah	Pendidikan yang pernah ditempuh oleh kepala rumah tangga.
Cara Memperoleh Makanan Pokok (X_5)	Nominal	1 = Tidak Berhutang 2 = Berhutang	Usaha yang dilakukan untuk memperoleh makanan pokok.
Jenis Protein yang Dikonsumsi (X_6)	Nominal	1 = Hewani atau Hewani dan Nabati 2 = Nabati atau Tidak Ada Sama Sekali	Jenis protein yang terkandung dalam lauk pauk yang dikonsumsi sehari-hari.

3.4 Langkah Analisis

Penelitian ini dalam pengolahan data pada kasus CART dan CART ARCING menggunakan bantuan *software CART for Windows Version 4.0* untuk memperoleh hasil pohon klasifikasi dan ketepatan klasifikasi. Terdapat lima langkah utama dalam penelitian ini, yakni (1) Menggambarkan karakteristik ketahanan pangan rumah tangga di Surabaya, (2) Membentuk pohon klasifikasi (CART), (3) Melakukan klasifikasi dengan menggunakan CART ARCING, (4) Mengevaluasi ketepatan klasifikasi CART dan CART ARCING, dan (5) Membandingkan hasil ketepatan klasifikasi CART dan CART ARCING.

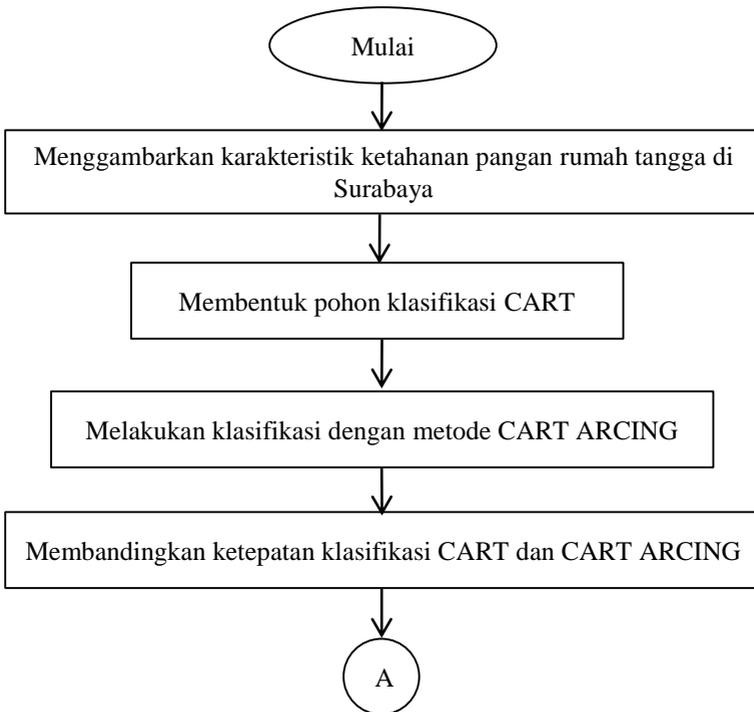
1. Menggambarkan karakteristik ketahanan pangan rumah tangga di Surabaya.
2. Membentuk pohon klasifikasi (CART) dengan langkah sebagai berikut.

- a. Melakukan pemisahan data menjadi dua bagian, yakni data *training* dan data *testing* dengan perbandingan 80:20.
 - b. Membentuk pohon klasifikasi dari data *training* dengan langkah-langkah sebagai berikut.
 - i. Menentukan pemilah dan pemilahan terbaik dari variabel prediktor
 - ii. Menentukan simpul terminal
 - iii. Memberikan label kelas pada simpul terminal
 - c. Menghentikan pembentukan pohon klasifikasi.
 - d. Melakukan pemangkasan pohon klasifikasi.
 - e. Memilih pohon klasifikasi optimal dengan *10-fold cross validation estimate* yang kemudian menjadi pohon klasifikasi terbaik.
 - f. Menghitung nilai ketepatan klasifikasi dari pohon klasifikasi terbaik dengan data *training* dan melakukan validasi dengan data *testing* pada pohon klasifikasi yang terbentuk.
3. Melakukan klasifikasi menggunakan CART ARCING dengan langkah sebagai berikut.
- a. Melakukan pemisahan data menjadi dua bagian, yakni data *training* dan data *testing*. Pembagian data *training* dan *testing* yang digunakan dalam penelitian ini yakni 80:20.
 - b. Mengambil sampel dengan pengembalian dari data *training* awal (T) dengan probabilitas $p^{(k)}(l) = \frac{1}{N}$ untuk mendapatkan data *training* baru $T^{(k)}$.
 - c. Membentuk *classifier* atau pemilah (C_k) dari data $T^{(k)}$ berdasarkan prosedur pada langkah (2).
 - d. Menjalankan data *training* awal (T) pada C_k , sehingga dihasilkan jumlah misklasifikasi $m(l)$.
 - e. Mengulangi langkah pada pengambilan sampel dengan memperbarui nilai probabilitas yang digunakan dengan Persamaan (2.9) dengan jumlah replikasi yang digunakan yakni 25, 50, 75, 100, 125, dan 150.

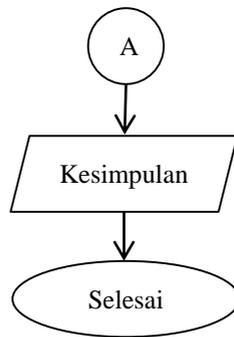
- f. Mengkombinasikan ke-K pemilah (C_1, C_2, \dots, C_K) dengan *simple voting* untuk mendapatkan model yang optimal.
 - g. Menjalankan data *testing* pada yang dihasilkan pada (f) untuk menghitung nilai ketepatan klasifikasi.
4. Membandingkan hasil ketepatan klasifikasi CART dan CART ARCING.
 5. Menarik kesimpulan.

3.5 Diagram Alir

Tahapan utama dalam penelitian ini berdasarkan langkah analisis yang telah dijelaskan sebelumnya digambarkan melalui diagram alir penelitian sebagai berikut.



Gambar 3.2 Diagram Alir Penelitian



Gambar 3.2 Diagram Alir Penelitian (Lanjutan)

(halaman ini sengaja dikosongkan)



BAB IV
HASIL DAN PEMBAHASAN



BAB IV

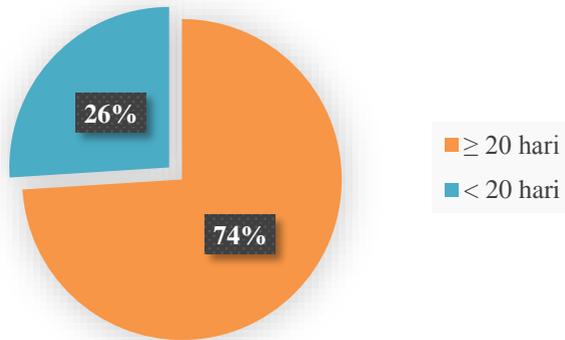
ANALISIS DAN PEMBAHASAN

Pada bab ini dibahas mengenai hasil analisis dan pembahasan berdasarkan pengolahan data yang telah dilakukan. Metode yang digunakan dalam analisis ini meliputi statistika deskriptif untuk menggambarkan karakteristik faktor-faktor ketahanan pangan rumah tangga di Surabaya serta CART untuk mengetahui hasil ketepatan klasifikasi ketahanan pangan rumah tangga di Surabaya serta mengetahui faktor-faktor yang berpengaruh dalam klasifikasi tersebut. Selain itu, juga dilakukan analisis CART ARCING untuk perbaikan hasil ketepatan klasifikasi ketahanan pangan rumah tangga di Surabaya yang dihasilkan dari metode CART.

4.1 Karakteristik Faktor-Faktor Ketahanan Pangan Rumah Tangga di Surabaya

Berdasarkan *Food and Agriculture Organization* (FAO) ketahanan pangan digambarkan melalui empat aspek utama, yakni ketersediaan pangan, stabilitas ketersediaan atau pemerataan pangan, aksesibilitas atau keterjangkauan pangan, dan kualitas atau keamanan pangan.

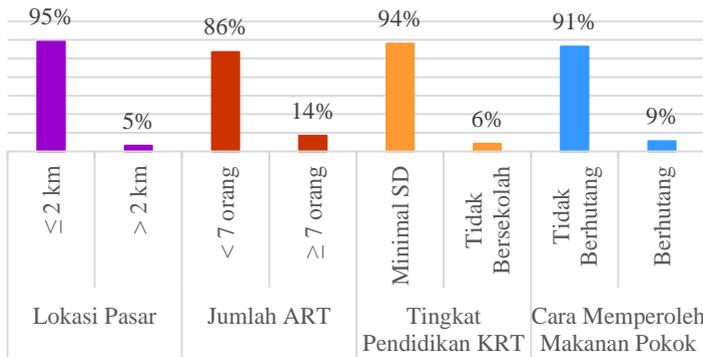
Pusat Penelitian LIPI menjelaskan bahwa ketersediaan pangan dalam rumah tangga yang dipakai dalam pengukuran mengacu pada pangan yang cukup dan tersedia dalam jumlah yang dapat memenuhi kebutuhan konsumsi rumah tangga dalam waktu satu bulan atau satu tahun. Suatu rumah tangga dikatakan memiliki ketersediaan pangan yang cukup jika memiliki persediaan makanan pokok minimal 20 hari dalam satu bulan (Pramita, 2016). Berikut merupakan presentase persediaan makanan pokok rumah tangga berdasarkan data yang tersedia.



Gambar 4.1 Persediaan Makanan Pokok Rumah Tangga di Surabaya

Berdasarkan data yang tersedia, diketahui bahwa sebanyak 92 dari 124 rumah tangga atau sebanyak 74% memiliki persediaan makanan pokok berupa beras sama dengan atau lebih dari 20 hari dalam satu bulan. Sedangkan sebanyak 32 dari 124 rumah tangga atau sebanyak 26% memiliki persediaan makanan pokok berupa beras kurang dari 20 hari dalam satu bulan.

Stabilitas ketersediaan pangan pada tingkat rumah tangga dapat diukur berdasarkan kecukupan ketersediaan pangan dan frekuensi makan anggota rumah tangga dalam sehari. Aspek ketiga dalam ketahanan pangan adalah aksesibilitas pangan. Aksesibilitas atau keterjangkauan terhadap pangan dapat diukur dengan cara melihat kemudahan rumah tangga dalam memperoleh pangan yang diukur berdasarkan akses fisik, akses sosial, dan akses ekonomi yang ditentukan oleh Badan Pusat Statistik pada tahun 2007. Akses fisik dapat digambarkan berdasarkan lokasi pasar terdekat. Akses sosial dapat digambarkan berdasarkan jumlah anggota keluarga dan tingkat pendidikan terakhir kepala rumah tangga. Sedangkan, akses ekonomi dapat digambarkan melalui cara memperoleh makanan pokok. Berikut merupakan karakteristik aksesibilitas ketahanan pangan rumah tangga di Surabaya berdasarkan data yang tersedia.

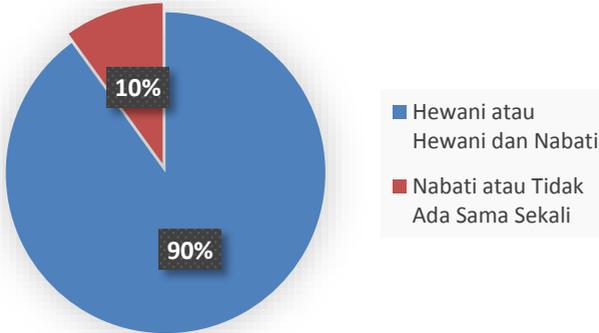


Gambar 4.2 Akses Fisik, Sosial, dan Ekonomi Rumah Tangga di Surabaya

Hasil perhitungan menunjukkan bahwa sebanyak 118 dari 124 rumah tangga atau sebanyak 95% rumah tangga memiliki tempat tinggal yang berjarak kurang dari sama dengan dua kilometer dari pasar terdekat, sedangkan sisanya sebanyak 5% bertempat tinggal dengan jarak dari pasar terdekat lebih dari dua kilometer. Hasil perhitungan juga menunjukkan bahwa sebanyak 107 dari 124 rumah tangga atau sebanyak 86% rumah tangga memiliki anggota keluarga kurang dari tujuh orang, sedangkan sisanya sebanyak 14% rumah tangga memiliki jumlah anggota keluarga lebih dari tujuh orang. Terdapat 116 dari 124 rumah tangga atau sebanyak 94% rumah tangga yang memiliki kepala rumah tangga dengan pendidikan terakhir minimal SD dan sisanya sebanyak 6% kepala rumah tangganya tidak bersekolah. Selain itu, berdasarkan cara memperoleh makanan pokok diketahui bahwa sebanyak 113 dari 124 rumah tangga atau sebanyak 91% rumah tangga mendapatkan makanan pokok dengan tidak berhutang sedangkan sisanya sebanyak 9% mendapatkan makanan pokok dengan cara berhutang.

Aspek terakhir dalam ketahanan pangan adalah kualitas pangan. Penentuan kualitas pangan dapat diukur dengan cara melihat data pengeluaran untuk konsumsi makanan (lauk pauk) sehari-hari yang mengandung protein hewani dan/atau nabati.

Gambar 4.4 merupakan gambaran jenis protein yang dikonsumsi oleh 124 rumah tangga yang ada.



Gambar 4.3 Jenis Protein yang Dikonsumsi Rumah Tangga di Surabaya

Berdasarkan Gambar 4.4 dapat diketahui bahwa sebanyak 90% rumah tangga atau sebanyak 112 dari 124 rumah tangga mengkonsumsi lauk pauk dengan protein yang terkandung yakni hewani atau hewani dan nabati. Sedangkan sisanya sebanyak 10% rumah tangga mengkonsumsi lauk pauk dengan protein nabati saja atau tidak mengkonsumsi protein sama sekali.

4.2 Klasifikasi Ketahanan Pangan Rumah Tangga di Surabaya dengan *Classification Trees*

Berikut merupakan hasil analisis ketahanan pangan rumah tangga di Surabaya. Metode klasifikasi yang digunakan dalam analisis ini adalah *classification trees*. *Classification trees* merupakan metode klasifikasi yang bertujuan untuk mengetahui ketepatan klasifikasi yang telah dihasilkan berdasarkan status suatu rumah tangga melalui pohon klasifikasi maksimal dan pohon klasifikasi optimal dengan proses pemilahan. Metode pemilahan yang digunakan dalam analisis ini adalah fungsi heterogenitas indeks gini. Data yang digunakan dalam analisis ini adalah sebanyak 124 buah, dimana 99 data digunakan sebagai data *training* dan sisanya sebanyak 25 data digunakan sebagai

data *testing*. Berdasarkan data *training* yang digunakan dalam analisis, terdapat 54 rumah tangga yang dikategorikan sebagai rumah tangga rawan pangan dan 45 rumah tangga yang dikategorikan sebagai rumah tangga tahan pangan. Karena jumlah data sampel pengamatan berukuran kecil (<3000), maka penentuan pohon klasifikasi optimal ditentukan menggunakan metode *V-folds cross validation estimate* dengan jumlah *fold* yang digunakan sebanyak 10 *fold* ($V=10$).

4.2.1 Kemungkinan Pemilah untuk Setiap Variabel Prediktor

Variabel terpenting diperlukan dalam tahap pembentukan pohon klasifikasi. Variabel tersebut akan berperan sebagai pemilah terbaik. Variabel yang berperan sebagai pemilah dipilih berdasarkan beberapa kemungkinan pemilah dari setiap variabel prediktor atau independen. Jika variabel independen yang digunakan memiliki jenis data kontinyu dengan ruang sampel berukuran n , maka akan terbentuk $n - 1$ jenis pemilahan yang berbeda. Jika variabel independen yang digunakan memiliki jenis data kategori nominal bertaraf L , maka akan terbentuk $2^{L-1} - 1$ pemilahan yang berbeda. Namun jika variabel independen yang digunakan memiliki jenis data kategori ordinal bertaraf L , maka akan diperoleh $L - 1$ jenis pemilahan. Berikut merupakan hasil perhitungan kemungkinan pemilah pada variabel prediktor yang digunakan dalam analisis ini.

Tabel 4.1 Kemungkinan Pemilah

Variabel	Nama Variabel	Skala Data	Jumlah Kategori	Kemungkinan Pemilah
X_1	Persediaan Makanan Pokok (Beras)	Ordinal	2	$2 - 1 = 1$ Pemilah
X_2	Jarak Rumah-Pasar	Ordinal	2	$2 - 1 = 1$ Pemilah
X_3	Jumlah Anggota Keluarga	Ordinal	2	$2 - 1 = 1$ Pemilah

Tabel 4.1 Kemungkinan Pemilah (Lanjutan)

Variabel	Nama Variabel	Skala Data	Jumlah Kategori	Kemungkinan Pemilah
X_4	Pendidikan Kepala Rumah Tangga	Ordinal	2	$2 - 1 = 1$ Pemilah
X_5	Cara Memperoleh Makanan Pokok	Nominal	2	$2^{2-1} - 1 = 1$ Pemilah
X_6	Jenis Protein yang Dikonsumsi	Nominal	2	$2^{2-1} - 1 = 1$ Pemilah

Berdasarkan Tabel 4.1 dapat diketahui bahwa keseluruhan variabel prediktor memiliki satu kemungkinan pemilah.

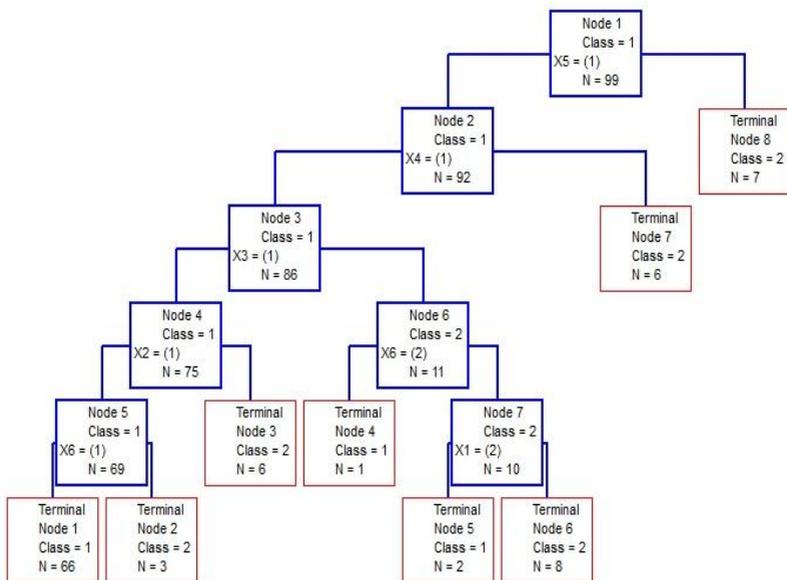
4.2.2 Pembentukan Pohon Klasifikasi Maksimal

Pembentukan pohon klasifikasi maksimal merupakan tahapan pertama dalam analisis *classification trees*. Langkah awal yang dilakukan dalam pembentukan pohon klasifikasi ini adalah pemilihan pemilah yakni dengan memilih variabel prediktor terpenting yang akan digunakan sebagai pemilah awal. Pemilihan pemilah tersebut didasari oleh skor kontribusi pada masing-masing variabel prediktor dalam pembentukan pohon klasifikasi maksimal. Variabel prediktor yang memiliki skor kontribusi terbesar akan dipilih sebagai pemilah awal. Besarnya kontribusi pada masing-masing variabel prediktor dapat dilihat pada Tabel 4.2.

Tabel 4.2 Skor Kontribusi Variabel Prediktor

Variabel	Nama Variabel	Skor
X_5	Cara Memperoleh Makanan Pokok	100,00
X_6	Jenis Protein yang Dikonsumsi	90,61
X_4	Pendidikan Kepala Rumah Tangga	84,07
X_3	Jumlah Anggota Keluarga	30,25
X_2	Jarak Rumah-Pasar	29,28
X_1	Persediaan Makanan Pokok (Beras)	13,72

Berdasarkan Tabel 4.2 dapat diketahui bahwa variabel X_5 atau cara memperoleh makanan pokok memiliki kontribusi terbesar yakni dengan skor 100,00. Hal ini berarti variabel X_5 dapat menurunkan tingkat heterogenitas paling besar, sehingga simpul akan dihasilkan lebih homogen. Selain itu, variabel X_5 juga merupakan variabel yang paling berpengaruh dalam pembentukan pohon klasifikasi ketahanan pangan. Cara memperoleh makanan pokok dapat ditinjau dari besarnya pendapatan. Jika dikaitkan dengan batasan masalah bahwa data yang digunakan merupakan data ketahanan pangan rumah tangga dari penderita tuberkulosis, besarnya pendapatan ini akan mempengaruhi cara memperoleh makanan pokok akibat adanya penderita tuberkulosis dalam rumah tangga tersebut yang membutuhkan pengobatan secara berkala. Selanjutnya, variabel X_5 akan digunakan sebagai pemilah awal atau biasa disebut dengan simpul induk (*parent nodes*).

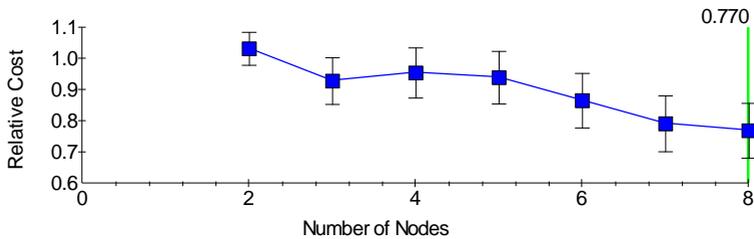


Gambar 4.4 Pohon Klasifikasi Maksimal

Setelah didapatkan variabel yang berperan sebagai simpul induk, selanjutnya dilakukan proses pemilahan dengan fungsi heterogenitas indeks gini untuk membentuk pohon klasifikasi maksimal. Pohon klasifikasi maksimal merupakan pohon klasifikasi yang memiliki jumlah simpul terminal paling banyak dan tingkat kedalaman paling tinggi. Gambar 4.4 merupakan topologi dari pohon klasifikasi maksimal yang dihasilkan berdasarkan analisis yang telah dilakukan. Jumlah simpul terminal yang dihasilkan dari pohon klasifikasi maksimal tersebut adalah sebanyak 8 dengan tingkat kedalaman sebesar 5.

4.2.3 Pemangkasan Pohon Klasifikasi Maksimal

Setelah didapatkan pohon klasifikasi maksimal, kemudian akan dilihat apakah perlu dilakukan pemangkasan (*pruning*) pada pohon klasifikasi maksimal tersebut. Pemangkasan tersebut bertujuan untuk menghindari adanya kasus *underfitting* atau *overfitting*. *Underfitting* merupakan suatu kasus dimana nilai akurasi yang dihasilkan kurang jika dibandingkan dengan nilai aktual, sedangkan *overfitting* merupakan suatu kasus dimana nilai akurasi yang dihasilkan melebihi kenyataan. Proses pemangkasan pohon klasifikasi maksimal dilakukan dengan pendekatan *cost complexity minimum*, serta digunakan metode *10-folds cross validation estimate* dalam menentukan pohon klasifikasi optimal. Metode ini membagi data menjadi 10 bagian, dimana 9 bagian berperan sebagai data *training* dan 1 bagian sebagai data *testing*. *V-fold cross validation estimate* memberikan kesempatan seluruh data untuk menjadi data *training* dan data *testing*. Pohon klasifikasi dibentuk berdasarkan data *training*, kemudian dilakukan perhitungan kesalahan klasifikasi dengan data *testing*. Hasil pemangkasan yang dilakukan memiliki *relative cost* tertentu. Pohon klasifikasi optimal merupakan pohon klasifikasi yang memiliki *relative cost* terkecil. Gambar 4.5 merupakan plot *relative cost* dari setiap pemangkasan yang dilakukan pada pohon klasifikasi ketahanan pangan rumah tangga di Surabaya.



Gambar 4.5 Plot *Relative Cost*

Berdasarkan plot *relative cost* tersebut dapat diketahui bahwa pohon klasifikasi maksimal yang terbentuk merupakan pohon klasifikasi optimal. Hal ini ditunjukkan oleh garis berwarna hijau yang berada pada simpul terminal ke-8 atau simpul terminal yang terakhir dalam Gambar 4.5 tersebut. *Relative cost* yang dimiliki oleh simpul ke-8 adalah sebesar 0,770. Tabel 4.3 memberikan informasi mengenai urutan dalam pembentukan pohon klasifikasi sampai didapatkan pohon klasifikasi optimal.

Tabel 4.3 Urutan Pembentukan Pohon Klasifikasi

<i>Tree Number</i>	<i>Terminal Nodes</i>	<i>Cross-Validation Relative Cost</i>	<i>Resubstitution Relative Cost</i>	<i>Complexity Parameter</i>
1**	8	0,770 ± 0,088	0,730	0,000
2	7	0,793 ± 0,090	0,733	0,002
3	6	0,867 ± 0,088	0,748	0,007
4	5	0,941 ± 0,084	0,770	0,011
5	4	0,956 ± 0,080	0,800	0,015
6	3	0,930 ± 0,075	0,841	0,020
7	2	1,033 ± 0,053	0,911	0,035
8	1	1,000 ± 0,000	1,000	0,044

** merupakan pohon klasifikasi optimal

Tabel 4.3 menunjukkan urutan pembentukan pohon klasifikasi. *Tree number* merupakan urutan pembentukan pohon klasifikasi. *Terminal nodes* merupakan banyaknya simpul terminal yang dihasilkan pada setiap urutan pembentukan pohon klasifikasi. *Cross-validation relative cost* merupakan rentang kesalahan

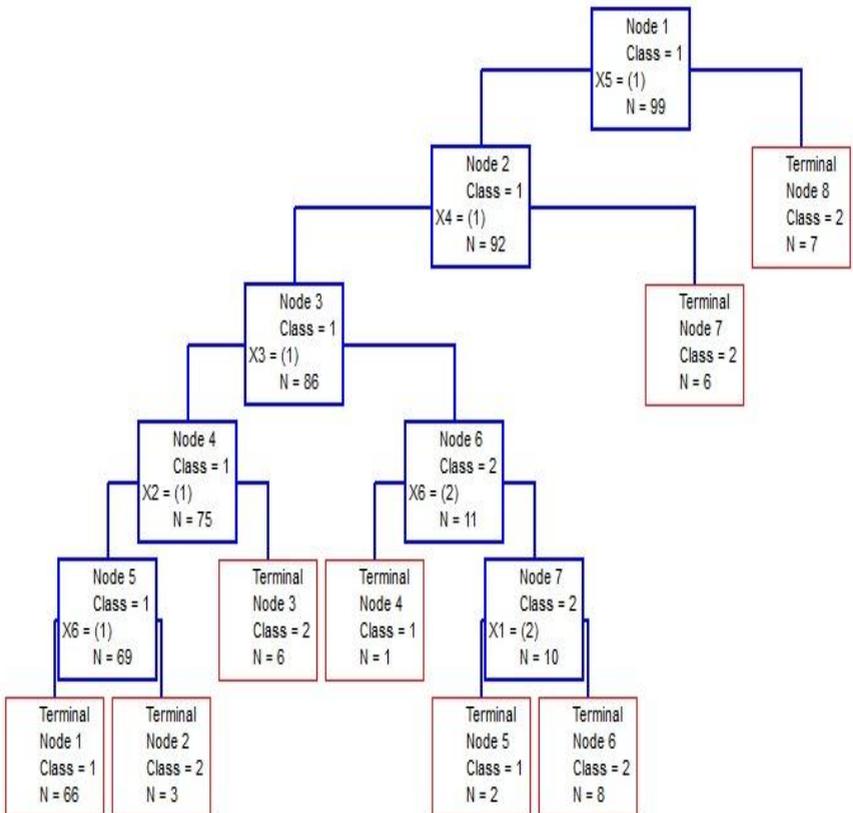
klasifikasi saat menggunakan *cross-validation estimate*. Sedangkan, *complexity parameter* merupakan *cost* bagi penambahan satu simpul terminal. Berdasarkan tabel tersebut dapat diketahui bahwa pohon pertama merupakan pohon klasifikasi maksimal dengan jumlah simpul terminal sebanyak 8 simpul. *Relative cost* yang dimiliki oleh pohon tersebut adalah di antara 0,682 dan 0,858 dengan nilai kompleksitas parameter sebesar 0,000. Pohon pertama tersebut memiliki nilai *cross-validation relative cost* terendah sehingga pohon tersebut juga merupakan pohon klasifikasi optimal. Sehingga dapat disimpulkan bahwa pemangkasan (*pruning*) pada pohon klasifikasi maksimal tidak perlu dilakukan.

4.2.4 Penentuan Pohon Klasifikasi Optimal

Pohon klasifikasi maksimal yang telah dihasilkan merupakan pohon klasifikasi yang sudah optimal karena memiliki *relative cost* terkecil, sehingga tidak perlu dilakukan pemangkasan (*pruning*). Langkah berikutnya dalam pembentukan *classification trees* dengan metode CART adalah menentukan pohon klasifikasi optimal yang kemudian digunakan untuk mengetahui nilai akurasi yang dihasilkan dari klasifikasi ketahanan pangan rumah tangga di Surabaya berdasarkan pemilahan indeks gini. Variabel-variabel yang berperan penting dalam pembentukan pohon klasifikasi dapat dilihat berdasarkan skor kontribusi pada Tabel 4.2. Berdasarkan tabel tersebut dapat diketahui bahwa variabel X_5 atau cara suatu rumah tangga dalam memperoleh makanan pokok memiliki kontribusi terbesar dalam pembentukan pohon klasifikasi yang dihasilkan. Proses pemilahan dengan metode pemilah indeks gini yang telah dibentuk oleh variabel tersebut ditampilkan pada Gambar 4.6.

Keseluruhan data *training* sebanyak 99 data dipilah menjadi dua simpul anak yakni simpul kanan dan simpul kiri berdasarkan cara suatu rumah tangga memperoleh makanan pokok. Jika suatu rumah tangga tidak berhutang dalam memperoleh makanan pokok (kategori 1) maka akan dimasukkan ke dalam simpul kiri dan dilabelkan menjadi kelas 1 (tahan pangan) dan jika

suatu rumah tangga berhutang dalam memperoleh makanan pokok (kategori 2) maka akan dimasukkan ke dalam simpul kanan dan dilabelkan menjadi kelas 2 (rawan pangan). Terdapat sebanyak 92 rumah tangga yang masuk dalam simpul kiri dan terdapat sebanyak 7 rumah tangga yang masuk dalam simpul kanan. Selanjutnya pemilahan dilakukan secara *recursive* hingga tidak dapat dipilah kembali dan menghasilkan simpul akhir atau simpul terminal dari hasil pemilahan yang telah dilakukan.



Gambar 4.6 Pohon Klasifikasi Optimal

Jumlah simpul terminal dan tingkat kedalaman yang dihasilkan oleh pohon klasifikasi optimal sama dengan pohon klasifikasi maksimal, yakni sebanyak 8 simpul terminal dan tingkat kedalaman sebesar 5. Langkah selanjutnya adalah memberi label kelas. Seluruh simpul terminal yang telah dihasilkan oleh pohon klasifikasi optimal memiliki karakteristik tertentu berdasarkan kelas pada variabel dependen yang telah diprediksi. Pemberian label kelas bertujuan untuk mengetahui karakteristik yang dihasilkan tersebut. Tabel 4.4 memberikan informasi mengenai kelas serta jumlah rumah tangga pada masing-masing simpul terminal.

Tabel 4.4 Label Kelas Simpul Terminal

Kelas	Simpul Terminal	Jumlah Rumah Tangga
1 (Tahan Pangan)	1	66
	4	1
	5	2
2 (Rawan Pangan)	2	3
	3	6
	6	8
	7	6
	8	7

Berdasarkan Tabel 4.4 dapat diketahui bahwa simpul terminal 1, 4, dan 5 memiliki label kelas 1. Hal ini berarti simpul-simpul tersebut dikategorikan sebagai rumah tangga tahan pangan. Sebaliknya, simpul terminal 2, 3, 6, 7, dan 8 memiliki label kelas 2 yang berarti simpul-simpul tersebut dikategorikan sebagai rumah tangga rawan pangan.

Simpul terminal 1 merupakan simpul yang diberi label kelas 1. Struktur sekuensial dari simpul terminal ini mengindikasikan bahwa rumah tangga mengkonsumsi protein hewani atau hewani dan nabati, memiliki jarak rumah ke pasar terdekat kurang dari sama dengan dua kilometer, memiliki jumlah anggota keluarga kurang dari 7 orang, memiliki kepala keluarga yang mengenyam pendidikan, serta tidak berhutang dalam memperoleh makanan pokok. Banyaknya rumah tangga yang masuk dalam kelompok ini adalah sebanyak 66 rumah tangga.

Simpul terminal 2 merupakan simpul yang diberi label kelas 2 dengan banyaknya rumah tangga yang masuk dalam kelompok ini adalah sebanyak 3 rumah tangga. Berdasarkan struktur sekuensialnya, rumah tangga yang masuk ke dalam kelompok ini diindikasikan mengkonsumsi protein nabati saja atau tidak mengkonsumsi protein sama sekali, memiliki jarak rumah ke pasar terdekat kurang dari sama dengan dua kilometer, memiliki jumlah anggota keluarga kurang dari 7 orang, memiliki kepala keluarga yang mengenyam pendidikan, serta tidak berhutang dalam memperoleh makanan pokok.

Label kelas yang dimiliki oleh simpul terminal 3 adalah kelas 2. Struktur sekuensial pada simpul terminal ini mengindikasikan bahwa rumah tangga yang masuk ke dalam kelompok ini memiliki rumah dengan jarak ke pasar terdekat lebih dari dua kilometer, memiliki jumlah anggota keluarga kurang dari 7 orang, memiliki kepala keluarga yang mengenyam pendidikan, serta tidak berhutang dalam memperoleh makanan pokok. Terdapat 6 rumah tangga yang masuk ke dalam kelompok ini.

Berbeda dengan simpul terminal 3, label kelas yang dimiliki oleh simpul terminal 4 adalah kelas 1. Rumah tangga pada simpul terminal ini memiliki karakteristik yakni mengkonsumsi protein berupa protein nabati saja atau tidak mengkonsumsi protein sama sekali, memiliki anggota keluarga lebih dari sama dengan 7 orang, memiliki kepala keluarga yang mengenyam pendidikan, serta tidak berhutang dalam memperoleh makanan pokok. Rumah tangga yang masuk ke dalam kelompok ini adalah sebanyak 1 rumah tangga.

Simpul terminal 5 juga memiliki label kelas 1 dengan jumlah rumah tangga pada simpul ini adalah sebanyak 2 rumah tangga. Karakteristik yang dimiliki oleh simpul terminal ini adalah memiliki persediaan makanan pokok berupa beras kurang dari 20 hari dalam sebulan, mengkonsumsi protein berupa protein hewani atau hewani dan nabati, memiliki anggota keluarga lebih dari sama dengan 7 orang, memiliki kepala keluarga yang me-

ngenyam pendidikan, serta tidak berhutang dalam memperoleh makanan pokok.

Struktur sekuensial simpul terminal 6 mengindikasikan rumah tangga dengan karakteristik memiliki persediaan makanan pokok berupa beras lebih dari sama dengan 20 hari per bulan, mengkonsumsi protein berupa protein hewani atau hewani dan nabati, memiliki anggota keluarga lebih dari sama dengan 7 orang, memiliki kepala keluarga yang mengenyam pendidikan, serta tidak berhutang dalam memperoleh makanan pokok. Rumah tangga yang masuk ke dalam kelompok ini adalah sebanyak 8 rumah tangga.

Terdapat 6 rumah tangga yang masuk dalam simpul terminal 7 yang memiliki label kelas 2 atau rumah tangga yang dika-tegorikan rawan pangan. Struktur sekuensial simpul terminal 7 mengindikasikan bahwa rumah tangga dalam kelompok ini memiliki karakteristik yakni memiliki kepala rumah tangga yang tidak mengenyam pendidikan dan tidak berhutang dalam memperoleh makanan pokok.

Simpul terminal yang terakhir adalah simpul terminal 8 dengan label kelas 2. Terdapat 7 rumah tangga dalam simpul terminal ini dengan karakteristik yang dimiliki adalah mengkonsumsi protein berupa protein nabati saja atau tidak mengkonsumsi protein sama sekali.

4.2.5 Perhitungan Hasil Ketepatan Klasifikasi *Classification Trees*

Langkah terakhir pada analisis CART adalah perhitungan hasil ketepatan klasifikasi untuk mengetahui tingkat akurasi dari klasifikasi ketahanan pangan rumah tangga yang telah dilakukan. Perhitungan ketepatan yang klasifikasi yang digunakan dalam analisis ini adalah *total accuracy rate*. *Total accuracy rate* (1-APER) merupakan proporsi observasi yang diprediksi secara benar. Perhitungan ini dilakukan pada data *training* serta data *testing*.

Tabel 4.5 menunjukkan hasil klasifikasi ketahanan pangan rumah tangga pada data *training*, yakni sebanyak 69 rumah tangga dikategorikan ke dalam rumah tangga tahan pangan dan sebanyak 30 rumah tangga dikategorikan ke dalam rumah tangga rawan pangan. Hasil ketepatan klasifikasi yang dihasilkan pada data *training* adalah sebesar 59,6%. Hal ini berarti terdapat 59,6% rumah tangga yang diklasifikasikan secara tepat pada pembentukan pohon klasifikasi optimal ini.

Tabel 4.5 Ketepatan Klasifikasi CART

Observasi		Prediksi		Total	Total Accuracy Rate
		Tahan Pangan	Rawan Pangan		
Data Training	Tahan Pangan	37	8	45	59,6%
	Rawan Pangan	32	22	54	
	Total	69	30	99	
Data Testing	Tahan Pangan	8	2	10	52,0%
	Rawan Pangan	10	5	15	
	Total	18	7	25	

Tabel 4.5 tersebut juga menunjukkan ketepatan klasifikasi untuk data *testing* yakni sebesar 52,0%. Hal ini menunjukkan bahwa model klasifikasi pohon optimal yang telah terbentuk memiliki keakuratan hasil prediksi sebesar 52,0%.

4.3 Klasifikasi Ketahanan Pangan Rumah Tangga di Surabaya dengan CART ARCING

Adaptive Resampling and Combining (ARCING) merupakan suatu metode yang dapat mengecilkan kasus kesalahan klasifikasi dengan cara menekan nilai varian dalam kelompok klasifikasinya. Pada metode ini dilakukan *resampling* pada data dengan Persamaan (2.9). *Resampling* tersebut dilakukan sebanyak kombinasi pohon yang telah ditetapkan secara bebas. Tabel 4.6

menyajikan hasil perhitungan ketepatan klasifikasi CART ARCING pada data *testing* dalam analisis ini.

Tabel 4.6 Ketepatan Klasifikasi CART ARCING

<i>Total Accuracy Rate (%)</i>	Kombinasi Pohon					
	25	50	75	100	125	150
<i>Data Training</i>	66,7	50,0	66,7	66,7	66,7	66,7
<i>Data Testing</i>	56,0	56,0	56,0	56,0	56,0	56,0

Terdapat sebanyak 6 kombinasi pohon yang digunakan dalam analisis CART ARCING ini, yakni 25, 50, 75, 100, 125, dan 150. Berdasarkan Tabel 4.6 dapat diketahui bahwa nilai ketepatan klasifikasi untuk data *training* telah konvergen pada angka 66,7% dan untuk data *testing* nilai ketepatan klasifikasinya telah konvergen pada angka 56%. Sehingga dapat disimpulkan bahwa nilai ketepatan klasifikasi ketahanan pangan rumah tangga di Surabaya pada data *testing* dengan metode CART ARCING adalah sebesar 56,0% untuk data *testing* dan 66,7% untuk data *training*.

4.4 Perbandingan Hasil Klasifikasi CART dan CART ARCING

Baik metode CART maupun CART ARCING menghasilkan nilai ketepatan klasifikasi untuk data *training* dan *testing*. Nilai ketepatan klasifikasi tersebut yang akan dibandingkan dan ditarik kesimpulannya. Tabel 4.7 memberikan informasi mengenai hasil ketepatan klasifikasi data *training* dan *testing* ketahanan pangan rumah tangga di Surabaya dengan menggunakan metode CART dan CART ARCING.

Tabel 4.7 Perbandingan Ketepatan Klasifikasi

<i>Total Accuracy Rate (%)</i>	Metode						
	CART	CART ARCING					
		25	50	75	100	125	150
<i>Data Training</i>	59,6	66,7	50,0	66,7	66,7	66,7	66,7
<i>Data Testing</i>	52,0	56,0	56,0	56,0	56,0	56,0	56,0

Berdasarkan Tabel 4.7 dapat diketahui bahwa nilai *total accuracy rate* yang dihasilkan dari metode CART adalah sebesar

59,6% untuk data *training* dan 52,0% untuk data *testing*. Sedangkan, tingkat akurasi total yang dihasilkan dari metode CART ARCING adalah sebesar 66,7% untuk data *training* dan 56,0% untuk data *testing*. Sehingga, dapat disimpulkan bahwa ketepatan klasifikasi dengan menggunakan metode CART ARCING menghasilkan nilai lebih tinggi. Metode ARCING dapat memperbaiki ketepatan klasifikasi ketahanan pangan rumah tangga di Surabaya sebesar 7,1% untuk data *training* dan 4,0% untuk data *testing*.

(halaman ini sengaja dikosongkan)



BAB V
KESIMPULAN DAN SARAN



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan mengenai pengklasifikasian ketahanan pangan rumah tangga di Surabaya, didapatkan kesimpulan sebagai berikut.

1. Hasil analisis karakteristik faktor-faktor ketahanan pangan dari 124 rumah rumah tangga yang diobservasi diketahui sebanyak 74% memiliki persediaan makanan pokok berupa beras sama dengan atau lebih dari 20 hari dalam satu bulan, 95% memiliki tempat tinggal yang berjarak kurang dari sama dengan dua kilometer dari pasar terdekat, 86% memiliki anggota keluarga kurang dari tujuh orang, 94% memiliki kepala rumah tangga dengan pendidikan terakhir minimal SD, dan 90% mengkonsumsi lauk pauk dengan protein yang terkandung yakni hewani atau hewani dan nabati.
2. Hasil klasifikasi ketahanan pangan rumah tangga di Surabaya dengan metode CART menghasilkan pohon optimal yang terdiri dari 8 simpul terminal dengan tingkat kedalaman sebesar 5. Dari 8 simpul terminal diduga sebanyak 3 simpul terminal untuk rumah tangga tahan pangan dan 5 simpul terminal untuk rumah tangga rawan pangan. Variabel yang paling berpengaruh pada pembentukan pohon klasifikasi optimal ini adalah cara memperoleh makanan pokok dengan ketepatan klasifikasi yang dihasilkan adalah 59,6% untuk data *training* dan 52,0% untuk data *testing*.
3. Pembentukan pohon klasifikasi dengan metode CART AR-CING menghasilkan nilai ketepatan klasifikasi yang lebih tinggi dibandingkan dengan metode CART, yakni sebesar 66,7% untuk data *training* dan 56,0% untuk data *testing*. Ketepatan klasifikasi tersebut lebih tinggi sebesar 7,1% untuk data *training* dan 4,0% untuk data *testing* dibandingkan dengan ketepatan klasifikasi yang dihasilkan oleh metode CART.

5.2 Saran

Saran yang dapat diberikan sebagai tindak lanjut dari penelitian ini yaitu sebaiknya data yang digunakan dalam analisis memiliki ukuran yang lebih besar (*big data*) dengan tujuan untuk meningkatkan ketepatan klasifikasi yang dilakukan. Penambahan variabel juga dapat dilakukan agar karakteristik yang dimiliki dapat lebih terlihat. Selain itu, penambahan *ensemble* atau pendekatan lain juga dapat dilakukan sehingga dapat diketahui pendekatan yang terbaik. Untuk pemerintah Surabaya, dapat dilakukan penambahan lapangan kerja baru dan pemberian bantuan pengobatan pada penderita tuberkulosis sehingga pendapatan masyarakat bertambah dan jumlah rumah tangga yang berhutang dalam memperoleh makanan pokok dapat berkurang.



DAFTAR PUSTAKA



DAFTAR PUSTAKA

- Badan Pusat Statistik. 2007. *Indeks Ketahanan Pangan*, <URL: <http://www.bps.go.id/KegiatanLain>>
- Breiman, L. 1998. Arcing Classifiers. *The Annals of Statistics*, 801-849.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. 1984. *Classification and Regression Tree*. New York - London: Chapman Hall.
- detikFinance. Maret, 2016. *Negara dengan Penduduk Terbanyak di Dunia, Indonesia masuk 4 Besar*, <URL: <https://www.finance.detik.com/ekonomi-bisnis/2517461/negara-dengan-penduduk-terbanyak-di-dunia-ri-masuk-4-besar>>
- Informasipedia. Juli, 2015. *Jumlah Penduduk Dunia Tahun 2015*. <URL: <http://informasipedia.com>>
- Julianto, P. A. Agustus, 2016. *KOMPAS*. <URL: <http://bisniskeuangan.kompas.com>>
- Lewis, R. J. 2000. An Introduction to *Classification and Regression Treess (CART) Analysis*. *Annual Meeting of the Society for Academic Emergency Medicine*.
- Maxwell, S., & Frankenberger, T. 1992. *Household Food Security Concepts, Indicators, and Measurements*. New York: Unicef and Ifad.
- Nugrayasa, O. September, 2015. *Tantangan Ketahanan Pangan Indonesia*. <URL: <http://setkab.go.id>>
- PoskotaNews. Februari, 2016. *19,4 Juta Jiwa Penduduk Indonesia Masih Kelaparan*. <URL: <http://www.poskotanews.com>>
- Pramita, H. 2016. *Pemodelan Penderita Penyakit Tuberkulosis Paru Menurut Stratifikasi Ketahanan Pangan Rumah Tangga di Wilayah Tengah Kota Surabaya dengan Metode Regresi Logistik Biner*.
- Purwantini, T. B. 2014. *Pendekatan Rawan Pangan dan Gizi: Besaran, Karakteristik, dan Penyebabnya*. Bogor: Pusat Sosial Ekonomi dan Kebijakan Pertanian.

- Puslit Kependudukan-LIPI. Februari, 2009. Ketahanan Pangan Rumah Tangga di Pedesaan: Konsep dan Ukuran. *Ketahanan Pangan dan Kemiskinan dalam Konteks Demografi*. <URL: www.ppk.lipi.go.id>
- Steinberg, Dan dan Golovya, Mikhail. 2007 CART 6.0 User's Guide. San Diego: Salford Systems.
- Sutton, C. D. 2005. *Classification and Regression Trees, Bagging, and Boosting. Handbook of Statistics*.
- Widyandoro, I. 2011. Klasifikasi Kesejahteraan Rumah Tangga di Provinsi Jawa Timur dengan Pendekatan CART ARCING.
- Witten, I. H. 2011. *Data Mining Practical Machine Learning Tools and Technique*. New York: Morgan Kauffman.
- Wulandari, S. P., Susilaningrum, D., & Latra, I. N. 2016. *Analisis Ketahanan Pangan Rumah Tangga Terhadap Kasus Penderita Penyakit Tuberculosis dengan Pendekatan Geographically Weighted Poisson Regression (Studi Kasus Pantai Pesisir Surabaya)*.



LAMPIRAN



LAMPIRAN

Lampiran A. Data Ketahanan Pangan Rumah Tangga di Surabaya

Data ke-	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
1	1	1	1	1	1	1	1
2	1	2	1	1	1	1	1
3	1	1	1	1	1	1	2
4	1	1	1	1	1	1	2
5	1	1	1	1	1	1	1
6	2	1	1	1	2	2	2
7	1	1	1	2	1	1	1
8	2	1	2	1	1	1	1
9	1	1	1	1	1	1	1
10	2	1	1	1	1	2	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
119	1	1	1	1	1	1	2
120	1	1	1	1	1	1	2
121	2	1	1	1	1	1	1
123	1	1	1	1	1	1	2
124	2	1	1	1	1	1	2

KETERANGAN

- Y : Ketahanan Pangan
X₁ : Persediaan Makanan Pokok (Beras)
X₂ : Jarak Rumah ke Pasar Terdekat
X₃ : Jumlah Anggota Keluarga
X₄ : Pendidikan Terakhir Kepala Rumah Tangga
X₅ : Cara Memperoleh Makanan Pokok
X₆ : Jenis Protein yang Dikonsumsi

Lampiran B. Output Variable Importance

=====			
VARIABLE IMPORTANCE			
=====			
	Relative Importance	Number Of Categories	Minimum Category

X5	100.000	2	1
X6	90.612	2	1
X4	84.066	2	1
X3	30.247	2	1
X2	29.280	2	1
X1	13.722	2	1

Lampiran C. Output Cross Validation Classification Table

=====				
CROSS VALIDATION CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class		Actual Total	
	1	2		

1	37.00	8.00	45.00	
2	32.00	22.00	54.00	

PRED. TOT.	69.00	30.00	99.00	
CORRECT	0.822	0.407		
SUCCESS IND.	0.368	-0.138		
TOT. CORRECT	0.596			
SENSITIVITY:	0.822	SPECIFICITY:	0.407	
FALSE REFERENCE:	0.464	FALSE RESPONSE:	0.267	
REFERENCE = "1", RESPONSE = "2"				

Lampiran D. Output Tree Sequence

```

=====
TREE SEQUENCE
=====

Dependent variable: Y

Tree   Terminal   Cross-Validated   Resubstitution   Complexity
      Node       Relative Cost     Relative Cost     Parameter
-----
1**    8       0.770 +/- 0.088   0.730            0.000
2      7       0.793 +/- 0.090   0.733            0.002
3      6       0.867 +/- 0.088   0.748            0.007
4      5       0.941 +/- 0.084   0.770            0.011
5      4       0.956 +/- 0.080   0.800            0.015
6      3       0.930 +/- 0.075   0.841            0.020
7      2       1.033 +/- 0.053   0.911            0.035
8      1       1.000 +/- 0.000   1.000            0.044

Initial misclassification cost = 0.500
Initial class assignment = 1

```

Lampiran E. Output Missclassification

```

=====
MISCLASSIFICATION BY CLASS
=====

(Cross Validation)

      Prior
Class  Prob  Wgt Count  Count  Wgt
      Misclass  Misclass  Cost
-----
1      0.500  45.00   45     7.00
      (45.00  45     8.00
2      0.500  54.00   54    31.00
      (54.00  54    32.00
      32    0.593)
-----
Total  1.000  99.00   99    38.00
      (99.00  99    40.00
      40)

```

Lampiran F. Informasi Simpul Terminal

```

=====
TERMINAL NODE INFORMATION
=====
[Breiman adjusted cost, lambda = 0.004]

```

Node	Class	Wgt	Count	N	Prob	Cost	Parent Complexity
1	1	66.00	66	0.678	0.410	0.007	[0.418]
	1	36.00	36	0.590			
	2	30.00	30	0.410			
2	2	3.00	3	0.030	0.375	0.007	[0.543]
	1	1.00	1	0.375			
	2	2.00	2	0.625			
3	2	6.00	6	0.059	0.375	0.015	[0.464]
	1	2.00	2	0.375			
	2	4.00	4	0.625			
4	1	1.00	1	0.011	0.000	0.011	[0.380]
	1	1.00	1	1.000			
	2	0.00	0	0.000			
5	1	2.00	2	0.020	0.455	0.002	[0.688]
	1	1.00	1	0.545			
	2	1.00	1	0.455			
6	2	8.00	8	0.078	0.286	0.002	[0.355]
	1	2.00	2	0.286			
	2	6.00	6	0.714			
7	2	6.00	6	0.057	0.194	0.035	[0.285]
	1	1.00	1	0.194			

	2	5.00	5	0.806		
8	2	7.00	7	0.067	0.167	0.044 [0.246]
	1	1.00	1	0.167		
	2	6.00	6	0.833		
Node	Learn					
1	66.000	36.000			30.000	
2	3.000	1.000			2.000	
3	6.000	2.000			4.000	
4	1.000	1.000			0.000	
5	2.000	1.000			1.000	
6	8.000	2.000			6.000	
7	6.000	1.000			5.000	
8	7.000	1.000			6.000	

Lampiran G. Output Ketepatan Klasifikasi CART untuk Data Training

=====			
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE			
=====			
Actual Class	Predicted Class		Actual Total
	1	2	
-----	-----	-----	-----
1	37.00	8.00	45.00
2	32.00	22.00	54.00
-----	-----	-----	-----
PRED. TOT.	69.00	30.00	99.00
CORRECT	0.822	0.407	
SUCCESS IND.	0.368	-0.138	
TOT. CORRECT	0.596		
SENSITIVITY:	0.822	SPECIFICITY:	0.407
FALSE REFERENCE:	0.464	FALSE RESPONSE:	0.267
REFERENCE = "1", RESPONSE = "2"			

Lampiran H. Output Ketepatan Klasifikasi CART untuk Data Testing

=====				
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class		Actual Total	
	1	2		
1	8.00	2.00	10.00	
2	10.00	5.00	15.00	

PRED. TOT.	18.00	7.00	25.00	
CORRECT	0.800	0.333		
SUCCESS IND.	0.400	-0.267		
TOT. CORRECT	0.520			
SENSITIVITY:	0.800	SPECIFICITY:	0.333	
FALSE REFERENCE:	0.556	FALSE RESPONSE:	0.286	
REFERENCE = "1", RESPONSE = "2"				

Lampiran I. Output Ketepatan Klasifikasi CART 25 Replikasi Data Training

```

=====
COMMITTEE SUMMARY
=====
                                % CORRECT
FIRST TREE ONLY                 0.500
BOOTSTRAP SET                   0.667

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

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

Actual          Predicted Class          Actual
Class           1            2            Total
-----
I 1              2.00         2.00         4.00
B                2.00         2.00
I 2              1.00         1.00         2.00
B                0.00         2.00
-----
INITIAL TREE
PRED. TOT.          3.00         3.00         6.00
CORRECT             0.500        0.500
SUCCESS IND.       -0.167        0.167
Tot. Correct       0.500
-----
BOOTSTRAP SET OF TREES
PRED. TOT.          2.00         4.00
CORRECT             0.500        1.000
SUCCESS IND.       -0.167        0.667
Tot. Correct       0.667
-----
INITIAL TREE
    SENSITIVITY:    0.500    SPECIFICITY:    0.500
FALSE REFERENCE:   0.333  FALSE RESPONSE: 0.667
REFERENCE = "1",  RESPONSE = "2"
-----
BOOTSTRAP SET OF TREES
    SENSITIVITY:    0.500    SPECIFICITY:    1.000
FALSE REFERENCE:   0.000  FALSE RESPONSE: 0.500
REFERENCE = "1",  RESPONSE = "2"

```

**Lampiran J. Output Ketepatan Klasifikasi CART 50 Replikasi
Data Training**

```

=====
COMMITTEE SUMMARY
=====
                % CORRECT
FIRST TREE ONLY          0.500
BOOTSTRAP SET           0.500

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

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

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

```

Lampiran K. Output Ketepatan Klasifikasi CART 75 Replikasi Data Training

```

=====
COMMITTEE SUMMARY
=====
                                % CORRECT
FIRST TREE ONLY                 0.500
BOOTSTRAP SET                   0.667

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

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

Actual      Predicted Class      Actual
Class              1          2          Total
-----
I 1              2.00         2.00         4.00
B                  2.00         2.00
I 2              1.00         1.00         2.00
B                  0.00         2.00
-----
INITIAL TREE
PRED. TOT.              3.00         3.00         6.00
CORRECT                 0.500         0.500
SUCCESS IND.           -0.167         0.167
Tot. Correct            0.500
-----
BOOTSTRAP SET OF TREES
PRED. TOT.              2.00         4.00
CORRECT                 0.500         1.000
SUCCESS IND.           -0.167         0.667
Tot. Correct            0.667
-----
INITIAL TREE
      SENSITIVITY:      0.500      SPECIFICITY:      0.500
FALSE REFERENCE:      0.333  FALSE RESPONSE:      0.667
REFERENCE = "1", RESPONSE = "2"
-----
BOOTSTRAP SET OF TREES
      SENSITIVITY:      0.500      SPECIFICITY:      1.000
FALSE REFERENCE:      0.000  FALSE RESPONSE:      0.500
REFERENCE = "1", RESPONSE = "2"

```

**Lampiran L. Output Ketepatan Klasifikasi CART 100 Replikasi
Data Training**

```

=====
COMMITTEE SUMMARY
=====
                % CORRECT
FIRST TREE ONLY          0.500
BOOTSTRAP SET           0.667

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

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

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

-----
INITIAL TREE
PRED. TOT.          3.00          3.00          6.00
CORRECT            0.500          0.500
SUCCESS IND.       -0.167          0.167
Tot. Correct       0.500

-----
BOOTSTRAP SET OF TREES
PRED. TOT.          2.00          4.00
CORRECT            0.500          1.000
SUCCESS IND.       -0.167          0.667
Tot. Correct       0.667

-----
INITIAL TREE
  SENSITIVITY:      0.500          SPECIFICITY:      0.500
FALSE REFERENCE:    0.333  FALSE RESPONSE:    0.667
REFERENCE = "1", RESPONSE = "2"

-----
BOOTSTRAP SET OF TREES
  SENSITIVITY:      0.500          SPECIFICITY:      1.000
FALSE REFERENCE:    0.000  FALSE RESPONSE:    0.500
REFERENCE = "1", RESPONSE = "2"

```

Lampiran M. Output Ketepatan Klasifikasi CART 125 Replikasi Data Training

```

=====
COMMITTEE SUMMARY
=====
                                % CORRECT
FIRST TREE ONLY                 0.500
BOOTSTRAP SET                   0.667

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

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

Actual      Predicted Class      Actual
Class              1          2          Total
-----
I 1              2.00         2.00         4.00
B                  2.00         2.00
I 2              1.00         1.00         2.00
B                  0.00         2.00
-----
INITIAL TREE
PRED. TOT.              3.00         3.00         6.00
CORRECT                 0.500         0.500
SUCCESS IND.            -0.167         0.167
Tot. Correct            0.500
-----
BOOTSTRAP SET OF TREES
PRED. TOT.              2.00         4.00
CORRECT                 0.500         1.000
SUCCESS IND.            -0.167         0.667
Tot. Correct            0.667
-----
INITIAL TREE
      SENSITIVITY:      0.500      SPECIFICITY:      0.500
FALSE REFERENCE:      0.333  FALSE RESPONSE:      0.667
REFERENCE = "1", RESPONSE = "2"
-----
BOOTSTRAP SET OF TREES
      SENSITIVITY:      0.500      SPECIFICITY:      1.000
FALSE REFERENCE:      0.000  FALSE RESPONSE:      0.500
REFERENCE = "1", RESPONSE = "2"

```

Lampiran N. Output Ketepatan Klasifikasi CART 150 Replikasi Data Training

```

=====
COMMITTEE SUMMARY
=====
                % CORRECT
FIRST TREE ONLY          0.500
BOOTSTRAP SET           0.667

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

I=INITIAL TREE, B=BOOTSTRAP SET OF TREES

Actual          Predicted Class                Actual
Class           1                2                Total
-----
I 1              2.00              2.00              4.00
B                2.00              2.00
I 2              1.00              1.00              2.00
B                0.00              2.00
-----
INITIAL TREE
PRED. TOT.              3.00              3.00              6.00
CORRECT                 0.500             0.500
SUCCESS IND.           -0.167             0.167
Tot. Correct           0.500
-----
BOOTSTRAP SET OF TREES
PRED. TOT.              2.00              4.00
CORRECT                 0.500             1.000
SUCCESS IND.           -0.167             0.667
Tot. Correct           0.667
-----
INITIAL TREE
  SENSITIVITY:         0.500      SPECIFICITY:         0.500
FALSE REFERENCE:      0.333  FALSE RESPONSE:       0.667
REFERENCE = "1", RESPONSE = "2"
-----
BOOTSTRAP SET OF TREES
  SENSITIVITY:         0.500      SPECIFICITY:         1.000
FALSE REFERENCE:      0.000  FALSE RESPONSE:       0.500
REFERENCE = "1", RESPONSE = "2"

```

**Lampiran O. Output Ketepatan Klasifikasi CART ARCING 25
Replikasi Data Testing**

=====				
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class		Actual Total	
	1	2		
1	7.00	3.00	10.00	
2	8.00	7.00	15.00	

PRED. TOT.	15.00	10.00	25.00	
CORRECT	0.700	0.467		
SUCCESS IND.	0.300	-0.133		
TOT. CORRECT	0.560			
SENSITIVITY:	0.700	SPECIFICITY:	0.467	
FALSE REFERENCE:	0.533	FALSE RESPONSE:	0.300	
REFERENCE = "1", RESPONSE = "2"				

**Lampiran P. Output Ketepatan Klasifikasi CART ARCING 50
Replikasi Data Testing**

=====				
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class		Actual Total	
	1	2		
1	7.00	3.00	10.00	
2	8.00	7.00	15.00	

PRED. TOT.	15.00	10.00	25.00	
CORRECT	0.700	0.467		
SUCCESS IND.	0.300	-0.133		
TOT. CORRECT	0.560			
SENSITIVITY:	0.700	SPECIFICITY:	0.467	
FALSE REFERENCE:	0.533	FALSE RESPONSE:	0.300	
REFERENCE = "1", RESPONSE = "2"				

Lampiran Q. Output Ketepatan Klasifikasi CART ARCING 75
Replikasi Data *Testing*

=====				
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class		Actual Total	
	1	2		
1	7.00	3.00	10.00	
2	8.00	7.00	15.00	

PRED. TOT.	15.00	10.00	25.00	
CORRECT	0.700	0.467		
SUCCESS IND.	0.300	-0.133		
TOT. CORRECT	0.560			
SENSITIVITY:	0.700	SPECIFICITY:	0.467	
FALSE REFERENCE:	0.533	FALSE RESPONSE:	0.300	
REFERENCE = "1", RESPONSE = "2"				

Lampiran R. Output Ketepatan Klasifikasi CART ARCING 100
Replikasi Data *Testing*

=====				
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class		Actual Total	
	1	2		
1	7.00	3.00	10.00	
2	8.00	7.00	15.00	

PRED. TOT.	15.00	10.00	25.00	
CORRECT	0.700	0.467		
SUCCESS IND.	0.300	-0.133		
TOT. CORRECT	0.560			
SENSITIVITY:	0.700	SPECIFICITY:	0.467	
FALSE REFERENCE:	0.533	FALSE RESPONSE:	0.300	
REFERENCE = "1", RESPONSE = "2"				

**Lampiran S. Output Ketepatan Klasifikasi CART ARCING 125
Replikasi Data Testing**

=====				
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class		Actual Total	
	1	2		
1	7.00	3.00	10.00	
2	8.00	7.00	15.00	

PRED. TOT.	15.00	10.00	25.00	
CORRECT	0.700	0.467		
SUCCESS IND.	0.300	-0.133		
TOT. CORRECT	0.560			
SENSITIVITY:	0.700	SPECIFICITY:	0.467	
FALSE REFERENCE:	0.533	FALSE RESPONSE:	0.300	
REFERENCE = "1", RESPONSE = "2"				

**Lampiran T. Output Ketepatan Klasifikasi CART ARCING 150
Replikasi Data Testing**

=====				
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class		Actual Total	
	1	2		
1	7.00	3.00	10.00	
2	8.00	7.00	15.00	

PRED. TOT.	15.00	10.00	25.00	
CORRECT	0.700	0.467		
SUCCESS IND.	0.300	-0.133		
TOT. CORRECT	0.560			
SENSITIVITY:	0.700	SPECIFICITY:	0.467	
FALSE REFERENCE:	0.533	FALSE RESPONSE:	0.300	
REFERENCE = "1", RESPONSE = "2"				

Lampiran U. Surat Keterangan Data Sekunder

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini menerangkan bahwa:

1. Mahasiswa Statistika FMIPA-ITS dengan identitas berikut:

Nama : Yusnada Asa Nurani
NRP : 1313 100 016

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

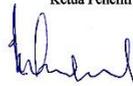
2. Tidak Keberatan/~~Keberatan~~* nama perusahaan dicantumkan dalam Tugas Akhir/Thesis mahasiswa Statistika yang akan di simpan di Perpustakaan ITS dan dibaca di lingkungan ITS.
3. Tidak Keberatan/~~Keberatan~~* bahwa hasil analisis data dari perusahaan dipublikasikan dalam E journal ITS yaitu Jurnal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

BIODATA PENULIS



Penulis bernama Yusnada Asa Nurani yang akrab disapa dengan Nada, merupakan anak kedua dari dua bersaudara. Penulis dilahirkan di Malang pada tanggal 1 Januari 1996. Penulis memulai pendidikan pertamanya di Taman Kanak-kanak Hidayatul Mubtadi'in Malang (2000-2002), kemudian melanjutkan pendidikan sekolah dasar di MIN I Malang (2002-2008). Setelah menamatkan pendidikan sekolah dasar, penulis melanjutkan pendidikan di MTsN I Malang (2008-2010) dan SMAN 1 Malang (2010-2013). Pada tahun 2013, penulis melanjutkan pendidikan di Perguruan Tinggi Negeri Institut Teknologi Sepuluh Nopember dan diterima sebagai mahasiswi jurusan statistika dengan NRP 1313100016. Selama masa perkuliahan, penulis juga aktif dalam kegiatan organisasi dan pernah menjadi Staff Departemen Dalam Negeri di BEM FMIPA ITS pada tahun 2014-2015, kemudian pada tahun 2015-2016 penulis menjabat sebagai Sekretaris Departemen Dalam Negeri di BEM FMIPA ITS. Selain itu, penulis juga aktif dalam kepanitiaan seperti Pekan Raya Stastistika (PRS). Penulis juga memiliki pengalaman magang di PDAM Tirta Dharma Kota Malang. Penulis menerima segala kritikan, masukan, dan saran yang bersifat membangun demi meningkatkan manfaat Tugas Akhir ini. Selain itu, penulis juga siap membantu terkait dengan metode dan ilmu statistik dengan menyampaikannya melalui email: yusnadaasanurani@ymail.com.