

## Klasifikasi Rumah Tangga Miskin Di Kecamatan Kaubun Tahun 2020 Dengan Menggunakan Metode *Improved Chi-Square Automatic Interaction Detection*

### *Classification of Poor Households in Kaubun District in 2020 Using the Improved Chi-Square Automatic Interaction Detection Method*

Pratiwi Dwi Yuliasari<sup>1</sup>, Rito Goejantoro<sup>2</sup>, dan Fidia Deny Tisna Amijaya<sup>3</sup>

<sup>1,2</sup>Laboratorium Statistika Komputasi Program Studi Statistika, FMIPA, Universitas Mulawarman

<sup>3</sup>Laboratorium Matematika Komputasi Program Studi Statistika, FMIPA, Universitas Mulawarman

E-mail: [pratiwidwiuliasari@gmail.com](mailto:pratiwidwiuliasari@gmail.com)

#### Abstract

*Classification was grouping samples based on similarities and differences by using categorical variables. The classification method used in this study was the Improved Chi-square Automatic Interaction Detection (I-CHAID) which was an improvement from the CHAID method. This research aims to provide an overview of poor households and classify poor households, and to compare the accuracy of the classification results for each data proportion. The data used is household data in Kaubun District in 2020 with poor and non-poor status. The results of this study indicate that there were 10 households with poor status, and households were classified as poor if the frequency of eating is less than 3 times a day, and the best classification accuracy results use the proportion of training data of 60% and testing data of 40%.*

**Keywords:** CHAID Method, Classification, I-CHAID Method, Poverty

#### Pendahuluan

Kemiskinan didefinisikan dengan berbagai konsep salah satunya yaitu Bank Dunia mendefinisikan kemiskinan merupakan suatu ketidakmampuan untuk mencapai kehidupan layak dengan pendapatan \$ 1,9 perhari, sedangkan menurut Badan Pusat Statistika (BPS) Indonesia mendefinisikan kemiskinan merupakan ketidakmampuan setiap manusia dalam memenuhi kebutuhan dasar (sandang, pangan, dan papan) dari sisi ekonomi, sehingga dapat disimpulkan bahwa penduduk yang tergolong miskin adalah penduduk yang rata-rata pengeluaran perkapita perbulannya dibawah garis kemiskinan (Kaunang, 2018).

Kemiskinan bukan hanya merupakan permasalahan secara nasional saja, akan tetapi kemiskinan juga menjadi permasalahan di seluruh daerah di Indonesia, salah satunya yaitu Kabupaten Kutai Timur. Berdasarkan data Badan Pusat Statistika (BPS) Kutai Timur, di tahun 2019 persentase masyarakat miskin berada diangka 9,5%, hal ini disebabkan karena tingginya jumlah pendatang yang masuk ke wilayah Kutai Timur untuk mencari pekerjaan, sehingga pendatang yang datang dalam kondisi tidak memiliki pekerjaan dan rumah dikategorikan masyarakat miskin (BPS Kutim, 2019). Kecamatan Kaubun merupakan salah satu wilayah pemukiman Transmigrasi di Kabupaten Kutai Timur dan salah satu daerah yang menjadi tujuan pendatang baru, juga merupakan Kecamatan baru hasil pemekaran dari Kecamatan Kaliurang pada tahun 2005 dan dengan jumlah rumah tangga terbanyak keempat

untuk wilayah pedesaan di Kabupaten Kutai Timur (Kecamatan Kaubun, 2010).

Klasifikasi merupakan pengelompokan sampel berdasarkan ciri-ciri persamaan dan perbedaan dengan menggunakan variabel yang berupa kategorik. Metode klasifikasi yang paling banyak digunakan adalah metode pohon klasifikasi (Faisal dkk, 2017).

Pohon klasifikasi merupakan salah satu metode yang digunakan untuk mengklasifikasikan data yang menggunakan aturan untuk menentukan kelas dari suatu objek yang memiliki nilai-nilai variabel independen, dan menggunakan representasi struktur pohon (*tree*) di mana setiap *node* mempresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas, dan *node* yang paling atas dari pohon klasifikasi disebut sebagai *root* dan pohon klasifikasi sendiri mempunyai banyak algoritma salah satunya adalah *Chi-square Automatic Interaction Detection* (CHAID) (Tohari dkk, 2017).

Metode *Chi-square Automatic Interaction Detection* (CHAID) merupakan metode untuk menyelidiki hubungan antara variabel dependen dan variabel independen yang digunakan untuk klasifikasi (Wirania dalam Damayanti dkk, 2018), sehingga metode ini sering dikenal sebagai metode pohon klasifikasi (*Classification Tree Method*). Pada metode CHAID digunakan uji independensi *Chi-square* yang merupakan uji nonparametrik karena sesuai untuk menguji hubungan variabel yang berbentuk kategorik dan lebih efektif diterapkan pada data dengan pengamatan yang banyak, sehingga metode ini

memiliki kelemahan yaitu menghasilkan bias dalam memilih variabel independen yang dapat mengubah interpretasi data pada pohon klasifikasi sehingga diperlukan adanya pengembangan dari metode CHAID yang disebut dengan metode *Improved Chi-square Automatic Interaction Detection* (I-CHAID) (Damayanti dkk, 2018).

Metode I-CHAID merupakan perbaikan dari metode CHAID dengan mengkolaborasi nilai *Chi-square* dengan *Tschuprow's T*. Kemudian metode I-CHAID sendiri menghasilkan pohon klasifikasi yang menyediakan informasi tentang derajat hubungan antara variabel dependen dan variabel independen dan juga menyediakan informasi karakteristik segmen sehingga hasil segmentasinya akan lebih baik (Damayanti dkk, 2018).

Terdapat beberapa penelitian yang membahas tentang I-CHAID diantaranya Muhajir (2016) melakukan penelitian Analisis Kredit Macet BMT (Baitul Mal Wa Tawil) dengan menggunakan metode I-CHAID, Ispriyanti dan kawan-kawan (2019) juga melakukan penelitian Klasifikasi Kemiskinan di Kota Semarang Menggunakan Algoritma *Chi-Square Automatic Interaction Detection* (CHAID) dan *Classification and Regression Tree* (CART), Damayanti dan kawan-kawan (2018) melakukan penelitian Perbandingan Hasil Pembentukan Pohon Klasifikasi Metode CHAID dan Metode *Improved CHAID*

Berdasarkan diuraikan diatas, maka penulis tertarik untuk melakukan penelitian dengan judul "Klasifikasi Rumah Tangga Miskin di Kecamatan Kaubun Tahun 2020 dengan Menggunakan Metode *Improved Chi-square Automatic Interaction Detection*".

**Data Mining**

*Data mining* merupakan suatu istilah yang dipakai untuk menguraikan penemuan pengetahuan didalam *database* dan merupakan salah satu bentuk dari pengolahan suatu data, data didefinisikan merupakan catatan atas kumpulan fakta dan diperoleh, disimpan, diolah, dipakai dan sebagainya dan *datamining* juga dapat diartikan sebagai proses yang menggunakan teknik statistika, matematika, kecerdasan buatan, dan *machine learning* untuk menguraikan, mengidentifikasi informasi yang bermanfaat, dan pengetahuan yang terkait dari berbagai *database* besar (Nengsih, 2019).

Menurut Hasket (2000), *data mining* memiliki fungsi-fungsi yang dapat dibagi kedalam tujuh kelompok, yaitu sebagai berikut:

1. Hubungan (*association*) merupakan proses untuk menemukan aturan asosiasi antara suatu kombinasi item dalam suatu waktu.
2. *Sequence* merupakan suatu proses untuk menemukan aturan asosiasi antara suatu

kombinasi item dalam suatu waktu dan diterapkan lebih dari satu periode.

3. *Clustering* merupakan proses pengelompokan sejumlah data atau obyek kedalam kelompok data sehingga setiap kelompok berisi data yang mirip.
4. Klasifikasi (*classification*) merupakan proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.
5. *Regression* merupakan proses pemetaan data dalam suatu nilai prediksi.
6. *Forecasting* merupakan proses pengestimasian nilai prediksi berdasarkan pola-pola didalam sekumpulan data.
7. *Solution* merupakan proses penemuan akar masalah dan *problem solving* dari persoalan bisnis yang dihadapkan atau paling tidak sebagai informasi dalam pengambilan keputusan.

(Vulandari, 2017)

**Klasifikasi**

Klasifikasi merupakan pengelompokan yang sistematis pada sejumlah objek, gagasan, buku atau benda-benda lain kedalam kelas atau golongan tertentu berdasarkan ciri-ciri yang sama (Wijaya, 2018). Klasifikasi memiliki teknik untuk mendapatkan data baru dengan memanipulasi data yang ada yang telah diklasifikasikan dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan yang digunakan pada data-data baru untuk diklasifikasikan, teknik ini menggunakan *supervised induction* yang memanfaatkan kumpulan pengujian dari data set yang terklasifikasi.

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*).

Matriks konfusi adalah tabel pencatat hasil kerja dari klasifikasi, dengan tabel matriks konfusi sebagai berikut:

**Tabel 1. Confusion Matrix**

Kelas asli (i)	Kelas hasil prediksi (j)	
	Kelas = 0	Kelas = 1
Kelas = 0	$f_{00}$	$f_{01}$
Kelas = 1	$f_{10}$	$f_{11}$

Dimana setiap sel  $f_{ij}$  dalam matriks menyatakan jumlah data dari kelas  $i$  yang hasil

prediksinya masuk ke kelas  $j$ , kemudian berdasarkan isi matriks konfusi, dapat diketahui jumlah data dari masing-masing kelas yang diprediksi secara benar yaitu  $(f_{00} + f_{11})$ , dan data yang diklasifikasikan secara salah, yaitu  $(f_{01} + f_{10})$ . Kuantitas matriks konfusi dapat diringkas menjadi dua nilai, yaitu akurasi dan laju eror, dengan mengetahui jumlah data yang diklasifikasi secara benar sehingga dapat diketahui akurasi hasil prediksi, dan dengan mengetahui jumlah data yang diklasifikasikan secara salah dapat diketahui laju eror dari prediksi yang dilakukan. Untuk menghitung keakuratan klasifikasi, berikut persamaannya:

$$\text{Akurasi} = \frac{\text{banyak prediksi benar}}{\text{total banyak prediksi}} \times 100\% \quad (1)$$

$$\text{Akurasi} = \frac{f_{00} + f_{11}}{f_{00} + f_{01} + f_{10} + f_{11}} \times 100\% \quad (2)$$

(Prasetyo, 2012)

### Pohon Keputusan

Pohon keputusan (*Decision Tree*) yaitu pohon klasifikasi yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari suatu permasalahan dan dalam pohon keputusan daerah yang digunakan untuk pengambilan keputusan yang sebelumnya kompleks dapat diubah menjadi lebih sederhana (Prasetyo, 2012).

Pohon keputusan terdapat 3 jenis node, yaitu:

1. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
2. *Internal node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
3. *Leaf node* atau *terminal node*, merupakan *node* akhir pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

(Zaman, 2016)

### Metode Improved Chi-Square Automatic Interaction Detection (I-CHAID)

CHAID berasal dari singkatan *Chi-square Automatic Interaction Detection* dan pertama kali diperkenalkan oleh G.V Kass tahun 1980 dengan judul "An Exploratory Technical For Investigating Large Quantities Of Categorical Data", CHAID merupakan sebuah metode yang digunakan untuk menyelidiki keterkaitan antara variabel dependen dan variabel independen yang digunakan dalam klasifikasi, dan metode CHAID secara umum dikenal sebagai metode pohon klasifikasi (*Classification Tree Method*) (Dalam Damayanti dkk, 2018).

Metode CHAID menggunakan uji independensi *Chi-square* yang merupakan uji nonparametrik karena ujinya sesuai untuk menguji hubungan antara variabel yang berbentuk kategori (Myers dalam Kunto dan Siti 2006), dan membagi data menjadi kelompok-kelompok berdasarkan

keterkaitan antara variabel dependen dengan variabel independen dan dapat digunakan pada data berbentuk kategorik dan lebih efektif diterapkan pada data dengan pengamatan yang banyak. Kelemahan dari metode CHAID adalah menghasilkan bias dalam memilih variabel independen yang dapat mengubah interpretasi data pada pohon klasifikasi sehingga perlu adanya pengembangan dari metode CHAID yang disebut dengan I-CHAID (Faridhan dkk dalam Damayanti dkk 2018).

Metode I-CHAID merupakan perbaikan dari metode CHAID dengan mengkolaborasikan uji *Chi-square* kedalam uji *Tschuprow's* dan I-CHAID menghasilkan pohon klasifikasi yang menyediakan informasi tentang derajat hubungan antara variabel dependen dan variabel independen serta informasi mengenai karakteristik segmen sehingga hasil segmentasinya akan lebih baik. Pohon klasifikasi yang dihasilkan merupakan *output* grafis sehingga membuat metode I-CHAID mudah untuk diinterpretasikan karena bisa dilihat langsung pola pemisahan dan penggabungan variabelnya. Tahap analisis yang dilakukan dalam metode I-CHAID hampir sama dengan tahap pada analisis yang ada didalam metode CHAID, perbedaannya terletak pada tahap kedua yakni setelah uji *Chi-square* lalu dilanjutkan dengan uji *Tschuprow's T* kemudian koreksi Bonferroni, dengan demikian pembentukan hasil segmentasi dan klasifikasi menjadi lebih baik (Muhajir, 2016).

### Algoritma Improved Chi-Square Automatic Interaction Detection (I-CHAID)

Kass (1980) dalam Muhajir (2016) menyatakan langkah analisis I-CHAID yaitu sebagai berikut:

1. Penggabungan (*Merging*) yaitu proses pemeriksaan signifikansi dari tiap-tiap kategori variabel independen terhadap variabel dependen.
2. Pemisahan (*Splitting*) yaitu memilih variabel independen yang digunakan sebagai pemisah *node* (*split node*) yang paling baik.
3. Penghentian (*Stopping*) yaitu penghentian proses pertumbuhan pohon klasifikasi.

### Uji Chi-square

Uji *Chi-square* adalah uji yang digunakan untuk membandingkan antara fakta yang diperoleh berdasarkan hasil observasi atau yang disebut juga dengan hasil kenyataan dan fakta yang didasarkan secara teoritis atau yang diharapkan (Damayanti, 2018).

Hipotesis

$H_0$  : Tidak terdapat hubungan antara variabel dependen dan variabel independen.

$H_1$  : Terdapat hubungan antara variabel dependen dan variabel independen.

Statistik Uji

1. Kontingensi berukuran  $2 \times 2$

$$\chi_{hitung}^2 = \frac{N(O_{11} \times O_{22} - O_{12} \times O_{21})^2}{n_{.1} \times n_{.2} \times n_{1.} \times n_{2.}} \quad (3)$$

2. Kontingensi berukuran  $B \times K$

$$\chi_{hitung}^2 = \sum_{ij=1}^N \left[ \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right] \quad (4)$$

dimana:

$$E_{ij} = \left( \frac{n_{i.} \times n_{.j}}{N} \right) \quad (5)$$

dimana:

$O_{ij}$  = jumlah pengamatan baris ke- $i$  dan kolom ke- $j$

$E_{ij}$  = nilai harapan pada baris ke- $i$  dan kolom ke- $j$

$n_{i.}$  = nilai total pada kolom ke- $j$

$n_{.j}$  = nilai total pada baris ke- $i$

$N$  = nilai total seluruh pengamatan

$i$  = jumlah baris

$j$  = jumlah kolom

Daerah Kritis

Menolak  $H_0$  jika  $\chi_{hitung}^2 > \chi_{\alpha; (B-1)(K-1)}^2$   
atau jika  $p\text{-value} < \alpha$

### Uji *Tschuprow's*

Menurut Bergsma (2013) uji *Tschuprow's* merupakan uji yang digunakan untuk melihat hubungan antara dua variabel nominal dengan koefisien estimasi sederhana, koefisien estimasi sederhana merupakan uji *Chi-square* ( $\chi^2$ ) dan nilai  $\chi^2$  tidak terbatas yang terletak pada interval  $[0, +\infty)$ , sehingga memiliki kekurangan yang menyebabkan terjadinya penekanan tinggi dalam mendiskripsikan analisis data, solusi untuk mengatasi kekurangan tersebut adalah dengan menormalisasikan nilai  $\chi^2$  yang berada pada interval  $[0,1]$  sehingga diperoleh persamaan berikut:

$$T = \frac{\chi^2}{W\sqrt{(V-1) \times (U-1)}} \quad (6)$$

Dimana:

$V$ = Jumlah kategori variabel independen

$U$ = Jumlah kategori variabel dependen

$W$ = Banyak sampel

Uji *Tschuprow's* dalam metode I-CHAID memiliki tujuan untuk menentukan variabel yang dapat digunakan dalam pembentukan pohon klasifikasi, uji *Tschuprow's* dapat dihubungkan dengan uji chi-square hitung sehingga dapat mempermudah dalam melakukan perhitungan dan menghasilkan segmentasi yang lebih baik (Damayanti dkk, 2018).

Pengkategorian kuat lemah hubungan antar variabel secara umum dapat mengikuti pengelompokan sebagai berikut:

**Tabel 2.** Pengkategorian Hubungan Antar Variabel

Besarnya Nilai Hubungan (Koefisien Korelasi)	Interpretasi Hubungan
$0,80 \leq KK \leq 1,00$	Tinggi
$0,60 \leq KK < 0,80$	Cukup
$0,40 \leq KK < 0,60$	Agak Rendah
$0,20 \leq KK < 0,40$	Rendah
$0,00 \leq KK < 0,20$	Sangat Rendah

(Cahyono, 2018)

### Koreksi Bonferroni

Menurut Sharp (2002) dalam Damayanti dkk (2018) koreksi Bonferroni merupakan suatu proses koreksi yang digunakan ketika uji statistik untuk menguji independensi atau menguji dependensi dilakukan secara bersamaan. Koreksi Bonferroni dilakukan setelah mendapatkan variabel independen yang signifikan dan variabel independen yang telah digabungkan serta dihitung nilai Bonferroninya berdasarkan bentuk variabel independennya, variabel independen yang digabungkan memiliki  $c$  kategori dan dikurangi  $g$  kategori setelah penggabungan, sehingga perhitungan Bonferroni dapat disimpulkan merupakan banyaknya cara yang mungkin dari  $c$  kategori yang dapat digabungkan menjadi  $g$  kategori. Perhitungan Koreksi Bonferroni untuk masing-masing bentuk variabel independen, yaitu sebagai berikut:

- a. Variabel independen monotonik

$$M = \binom{c-1}{g-1} = \frac{(c-1)!}{(g-1)!(c-1)-(g-1)!} \quad (7)$$

- b. Variabel independen *free* (bebas)

$$M = \sum_{i=0}^{g-1} (-1)^i \frac{(g-1)^c}{i!(g-i)!} \quad (8)$$

- c. Variabel independen *floating* (mengambang)

$$M = \binom{c-2}{g-2} + g \binom{c-2}{g-1} \quad (9)$$

Persamaan (7), (8), dan (9) digunakan untuk mencari nilai dari koreksi Bonferroni pada masing-masing variabel independen dan setelah mendapat nilai koreksi Bonferroni pada variabel independen maka dilanjutkan ke persamaan berikut ini:

$$1 - (1 - \alpha)^M > \alpha \quad (10)$$

Dimana:

$\alpha$  = Comparison wise error rate (CWER)

$M$  = Nilai koreksi Bonferroni

**Kemiskinan**

Kemiskinan memiliki banyak definisi dan konsep, sehingga kemiskinan merupakan masalah yang bersifat multidimensi yang dapat ditinjau dari berbagai sudut pandang, Bank Dunia (2000) membagi dimensi kemiskinan kedalam empat hal pokok, yaitu *lack of opportunity*, *low capabilities*, *low level security*, dan *low capacity*, kemudian kemiskinan dapat dikaitkan juga dengan keterbatasan hak-hak sosial, ekonomi dan politik sehingga menyebabkan kerentanan, keterpurukan, dan ketidakberdayaan, sehingga mendefinisikan kemiskinan merupakan kehilangan kesejahteraan (*deprivation of well being*) dengan inti permasalahan kemiskinan adalah batasan-batasan tentang kesejahteraan itu sendiri, kemudian dalam teori ekonomi, kemiskinan dilihat dari segi tingkat kesejahteraan dimana tingginya tingkat kesejahteraan diartikan sebagai kemampuan untuk mengakses sumber daya yang tersedia, sehingga kemiskinan dapat diartikan ketidakmampuan dalam memenuhi kebutuhan hidup (dalam Usman dkk, 2006).

**Indikator Rumah Tangga Miskin**

Badan Pusat Statistik (BPS) pada tahun 2005 melakukan pendataan untuk penargetan Program Bantuan Langsung Tunai dengan berpedoman pada Instruksi Presiden Nomor 12 Tahun 2005, sistem pendataan ini disebut pendataan Sosial-Ekonomi Penduduk Tahun 2005, atau lebih dikenal sebagai PSE05. Dalam PSE05, sebuah rumah tangga dikategorikan miskin apabila:

1. Luas lantai bangunan tempat tinggal kurang dari 8 m<sup>2</sup> per orang.
2. Lantai bangunan tempat tinggalnya terbuat dari tanah/ bambu/ kayu murah.
3. Dinding bangunan tempat tinggalnya terbuat dari bambu/ rumbia/ kayu berkualitas rendah/ tembok tanpa diples ter.
4. Tidak memiliki fasilitas buang air besar atau bersama-sama rumah tangga lainnya menggunakan satu jamban.
5. Sumber penerangan rumah tangga tidak menggunakan listrik.
6. Air minum berasal dari sumber atau mata air tidak terlindung seperti sungai atau air hujan.
7. Bahan bakar untuk memasak sehari-hari adalah kayu bakar/ arang/ minyak tanah.
8. Hanya mengkonsumsi daging, ayam, susu satu kali dalam satu minggu.
9. Hanya membeli satu pasang pakaian dalam satu tahun.
10. Hanya mampu makan satu atau dua kali dalam satu hari.
11. Tidak sanggup membayar biaya pengobatan di puskesmas/ poliklinik.
12. Sumber penghasilan kepala keluarga adalah: petani dengan luas lahan 0,5 ha, buruh tani,

nelayan, buruh bangunan, buruh perkebunan, atau pekerjaan lainnya dengan pendapatan dibawah Rp 600.000 per bulan.

13. Pendidikan terakhir kepala keluarga yaitu tidak sekolah, tidak tamat sekolah dasar (SD), hanya tamatan SD.
14. Tidak memiliki tabungan/ barang yang mudah dijual dengan nilai minimal Rp 600.000 misalkan sepeda motor (kredit/non kredit), emas, hewan ternak, kapal, motor atau barang modal lainnya.

(Isdijoso dkk, 2016)

**Variabel Penelitian**

Variabel dalam penelitian ini meliputi variabel dependen dan variabel independen. Variabel dependen dalam penelitian ini adalah status rumah tangga dan variabel independen dalam penelitian ini adalah sebagai berikut:

**Tabel 3.** Variabel Penelitian

Variabel	Definisi Variabel	Kategori
Y	Status Rumah Tangga	1. Miskin 2. Tidak Miskin
X <sub>1</sub>	Luas lantai tempat tinggal	1. < 8 m <sup>2</sup> (per orang) 2. > 8 m <sup>2</sup> (per orang)
X <sub>2</sub>	Jenis lantai tempat tinggal	1. Tanah/bambu/kayu murah 2. Selain tanah/bambu/kayu murah
X <sub>3</sub>	Jenis dinding tempat tinggal	1. Bambu/rumbia/kayu kualitas rendah/ tembok tanpa plester 2. Selain bambu/rumbia/kayu kualitas rendah/tembok tanpa plester
X <sub>4</sub>	Sumber penerangan tempat tinggal	1. Bukan listrik 2. Listrik
X <sub>5</sub>	Sumber air bersih	1. Air hujan/ sumur tidak terlindungi 2. Ledeng/ sumur terlindungi/ PAM
X <sub>6</sub>	Kepemilikan Jamban/WC	1. Tidak Memiliki/ Bersama-sama 2. Sendiri
X <sub>7</sub>	Bahan bakar memasak	1. Kayu Bakar 2. Gas
X <sub>8</sub>	Mengkonsumsi daging/ ayam/ susu dalam satu minggu	1. Tidak pernah mengkonsumsi/ hanya satu kali 2. Pernah mengkonsumsi/ lebih dari satu kali
X <sub>9</sub>	Kemampuan membeli pakaian baru dalam satu tahun	1. Tidak pernah membeli/ hanya membeli satu pasang pakaian 2. Pernah membeli/ lebih dari satu pasang pakaian

$X_{10}$	Frekuensi makan dalam dalam satu hari	1. < 3 kali sehari 2. $\geq 3$ kali sehari
$X_{11}$	Tempat pengobatan	1. Pengobatan sendiri 2. Rumah Sakit/ Puskesmas/ Poliklinik

**Tabel 3.** Variabel Penelitian (Lanjutan)

Variabel	Definisi Variabel	Kategori
$X_{12}$	Penghasilan kepala rumah tangga	1. Penghasilan < Rp 600.000 per bulan 2. Penghasilan $\geq$ Rp 600.000 per bulan
$X_{13}$	Pendidikan tertinggi kepala rumah tangga	1. Tidak sekolah/ tamat SD 2. Tamat (SMP/SMA/Perguruan Tinggi) sederajat
$X_{14}$	Kepemilikan aset/tabungan	1. Tidak Memiliki aset/tabungan 2. Memiliki aset/tabungan

**Metode Penelitian**

Adapun langkah-langkah dalam menganalisis menggunakan metode I-CHAID dengan menggunakan  $\alpha$  sebesar 0,05 adalah sebagai berikut:

- a. Memisahkan data set menjadi *training data* dan *testing data*.

Pada tahap ini data set dibagi menjadi dua antara lain *training data* yaitu data yang akan dijadikan *input* dalam proses untuk membentuk pohon klasifikasi untuk menemukan aturan klasifikasi, dan *testing data* yaitu data yang akan digunakan untuk menguji aturan klasifikasi yang telah terbentuk. Berikut kombinasi proporsi pembagian *training data* dan *testing data*, ditampilkan pada Tabel 4 sebagai berikut:

<i>Training Data</i>	<i>Testing Data</i>
50%	50%
60%	40%
70%	30%
80%	20%
90%	10%

- b. Melakukan Proses Analisis I-CHAID  
Kass (1980) dalam Muhajir (2016) menyatakan langkah analisis I-CHAID dengan *Tschuprow's T* yaitu sebagai berikut:

1) Penggabungan (*Merging*)

Pada tahap pertama adalah proses pemeriksaan signifikansi dari tiap-tiap kategori

variabel independen terhadap variabel dependennya. Tahap penggabungan secara detail untuk setiap variabel independen dalam menggabungkan kategori-kategori tidak signifikan seperti berikut:

- a. Membuat tabel kontingensi dua arah untuk masing-masing variabel independen terhadap variabel dependennya.
- b. Langkah selanjutnya menghitung statistik *chi-square* untuk setiap pasangan kategori yang dapat dipilih untuk digabung menjadi satu, kemudian menghitung statistik *Tschuprow's* untuk setiap pasang kategori yang dapat dipilih untuk digabung menjadi satu, untuk menguji kebebasannya dalam sebuah sub tabel kontingensi  $2 \times 2$  atau  $2 \times J$  yang menunjukkan bahwa jumlah kelompok adalah sama persis dengan jumlah nilai kelas.
- c. Menghitung *p-value* berpasangan bersamaan, untuk masing-masing nilai *chi-square* berpasangan. Gabungkan sebuah pasangan kategori yang paling mirip yaitu pasangan yang mempunyai nilai *chi-square* berpasangan terkecil dan *p-value* terbesar menjadi sebuah kategori tunggal, diantara pasangan-pasangan yang tidak signifikan dan selanjutnya diteruskan kelangkah berikutnya. Misalkan apabila terdapat pasangan nilai *p-value* lebih besar dari taraf signifikansi, maka pasangan tersebut digabungkan. Misal pasangan kategori  $X_1$  dan  $X_2$  tidak signifikan, maka pasangan tersebut digabungkan menjadi satu variabel baru  $X_{1,2}$ .
- d. Membuat kategori baru yang terbentuk setelah digabungkan dengan kategori lainnya dalam variabel independen diperiksa kembali kesignifikansiannya dan kekuatan hubungannya. Apabila terdapat pasangan yang belum signifikan, maka langkah 3 diulang lagi. Apabila ternyata sudah semua signifikan lanjut kelangkah selanjutnya.
- e. Menghitung *p-value* terkoreksi Bonferroni didasarkan pada tabel yang telah digabung.

2) Pemisahan (*Splitting*)

Pada tahap pemisahan akan memilih variabel independen yang akan digunakan sebagai pemisah *node* (*split node*) yang paling baik. Proses pemilihan dikerjakan dengan cara membandingkan *p-value* pada tiap-tiap variabel independen. Adapun langkah-langkah pemisahan sebagai berikut:

- a. Langkah yang pertama adalah memilih variabel independen yang mempunyai *p-value* paling kecil (nilai *chi-square* hitung paling besar) atau paling signifikan dan nilai *Tschuprow's* paling besar atau hubungan antar variabel independen dengan variabel dependennya paling kuat yang nantinya digunakan sebagai *split node*.

b. Apabila nilai *p-value* kurang dari sama dengan taraf signifikansi, maka *split node* menggunakan variabel independen ini. Apabila tidak terdapat variabel independen dengan *p-value* paling kecil (nilai *chi-square* hitung paling besar) atau paling signifikan dan nilai *Tschuprow's* paling besar atau hubungan antar variabel independen dengan variabel dependennya paling kuat, maka tidak perlu dilakukan *split* dan *node* ditentukan sebagai terminal *node* (akhir *node*).

3) Penghentian (*Stopping*)

Pembentukan sub kelompok berikutnya, dengan mengulangi langkah penggabungan. Pada tahap penghentian dilakukan penghentian proses pertumbuhan dengan peraturan sebagai berikut:

- a. Variabel independen yang signifikan menunjukkan perbedaan terhadap variabel dependen sudah tidak ada lagi.
- b. Proses pertumbuhan dapat terhenti ketika pohon sekarang mencapai batas nilai maksimum pohon dari spesifikasi. Contoh peneliti menetapkan batas kedalaman pertumbuhan pohon klasifikasi adalah 3, maka pertumbuhan pohon klasifikasi terhenti ketika pertumbuhan pohon sudah mencapai kedalaman 3.
- c. *Node* tidak dapat dipisahkan apabila ukuran dari *child node* kurang dari nilai ukuran *child node* minimum spesifikasi, atau berisi pengamatan-pengamatan dengan banyak yang terlalu sedikit. Contoh ditetapkan ukuran minimal *child node* adalah 50, maka *node* tersebut tidak akan dipecah ketika pemisahan menghasilkan ukuran *child node* kurang dari 50.

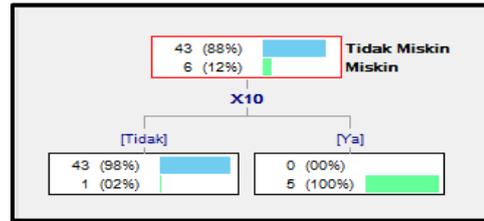
Semua proses didalam metode I-CHAID menggunakan bantuan *software Sipina 3.12*.

**Hasil dan Pembahasan Analisis I-CHAID**

Telah diketahui bahwa dalam penelitian ini menggunakan 14 variabel independen dan menggunakan 2 kategori, sehingga tidak terjadi tahap penggabungan. Proses metode I-CHAID dapat di klasifikasikan dengan diagram pohon klasifikasi I-CHAID. Berikut hasil dari klasifikasi menggunakan metode I-CHAID:

1. *Training Data 50%*

Berikut pohon klasifikasi yang terbentuk berdasarkan *training data 50%*, yang dapat dilihat pada Gambar 1.



**Gambar 1.** Pohon Klasifikasi Analisis I-CHAID *Training Data 50%*

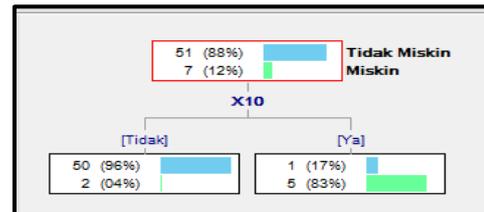
Berdasarkan pohon keputusan yang terbentuk pada Gambar 1 maka segmentasi status rumah tangga di Kecamatan Kaubun tahun 2020 terdapat pada Tabel 5. Berdasarkan Gambar 1 dapat diketahui pada segmen 1 rumah tangga yang memiliki status rumah tangga miskin dengan frekuensi makan 3 kali atau lebih dalam satu sebesar 2% dan status rumah tangga tidak miskin dengan frekuensi makan 3 kali atau lebih dalam satu hari sebesar 98%, dan segmen 2 rumah tanggayang memiliki status rumah tangga miskin dengan frekuensi makan kurang dari 3 kali dalam satu hari sebesar 100%.

**Tabel 5.** Segmen Status Rumah Tangga dengan I-CHAID *Training Data 50%*

Segmen	Keterangan
Segmen ke-1	Status rumah tangga berdasarkan frekuensi makan kurang dari 3 kali dalam satu hari.
Segmen ke-2	Status rumah tangga berdasarkan frekuensi makan 3 kali atau lebih dalam satu hari.

2. *Training Data 60%*

Berikut pohon klasifikasi yang terbentuk berdasarkan *training data 60%*, yang dapat dilihat pada Gambar 2.



**Gambar 2** Pohon Klasifikasi Analisis I-CHAID *Training Data 60%*

Berdasarkan pohon keputusan yang terbentuk pada Gambar 2 maka segmentasi status rumah tangga di Kecamatan Kaubun tahun 2020, sebagai berikut.

**Tabel 6.** Segmen Status Rumah Tangga dengan I-CHAID *Training Data 60%*

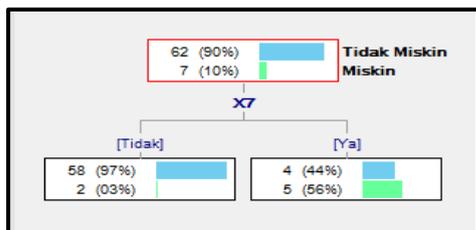
Segmen	Keterangan
Segmen ke-1	Status rumah tangga berdasarkan frekuensi makan kurang dari 3 kali dalam satu hari.
Segmen ke-2	Status rumah tangga berdasarkan frekuensi makan 3 kali atau lebih dalam satu hari.

Berdasarkan Gambar 2 dapat diketahui pada segmen 1 rumah tangga miskin dengan frekuensi makan 3 kali atau lebih dalam satu hari sebesar 4% dan rumah tangga tidak miskin dengan frekuensi makan 3 kali atau lebih dalam satu hari

sebesar 96%, dan segmen 2 rumah tangga miskin dengan frekuensi makan kurang dari 3 kali dalam satu hari sebesar 83% dan rumah tangga tidak miskin dengan frekuensi makan kurang dari 3 kali dalam satu hari sebesar 17%.

3. Training Data 70%

Berikut pohon klasifikasi yang terbentuk berdasarkan *training data* 70%, yang dapat dilihat pada Gambar 3. Berdasarkan pohon keputusan yang terbentuk pada Gambar 3 maka segmentasi status rumah tangga di Kecamatan Kaubun tahun 2020 pada Tabel 7.



Gambar 3 Pohon Klasifikasi Analisis I-CHAID Training Data 70%

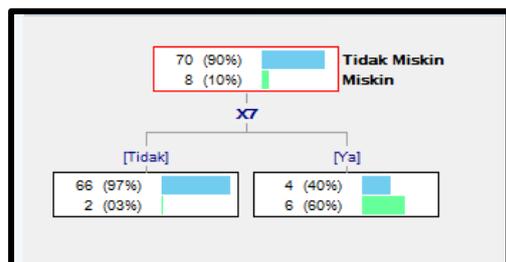
Tabel 7. Segmen Status Rumah Tangga dengan I-CHAID Training Data 70%

Segmen	Keterangan
Segmen ke-1	Status rumah tangga berdasarkan bahan bakar memasak menggunakan kayu bakar.
Segmen ke-2	Status rumah tangga berdasarkan bahan bakar memasak menggunakan gas.

Berdasarkan Gambar 3 dapat diketahui pada segmen 1 rumah tangga miskin dengan bahan bakar memasak menggunakan gas sebesar 3% dan rumah tangga tidak miskin dengan bahan bakar memasak menggunakan gas sebesar 97%, dan segmen 2 rumah tangga miskin dengan bahan bakar memasak menggunakan kayu bakar sebesar 56% dan status rumah tangga tidak miskin dengan bahan bakar memasak menggunakan kayu bakar sebesar 44%.

4. Training Data 80%

Berikut pohon klasifikasi yang terbentuk berdasarkan *training data* 80%, yang dapat dilihat pada Gambar 4.



Gambar 4. Pohon Klasifikasi Analisis I-CHAID Training Data 80%

Berdasarkan pohon keputusan yang terbentuk pada Gambar 4 maka segmentasi status rumah

tangga di Kecamatan Kaubun tahun 2020, sebagai berikut.

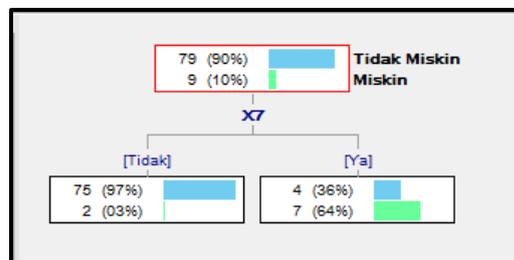
Tabel 8. Segmen Status Rumah Tangga Dengan I-CHAID untuk Training Data 80%

Segmen	Keterangan
Segmen ke-1	Status rumah tangga berdasarkan bahan bakar memasak menggunakan gas.
Segmen ke-2	Status rumah tangga berdasarkan bahan bakar memasak menggunakan kayu bakar.

Berdasarkan Gambar 4 dapat diketahui pada segmen 1 rumah tangga miskin dengan bahan bakar memasak menggunakan gas sebesar 3% dan rumah tangga tidak miskin dengan bahan bakar memasak menggunakan gas sebesar 97%, dan segmen 2 rumah tangga miskin dengan bahan bakar memasak menggunakan kayu bakar sebesar 60% dan rumah tangga tidak miskin dengan bahan bakar memasak menggunakan kayu bakar sebesar 40%.

5. Training Data 90%

Berikut pohon klasifikasi yang terbentuk berdasarkan *training data* 90%, yang dapat dilihat pada Gambar 5.



Gambar 5. Pohon Klasifikasi Analisis I-CHAID Training Data 90%

Berdasarkan pohon keputusan yang terbentuk pada Gambar 5 maka segmentasi status rumah tangga di Kecamatan Kaubun tahun 2020, sebagai berikut.

Tabel 9. Segmen Status Rumah Tangga dengan I-CHAID Training Data 90%

Segmen	Keterangan
Segmen ke-1	Status rumah tangga berdasarkan bahan bakar memasak menggunakan gas.
Segmen ke-2	Status rumah tangga berdasarkan bahan bakar memasak menggunakan kayu bakar.

Berdasarkan Gambar 5 dapat diketahui pada segmen 1 rumah tangga miskin dengan bahan bakar memasak menggunakan gas sebesar 3% dan rumah tangga tidak miskin dengan bahan bakar memasak menggunakan gas sebesar 97%, dan segmen 2 rumah tangga miskin dengan bahan bakar memasak menggunakan kayu bakar sebesar 64% dan rumah tangga tidak miskin dengan bahan bakar memasak menggunakan kayu bakar sebesar 36%.

Ketepatan Klasifikasi

Berikut ketepatan klasifikasi untuk setiap *testing data*, yang dapat dilihat pada tabel 10, sebagai berikut:

**Tabel 10.** Ketepatan Klasifikasi Untuk Setiap *Testing Data*

<i>Testing Data</i>	Ketepatan Klasifikasi	Jumlah Rumah Tangga Tepat Klasifikasi
50%	96%	94
40%	100%	98
30%	86%	84
20%	90%	88
10%	80%	78

### Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang dilakukan, diperoleh kesimpulan sebagai berikut:

1. Rumah tangga dengan status rumah tangga miskin di Kecamatan Kaubun dari 98 jumlah rumah tangga terdapat 10 rumah tangga dengan status rumah tangga miskin.
2. Hasil analisis klasifikasi pada status rumah tangga miskin di Kecamatan Kaubun menggunakan metode *Improved Chi-square Automatic Interaction Detection* yaitu didapatkan rumah tangga miskin diklasifikasikan berdasarkan frekuensi makan kurang dari 3 kali dalam satu hari.
3. Hasil ketepatan klasifikasi menunjukkan bahwa proporsi terbaik untuk mengklasifikasikan status rumah tangga miskin di Kecamatan Kaubun menggunakan *training data* 60% dan *testing data* 40%.

### Daftar Pustaka

- Badan Pusat Statistik. 2016. *Penghitungan dan Analisis Kemiskinan Makro Indonesia 2016*. Jakarta: BPS.
- Bergsma, Wicher. 2013. A bias-correction for Cramér's V and Tschuprow's T. *Journal of the Korean Statistical Society*, 323-328.
- Chayono, Tri. 2018. *Statistika Terapan dan Indikator Kesehatan*. Yogyakarta: CV. Budi Utama.
- Damayanti, Chici, Dadan Kusnandar, dan Yudhi. 2018. Perbandingan Hasil Pembentukan Pohon Klasifikasi Metode CHAID dan Improved CHAID. *Buletin Ilmiah Mat, Stat, dan Terapan (Bimaster)*, 379-384.
- Faisal, Muhammad, Yuki Novia Nasution, dan Fidia Deny Tisna Amijaya. 2017. Perbandingan Kinerja Metode Klasifikasi Chi-square Automatic Interaction Detection (CHAID) dengan Metode Klasifikasi Algoritma C4.5 pada Studi Kasus : Penderita Diabetes Melitus Tipe 2 Di Samarinda Tahun 2015. *Eksponensial*, 119-124.
- Isdijoso, Widjajanti, Asep Suryahadi, dan Akhmadi. 2016. *Kertas Kerja SMERU Penetapan Kriteria dan Variabel Pendapatan Penduduk Miskin yang Komprehensif dalam Rangka Perlindungan Penduduk Miskin di Kabupate/Kota*. The SMERU Research Institute.
- Ispriyanti, Dwi, Alan Prahutama, dan Mustafid. 2019. Klasifikasi Kemiskinan di Kota Semarang Menggunakan Algoritma *chi-square Automatic Interaction Detection* (CHAID) dan *Classification and Regression Tree* (CART). *Media Statistika*, 63-72.
- Kass, G.V. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 119-127.
- Kaunang, Fergie Joanda. 2018. Penerapan Algoritma J48 Decisoan Tree Untuk Analisis Tingkat Kemiskinan di Indonesia. *Cogito Smart Journal*, 348-354.
- Kecamatan Kaubun. 2010. *Sekilas Tentang Kecamatan Kaubun*. <https://Kecamatankaubun.wordpress.com/2010/08/02/sekilas-tentang-Kecamatan-kaubun/>. Diakses akses: 15 Agustus 2020.
- Kunto, Yohanes Sondang, dan Siti Nurul Hasanah. 2006. Analisis CHAID Sebagai Alat Bantu Statistika Untuk Segmentasi Pasar (Studi Kasus Pada Koperasi Syari'ah Al-Hidayah). *Manajemen Pemasaran*, 88-98.
- Muhajir, Muhammad. 2016. Metode Improved CHAID (Chi-square Automatic Interaction Detection) pada Analisis Kredit Macet BMT (Baitul Mal Wa Tamwil). *Ilmu-Ilmu MIPA*, 55-63.
- Nengsih, Warnia. 2019. *Implementasi Data Mining dalam Kasus Terapan*. Riau: Politeknik Catex Riau.
- Prasetyo, Eko. 2012. *Data Mining : Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: CV. Andi Offset.
- Sharp, A, Romaniuk J, dan Cierpicki S. 2002. The Performance of Segmentation Variables: A Comparative Study. *Senior Research Asssociates*, 2398-2411.
- Tohari, Anas, Yuliana susanti, dan Tri Atmojo Kusmayadi. 2017. Penerapan Metode CHAID (*Chi-square Automatic Interaction Detection*) dan Exhaustive CHAID pada Klasifikasi Produksi Jagung di Pulau Jawa. *Universitas Sebelas Maret*, 01-08.
- Usman, Bonar M. Sinaga, dan Hermanto Siregar. 2006. Analisis Determinan Kemiskinan Sebelum dan Sesudah Desentralisasi Fiskal. *Sosial Ekonomi Pertanian*, 01-17.
- Vulandari, Retno Tri. 2017. *Data Mining Teori dan Aplikasi Rapidminer*. Surakarta: Penerbit Gava Media.

- Wijaya, Apriani Candra, Nelly Astuti Hasibuan, dan Putri Ramdhani. 2018. Implementasi Algoritma C5.0 dalam Klasifikasi Pendapatan Masyarakat pada Studi Kasus: Kelurahan Mesjid Kecamatan Medan Kota. *Majalah Ilmiah INTI*, 192–198.
- Wirania, Yustia, M. Novitasari Mara, dan Dadan Kusnandar. 2013. Pembentukan pohon klasifikasi dengan metode CHAID. *Buletin Ilmiah Mat, Stat, dan Terapan (Bimaster)*, 45-50.
- Zaman, Khairul. 2016. Penerapan Data Mining Menggunakan Algoritma C4.5 untuk Menentukan Kelayakan Penerima Bantuan Rehabilitasi Sosial Rumah Tidak Layak Huni (Studi Kasus: Di Pemerintahan Kabupaten Solok Selatan). *Komunikasi Teknologi Informatika*, 12–24.