

**KLASIFIKASI KEMISKINAN DI KOTA SEMARANG MENGGUNAKAN
ALGORITMA *CHISQUARE AUTOMATIC INTERACTION DETECTION* (CHAID)
DAN *CLASSIFICATION AND REGRESSION TREE* (CART)**

Dwi Ispriyanti, Alan Prahutama, Mustafid

Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

e-mail: alan.prahutama@gmail.com

DOI: 10.14710/medstat.12.1.63-72

Article Info:

Received: 17 January 2019

Accepted: 22 July 2019

Available Online: 24 July 2019

Keywords:

Classification, poverty level, CHAID, CART

Abstract: Decreasing poverty level is the first goal of Sustainable Development Goals (SDGs). Poverty in Central Java from 2002 to 2017 has decreased, as well as the city of Semarang. Therefore, it is necessary to examine the factors that determine the decline in poverty classification in the city of Semarang. The classification analysis in statistics uses one classification tree. Several methods using classification trees include CART, CHAID, C45 and ID3 algorithms. In this study the methods used were CART and CHAID Algorithms. CART and CHAID algorithms are binary classification trees. The CART separation rules use split goodness op, while CHAID uses CHI-Square. In the analysis, the value of using CART was 95.2% while CHAID was 95.2%. While the factors that influence poverty classification using CHAID include the acceptance of poor rice, the main building materials of the house walls, and the main fuel for cooking. Whereas with the CART Algorithm the variables that influence are the main fuels for cooking, poor rice receipts, the number of household members, final disposal sites, sources of drinking water, the household head's business field, roofing materials, and building walls.

1. PENDAHULUAN

Penurunan kemiskinan merupakan salah satu tujuan pertama dari *Sustainable Development Goal's* (SDG's). SDG's merupakan program hasil konferensi dari Perserikatan Bangsa-Bangsa (PBB) di tahun 2012. SDG's menetapkan 17 program antara lain pengentasan kemiskinan; mengakhiri kelaparan; kehidupan sehat dan sejahtera; kesetaraan gender; air bersih dan sanitasi yang layak; Energi bersih dan terjangkau; pekerjaan layak dan pertumbuhan ekonomi; industri, inovasi dan infrastruktur; berkurangnya kesenjangan; kota dan komunitas berkelanjutan; konsumsi dan produksi yang bertanggung jawab; penanganan perubahan iklim; ekosistem laut; ekosistem daratan; perdamaian, keadilan dan kelembagaan yang tangguh; serta kemitraan untuk mencapai

tujuan. Dalam penerapan SDG's, pemerintah Indonesia telah memasukan sebagian besar program SDG's kedalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN).

Menurut Badan Pusat Statistik (2014), kelompok rumah tangga berdasarkan pengeluaran per kapita yang sudah dikonversikan berdasarkan garis kemiskinan terdiri dari dua kelompok yaitu kelompok rumah tangga miskin dan kelompok rumah tangga tidak miskin. Garis Kemiskinan digunakan sebagai batas untuk mengelompokkan rumah tangga miskin dan rumah tangga tidak miskin. Rumah tangga miskin adalah rumah tangga yang memiliki rata-rata pengeluaran perkapita per bulan di bawah Garis Kemiskinan (GK). Melihat masih tingginya tingkat kemiskinan maka perlu adanya suatu cara untuk mengidentifikasi kelompok rumah tangga. Salah satunya yaitu dengan melihat faktor-faktor yang dapat menjadi ciri dari kelompok rumah tangga tersebut.

Pada dasarnya pohon klasifikasi dibedakan menjadi dua yaitu pohon klasifikasi biner dan pohon klasifikasi non-biner. Pohon klasifikasi biner merupakan pohon klasifikasi yang menghasilkan dua simpul disetiap penyekatannya (*split*). Sedangkan pohon klasifikasi non-biner bisa menghasilkan lebih dari dua simpul untuk setiap penyekatannya. Metode klasifikasi biner antara lain *Classification and Regression Tree* (CART) dan *Quick, Unbiased, and Efficient Statistical Tree* (QUEST). Sedangkan metode klasifikasi non-biner antara lain *Fast and Automatic Classification Tree* (FACT), C4.5, Algoritma *Chi-Squared Automatic Interaction Detection* (CHAID), FIRM, dan CRUISE (Kim and Loh, 2001).

Algoritma CHAID merupakan algoritma yang dikembangkan dari *Automatic Interaction Detector* (AID) oleh G.V. Kass pada tahun 1980. CHAID merupakan teknik *iterative* yang menguji setiap variabel independen dalam penentuan kelas dari variabel dependen. Setiap variabel independen diuji menggunakan statistik uji Chi-Square terhadap variabel dependennya. Variabel independen dengan nilai Chi-Square terbesar yang digunakan sebagai variabel pemisah. Pada pembentukan pohon klasifikasi menggunakan algoritma CHAID terdapat tiga prosedur antara lain penggabungan, pemisahan dan penghentian (Rokach and Maimon, 2008)

Algoritma CART merupakan algoritma klasifikasi biner yang menggunakan prinsip *Binary Recursive Partitioning*. *Binary* artinya simpulnya dipecah menjadi dua bagian. *Recursive* berarti bahwa proses pemilahan tersebut diulang kembali pada setiap simpul anak sebagai hasil pemilahan terdahulu, sedangkan simpul-simpul anak tersebut sekarang menjadi simpul induk. Proses pemilahan ini akan terus dilakukan sampai tidak ada kesempatan lagi untuk melakukan pemilahan berikutnya. *Partitioning* berarti bahwa data sampel yang dimiliki dipecah ke dalam bagian-bagian atau partisi-partisi yang lebih kecil (Lewis, 2000).

Pada penelitian ini mengkaji klasifikasi kemiskinan di kota Semarang yang mengkategorikan rumah tangga miskin dan tidak miskin dilihat dari faktor yang mempengaruhinya. Metode klasifikasi yang digunakan adalah Algoritma CHAID dan CART.

2. TINJAUAN PUSTAKA

2.1. *Chi-Squared Automatic Interaction Detection* (CHAID)

Pada algoritma CHAID variabel independen dibedakan menjadi tiga yaitu variabel bebas, monotonik, dan variabel mengambang. Algoritma CHAID mensyaratkan bahwa semua variabel independennya bersifat kategorik. Algoritma CHAID dibagi menjadi tiga

tahap, yaitu penggabungan (*merging*), pemisahan (*splitting*) dan penghentian (*stopping*) (Onder and Uyar, 2017). Diagram pohon dimulai dari simpul akar melalui tiga tahap tersebut pada setiap simpul yang terbentuk dan secara berulang. Berikut algoritma dalam metode CHAID (Breiman, *et. al*, 1984) adalah sebagai berikut:

a. Penggabungan (*Merging*)

Pada tahap penggabungan, dilakukan di setiap kategori pada variabel independen terhadap variabel dependen. Pengujian yang dilakukan pada tahap penggabungan menggunakan statistik uji Chi-Square. Misalnya variabel independent (X_1) terdiri dari 3 kategori (X_{1-1} , X_{1-2} dan X_{1-3}), maka pengujian yang dilakukan adalah menguji setiap penggabungan (X_{1-1} dan X_{1-2} digabungkan; X_{1-2} dan X_{1-3} digabungkan; dan X_{1-1} dan X_{1-3} digabungkan) terhadap variabel dependen. Adapun prosedur pengujian statistik Chi Square adalah sebagai berikut:

Hipotesis yang digunakan pada pengujian *Chi-Square* adalah :

- H_0 : Terdapat hubungan antara variabel independen dengan variabel dependen
- H_1 : Tidak Terdapat hubungan antara variabel independen dengan variabel dependen

Statistik Uji yang digunakan adalah sebagai berikut (Wikass, 1980):

$$\chi^2_{hitung} = \sum_{i=1}^r \sum_{j=1}^c \left[\frac{(n_{ij} - E_{ij})^2}{E_{ij}} \right]; \quad E_{ij} = \frac{n_i \cdot n_j}{n}$$

n_{ij} adalah banyaknya pengamatan yang termasuk dalam kategori ke-i dari variabel independen dan kategori ke-j dari variabel dependen; E_{ij} merupakan frekuensi harapan pengamatan yang termasuk dalam kategori ke-i dari variabel independen dan kategori ke-j dari variabel dependen; r merupakan jumlah kategori dalam variabel independen dan c merupakan jumlah kategori dalam variabel dependen. n_i merupakan banyaknya pengamatan yang termasuk dalam kategori ke-i dari variabel independen; n_j merupakan banyaknya pengamatan yang termasuk dalam kategori ke-j dari variabel dependen. Kriteria pengujian, Tolak H_0 jika $\chi^2_{hitung} > \chi^2_{\alpha; (r-1)(c-1)}$.

b. Pemisahan (*Splitting*)

Tahap pemisahan dilakukan dengan melihat setiap variabel independen. Variabel independen yang dipilih sebagai variabel pemisah adalah variabel independen yang paling signifikan berdasarkan pengujian Chi Square atau variabel dengan nilai Chi-Square terbesar.

c. Penghentian (*stopping*)

Tahap penghentian dilakukan jika tidak adalagi variabel independen yang signifikan. Selain itu penghentian bisa dilakukan jika pohon keputusan mencapai maksimum dari nilai yang sudah ditentukan (kedalaman pohon). Selain itu proses berhenti jika ukuran dari simpul anak kurang dari ukuran minimum yang ditentukan.

3.1. Classification and regression Tree (CART)

CART termasuk kedalam metode klasifikasi nonparametric sehingga tidak membutuhkan asumsi. CART terdiri dari dua bagian yaitu klasifikasi dan regresi. Apabila variabel dependen berupa data diskrit, maka akan menghasilkan pohon klasifikasi sedangkan apabila variabel dependen berupa data kontinue maka menghasilkan model regresi. Algoritma CART dalam membuat pohon klasifikasi bersifat *Binary Recursive Partitioning*. *Binary* artinya setiap simpul induk akan selalu mengalami pemecahan

kedalam dua simpul anak. *Recursive* berarti bahwa proses pemecahan tersebut diulang kembali pada setiap simpul anak sebagai hasil pemecahan terdahulu, sedangkan simpul-simpul anak tersebut sekarang menjadi simpul induk. Proses pemecahan ini akan terus dilakukan sampai tidak ada kesempatan lagi untuk melakukan pemecahan berikutnya. Sedangkan *Partitioning* berarti bahwa ukuran sampel yang dimiliki dipecah kedalam bagian-bagian atau partisi-partisi yang lebih kecil (Larose, 2005).

Algoritma CART mempunyai empat langkah utama. Langkah pertama adalah *tree building process* yaitu proses pembentukan atau pembuatan sebuah pohon klasifikasi. Terdiri dari proses pemecahan simpul induk menjadi dua buah simpul anak melalui aturan pemecahan (*splitting rules*) tertentu dan dilakukan secara berulang serta proses *class assignment* yaitu proses mengidentifikasi simpul-simpul yang terbentuk pada suatu kelas tertentu melalui aturan pengidentifikasian. Langkah kedua adalah proses penghentian pembuatan atau pembentukan pohon klasifikasi (*stopping the trees building process*). Pada tahap ini pohon terakhir atau *maximal tree* (T_{max}) telah terbentuk. Langkah ketiga adalah *pruning the tree* yaitu proses pemangkasan atau pemotongan T_{max} menjadi pohon yang lebih kecil (T) apabila pohon terlalu besar dan banyak. Langkah terakhir adalah interpretasi dan menghitung akurasi pohon.

Proses pemecahan pada masing-masing simpul induk didasarkan pada *goodness of split criterion* (kreteria uji pemecah terbaik). Dalam penelitian ini digunakan *twoing criterion*, yaitu (Lewis, 2000):

$$\Phi(s, t) = \frac{P_L P_R}{4} \left[\sum_j |p(j|t_L) - p(j|t_R)| \right]^2$$

dengan :

t_L = simpul kiri dari simpul t

t_R = simpul kanan dari simpul t

P_L = peluang objek yang berada pada simpul kiri

P_R = peluang objek yang berada pada simpul kanan

$p(j|t_L)$ = peluang bahwa sebuah objek adalah anggota kelas j jika diketahui objek ini berada dalam simpul kiri

$p(j|t_R)$ = peluang bahwa sebuah objek adalah anggota kelas j jika diketahui objek ini berada dalam simpul kanan

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data Susenas 2016 di kota Semarang. Data yang diambil sebanyak 930 rumah tangga. Klasifikasi kemiskinan didasarkan pada pengeluaran per kapita (makanan dan nonmakanan). Jika pengeluaran per kapita selama sebulan kurang dari batas garis kemiskinan maka dikategorikan miskin. Jika ini melebihi dari nilai garis kemiskinan dikatakan tidak miskin. Nilai garis kemiskinan di kota Semarang pada tahun 2016 bernilai Rp 382.160.

Adapun variabel yang digunakan adalah sebagai berikut variabel dependen bernilai kategorik (1= miskin; 0=tidak miskin); X_1 : Jenis Kelamin kepala RT (1=Laki-Laki; 2=perempuan); X_2 : umur kepala RT; X_3 : banyaknya anggota RT; X_4 :Ijazah tertinggi kepala

RT (1: Tidak ada ijazah; 2: SD; 3: SMP; 4:SMA; 5: D1/D2/D3; 6: S1; 7: S2/S3); X₅: Apakah kepala RT bekerja (1: Ya; 2=Tidak); X₆=lapangan usaha (0: Tidak bekerja; 1=pertanian; 2=pertambangan; 3=Industri pengolahan; 4: Listrik dan Gas; 5: bangunan; 6: perdagangan, hotel dan rumah makan; 7: Transportasi dan pergudangan; 8: Keuangan dan asuransi; 9: Jasa); X₇=status pekerjaan (0: Tidak bekerja; 1:berusaha sendiri; 2: berusaha dibantu buruh tidak tetap; 3: berusaha dibantu buruh tetap; 4: Buruh/karyawan/pegawai; 5: pekerja bebas; 6: pekerja keluarga); X₈=status kepemilikan rumah (1: milik sendiri; 2: Kontrak/Sewa; 3: Bebas sewa; 4: Dinas; 5: Lainnya); X₉=bahan bangunan atap (1: Beton; 2: Genteng Keramik; 3: Genteng Metal; 4: Genteng Tanah Liat; 5: Asbes; 6: Seng; 7: Bambu; 8: Kayu/Sirap; 9: Jerami; 10: Lainnya); X₁₀=Bahan Utama dinding Rumah (1: Tembok; 2: kayu; 3: batang kayu; 4: Lainnya); X₁₁=Bahan Utama Lantai Rumah (1: Marmer/Granit; 2: Keramik; 3:Ubin/Tegel; 4: Semen/bata merah; 5: Tanah; 6: Lainnya); X₁₂= Penggunaan Fasilitas tempat buang air besar (1:Ada digunakan hanya ART sendiri; 2: Ada, digunakan bersama dengan ART lainnya; 3: Ada, di MCK Umum; 4: Tidak ada); X₁₃= Tempat pembuangan akhir tinja (1:tangki dengan dasar semen; 2: Tangki tanpa dasar semen; 3: IPAL; 4: Kolam/Sawah/Sungai/Danau; 5: Lubang Tanah; 6:Lainnya); X₁₄=Sumber Air Minum (1: Air kemasan bermerk; 2: Air isi ulang; 3:Leding Meteran; 4: Leding eceran; 5: sumur bor/pompa; 6:Sumur terlindung; 7: Sumur Tak terlindung; 8:Mata air terlindung; 9: Lainnya); ; X₁₅=Bahan bakar utama memasak (0: tidak memasak dirumah; 1:Listrik; 2: elpiji 5.5kg; 3: elpiji 12kg; 4:elpiji 3 kg; 5: minyak tanah; 6: kayu bakar); X₁₆=menerima beras miskin (1: Ya; 2: Tidak).

Berikut disajikan statistika deskriptif untuk masing-masing variabel. Pembentukan pohon klasifikasi menggunakan CHAID dengan analisis sebagai berikut:

1. Penentuan variabel pemisah/pemilah terbaik menggunakan uji Chi-Square. Variabel pemisah pertama adalah dengan nilai Chi-Square paling besar (paling signifikan).
2. Penentuan *Merging* untuk variabel independen yang lebih dari dua kategori dilakukan dengan analisis Chi Square dan dilakukan koreksi bonferoni
3. Tahap penghentian pohon klasifikasi antara lain jumlah sampel minimal setiap simpul adalah 5 dengan taraf signifikansi 5%. Kedalaman pohon adalah maksimal 5

Pembentukan Pohon klasifikasi menggunakan CART sama halnya dengan pembentukan menggunakan CHAId, namun kriteria pemilihannya tidak didasarkan pada statistika uji Chi Square, melainkan menggunakan *Twoing criteria*.

4. HASIL DAN PEMBAHASAN

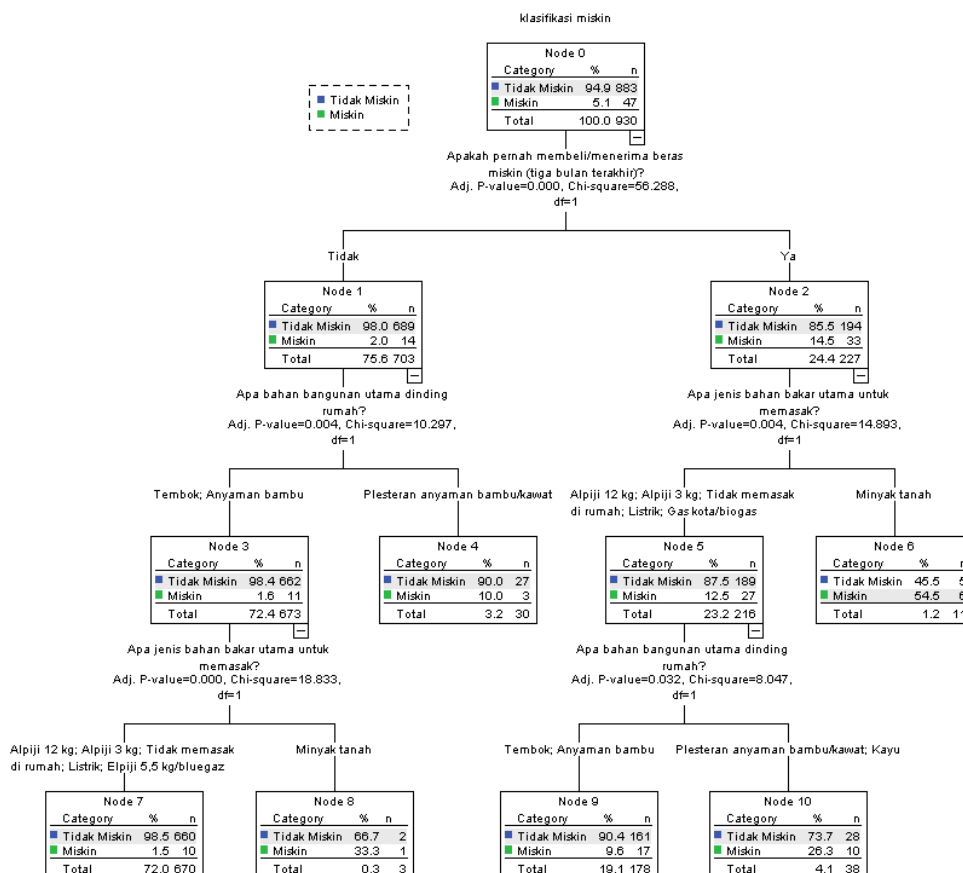
Berdasarkan dari data SUSENAS 2016 Kota Semarang dengan jumlah sampel 930 rumah tangga, didapatkan kategori rumah tangga miskin sebagai berikut:

1. Jumlah anggota rumah tangga rata-rata 4-5 orang
2. Ijazah tertinggi kepala rumah tangga adalah SD/MI dan tidak berijazah
3. Lapangan usaha kebanyakan perdagangan, pertanian dan bangunan
4. Bahan bangunan atap tidak ada yang terbuat dari beton
5. Dinding rumah didominasi juga oleh plasteran/anyaman bambu
6. Tidak ada lantai rumah yang terbuat dari marmer/granit
7. Tempat pembuangan akhir tinja kebanyakan terbuat dari tangki tanpa dasar semen
8. Sumber utama air minum adalah air isi ulang dan air sumur pompa bor, tidak ada yang menggunakan air kemasan bermerk
9. Tidak ada yang menggunakan bahan bakar elpiji 12 kg
10. 70.2% menerima beras miskin

4.1 Klasifikasi Kemiskinan di Kota Semarang menggunakan CHAID

Berdasarkan Gambar 1 dapat diinterpretasikan sebagai berikut:

1. Rumah tangga yang tidak mendapatkan beras miskin dengan bangunan utama dinding berupa plesteran anyaman bamboo/kawat dikategorikan tidak miskin
2. Rumah tangga yang tidak mendapatkan beras miskin, dinding rumah tinggal berupa tembok atau anyaman bamboo, sedangkan bahan bakar utama untuk memasak berupa minyak tanah dikategorikan masuk kedalam rumah tangga tidak miskin
3. Rumah tangga yang tidak mendapatkan beras miskin, dinding rumah tinggal berupa tembok atau anyaman bamboo, sedangkan bahan bakar utama untuk memasak berupa gas elpiji 12 kg, 3 kg, tidak memasak, listrik, blue gas 5.5 kg dikategorikan masuk kedalam rumah tangga tidak miskin
4. Rumah tangga yang menerima beras miskin, dengan bahan bakar untuk memasak berupa minyak tanah dikategorikan kedalam rumah tangga miskin
5. Rumah tangga yang menerima beras miskin dengan bakar untuk memasak berupa elpiji 12 kg, 3 kg, listrik, tidak memasak, biogas dan bahan utama dinding rumah tinggal terbuat dari tembok dan anyaman bamboo dikategorikan kedalam rumah tangga tidak miskin.
6. Rumah tangga yang menerima beras miskin dengan bakar untuk memasak berupa elpiji 12 kg, 3 kg, listrik, tidak memasak, biogas dan bahan utama dinding rumah tinggal terbuat dari plesteran anyaman bamboo atau kayu dikategorikan kedalam rumah tangga tidak miskin.



Gambar 1 Hasil Pohon Klasifikasi Menggunakan CHAID

Berikut disajikan nilai *Gain Node* untuk rumah tangga dengan kategori tidak miskin

Tabel 1 Gain Node untuk Rumah Tangga Tidak Miskin

<i>Node</i>	<i>Node</i>		<i>Gain</i>		<i>Response</i>	<i>Index</i>
	<i>N</i>	<i>Percent</i>	<i>N</i>	<i>Percent</i>		
7	670	72,0%	660	74,7%	98,5%	103,8%
9	178	19,1%	161	18,2%	90,4%	95,3%
4	30	3,2%	27	3,1%	90,0%	94,8%
10	38	4,1%	28	3,2%	73,7%	77,6%
8	3	0,3%	2	0,2%	66,7%	70,2%
6	11	1,2%	5	0,6%	45,5%	47,9%

Berdasarkan Tabel 1 terlihat bahwa, nilai *Gain* yang melebihi 100% terdapat pada *node* 7. Segmen pada *node* 7 adalah Rumah tangga yang tidak mendapatkan beras miskin, dinding rumah tinggal berupa tembok atau anyaman bambu, sedangkan bahan bakar utama untuk memasak berupa gas elpiji 12 kg, 3 kg, tidak memasak, listrik, *blue gas* 5.5 kg dikategorikan masuk kedalam rumah tangga tidak miskin. Nilai kelebihanannya sebesar 3.8%, artinya kategori rumah tangga tidak miskin pada segmen ini nilai 3.8% lebih tinggi daripada kemungkinan status rumah tangga tidak miskin diseluruh sampel.

Tabel 2 Gain Node untuk Rumah Tangga Miskin

<i>Node</i>	<i>Node</i>		<i>Gain</i>		<i>Response</i>	<i>Index</i>
	<i>N</i>	<i>Percent</i>	<i>N</i>	<i>Percent</i>		
6	11	1,2%	6	12,8%	54,5%	1079,3%
8	3	0,3%	1	2,1%	33,3%	659,6%
10	38	4,1%	10	21,3%	26,3%	520,7%
4	30	3,2%	3	6,4%	10,0%	197,9%
9	178	19,1%	17	36,2%	9,6%	189,0%
7	670	72,0%	10	21,3%	1,5%	29,5%

Tabel 2 merupakan *gain node* untuk kategori rumah tangga miskin. Nilai *index* yang melebihi 100% terdapat pada *node* 6, *node* 8, *node* 10, *node* 4 dan *node* 9.

Pada *node* 6 mempunyai segmen adalah Rumah tangga yang menerima beras miskin, dengan bahan bakar untuk memasak berupa minyak tanah dikategorikan kedalam rumah tangga miskin. Artinya segmen pada *node* ini nilainya 921% lebih tinggi daripada kemungkinan status rumah tangga miskin diseluruh sampel. Segmen pada *node* 8 adalah Rumah tangga yang tidak mendapatkan beras miskin, dinding rumah tinggal berupa tembok atau anyaman bambu, sedangkan bahan bakar utama untuk memasak berupa minyak tanah dikategorikan masuk kedalam rumah tangga tidak miskin. Artinya segmen pada *node* ini tingkat kategori rumah tangga miskin 659,6% lebih tinggi daripada kemungkinan status rumah tangga tidak miskin diseluruh sampel. Begitu juga pada *node* 8, *node* 1, *node* 4 dan *node* 9. Nilai kategori miskin pada *node-node* tersebut perlu diperhatikan, karena nilainya diatas rata-rata kategori rumah tangga miskin sampel seluruhnya.

Tabel 3 Hasil Akurasi Klasifikasi Kemiskinan Di Kota Semarang Menggunakan CHAID

<i>Observed</i>	<i>Predicted</i>		
	Tidak Miskin	Miskin	<i>Percent Correct</i>
Tidak Miskin	878	5	99,4%
Miskin	41	6	12,8%
<i>Overall Percentage</i>	98,8%	1,2%	95,1%

4.2. Klasifikasi Kemiskinan di Kota Semarang menggunakan CART

Berdasarkan Lampiran 1 dapat dikaji beberapa hal terkait klasifikasi rumah tangga miskin dan tidak miskin adalah sebagai berikut:

1. Rumah tangga yang menggunakan minyak sebagai bahan bakar utama untuk memasak, 50% dikategorikan miskin dan 50% dikategorikan tidak miskin.
2. Rumah tangga dengan bahan bakar utama untuk memasak gas elpigi 3kg, 12 kg, tidak memasak, listrik, elpigi 5,5 kg *blue gas*, bio gas serta rumah tangga tersebut menerima atau tidak menerima beras miskin dikategorikan kedalam rumah tangga tidak miskin.
3. Rumah tangga dengan bahan bakar utama untuk memasak gas elpigi 3kg, 12 kg, tidak memasak, listrik, elpigi 5,5 kg *blue gas*, bio gas serta rumah tangga tersebut tidak menerima beras miskin dan jumlah anggota keluarganya lebih dari 12 dikategorikan kedalam rumah tangga miskin.
4. Rumah tangga dengan bahan bakar utama untuk memasak gas elpigi 3kg, 12 kg, tidak memasak, listrik, elpigi 5,5 kg *blue gas*, *bio gas* serta rumah tangga tersebut tidak menerima beras miskin dan jumlah anggota keluarganya kurang dari atau sama dengan 12 dikategorikan kedalam rumah tangga tidak miskin.
5. Rumah tangga dengan bahan bakar utama untuk memasak gas elpigi 3kg, 12 kg, tidak memasak, listrik, elpigi 5,5 kg *blue gas*, bio gas serta rumah tangga tersebut menerima beras miskin dan pekerjaan kepala rumah tangga dibidang pertanian padi/palawija serta penggalan atau pertambangan dikategorikan kedalam rumah tangga miskin.
6. Rumah tangga dengan bahan bakar utama untuk memasak gas elpigi 3kg, 12 kg, tidak memasak, listrik, elpigi 5,5 kg *blue gas*, bio gas serta rumah tangga tersebut menerima beras miskin dan pekerjaan kepala rumah tangga selain dibidang pertanian padi/palawija serta penggalan atau pertambangan dikategorikan kedalam rumah tangga tidak miskin.

Berikut hasil tingkat akurasi hasil klasifikasi kemiskinan di kota Semarang menggunakan CART

Tabel 4 Hasil Akurasi Klasifikasi Kemiskinan Di Kota Semarang Menggunakan CART

<i>Observed</i>	<i>Predicted</i>		
	Tidak Miskin	Miskin	<i>Percent Correct</i>
Tidak Miskin	882	1	99,9%
Miskin	44	3	6,4%
<i>Overall Percentage</i>	99,6%	0,4%	95,2%

Tabel 4 menunjukkan hasil akurasi klasifikasi menggunakan CART. Berdasarkan hasil klasifikasi akurasinya mencapai 95,2%. Hasil klasifikasi menggunakan CART lebih bagus dibandingkan dengan menggunakan CHAID Selanjutnya adalah menganalisis variabel independen yang dianggap kurang sesuai dan sangat mempengaruhi klasifikasi kemiskinan di kota Semarang. Variabel yang selalu ada dalam penentuan miskin dan tidak miskin pada metode CART, dan CHAID adalah variabel penerimaan beras miskin. Kategori yang dianggap tidak sesuai adalah menerima beras miskin akan tetapi dikategorikan kedalam

rumah tangga tidak miskin. Oleh karena itu variabel penerimaan beras miskin perlu dikaji lebih lanjut terhadap faktor-faktor yang menjadi penentu dalam penerimaan beras miskin. Selanjutnya variabel penerimaan beras miskin akan menjadi variabel dependen yang akan dianalisis.

5. KESIMPULAN

Berdasarkan hasil analisis yang dilakukan pemodelan menggunakan CART menghasilkan akurasi yang lebih bagus dibandingkan dengan CHAID. Variabel yang menjadi penentu klasifikasi kemiskinan untuk Algoritma CART lebih beragam dibandingkan dengan CHAID. Klasifikasi kemiskinan menggunakan CART menghasilkan pohon klasifikasi yang lebih beragam dibandingkan dengan CHAID. Sedangkan faktor-faktor yang mempengaruhi klasifikasi kemiskinan menggunakan CHAID antara lain penerimaan beras miskin, bahan bangunan utama dinding rumah, dan bahan bakar utama untuk memasak. Sedangkan dengan Algoritma CART variabel yang mempengaruhi adalah bahan bakar utama untuk memasak, penerimaan beras miskin, banyaknya anggota rumah tangga, tempat pembuangan akhir, sumber air minum, lapangan usaha kepala rumah tangga, bahan bangunan atap, bahan bangunan dinding rumah.

UCAPAN TERIMA KASIH

Terima kasih kepada Universitas Diponegoro atas dukungan dana untuk skim penelitian Riset Pengembangan dan Penerapan (RPP) pendanaan tahun 2018 dengan nomor kontrak Nomor: 474-26/UN7.P4.3/PP/2018 Tanggal 2 Juli 2018.

DAFTAR PUSTAKA

- Badan Pusat Statistik. 2014. *Perhitungan Dan Analisis Kemiskinan Makro Indonesia Tahun 2014*. Jakarta: Badan Pusat Statistik
- Breiman, L., Friedman, J.H., Olshen, R.A., dan Stone, C.J. 1984. *Classification And Regression Trees*, New York: Chapman & Hall.
- Wikass, G.V. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29(2), 119-127.
- Kim, H. dan Loh, W.Y. 2001. Classification Trees with Unbiased Multiway Splits. *Am. Statist. Assoc.* 96, 590-604.
- Larose, D. T. 2005. *Discovering Knowledge in Data "An Introduction to Data Mining"*. New Jersey: John Wiley dan Sons, Inc.
- Lewis, R.J. 2000. *An Introduction to Classification And Regression Tree (CART) Analysis*, Annual Meeting of the Society For Academic Emergency Medicine in San Fransisco, California, Department of Emergency Medicine, California.
- Onder, E. dan Uyar, S. 2017. CHAID analysis to determine Socioeconomic Variables that explain student academic success. *Universal Journal of Educational Research*, 4(4), 608-619.
- Rokach, L. dan Maimon, O. 2008. *Data Mining with Decision Trees "Theory and Application"*. USA: World Scientific Publishing Co. Pte. Ltd.

Lampiran: Pohon Klasifikasi dengan Algoritma CART

