

Received: 27 October 2022

Revised: 21 December 2022

Accepted: 29 December 2022

Published: 31 December 2022

Klasifikasi Pemilih dalam Pemilu 2019 di Indonesia Menggunakan Regresi Logistik Multinomial dan *Chi-Square Automatic Decision Tree* (CHAID)

Curie Nabilah Nasution^{1, a)}, Yekti Widyaningsih^{1, b)}

¹*Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia, Depok*

E-mail: ^{a)} curie.nabilah@sci.ui.ac.id, ^{b)} yekti@sci.ui.ac.id (Penulis Koresponden)

Abstract

General Elections (Elections) in Indonesia are conducted once every 5 years. This paper focused on the election of the president and his representatives. Voters in elections are regulated by law. There are many aspects of a voter's background that influence voters in making decisions during elections. In this paper, the focus is on voter background factors in geographical, demographic, and socio-economic aspects. The geographical aspect is the region where voters live; demographic aspects are gender, age, recent education, marital status, and religion; socioeconomic aspects are the household expenses, household income, personal expenses, type of work, type of residence, and status of residence. There are 3 categories of voter decisions in the election, namely candidate pair A, candidate pair B, and refuse to vote. The objectives of this paper are to find out what variables significantly influence voters in making decisions during elections, and to classify voters based on the significance of variables in geographical, demographic, and socio-economic aspects related to voter choice in elections. Multinomial logistic regression analysis is used to answer the first objective, while the CHAID (Chi-Square Automatic Decision Tree) decision tree is used to answer the second objective. Through multinomial logistic regression analysis, it can be seen that the variable type of region, age, recent education, religion, household expenditure, household income, house type and home status influence voter decisions in elections. Through the CHAID decision tree, the results obtained are 5 types of voters based on a decision tree of several independent variables that are significant to the dependent variable.

Keywords: chi-square, classification, election, multinomial logistic regression

Abstrak

Pemilihan Umum (Pemilu) di Indonesia dilakukan setiap 5 tahun sekali. Penelitian ini berfokus pada pemilihan presiden dan wakilnya. Pemilih dalam pemilu diatur dengan undang-undang. Ada banyak aspek latar belakang pemilih yang mempengaruhi pemilih dalam mengambil keputusan saat pemilu. Dalam penelitian ini, fokusnya adalah pada faktor latar belakang pemilih dalam aspek geografis, demografis, dan sosial ekonomi. Aspek geografis adalah wilayah tempat tinggal

pemilih; aspek demografis adalah jenis kelamin, usia, pendidikan terakhir, status perkawinan, dan agama; Aspek sosial ekonomi adalah pengeluaran rumah tangga, pendapatan rumah tangga, pengeluaran pribadi, jenis pekerjaan, jenis tempat tinggal, dan status tempat tinggal. Ada 3 kategori keputusan pemilih dalam pemilu, yaitu pasangan calon A, pasangan calon B, dan menolak memilih. Tujuan dari penelitian ini adalah untuk mengetahui variabel-variabel apa yang secara signifikan mempengaruhi pemilih dalam pengambilan keputusan pada saat pemilihan umum, dan untuk mengklasifikasikan pemilih berdasarkan signifikansi variabel dalam aspek geografis, demografis, dan sosial ekonomi yang terkait dengan pilihan pemilih dalam pemilu. Analisis regresi logistik multinomial digunakan untuk menjawab tujuan pertama, sedangkan pohon keputusan CHAID (Chi-Square Automatic Decision Tree) digunakan untuk menjawab tujuan kedua. Melalui analisis regresi logistik multinomial dapat diketahui bahwa variabel jenis daerah, umur, pendidikan terakhir, agama, pengeluaran rumah tangga, pendapatan rumah tangga, jenis rumah dan status rumah berpengaruh terhadap keputusan pemilih dalam pemilihan umum. Melalui pohon keputusan CHAID, diperoleh hasil 5 tipe pemilih berdasarkan pohon keputusan dari beberapa variabel bebas yang signifikan terhadap variabel terikat.

Kata-kata kunci: *chi-square*, klasifikasi, pemilihan, regresi logistik multinomial

PENDAHULUAN

Pemilihan umum Presiden dan Wakil Presiden Republik Indonesia telah menarik perhatian yang cukup besar dari seluruh masyarakat di Indonesia. Hal ini menjadi bahan diskusi oleh berbagai politisi, akademisi, dan peneliti. Para ahli statistik telah mencoba melakukan penelitian mengenai perkiraan hasil pemilu dengan menggunakan berbagai teknik dan model statistik. Kemenangan presiden dan wakil presiden dalam setiap pemilihan umum di Indonesia dipengaruhi oleh beberapa faktor seperti demografi pasangan calon presiden, partai politik pendukung, dan visi misi pasangan calon presiden dan wakil presiden, dan program kerja yang ditawarkan oleh pasangan calon presiden dan wakil presiden.

Pemilih dalam pemilu di Indonesia diatur oleh undang-undang yang berlaku. Pemilih tersebar di seluruh wilayah Indonesia dan setiap pemilih memiliki latar belakang yang berbeda-beda. Hal ini menarik untuk menguji apakah latar belakang pemilih dapat mempengaruhi keputusan pemilih dalam pemilu. Dalam penelitian ini, variabel bebas yang digunakan adalah latar belakang pemilih berdasarkan aspek geografis, demografis, dan sosial ekonomi. Aspek geografis dalam penelitian ini adalah wilayah tempat tinggal pemilih. Aspek demografi dalam penelitian ini adalah jenis kelamin, usia, pendidikan terakhir, status perkawinan, dan agama. Terakhir, aspek sosial ekonomi dari penelitian ini adalah pendapatan rumah tangga, pengeluaran rumah tangga, pendapatan pribadi, jenis pekerjaan, jenis rumah, dan status rumah.

Misalnya dalam suatu pemilu, 2 pasangan calon presiden dan wakil presiden diberi inisial pasangan calon presiden dan wakil presiden A dan pasangan calon presiden dan wakil presiden B. Jika variabel-variabel yang telah ditetapkan bisa dimodelkan untuk mengetahui perbedaan keputusan pemilih untuk memilih calon presiden dan wakil presiden A, calon presiden dan wakil presiden B, atau menolak memilih salah satu pasangan calon, hal ini berarti fungsi pembeda antara ketiga pilihan tersebut dapat diperkirakan atau diprediksi. Tujuan penulisan ini secara umum adalah mengelompokkan responden menjadi 3 kelas berdasarkan karakteristiknya. Secara lebih spesifik, tujuan dari penulisan ini adalah untuk mengetahui variabel-variabel apa saja yang signifikan terhadap pemilih dalam pengambilan keputusan pada saat pemilu dan mengklasifikasikan pemilih berdasarkan signifikansi variabel dalam aspek geografis, demografis, dan sosial ekonomi yang terkait dengan pilihan pemilih dalam pemilu.

Menurut penulis, dari banyak metode statistik yang tersedia, penulis memilih regresi logistik multinomial dan pohon keputusan otomatis Chi-square (CHAID) untuk mencapai tujuan penelitian. Analisis regresi logistik multinomial adalah metode statistik untuk menganalisis hubungan antara variabel dependen dan variabel independen dimana variabel dependen adalah kategoris dan ada dua atau lebih kategori [8]. Melalui regresi logistik dapat diketahui variabel apa saja yang berhubungan dengan keputusan pemilih dalam pemilu. Sedangkan metode Chi-Square Automatic Interaction Detection

(CHAID) merupakan metode untuk mengklasifikasikan data kategorik dengan membagi suatu kumpulan data menjadi beberapa sub kelompok secara signifikan sesuai dengan variabel terikatnya [1]. Melalui metode CHAID dapat diketahui karakter masing-masing pemilih calon presiden dan wakil presiden A, pemilih presiden dan calon wakil presiden B, bahkan pemilih yang memutuskan untuk menolak memilih salah satu calon.

METODOLOGI

Bahan dan Data

Pada bagian ini akan dibahas variabel-variabel yang digunakan dalam makalah dan teknik pengambilan sampel yang digunakan dalam penelitian ini.

TABEL 1. Definisi Operasional Variabel Independen

<i>Jenis Variabel</i>	<i>Definisi</i>	<i>Kategori</i>	<i>Skala</i>
Tipe Area	Tempat tinggal responden.	1. Pedesaan 2. Ibukota Provinsi 3. Kotamadya di luar Ibukota Provinsi 4. Kabupaten di luar Ibukota Provinsi	Nominal
Umur	Umur responden saat disurvei.	Pembagian kelompok umur berdasarkan keputusan Kementerian Kesehatan (2009) 1. Remaja (20 - 25 tahun) 2. Dewasa (26 - 45 tahun) 3. Lanjut Usia (46 – 65 tahun)	Nominal
Jenis Kelamin	Ciri biologis yang dimiliki oleh responden.	1. Pria 2. Wanita	Nominal
Pendidikan Terakhir	Jenjang sekolah formal yang diselesaikan oleh responden dan dapat pengakuan menyelesaikan jenjang tersebut.	1. Universitas/Akademi/Sarjana Muda 2. SMP/SMA 3. ≤ SD (termasuk responden yang tidak menyelesaikan SD atau sama sekali tidak sekolah formal)	Nominal
Status Pernikahan	Ikatan perkawinan responden yang diakui dan dilakukan secara hukum dan agama yang sesuai dan hidup sebagai suami istri.	1. Belum menikah 2. Sudah menikah 3. Janda/Duda	Nominal
Agama	Keyakinan dan kegiatan keagamaan yang dilakukan oleh responden.	1. Muslim 2. non-Muslim	Nominal
Pengeluaran Rumah Tangga	Uang atau biaya yang rutin dikeluarkan dalam satu bulan oleh seluruh anggota dalam satu rumah tangga.	1. < Rp 1,5 juta 2. Rp 1,5 juta – 3 juta 3. Rp 3 juta	Nominal
Pendapatan Rumah Tangga	Total uang yang dihasilkan atau didapatkan oleh seluruh anggota di satu rumah tangga.	1. < Rp 1,5 juta 2. Rp 1,5 juta – 3 juta 3. Rp 3 juta	Nominal
Pendapatan Pribadi	Total uang yang dihasilkan atau didapatkan oleh satu anggota di suatu rumah tangga yang menjadi responden survey ini.	1. < Rp 1,5 juta 2. Rp 1,5 juta – 3 juta 3. Rp 3 juta	Nominal
Jenis Pekerjaan	Segala aktivitas yang dikerjakan yang menghasilkan uang yang dijadikan sumber atau pokok pencaharian dalam kehidupan.	1. Karyawan 2. Wira Swasta / Wira Usaha / Pedagang 3. Buruh	Nominal

		4. Non Karyawan (termasuk pelajar, ibu rumah tangga, dan orang yang tidak bekerja)	
Jenis Rumah	Tipe dari rumah yang ditinggali responden, dilihat dari bahan dan bentuk rumah.	1. Permanen 2. Non-Permanen	Nominal
Status Rumah	Status kepemilikan rumah responden yang menjadi tempat tinggal.	3. Milik Sendiri 4. Bukan Milik Sendiri	Nominal

Variabel dependen dalam tulisan ini adalah pilihan pemilih pada pemilihan presiden tahun 2019. Variabel ini ada 3 kategori, yaitu calon presiden dan wakil presiden A, calon presiden dan wakil presiden B, dan menolak memilih atau tidak tahu memilih.

Data diambil berdasarkan survei dari Danareksa *Research Institute*. Survei yang dilakukan adalah survei kepercayaan konsumen terhadap perekonomian Indonesia yang menambah pertanyaan seputar pemilihan presiden yang menjadi topik hangat saat itu. Survei dilakukan pada bulan Oktober 2018 selama 2 minggu di 6 provinsi yang memiliki produk domestik regional bruto tertinggi pada tahun 2019 (berdasarkan data BPS). Enam provinsi yang dimaksud adalah Sumatera Utara, DKI Jakarta, Jawa Tengah, Jawa Barat, Jawa Timur, dan Sulawesi Selatan. Dari masing-masing provinsi diambil sampel dengan menggunakan beberapa teknik sampel, yaitu:

Multi Stage Random Sampling

Metode ini sering digunakan dalam pengambilan sampel dari populasi yang tersebar secara geografis, sehingga sulit untuk menentukan kerangka sampel dari semua elemen dalam populasi tersebut. Syarat menggunakan metode ini adalah populasinya homogen, sangat besar, menempati area yang luas, dan memiliki biaya penelitian yang terbatas.

Simple Random Sampling

Sampel diambil dengan teknik sedemikian rupa sehingga setiap unit penelitian mendapat kesempatan yang sama untuk dipilih sebagai sampel. Pemilihan sampel menggunakan metode undian atau tabel acak.

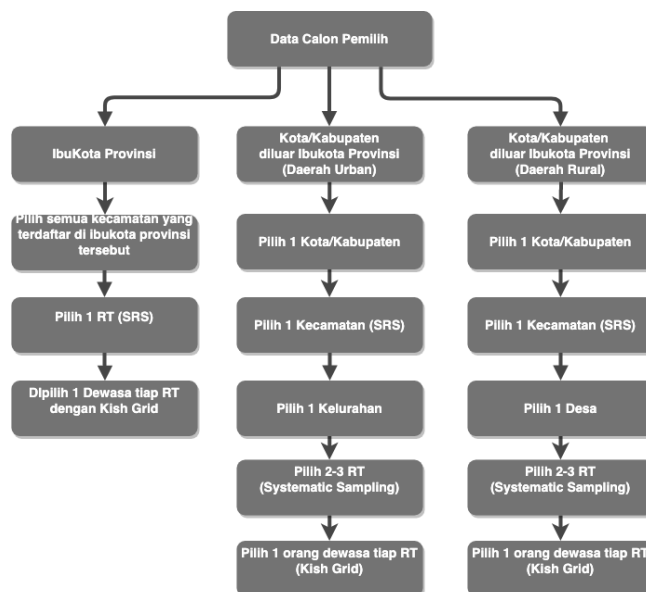
Systematic Sampling

Pengambilan sampel sistematis adalah metode pengambilan sampel, di mana hanya elemen pertama yang ditentukan secara acak, sedangkan sampel berikutnya dipilih secara sistematis mengikuti suatu pola. Syarat menggunakan sampling sistematis adalah populasi yang memiliki urutan atau populasi berupa daftar yang dapat diberi nomor.

Kish Grid

Kish grid adalah suatu metode pengambilan sampel untuk pemilihan responden secara acak dari suatu rumah tangga [6]. Metode digunakan supaya tidak ada bias dalam seleksi ketika menggunakan metode yang salah dalam memilih responden. Aspek yang paling penting dalam metode kish grid ini adalah metode yang menetapkan probabilitas yang sama untuk semua responden yang akan diwawancarai.

Berikut adalah gambar diagram sampling dalam survei ini.



GAMBAR 1. Diagram Sampling

Metode Penelitian

Dalam bagian ini akan dibahas mengenai metode statistika yang dipakai untuk mencapai tujuan penelitian yang sudah dijelaskan pada bagian pendahuluan.

Multinomial Logistic Regression

Multinomial logistic regression adalah suatu metode untuk menganalisis hubungan antara beberapa variabel bebas dengan variabel terikat kategoris yang terdiri lebih dari 2 kategori. Variabel terikat dapat dikodekan 1, 2, 3, sampai dengan k, dengan $y = 1$ sebagai kategori acuan atau *base level*. Dapat dikatakan bahwa model regresi logistik multinomial merupakan bentuk umum dari regresi logistik binomial. Fungsi logit adalah model yang dihasilkan dalam regresi logistik multinomial. Secara umum, fungsi logit untuk variabel dependen dengan 3 kategori dan kategori 1 sebagai level dasar dinotasikan sebagai berikut: dinotasikan sebagai berikut:

$$g_3(x_i) = \ln \left[\frac{\Pr(Y_i = 3 | x_i)}{\Pr(Y_i = 1 | x_i)} \right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1p}x_p \tag{2.1}$$

$$g_2(x_i) = \ln \left[\frac{\Pr(Y_i = 2 | x_i)}{\Pr(Y_i = 1 | x_i)} \right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2p}x_p \tag{2.2}$$

Persamaan untuk regresi logistik multinomial adalah untuk mengestimasi parameter. Menurut Hosmer dan Lemeshow (2000) persamaan untuk logistik multinomial secara umum adalah sebagai berikut:

$$(Y = i | x) = \pi_i(x_i) = \frac{\exp[g_i(x)]}{1 + \sum_{k=1}^{K-1} \exp[g_k(x)]} = \frac{\exp [\beta_{i0} + \beta_{i1}x_1 + \beta_{i2}x_2 + \dots + \beta_{ip}x_p]}{1 + \sum_{k=1}^{K-1} \exp [\beta_{k0} + \beta_{k1}x_1 + \beta_{k2}x_2 + \dots + \beta_{kp}x_p]} \tag{2.3}$$

dengan penjelasan notasi sebagai berikut:

- $i = 0, 1, \dots, K-1$
- $P(Y=i | x)$: Peluang bersyarat dari variabel dependen Y untuk kategori i diberikan nilai vektor x.
- $\pi_i(x)$: Persamaan regresi logistik untuk variabel dependen Y pada kategori ke-i.
- $g_i(x)$: Logit pada variabel dependen Y untuk kategori ke-i
- x : Vektor variabel independen
- β : Parameter model logit

Chi-Square Automatic Decision Tree (CHAID)

Teknik analisis *Chi Square Automatic Interaction Detection (CHAID)* pertama kali dikembangkan di Afrika Selatan dan diperkenalkan pada tahun 2018 oleh Gordon V. Kass dalam suatu artikel dengan judul “An Exploratory Technique for Investigating Large Quantities of Categorical Data”. Menurut Lehmann dan Eherler, metode *CHAID* adalah suatu metode untuk mengklasifikasikan data kategorik dengan membagi suatu set data menjadi beberapa subgrup-subgrup secara signifikan sesuai dengan variabel dependennya. Hasil dari metode ini adalah berbentuk sebuah diagram pohon yang menggambarkan hubungan antara variabel dependen dan variabel independen yang paling signifikan.

Hal pertama yang dilakukan dalam metode *CHAID* adalah membagi node 0 yang berisi jumlah pengamatan untuk setiap level dari variabel dependen berdasarkan kategori-kategori pada variabel independen 1 yang paling signifikan terhadap variabel dependen tersebut. Langkah kedua adalah membagi node berikutnya berdasarkan variabel independen 2 yang signifikan terhadap variabel dependen dan variabel independen 1. Langkah ini berlanjut pada node-node hasil pembagian berikutnya berdasarkan variabel-variabel yang independen selanjutnya, hingga tidak ada lagi variabel independen yang signifikan.

Pada metode *CHAID*, uji Chi-Square merupakan langkah awal untuk mengetahui variabel independen mana yang paling signifikan mempengaruhi variabel dependen. Variabel bebas yang paling signifikan mempengaruhi variabel terikat adalah variabel yang menjadi pemisah pertama dari node induk pada pohon keputusan *CHAID*. Variabel independen signifikan berikutnya adalah variabel pemisah pada node split awal. Seberapa signifikannya variabel dalam *CHAID* dilihat dari nilai chi-square hitung terbesar dengan p-value terkecil. Uji Chi-square dilakukan dengan menggunakan tabulasi silang, kemudian dicari nilai taksiran probabilitas masing-masing nilai pada tabel tabulasi silang. Hipotesis yang digunakan dalam uji Chi-square adalah:

$H_0 : \hat{p}_{ij} = \hat{p}_i \times \hat{p}_j$ (Variabel i dan variabel j tidak memiliki hubungan secara signifikan)

$H_1 : \hat{p}_{ij} \neq \hat{p}_i \times \hat{p}_j$ (Variabel i dan variabel j memiliki hubungan secara signifikan)

Statistik uji yang digunakan dalam Uji Chi-Square untuk pengujian hipotesis adalah sebagai berikut:

$$W_{ij} = \sum_{i=1}^m \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{2.4}$$

dengan penjelasan notasi sebagai berikut:

$i = 1, 2, \dots, m; j = 1, 2, \dots, k$

$O_{ij} = n_{ij}$: nilai sebenarnya dari pengamatan

$E_{ij} = n \hat{p}_{ij} = n \times \frac{n_{i*} n_{*j}}{n^2} = \frac{n_{i*} n_{*j}}{n}$: nilai harapan setiap sel dalam tabel *crosstab*

n_{ij} : total pengamatan variabel pertama Ai dan variabel kedua Bj

$n_{i.}$: total pengamatan variabel pertama Ai

$n_{.j}$: total pengamatan variabel kedua Bj

Aturan keputusan yang menolak H_0 dalam uji Chi-Square adalah jika $W > \chi_{(\alpha, (m-1), (k-1))}$ atau p-value $< \alpha$.

Hal penting berikutnya dalam metode *CHAID* adalah Koreksi Bonferroni. Koreksi Bonferroni adalah proses koreksi yang digunakan ketika beberapa uji statistik untuk melihat bebas atau tidak bebasnya suatu variabel dilakukan secara bersamaan.

HASIL DAN PEMBAHASAN

Pada bagian ini berisi pembahasan hasil penelitian dari pengolahan data menggunakan regresi logistik multinomial dan *CHAID*.

Regresi Logistik Multinomial

Nilai variabel dependen pada data survei pemilu 2019 terdiri dari 3 kategori yaitu “memilih Paslon A”, “memilih Paslon B”, dan “tidak memilih”. Kategori “Tidak Memilih” menjadi base level atau kategori referensi. Sehingga, model regresi multinomial yang dihasilkan terdiri dari 2 model untuk setiap aspeknya yaitu $p_1(x)$ dan $p_2(x)$. $p_1(x)$ adalah fungsi logit untuk variabel dependen Y untuk kategori ke-1, sedangkan $p_2(x)$ adalah fungsi logit untuk variabel dependen Y untuk kategori ke-2. Untuk melihat variabel apa saja yang signifikan mempengaruhi keputusan pemilih pada pemilihan calon presiden dan wakil presiden tahun 2019, biasanya dapat dilihat melalui p-value yang lebih kecil dari 0.1, atau 0.05. Pada penelitian ini, penulis memilih menggunakan p-value lebih kecil dari 0.1.

TABEL 2. Estimasi Parameter

Y	Variabel X	Nilai Taksiran β	Wald	P-value	$\exp(\beta)$
	Intersep (β_{10})	5,739	4,551	0,000	310,8
	Tipe Area				
	Ibukota Provinsi (x_{11})	-0,338	-1,532	0,125	0,713
	Kotamadya di luar Ibukota Provinsi (x_{12})	0,036	0,104	0,917	1,037
	Kabupaten di luar Ibukota Provinsi (x_{13})	1,211	4,057	0,000	3,359
	Umur				
	Dewasa (x_{14})	-0,328	-0,640	0,522	0,720
	Lanjut Usia (x_{15})	-0,270	-0,507	0,612	0,763
	Jenis Kelamin				
	Perempuan (x_{16})	-0,346	-1,489	0,136	0,707
	Pendidikan Terakhir				
	SMP_SMA (x_{17})	-0,478	-2,245	0,024	0,620
	Universitas/Akademi/Sarjana Muda (x_{18})	-0,620	-1,436	0,151	0,538
	Status Pernikahan				
	Menikah (x_{19})	-0,359	-0,757	0,449	0,698
	Janda/Duda (x_{110})	0,474	0,706	0,480	1,607
	Agama				
Pr($Y_i=1 x$)	Muslim (x_{111})	-3,067	-3,010	0,002	0,047
	Pengeluaran Rumah Tangga				
	Rp 1,5 juta – 3 juta (x_{112})	0,720	2,183	0,029	2,056
	> Rp 3 juta (x_{113})	0,994	2,030	0,042	2,701
	Pendapatan Rumah Tangga				
	Rp 1,5 juta – 3 juta (x_{114})	0,417	1,123	0,261	1,517
	> Rp 3 juta (x_{115})	1,232	2,568	0,010	3,430
	Penghasilan Pribadi				
	Rp 1,5 juta – 3 juta (x_{116})	0,099	0,379	0,704	1,104
	> Rp 3 juta (x_{117})	-0,548	-1,249	0,211	0,578
	Jenis Pekerjaan				
	Wiraswasta/Wirausaha/Pedagang (x_{118})	-0,474	-0,959	0,338	0,623
	Buruh (x_{119})	-0,345	-0,662	0,508	0,708
	Non Karyawan (x_{120})	0,043	0,084	0,933	1,044
	Jenis Rumah				
	Permanen (x_{121})	-0,567	-2,044	0,041	0,567
	Status Rumah				
	Milik Sendiri (x_{122})	-0,636	-2,697	0,007	0,530
Pr($Y_i=2 x$)	Intersep (β_{20})	2,682	2,037	0,000	14,61

Tipe Area				
Ibukota Provinsi (x ₂₁)	-0,038	-0,161	0,872	0,963
Kotamadya di luar Ibukota Provinsi (x ₂₂)	0,745	2,131	0,033	2,106
Kabupaten di luar Ibukota Provinsi (x₂₃)	1,208	3,846	0,000	3,346
Umur				
Dewasa (x ₂₄)	-0,405	-0,776	0,438	0,667
Lanjut Usia (x ₂₅)	-0,739	-1,352	0,176	0,478
Jenis Kelamin				
Perempuan (x ₂₆)	-0,843	-3,437	0,000	0,430
Pendidikan Terakhir				
SMP_SMA (x ₂₇)	-0,359	-1,591	0,111	0,698
Universitas/Akademi/Sarjana Muda (x ₂₈)	-0,191	-0,437	0,662	0,836
Status Pernikahan				
Menikah (x ₂₉)	-0,147	-0,301	0,763	0,864
Janda/Duda (x ₂₁₀)	1,077	1,556	0,120	2,935
Agama				
Muslim (x ₂₁₁)	-0,715	-0,665	0,506	0,489
Pengeluaran Rumah Tangga				
Rp 1,5 juta – 3 juta (x ₂₁₂)	0,663	1,858	0,063	1,941
> Rp 3 juta (x ₂₁₃)	1,112	2,173	0,030	3,043
Pendapatan Rumah Tangga				
Rp 1,5 juta – 3 juta (x ₂₁₄)	0,549	1,350	0,177	1,732
> Rp 3 juta (x ₂₁₅)	1,374	2,675	0,007	3,950
Penghasilan Pribadi				
Rp 1,5 juta – 3 juta (x ₂₁₆)	-0,122	-0,446	0,656	0,885
> Rp 3 juta (x ₂₁₇)	-0,561	-1,253	0,210	0,570
Jenis Pekerjaan				
Wiraswasta/Wirusaha/Pedagang (x ₂₁₈)	-0,412	-0,824	0,410	0,662
Buruh (x ₂₁₉)	-0,406	-0,765	0,444	0,666
Non Karyawan (x ₂₂₀)	0,264	0,507	0,612	1,303
Jenis Rumah				
Permanen (x ₂₂₁)	-0,300	-1,025	0,305	0,741
Status Rumah				
Milik Sendiri (x ₂₂₂)	-0,773	-3,202	0,001	0,462

Berdasarkan TABEL 2, huruf yang dicetak tebal menunjukkan variabel yang signifikan. Model yang dihasilkan berdasarkan TABEL 2 adalah sebagai berikut:

$$\begin{aligned}
 p_1(x) &= \ln \left(\frac{\Pr(Y_i = 1|x)}{\Pr(Y_i = 3|x)} \right) \\
 &= 5,739 - 0,338x_{11} + 0,036x_{12} + \mathbf{1,211}x_{13} - 0,328x_{14} - 0,270x_{15} - 0,346: \\
 &\quad - \mathbf{0,478}x_{17} - 0,620x_{18} - 0,359x_{19} + 0,474x_{110} - \mathbf{3,067}x_{111} + \mathbf{0,720}x_{112} \quad (3.1) \\
 &\quad + \mathbf{0,994}x_{113} + 0,417x_{114} + \mathbf{1,232}x_{115} + 0,099x_{116} - 0,548x_{117} - 0,474x_{11} \\
 &\quad - 0,345x_{119} + 0,043x_{120} - \mathbf{0,567}x_{121} - \mathbf{0,636}x_{122}
 \end{aligned}$$

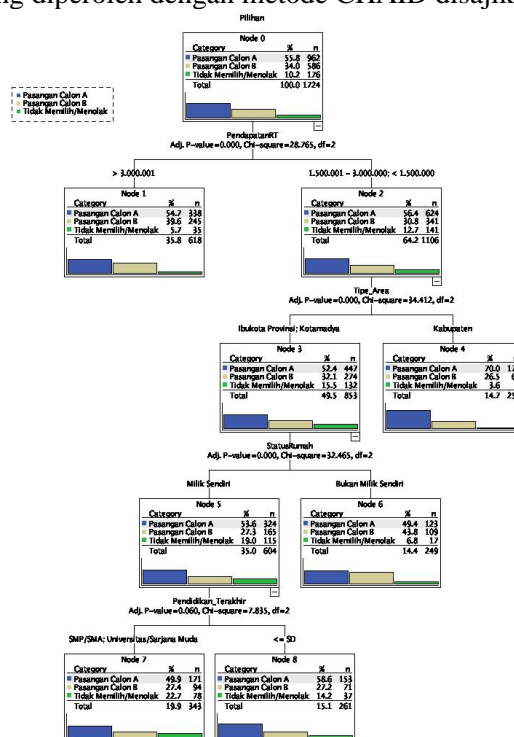
$$\begin{aligned}
 p_2(x) &= \ln \left(\frac{\Pr(Y_i = 2|x)}{\Pr(Y_i = 3|x)} \right) \\
 &= 2,682 - 0,038x_{21} + \mathbf{0,745}x_{22} + \mathbf{1,208}x_{23} - 0,405x_{24} - 0,739x_{25} - \mathbf{0,843}x \\
 &\quad - 0,359x_{27} - 0,191x_{28} - 0,147x_{29} + 1,077x_{210} - 0,715x_{211} + \mathbf{0,663}x_{212} \\
 &\quad + \mathbf{1,112}x_{213} + 0,549x_{214} + \mathbf{1,374}x_{215} - 0,122x_{216} - 0,561x_{217} - 0,412x_{218} \\
 &\quad - 0,406x_{219} + 0,264x_{220} - 0,300x_{221} - \mathbf{0,773}x_{222}
 \end{aligned} \tag{3.2}$$

Dari TABEL 2 terlihat bahwa variabel yang signifikan adalah tipe area, jenis kelamin, pendidikan terakhir, agama, pengeluaran rumah tangga, penghasilan rumah tangga, tipe rumah, dan status rumah, Interpretasi untuk beberapa variabel signifikan adalah sebagai berikut:

- Peluang pemilih yang tinggal di wilayah kabupaten di luar ibu kota provinsi untuk memilih Calon A pada Pilpres 2019 lebih mungkin terjadi 3,359 kali dibandingkan pemilih yang tinggal di pedesaan untuk menolak memilih pada Pilpres 2019,
- Peluang pemilih perempuan memilih Calon B pada Pemilu 2019 lebih besar kemungkinannya terjadi 0,430 kali dibandingkan pemilih laki-laki yang tidak memilih pada Pilpres 2019.

Chi-Square Automatic Decision Tree (CHAID)

Dari regresi logistik multinomial diperoleh variabel yang signifikan. Pada bagian ini akan dibuat pohon keputusan berdasarkan variabel signifikan yang kita peroleh dari regresi logistik multinomial dengan metode CHAID. Dengan menggunakan tabel *crosstab*, kita dapat mengetahui variabel mana yang paling signifikan dan variabel ini akan menjadi *node* pemisah awal di CHAID. Kriteria variabel yang paling signifikan adalah variabel yang memiliki nilai chi-square hitung terbesar dan p-value terkecil. Pohon keputusan yang diperoleh dengan metode CHAID disajikan pada Gambar 2.



GAMBAR 2. Pohon Keputusan dengan CHAID

Hasil analisis didasarkan pada pohon keputusan CHAID yang memisahkan data masukan ke dalam kelompok-kelompok berdasarkan variabel bebas yang memiliki nilai signifikansi kurang dari 0.1. Dari pohon keputusan yang telah terbentuk, terdapat 5 tipe pemilih yang disajikan pada TABEL 3.

TABEL 3. Jenis Pemilih berdasarkan pohon keputusan CHAID

No.	Tipe Pemilih	Jumlah Pemilih yang Memilih Calon Kandidat A		Jumlah Pemilih yang Memilih Calon Kandidat B		Jumlah Pemilih yang Menolak Memilih	
			%		%		%
1.	Pemilih yang memiliki pendapatan rumah tangga lebih dari Rp 3.000.000,00.	338	54,7 %	245	39,6 %	35	5,7 %
2.	Pemilih yang memiliki pendapatan rumah tangga antara Rp 1.500.001,00 sampai dengan Rp 3.000.000,00 atau kurang dari Rp 1.500.000,00 berdomisili di ibu kota provinsi atau kota di luar ibu kota, memiliki rumah sendiri dan pendidikan terakhir SMP, SMA, atau Perguruan Tinggi.	171	49,9%	94	27,4%	78	22,7 %
3.	Pemilih yang memiliki pendapatan rumah tangga antara Rp 1.500.001,00 sampai dengan Rp 3.000.000,00 atau kurang dari Rp 1.500.000,00, berdomisili di ibu kota provinsi atau kota di luar ibu kota, memiliki rumah sendiri dan pendidikan terakhir kurang atau sama dengan sekolah dasar.	153	58,6 %	71	14,2 %	37	14,2 %
4.	Pemilih yang memiliki pendapatan rumah tangga antara Rp 1.500.001,00 sampai dengan Rp 3.000.000,00 atau kurang dari Rp 1.500.000,00, berdomisili di ibu kota provinsi atau kota di luar ibu kota, dan memiliki rumah bukan milik sendiri.	123	49,4 %	109	43,8 %	17	6,8 %
5.	Pemilih Pemilih yang pendapatan rumah tangganya antara Rp 1.500.000,00 sampai dengan Rp 3.000.000,00 atau kurang dari Rp 1.500.000,00 dan berdomisili di kabupaten di luar ibu kota.	177	70,0 %	67	26,5 %	9	3,6 %

Berdasarkan TABEL 3, kelompok yang memiliki persentase terbesar yang memilih presiden dan wakil presiden dari pasangan calon A adalah pemilih tipe ketiga, sedangkan kelompok yang memiliki persentase terbesar yang memilih pasangan calon presiden dan wakil presiden B adalah pemilih tipe keempat. Untuk kelompok dengan persentase pemilih yang menolak memilih terbesar adalah pemilih tipe kedua. Untuk melihat keakuratan pohon keputusan dalam mengklasifikasikan pemilih pada pemilihan presiden dan wakil presiden 2019 menggunakan CHAID, tabel akurasi klasifikasi disajikan pada TABEL 4 dan TABEL 5 di bawah ini:

TABEL 4. Tabel Risiko CHAID

Risk	
Estimate	Std. Error
.423	.012

Growing Method:
CHAID
Dependent Variable:
Pilihan

TABEL 5. Tabel Klasifikasi CHAID

Observed	Predicted			Percent Correct
	Pasangan Calon A	Pasangan Calon B	Tidak Memilih/Menolak	
Pasangan Calon A	904	26	32	94.0%
Pasangan Calon B	520	53	13	9.0%
Tidak Memilih/Menolak	137	2	37	21.0%
Overall Percentage	90.5%	4.7%	4.8%	57.7%

Growing Method: CHAID
Dependent Variable: Pilihan

Estimasi risiko sebesar 0,423 menunjukkan bahwa kategori yang diprediksi oleh pohon keputusan salah untuk 42,3% kasus. Hasil pada tabel klasifikasi hampir sama dengan estimasi risiko. Tabel tersebut menunjukkan bahwa model mengklasifikasikan sekitar 57,7% pemilih dengan benar.

KESIMPULAN DAN SARAN

Berdasarkan analisis yang dilakukan, ada beberapa hal yang dapat disimpulkan. Melalui regresi logistik multinomial dapat diketahui variabel-variabel yang berpengaruh signifikan terhadap keputusan pemilih pada pemilihan presiden dan wakil presiden 2019 di Indonesia. Variabel signifikan yang diperoleh adalah jenis wilayah, pendidikan terakhir, jenis kelamin, agama, pendapatan rumah tangga, pengeluaran rumah tangga, jenis rumah, dan status rumah. Setelah menemukan variabel yang signifikan, selanjutnya yang dilakukan adalah membuat pohon keputusan dengan menggunakan metode CHAID. Melalui CHAID, dapat diketahui tipe pemilih mana yang mendominasi memilih calon a, calon b, atau menolak memilih. Pemilih kandidat A didominasi oleh pemilih tipe ketiga, pemilih kandidat B didominasi oleh pemilih tipe keempat, dan pemilih yang menolak memilih didominasi oleh pemilih tipe kedua. Akurasi klasifikasi menggunakan metode CHAID adalah 57,7%.

UCAPAN TERIMA KASIH

Penelitian ini didukung oleh Direktorat Riset dan Pengembangan Universitas Indonesia, melalui Program Hibah Publikasi Terindeks Internasional (PUTI) Q2 Tahun Anggaran 2022-2023 Nomor: NKB-668/UN2.RST/HKP.05.00/2022 sebagai luaran tambahan. Ucapan terima kasih juga ditujukan kepada *Danareksa Research Institute* yang telah menyediakan data yang digunakan dalam penulisan ini.

REFERENSI

- Adha , Muhammad Rizky. (2018). Regresi Logistik Multinomial dan Regresi Spline untuk Pemodelan Risiko Kredit [skripsi]. Depok (ID): UI.
- Bagozzi, R.P. (1994). *Advanced Methods of Marketing Research*. Blackwell Publishers Ltd., Oxford.
- Garson, G. David. (2016). *Logistic Regression: Binary & Multinomial*. USA: Statistical Publishing Associates.
- Gallagher, C.A., Monroe, H.M., Fish, J.L. (2000). "An Iterative Approach to Classification Analysis". https://www.casact.org/sites/default/files/database/dpp_dpp90_90dpp237.pdf
- Hosmer, D.W. dan Lemeshow, S. (2000). *Applied Logistic Regression*, John Wiley & Sons, USA.
- Lehmann T and Eherler D 2001 Responder Pro- filing with CHAID and Dependency Analysis [Online]. Available: <https://pdfs.semanticscholar.org/4536/a9350179566c6e47b0ff9b283457aa6c75b0.pdf>
- Kementrian Kesehatan Indonesia 2009 Kategori Usia [Online]. Available: <http://kategori-umurmenurut-Depkes.html>
- Kunto, Y.S. dan Hasana, S.N. (2006). Analisis CHAID Sebagai Alat Bantu Statistika Segmentasi Pasar. *Jurnal Manajemen*. 1(2), 88-97. <https://jurnalpemasaran.petra.ac.id/index.php/mar/article/view/16538>
- Mendenhall, W., dan Sincich, Terry. (2012). *A Second Course in Statistics Regression Analysis*. Pearson Education Inc.
- Sharma, Mohit. (2018). *Multinomial Logistic Regression Using R*. <https://datasciencebeginners.com/2018/12/20/multinomial-logistic-regression-using-r/>.
- Stephanie. (2017). *Kish Grid: What It Is and How to Use It*. <https://www.statisticshowto.datasciencecentral.com/kish-grid/>.

Zara, Novita. (2017). Identifikasi Faktor – Faktor yang Memengaruhi Keberhasilan Akademik Mahasiswa Menggunakan Regresi Logistik Multinomial [skripsi]. Bogor (ID): IPB