

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

Zikrariza Kurnia Malta dan Sutikno

Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data,

Institut Teknologi Sepuluh Nopember (ITS)

*e-mail*: Sutikno@statistika.its.ac.id

**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 35 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 marmer/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*.

## I. PENDAHULUAN

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 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 35 variabel yang berhubungan dengan kesejahteraan. TNP2K menetapkan sebanyak 40% ruta 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 ruta tersebut lebih rendah dibandingkan dengan tingkat konsumsi ruta 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 CART (*Classification and Regression Tree*). Metode CART dengan variabel respon numerik 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 [1] 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 [1]. Penelitian lain dilakukan oleh Inayah, Purnami, dan Otok yang meneliti tentang rumah tangga miskin di Kabupaten Jombang berdasarkan faktor yang mempengaruhi [2]. Selain itu juga penelitian yang meneliti tentang kemiskinan yaitu penelitian oleh Mahalisa tentang analisis kesejahteraan rumah tangga di Banjarmasin menggunakan metode pohon gabungan pada CART [3]. 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 mengentaskan kemiskinan di Surabaya.

## II. TINJAUAN PUSTAKA

### A. 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 [4].

Statistika deskriptif bisa disajikan dalam bentuk tabel, boxplot, dan cross tabulation sehingga informasi yang disampaikan akan lebih mudah dipahami oleh pembaca. *Box plot* adalah cara standar untuk menampilkan distribusi data berdasarkan lima rangkuman (minimum, kuartil pertama, median, kuartil ketiga, dan maksimum). Sedangkan *cross tabulation* adalah metode statistika yang menggambarkan dua atau lebih variabel secara bersama-sama yang hasilnya berupa tabel kontingensi.

### B. 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 [5]. Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi atau akan menghasilkan peluang kesalahan alokasi yang hasilnya kecil [6]. Pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi.

Menurut [7], 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:

- 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.

- Kriteria *Goodness of Split* merupakan alat evaluasi bagi pemilahan yang dilakukan pemilah  $s$  pada simpul  $t$  berupa persamaan

$$\Phi(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (1)$$

dimana:

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

$P_L$  : proporsi data yang masuk ke simpul kiri

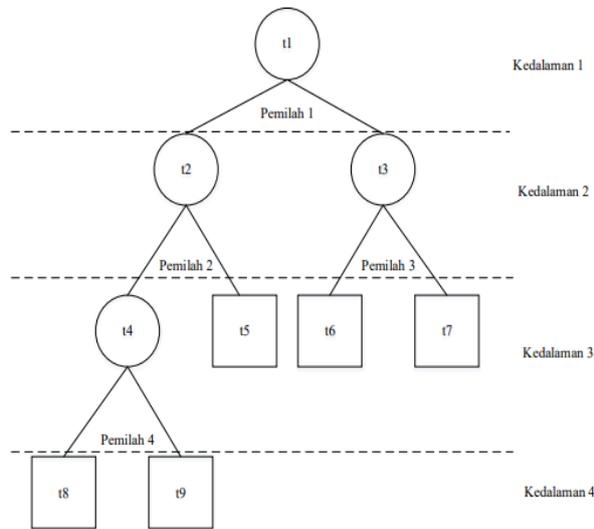
$P_R$  : proporsi data yang masuk ke simpul kanan

- Ukuran yang digunakan untuk menentukan ukuran pohon yang layak yaitu dengan menggunakan sampel uji (test sample) atau sampel validasi silang lipat- $v$ .
- 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 [8].

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 1. Struktur pohon klasifikasi [7].

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 [9]. Tujuan utama CART adalah mendapatkan suatu kelompok data yang akurat sebagai bentuk karakteristik dari suatu pengklasifikasian.

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 treshold 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

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

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

$$i(t) = \frac{P_L P_R}{4} [\sum_j |p(j|t_L) - p(j|t_R)|]^2 \tag{3}$$

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. Proses Pemangkasan Pohon Klasifikasi

Pemangkasan pohon klasifikasi atau bisa disebut *prunning* 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 [7]. Ukuran *cost complexity* ditulis pada persamaan 4.

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}| \tag{4}$$

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.

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.

C. Ketepatan Klasifikasi

Ukuran ketepatan klasifikasi bisa dihitung sesuai persamaan 2.5 dimana *APER* (*Apparent Error Rate*) menunjukkan ukuran kesalahan klasifikasi sejumlah dataset pengamatan oleh suatu fungsi klasifikasi [6]. Perhitungan tersebut bisa didapatkan dari tabel prediksi klasifikasi dengan 4 level kelas variabel respon.

$$Akurasi\ Klasifikasi = 1 - APER = \frac{n_{11} + n_{22} + n_{33} + n_{44}}{N} \tag{5}$$

Tabel 1.

Contoh Tabel Prediksi Klasifikasi dengan 4 Level Kelas Variabel Respon

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

D. 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 [10]. Dalam mengukur kesejahteraan rumah tangga diperlukan indikator moneter, indikator yang banyak digunakan adalah pendapatan dan pengeluaran [10]. 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) [10]. 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.

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 asset.

III. METODOLOGI PENELITIAN

A. Sumber data

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

B. Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini adalah karakteristik kesejahteraan rumah tangga yang terdiri dari variabel kategori dan variabel kontinu sebagai berikut.

Tabel 2.  
Variabel Penelitian

Variabel	Deskripsi	Skala Data
X <sub>1</sub>	Kepemilikan Mobil	Nominal
X <sub>2</sub>	Kepemilikan Kulkas	Nominal
X <sub>3</sub>	Kepemilikan Sepeda Motor	Nominal
X <sub>4</sub>	Kepemilikan Komputer	Nominal
X <sub>5</sub>	Kepemilikan Telepon Rumah	Nominal
X <sub>6</sub>	Bahan Bakar / Energi Untuk Memasak: Kayu Bakar, Tidak memasak di rumah	Nominal
X <sub>7</sub>	Sumber Air Minum: Air Kemasan Bermerk	Nominal
X <sub>8</sub>	Sumber Air Minum: Air Isi Ulang, Ledeng Meteran	Nominal
X <sub>9</sub>	Sumber Air Minum: Ledeng Eceran, Sumur Bor/Pompa	Nominal
X <sub>10</sub>	Jumlah Anggota Rumah Tangga (ART)	
X <sub>11</sub>	Status Penguasaan Bangunan Tempat Tinggal yang Ditempati: Kontrak/Sewa	Nominal
X <sub>12</sub>	Status Penguasaan Bangunan Tempat Tinggal yang Ditempati: Dinas	Nominal
X <sub>13</sub>	Luas Lantai per Kapita	Rasio
X <sub>14</sub>	Jumlah ART yang Menamatkan D1/D2/D3	Rasio
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	Rasio

	Pertanian dengan Status Sebagai Berusaha Sendiri		
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 Sebagai Pekerja Bebas	Rasio	
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	
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	Nominal	
X <sub>30</sub>	Jenis Lantai Terluas: Ubin/Tegel/Braso	Nominal	
X <sub>31</sub>	Jenis Lantai Terluas: Kayu/Papan Kualitas Tinggi, Semen/Bata Merah, Bambu, Kayu/Papan Kualitas Rendah	Nominal	
X <sub>32</sub>	Jenis Dinding Terluas: Berkategori Lain	Nominal	
X <sub>33</sub>	Tipe toilet: Pribadi dengan Jenis Kloset Leher Angsa	Nominal	

C. Langkah Penelitian

Berikut ini adalah langkah-langkah penelitian yang dilakukan dalam penelitian ini.

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 dan 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 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

IV. ANALISIS DAN PEMBAHASAN

A. *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. Selanjutnya yaitu menandai data yang tidak dipakai dengan tanda "NULL" dan akan dieliminasi sehingga data yang muncul hanyalah data yang dibutuhkan. 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.

C. *Pohon Klasifikasi*

Setelah dilakukan penyajian data menggunakan statistika deskriptif, dilakukan analisis klasifikasi status kesejahteraan rumah tangga di Kota Surabaya menggunakan pohon klasifikasi karena variabel respon merupakan variabel 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.

Pada Tabel 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.

Tabel 5.  
Perbandingan Ketepatan Klasifikasi dari Metode *Test Sample*

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

Pada topologi pohon klasifikasi 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.

Dari setiap simpul terminal dapat diketahui karakteristik setiap tingkat kesejahteraan rumah tangga. Karakteristik yang dipilih yaitu karakteristik dari simpul terminal yang memiliki persentase terbesar pada setiap tingkat kesejahteraan. 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 6.  
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>Jumlah ART 8,9,10,11 orang</li> <li>Tidak memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>Sumber air minum bukan air kemasan bermerk</li> <li>Tidak memiliki telepon rumah</li> <li>Memiliki kulkas</li> </ol>
Desil 2	<ol style="list-style-type: none"> <li>Memiliki luas lantai &gt;45 m<sup>2</sup></li> <li>Tidak memiliki dinding terluas berkategori lain</li> <li>Sumber air minum bukan air kemasan bermerk</li> <li>Status penguasaan bangunan tempat tinggal yang ditempati: kontrak/sewa</li> <li>Memiliki toilet bersama dengan jenis kloset leher angsa</li> <li>Sumber air minum bukan ledeng eceran ataupun sumur bor/pompa</li> <li>Tidak memiliki sepeda motor</li> <li>Jenis lantai terluas bukan marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>Jumlah ART 3</li> <li>Tidak memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>Tidak memiliki kulkas</li> </ol>
Desil 3	<ol style="list-style-type: none"> <li>Status penguasaan bangunan tempat tinggal yang ditempati: kontrak/sewa</li> <li>Tidak memiliki jenis dinding terluas berkategori lain</li> <li>Memiliki sumber air minum ledeng eceran atau sumur bor/pompa</li> <li>Tidak memiliki toilet bersama dengan jenis kloset leher angsa</li> <li>Jenis lantai terluas bukan marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>Jumlah ART 1</li> <li>Tidak memiliki kulkas</li> </ol>
Desil 4	<ol style="list-style-type: none"> <li>Memiliki telepon rumah</li> <li>Memiliki toilet pribadi dengan jenis kloset leher angsa</li> <li>Jenis lantai terluas marmer/granit, keramik, ataupun parket/vinil/permadani</li> <li>Sumber air minum bukan air kemasan bermerk</li> <li>Jumlah ART 4</li> <li>Memiliki kulkas</li> </ol>

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

Tabel 7.  
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 Kloset Leher Angsa	0.055	46.6%
Kepemilikan Sepeda Motor	0.046	38.6%

Variabel terpenting dalam pemilihan 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%. 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.

Tabel 8.  
Cross tabulation pohon klasifikasi optimal  
% Ketepatan Klasifikasi

Observasi	% 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%

Dari tabel kesesuaian klasifikasi tersebut dapat diketahui juga apakah kesesuaian klasifikasi dari pohon tersebut sudah sesuai, over estimated, atau under estimated dengan menghitung jumlah dari nilai-nilai tabel yang sesuai dengan kriteria. Over estimated dihitung dari nilai tabel dari desil 1 ke desil 2, desil 1 ke desil 3, desil 1 ke desil 4, desil 2 ke desil 3, desil 2 ke desil 4, dan desil 3 ke desil 4. Sedangkan under estimated dihitung dari nilai tabel dari desil 4 ke desil 1, desil 4 ke desil 2, desil 4 ke desil 3, desil 3 ke desil 1, desil 3 ke desil 2, dan desil 2 ke desil 1. Perhitungan manual over estimated dan under estimated sebagai berikut

$$\begin{aligned} \text{Over estimated} &= \frac{[n_{12} + n_{13} + n_{14} + n_{23} + n_{24} + n_{34}]}{N} \\ &= \frac{[4185 + 491 + 190 + 5222 + 361 + 3517]}{104424} \\ &= 13.3\% \end{aligned}$$

Sedangkan untuk perhitungan under estimated yaitu sebagai berikut

$$\begin{aligned} \text{Under estimated} &= \frac{[n_{41} + n_{42} + n_{43} + n_{31} + n_{32} + n_{21}]}{N} \\ &= \frac{[319 + 1875 + 7259 + 1289 + 7191 + 5557]}{104424} \\ &= 22.4\% \end{aligned}$$

Pada perhitungan diatas dapat diketahui jika kemungkinan pohon klasifikasi ini mengalami over estimated hanya sebesar 13.3% dan mengalami under estimated 22.4%. Nilai tersebut jauh dibawah tingkat akurasi right estimated yaitu sebesar 64.1%. Sehingga dapat diketahui bahwa pohon klasifikasi yang terbentuk merupakan pohon klasifikasi estimasi yang tepat dengan tingkat akurasi sebesar 64.1%.

## V. KESIMPULAN DAN SARAN

### A. 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 m<sup>2</sup>, 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%.

5. 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.

6. Berdasarkan hasil perhitungan kesesuaian estimasi, didapatkan hasil jika pohon klasifikasi tersebut merupakan pohon klasifikasi dengan *right estimated*.

#### B. 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

- [1] N. Aritonang and et al, "Klasifikasi Status Kerja pada Angkatan Kerja Kota Semarang tahun 2014 Menggunakan Metode CHAID dan CART," *J. Gaussian Univ. Diponegoro*, vol. 5, no. 1, pp. 183–192, 2016.
- [2] S. Purnami and et al, "Klasifikasi Rumah Tangga Miskin di Kabupaten Jombang Berdasarkan Faktor-faktor yang Mempengaruhi dengan Pendekatan CART," *J. Sains dan Seni Pomits*, vol. 3, no. 2, pp. 2337–3520, 2014.
- [3] G. Mahalisa, "Metode Pohon Gabungan Pada CART Untuk Analisa Kesejahteraan Rumah Tangga di Banjarmasin," *J. Technol.*, vol. 7, no. 1, pp. 15–24, 2016.
- [4] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability & Statistics for Engineers & Scientists*, 9th ed. USA: Pearson Education Inc, 2012.
- [5] J. Han, *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall, 2007.
- [6] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Pearson Education, Inc, 2007.
- [7] L. Breiman, J. Friedman, R. Olshen, and Stone, *Classification and Regression Trees*. New York: Chapman Hall, 2007.
- [8] R. Lewis, *Introduction to Classification and Regression Tree (CART) Analysis*. 2000.
- [9] S. Du and et al, *Graphical Exploratory Data Analysis*. New York: Springer-Verlag, 1986.
- [10] [BPS] Badan Pusat Statistika, "Kemiskinan dan Kesejahteraan." [Online]. Available: <http://surabayakota.bps.go.id>.