

# Characteristics of Drinking Water Conditions According to Urban and Rural Areas in Indonesia Using the CHAID Method

Aulia Wanda, Yenni Kurniawati\*

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

\*Corresponding author: yennikurniawati@fmipa.unp.ac.id

Submitted : 31 Januari 2024

Revised : 12 Februari 2024

Accepted : 16 Februari 2024

## ABSTRACT

*Drinking water is the basic needs of society instead of food, clothing and shelter. Decent and clean drinking water is protected drinking water, including tap water, public taps, public hydrants, water terminals, rainwater reservoirs, or protected springs and wells, drilled wells/pumps with the closest distance being 10 meters from the location of waste disposal, waste storage and rubbish disposal. Proper drinking water must be good in physical characteristics and free from contamination of harmful elements. In order to meet the requirements of safety in terms of health for drinking water, this can be done by treating the water before consumption. Access to safe drinking water for Indonesians in urban and rural areas is quite different. This situation causes the treatment of drinking water in urban and rural areas to be different. Therefore, to determine the condition of drinking water in Indonesia based on the treatment given to drinking water by urban and rural areas, Chi-Square Automatic Interaction Detection (CHAID) analysis was used. CHAID analysis was used on categorical type variables. Prior to the analysis stage, there was a data mining process to gain knowledge from the data clusters and handle missing data in the data clusters. Handling missing data on categorical variables was done by mode imputation. Using the CHAID method, household drinking water was generally characterized by treating water before drinking more in rural areas than urban areas.*

**Keywords:** *drinking water, CHAID, data mining, missing data.*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

## I. PENDAHULUAN

Air adalah zat yang mempunyai peranan sangat besar bagi keberlangsungan hidup setiap makhluk. Tubuh manusia terdiri dari air sebanyak tiga perempat bagian. Jika tidak minum selama empat sampai lima hari, manusia tidak akan dapat bertahan hidup (Wandriavel dkk, 2012). Menurut Bambang dkk (2022) air minum merupakan kebutuhan pokok masyarakat disamping kebutuhan pangan, sandang, dan papan. Ketersediaan dan kualitas dari air minum perlu diperhatikan, baik dari segi jumlah serta kelayakannya yang harus memenuhi syarat.

Seluruh masyarakat di dunia mendapatkan hak untuk memperoleh air yang layak dan bersih. Air dinyatakan oleh Perserikatan Bangsa-Bangsa (PBB) termasuk ke dalam hak asasi manusia (Arsyina dkk, 2019). Terkait air juga masuk dalam salah satu tujuan dari pembangunan berkelanjutan atau (*Sustainable Development Goals/SDG's*) yang diagendakan oleh 193 negara anggota PBB dengan harapan harus tuntas di tahun 2030. Tujuannya yaitu memastikan bahwa seluruh masyarakat memperoleh air yang bersih serta sanitasinya layak (Nurzanah dkk, 2020). Dengan begitu, pemerataan terhadap akses air yang bersih dan layak dapat dirasakan oleh setiap lapisan masyarakat di dunia.

Air minum yang layak berasal dari air yang bersih. Menggunakan air bersih sebagai minum dapat mengurangi penyakit seperti diare, kolera, disentri, tipes, cacingan, penyakit kulit, dan keracunan (Kementerian Kesehatan Republik Indonesia, 2020). Air minum dikatakan aman untuk kesehatan jika telah memenuhi syarat dari segi fisik, mikrobiologis, kimia, dan radioaktif (Kementerian Kesehatan Republik Indonesia, 2018). Air minum layak dan bersih adalah air minum dengan sumber terlindung, diantaranya air leding (keran), keran umum, *hydrant* umum, terminal air, penampungan air hujan (PAH), atau mata air dan sumur terlindung, sumur bor/pompa dengan jarak terdekat adalah sepuluh meter dari lokasi pembuangan kotoran, penampungan limbah, dan pembuangan sampah (Badan Pusat Statistik, 2017). Oleh karena itu, selain memperhatikan dari segi aman menurut kesehatan tetapi juga perlu memperhatikan dari mana asal sumber air minum tersebut.

Kondisi akses air minum di dunia saat ini belum semua masyarakat yang memperoleh layanan air minum dengan baik dan aman. *World Health Organization* (WHO) dan *United Nations Children's Fund* (UNICEF) (2021) menyatakan bahwa pada tahun 2020 ada sekitar 2 miliar masyarakat dunia tidak memperoleh layanan air untuk minum yang dikelola secara aman, termasuk di dalamnya 1,2 miliar orang dengan layanan dasar, 282 juta orang dengan layanan yang terbatas, dan 367 juta orang mengonsumsi sumber air minum tidak layak, serta 122 juta orang minum air permukaan langsung. Di Indonesia, pada tahun 2020, sebanyak 90,2% rumah tangga yang mengakses sumber air minum dari sumber yang layak berhasil dicapai melalui upaya yang dilakukan pemerintah terhadap air minum aman selama bertahun-tahun, tetapi ada perbedaan persentase pada akses di perkotaan dengan pedesaan (Badan Pusat Statistik, 2020). Kemudian, jika dilihat pada tahun 2017, rumah tangga yang mengakses air minum layak hanya sebesar 72% dengan persentase di perkotaan 80,8% dan di pedesaan 62% (Badan Pusat Statistik, 2023b). Hal ini menunjukkan perbedaan yang cukup jauh baik dari tahun 2017 dengan 2020 dan juga antara wilayah pedesaan dengan perkotaan.

Perbedaan akses air minum di perkotaan dan pedesaan terbukti pada penelitian terdahulu. Berdasarkan hasil penelitian dari Rahim dan Muchlisoh menunjukkan bahwa rumah tangga yang tempat tinggalnya adalah di wilayah perkotaan kecenderungannya lebih tinggi untuk menggunakan sumber air layak jika dibandingkan dengan wilayah pedesaan (Rahim dan Muchlisoh, 2020). Selain air minum yang layak, perlu diperhatikan bahwa air minum tersebut aman yang telah memenuhi syarat kesehatan. Air minum yang layak harus baik secara ciri-ciri fisik dan terbebas dari kontaminasi unsur-unsur yang berbahaya. Cara yang dilakukan masyarakat untuk membuat air minum memenuhi syarat aman dari segi kesehatan adalah memberikan perlakuan terhadap air minum sebelum dikonsumsi. Sebab rumah tangga yang mempunyai akses untuk air, bisa saja air tersebut mempunyai kualitas mikrobiologis yang tidak tentu (Badan Pusat Statistik, 2023a). Masyarakat dapat melakukan pengolahan air minum dengan berbagai cara, diantaranya adalah dengan merebus, menyaring dengan kain, memberi bahan kimia, disinfeksi dengan matahari, filtrasi keramik, dan lain sebagainya (Herlambang, 2010). Sebagaimana dalam penelitian Lestari dan Indriani mengenai identifikasi pola kepemilikan sumber air minum rumah tangga di Jawa Timur, mengungkapkan bahwa sebagian besar rumah tangga di daerah perkotaan sudah mempunyai sumber air minum yang layak. Ini disebabkan oleh akses sumber air minum yang aman jauh lebih mudah diakses di perkotaan daripada pedesaan. Namun, rumah tangga di perkotaan tidak mengolah air minum untuk dikonsumsi (Lestari dan Indriani, 2022). Sehingga, perbedaan dalam pengaksesan air minum di perkotaan dan pedesaan menunjukkan perlakuan dalam untuk air minum di masyarakat juga berbeda. Dengan memberikan perlakuan terhadap air, membuat air minum lebih aman untuk dikonsumsi dan memenuhi syarat kesehatan.

Pada penelitian ini, dilakukan proses *data mining* serta penanganan *missing data*, kemudian dilakukan analisis *Chi-Square Automatic Interaction Detection* (CHAID) guna mengetahui bagaimana kondisi air minum masyarakat di Indonesia dengan perlakuan yang diberi pada air minum berdasarkan jenis tempat tinggal (perkotaan atau pedesaan). Menurut Miftahuddin (2012) metode CHAID (*Chi-Square Automatic Interaction Detection*) merupakan bagian dari *Automatic Interaction Detection* (AID) yang dibuat untuk melakukan penyelidikan terhadap struktur kaitan antar variabel respon kategorik dengan variabel penduga kategorik. Metode CHAID termasuk ke dalam teknik eksplorasi secara non parametrik yang mampu menganalisis data yang besar ukurannya, dan dapat membagi data jadi kelompok yang saling bebas serta mampu menjelaskan variabel responnya. Metode CHAID ini cukup tepat dalam melakukan pendugaan terhadap variabel penduga dan hubungan yang signifikan terhadap variabel respon serta bagaimana hubungan antara variabel penduga tersebut.

Penelitian terdahulu yang menggunakan metode CHAID diantaranya adalah penelitian yang dilakukan oleh Amalita dkk tentang karakteristik penerima beasiswa bidikmisi di FMIPA UNP. Dari penelitian tersebut diperoleh 5 kelompok karakteristik mahasiswa penerima beasiswa bidikmisi FMIPA UNP untuk tahun 2016 dengan karakteristik mahasiswa yang menjadi penerima beasiswa bidikmisi adalah pekerjaan ayah sebagai petani dan lainnya serta mempunyai luas rumah kurang atau sama dengan 50 m<sup>2</sup> (Amalita dkk, 2019). Penelitian lain dilakukan oleh Suhendra dkk tentang ketepatan klasifikasi pemberian Kartu Keluarga Sejahtera (KKS) di Kota Semarang menggunakan metode regresi logistik biner dan metode CHAID. Diperoleh faktor-faktor yang mempengaruhi pemberian KKS menggunakan metode logistik biner adalah variabel jumlah anggota keluarga dan pendidikan kepala keluarga, sedangkan menggunakan CHAID variabel yang berpengaruh yaitu jumlah anggota keluarga, status perkawinan, usia kepala keluarga, pendidikan kepala keluarga dan kepemilikan/penguasaan HP. Dari tingkat akurasi diperoleh akurasi yang paling tinggi adalah menggunakan metode CHAID sebesar 90,2% sedangkan logistik biner hanya 88% (Suhendra dkk, 2020). Dari hasil ini menunjukkan bahwa metode CHAID lebih baik daripada metode regresi logistik biner.

Oleh karena itu, pada penelitian ini digunakan metode CHAID yang bertujuan untuk mengelompokkan karakteristik air minum di wilayah perkotaan maupun pedesaan di Indonesia dengan melihat bagaimana perlakuan yang diberi pada air. Data yang digunakan adalah data Survei Demografi Kesehatan Indonesia (SDKI) tahun 2017.

Adapun variabelnya yaitu jenis tempat tinggal, letak sumber air, air diolah dengan direbus, air diolah dengan diberi pemutih/khlor, dan air diolah dengan disaring dengan kain.

## II. METODE PENELITIAN

### A. Sumber Data dan Variabel Penelitian

Penelitian ini memakai data sekunder yaitu SDKI 2017. Sumbernya berasal dari Survei Demografi dan Kesehatan Indonesia yang dilakukan pada tahun 2017 oleh Badan Pusat Statistik (BPS) bersama Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN), dan Kementerian Kesehatan (Kemenkes). Pada Tabel 1 dapat dilihat variabel yang digunakan.

**Tabel 1.** Variabel beserta penjelasan

Variabel	Tipe Data	Penjelasan	Kategori
Y: Jenis tempat tinggal	Kategorik	menjelaskan tentang jenis tempat tinggal dari rumah tangga yang diwawancarai dalam survei.	Perkotaan: (1) Pedesaan: (2)
X <sub>1</sub> : Letak sumber air	Kategorik	menjelaskan lokasi keberadaan dari sumber air yang dimiliki oleh rumah tangga.	Di rumah sendiri: (1) Di halaman sendiri: (2) Tempat lain: (3)
X <sub>2</sub> : Air diolah dengan direbus	Kategorik	menjelaskan tentang pengolahan air minum secara direbus.	Tidak: (0) Ya: (1) Tidak Tahu: (8)
X <sub>3</sub> : Air diolah dengan diberi pemutih/khlor	Kategorik	menjelaskan tentang pengolahan air minum dengan diberi pemutih/khlor.	Tidak: (0) Ya: (1) Tidak Tahu: (8)
X <sub>4</sub> : Air disaring dengan kain	Kategorik	menjelaskan tentang pengolahan air minum dengan cara disaring menggunakan kain.	Tidak: (0) Ya: (1) Tidak Tahu: (8)

### B. Teknik Analisis Data

Teknik analisis data yang digunakan adalah metode CHAID dengan melakukan proses *data mining* sebelum masuk ke tahapan analisis. Adapun *software* yang digunakan yaitu *R-Studio*. Berikut adalah langkah-langkahnya.

- Memilih data yang akan digunakan untuk analisis.
- Melakukan proses *data mining* untuk identifikasi *missing data* dan penanganan menggunakan imputasi modus.  
Proses *data mining* merupakan kegiatan yang dilakukan terhadap suatu gugus data dimana dilakukan pencarian secara berulang serta intensif dengan tujuannya yaitu memperoleh pengetahuan yang terdapat pada gugus data tersebut yang sebelumnya tidak berarti penting (Pramana dkk, 2023). *Missing data* adalah permasalahan yang terjadi pada suatu gugus data yang menyebabkan beberapa bagian dari data hilang atau tidak lengkap. Penyebabnya adalah terdapat responden yang menolak untuk bekerja sama secara langsung atau menolak menjawab beberapa pertanyaan dengan alasan tertentu, kesalahan dalam penginputan hasil wawancara dan kesalahan teknis lainnya (Longford, 2005). *Missing data* memiliki tiga mekanisme, yaitu *Missing Completely at Random* (MCAR), *Missing at Random* (MAR), dan *Missing not at Random* (MNAR) (Little dan Rubin, 2020). Penanganan untuk *missing data* bisa dilakukan dengan metode imputasi. Metode imputasi misalnya dengan rata-rata atau nilai modus, sesuai dengan jenis dari data. Jika kategorik, dapat menggantinya dengan nilai modus (Hendrawati, 2015). Oleh karena itu, pada penanganan *missing data* dari penelitian ini digunakan imputasi modus karena jenis datanya adalah kategorik.
- Melakukan analisis dengan metode CHAID.

Pada metode CHAID, menurut Gallagher (2000) dalam Suhendra dkk (2020) variabel penduga kategorik akan dibedakan menjadi tiga bentuk. Pertama yaitu variabel monotonik, dimana kategori yang terdapat pada variabel penduga bisa digabung dengan syarat keduanya berdekatan atau sesuai urutan sebenarnya (data ordinal). Kemudian, yang kedua yaitu variabel bebas, dimana kategori dari variabel penduga bisa digabung dalam kondisi berdekatan atau tidak (data nominal). Ketiga adalah variabel mengambang, dimana kategori pada variabel dalam variabel penduga dapat diberi perlakuan seperti monotonik, tetapi tidak bisa untuk kategori *missing value*, yang bisa dikombinasikan dengan kategori lainnya.

Terdapat tiga tahapan dalam metode CHAID menurut Bagozzi (1994) dalam Putri dkk (2023). Tahapan pertama yaitu penggabungan, dimana dilakukan pemeriksaan signifikansi dari setiap kategori variabel penduga dengan variabel respon. Pada tahapan ini dilakukan perhitungan statistik *chi-square*. Kedua adalah pemisahan yaitu menentukan variabel penduga untuk *split node* yang tepat, dengan melihat perbandingan nilai *p-value* (pada tahap sebelumnya) untuk masing-masing dari variabel respon. Ketiga yaitu penghentian, dimana pada tahapan ini dilakukan

penghentian apabila pada proses pembentukan pohonnya, tidak terdapat lagi variabel penduga yang signifikan untuk melihat bedanya dengan variabel respon.

Melakukan uji *chi-square* bisa menggunakan langkah berikut.

a. Membuat hipotesis dengan ketentuan:

$H_0$ : Kedua kriteria klasifikasi bebas (independen)

$H_1$ : Kedua kriteria klasifikasi tidak bebas (dependen)

b. Menghitung statistik uji *chi-square* dengan rumus yang terdapat pada persamaan (1).

$$X_{hitung}^2 = \sum_{i=1}^r \sum_{j=1}^c \left[ \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \right] \quad (1)$$

Dengan,

$$E_{ij} = \frac{n_{i.}n_{.j}}{n} \quad \text{untuk } i = 1, 2, \dots, r \text{ dan } j = 1, 2, \dots, c$$

Keterangan:

$n_{ij}$ : banyak observasi pada baris ke- $i$  dan kolom ke- $j$

$E_{ij}$ : frekuensi harapan pada baris ke- $i$  dan kolom ke- $j$

c. Menentukan kriteria penolakan  $H_0$ , jika  $X_{hitung}^2 > X_{\alpha:(r-1)(c-1)}^2$

d. Menginterpretasikan pohon klasifikasi dan mengecek keakurasian dari model yang terbentuk.

### III. HASIL DAN PEMBAHASAN

#### A. Proses Data Mining

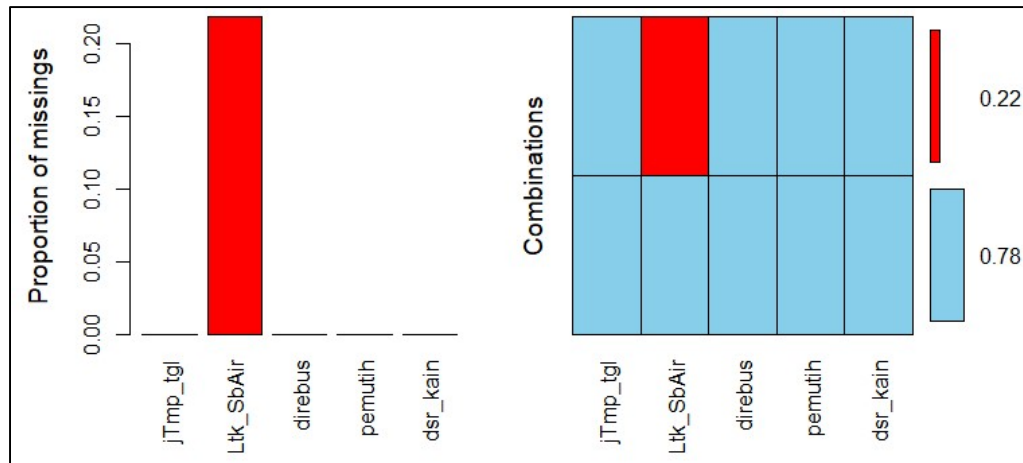
Proses yang dilakukan pada data mining dimulai dari *import* data, memilih variabel dan membuat *data frame* baru, dilanjutkan dengan mengubah tipe variabel yang semula terbaca oleh *software R-studio* berupa numerik dan diubah menjadi kategorik. Setelah itu masuk ke tahapan identifikasi *missing data*. Dalam pengidentifikasian ini dilakukan dengan melihat hasil *summary*, metode tabulasi, penggunaan fungsi *aggr*, dan *matrixplot*. Berikut dapat dilihat ringkasan statistik sebelum dilakukan proses *data mining* yang terdapat pada Tabel 2.

Tabel 2. Ringkasan statistik sebelum proses *data mining*

Jenis tempat tinggal	Letak sumber air	Air diolah dengan direbus	Air diolah dengan diberi pemutih	Air disaring dengan kain
1 : 101076	1 : 53124	0 : 63371	0 : 196346	0 : 182593
2 : 96647	2 : 49439	1 : 134121	1 : 1146	1 : 14899
	3 : 51794	8 : 61	8 : 61	8 : 61
	9 : 148	9 : 170	9 : 170	9 : 170
	NA : 43218			

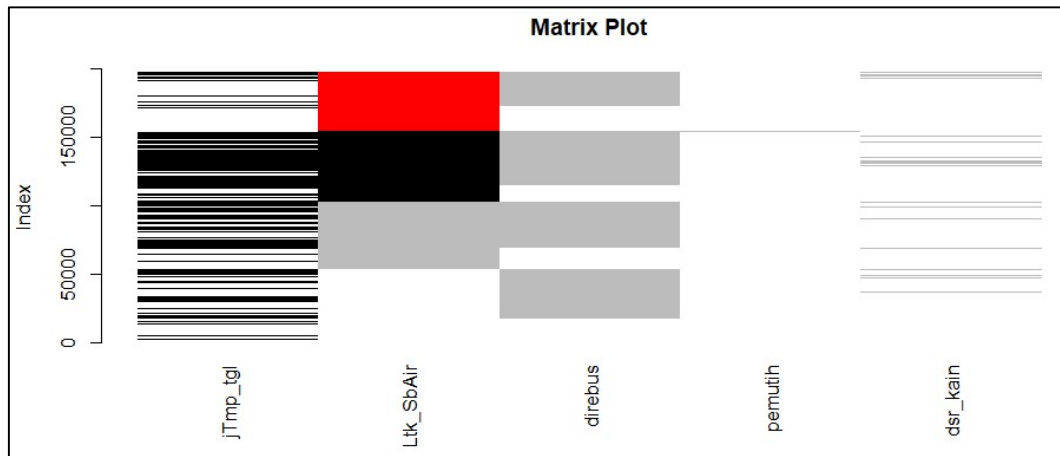
Berdasarkan Tabel 2, terlihat bahwasannya terdapat *missing data* yang ditandai dengan *Not Available* (NA) sebanyak 43218 amatan pada variabel letak sumber air. Kemudian, terlihat ada data yang berkategori 9 pada variabel letak sumber air, air diolah dengan direbus, air diolah dengan diberi pemutih, dan air disaring dengan kain. Pada kasus seperti ini disebut dengan data *noise* atau data yang tidak masuk ke dalam kategori. Untuk kasus data *noise*, penanganan yang dilakukan adalah menghapus amatan yang mengandung data *noise* tersebut.

Selanjutnya adalah mengidentifikasi *missing data* berupa proporsi yang ada pada Gambar 1. Terlihat bahwasannya proporsi *missing data* yang ada pada variabel letak sumber air adalah sebanyak 22%. Dengan begitu, penanganan untuk data hilang ini dapat dilakukan dengan imputasi modus karena proporsi *missing data* tidak terlalu tinggi. Akan tetapi, dalam menanganani *missing data*, perlu juga dilakukan pengidentifikasian mekanisme dari *missing data* yang terdapat pada variabel letak sumber air. Pengidentifikasian ini dilakukan dengan menggunakan *matrixplot* yang dapat dilihat pada Gambar 2.



Gambar 1. Identifikasi missing data dengan fungsi aggr dalam bentuk proporsi

Matrixplot pada Gambar 2 menunjukkan bahwa missing data terdapat di variabel letak sumber air yang ditandai dengan bagian yang berwarna merah. Pada matrixplot tersebut mengindikasikan bahwa variabel letak sumber air mengalami missing data tidak dipengaruhi oleh variabel lainnya. Hal ini dikarenakan pada bagian yang sejajar dengan mengalami missing data, amatannya tidak menunjukkan warna hitam yang pekat. Oleh karena itu, dapat dikatakan mekanisme missing data pada variabel ini adalah karena variabel itu sendiri atau dikenal dengan nama Missing not at Random (MNAR).



Gambar 2. Matrixplot diurutkan dari variabel letak sumber air

Setelah dilakukan penanganan, dilihat lagi hasil summary-nya. Berikut dapat dilihat pada Tabel 3 bahwasannya sudah tidak ada lagi missing data dan noise-nya. Baru dilanjutkan ke tahapan visualisasi dan analisis menggunakan metode CHAID.

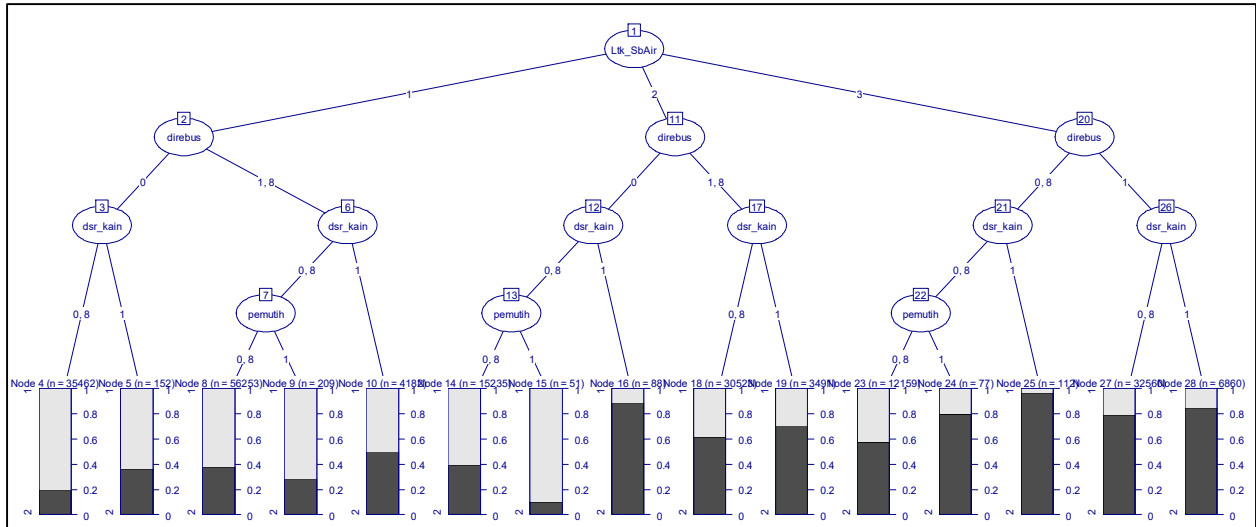
Tabel 3. Ringkasan statistik setelah dilakukan proses data mining

Jenis tempat tinggal	Letak sumber air	Air diolah dengan direbus	Air diolah dengan diberi pemutih	Air disaring dengan kain
1 : 100907	1 : 96258	0 : 63328	0 : 196207	0 : 182468
2 : 96507	2 : 49388	1 : 134025	1 : 1146	1 : 14885
	3 : 51768	8 : 61	8 : 61	8 : 61

Berdasarkan Tabel 3, terlihat tidak ada lagi missing data maupun data noise, sehingga dapat dilanjutkan ke tahapan analisis menggunakan metode CHAID.

**B. Analisis dengan Metode CHAID**

Hasil analisis menggunakan metode CHAID adalah berupa pohon klasifikasi. Dari analisis yang telah dilakukan pada *software* R-Studio, terbentuklah pohon klasifikasi yang dapat dilihat pada Gambar 3. Pohon klasifikasi tersebut menunjukkan karakteristik air minum dari tiap kelompok yang ada disertai dengan keterangan mana yang dominan antara wilayah pedesaan atau perkotaan terhadap karakteristik tersebut dari *node* atau cabang yang terbentuk.



**Gambar 3.** Pohon klasifikasi dari model CHAID

Berdasarkan pohon klasifikasi pada Gambar 3 untuk lebih jelas mengetahui karakteristik air minum menurut pedesaan dan perkotaan di Indonesia dapat dilihat pada Tabel 4.

**Tabel 4.** Karakteristik air minum dan persentase dari dugaan jenis tempat tinggal

Kelompok	Karakteristik	Dugaan jenis tempat tinggal		Perlakuan terhadap air minum
		Pedesaan	Perkotaan	
1	Rumah tangga dengan karakteristik air minum tidak disaring menggunakan kain dan tidak direbus dengan sumber air berada di rumah sendiri.	20%	80%	Tidak ada
2	Rumah tangga dengan karakteristik air minum disaring menggunakan kain dan tidak direbus dengan sumber air berada di rumah sendiri.	38%	62%	Ada
3	Rumah tangga dengan karakteristik air minum tidak diberi pemutih/khlor, tidak disaring menggunakan kain, tetapi direbus dengan lokasi sumber air di tempat lain.	39%	61%	Ada
4	Rumah tangga dengan karakteristik air diberi pemutih/khlor, tidak disaring menggunakan kain, tetapi direbus dan tidak tahu direbus dengan sumber air berada di rumah sendiri.	23%	77%	Ada
5	Rumah tangga dengan karakteristik air minum yang disaring dengan kain, direbus dan tidak tahu direbus dengan sumber air berada di rumah sendiri.	50%	50%	Ada
6	Rumah tangga dengan karakteristik air minum tidak diberi pemutih/khlor, tidak disaring menggunakan kain, dan tidak direbus dengan sumber air berada di halaman sendiri.	40%	60%	Tidak ada
7	Rumah tangga dengan karakteristik air minum diberi pemutih/khlor, tidak disaring menggunakan kain, dan tidak direbus dengan sumber air berada di halaman sendiri.	10%	90%	Ada



Kelompok	Karakteristik	Dugaan jenis tempat tinggal		Perlakuan terhadap air minum
		Pedesaan	Perkotaan	
8	Rumah tangga dengan karakteristik air minum disaring menggunakan kain tetapi tidak direbus dengan sumber air berada di halaman sendiri.	88%	12%	Ada
9	Rumah tangga dengan karakteristik air minum tidak disaring menggunakan kain, tetapi direbus dengan sumber air berada di halaman sendiri.	61%	39%	Ada
10	Rumah tangga dengan karakteristik air minum disaring menggunakan kain, direbus dan tidak tahu kalau direbus, dengan sumber air berada di halaman sendiri.	72%	28%	Ada
11	Rumah tangga dengan karakteristik air minum diberi pemutih/khlor, tidak disaring dengan kain, dan tidak direbus dengan sumber air berada di tempat lain.	58%	42%	Ada
12	Rumah tangga dengan karakteristik air diberi pemutih/khlor, tidak disaring menggunakan kain, dan tidak direbus dengan sumber air berada di tempat lain.	80%	20%	Ada
13	Rumah tangga dengan karakteristik air minum disaring menggunakan kain dan tidak direbus dengan sumber air berada di tempat lain.	98%	2%	Ada
14	Rumah tangga dengan karakteristik air minum tidak disaring menggunakan kain tetapi direbus, dengan sumber air berada di tempat lain.	80%	20%	Ada
15	Rumah tangga dengan karakteristik air minum disaring menggunakan kain dan direbus dengan sumber air berada di tempat lain.	85%	25%	Ada

Pada Gambar 3 dan Tabel 4, menunjukkan bahwasannya terdapat 15 kelompok untuk karakteristik air minum menurut perkotaan dan pedesaan di Indonesia. Kelompok yang memiliki perlakuan terhadap air minum adalah kelompok 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, dan 15. Sementara itu, kelompok yang tidak ada perlakuan terhadap air minum adalah kelompok 1 dan 6. Karakteristik dari air minum yang tidak ada perlakuan sama sekali lebih banyak terdapat pada rumah tangga di wilayah perkotaan dengan persentase kelompok 1 adalah 80% (letak sumber air berada di rumah sendiri) dan kelompok 6 sebanyak 60% (letak sumber air berada di halaman sendiri).

Air minum yang diberi perlakuan untuk wilayah pedesaan, persentase dugaan yang berada di atas 50% adalah kelompok 8, 9, 10, 11, 12, 13, 14, dan 15. Karakteristiknya yaitu adanya perlakuan dengan cara direbus, diberi pemutih/khlor, dan disaring menggunakan kain (salah satu atau ketiganya) dengan lokasi sumber air berada di halaman dan rumah sendiri. Sementara itu, di wilayah perkotaan persentase dugaan di atas 50% terdapat pada kelompok 2, 3, 4, dan 7. Karakteristiknya adalah diberi perlakuan dengan sumber air berada di rumah sendiri, di halaman, dan di tempat lain. Selain itu, untuk persentase yang imbang antara perkotaan dan pedesaan terdapat pada kelompok 5 dengan karakteristik air minum pada rumah tangga itu disaring menggunakan kain, direbus dan tidak tahu direbus, dengan sumber air berada di rumah sendiri.

Setelah model terbentuk, dilakukan pengecekan keakurasian dengan membuat tabel *confusion matrix* yang dapat dilihat pada Tabel 5. Tujuannya adalah untuk melihat seberapa baik prediksi yang dilakukan oleh model dalam menduga karakteristik air minum menurut pedesaan dan perkotaan. Diperoleh hasil prediksi kategori perkotaan pada kategori perkotaan adalah sebesar 72784, sedangkan salah prediksinya adalah sebesar 28123. Kesalahan prediksi untuk kategori pedesaan adalah sebesar 34578 dan prediksi benarnya adalah 61929.

**Tabel 5. Confusion Matrix**

	Prediksi	
	Perkotaan	Pedesaan
Perkotaan	72784	28123
Pedesaan	34578	61929

Selanjutnya adalah melihat tingkat keakurasian dari model yang terbentuk yang dapat dilihat pada Tabel 6 pengecekan keakurasian dari model yang terbentuk.

**Tabel 6.** Evaluasi model

Evaluasi Model	Nilai
<i>Accuracy</i>	0.6824
<i>Sensitivity</i>	0.6779
<i>Specitivity</i>	0.6877
<i>Balanced Accuracy</i>	0.6828

Dari tingkat akurasinya, menunjukkan bahwa hasil prediksi dengan data tersebut adalah sebesar 0,6824 atau 68,24%. Ini berarti model yang dihasilkan agak baik untuk mengkorelasikan hasilnya dengan atribut dalam data yang digunakan. Lalu, dilihat nilai sensitivitasnya. Nilai sensitivitas merupakan nilai yang digunakan untuk mengukur proporsi dari jumlah amatan positif yang tepat prediksi. Berdasarkan hasilnya, sensitivitas yang diperoleh yaitu sebesar 0,6779 atau 67,79%. Hal ini menunjukkan bahwa nilai sensitivitasnya agak baik untuk menghitung proporsinya. Selanjutnya adalah *specificity* yang digunakan untuk mengukur jumlah amatan negatif yang tepat diprediksi. Nilainya adalah sebesar 0,6877 atau 68,77%. Hal ini menunjukkan bahwa nilai *specificity* cukup baik dalam mengukur proporsi jumlah amatan negatif yang tepat prediksi. Kemudian, untuk *balanced accuracy*, nilainya adalah sebesar 0,6828 atau 68,28%. Hal ini menunjukkan bahwa akurasi dari proporsi jumlah amatan kelas positif yang diprediksi cukup baik.

#### IV. KESIMPULAN

Berdasarkan hasil dari proses data mining, terdapat *missing data* pada variabel letak sumber air dengan mekanismenya adalah MNAR (*Missing Not At Random*). Penanganan yang digunakan untuk *missing data* tersebut adalah imputasi modus karena datanya bertipe kategorik dan proporsinya yaitu 22%. Kemudian, dari hasil analisis menggunakan metode CHAID, diperoleh sebanyak 15 kelompok. Kelompok yang memiliki perlakuan terhadap air minum terdiri atas 13 kelompok sedangkan untuk yang tidak ada perlakuan sama sekali hanya 2 kelompok. Karakteristik dari air minum yang tidak diberi perlakuan sama sekali ini terdapat pada rumah tangga di wilayah perkotaan dengan letak sumber air di rumah sendiri dan di halaman sendiri. Akan tetapi, untuk wilayah perkotaan yang lainnya ada diberi perlakuan pada air minum dan rumah tangga yang lebih banyak memberi perlakuan pada air minum adalah yang tinggi di wilayah pedesaan.

Tingkat keakurasian dari model yang terbentuk adalah sebesar 68,24% yang mengindikasikan bahwa model cukup mampu mengkorelasikan hasilnya dengan atribut dalam data yang digunakan. Kemudian, untuk sensitivitas adalah 67,79% yang mengindikasikan bahwa model yang terbentuk agak baik untuk menghitung proporsi. Selanjutnya adalah spesitivitas dengan nilai sebesar 68,77% yang menunjukkan bahwa model cukup baik dalam mengukur proporsi jumlah amatan negatif yang tepat prediksi. Lalu, untuk *balanced accuracy*, nilainya adalah 68,28% yang menunjukkan bahwa akurasi dari proporsi jumlah amatan kelas positif yang diprediksi cukup baik.

Peneliti selanjutnya disarankan menangani *missing data* dengan cara yang lain agar bisa melihat tingkat keakurasian dari model yang dihasilkan oleh metode CHAID sehingga dapat dibandingkan mana yang lebih baik untuk menentukan karakteristik air minum rumah tangga menurut perkotaan dan pedesaan di Indonesia.

#### DAFTAR PUSTAKA

- Amalita, N., Kurniawati, Y., dan Fitria, D. (2019), "Characteristics of bidikmisi's scholarship awardee in FMIPA UNP using chi-squared automatic interaction detection", *Journal of Physics: Conference Series*, Vol. 1317, No. 1, hal. 1–6.
- Arsyina, L., Wispriyono, B., Ardiansyah, I., dan Pratiwi, L. D. (2019), "Hubungan Sumber Air Minum dengan Kandungan Total Coliform dalam Air Minum Rumah Tangga", *Jurnal Kesehatan Masyarakat Indonesia*, Vol. 14, No. 2, hal. 18–23.
- Badan Pusat Statistik. (2017). *Statistik Kesejahteraan Rakyat 2017*. Jakarta: Badan Pusat Statistik.
- Badan Pusat Statistik. (2020). *Statistik Kesejahteraan Rakyat 2020*. Jakarta: Badan Pusat Statistik.
- Badan Pusat Statistik. (2023a). *Statistik Perumahan dan Permukiman 2022*. Jakarta: Badan Pusat Statistik.



- Badan Pusat Statistik. (2023b). *Persentase Rumah Tangga menurut Provinsi, Tipe Daerah dan Sumber Air Minum Layak (Persen), 2021-2023*, <https://www.bps.go.id/id/statistics-table/2/ODU0IZI=/persentase-rumah-tangga-menurut-provinsi--tipe-daerah-dan-sumber-air-minum-layak--persen-.html>. (diakses 30 Januari 2024)
- Bambang, S., Haryanti, S., dan Pangestuti, R. Y. (2022), Pelatihan Pengolahan Air Bersih di Bengkel Kerja Kesehatan Lingkungan Badegan Bantul. *J-ABDI: Jurnal Pengabdian Kepada Masyarakat*, Vol. 1, No. 11, hal. 2963–2968.
- Hendrawati, T. (2015), Kajian Metode Imputasi dalam Menangani Missing Data, *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS*, hal. 637–642.
- Herlambang, A. (2010), Teknologi Penyediaan Air Minum untuk Keadaan Tanggap Darurat, *Jurnal Air Indonesia*, Vol. 6, No. 1, hal. 52–63.
- Kementerian Kesehatan Republik Indonesia. (2018). *Profil Kesehatan Indonesia Tahun 2017*. Jakarta: Kementerian Kesehatan Republik Indonesia.
- Kementerian Kesehatan Republik Indonesia. (2020), *Manfaat Air Bersih dan Menjaga Kualitasnya*, <https://ayosehat.kemkes.go.id/manfaat-air-bersih-dan-menjaga-kualitasnya>. (diakses 28 Januari 2024)
- Lestari, N., dan Indriani, D. (2022), Identifikasi Pola Kepemilikan Sumber Air Minum Rumah Tangga di Jawa Timur, *Media Gizi Kesmas*, Vol. 11, No. 1, hal. 88–94.
- Little, R. J. A., dan Rubin, D. B. (2020). *Statistical Analysis with Missing Data (3rd edition)*. John Wiley & Sons.
- Longford, N. T. (2005). *Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician*. Springer Science & Business Media.
- Miftahuddin. (2012), Penggunaan Metode Chaid (Chi Square-Automatic Interaction Detection) Pada Pohon Klasifikasi Menggunakan Satu Peubah Respon Dengan Perbandingan Taraf Nyata, *Jurnal Matematika, Statistika, dan Komputasi*, Vol. 9, No. 1, hal. 11–22.
- Nurzanah, T. N., Zakianis, Z., Wispriyono, B., dan Anwar, A. (2020), Sanitasi dan Air Minum di Daerah Perkotaan dan Pedesaan di Provinsi Bengkulu (Analisis Data Potensi Desa 2018), *Jurnal Ekologi Kesehatan*, Vol. 18, No. 3, hal. 159–170.
- Pramana, S., Yuniarto, B., Santoso, I., Nooraeni, R., dan Suadaa, L. H. (2023). *Data Mining dengan R (2 ed.)*. Bogor: In Media.
- Putri, F. W., Vionanda, D., Putra, A. A., dan Fitri, F. (2023), Comparison of Error Prediction Methods in Classification Modeling with CHAID Methods for Balanced Data. *UNP Journal of Statistics and Data Science*, Vol. 1, No. 5, hal. 456–463.
- Rahim, Z. M., dan Muchlisoh, S. (2020), Faktor yang Mempengaruhi Penggunaan Sumber Air Minum Layak di Bengkulu Tahun 2018, *Seminar Nasional Official Statistics 2020*, hal. 1137–1146.
- Suhendra, M. A., Ispriyanti, D., dan Sudarno. (2020), Ketepatan Klasifikasi Pemberian Kartu Keluarga Sejahtera di Kota Semarang Menggunakan Metode Regresi Logistik Biner dan Metode CHAID, *Jurnal Gaussian*, Vol. 9, No. 1, hal. 64–74.
- Wandrivel, R., Suharti, N., dan Lestari, Y. (2012), Kualitas Air Minum Yang Diproduksi Depot Air Minum Isi Ulang Di Kecamatan Bungus Padang Berdasarkan Persyaratan Mikrobiologi, *Jurnal Kesehatan Andalas*, Vol. 1, No. 3, hal. 129-133.
- World Health Organization (WHO), dan United Nations Children’s Fund (UNICEF). (2021). *Progress on Household Drinking Water, Sanitation and Hygiene 2000-2020: Five Years into the SDGs*.