



**TUGAS AKHIR - KS184822**

**ANALISIS KARAKTERISTIK TINGKAT KESEJAHTERAAN  
DI KOTA SURABAYA MENGGUNAKAN METODE  
POHON KLASIFIKASI**

**ZIKRARIZA KURNIA MALTA  
NRP 062115 4000 0042**

**Dosen Pembimbing  
Dr. Sutikno, M.Si**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**



**TUGAS AKHIR - KS184822**

**ANALISIS KARAKTERISTIK TINGKAT KESEJAHTERAAN  
DI KOTA SURABAYA MENGGUNAKAN METODE  
POHON KLASIFIKASI**

**ZIKRARIZA KURNIA MALTA  
NRP 062115 4000 0042**

**Dosen Pembimbing  
Dr. Sutikno, M.Si**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**



**FINAL PROJECT - KS184822**

**CHARACTERISTICS ANALYSIS OF WELFARE LEVELS  
IN SURABAYA CITY USING CLASSIFICATION  
TREE METHOD**

**ZIKRARIZA KURNIA MALTA  
SN 062115 4000 0042**

**Supervisor  
Dr. Sutikno, M.Si**

**UNDERGRADUATE PROGRAMME  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**

## LEMBAR PENGESAHAN

### ANALISIS KARAKTERISTIK TINGKAT KESEJAHTERAAN DI KOTA SURABAYA MENGUNAKAN METODE Pohon KLASIFIKASI

#### TUGAS AKHIR

Diajukan untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Statistika  
pada  
Program Studi Sarjana Departemen Statistika  
Fakultas Matematika, Komputasi, dan Sains Data  
Institut Teknologi Sepuluh Nopember

Oleh :

**Zikrariza Kurnia Malta**  
NRP. 062115 4000 0042

Disetujui oleh Pembimbing:

**Dr. Sutikno, M.Si.**

NIP. 19710313 199702 1 001

(  )



SURABAYA, JULI 2019

# ANALISIS KARAKTERISTIK TINGKAT KESEJAHTERAAN DI KOTA SURABAYA MENGUNAKAN METODE POHON KLASIFIKASI

Nama Mahasiswa : Zikrariza Kurnia Malta  
NRP : 062115 4000 0042  
Departemen : Statistika-FMKSD-ITS  
Dosen Pembimbing : Dr. Sutikno, M.Si

## Abstrak

*Terdapat beberapa upaya pengentasan kemiskinan yang dilakukan Pemerintah Kota Surabaya, salah satunya melalui kerjasama dengan Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) yang mengeluarkan data rumah tangga dengan tingkat pendapatan rendah di Surabaya yang didapatkan melalui metode Proxy Means Testing (PMT). Namun, dari hasil tersebut tidak dijelaskan apa saja karakteristik setiap tingkat kesejahteraan. Karakteristik yang terdapat pada setiap desil dapat diketahui dengan melakukan pengklasifikasian. Salah satu metode statistika yang dapat digunakan yaitu pohon klasifikasi. Variabel respon yang digunakan yaitu tingkat kesejahteraan. Data yang digunakan yaitu 149.186 data rumah tangga dan 37 variabel indikator kesejahteraan. Selanjutnya dilakukan metode test sample dengan testing 5%-30% menggunakan Indeks Gini dan Indeks Twoing. Metode terbaik yang digunakan yaitu metode test sample dengan testing 30% menggunakan Indeks Gini. Hasil yang didapatkan yaitu pohon klasifikasi optimal dengan 391 simpul dan 191 simpul terminal. Karakteristik setiap desil didapatkan dari persentase tertinggi pada simpul terminal setiap tingkat kesejahteraan. Lima variabel terpenting dalam pemilahan pohon klasifikasi yaitu jumlah anggota rumah tangga, jenis lantai terluas dengan bahan marmor/granit, keramik, parket/vinil/permadani, kepemilikan kulkas, tipe toilet pribadi dengan jenis kloset leher angsa, dan kepemilikan sepeda motor. Pohon klasifikasi tersebut merupakan pohon klasifikasi dengan right estimated dengan tingkat keakuratan klasifikasi sebesar 64.1%.*

**Kata Kunci:** Indikator Kesejahteraan, PMT, Pohon Klasifikasi, Rumah Tangga, Test Sample.

*(Halaman ini sengaja dikosongkan)*

# CHARACTERISTICS ANALYSIS OF WELFARE LEVELS IN SURABAYA CITY USING CLASSIFICATION TREE METHOD

**Name** : Zikrariza Kurnia Malta  
**Student Number** : 062115 4000 0042  
**Department** : Statistics-FMKSD-ITS  
**Supervisor** : Dr. Sutikno, M.Si

## Abstract

*There are several poverty alleviation efforts carried out by the Surabaya City Government, one of which is in collaboration with Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) which issues household data with low income levels in Surabaya obtained through the Proxy Means Testing (PMT) method. However, the results do not explain what the characteristics of each level of welfare are. The characteristics found in each decile can be known by classifying. One of the statistical methods that can be used is the classification tree. The response variable used is the level of welfare. The data used were 149,186 household data and 35 welfare indicator variables. Then the test sample method is carried out by testing 5% -30% using the Gini Index and Twoing Index. The best method used is the test sample method with testing 30% using the Gini Index. The results obtained are optimal classification trees with 391 nodes and 191 terminal nodes. The characteristics of each decile are obtained from the highest percentage of terminal nodes for each level of welfare. The five most important variables in sorting classification trees are the number of household members, the widest type of floor with marble / granite material, ceramics, parquet / vinyl / rug, ownership of the refrigerator, type of private toilet with the type of goose neck toilet, and motorcycle ownership. The classification tree is a right estimated classification tree with a classification accuracy rate of 64.1%.*

**Keywords:** Classification Trees, Households, PMT, Sample Tests, Welfare Indicators.

*(Halaman ini sengaja dikosongkan)*



## KATA PENGANTAR

Puji syukur kehadirat Allah SWT, atas limpahan rahmat yang tidak pernah berhenti sehingga penulis dapat menyelesaikan penyusunan Tugas Akhir yang berjudul

### **“Analisis Karakteristik Tingkat Kesejahteraan di Kota Surabaya Menggunakan Metode Pohon Klasifikasi”**

dengan baik. Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini tidak terlepas dari bantuan dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT yang telah melimpahkan segala rahmatNya sehingga penulis mampu menyelesaikan Tugas Akhir dengan lancar.
2. Ayah, ibu, eyang, adik-adik, dan seluruh keluarga dirumah atas segala doa, kasih sayang, dan dukungan yang diberikan kepada penulis.
3. Bapak Dr. Suhartono selaku Kepala Departemen Statistika yang telah memberikan banyak fasilitas untuk kelancaran penyelesaian Tugas Akhir ini.
4. Bapak Dr. Sutikno, M.Si selaku dosen pembimbing yang telah sabar dalam memberikan bimbingan dan saran, serta meluangkan segala kesempatan dan waktu yang ada untuk memberikan bimbingan terhadap Tugas Akhir ini.
5. Bapak Dr. Bambang Widjanarko Otok, M.Si dan Bapak Jerry Dwi Trijoyo Purnomo, S.Si, M.Si, Ph.D selaku dosen penguji atas saran dan kritiknya yang sangat membangun.
6. Bapak Dr. Muhammad Mashuri, M.T. selaku dosen wali atas dukungan, nasehat, dan saran yang diberikan pada saat perwalian.
7. Teman-teman terdekat Ela Wulansari, Riska Fitri Amalia, Anindya Shafira Putri, Novita Dwi Utami, Naomi Puspita,

Maila S. Baladina, Ayu Febriana Dwi Rositawati, Iftitah Ayundari, Dissa, Ainun, Mbak Halumma, Rhavida, Juwitasari, dan Imroatus yang sudah menyemangati, memberikan tutor, dan selalu menemani penulis dalam pengerjaan tugas akhir ini.

8. Teman-teman seperjuangan Statistika 2015 Vivacious  $\Sigma 26$  yang lebih dekat dari tetangga dan lebih besar dari keluarga.
9. Teman-teman KSE ITS dan KSE se-Nusantara yang sudah memberi warna dalam hari-hari pengerjaan tugas akhir ini.
10. Keluarga UKM Cinta Rebana khususnya teman-teman CR 2015 yang sudah memberi support dan semangat dalam mengerjakan tugas akhir
11. Serta semua pihak yang telah mendukung dan tidak dapat penulis sebutkan satu persatu.

Penulis sangat berharap hasil Tugas Akhir ini dapat bermanfaat bagi kita semua serta saran dan kritik yang bersifat membangun guna perbaikan di masa mendatang.

Surabaya, Juli 2019

**Penulis**

## DAFTAR ISI

	Halaman
<b>HALAMAN JUDUL</b> .....	i
<b>LEMBAR PENGESAHAN</b> .....	v
<b>ABSTRAK</b> .....	vii
<b>ABSTRACT</b> .....	ix
<b>KATA PENGANTAR</b> .....	xi
<b>DAFTAR ISI</b> .....	xiii
<b>DAFTAR GAMBAR</b> .....	xv
<b>DAFTAR TABEL</b> .....	xvii
<b>DAFTAR LAMPIRAN</b> .....	xix
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan.....	3
1.4 Manfaat.....	3
<b>BAB II TINJAUAN PUSTAKA</b> .....	5
2.1 Statistika Deskriptif.....	5

2.2 Pohon Klasifikasi.....	7
2.2.1 Pembentukan Pohon Klasifikasi.....	10
2.2.2 Pemangkasan Pohon Klasifikasi.....	11
2.2.3 Pemilihan Pohon Optimal.....	12
2.3 Ketepatan Klasifikasi.....	12
2.4 Konsep dan Definisi Tingkat Kesejahteraan.....	13
<b>BAB III METODOLOGI PENELITIAN.....</b>	<b>15</b>
3.1 Sumber Data.....	15
3.2 Variabel Penelitian.....	15
3.3 Struktur Data.....	18
3.4 Langkah Penelitian.....	18
3.5 Diagram Alir.....	20
<b>BAB IV ANALISIS DAN PEMBAHASAN.....</b>	<b>21</b>
4.1 <i>Pre-processing</i> Data.....	21
4.2 Karakteristik Rumah Tangga di Surabaya .....	22
4.3 Pohon Klasifikasi.....	35
<b>BAB V KESIMPULAN DAN SARAN.....</b>	<b>49</b>
5.1 Kesimpulan.....	49

5.2 Saran.....	52
<b>DAFTAR PUSTAKA.....</b>	<b>53</b>
<b>LAMPIRAN.....</b>	<b>55</b>
<b>BIODATA PENULIS.....</b>	<b>63</b>

## DAFTAR GAMBAR

<b>Gambar 2.1</b>	Visualisasi Box plot.....	6
<b>Gambar 2.2</b>	Struktur Pohon Klasifikasi .....	9
<b>Gambar 3.1</b>	Peta Surabaya.....	15
<b>Gambar 3.2</b>	Diagram Alir Langkah Analisis .....	20
<b>Gambar 4.1</b>	Presentase tiap desil pada variabel status kesejahteraan.....	21
<b>Gambar 4.2</b>	Proporsi rumah tangga kepemilikan aset rumah tangga.....	23
<b>Gambar 4.3</b>	Persentase rumah tangga yang memakai kayu bakar dan tidak memasak di rumah.....	24
<b>Gambar 4.4</b>	Proporsi rumah tangga pemakai sumber air minum ledeng eceran dan sumur bor/pompa, air isi ulang dan ledeng meteran, dan air kemasan bermerk.....	25
<b>Gambar 4.5</b>	Persentase tangga dengan status penguasaan bangunan tempat tinggal pada kategori kontrak/sewa dan dinas..	26
<b>Gambar 4.6</b>	Proporsi rumah tangga yang menggunakan jenis lantai terluas dari ketiga kategori.....	27
<b>Gambar 4.7</b>	Box plot Sebaran data variabel luas lantai per kapita..	28
<b>Gambar 4.8</b>	Persentase rumah tangga yang menggunakan jenis dinding terluas berkategori lain .....	29
<b>Gambar 4.9</b>	Proporsi rumah tangga yang menggunakan 3 tipe toilet sesuai variabel penelitian .....	29
<b>Gambar 4.10</b>	Proporsi anggota rumah tangga yang masih bersekolah dan sudah menamatkan pendidikan dari variabel kategori pendidikan .....	31
<b>Gambar 4.11</b>	Proporsi anggota rumah tangga yang bekerja dan tidak bekerja pada variabel kategori pekerjaan .....	33
<b>Gambar 4.12</b>	Box plot sebaran data jumlah anggota rumah tangga pada setiap rumah tangga.....	34

*(Halaman ini sengaja dikosongkan)*

## DAFTAR TABEL

<b>Tabel 2.1</b>	Tabulasi Silang Variabel x dengan Variabel y .....	6
<b>Tabel 2.2</b>	Contoh Tabel Prediksi Klasifikasi dengan 4 Level Kelas Variabel Respon .....	13
<b>Tabel 3.1</b>	Variabel Penelitian .....	16
<b>Tabel 3.2</b>	Struktur Data Penelitian.....	18
<b>Tabel 4.1</b>	Statistika Deskriptif dari Variabel Luas Lantai per Kapita .....	27
<b>Tabel 4.2</b>	Jumlah anggota rumah tangga yang masih bersekolah dan sudah menamatkan pendidikan pada variabel kategori pendidikan .....	30
<b>Tabel 4.3</b>	Jumlah anggota rumah tangga yang bekerja dari variabel kategori pekerjaan.....	32
<b>Tabel 4.4</b>	Statistika deskriptif dari variabel jumlah anggota rumah tangga dan anggota rumah tangga laki-laki.....	33
<b>Tabel 4.5</b>	Perbandingan Ketepatan Klasifikasi dari Metode Test Sample .....	35
<b>Tabel 4.6</b>	Lima simpul terminal dengan persentase terbesar pada setiap tingkat kesejahteraan .....	37
<b>Tabel 4.7</b>	Karakteristik setiap tingkat kesejahteraan berdasarkan persentase terbesar simpul node pada setiap tingkat kesejahteraan .....	45
<b>Tabel 4.8</b>	Lima variabel dengan nilai <i>importance</i> tertinggi.....	46
<b>Tabel 4.9</b>	Kesesuaian Klasifikasi Pohon Klasifikasi Pada Setiap Tingkat Kesejahteraan Rumah Tangga di Kota Surabaya	47



*(Halaman ini sengaja dikosongkan)*

## DAFTAR LAMPIRAN

<b>Lampiran 1.</b>	Data Hasil <i>Pre-processing</i> .....	55
<b>Lampiran 2.</b>	<i>Output</i> Model Summary.....	56
<b>Lampiran 3.</b>	<i>Cross Tabulation Data Learning dan Testing</i> .....	56
<b>Lampiran 4.</b>	Variabel Penting dalam Pembentukan Pohon ....	57
<b>Lampiran 5.</b>	Diagram batang Normalized Importance .....	60
<b>Lampiran 6.</b>	Topografi Pohon Klasifikasi Kedalaman 3.....	61
<b>Lampiran 7.</b>	Surat Pernyataan Data.....	62

*(Halaman ini sengaja dikosongkan)*

# **BAB I**

## **PENDAHULUAN**

Bab ini membahas mengenai latar belakang yang mendasari dilakukannya penelitian ini. Selanjutnya merumuskan masalah dan menetapkan tujuan penelitian. Selain itu, pada bab ini juga dibahas mengenai manfaat dan batasan masalah dari penelitian yang dilakukan.

### **1.1 Latar Belakang**

Kemiskinan menjadi salah satu masalah utama pembangunan di Indonesia. Kemiskinan merupakan sebuah kondisi yang berada di bawah garis nilai standar kebutuhan minimum, baik untuk makanan dan non makanan yang disebut garis kemiskinan. Badan Pusat Statistik (BPS) mendefinisikan kemiskinan sebagai ketidakmampuan individu dalam memenuhi kebutuhan dasar minimal untuk layak hidup. Kemiskinan dapat memunculkan berbagai permasalahan seperti semakin banyaknya kejahatan, menurunnya tingkat pendidikan dan kesehatan masyarakat karena minimnya pendapatan masyarakat miskin. Badan Pusat Statistik menggunakan konsep kemampuan memenuhi kebutuhan dasar (basic needs approach) untuk mengukur kemiskinan. Dengan pendekatan ini, kemiskinan dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. BPS mencatat bahwa jumlah penduduk miskin di Indonesia pada tahun 2018 mencapai 25,95 juta jiwa atau sekitar 9,82 persen dari jumlah penduduk Indonesia. Jumlah ini mengalami penurunan dibandingkan

dengan dengan tahun 2017 mencapai 27,77 juta jiwa atau sekitar 10,64 persen dari total penduduk Indonesia.

Masalah kemiskinan selalu menjadi agenda utama dalam pembangunan di Indonesia, termasuk salah satunya di Kota Surabaya. Upaya pengentasan kemiskinan yang dilakukan Pemerintah Kota Surabaya salah satunya melalui kerjasama dengan Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) yang mengeluarkan data rumah tangga dengan tingkat pendapatan rendah di Surabaya yang didapatkan melalui metode *Proxy Means Testing* (PMT). Metode ini merupakan metode yang digunakan untuk memprediksikan pendapatan sebuah rumah tangga dengan cara mengumpulkan informasi sederhana tentang aset yang dimiliki oleh suatu rumah tangga. Dengan metode PMT, TNP2K mendapatkan estimasi konsumsi/pengeluaran rumah tangga melalui variabel-variabel yang digunakan sebagai indikator penentu tingkat kesejahteraan suatu rumah tangga. Dalam perhitungan menggunakan metode PMT yang dilakukan oleh TNP2K secara nasional digunakan sebanyak 41 variabel yang berhubungan dengan kesejahteraan. Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) menetapkan sebanyak 40% rupa di Kota Surabaya hasil perhitungan dengan metode PMT memiliki tingkat konsumsi atau pendapatan terendah. Hasil tersebut kemudian diklasifikasikan ke dalam empat desil atau empat golongan. Desil satu atau golongan satu menunjukkan bahwa tingkat konsumsi rupa tersebut lebih rendah dibandingkan dengan tingkat konsumsi rupa pada desil dua atau golongan dua. Namun, dari hasil tersebut tidak dijelaskan apa saja karakteristik setiap tingkat kesejahteraan di Kota Surabaya dan variabel apa saja yang membedakan kriteria setiap tingkat kesejahteraan tersebut.

Karakteristik yang terdapat pada setiap desil dapat diketahui dengan melakukan pengklasifikasian. Salah satu metode statistika yang dapat digunakan yaitu *Classification and Regression Tree* (CART). Metode CART dengan variabel respon numeric menggunakan metode pohon regresi sedangkan jika menggunakan variabel respon kategorik menggunakan pohon klasifikasi. Pada kasus ini, variabel respon merupakan variabel kategorik sehingga digunakan pohon klasifikasi pada penelitian ini. Metode ini merupakan metode nonparametrik berupa pohon keputusan yang dapat mengklasifikasikan suatu observasi menjadi sub kelompok dengan karakteristik lebih rinci dan hasil lebih mudah divisualisasikan interpretasinya karena dalam bentuk gambar. Beberapa penelitian yang menggunakan metode pohon klasifikasi telah banyak dilakukan, diantaranya oleh Aritonang, Rusgiyono, dan Rahmawati (2016) yang meneliti tentang status kerja pada angkatan kerja Kota Semarang yang menunjukkan bahwa metode pohon klasifikasi mampu mengeksplorasi variabel-variabel independen yang berpengaruh dan merepresentasikan hasil variabel dependen dari pengelompokan data yang terbentuk. Penelitian lain dilakukan oleh Inayah, Purnami, dan Otok (2014) yang meneliti tentang rumah tangga miskin di Kabupaten Jombang berdasarkan faktor yang mempengaruhi. Selain itu juga penelitian yang meneliti tentang kemiskinan yaitu penelitian oleh Mahalisa (2016) tentang analisis kesejahteraan rumah tangga di Banjarmasin menggunakan metode pohon gabungan pada CART. Metode ini dapat memetakan rumah tangga miskin berdasarkan kriteria variabel yang mempengaruhi terbentuknya suatu desil, sehingga nantinya diharapkan dapat dijadikan dasar pertimbangan keputusan, kebijakan, maupun perbaikan dalam meningkatkan kualitas kesejahteraan penduduk dan mengentas kemiskinan di Surabaya.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang di atas, permasalahan dari penelitian ini adalah bagaimana karakteristik setiap tingkat kesejahteraan di Kota Surabaya dan variabel apa yang berpengaruh terhadap setiap tingkat kesejahteraan rumah tangga di Kota Surabaya.

## **1.3 Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah diuraikan sebelumnya, tujuan dari penelitian ini adalah untuk mengetahui karakteristik tingkat kesejahteraan di Kota Surabaya dan mendapatkan model terbaik dalam memodelkan tingkat kesejahteraan di Kota Surabaya dengan metode pohon klasifikasi.

## **1.4 Manfaat Penelitian**

Hasil penelitian ini diharapkan dapat dijadikan bahan pertimbangan oleh Pemerintah Kota Surabaya dalam pembuatan kebijakan berkaitan dengan pengentasan kemiskinan. Selain itu penelitian ini diharapkan dapat memberikan tambahan wawasan kepada peneliti terkait penerapan ilmu statistika tentang metode pohon klasifikasi untuk menyelesaikan permasalahan riil.

## **BAB II**

### **TINJAUAN PUSTAKA**

Bab ini membahas mengenai landasan teori yang digunakan untuk membantu dalam menyelesaikan permasalahan pada penelitian ini yaitu mendapatkan karakteristik setiap level status kesejahteraan rumah tangga di Kota Surabaya. Beberapa landasan teori yang dibahas pada bab ini adalah statistika deskriptif, pohon klasifikasi, ketepatan pohon klasifikasi, dan konsep serta definisi tingkat kesejahteraan.

#### **2.1 Statistika Deskriptif**

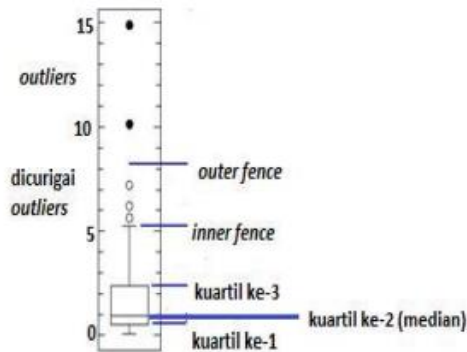
Statistika deskriptif adalah salah satu metode statistika yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna. Statistika deskriptif hanya memberikan informasi mengenai data yang dipunyai dan sama sekali tidak menarik inferensia atau kesimpulan apapun. Dengan statistika deskriptif, kumpulan data yang diperoleh akan tersaji dengan ringkas dan rapi serta dapat memberikan informasi inti dari kumpulan data yang ada. Informasi yang dapat diperoleh dari statistika deskriptif ini antara lain ukuran pemusatan data dan ukuran penyebaran data (Walpole dkk, 2007).

Pada variabel berskala numerik digunakan ukuran pemusatan data rata-rata (mean) dan median. Sedangkan untuk variabel berskala kategorik menggunakan ukuran pemusatan data modus atau nilai yang sering muncul pada suatu kelompok data. Data dengan frekuensi terbanyak menunjukkan modus. Selain itu juga statistika deskriptif bisa disajikan dalam bentuk tabel,



boxplot, dan cross tabulation sehingga informasi yang disampaikan akan lebih mudah dipahami oleh pembaca.

Box plot atau dikenal sebagai diagram box and-whisker merupakan suatu box (kotak berbentuk bujur sangkar). Box plot adalah cara standar untuk menampilkan distribusi data berdasarkan lima rangkuman (minimum, kuartil pertama, median, kuartil ketiga, dan maksimum). Box plot dapat diilustrasikan seperti Gambar 2.1. Dalam penggambarannya, box plot dapat diilustrasikan secara horizontal maupun vertikal.



**Gambar 2.1** Visualisasi Box plot  
(Sumber : <http://www.physics.csbsju.edu>)

*Cross tabulation* adalah metode statistika yang menggambarkan dua atau lebih variabel secara bersama-sama yang hasilnya berupa tabel kontingensi. Tabel kontingensi dapat menunjukkan hubungan antara variabel kategorikal. Sebuah tabel dibuat dengan  $I$  baris untuk kategori  $x$  dan  $J$  kolom untuk kategori  $y$ , maka sel dari tabel tersebut menunjukkan  $IJ$  hasil yang

mungkin (Agresti, 2013). Tabel 2.1 menunjukkan *cross tabulation* berukuran  $I \times J$ .

**Tabel 2.1** Tabulasi Silang Variabel x dengan Variabel y

Variabel x	Variabel y						Total
	1	2	3	j	...	J	
<b>1</b>	$n_{11}$	$n_{12}$	$n_{13}$	$n_{1j}$	...	$n_{1J}$	$n_{1.}$
<b>2</b>	$n_{21}$	$n_{22}$	$n_{23}$	$n_{2j}$	...	$n_{2J}$	$n_{2.}$
i	$n_{i1}$	$n_{i2}$	$n_{i3}$	$n_{ij}$	...	$n_{iJ}$	$n_{i.}$
:	:	:	:			:	:
<b>I</b>	$n_{I1}$	$n_{I2}$	$n_{I3}$	$n_{Ij}$	...	$n_{IJ}$	$n_{I.}$
<b>Total</b>	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$		$n_{.J}$	$n_{..}$

## 2.2 Pohon Klasifikasi

Klasifikasi adalah suatu proses untuk menemukan suatu model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan dapat menggunakan model untuk memprediksi suatu kelas objek yang label kelasnya tidak diketahui (Han, Kamber, dan Pei, 2007). Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi atau akan menghasilkan peluang kesalahan alokasi yang hasilnya kecil (Johnson dan Winchern, 2007). Pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi.

Menurut Breiman dkk. (1984), metode pohon regresi dan pohon klasifikasi adalah metode yang digunakan untuk menggambarkan hubungan antara peubah respon dengan satu set peubah prediktor. Pohon klasifikasi bertujuan untuk menghasilkan pengklasifikasian yang akurat dan menjelaskan prediksi data baru dalam tiap kategori yang terdapat dalam respon. Pembentukan pohon klasifikasi memerlukan 4 komponen, yaitu:

- a. Menentukan pemilahan bagi ruang peubah prediktor. Amatan dengan jawaban “ya” masuk ke ruang A, sedangkan jawaban “tidak” masuk ke ruang komplemen A. Ruang yang terbentuk dari jawaban tersebut disebut simpul.
- b. Kriteria *Goodness of Split* merupakan alat evaluasi bagi pemilahan yang dilakukan pemilah  $s$  pada simpul  $t$  berupa persamaan

$$\Phi(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (2.1)$$

dimana:

$i(t)$  : fungsi keheterogenan pada simpul  $t$

$P_L$  : proporsi data yang masuk ke simpul kiri

$P_R$  : proporsi data yang masuk ke simpul kanan

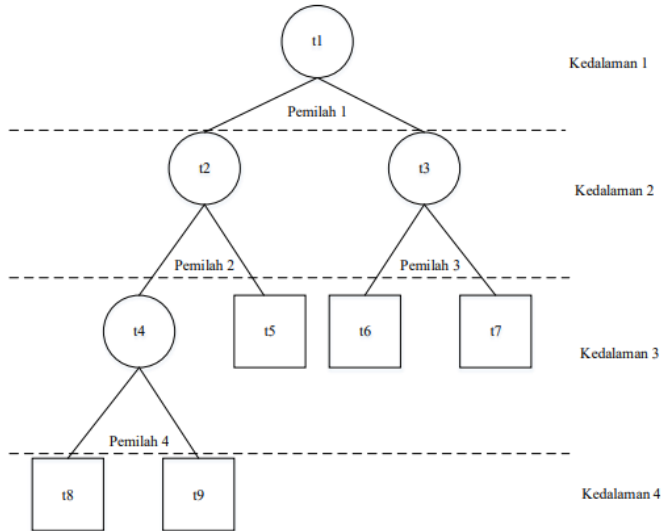
- c. Ukuran yang digunakan untuk menentukan ukuran pohon yang layak yaitu dengan menggunakan sampel uji (test sample) atau sampel validasi silang lipat- $v$ .
- d. Aturan penandaan label kelas pada setiap simpul terminal

Metode pohon klasifikasi merupakan metode nonparametrik sehingga tidak diperlukan pemenuhan asumsi kenormalan data. Struktur data dapat dilihat secara visual sehingga memudahkan eksplorasi data dan pengambilan keputusan berdasarkan model yang diperoleh. Namun perhitungan metode pohon klasifikasi

sulit dilakukan secara manual karena banyaknya kombinasi untuk mencari peubah yang paling dominan sehingga digunakan software SPSS, tidak ada penetapan validasi silang lipat-v (*v-fold cross validation*) dan jenis fungsi keheterogenan simpul. Pohon klasifikasi ini dibentuk melalui penyekatan berulang yang akan membentuk partisi-partisi kelas pengamatan yang lebih homogen dengan karakteristik tertentu (*binary recursive partitioning*). Proses disebut *binary* karena setiap simpul asal akan selalu mengalami pemecahan kedalam tepat dua simpul dalam. Sedangkan *recursive* berarti bahwa proses pemecahan tersebut diulang kembali pada setiap simpul dalam sebagai hasil pemecahan sebelumnya, sedangkan simpul-simpul dalam tersebut sekarang menjadi simpul asal. Proses pemecahan ini akan terus dilakukan sampai tidak dapat dilakukan pemecahan berikutnya. Istilah *partitioning* mengartikan bahwa *learning* sample yang dimiliki dipecah kedalam bagian-bagian atau partisi-partisi yang lebih kecil (Lewis, 2000).

Klasifikasi menggunakan metode CART memiliki tiga tahapan. Pertama, metode ini membentuk pohon klasifikasi dengan prosedur pembentukan menggunakan pemilahan simpul secara berulang (*recursive*). Tahap kedua adalah pemangkasan pohon klasifikasi (*pruning*) yang menghasilkan rangkaian pohon klasifikasi yang lebih sederhana. Tahap terakhir, penentuan pohon klasifikasi optimal, dimana pohon klasifikasi tersebut dapat mempresentasikan informasi dari data namun tidak berlebihan (*overfitting*). Akan tetapi, pada penelitian ini terlebih dahulu dilakukan perubahan beberapa nilai dalam variabel menjadi bentuk data kategorik/symbol untuk mempermudah proses pengolahan data. Terakhir membagi data menjadi data training dan testing. Tahapan dalam pembangunan pohon klasifikasi menggunakan metode CART yaitu (1) pembentukan pohon klasifikasi, (2) proses pemangkasan pohon klasifikasi, (3) pemilihan pohon optimal.

Berikut ini merupakan struktur pohon pada metode CART



**Gambar 2.2** Struktur pohon klasifikasi

(Sumber : Breiman dkk, 1993)

Ilustrasi struktur pohon klasifikasi yang ditunjukkan pada Gambar 2.1. Simpul awal yang mengandung seluruh data dengan notasi  $t_1$ . Pada Gambar 2.1 simpul dalam (internal node) dinotasikan dengan  $t_2$ ,  $t_3$ , dan  $t_4$ , sedangkan simpul akhir (terminal node) dinotasikan dengan  $t_5$ ,  $t_6$ ,  $t_7$ ,  $t_8$ , dan  $t_9$  dimana setelahnya tidak ada lagi pemilahan, artinya simpul anak yang dihasilkan telah homogen. Setiap simpul berada pada kedalaman (depth) tertentu, dimulai dari simpul awal  $t_1$  yang berada pada kedalaman 1,  $t_2$  dan  $t_3$  berada pada kedalaman 2, dan begitu seterusnya hingga dapat simpul  $t_4$ ,  $t_5$ ,  $t_6$ ,  $t_8$ , dan  $t_9$  yang berada pada kedalaman 4. Metode CART akan menghasilkan output berbeda tergantung dari skala data variabel respon. Jika variabel respon bertipe kategorik maka CART akan menghasilkan pohon klasifikasi (*classification trees*). Sedangkan jika variabel respon

bertipe kontinu atau numerik maka CART akan menghasilkan pohon regresi (*regression trees*). Metode CART efektif bila diterapkan pada data dengan pengamatan yang relatif banyak (Du, Steyn, dan Stumph, 1986). Tujuan utama CART adalah mendapatkan suatu kelompok data yang akurat sebagai bentuk karakteristik dari suatu pengklasifikasian.

### 2.2.1 Pembentukan Pohon Klasifikasi

Pada tahap ini akan dilakukan proses pembentukan CART yang diawali dengan menentukan variabel yang akan dijadikan pemilah tiap simpul dalam pohon klasifikasi. Sehingga setiap node dari algoritma yang digunakan untuk membangun pohon harus dapat menentukan variabel yang menjadi split (pemilah) terbaik. Karena hal tersebut menjadi pertimbangan hasil perhitungan semua variabel dalam menentukan variabel pemilah terbaik.

Suatu pemilah berasal dari satu kemungkinan pemilah variabel prediktor yang mungkin. Pemilihan pemilah bisa ditentukan menggunakan indeks Gini atau indeks *Twoing* yang mengukur keheterogenan simpul. Lalu masing-masing variabel dan semua kemungkinan threshold atau nilai variabel yang menjadi pemilah dihitung nilai *goodness of split* dan yang menghasilkan nilai maksimum akan dipilih sebagai pemilah terbaik. Fungsi keheterogenan indeks gini ditulis pada persamaan 2.2

$$i(t) = 1 - \sum_{j=1} p^2(j|t) \quad (2.2)$$

dengan  $i(t)$  merupakan nilai fungsi keheterogenan simpul  $t$ , dan  $p(j|t)$  merupakan proporsi kelas  $j$  pada simpul  $t$ .

Fungsi keheterogenan indeks *Twoing* dituliskan pada persamaan 2.3

$$i(t) = \frac{P_L P_R}{4} [\sum_j |p(j|t_L) - p(j|t_R)|]^2 \quad (2.3)$$

dengan  $P_L$  merupakan jumlah catatan pada calon kiri  $t_L$  dan  $P_R$  merupakan jumlah catatan pada calon kanan  $t_R$ . Sedangkan  $(j|t_L)$  jumlah catatan berkategori  $j$  pada calon cabang kiri  $t_L$  pada simpul  $t$  dan  $(j|t_R)$  merupakan jumlah catatan berkategori  $j$  pada calon cabang kanan  $t_R$  pada simpul  $t$ .

Lalu dilanjutkan dengan penentuan simpul dan pemberian label kelas. Prosedur binary recursive partitioning akan berhenti atau dengan kata lain terbentuk simpul terminal yaitu apabila kondisi simpul tersebut memenuhi salah satu kriteria: (1) hanya ada satu pengamatan ( $n=1$ ) dalam tiap simpul anak atau adanya batasan minimum  $n$  pengamatan yang diinginkan peneliti, (2) semua pengamatan dalam setiap simpul anak mempunyai distribusi yang identik terhadap variabel prediktor sehingga tidak mungkin untuk dipilih lagi, (3) adanya batasan jumlah level atau tingkat kedalaman pohon maksimal yang ditetapkan peneliti. Setiap simpul terminal perlu diberi label kelas, sehingga nantinya dapat diketahui karakteristik dari klasifikasi pengamatan untuk setiap kelas variabel respon yang terbentuk. Pemberian label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah anggota kelas terbanyak pada simpul terminal tersebut

### 2.2.2 Proses Pemangkasan Pohon Klasifikasi

Pemangkasan pohon klasifikasi atau bisa disebut *pruning* perlu dilakukan karena semakin banyak pemilahan yang dilakukan mengakibatkan makin kecilnya tingkat kesalahan prediksi atau dengan kata lain nilai prediksi melebihi nilai yang sebenarnya (*overfitting*). Pemangkasan pohon dilakukan dengan

menentukan *cost complexity* minimum (Breiman dkk, 1984). Ukuran *cost complexity* ditulis pada persamaan 2.4.

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}| \quad (2.4)$$

keterangan:

$R_\alpha(T)$  : Ukuran kompleksitas suatu pohon  $T$  pada kompleksitas  $\alpha$ .

$R(T)$  : Penduga pengganti (*resubstitution estimate*) pohon atau ukuran kesalahan klasifikasi pohon  $T$ .

$\alpha$  : Parameter *cost complexity* bagi penambah satu simpul terminal pada pohon  $T$ .

$|\tilde{T}|$  : Banyaknya simpul terminal pada pohon  $T$ .

*Cost complexity pruning* digunakan untuk menentukan pohon bagian  $T(\alpha)$  yang dapat meminimumkan  $R_\alpha(T)$  pada pohon bagian atau setiap nilai  $\alpha$ . Nilai parameter kompleksitas ( $\alpha$ ) akan secara perlahan meningkat selama proses pemangkasan. Selanjutnya, pencarian pohon bagian  $T(\alpha) < T_{maks}$  yang dapat meminimumkan  $R_\alpha(T)$ . Jika  $R(T)$  digunakan sebagai kriteria penentuan pohon optimal maka akan cenderung pohon terbesar adalah  $T_1$ , sebab semakin besar pohon, maka semakin kecil nilai  $R(T)$  nya.

### 2.2.3 Pemilihan Pohon Optimal

Pada tahap ini, perlu dilakukan pohon klasifikasi optimal yang berukuran sederhana tetapi memberikan nilai penduga pengganti cukup kecil. Ukuran pohon klasifikasi yang besar akan menyebabkan nilai kompleksitas kesalahan (*error complexity*) yang tinggi, sehingga diperlukan pohon klasifikasi optimal yang



memiliki kesalahan prediksi kecil. Hal ini dikarenakan ukuran pohon yang besar akan menyebabkan struktur data yang digambarkan cenderung kompleks, sehingga perlunya memilih pohon optimal yang berukuran sederhana tetapi memberikan nilai kesalahan pengklasifikasian cukup kecil. Pada tiap simpul terminal yang terbentuk, akan ditunjukkan persentase dari jumlah anggota yang ada di setiap simpul terminalnya.

### 2.3 Ketepatan Klasifikasi

Ukuran ketepatan klasifikasi bisa dihitung sesuai persamaan 2.5 dimana *Apparent Error Rate* (APER) menunjukkan ukuran kesalahan klasifikasi sejumlah data pengamatan oleh suatu fungsi klasifikasi (Johnson dan Wichern, 2007). Perhitungan tersebut bias didapatkan dari table prediksi klasifikasi dengan 4 level kelas variabel respon.

$$\text{Akurasi Klasifikasi} = 1 - \text{APER} = \frac{[n_{11} + n_{22} + n_{33} + n_{44}]}{N} \quad (2.5)$$

**Tabel 2.2** Contoh *Tabel Prediksi Klasifikasi dengan 4 Level Kelas Variabel Respon*

Kondisi Aktual	Prediksi Klasifikasi				Total
	Kelas 1	Kelas 2	Kelas 3	Kelas 4	
<b>Kelas 1</b>	$n_{11}$	$n_{12}$	$n_{13}$	$n_{14}$	$n_{1.}$
<b>Kelas 2</b>	$n_{21}$	$n_{22}$	$n_{23}$	$n_{24}$	$n_{2.}$
<b>Kelas 3</b>	$n_{31}$	$n_{32}$	$n_{33}$	$n_{34}$	$n_{3.}$
<b>Kelas 4</b>	$n_{41}$	$n_{42}$	$n_{43}$	$n_{44}$	$n_{4.}$
<b>Total</b>	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	$N$

## **2.4 Konsep dan Definisi Tingkat Kesejahteraan**

Kesejahteraan sosial adalah kondisi sosial ekonomi yang memungkinkan bagi setiap warga negara untuk dapat memenuhi kebutuhan yang bersifat jasmani, rohani dan sosial sesuai dengan harkat dan martabat manusia. Tingkat kesejahteraan masyarakat antara lain dapat diukur melalui besarnya pendapatan atau pengeluaran. Pengeluaran untuk kebutuhan konsumsi dapat mencerminkan tingkat kemampuan ekonomi masyarakat, dan kemampuan daya beli masyarakat dapat memberikan gambaran tentang tingkat kesejahteraan masyarakat. Semakin tinggi daya beli masyarakat menunjukkan meningkatnya kemampuan dalam memenuhi kebutuhan hidupnya dan selanjutnya akan berdampak meningkatnya kesejahteraan masyarakat (BPS, 2014). Dalam mengukur kesejahteraan rumah tangga diperlukan indikator moneter, indikator yang banyak digunakan adalah pendapatan dan pengeluaran (BPS, 2014). Indikator pengeluaran, dalam hal ini disebut juga konsumsi, dipilih karena sifatnya tetap dan relatif stabil terhadap berfluktuasinya pendapatan dari tahun ke tahun.

Tingkat kesejahteraan sangat erat kaitannya dengan kemiskinan. Kemiskinan merupakan salah satu masalah pembangunan kesejahteraan sosial dimana merupakan bentuk ketidakmampuan untuk meraih kesejahteraan dipandang dari sisi ekonomi dalam memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan (GK) (BPS, 2014). Sementara itu Badan Perencanaan dan Pembangunan Nasional (BAPPENAS) (2004) mendefinisikan kemiskinan sebagai kondisi di mana seseorang atau sekelompok orang, laki-laki dan perempuan, tidak mampu memenuhi hak dasarnya untuk

mempertahankan dan mengembangkan kehidupan yang bermartabat.

Kemiskinan absolut merupakan kondisi seseorang memiliki pendapatan di bawah garis kemiskinan atau tak cukup memenuhi kebutuhan pangan, sandang, papan, kesehatan, perumahan, dan pendidikan yang dibutuhkan untuk bisa hidup dan bekerja. Kebutuhan mendasar minimum diartikan sebagai ukuran keuangan dalam bentuk uang dan nilainya dinamakan dengan istilah garis kemiskinan. Penduduk yang memiliki rata-rata pendapatan/ pengeluaran per kapita per bulan dibawah garis kemiskinan digolongkan sebagai penduduk miskin.

Terdapat empat belas indikator kemiskinan untuk memenuhi berbagai program pelayanan dasar data rumah tangga yang meliputi luas lantai rumah, jenis lantai rumah, jenis dinding rumah, fasilitas tempat buang air besar, sumber air minum, penerangan yang digunakan, bahan bakar yang digunakan, frekuensi makan dalam sehari, kebiasaan membeli daging/ayam/susu, kemampuan membeli satu set pakaian, kemampuan berobat ke puskesmas/ poliklinik, lapangan pekerjaan kepala rumah tangga, pendidikan kepala rumah tangga, dan kepemilikan aset.

## **BAB III**

### **METODE PENELITIAN**

Bab ini membahas mengenai kerangka konsep penelitian yang menghubungkan antara konsep kesejahteraan menurut BPS dan penelitian ini, sumber data yang digunakan, variabel penelitian, dan langkah analisis yang dilakukan dalam penelitian.

#### **3.1 Sumber Data**

Sumber data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari TNP2K melalui Badan Perencanaan Pembangunan Kota (BAPEKKO) Surabaya (data penelitian Sutikno, 2018) mengenai indikator tingkat kesejahteraan penduduk Kota Surabaya. Unit penelitian yang digunakan adalah rumah tangga di Kota Surabaya. Lokasi penelitian yaitu kecamatan di Kota Surabaya.



**Gambar 3.1** Peta Surabaya  
(Sumber : <http://insurabaya.blogspot.co.id>)

### 3.2 Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini adalah karakteristik kesejahteraan rumah tangga. Variabel respon yang digunakan yaitu Tingkat kesejahteraan dengan 4 level respon yaitu desil 1, desil 2, desil 3, dan desil 4. Data tingkat kesejahteraan ini merupakan data konsumsi rumah tangga yang dibagi menjadi sepuluh bagian data. Namun pada penelitian ini hanya digunakan 40% data konsumsi perkapita terendah penduduk Kota Surabaya sehingga variabel respon yang dipakai hanya desil 1 hingga 4. Desil 1 menunjukkan 1-10% data tingkat konsumsi perkapita masyarakat terendah di Kota Surabaya. Sedangkan desil 2 menunjukkan 11-20% data tingkat konsumsi perkapita masyarakat terendah di Kota Surabaya, dan seterusnya. Variabel prediktor yang digunakan dalam penelitian yaitu sebagai berikut

**Tabel 3.1** Variabel Penelitian

<b>Variabel</b>	<b>Deskripsi</b>	<b>Keterangan</b>	<b>Skala</b>
X <sub>1</sub>	Kepemilikan Mobil	0. Tidak 1. Ya	Nominal
X <sub>2</sub>	Kepemilikan Kulkas	0. Tidak 1. Ya	Nominal
X <sub>3</sub>	Kepemilikan Sepeda Motor	0. Tidak 1. Ya	Nominal
X <sub>4</sub>	Kepemilikan Komputer	0. Tidak 1. Ya	Nominal
X <sub>5</sub>	Kepemilikan Telepon Rumah	0. Tidak 1. Ya	Nominal
X <sub>6</sub>	Bahan Bakar / Energi Untuk Memasak: Kayu Bakar, Tidak memasak di	0. Tidak 1. Ya	Nominal

	rumah			
X <sub>7</sub>	Sumber Air Minum: Air Kemasan Bermerk	0. Tidak 1. Ya		Nominal
X <sub>8</sub>	Sumber Air Minum: Air Isi Ulang, Ledeng Meteran	0. Tidak 1. Ya		Nominal
X <sub>9</sub>	Sumber Air Minum: Ledeng Eceran, Sumur Bor/Pompa	0. Tidak 1. Ya		Nominal
X <sub>10</sub>	Jumlah Anggota Rumah Tangga (ART)			
X <sub>11</sub>	Status Penguasaan Bangunan Tempat Tinggal yang Ditempati: Kontrak/Sewa	0. Tidak 1. Ya		Nominal
X <sub>12</sub>	Status Penguasaan Bangunan Tempat Tinggal yang Ditempati: Dinas	0. Tidak 1. Ya		Nominal
X <sub>13</sub>	Luas Lantai per Kapita	-		Rasio
X <sub>14</sub>	Jumlah ART yang Menamatkan D1/D2/D3	-		Rasio

**Tabel 3.1** Variabel Penelitian (Lanjutan)

Variabel	Deskripsi	Keterangan	Skala
X <sub>15</sub>	Jumlah ART yang Menamatkan S1/S2/S3	-	Rasio

X <sub>16</sub>	Jumlah ART yang Menamatkan SLTA	-	Rasio
X <sub>17</sub>	Jumlah ART yang Menamatkan SLTP	-	Rasio
X <sub>18</sub>	Jumlah ART Laki-laki	-	Rasio
X <sub>19</sub>	Jumlah ART yang Bersekolah di SLTA	-	Rasio
X <sub>20</sub>	Jumlah ART yang Bekerja di Pertanian dengan Status Sebagai Berusaha Sendiri	-	Rasio
X <sub>21</sub>	Jumlah ART yang Bekerja di Pertanian dengan Status Sebagai Pekerja Tak Dibayar	-	Rasio
X <sub>22</sub>	Jumlah ART yang Bekerja di Industri dengan Status Sebagai Berusaha Sendiri	-	Rasio
X <sub>23</sub>	Jumlah ART yang Bekerja di Industri dengan Status Sebagai Berusaha Sendiri dengan Buruh Dibayar	-	Rasio
X <sub>24</sub>	Jumlah ART yang Bekerja di Industri dengan Status Pegawai	-	Rasio
X <sub>25</sub>	Jumlah ART yang Bekerja di Industri dengan Status	-	Rasio

	Sebagai Pekerja Bebas		
X <sub>26</sub>	Jumlah ART yang Bekerja di Perusahaan Listrik dan Gas dengan Status Sebagai Pegawai	-	Rasio
X <sub>27</sub>	Jumlah ART yang Bekerja di Perusahaan Listrik dan Gas dengan Status Sebagai Pekerja Tak Dibayar	-	Rasio

**Tabel 3.1** Variabel Penelitian (Lanjutan)

Variabel	Deskripsi	Keterangan	Skala
X <sub>28</sub>	Jumlah ART yang Bekerja di Konstruksi dengan Status Sebagai Pekerja Bebas	-	Rasio
X <sub>29</sub>	Jenis Lantai Terluas: Marmer/Granit, Keramik, Parket/Vinil/Permadani	0. Tidak 1. Ya	Nominal
X <sub>30</sub>	Jenis Lantai Terluas: Ubin/Tegel/Braso	0. Tidak 1. Ya	Nominal
X <sub>31</sub>	Jenis Lantai Terluas: Kayu/Papan Kualitas Tinggi, Semen/Bata Merah, Bambu,	0. Tidak 1. Ya	Nominal



	Kayu/Papan Kualitas Rendah			
$X_{32}$	Jenis Dinding Terluas: Berkategori Lain	0. Tidak 1. Ya		Nominal
$X_{33}$	Tipe toilet: Pribadi dengan Jenis Kloset Leher Angsa	0. Tidak 1. Ya		Nominal
$X_{34}$	Tipe toilet: Pribadi dengan Jenis Kloset Plengsengan	0. Tidak 1. Ya		Nominal
$X_{35}$	Tipe toilet: Bersama dengan Jenis Kloset Leher Angsa	0. Tidak 1. Ya		

### 3.3 Struktur Data

Struktur data pada penelitian ini ditunjukkan pada Tabel 3.2 berikut.

**Tabel 3.2** Struktur Data Penelitian

Pengamatan ke-n	Respon (Y)	Variabel X				
		$X_1$	$X_2$	$X_3$	...	$X_k$
1	$Y_1$	$X_{11}$	$X_{12}$	$X_{13}$	...	$X_{1k}$
2	$Y_2$	$X_{21}$	$X_{22}$	$X_{23}$	...	$X_{2k}$
:	:	:	:	:		:
n	$Y_n$	$X_{n1}$	$X_{n2}$	$X_{n3}$	...	$X_{nk}$

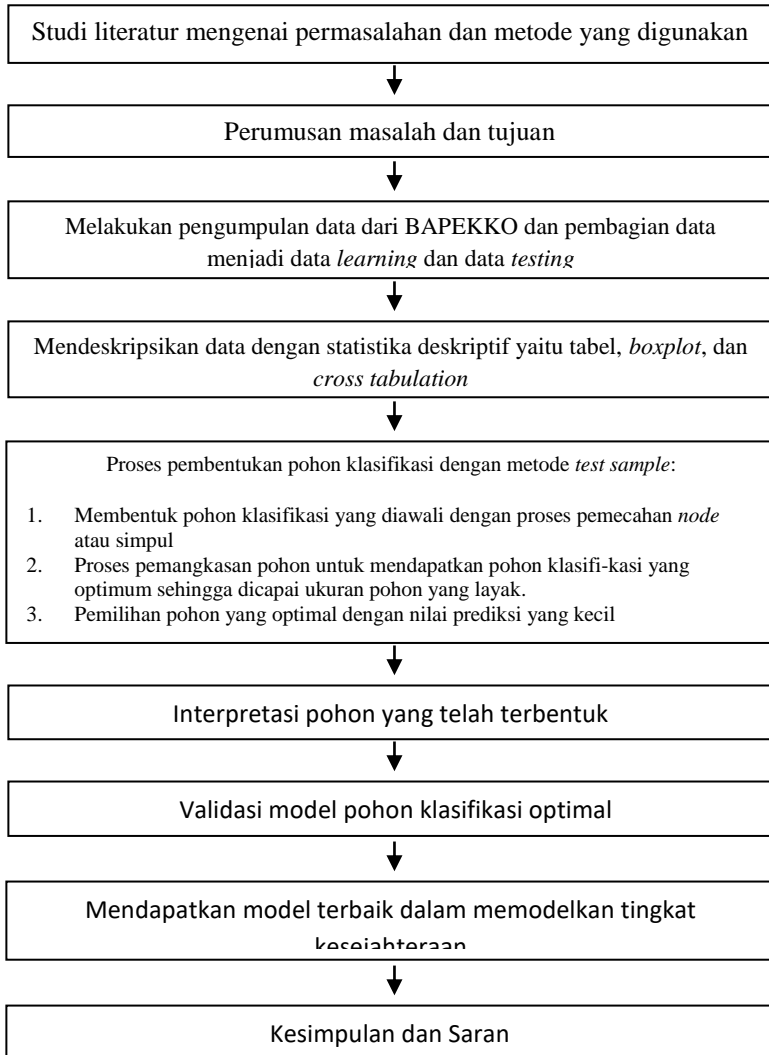
### 3.4 Langkah Penelitian

Berikut ini adalah langkah-langkah penelitian yang dilakukan adalah sebagai berikut.

1. Melakukan studi literatur mengenai permasalahan dan metode yang digunakan.
2. Merumuskan masalah dan tujuan.
3. Melakukan pengumpulan data dan *pre-processing* data
4. Melakukan pengolahan dan penyajian data dalam bentuk statistika deskriptif sehingga dapat memahami karakteristik dari variabel.
5. Melakukan pembagian data menjadi data *learning* dan data *testing*
6. Melakukan pembentukan pohon klasifikasi menggunakan metode *test sample*. Tahapan pembuatan CART adalah sebagai berikut :
  - a. Membentuk pohon klasifikasi yang diawali dengan proses pemecahan *node* atau simpul yang diawali dengan mencari semua kemungkinan pemilahan pada tiap variabel prediktor sehingga dapat memilih pemilahan terbaik yang memak-simumkan ukuran kehomogenan dalam masing-masing simpul (persamaan 2.2 dan 2.3)
  - b. Proses pemangkasan pohon untuk mendapatkan pohon klasifikasi yang optimum sehingga dicapai ukuran pohon yang layak. Besarnya *resubstitution estimate* pohon T pada parameter kompleksitas  $\alpha$  (persamaan 2.4)
  - c. Pemilihan pohon yang optimal dengan memiliki nilai prediksi yang kecil serta menginterpretasi pohon regresi yang telah terbentuk.
7. Melakukan validasi model pohon klasifikasi optimal dan menghitung keakuratan pohon klasifikasi
8. Mendapatkan model terbaik dalam memodelkan tingkat kesejahteraan
9. Menarik kesimpulan dan saran.

### 3.5 Diagram Alir

Berikut ini adalah langkah-langkah penelitian yang dilakukan



**Gambar 3.2** Diagram Alir Langkah Analisis

## **BAB IV**

### **ANALISIS DAN PEMBAHASAN**

Bab ini membahas mengenai analisis karakteristik setiap tingkat kesejahteraan rumah tangga di Kota Surabaya menggunakan pohon klasifikasi. Sebelum itu, dilakukan *pre-processing*, statistika deskriptif, dan pembagian data *learning* dan *testing* sebelum menganalisis lebih lanjut.

#### **4.1 *Pre-processing* Data**

Sebelum dilakukan pengolahan data, dilakukan *pre-processing* pada data. Pada data TNP2K ini terdapat dua bagian data yaitu data individu dan data rumah tangga. Data individu terdiri atas 514.418 data sedangkan data rumah tangga terdiri atas 149.189 data. *Pre-processing* yang dilakukan yaitu mengubah data individu menjadi data rumah tangga atau dengan kata lain menggabungkan data individu menjadi data yang lebih sedikit yaitu ke data rumah tangga. Langkah pertama yang dilakukan yaitu menyortir data individu yang dibutuhkan lalu dilakukan penggabungan data. Setelah menyortir data dan variabel yang dibutuhkan, lalu dilakukan pemisahan data. Sebanyak 514.418 data individu dipisah menjadi beberapa file excel agar bisa diolah lebih cepat. Setelah itu dilakukan untuk pemilahan data sesuai kriteria dan menghitung jumlah data yang masuk dalam satu rumah tangga yang sama, sehingga dapat dikatakan menghitung jumlah anggota rumah tangga yang sesuai dengan kriteria variabel namun masih dalam satu rumah tangga.

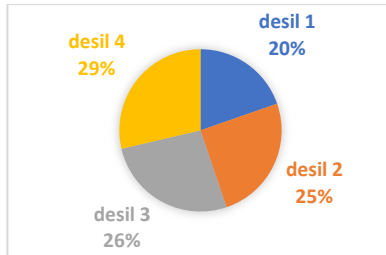
Tahap selanjutnya yang dilakukan yaitu menandai data yang tidak dipakai dengan tanda “NULL” dan akan dieliminasi sehingga data yang muncul hanyalah data yang dibutuhkan. Sehingga data individu sebanyak 514.418 individu yang sudah

diolah tersebut bisa dijadikan satu menjadi 149.189 data rumah tangga. Namun setelah dilakukan penggabungan data, ternyata terdapat 3 data yang *missing* sehingga diperlukan penghapusan data per baris. Maka dari itu data yang dipakai yaitu 149.186 data rumah tangga. Setelah dilakukan pengecekan *missing value*, dilakukan pengecekan data dan kode data pada setiap variabel. Pada saat pengecekan ditemukan satu variabel yang berisi data 0 saja sehingga tidak bisa dipakai untuk analisis lebih lanjut. Oleh karena itu variabel ini dihapuskan. Data tersebut merupakan jumlah anggota rumah tangga yang bekerja di perusahaan listrik dan gas dengan status sebagai berusaha sendiri dengan buruh tak dibayar. Dari hasil *pre-processing* tersebut maka data yang dipakai untuk analisis lebih lanjut yaitu 149.186 data dengan 35 variabel prediktor.

#### **4.2 Karakteristik Rumah Tangga di Surabaya**

Sebelum melakukan analisis lebih lanjut, eksplorasi data variabel yang diteliti menjadi hal yang penting untuk dilakukan terlebih dahulu. Tujuan eksplorasi data adalah untuk mengetahui gambaran umum karakteristik rumah tangga di Surabaya berdasarkan indikator tingkat kesejahteraan.

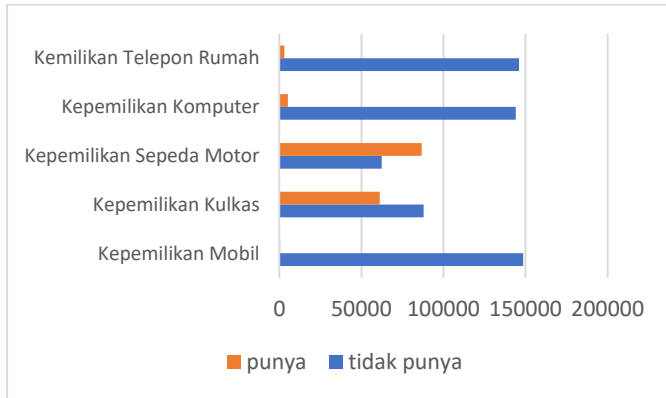
Pada penelitian ini, digunakan variabel respon yaitu status kesejahteraan yang terdiri dari desil 1, desil 2, desil 3, dan desil 4. Desil satu atau golongan satu menunjukkan bahwa tingkat konsumsi rata tersebut lebih rendah dibandingkan dengan tingkat konsumsi rata pada desil dua atau golongan dua. Proporsi anggota rumah tangga setiap desil ditunjukkan pada Gambar 4.1.



**Gambar 4.1** Persentase tiap desil pada variabel Status Kesejahteraan

Pada Gambar 4.1 diketahui jika persentase terbesar berada pada status kesejahteraan desil 4. Hal tersebut menunjukkan bahwa dari total 40% rumah tangga dengan tingkat konsumsi atau pendapatan rendah tersebut, proporsi rumah tangga desil 4 (29%) masih lebih besar dari desil lainnya sehingga diketahui masih banyak rumah tangga yang memiliki status kesejahteraan hidup yang baik mengingat desil 4 merupakan status kesejahteraan paling tinggi dari ketiga desil lainnya. Pada diagram juga diketahui bahwa rumah tangga yang mempunyai status kesejahteraan desil 1 hanya 20% dari total keseluruhan, sedangkan terdapat 80% rumah tangga yg memiliki status kesejahteraan diatas desil 1. Hal tersebut memberikan bukti bahwa dari 40% rumah tangga yang memiliki tingkat konsumsi rendah dan pendapatan rendah tersebut, sebagian besar memiliki tingkat kesejahteraan hidup yang tidak tergolong buruk.

Pada penelitian ini digunakan beberapa variabel yang menjadi indikator. Faktor pertama yang dianalisis yaitu variabel pada kategori kepemilikan aset. Pada kategori ini hanya ada 5 variabel yang diteliti yaitu kepemilikan telepon rumah, kepemilikan komputer, kepemilikan sepeda motor, kepemilikan kulkas, dan kepemilikan mobil.



**Gambar 4.2** Proporsi rumah tangga kepemilikan aset rumah tangga

Pada Gambar 4.2 diketahui jika sebagian besar rumah tangga tidak memiliki telepon rumah, komputer, dan mobil. Hal tersebut dapat dikaitkan dengan keadaan pada masa sekarang bahwa *handphone* seakan sudah menjadi kewajiban bagi setiap orang untuk memilikinya sehingga sedikit rumah tangga yang memasang telepon rumah. Selain itu, banyak orang yang mengganti komputer dengan laptop karena lebih mudah untuk dibawa kemana-mana. Sedangkan untuk mobil, mobil merupakan barang tersier dengan tingkat urgensitas untuk membeli mobil sangat kecil sehingga tidak banyak rumah tangga yang membeli mobil. Kebanyakan rumah tangga membeli sepeda motor untuk kebutuhan sehari-hari, seperti yang terlihat pada Gambar 4.2. Untuk kepemilikan kulkas terlihat bahwa masih banyak rumah tangga yang tidak memiliki kulkas. Namun jumlah rumah tangga

yang memiliki kulkas dan tidak memiliki kulkas tidaklah terpaut sangat jauh. Kulkas merupakan hal yang cukup vital untuk setiap rumah tangga apalagi dengan keadaan Kota Surabaya yang panas sehingga tidak heran jika cukup banyak rumah tangga di Kota Surabaya ini yang memiliki kulkas.

Faktor selanjutnya yaitu berasal dari kategori bahan bakar/energi untuk memasak. Pada kategori ini variabel yang diteliti hanya bahan bakar dengan kategori kayu bakar dan tidak memasak di rumah.



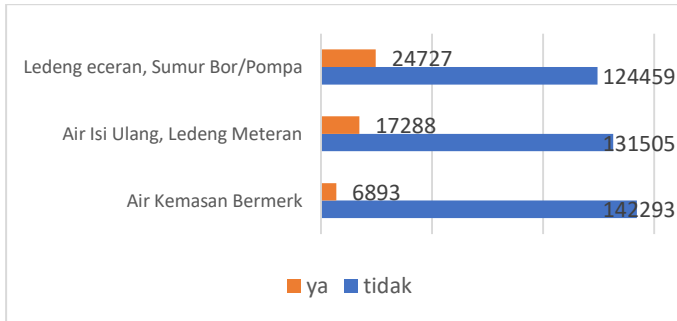
**Gambar 4.3** Persentase rumah tangga yang memakai kayu bakar dan tidak memasak di rumah

Pada Gambar 4.3 diketahui jika sebagian besar rumah tangga sudah tidak menggunakan kayu bakar untuk memasak di rumah dan sebagian besar rumah tangga memasak di rumah. Dilihat dari perkembangan jaman, saat ini banyak rumah tangga yang menggunakan bahan bakar elpiji untuk memasak kebutuhan sehari-hari karena lebih praktis.

Lalu faktor lain yang diteliti yaitu sumber air minum. Pada penelitian ini hanya 3 kategori yang digunakan. Pertama yaitu sumber air minum yang berasal dari ledeng eceran, sumur,



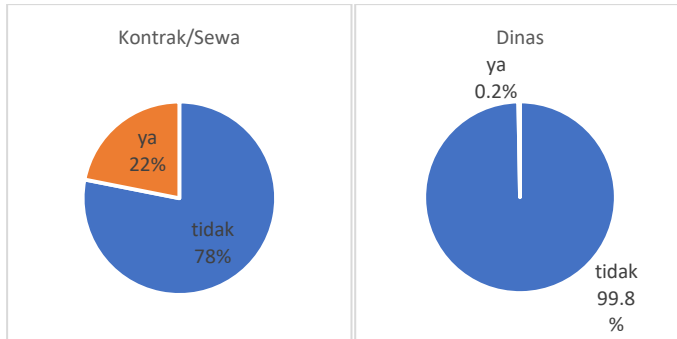
dan bor pompa, kedua yaitu sumber air minum berasal dari air isi ulang dan ledeng meteran, dan yang ketiga yaitu sumber air minum berasal dari air kemasan bermerk.



**Gambar 4.4** Proporsi rumah tangga pemakai sumber air minum ledeng eceran dan sumur bor/pompa, air isi ulang dan ledeng meteran, dan air kemasan bermerk

Berdasarkan diagram batang pada Gambar 4.4 diketahui proporsi pemakai sumber air minum terbanyak yaitu berada pada kategori ledeng eceran atau sumur bor/pompa. Lalu disusul dengan air isi ulang dan ledeng meteran, dan yang terakhir air kemasan bermerk. Dari ketiga kategori sumber air minum tersebut dapat diketahui bahwa 40% rumah tangga yang memiliki konsumsi atau pendapatan rendah ini banyak yang memakai ledeng eceran dan sumur bor/pompa.

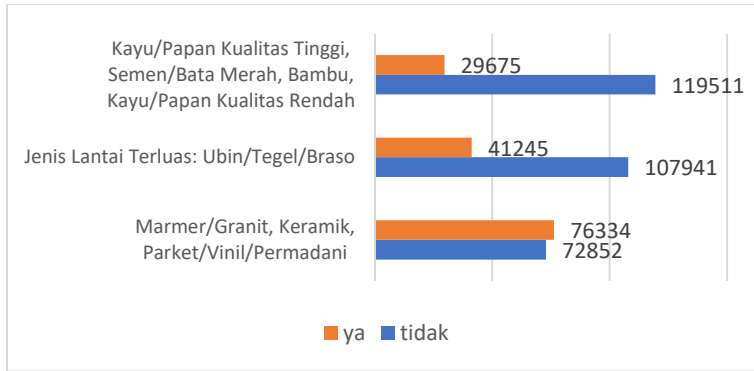
Selanjutnya, variabel lain yang digunakan yaitu variabel status penguasaan bangunan tempat tinggal yang ditempati.



**Gambar 4.5** Persentase tangga dengan status penguasaan bangunan tempat tinggal pada kategori kontrak/sewa dan dinas

Terdapat dua variabel yang diteliti pada kategori ini yaitu rumah kontrak/sewa dan rumah dinas. Berdasarkan hasil diagram pada Gambar 4.5 diketahui jika pada kategori kontrak/sewa terdapat 22% rumah tangga yang mengontrak/sewa dan 78% tidak. Sedangkan untuk kategori rumah dinas, sebanyak 0.2% rumah tangga yang menempati rumah dinas. Sebagian besar rumah tangga tidak menempati rumah dinas dan tidak mengontrak.

Selanjutnya yaitu pada kategori bangunan. Pertama, variabel yang digunakan yaitu jenis lantai terluas setiap rumah tangga di Kota Surabaya. Ada 3 kategori variabel yang dipakai dalam penelitian ini. Kategori pertama yaitu kayu/papan kualitas tinggi, semen/bata merah, bambu, kayu/papan kualitas rendah. Kategori kedua ubin/tegel/braso, dan kategori ketiga yaitu marmer/granit, keramik, parket/vinil/permadani.



**Gambar 4.6** Proporsi rumah tangga yang menggunakan jenis lantai terluas dari ketiga kategori

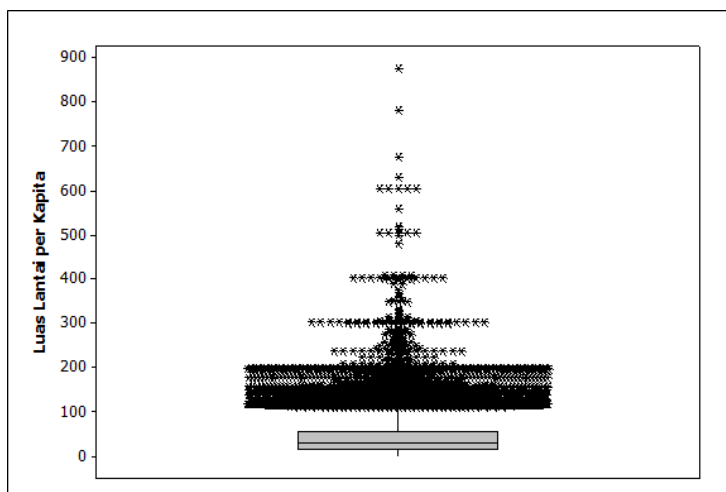
Pada Gambar 4.6 diketahui jika jumlah rumah tangga yang memiliki jenis lantai terluas marmar/granit, keramik, parket/vinil/permadani memiliki jumlah yang besar diantara ketiga kategori dengan jumlah sebanyak 76.334 rumah tangga. Setelah itu disusul oleh rumah tangga yang menggunakan ubin/tegel/braso dan kayu/papan kualitas tinggi, semen/bata merah, bambu, kayu/papan kualitas rendah. Pada informasi diatas dapat diketahui jika sebagian besar masyarakat di Kota Surabaya memiliki rumah dengan jenis lantai berbahan bagus yaitu marmar/granit, keramik, parket/vinil/permadani.

Variabel yang juga berhubungan dengan lantai yaitu luas lantai per kapita. Karakteristik dari variabel ini disajikan dalam tabel 4.1.

**Tabel 4.1** Statistika deskriptif dari variabel luas lantai per kapita

Mean	Stdev	Varian	Min	Max	Median	Modus
36,90	32,37	1048,03	4	874	30	12

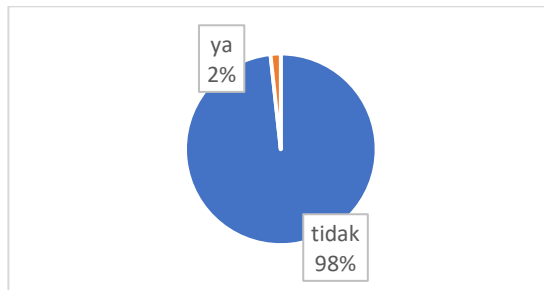
Pada variabel ini diketahui jika rata-rata luas lantai per kapita rumah tangga di Kota Surabaya yaitu  $36,9094 \text{ m}^2$ . Namun varians pada data ini sangat besar, yaitu 1048,034 yang artinya keragaman data sangat besar. Keragaman pada data dapat dilihat dari *boxplot* untuk variabel luas lantai per kapita.



**Gambar 4.7** Box plot sebaran data variabel luas lantai per kapita

Pada Gambar 4.7 diketahui bahwa masih banyak data outlier pada variabel luas lantai per kapita. Dari *boxplot* didapatkan informasi jika rumah yang digunakan oleh beberapa rumah tangga di Surabaya masih tergolong besar, terlihat dari banyaknya *outlier* yang terlihat, terutama *outlier* pada angka 100-200 m<sup>2</sup>

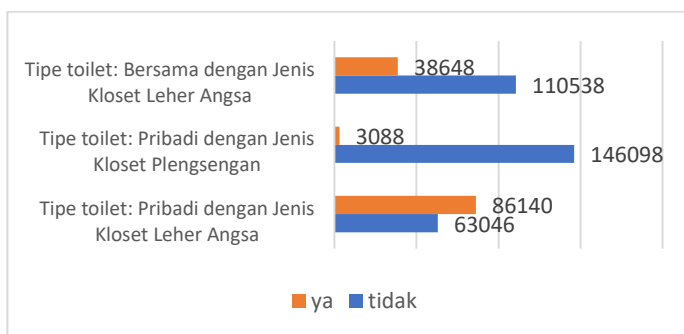
Selanjutnya yaitu variabel yang juga erat kaitannya dengan bangunan yaitu jenis dinding terluas berkategori lain. Pada variabel ini yang dimaksud dengan jenis dinding terluas berkategori lain adalah jenis dinding selain tembok, plesteran anyaman bambu, kayu, anyaman bambu, batang kayu, dan bambu.



**Gambar 4.8** Persentase rumah tangga yang menggunakan jenis dinding terluas berkategori lain

Berdasarkan hasil diagram pada Gambar 4.8 diketahui jika sebagian besar rumah tangga di Kota Surabaya menggunakan jenis dinding terluas tidak berkategori lain yang artinya jenis dinding yang digunakan yaitu antara lain tembok, plesteran anyaman bambu, kayu, anyaman bambu, batang kayu, atau bambu.

Variabel lain yang berhubungan dengan bangunan yaitu tipe toilet pada setiap rumah tangga. Pada penelitian ini terdapat 3 kategori tipe toilet yang diteliti yaitu tipe toilet bersama dengan jenis kloset leher angsa, tipe toilet pribadi dengan jenis kloset plengsengan, dan tipe toilet pribadi dengan jenis kloset leher angsa.



**Gambar 4.9** Proporsi rumah tangga yang menggunakan 3 tipe toilet sesuai variabel penelitian

Diagram batang pada Gambar 4.9 menunjukkan bahwa tipe toilet pribadi dengan jenis kloset leher angsa banyak digunakan oleh rumah tangga di Kota Surabaya. Sebanyak 86.140 rumah tangga menggunakan tipe toilet pada kategori ini. Selanjutnya jumlah terbesar disusul oleh tipe toilet bersama dengan jenis kloset leher angsa dan yang terakhir tipe toilet pribadi dengan jenis kloset plengsengan.

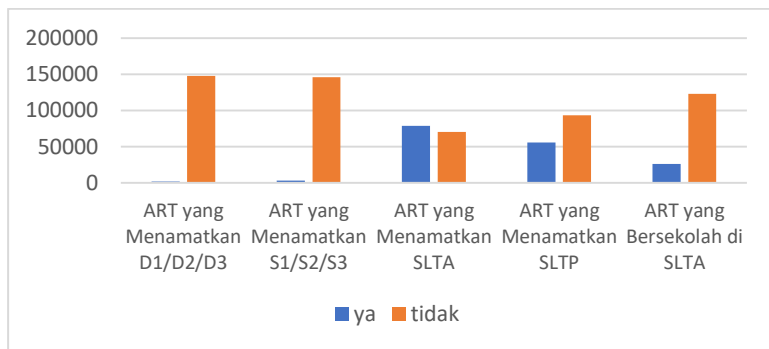
Variabel selanjutnya yaitu berhubungan dengan pendidikan, yaitu jumlah anggota rumah tangga yang masih bersekolah atau sudah menamatkan pendidikan Sekolah Dasar hingga Perguruan Tinggi. Pada penelitian ini, variabel yang digunakan yaitu jumlah anggota rumah tangga yang menamatkan D1/D2/D3, jumlah

anggota rumah tangga yang menamatkan S1/S2/S3, jumlah anggota rumah tangga yang menamatkan SLTA, jumlah anggota rumah tangga yang menamatkan SLTP, dan jumlah anggota rumah tangga yang masih bersekolah di SLTA. Berikut merupakan jumlah anggota rumah tangga dari berbagai variabel pada kategori pendidikan tersebut.

**Tabel 4.2** Jumlah anggota rumah tangga yang masih bersekolah dan sudah menamatkan pendidikan pada variabel kategori pendidikan

<b>Jumlah ART</b>	<b>Jumlah ART yang Menamatkan D1/D2/D3</b>	<b>Jumlah ART yang Menamatkan S1/S2/S3</b>	<b>Jumlah ART yang Menamatkan SLTA</b>	<b>Jumlah ART yang Menamatkan SLTP</b>	<b>Jumlah ART yang Bersekolah di SLTA</b>
0	147.509	146.038	70.437	93.400	122.977
1	1.571	2.893	45.293	40.587	22.510
2	103	242	23.599	12.522	3.441
3	3	11	7.585	2.268	244
4		2	1.828	349	14
5			358	54	
6			72	5	
7			13	1	
8			1		

Berdasarkan Tabel 4.2 diketahui jumlah anggota rumah tangga yang sudah menamatkan atau masih bersekolah di jenjang pendidikan sesuai kriteria setiap variabel. Variabel dengan jumlah anggota rumah tangga tertinggi (8 orang) yaitu variabel jumlah anggota rumah tangga yang menamatkan SLTP sedangkan jumlah anggota rumah tangga terkecil (3 orang) yaitu jumlah anggota rumah tangga yang menamatkan D1/D2/D3. Untuk proporsi anggota rumah tangga yang manamatkan sekolah, bersekolah, dan tidak menamatkan atau tidak bersekolah, dapat dilihat pada Gambar 4.10.



**Gambar 4.10** Proporsi anggota rumah tangga yang masih bersekolah dan sudah menamatkan pendidikan dari variabel kategori pendidikan

Pada Gambar 4.10 diketahui bahwa banyak anggota rumah tangga yang menamatkan pendidikan SLTA dan SLTP daripada menamatkan Perguruan Tinggi. Terlihat pada diagram batang bahwa jumlah anggota rumah tangga yang menamatkan SLTA dan SLTP jauh berbeda dibanding dengan anggota rumah tangga yang menamatkan pendidikan D1/D2/D3 ataupun S1/S2/S3. Selain itu juga diketahui jika banyak anggota rumah tangga yang



lulus SMA namun tidak melanjutkan pendidikannya ke jenjang perguruan tinggi.

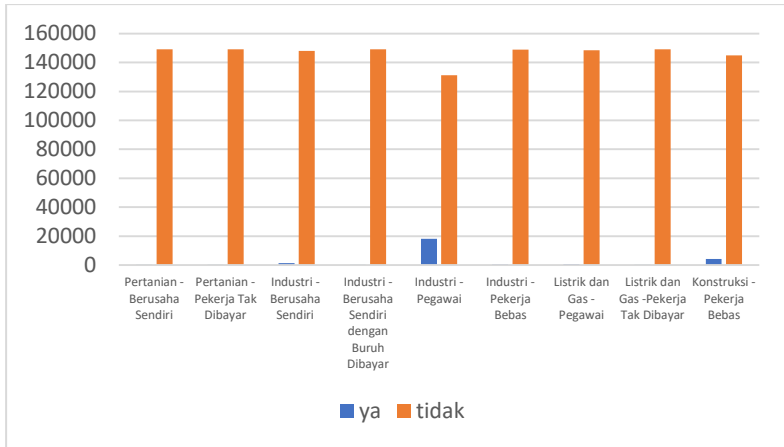
Kategori selanjutnya dari variabel penelitian ini yaitu kategori pekerjaan. Terdapat 9 variabel penelitian pada kategori ini. Variabel tersebut yaitu jumlah anggota rumah tangga yang bekerja di pertanian dengan status sebagai berusaha sendiri, jumlah anggota rumah tangga yang bekerja di pertanian dengan status sebagai pekerja tak dibayar, jumlah anggota rumah tangga yang bekerja di industri dengan status sebagai berusaha sendiri, jumlah anggota rumah tangga yang bekerja di industri dengan status sebagai berusaha sendiri dengan buruh dibayar, jumlah anggota rumah tangga yang bekerja di industri dengan status sebagai pegawai, jumlah anggota rumah tangga yang bekerja di industri dengan status sebagai pekerja bebas, jumlah anggota rumah tangga yang bekerja di perusahaan listrik dan gas dengan status sebagai pegawai, jumlah anggota rumah tangga yang bekerja di perusahaan listrik dan gas dengan status sebagai pekerja tak dibayar, dan jumlah anggota rumah tangga yang bekerja di konstruksi dengan status sebagai pekerja bebas. Berikut merupakan jumlah anggota rumah tangga yang bekerja sesuai kriteria variabel

**Tabel 4.3** Jumlah anggota rumah tangga yang bekerja dari variabel kategori pekerjaan

	Tidak ada	1 org	2 org	3 org	4 org	5 org
Pertanian-Berusaha Sendiri	149034	152				

Pertanian-Pekerja Tak Dibayar	149107	75	4			
Industri- Berusaha Sendiri	147876	1266	41	3		
Industri-Berusaha Sendiri dengan Buruh Dibayar	149144	41	1			
Industri-Pegawai	131094	15455	2332	266	36	3
Industri-Pekerja Bebas	148778	373	31	4		
Listrik dan Gas-Pegawai	148393	760	30	3		
Listrik dan Gas-Pekerja Tak Dibayar	149185	1				
Konstruksi-Pekerja Bebas	144995	3821	375	31	3	1

Berdasarkan Tabel 4.3 diketahui jika jumlah anggota rumah tangga terbanyak yang bekerja yaitu 5 orang terdapat pada variabel bekerja di industri sebagai pegawai dan bekerja di konstruksi sebagai pekerja bebas. Proporsi dari banyaknya orang yang bekerja dan tidak, dapat dilihat pada diagram batang pada Gambar 4.11



**Gambar 4.11** Proporsi anggota rumah tangga yang bekerja dan tidak bekerja pada variabel kategori pekerjaan

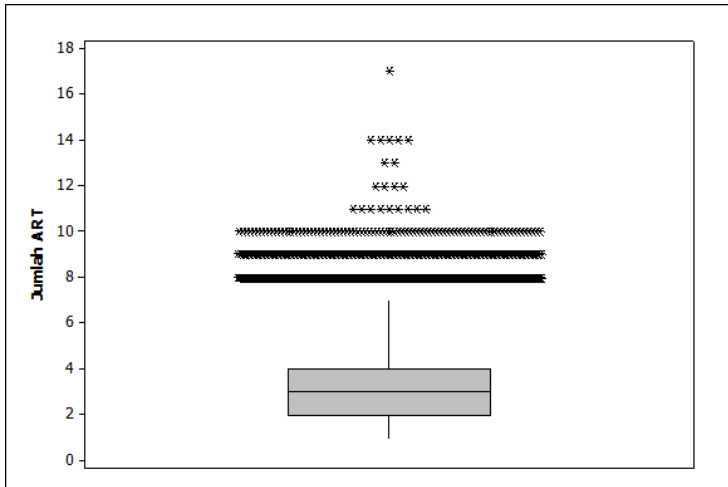
Pada Gambar 4.11 diketahui jika diagram yang paling menonjol dari kategori bekerja yaitu jumlah anggota rumah tangga yang bekerja di industri sebagai pegawai. Hal tersebut bisa dikarenakan di Kota Surabaya banyak sekali industri sehingga banyak anggota rumah tangga yang bekerja sebagai pegawai di perusahaan industri.

Variabel selanjutnya yang digunakan yaitu jumlah anggota rumah tangga dan jumlah anggota rumah tangga berjenis kelamin laki-laki. Berikut merupakan karakteristik dari kedua variabel tersebut.

**Tabel 4.4** Statistika deskriptif dari variabel jumlah anggota rumah tangga dan anggota rumah tangga laki-laki

	Mean	Stdev	Varia ns	Mi n	Ma x	Medi an	Mod us
Jumlah ART laki-laki	1,721 7	1,109 1	1,230 1	0	8	2	1
Jumlah ART	3,448 1	1,579 1	2,493 5	1	17	3	4

Diketahui bahwa rata-rata jumlah anggota rumah tangga dan jumlah anggota rumah tangga laki-laki dalam setiap rumah tangga yaitu 4 dan 2 orang. Untuk jumlah anggota rumah tangga, jumlah minimal yaitu 1 dan jumlah maksimal 17, sedangkan untuk jumlah anggota rumah tangga laki-laki yaitu minimal 0 orang dan jumlah maksimal 8 orang. Varians data jumlah anggota rumah tangga lebih besar daripada jumlah anggota rumah tangga laki-laki. Hal tersebut juga terlihat dari jarak angka minimum dan maksimum data sangat jauh. Untuk mengetahui bagaimana sebaran data pada variabel jumlah anggota rumah tangga maka disajikan boxplot pada gambar 4.12.



**Gambar 4.12** Box plot sebaran data jumlah anggota rumah tangga pada setiap rumah tangga

*Boxplot* pada Gambar 4.12 diketahui jika masih terdapat banyak *outlier* pada data jumlah anggota rumah tangga. Hal tersebut memberikan informasi bahwa masih banyak keluarga di Indonesia yang mempunyai keturunan dalam jumlah yang banyak dalam satu keluarga, sehingga program Keluarga Berencana (KB) 2 anak cukup harus lebih digencarkan.

### 4.3 Pohon Klasifikasi

Setelah dilakukan penyajian data menggunakan statistika deskriptif, dilakukan analisis klasifikasi status kesejahteraan rumah tangga di Kota Surabaya menggunakan pohon klasifikasi karena variable respon merupakan variable kategorik dengan 4 level yaitu desil 1, desil 2, desil 3, dan desil 4. Metode yang digunakan yaitu metode *Test Sample* menggunakan fungsi keheterogenan Indeks Gini dan Indeks *Twoing*. Sebelum

dilakukan analisis, dilakukan pembagian data *learning* dan *testing* dengan persentase 5%,10%,15%,20%,25% dan 30% data pada setiap indeks. Pembagian data ini bertujuan untuk mendapatkan hasil ketepatan klasifikasi terbaik dari semua kemungkinan *test sample*. Data *learning* digunakan untuk membentuk pohon klasifikasi sedangkan data *testing* digunakan untuk melihat keakuratan klasifikasi dari suatu pohon klasifikasi.

**Tabel 4.5** Perbandingan Ketepatan Klasifikasi dari Metode *Test Sample*

	Learnin g	Testing	% Ketepatan Klasifikasi		Ke- dala- man	Jumla h simpu l (Node )
			Learning	Testing		
<i>Test</i>	95%	5%	67.9	60.1	18	369
<i>Sampl</i>	90%	10%	67.5	62.2	15	495
<i>e</i>	85%	15%	66.8	63.2	15	481
denga	80%	20%	66.2	63.6	16	447
n	75%	25%	65.5	64	13	399
Indeks Gini	<b>70%</b>	<b>30%</b>	<b>65.1</b>	<b>64.1</b>	<b>13</b>	<b>381</b>
<i>Test</i>	95%	5%	69.2	60.5	17	379
<i>Sampl</i>	90%	10%	68.8	61.8	20	663
<i>e</i>	85%	15%	68.7	62.5	19	867
denga	80%	20%	69.3	63.2	21	1089
n	75%	25%	68.6	63.5	21	1053
Indeks Gini	70%	30%	68	63.5	21	1101

Pada Tabel 4.5 diketahui jika persentase ketepatan klasifikasi data testing terbesar dengan menggunakan metode *test sample* 30% dengan Indeks Gini. Metode tersebut menghasilkan nilai yang lebih besar daripada nilai-nilai ketepatan klasifikasi dari *test sample* yang lain. Nilai keakuratan klasifikasi pada metode *test sample* yang terpilih yaitu 64.1%, sehingga dalam penelitian ini dipilih metode *test sample* 30% dengan Indeks Gini untuk pembentukan pohon klasifikasi tingkat kesejahteraan rumah tangga di Kota Surabaya.

Pohon klasifikasi yang terbentuk sangat besar dan kompleks mengingat data dan variable yang digunakan cukup besar sehingga perlu dilakukan pemangkasan pohon klasifikasi (*pruning*). Pemangkasan pohon klasifikasi maksimal perlu dilakukan untuk mendapatkan pohon klasifikasi yang optimal. Pohon klasifikasi yang optimal dianggap sebagai ukuran pohon yang layak untuk klasifikasi status kesejahteraan rumah tangga di Kota Surabaya. Hasil dari pohon klasifikasi optimal yaitu pohon dengan jumlah simpul sebanyak 381 simpul dan kedalaman pohon 13. Topologi pohon klasifikasi optimal tersebut ditampilkan pada Lampiran 6

Pada Gambar 4.13 diketahui jika pohon klasifikasi optimal yang sudah di *pruning* menghasilkan 191 simpul terminal. Pemilah dengan nilai *improvement* tertinggi digunakan sebagai pemilah awal dalam pohon. Pada pohon klasifikasi, saat dipilah ke sebelah kiri dan kanan oleh sebuah pemilah akan menghasilkan nilai *improvement*. Nilai *improvement* merupakan kebaikan pemilah melalui nilai penurunan keheterogenan dalam pohon. Pemilah awal dari pohon yaitu variabel jumlah anggota rumah tangga dengan nilai *improvement* sebesar 0.058. Hal ini

menyatakan bahwa variabel jumlah anggota rumah tangga merupakan variabel yang paling dominan dalam pembentukan pohon klasifikasi.

Rumah tangga yang berjumlah 149.186, pada simpul pertama (simpul 0) dipilah menjadi kelompok kiri dan kelompok kanan oleh variabel pemilah yaitu jumlah anggota rumah tangga. Jumlah anggota rumah tangga 1, 2, 3, 13, 14, 17 sebanyak 23.118 rumah tangga mengelompok pada simpul 2 (kanan) sedangkan jumlah anggota rumah tangga 4, 5, 6, 7, 8, 9, 10, 11, 12 sebanyak 21.644 rumah tangga mengelompok pada simpul kiri (simpul 1). Lalu jumlah anggota rumah tangga 4, 5, 6, 7, 8, 9, 10, 11, 12 sebanyak 21.644 rumah tangga pada simpul 1 dipilah menjadi kelompok kiri dan kanan oleh variabel pemilah kepemilikan kulkas. Rumah tangga yang memiliki kulkas sebanyak 10.523 rumah tangga mengelompok pada simpul 3, sedangkan rumah tangga yang tidak memiliki kulkas sebanyak 11.121 rumah tangga mengelompok pada simpul 4. Simpul-simpul tersebut terus bercabang hingga menghasilkan 191 simpul terminal. Berikut merupakan karakteristik 5 simpul terminal dengan persentase terbesar pada setiap tingkat kesejahteraan.

**Tabel 4.6** 5 simpul terminal dengan persentase terbesar pada setiap tingkat kesejahteraan

Perse n tase (%)	Simpul	n	Status Keseja heraan	Karakteristik
87.8	178	43	Desil 1	<ol style="list-style-type: none"> <li>1. Jumlah ART 8,9,10,11 orang</li> <li>2. Tidak memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>3. Sumber air minum bukan air</li> </ol>



				kemasan bermerk 4. Tidak memiliki telepon rumah 5. Memiliki kulkas
84.9	43	206 3	Desil 1	1. Tidak memiliki telepon rumah 2. Tidak memiliki toilet pribadi dengan jenis kloset leher angsa 3. Jumlah ART 5,6,7,8,9,10,11,12 orang 4. Tidak memiliki kulkas
79.4	127	719	Desil 1	1. Tidak memiliki telepon rumah 2. Tidak memiliki sepeda motor 3. Jenis Lantai terluas bukan marmer/granit, keramik, atau parket/vinil/permadani 4. Tidak memiliki toilet pribadi dengan jenis kloset leher angsa 5. Jumlah ART 4 orang 6. Tidak memiliki kulkas
72	77	983	Desil 1	1. Tidak memiliki telepon rumah 2. Jenis Lantai terluas bukan marmer/granit, keramik, atau parket/vinil/permadani

**Tabel 4.6** 5 simpul terminal dengan persentase terbesar pada setiap tingkat kesejahteraan (lanjutan)

Persentase (%)	Simpul	n	Status Kesejahteraan	Karakteristik
				3. Memiliki toilet pribadi dengan jenis kloset leher angsa 4. Jumlah ART 5,6,7,8,9,10,11,12 org 5. Tidak memiliki kulkas
71.4	325	10	Desil 1	1. Luas lantai per kapita $\leq 110 \text{ m}^2$ 2. Sumber air minum ledeng

				<p>eceran atau sumur bor/pompa</p> <ol style="list-style-type: none"> <li>3. Tidak memiliki telepon rumah</li> <li>4. Jenis lantai terluas bukan marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>5. Tidak memiliki sepeda motor</li> <li>6. Jumlah ART 3,14 orang</li> <li>7. Memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>8. Tidak memiliki kulkas</li> </ol>
100	380	5	Desil 2	<ol style="list-style-type: none"> <li>1. Memiliki luas lantai <math>&gt;45 \text{ m}^2</math></li> <li>2. Tidak memiliki dinding terluas berkategori lain</li> <li>3. Sumber air minum bukan air kemasan bermerk</li> <li>4. Status penguasaan bangunan tempat tinggal yang ditempati: kontrak/sewa</li> <li>5. Memiliki toilet bersama dengan jenis kloset leher angsa</li> <li>6. Sumber air minum bukan ledeng eceran ataupun sumur bor/pompa</li> <li>7. Tidak memiliki sepeda motor</li> <li>8. Jenis lantai terluas bukan marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>9. Jumlah ART 3 orang</li> <li>10. Tidak memiliki toilet pribadi dengan jenis kloset leher angsa.</li> <li>11. Tidak memiliki kulkas</li> </ol>
85.3	337	383	Desil 2	<ol style="list-style-type: none"> <li>1. Tidak memiliki telepon rumah</li> <li>2. Tidak memiliki sepeda motor</li> </ol>

**Tabel 4.6** 5 simpul terminal dengan persentase terbesar pada setiap tingkat kesejahteraan (lanjutan)

Perse n tase (%)	Simpul	n	Status Keseja heraan	Karakteristik
				<ul style="list-style-type: none"> <li>3. Tidak memiliki jenis dinding terluas berkategori lain</li> <li>4. Memiliki sumber air minum ledeng eceran atau sumur bor/pompa</li> <li>5. Jenis lantai terluas bukan marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>6. Jumlah ART 2</li> <li>7. Tidak memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>8. Tidak memiliki kulkas</li> </ul>
85	302	17	Desil 2	<ul style="list-style-type: none"> <li>1. Memiliki sumber air minum ledeng eceran atau sumur bor/pompa</li> <li>2. Tidak memiliki sepeda motor</li> <li>3. Tidak memiliki komputer</li> <li>4. Tidak memiliki telepon rumah</li> <li>5. Memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>6. Memiliki jenis lantai terluas marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>7. Sumber air minum bukan air kemasan bermerk</li> <li>8. Jumlah ART 4</li> <li>9. Memiliki kulkas</li> </ul>
84.6	298	11	Desil 2	<ul style="list-style-type: none"> <li>1. Memiliki jenis dinding terluas berkategori lain</li> <li>2. Sumber air minum ledeng eceran, atau sumur bor/pompa</li> </ul>

				<ol style="list-style-type: none"> <li>3. Tidak memiliki sepeda motor</li> <li>4. Jenis lantai terluas bukan marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>5. Jumlah ART 3</li> <li>6. Tidak memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>7. Tidak memiliki kulkas</li> </ol>
--	--	--	--	---

**Tabel 4.6** 5 simpul terminal dengan persentase terbesar pada setiap tingkat kesejahteraan (lanjutan)

Perse n tase (%)	Simpul	n	Status Keseja heraan	Karakteristik
84.4	343	27	Desil 2	<ol style="list-style-type: none"> <li>1. Tidak memiliki toilet bersama dengan jenis kloset leher angsa</li> <li>2. Memiliki sumber air minum ledeng eceran atau sumur bor/pompa</li> <li>3. Tidak memiliki sepeda motor</li> <li>4. Tidak memiliki telepon rumah</li> <li>5. Memiliki jenis lantai terluas marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>6. Jumlah ART 3 orang</li> <li>7. Tidak memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>8. Tidak memiliki kulkas</li> </ol>
90.5	204	95	Desil 3	<ol style="list-style-type: none"> <li>1. Status penguasaan bangunan tempat tinggal yang ditempati: kontrak/sewa</li> <li>2. Tidak memiliki jenis dinding terluas berkategori lain</li> <li>3. Memiliki sumber air minum ledeng eceran atau sumur</li> </ol>

				bor/pompa 4. Tidak memiliki toilet bersama dengan jenis kloset leher angsa 5. Jenis lantai terluas bukan marmer/granit, keramik, ataupun parket/vinil/permadani 6. Jumlah ART 1 orang 7. Tidak memiliki kulkas
87.2	281	41	Desil 3	1. Tidak memiliki toilet bersama dengan jenis kloset leher angsa 2. Sumber air minum ledeng eceran atau sumur bor/pompa 3. Tidak memiliki sepeda motor 4. Jenis lantai terluas marmer/granit, keramik, ataupun parket/vinil/permadani 5. Jumlah ART 2 6. Tidak memiliki toilet pribadi dengan jenis kloset leher angsa

**Tabel 4.6 5** simpul terminal dengan persentase terbesar pada setiap tingkat kesejahteraan (lanjutan)

Perse n tase (%)	Simpul	n	Status Kesej ahtera an	Karakteristik
				7. Tidak memiliki kulkas
85.7	260	1 2	Desil 3	1.Sumber air minum ledeng eceran atau sumur bor/pompa 2.Memiliki sepeda motor 3.Jenis dinding terluas bukan berkategori lain 4.Sumber air minum bukan air

				<p>kemasan bermerk</p> <p>5. Memiliki toilet bersama dengan jenis kloset leher angsa</p> <p>6. Jenis lantai terluas bukan marmmer/granit, keramik, ataupun parket/vinil/permadani</p> <p>7. Jumlah ART 1</p> <p>8. Tidak memiliki kulkas</p>
85.7	353	1 8	Desil 3	<p>1. Luas lantai per kapita <math>\leq 7.5</math> m<sup>2</sup></p> <p>2. Jenis lantai terluas: kayu/papan kualitas tinggi, semen/bata merah, bamboo, kayu/papan kualitas rendah</p> <p>3. Jenis lantai terluas bukan ubin/tegel/braso</p> <p>4. Status penguasaan bangunan tempat tinggal yang ditempati bukan kontrak/sewa</p> <p>5. Jenis dinding terluas bukan berkategori lain</p> <p>6. Sumber air minum: ledeng eceran atau sumur bor/pompa</p> <p>7. Tidak memiliki toilet bersama dengan jenis kloset leher angsa</p> <p>8. Jenis lantai terluas bukan marmmer/granit, keramik, ataupun parket/vinil/permadani</p> <p>9. Jumlah ART 1 orang</p> <p>10. Tidak memiliki kulkas</p>

**Tabel 4.6** 5 simpul terminal dengan persentase terbesar pada setiap tingkat kesejahteraan (lanjutan)

Perse n tase (%)	Simpul	n	Status Kesej ahtera an	Karakteristik
84.5	262	37 2	Desil 3	<ol style="list-style-type: none"> <li>1. Jenis lantai terluas: ubin/tegal/braso</li> <li>2. Tidak memiliki sepeda motor</li> <li>3. Jenis dinding terluas bukan berkategori lain</li> <li>4. Sumber air minum bukan air kemasan bermerk</li> <li>5. Memiliki toilet bersama dengan jenis kloset leher angsa</li> <li>6. Jenis lantai terluas bukan marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>7. Jumlah ART 1</li> <li>8. Tidak memiliki kulkas</li> </ol>
100	102	65	Desil 4	<ol style="list-style-type: none"> <li>1. Memiliki telepon rumah</li> <li>2. Memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>3. Jenis lantai terluas marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>4. Sumber air minum bukan air kemasan bermerk</li> <li>5. Jumlah ART 4</li> </ol>

				6. Memiliki kulkas
100	92	29	Desil 4	<ol style="list-style-type: none"> <li>1. Sumber air minum merupakan air kemasan bermerk</li> <li>2. Memiliki toilet bersama dengan jenis kloset leher angsa</li> <li>3. Tidak memiliki jenis lantai terluas: marmar/granit, keramik, parket/vinil/permadani</li> <li>4. Jumlah ART 1</li> <li>5. Tidak memiliki kulkas</li> </ol>
100	210	24	Desil 4	<ol style="list-style-type: none"> <li>1. Sumber air minum merupakan air kemasan</li> <li>2. Jenis lantai terluas bukan marmar/granit, keramik, parket/vinil/permadani</li> </ol>

**Tabel 4.6** 5 simpul terminal dengan persentase terbesar pada setiap tingkat kesejahteraan (lanjutan)

Perse n tase (%)	Simpul	n	Status Keseja hteraan	Karakteristik
				<ol style="list-style-type: none"> <li>3. Memiliki sepeda motor</li> <li>4. Jumlah ART 2</li> <li>5. Memiliki tipe toilet pribadi dengan jenis kloset leher angsa</li> <li>6. Tidak memiliki kulkas</li> </ol>
100	216	20	Desil	<ol style="list-style-type: none"> <li>1. Memiliki telepon rumah</li> </ol>



			4	<ol style="list-style-type: none"> <li>2. Jenis lantai terluas bukan marmer/granit, keramik, parket/vinil/permadani</li> <li>3. Memiliki sepeda motor</li> <li>4. Jumlah ART 3, 14</li> <li>5. Memiliki tipe toilet pribadi dengan jenis kloset leher angsa</li> <li>6. Tidak memiliki kulkas</li> </ol>
100	300	18	Desil 4	<ol style="list-style-type: none"> <li>1. Memiliki mobil</li> <li>2. Memiliki sepeda motor</li> <li>3. Tidak memiliki komputer</li> <li>4. Tidak memiliki telepon rumah</li> <li>5. Memiliki tipe toilet pribadi dengan jenis kloset leher angsa</li> <li>6. Jenis lantai terluas berbahan marmer/granit, keramik, parket/vinil/permadani</li> <li>7. Sumber air minum bukan air kemasan bermerk</li> <li>8. Jumlah ART 4 orang</li> <li>9. Memiliki kulkas</li> </ol>
100	144	18	Desil 4	<ol style="list-style-type: none"> <li>1. Memiliki jenis dinding terluas berkategori lain</li> <li>2. Memiliki sumber air minum ledeng eceran, sumur bor/pompa</li> <li>3. Tidak memiliki tipe toilet bersama dengan jenis kloset leher angsa</li> </ol>

**Tabel 4.6** 5 simpul terminal dengan persentase terbesar pada setiap tingkat kesejahteraan (lanjutan)

Perse n tase (%)	Simpul	n	Status Kesej ahtera an	Karakteristik
				<ul style="list-style-type: none"> <li>4. Jenis lantai terluas bukan marmar/granit, keramik, parket/vinil/permadani</li> <li>5. Jumlah ART 1</li> <li>6. Tidak memiliki kulkas</li> </ul>
100	222	18	Desil 4	<ul style="list-style-type: none"> <li>1. Sumber air minum merupakan air kemasan bermerk</li> <li>2. Memiliki sepeda motor</li> <li>3. Memiliki jenis lantai terluas berbahan marmar/granit, keramik, parket/vinil/permadani</li> <li>4. Jumlah ART 2</li> <li>5. Tidak memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>6. Tidak memiliki kulkas</li> </ul>

Pada Tabel 4.6 diketahui persentase terbesar setiap tingkat kesejahteraan rumah tangga di Kota Surabaya. Pada desil 1, persentase terbesar pada simpul terminal 178 dengan jumlah 43 rumah tangga dan persentase sebesar 87.8% sedangkan pada desil 2, persentase terbesar pada simpul terminal 380 dengan jumlah 5 rumah tangga dan persentase 100%. Untuk desil 3 diketahui jika persentase terbesar yaitu 90.5% yang berasal dari simpul terminal 240 dengan jumlah rumah tangga sebanyak 95 rumah tangga.

Sedangkan pada desil 4, didapatkan 15 simpul terminal dengan persentase 100%, maka dari itu dipilih persentase dengan jumlah rumah tangga terbanyak yaitu 65 rumah tangga pada simpul terminal 102 untuk menjadi suatu karakteristik desil 4. Berikut merupakan hasil karakteristik berdasarkan persentase simpul terminal terbesar.

**Tabel 4.7** Karakteristik setiap tingkat kesejahteraan berdasarkan persentase terbesar simpul node pada setiap tingkat kesejahteraan

Status Kesejahteraan	Karakteristik
Desil 1	<ol style="list-style-type: none"> <li>1. Jumlah ART 8,9,10,11</li> <li>2. Tidak memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>3. Sumber air minum bukan air kemasan bermerk</li> <li>4. Tidak memiliki telepon rumah</li> <li>5. Memiliki kulkas</li> </ol>
Desil 2	<ol style="list-style-type: none"> <li>1. Memiliki luas lantai <math>&gt;45 \text{ m}^2</math></li> <li>2. Tidak memiliki dinding terluas berkategori lain</li> <li>3. Sumber air minum bukan air kemasan bermerk</li> <li>4. Status penguasaan bangunan tempat tinggal yang ditempati: kontrak/sewa</li> <li>5. Memiliki toilet bersama dengan jenis kloset leher angsa</li> <li>6. Sumber air minum bukan ledeng eceran ataupun sumur bor/pompa</li> <li>7. Tidak memiliki sepeda motor</li> <li>8. Jenis lantai terluas bukan marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>9. Jumlah ART 3</li> <li>10. Tidak memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>11. Tidak memiliki kulkas</li> </ol>

Desil 3	<ol style="list-style-type: none"> <li>1. Status penguasaan bangunan tempat tinggal yang ditempati: kontrak/sewa</li> <li>2. Tidak memiliki jenis dinding terluas berkategori lain</li> <li>3. Memiliki sumber air minum ledeng eceran atau sumur bor/pompa</li> <li>4. Tidak memiliki toilet bersama dengan jenis kloset leher angsa</li> <li>5. Jenis lantai terluas bukan marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>6. Jumlah ART 1</li> <li>7. Tidak memiliki kulkas</li> </ol>
Desil 4	<ol style="list-style-type: none"> <li>1. Memiliki telepon rumah</li> <li>2. Memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>3. Jenis lantai terluas marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>4. Sumber air minum bukan air kemasan bermerk</li> <li>5. Jumlah ART 4</li> <li>6. Memiliki kulkas</li> </ol>

Dari hasil pengolahan pohon klasifikasi diatas didapatkan beberapa variabel terpenting dalam pemilihan pohon klasifikasi. Berikut merupakan 5 variabel dengan nilai *importance* tertinggi.

**Tabel 4.8** Lima variabel dengan nilai *importance* tertinggi

Variabel	<i>Importance</i>	<i>Normalized importance</i>
Jumlah Anggota Rumah Tangga	0.119	100%
Jenis lantai terluas: Marmer/Granit, Keramik, Parket/Vinil/Permadani	0.079	66.2%
Kepemilikan Kulkas	0.074	62.4%
Tipe toilet: Pribadi dengan Jenis	0.055	46.6%

Kloset Leher Angsa		
Kepemilikan Sepeda Motor	0.046	38.6%

Variabel terpenting dalam pemilahan pohon klasifikasi optimal yang mengklasifikasikan rumah tangga di Kota Surabaya berturut-turut yaitu jumlah anggota rumah tangga dengan nilai *importance* 0,119, jenis lantai terluas dengan bahan marmer/ granit, keramik, atau parket/ vinil/ permadani dengan nilai *importance* 0,079, dan kepemilikan kulkas dengan nilai *importance* 0,074. Sedangkan variabel yang lain mendapatkan skor dibawah skor ketiga variabel tersebut. *Normalized importance* merupakan persentase kepentingan dari variabel tersebut, yang dihitung dengan membagi nilai *importance* variabel dengan nilai *importance* tertinggi. Dari hasil tersebut diketahui bahwa variabel yang berpengaruh dalam pembentukan pohon klasifikasi yaitu jumlah anggota rumah tangga, jenis lantai terluas dengan bahan marmer/ granit, keramik, atau parket/ vinil/ permadani, kepemilikan kulkas, tipe toilet pribadi dengan jenis kloset leher angsa, dan kepemilikan sepeda motor.

Pohon klasifikasi tingkat kesejahteraan rumah tangga di Kota Surabaya memiliki ketepatan klasifikasi sebesar 64.1%. Persentase untuk menduga rumah tangga desil 1 secara benar yaitu 76.3%, menduga rumah tangga desil 2 secara benar yaitu 57.5%, rumah tangga desil 3 secara benar yaitu 56.7% dan rumah tangga desil 4 secara benar yaitu 68.4%.

**Tabel 4.9** Kesesuaian klasifikasi pohon klasifikasi pada setiap tingkat kesejahteraan rumah tangga di Kota Surabaya

<i>Observasi</i>	% Ketepatan Klasifikasi				% Kebenaran
	Desil 1	Desil 2	Desil 3	Desil 4	
Desil 1	15708	4185	491	190	76.3%
Desil 2	5557	15098	5222	361	57.5%
Desil 3	1289	7191	15727	3517	56.7%
Desil 4	319	1875	7259	20435	68.4%
Persen total	21.9%	27.1%	27.5%	23.5%	64.1%

Model pohon klasifikasi penelitian ini lebih baik dalam menduga rumah tangga pada desil 1, lalu selanjutnya baik dalam menduga desil 4, desil 2, dan yang terakhir desil 3. Dari tabel kesesuaian klasifikasi tersebut dapat diketahui bahwa pohon klasifikasi yang terbentuk merupakan pohon klasifikasi estimasi yang tepat dengan tingkat akurasi sebesar 64,1%.

## **BAB V**

### **KESIMPULAN DAN SARAN**

Bab ini berisi tentang kesimpulan dari analisis yang telah dilakukan sebelumnya dan saran yang ditujukan bagi penelitian selanjutnya.

#### **5.1 Kesimpulan**

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, kesimpulan yang didapatkan dari penelitian ini adalah sebagai berikut.

1. Karakteristik variabel respon pada penelitian ini yaitu persentase status kesejahteraan rumah tangga desil 4 merupakan persentase terbesar dari semua desil. Pada variabel kepemilikan aset, sepeda motor dan kulkas banyak dimiliki oleh setiap rumah tangga. Dari segi bahan bakar untuk memasak, sebagian besar rumah tangga tidak menggunakan kayu bakar untuk memasak dan mereka memasak di rumah. Dari segi sumber air minum, pemakai terbanyak yaitu bersumber dari ledeng eceran, sumur bor/pompa. dari segi status penguasaan bangunan tempat tinggal, persentase rumah tangga mengontrak/sewa lebih banyak daripada rumah tangga yang menempati rumah dinas. Untuk kategori bangunan, jenis lantai terluas yang dipakai yaitu marmer/ granit, keramik, parket/ vinil/ permadani dengan rata-rata luas lantai yaitu 36.9 m<sup>2</sup>. Jenis dinding yang dipakai yaitu bukan berkategori lain. Tipe toilet yang banyak digunakan yaitu toilet pribadi dengan jenis kloset leher angsa. Untuk kategori pendidikan, banyak anggota rumah tangga yang hanya menamatkan pendidikannya hingga SLTA dan SLTP serta tidak melanjutkan ke jenjang yang lebih tinggi yaitu perguruan tinggi. Pada kategori pekerjaan, anggota rumah tangga banyak yang bekerja pada perusahaan industri sebagai

pegawai. Rata-rata jumlah anggota rumah tangga yaitu 4 orang dan jumlah anggota rumah tangga yang berjenis kelamin laki-laki yaitu 2 orang. Masih banyak keluarga di Indonesia yang mempunyai keturunan dalam jumlah banyak sehingga program KB harus lebih digencarkan.

2. Hasil perbandingan metode test sample, metode terbaik yang memiliki ketepatan klasifikasi terbesar yaitu metode *test sample* dengan *testing* 30% menggunakan Indeks Gini, dengan ketepatan klasifikasi sebesar 64.1%. Pohon klasifikasi yang terbentuk mempunyai 391 simpul dan 191 simpul terminal dengan pemilah awal adalah variabel jumlah anggota rumah tangga.
3. Karakteristik yang dipilih merupakan karakteristik dengan persentase terbesar pada simpul terminal dalam setiap tingkat kesejahteraan. Berdasarkan hasil pohon klasifikasi, karakteristik yang membedakan pada masing-masing tingkat kesejahteraan adalah sebagai berikut.

- a. Desil 1

Karakteristik pada desil 1 diantaranya jumlah anggota rumah tangga berkisar antara 8-11 orang, tidak memiliki toilet pribadi dengan jenis kloset leher angsa, sumber air minum bukan air kemasan bermerk, tidak memiliki telepon rumah, dan memiliki kulkas. Tingkat keakuratan 87.8%.

- b. Desil 2

Pada desil 2, karakteristik yang didapatkan yaitu memiliki luas lantai  $>45 \text{ m}^2$ , tidak memiliki dinding terluas berkategori lain (berdinding tembok, plesteran anyaman bambu, kayu, anyaman bambu, batang kayu, atau bambu), sumber air minum bukan air kemasan bermerk, tempat tinggal yang ditempati merupakan kontrak/sewa, memiliki toilet bersama dengan jenis kloset leher angsa, memiliki sumber air minum bukan ledeng eceran maupun sumur bor/pompa, tidak memiliki sepeda motor, jenis lantai terluas bukan



marmer/granit, keramik, ataupun parket/vinil/permadani, jumlah anggota rumah tangga 3 orang, tidak memiliki toilet pribadi dengan jenis kloset leher angsa, dan tidak memiliki kulkas. Tingkat keakuratan 100%.

c. Desil 3

Rumah tangga pada tingkat kesejahteraan ini mengontrak/menyewa tempat tinggal, tidak memiliki jenis dinding terluas berkategori lain (berdinding tembok, plesteran anyaman bambu, kayu, anyaman bambu, batang kayu, atau bambu), memiliki sumber air minum ledeng eceran atau sumur bor/pompa, tidak memiliki toilet bersama dengan jenis kloset leher angsa, jenis lantai terluas bukan marmer/granit, keramik, ataupun parket/vinil/permadani, jumlah anggota rumah tangga 1 orang, dan tidak memiliki kulkas. Tingkat keakuratan 90.5%.

d. Desil 4

Desil 4 memiliki karakteristik yaitu memiliki telepon rumah, memiliki toilet pribadi dengan jenis kloset leher angsa, jenis lantai terluas marmer/granit, keramik, ataupun parket/vinil/permadani, sumber air minum bukan air kemasan bermerk, jumlah anggota rumah tangga 4, dan memiliki kulkas. Tingkat keakuratan 100%.

4. Terdapat beberapa variabel penting dalam pembentukan pohon klasifikasi. Lima variabel terpenting dalam pemilahan pohon klasifikasi yaitu jumlah anggota rumah tangga, jenis lantai terluas dengan bahan marmer/granit, keramik, parket/vinil/permadani, kepemilikan kulkas, tipe toilet pribadi dengan jenis kloset leher angsa, dan kepemilikan sepeda motor.
5. Berdasarkan hasil perhitungan kesesuaian estimasi, didapatkan hasil jika pohon klasifikasi tersebut merupakan pohon klasifikasi dengan *right estimated*.

## **5.2 Saran**

Saran yang dapat diberikan dari hasil penelitian ini, untuk Pemerintah Kota Surabaya dapat menjadi acuan dan pertimbangan untuk program pengentasan kemiskinan berikutnya. Sedangkan untuk penelitian selanjutnya diharapkan untuk mencoba menggunakan metode dan software lain dengan hasil yang lebih akurat sehingga karakteristik setiap desil bisa terdefinisi dengan benar dan memiliki tingkat akurasi yang tinggi.

## DAFTAR PUSTAKA

- Agresti, A. (2013). *Categorical Data Analyst (2nd ed.)*. New Jersey: John Wiley & Sons.
- Akbar, M. S., Yuanita, D., dan Harini, S. (2009). Pendekatan CART untuk Mendapatkan Faktor yang Mempengaruhi Terjangkitnya Penyakit Demam Tifoid di Aceh Utara. *Jurnal Cauchy*, 1, No. 2, 71-77.
- Aprilia, T., Gusriani, N., dan Parmhati, K. (2015). Klasifikasi Ketepatan masa Studi Mahasiswa FMIPA Unpad Angkatan 2001-2006 dengan Menggunakan Metode Classification and Regression Trees (CART). *Jurnal Matematika Integratif*, 11, No 1, 7-14.
- Aritonang, N., Rusgiyono, A., dan Rahmawati, R. (2016). Klasifikasi Status Kerja pada Angkatan Kerja Kota Semarang tahun 2014 Menggunakan Metode CHAID dan CART. *Jurnal Gaussian Universitas Diponegoro*, 5, No.1, 183-192.
- BPS. (2014). Kemiskinan dan Kesejahteraan. <http://surabayakota.bps.go.id>. Dikutip 14 Februari 2018.
- Breiman, L., Friedman, J., Olshen, R., dan Stone, C. (1984). *Classification and Regression Trees*. New York: Chapman Hall.
- Du, S. T., Steyn, A., dan Stumph, R. (1986). *Graphical Exploratory Data Analysis*. New York: Springer-Verlag.
- Hair, J., Black, W., Babin, B. J., dan Anderson, R. (2010). *Multivariate Data Analysis, 7th ed.* New Jersey: Prentice Hall.
- Han, J., Kamber, M., dan Pei, J. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Inayah, R., Purnami, S. W., dan Otok, B. W. (2014). Klasifikasi Rumah Tangga Miskin di Kabupaten Jombang Berdasarkan Faktor-faktor yang Mempengaruhi dengan Pendekatan CART. *Jurnal Sains dan Seni Pomits*, 3, No. 2, 2337-3520.

- Jiwadiana, I. G., Sukarsa, I. K., & Srinadi, I. G. (2015). Klasifikasi Karakteristik Kecelakaan Lalu Lintas di Kota Denpasar dengan Pendekatan Classification and Regression Trees (CART). *Jurnal Matematika*, 4, No.4, 146-151.
- Johnson, R. A., dan Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis (sixth ed.)*. New Jersey: Prentice Hall.
- Lewis, R. (2000). *An Introduction to Classification and Regression Tree (CART) Analysis (A. M. Medicine, Ed.)*. San Fransisco.
- Mahalisa, G. (2016). Metode Pohon Gabungan Pada CART Untuk Analisa Kesejahteraan Rumah Tangga di Banjarmasin. *Jurnal Technologia*, 7, No. 1, 15-24.
- Putri, E., Suparti, dan Rahmawati. (2014). Perbandingan Metode Klasifikasi Naïve Bayes dan K-Nearest Neighbor pada Analisis Data Status Kerja di Kabupaten Demak Tahun 2012. *Jurnal Gaussian*, 3, No. 4, 831-838.
- Seftiana, D. (2015). Klasifikasi Rumah Tangga Sangat Miskin di Kabupaten Jombang dengan Pendekatan *Random Forest Classification and Regression Tree*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Walpole, R. E. (1995). *Pengantar Statistika (Terjemahan) (3 ed.)*. Jakarta: PT. Gramedia Pustaka Utama.

## LAMPIRAN

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

No Urut Ruta	Status Kesejah teraan	Kepemilikan Mobil	Kepemilikan Kulkas	...	Tipe toilet: Bersama dengan Jenis Kloset Leher Angsa
1	4	0	1	...	0
2	3	0	1	...	0
3	3	0	1	...	1
4	3	0	1	...	0
5	3	0	0	...	0
6	4	0	1	...	0
7	4	0	0	...	1
8	3	0	0	...	0
9	1	0	0	...	1
10	2	0	1	...	0
11	2	0	1	...	0
12	3	0	0	...	0
13	2	0	1	...	0
14	2	0	0	...	1
15	3	0	1	...	0
16	3	0	0	...	0
17	2	0	0	...	0
18	4	0	3	...	0
19	2	0	1	...	0
⋮	⋮	⋮	⋮	⋮	⋮
149183	4	0	1	...	0
149184	1	0	0	...	0
149185	1	0	1	...	0

## Lampiran 2. Output Model Summary

### Model Summary

Specific ations	Growing Method	CRT	
	Dependent Variable	Status Kesejahteraan	
	Number of Nodes		381
	Number of Terminal Nodes		191
	Depth		13

## Lampiran 3. Cross Tabulation Data Learning dan Testing

### Classification

Sample	Observed	Predicted				Percent Correct
		1	2	3	4	
Training	1	6751	1732	208	81	77.0%
	2	2361	6586	2123	186	58.5%
	3	497	2952	6945	1479	58.5%

	4	131	835	3031	8864	68.9%
	Overall Percentage	21.8%	27.0%	27.5%	23.7%	65.1%
Test	1	15708	4185	491	190	76.3%
	2	5557	15098	5222	361	57.5%
	3	1289	7191	15727	3517	56.7%
	4	319	1875	7259	20435	68.4%
	Overall Percentage	21.9%	27.1%	27.5%	23.5%	64.1%

Growing Method: CRT

Dependent Variable: Status Kesejahteraan

#### **Lampiran 4.** Variabel Penting dalam Pembentukan Pohon Klasifikasi

##### **Independent Variable Importance**

Independent Variable	Importance	Normalized Importance
Jumlah ART	.119	100.0%

Jenis Lantai Terluas: Marmer/Granit, Keramik, Parket/Vinil/Permadani	.079	66.2%
Kepemilikan Kulkas	.074	62.4%
Tipe toilet: Pribadi dengan Jenis Kloset Leher Angsa	.055	46.6%
Kepemilikan Sepeda Motor	.046	38.6%
Sumber Air Minum: Air Kemasan Bermerk	.040	33.7%
Jumlah ART Laki-laki	.039	32.7%
Luas Lantai per Kapita	.035	29.5%
Jenis Lantai Terluas: Ubin/Tegel/Braso	.034	28.5%
Sumber Air Minum: Ledeng eceran, Sumur Bor/Pompa	.034	28.4%
Tipe toilet: Bersama dengan Jenis Kloset Leher Angsa	.029	24.7%
Kemilikan Telepon Rumah	.028	23.3%



Status Penguasaan Bangunan		
Tempat Tinggal yang ditempati: Kontrak/Sewa	.019	16.0%
Jenis Lantai Terluas: Kayu/Papan Kualitas Tinggi, Semen/Bata Merah, Bambu, Kayu/Papan Kualitas Rendah	.015	13.0%
Jumlah ART yang Menamatkan SLTA	.013	10.7%
Kepemilikan Komputer	.011	9.0%
Jumlah ART yang Menamatkan SLTP	.007	6.2%
Kepemilikan Mobil	.006	4.9%
Sumber Air Minum: Air Isi Ulang, Ledeng Meteran	.006	4.9%
Jenis Dinding Terluas: Berkategori Lain	.005	4.2%
Jumlah ART yang Bersekolah di SLTA	.003	2.4%
Jumlah ART yang Bekerja di Konstruksi dengan Status Sebagai Pekerja Bebas	.003	2.2%

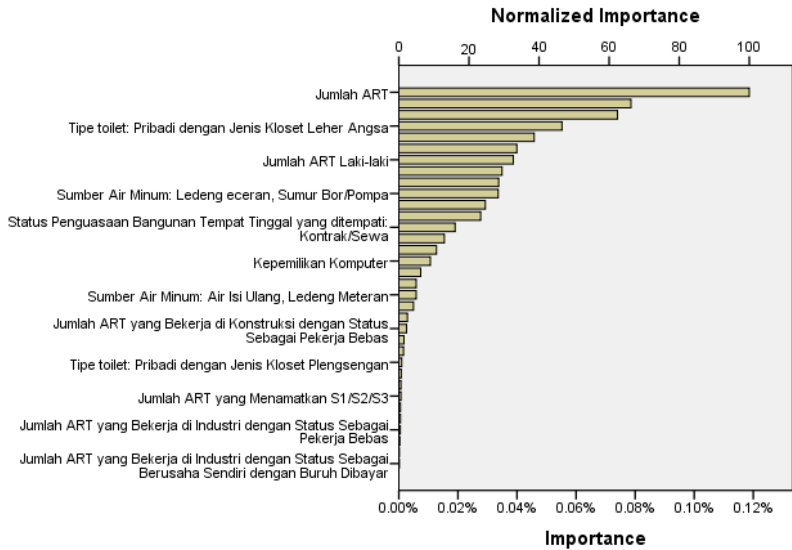
Bahan Bakar/ Energi untuk Memasak: Kayu Bakar Tidak Memasak di rumah	.002	1.5%
Jumlah ART yang Bekerja di Industri dengan Status Pegawai	.002	1.3%
Tipe toilet: Pribadi dengan Jenis Kloset Plengsengan	.001	0.8%
Status Penguasaan Bangunan Tempat Tinggal yang ditempati: Dinas	.001	0.7%
Jumlah ART yang Bekerja di Pertanian dengan Status Sebagai Berusaha Sendiri	.001	0.6%
Jumlah ART yang Menamatkan S1/S2/S3	.001	0.6%
Jumlah ART yang Bekerja di Perusahaan Listrik dan Gas dengan Status Sebagai Pegawai	.001	0.5%
Jumlah ART yang Menamatkan D1/D2/D3	.001	0.5%

Jumlah ART yang Bekerja di Industri dengan Status Sebagai Pekerja Bebas	.000	0.4%
Jumlah ART yang Bekerja di Industri dengan Status Sebagai Berusaha Sendiri	.000	0.3%
Jumlah ART yang Bekerja di Pertanian dengan Status Sebagai Pekerja Tak Dibayar	.000	0.1%
Jumlah ART yang Bekerja di Industri dengan Status Sebagai Berusaha Sendiri dengan Buruh Dibayar	.000	0.1%

Growing Method: CRT

Dependent Variable: Status Kesejahteraan

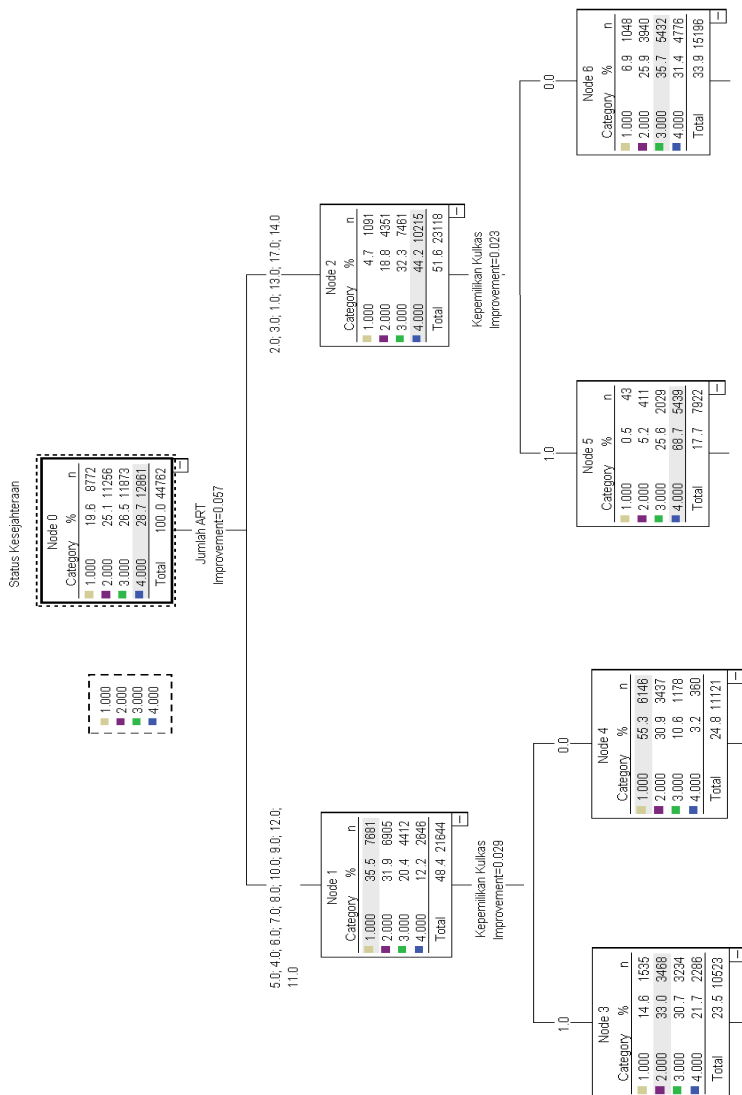
## Lampiran 5. Diagram batang *Normalized Importance*



Growing Method: CRT

Dependent Variable: Status Kesejahteraan

## Lampiran 6. Topografi Pohon Klasifikasi Kedalaman 3



## Lampiran 7. Surat Pernyataan Data

### SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika FMKSD ITS:

Nama : Zikrariza Kurnia Malta

NRP : 0621154000042

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 : Penelitian Pembimbing dari BAPEKKO

Keterangan : Penyusunan Model PMT di Kota Surabaya

Surat Pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Mengetahui  
Pembimbing Tugas Akhir



(Dr. Sutikno, S.Si, M.Si)  
NIP. 19710313 199702 1 001

Surabaya, Juli 2019



(Zikrariza Kurnia Malta)  
NRP. 0621154000042

## BIODATA PENULIS



Penulis lahir di Malang pada tanggal 31 Desember tahun 1996. Pendidikan formal yang telah ditempuh penulis adalah SD Negeri Kasin Malang pada tahun 2003 hingga 2009, SMP Negeri 3 Malang pada tahun 2009 hingga 2012 dan SMA Negeri 4 Malang pada tahun 2012 hingga 2015. Lalu pada tahun 2015 penulis melanjutkan Pendidikan di Departemen Statistika Institut Teknologi Sepuluh Nopember

(ITS) Surabaya. Pada masa perkuliahan, penulis aktif di berbagai organisasi dan kegiatan. Salah satu organisasi yang diikuti oleh penulis yaitu Paguyuban Beasiswa Karya Salemba Empat (KSE) ITS pada tahun 2018-2019. Pada paguyuban tersebut penulis berperan sebagai sekretaris departemen Pemberdayaan Sumber Daya Anggota (PSDA). Selain itu penulis juga mengikuti beberapa UKM, salah satunya yaitu UKM Cinta Rebana. Penulis berperan aktif menjadi pengurus di UKM tersebut sebagai sekretaris departemen Media Informasi (Medfo) pada tahun 2017-2018. Penulis juga berpengalaman melakukan kerja praktik di Bank Indonesia Surabaya di bagian Sistem Pembayaran-Pengelolaan Uang Rupiah (SP-PUR). Jika ada pembaca yang ingin berdiskusi lebih lanjut dengan penulis, dapat menghubungi alamat email [zikrarizakm@gmail.com](mailto:zikrarizakm@gmail.com).