

Metode *Exhaustive* CHAID untuk Klasifikasi Rumah Tangga Penerima KKS di Jawa Barat Tahun 2024

Adisty Tritania Maharani *, Reny Rian Marlina

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

adistymhrmn@gmail.com, renyrianmarlina@unisba.ac.id

Abstract. The poverty rate in West Java in 2024 reached 7.78% or around 3.8 million people. The distribution of social assistance through the Family Welfare Card (KKS) program still suffers from inaccuracy in targeting, with a potential loss to the state of Rp523 billion per month according to the BPK. This study uses the *Exhaustive* CHAID method to accurately classify KKS recipient households. The data used comes from the 2024 SUSENAS survey with 13,861 respondents, divided into 11,045 training data and 2,816 testing data, involving 12 independent variables. The resulting decision tree has a depth of three levels and 10 terminal *nodes*. The five most significant variables are the highest level of education of the household head (KRT), the number of health insurance policies, the type of floor material, the number of household members covered, and the status of home ownership. Households with low education levels, health insurance, and semi-permanent floors tend to receive KKS. Conversely, households with high levels of education, clear ownership status, and relatively few members were more likely to be classified as not receiving KKS. Model performance evaluation shows an accuracy of 64.72% on the training data and 66.23% on the testing data. The precision, recall, and F1-score values on the testing data are 57.3%, 57.76%, and 57.53%, respectively. The Matthews Correlation Coefficient (MCC) value of 0.2950 indicates a positive correlation, although it is still low, possibly due to an unbalanced class distribution. These results suggest that *Exhaustive* CHAID is sufficiently effective in classifying KKS recipients and can be used to improve the accuracy of social assistance distribution.

Keywords: *Classification, Exhaustive CHAID, KKS.*

Abstrak. Tingkat kemiskinan di Jawa Barat tahun 2024 mencapai 7,78% atau sekitar 3,8 juta jiwa. Penyaluran bantuan sosial melalui program Kartu Keluarga Sejahtera (KKS) masih mengalami ketidaktepatan sasaran, dengan potensi kerugian negara sebesar Rp523 miliar per bulan menurut BPK. Penelitian ini menggunakan metode *Exhaustive CHAID* untuk mengklasifikasikan rumah tangga penerima KKS secara tepat. Data yang digunakan berasal dari SUSENAS 2024 dengan 13.861 responden, dibagi menjadi 11.045 data *training* dan 2.816 data *testing*, serta melibatkan 12 variabel independen. Pohon keputusan yang dihasilkan memiliki kedalaman tiga tingkat dan 10 terminal *node*. Lima variabel paling signifikan yaitu pendidikan terakhir kepala rumah tangga (KRT), jumlah jaminan kesehatan, bahan lantai rumah, jumlah anggota rumah tangga ditanggung, dan status kepemilikan rumah. Rumah tangga dengan pendidikan dasar, memiliki jaminan kesehatan, dan lantai semi permanen cenderung menerima KKS. Sebaliknya, rumah tangga dengan pendidikan terakhir KRT yang tinggi, status kepemilikan rumah tangga nya jelas, serta jumlah anggota rumah tangga yang relatif sedikit, lebih dominan diklasifikasikan sebagai tidak menerima KKS. Evaluasi performa model menunjukkan akurasi sebesar 64,72% pada data *training* dan 66,23% pada data *testing*. Nilai *precision*, *recall*, dan *F1-score* pada data *testing* masing-masing sebesar 57,3%, 57,76%, dan 57,53%. Nilai *Matthews Correlation Coefficient* (MCC) sebesar 0,2950 menunjukkan adanya korelasi positif meskipun masih rendah, kemungkinan karena distribusi kelas yang tidak seimbang. Hasil ini menunjukkan bahwa *Exhaustive CHAID* cukup efektif dalam mengklasifikasikan penerima KKS dan dapat digunakan untuk meningkatkan ketepatan penyaluran bantuan sosial.

Kata Kunci: *Klasifikasi, Exhaustive CHAID, KKS.*

A. Pendahuluan

Kemiskinan merupakan persoalan sosial yang kompleks dan memiliki dampak yang luas terhadap kesejahteraan masyarakat (Mandjarreki, 2021). Pemerintah Indonesia telah melakukan berbagai upaya untuk menanggulangi masalah ini, salah satunya melalui peluncuran Program Kartu Keluarga Sejahtera (KKS). Program ini mengalami transformasi dari Kartu Perlindungan Sosial (KPS) yang disalurkan secara tunai menjadi KKS dengan penyaluran non-tunai berbasis data dari Basis Data Terpadu (BDT), guna meningkatkan ketepatan sasaran (Afrillia, 2021). Provinsi Jawa Barat, dengan jumlah penduduk lebih dari 49 juta jiwa pada tahun 2024, merupakan salah satu wilayah dengan tingkat kepadatan tertinggi di Indonesia (GoodStats Data, 2024). Berdasarkan data Badan Pusat Statistik (BPS), tercatat bahwa persentase penduduk miskin di provinsi ini mencapai 7,78% atau sekitar 3,8 juta jiwa, dengan tingkat ketimpangan pendapatan yang ditunjukkan oleh nilai Gini Ratio sebesar 0,421 (BPS Jawa Barat, 2024).

Namun demikian, pelaksanaan program KKS masih menghadapi permasalahan klasik, yaitu ketidaktepatan sasaran. Berdasarkan audit Badan Pemeriksa Keuangan (BPK), pada periode awal 2021 hingga awal 2023, ditemukan bahwa penyaluran bantuan sosial yang tidak tepat sasaran menyebabkan kerugian negara sekitar Rp523 miliar setiap bulan. Hal ini disebabkan oleh data penerima bantuan yang belum diperbarui, distribusi KKS yang belum merata, serta pengembalian dana ke rekening pemerintah yang belum ditindaklanjuti (Gandhawangi, 2023). Oleh karena itu, BPK merekomendasikan kepada Kementerian Sosial untuk meningkatkan pengawasan serta pengendalian terhadap proses penyaluran agar lebih akurat dan tepat sasaran (Singgalang, 2024).

Untuk menjawab tantangan tersebut, diperlukan pendekatan analitik yang mampu mengklasifikasikan rumah tangga secara lebih akurat sebagai dasar penentuan penerima bantuan. Metode analisis diskriminan merupakan salah satu teknik statistik yang sering digunakan dalam klasifikasi, karena kemampuannya membentuk fungsi klasifikasi dari kombinasi linier variabel prediktor (Hair Jr et al., 2014). Akan tetapi, metode ini memiliki keterbatasan, terutama ketika digunakan pada data sosial ekonomi yang kompleks dan sering tidak memenuhi asumsi-asumsi klasik seperti normalitas multivariat dan kesamaan matriks kovarian antar kelompok. Di sisi lain, metode pohon keputusan seperti *Classification and Regression Tree* (CART) juga digunakan untuk keperluan klasifikasi, tetapi struktur biner yang dihasilkan menjadikan interpretasi kebijakan menjadi kurang efisien, terutama dalam menangani variabel kategorik yang kompleks (Hastie et al., 2009).

Sebagai alternatif, metode *Chi-Square Automatic Interaction Detector* (CHAID) dapat digunakan untuk membentuk pohon keputusan yang membagi data berdasarkan uji *Chi-Square* terhadap variabel kategorik (Kass, 1980). CHAID memiliki keunggulan dalam membentuk cabang lebih dari dua pada setiap *node*, sehingga struktur pohon menjadi lebih ringkas. Meski demikian, kelemahannya terletak pada proses penggabungan kategori yang tidak menyeluruh, yang dapat mengakibatkan hilangnya informasi penting (Saputra, 2024). Oleh karena itu, dikembangkan metode *Exhaustive* CHAID yang menyempurnakan CHAID dengan mencoba seluruh kemungkinan penggabungan kategori sebelum menentukan pemisahan terbaik. Meskipun membutuhkan waktu komputasi yang lebih lama, metode ini mampu menghasilkan klasifikasi yang lebih akurat dan mudah diinterpretasikan (Tohari et al., 2017).

B. Metode

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) Provinsi Jawa Barat, yaitu data hasil Survei Sosial Ekonomi Nasional (SUSENAS) periode Maret tahun 2024. Data yang diambil merupakan data rumah tangga (RT) yang dikumpulkan oleh BPS melalui survei terhadap Kepala Rumah Tangga (KRT) sebagai responden utama sebanyak 13.861 orang.

Tahap Penggabungan

Pada tahap penggabungan ini dilakukan (1) membentuk tabel kontigensi dua arah untuk masing-masing kategori. (2) Menghitung statistika uji *Chi-Square* setiap pasang kategori yang dapat dipilih untuk digabung menjadi satu, dimana langkah melakukan uji *Chi-Square* adalah sebagai berikut:

Hipotesis untuk pengujian *Chi-Square*:

H_0 : Variabel respon dan variabel prediktor saling bebas

H_1 : Variabel respon dan variabel prediktor tidak saling bebas

Taraf signifikansi 5% dan statistik uji nya:

$$\chi^2 = \sum_{i=1}^a \sum_{j=1}^b \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad \dots (1)$$

Persamaan yang digunakan untuk menghitung frekuensi harapan pada statistik uji *Chi-Square* adalah sebagai berikut

$$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n_{..}} \quad \dots (2)$$

Kriteria penerimaan: Terima H_0 jika $X_{Hitung}^2 \leq X_{\alpha; df=(a-1)(b-1)}^2$ dan terima H_1 jika $X_{Hitung}^2 > X_{\alpha; df=(a-1)(b-1)}^2$ atau dengan kriteria lain yang membandingkan nilai signifikansi (*P-Value*) dengan taraf signifikansi (α). Dengan kriteria H_0 ditolak jika $P - Value(sig.) < \alpha$

(3) Dari pasangan kategori yang tidak signifikan, gabungkan pasangan yang paling mirip yakni dengan nilai *Chi-Square* terkecil dan *P-Value* terbesar pada variabel prediktor dengan lebih dari dua kategori. (4) Setelah penggabungan, periksa kembali signifikansi kategori baru. Jika masih ada pasangan tidak signifikan, ulangi proses penggabungan hingga seluruh pasangan signifikan. (5) Jika terjadi penggabungan, hitung *P-Value* yang telah dikoreksi menggunakan metode Bonferroni sesuai jenis variabel independen (Gallagher et al., 2000)

Variabel independen skala monotonik:

$$M = c - k \quad \dots (3)$$

Variabel independen skala bebas:

$$M = \sum_{i=0}^{c-k-1} (-1)^i \frac{(c-k-i)^{c-k+1}}{i! (c-k-i)!} \quad \dots (4)$$

Variabel independen skala mengambang:

$$M = 1 + \sum_{k=1}^{c-3} N(c-k+1, c-k) \quad \dots (5)$$

Tahap Pemisahan

Pemilihan variabel independen terbaik dilakukan dengan mempertimbangkan nilai *P-Value* terkecil. Selanjutnya, pembagian kelompok dilakukan berdasarkan variabel independen tersebut, dengan memanfaatkan kategori-kategori yang telah digabung secara optimal untuk membentuk subkelompok baru dari kelompok induk. Namun, jika tidak terdapat variabel independen dengan nilai *P-Value* yang signifikan, maka proses pembagian kelompok tidak dilakukan.

Tahap Penghentian

Tahap penghentian dalam pembentukan pohon keputusan dilakukan saat pohon tidak dapat tumbuh lebih lanjut. Hal ini terjadi jika simpul hanya berisi satu kelas, kedalaman maksimum tercapai, jumlah kasus di simpul kurang dari batas minimal untuk menjadi simpul orang tua, pemisahan menghasilkan simpul anak dengan jumlah kasus di bawah batas minimal, atau tidak ada variabel independen yang signifikan.

Keakuratan Klasifikasi

Untuk mengevaluasi hasil ketepatan klasifikasi, digunakan metode *confusion matrix* yang berfungsi untuk membandingkan antara hasil prediksi dengan kondisi observasi dari data yang dianalisis (Fanggidae et al., 2021). Melalui *confusion matrix*, dapat diperoleh informasi penting terkait tingkat keberhasilan dan kesalahan model klasifikasi, seperti nilai akurasi, presisi, *recall*, dan *F1-score* serta MCC (*Matthews Correlation Coefficient*).

Tabel 1. Confusion Matrix

Confusion Matrix	Prediksi
------------------	----------

Observasi	Positive		Negative
	Positive	TP	FN
	Negative	FP	TN

Maka perhitungan untuk mengetahui informasi penting terkait tingkat keberhasilan dan kesalahan model klasifikasi sebagai berikut.

Akurasi:

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad \dots (6)$$

Presisi:

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad \dots (7)$$

Recall:

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad \dots (8)$$

F1-Score:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \dots (9)$$

MCC:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad \dots (10)$$

C. Hasil Penelitian dan Pembahasan

Tahap penggabungan

Dalam metode *Exhaustive CHAID*, setiap pasangan kategori dari variabel prediktor yang telah dikategorikan dianalisis secara menyeluruh menggunakan uji *Chi-Square*. Penggabungan dilakukan dengan memprioritaskan pasangan yang memiliki nilai *P-Value* paling mendekati. Hasil penggabungan kategori dan nilai *Chi-Square* ditunjukkan pada Tabel 2. berikut.

Tabel 2. Hasil Uji *Chi-Square* dan penggabungan Kategori Node 0

Variabel Prediktor	Kategori			X^2_{Hitung}	<i>P-Value</i>
	I	II	III		
Usia KRT (X_1)	Produktif	Tidak Produktif		16,3023	< 0,0001
Pendidikan Terakhir KRT (X_2)	Dasar	Menengah	Tinggi	546,8973	< 0,0001
Status Pekerjaan KRT (X_3)	Bekerja	Tidak Bekerja		2,0287	0,15435
Jumlah Anggota RT Ditanggung (X_4)	≤ 3	> 3		38,9304	< 0,0001
Pengeluaran per Kapita per Bulan (X_5)	Mampu	Tidak Mampu		1,5423	0,21428
Jumlah Jamkes (X_6)	Memiliki 1 Jamkes	Memiliki ≥ 2 Jamkes; Tidak Memiliki Jamkes		223,9993	< 0,0001
Status Kepemilikan Rumah (X_7)	Milik Sendiri; Bebas Sewa	Kontrak Sewa; Dinas		32,0239	< 0,0001
Sumber Penerangan Utama RT (X_8)	PLN Meteran	PLN Bukan Meteran; Listrik		16,9044	< 0,0001

Variabel Prediktor	Kategori			X^2_{Hitung}	$P-Value$
	I	II	III		
		Bukan PLN/Tanpa Listrik			
Luas Lantai Bangunan per Kapita (m^2) (X_9)	$\geq 7,2$	$< 7,2$		22,7172	$< 0,0001$
Bahan Utama Dinding RT (X_{10})	Tembok	Plesteran Anyaman/Kawat; Kayu/Papan; Bambu		77,6666	$< 0,0001$
Bahan Utama Lantai RT (X_{11})	Permanen; Tidak Permanen	Semi Permanen		114,1006	$< 0,0001$
Fasilitas BAB RT (X_{12})	Pribadi; Tidak Ada Fasilitas	Umum		19,1967	$< 0,0001$

Tabel 2. menunjukkan bahwa proses penggabungan kategori pada *node 0* terjadi pada variabel jumlah jamkes (X_6), dimana kategori memiliki ≥ 2 jamkes dan tidak memiliki jamkes digabungkan sehingga tersisa dua kategori. Pasangan kategori yang memiliki $P-Value$ yang lebih dari 0,05 digabungkan menjadi satu kategori baru. Begitu pula untuk variabel prediktor lainnya jika memiliki lebih dari dua kategori dilakukan penggabungan kategori.

Selanjutnya, nilai $P-Value$ dikalikan dengan koreksi Bonferroni untuk setiap variabel prediktor yang mengalami penggabungan kategori. Variabel prediktor yang mengalami proses penggabungan merupakan variabel bebas. Hasil perhitungan koreksi Bonferroni pada variabel prediktor menggunakan metode *Exhaustive CHAID* sesuai dengan Persamaan (4), disajikan pada Tabel 3. berikut.

Tabel 3. Nilai Koreksi Bonferroni Variabel Prediktor *Node 0*

Variabel Prediktor	Jenis Variabel	Nilai Koreksi Bonferroni
Jumlah Jamkes (X_6)	Bebas	1
Status Kepemilikan Rumah (X_7)	Bebas	3
Sumber Penerangan Utama RT (X_8)	Bebas	1
Bahan Utama Dinding RT (X_{10})	Bebas	3
Bahan Utama Lantai RT (X_{11})	Bebas	1
Fasilitas BAB RT (X_{12})	Bebas	1

Nilai koreksi Bonferroni untuk masing-masing variabel prediktor yang mengalami penggabungan kategori disajikan pada Tabel 3. Nilai ini kemudian dikalikan dengan $P-Value$ dari masing-masing variabel prediktor untuk memperoleh $P-Value$ yang telah dikoreksi menggunakan metode Bonferroni. Hasil $P-Value$ yang telah dikoreksi tersebut ditampilkan pada Tabel 4.

Tabel 4. Nilai $P-Value$ Terkoreksi Bonferroni untuk Variabel Prediktor

Variabel Prediktor	Kategori		X^2_{Hitung}	$P-Value$
	I	II		
Jumlah Jamkes (X_6)	Memiliki 1 Jamkes	Memiliki ≥ 2 Jamkes; Tidak Memiliki Jamkes	223,9993	$< 0,0001$

Variabel Prediktor	Kategori		X^2_{Hitung}	<i>P-Value</i>
	I	II		
Status Kepemilikan Rumah (X_7)	Milik Sendiri; Bebas Sewa	Kontrak Sewa; Dinas	32,0239	< 0,0001
Sumber Penerangan Utama RT (X_8)	PLN Meteran	PLN Bukan Meteran; Listrik Bukan PLN/Tanpa Listrik	16,9044	< 0,0001
Bahan Utama Dinding RT (X_{10})	Tembok	Plesteran Anyaman/Kawat; Kayu/Papan; Bambu	77,6666	< 0,0001
Bahan Utama Lantai RT (X_{11})	Permanen; Tidak Permanen	Semi Permanen	114,1006	< 0,0001
Fasilitas BAB RT (X_{12})	Pribadi; Tidak Ada Fasilitas	Umum	19,1967	< 0,0001

Nilai *P-Value* yang telah dikoreksi menggunakan metode Bonferroni untuk masing-masing variabel prediktor disajikan pada Tabel 4. dapat disimpulkan bahwa setiap variabel prediktor memiliki tidak saling bebas terhadap status penerimaan KKS.

Tahap Pemisahan

Tahap pemisahan bertujuan untuk menentukan kategori yang paling tepat sebagai pemisah simpul pada pohon klasifikasi. Proses ini dilakukan dengan membandingkan nilai *P-Value* yang diperoleh dari tahap penggabungan untuk setiap kategori. Langkah awal dalam analisis *Exhaustive CHAID* adalah melakukan uji *Chi-Square* guna mengidentifikasi variabel prediktor dengan *P-Value* terkecil atau nilai X^2_{Hitung} tertinggi, yang kemudian digunakan sebagai pemisah simpul.

Tabel 5. Hasil Uji *Chi-Square* antara KKS dengan Variabel Prediktor

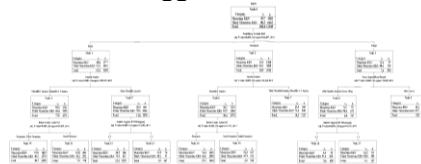
Pasangan Variabel	X^2_{Hitung}	X^2_{Tabel}	<i>P-Value</i>	Keterangan
Usia KRT*KKS	16,3023	3,481	< 0,0001	Tidak Saling Bebas
Pendidikan Terakhir KRT*KKS	546,897	5,591	< 0,0001	Tidak Saling Bebas
Status Pekerjaan KRT*KKS	2,0287	3,481	0,15435	Saling Bebas
Jumlah Anggota RT Ditanggung*KKS	38,9304	3,481	< 0,0001	Tidak Saling Bebas
Pengeluaran per Kapita per Bulan*KKS	1,5423	3,481	0,21428	Saling Bebas
Jumlah Jamkes*KKS	223,999	5,591	< 0,0001	Tidak Saling Bebas
Status Kepemilikan Rumah*KKS	32,0239	7,815	< 0,0001	Tidak Saling Bebas
Sumber Penerangan Utama RT*KKS	16,9044	5,591	< 0,0001	Tidak Saling Bebas
Luas Lantai Bangunan per Kapita (m ²)*KKS	22,7172	3,481	< 0,0001	Tidak Saling Bebas
Bahan Utama Dinding RT*KKS	77,6666	7,815	< 0,0001	Tidak Saling Bebas
Bahan Utama Lantai RT*KKS	114,101	5,591	< 0,0001	Tidak Saling Bebas
Fasilitas BAB RT*KKS	19,1967	5,591	< 0,0001	Tidak Saling Bebas

Tabel 5. menyajikan hasil uji *Chi-Square* dengan tingkat signifikansi 5% (0,05). Variabel prediktor yang tidak saling bebas dengan variabel respon ditunjukkan oleh nilai *P-Value* yang lebih kecil dari 0,05 atau nilai X^2_{Hitung} yang lebih besar dari X^2_{Tabel} . Dari dua belas variabel prediktor yang dianalisis, sepuluh di antaranya terbukti tidak saling bebas terhadap KKS. Di antara sepuluh variabel tersebut, variabel pendidikan terakhir (X_2) memiliki hubungan tidak saling bebas paling tinggi

terhadap KKS dengan nilai X^2_{hitung} sebesar 546,897. Oleh karena itu, variabel pendidikan terakhir (X_2) dipilih sebagai pemisah simpul utama dalam pembentukan pohon klasifikasi pada *node* 0. Untuk simpul-simpul berikutnya, proses pembentukan pohon klasifikasi dilanjutkan dengan menerapkan algoritma *Exhaustive* CHAID yang sama, melalui tahapan penggabungan dan tahap pemisahan.

Tahap Penghentian

Tahap penghentian dilakukan jika suatu proses pertumbuhan pohon harus dihentikan. Proses pertumbuhan pohon klasifikasi metode *Exhaustive* CHAID terhenti pada *node* 7, *node* 9, *node* 10, *node* 11, *node* 12, *node* 13, *node* 14, *node* 15, *node* 16, dan *node* 17 sehingga menjadi simpul terminal. Berikut pohon klasifikasi yang terbentuk menggunakan metode *Exhaustive* CHAID pada Gambar 1.



Gambar 1. Pohon Klasifikasi *Exhaustive* CHAID

Hasil klasifikasi metode *Exhaustive* CHAID pada data rumah tangga penerima KKS di Jawa Barat tahun 2024 terhadap faktor faktor yang memiliki hubungan paling erat diperoleh bahwa terdapat lima variabel prediktor yang mencirikan rumah tangga penerima KKS yaitu pendidikan terakhir KRT, jumlah jamkes, bahan utama lantai RT, jumlah anggota RT ditanggung serta status kepemilikan rumah. Diagram pohon *Exhaustive* CHAID yang terbentuk pada Gambar 1. menghasilkan 18 *nodes*. 18 *nodes* yang terdiri dari satu simpul induk (*node* 0), dan tujuh belas simpul cabang (*node* 1 sampai *node* 17).

Tabel 6. Segmen dan Presentase KKS

Pelabelan Kelas	Segmen	Karakteristik	Persentase
Menerima KKS	1	Rumah tangga dengan KRT yang memiliki pendidikan terakhirnya dasar, memiliki 1 jamkes atau memiliki ≥ 2 jamkes, dan bahan utama lantai rumah permanen atau tidak permanen.	53,8%
	2	Rumah tangga dengan KRT yang memiliki pendidikan terakhirnya dasar, memiliki 1 jamkes atau memiliki ≥ 2 jamkes, dan bahan utama lantai rumah semi permanen	63,8%
	6	Rumah tangga dengan KRT yang memiliki pendidikan terakhirnya menengah, memiliki 1 jamkes, dan bahan utama lantai rumah semi permanen atau tidak permanen.	52,5%
Tidak Menerima KKS	3	Rumah tangga dengan KRT yang memiliki pendidikan terakhirnya dasar, tidak memiliki jamkes, dan mempunyai ≤ 3 anggota rumah tangga yang ditanggung.	71,5%
	4	Rumah tangga dengan KRT yang memiliki pendidikan terakhirnya dasar, tidak memiliki jamkes, dan mempunyai > 3 anggota rumah tangga yang ditanggung.	58,1%

Pelabelan Kelas	Segmen	Karakteristik	Persentase
	5	Rumah tangga dengan KRT yang memiliki pendidikan terakhirnya menengah, memiliki 1 jamkes, dan bahan utama lantai rumah permanen.	65,6%
	7	Rumah tangga dengan KRT yang memiliki pendidikan terakhirnya menengah, memiliki ≥ 2 jamkes atau tidak memiliki jamkes.	74,5%
	8	Rumah tangga dengan KRT yang memiliki pendidikan terakhirnya tinggi, status rumah nya milik sendiri, kontrak/sewa, atau dinas, dan mempunyai ≤ 3 anggota rumah tangga yang ditanggung.	91,6%
	9	Rumah tangga dengan KRT yang memiliki pendidikan terakhirnya tinggi, status rumah nya milik sendiri, kontrak/sewa, atau dinas, dan mempunyai > 3 anggota rumah tangga yang ditanggung.	85,8%
	10	Rumah tangga dengan KRT yang memiliki pendidikan terakhirnya tinggi, status rumah nya bebas sewa.	69,7%

Tabel 6. diatas dapat diketahui bahwa rumah tangga dengan ciri menerima KKS adalah segmen ke-1, ke-2, dan ke-6. Sedangkan rumah tangga dengan ciri tidak menerima KKS adalah segmen ke-3, ke-4, ke-5, ke-7, ke-8, ke-9, dan ke-10.

Dari hasil segmentasi tersebut, dapat disimpulkan bahwa rumah tangga dengan pendidikan terakhir KRT dasar, memiliki jaminan kesehatan, dan kondisi tempat tinggal dengan bahan utama lantai semi permanen, lebih cenderung diklasifikasikan sebagai menerima KKS dilihat dari presentase terbesar pada pelabelan menerima KKS. Sebaliknya, rumah tangga dengan pendidikan terakhir KRT yang tinggi, status kepemilikan rumah tangga nya jelas, serta jumlah anggota rumah tangga yang relatif sedikit, lebih dominan diklasifikasikan sebagai tidak menerima KKS dilihat dari presentase terbesar pada pelabelan tidak menerima KKS.

Keakuratan Klasifikasi Metode *Exhaustive CHAID*

Keakuratan pohon klasifikasi metode *Exhaustive CHAID* dapat dilihat dapat dilihat dengan menggunakan matriks konfusi. Matriks konfusi digunakan dalam perhitungan akurasi, presisi, *recall*, *F1-Score*, dan juga MCC dalam mengklasifikasikan rumah tangga penerima KKS di Jawa Barat tahun 2024.

Tabel 7. Matriks Konfusi Klasifikasi *Exhaustive CHAID*

Matriks Konfusi		Data Training Prediksi		Data Testing Prediksi	
		Menerima KKS	Tidak Menerima KKS	Menerima KKS	Tidak Menerima KKS
Observasi	Menerima KKS	2488	1894	644	471
	Tidak Menerima KKS	2003	4660	480	1221
Akurasi		0,6472		0,6623	

Matriks Konfusi	Data Training Prediksi		Data Testing Prediksi	
	Menerima KKS	Tidak Menerima KKS	Menerima KKS	Tidak Menerima KKS
Presisi		0,5540		0,5730
Recall		0,5678		0,5776
F1-Score		0,5608		0,5753
MCC		0,2661		0,2950

Tabel 7. menunjukkan perhitungan akurasi, presisi, *recall*, *F1-Score*, dan juga MCC prediksi dari klasifikasi rumah tangga penerima KKS di Jawa Barat tahun 2024 menggunakan Persamaan (6) hingga Persamaan (10). Hasil perhitungan pada data *training*, model mencapai akurasi sebesar 64,72%. Selain akurasi, evaluasi juga memperhatikan nilai presisi, *recall*, dan *F1-score* yang masing-masing sebesar 55,40%, 56,78%, dan 56,08% pada data *training*. Sementara pada data *testing* akurasi meningkat menjadi 66,23%, hal ini menunjukkan bahwa model mampu mempertahankan performanya saat diterapkan pada data baru. Selain akurasi, evaluasi juga memperhatikan nilai presisi, *recall*, dan *F1-score* yang masing-masing sebesar 57,3%, 57,76%, dan 57,53% pada data *testing*, menandakan keseimbangan antara ketepatan dan kelengkapan klasifikasi. Nilai MCC (*Matthews Correlation Coefficient*) sebesar 0,2950 juga menunjukkan adanya korelasi positif, meskipun masih dalam kategori rendah, yang kemungkinan disebabkan oleh ketidakseimbangan distribusi kelas antara menerima KKS dan tidak menerima KKS. Secara keseluruhan, model *Exhaustive CHAID* dinilai cukup efektif dalam mengklasifikasikan rumah tangga penerima KKS.

D. Kesimpulan

Dari hasil segmentasi tersebut, dapat disimpulkan bahwa rumah tangga dengan pendidikan terakhir KRT yang dasar, memiliki jaminan kesehatan, dan kondisi tempat tinggal dengan bahan utama lantai semi permanen lebih cenderung diklasifikasikan sebagai menerima KKS. Sebaliknya, rumah tangga dengan pendidikan terakhir KRT yang tinggi, status kepemilikan rumah nya jelas, serta jumlah tanggungan anggota rumah tangga yang relatif sedikit lebih dominan diklasifikasikan sebagai tidak menerima KKS. Model klasifikasi penerima Kartu Keluarga Sejahtera (KKS) di Provinsi Jawa Barat tahun 2024 dengan metode *Exhaustive CHAID* menunjukkan hasil yang cukup baik. Pada data *training*, model mencapai akurasi 64,72%, dengan presisi, *recall*, dan *F1-score* masing-masing sebesar 55,40%, 56,78%, dan 56,08%. Pada data *testing*, akurasi meningkat menjadi 66,23%, dengan presisi 57,3%, *recall* 57,76%, dan *F1-score* 57,53%, menunjukkan performa yang konsisten dan seimbang. Nilai MCC sebesar 0,2950 mengindikasikan korelasi positif yang masih rendah, kemungkinan akibat ketidakseimbangan kelas. Secara keseluruhan, *Exhaustive CHAID* cukup efektif dalam mengklasifikasikan rumah tangga penerima KKS.

Daftar Pustaka

- Afrillia. (2021). Efektivitas Program Kartu Keluarga Sejahtera (KKS) Kelurahan Sanipah Kecamatan Samboja Kabupaten Kutai Kartanegara. *Ejournal.Ip.Fisip-Unmul.Ac.Id*, 2021(1), 1–10. <https://kaltim.tribunnews.com/2017/08/01/angka-kemiskinan-di-samboja-kukar->
- Ajeng Mega Pratiwi, & Mutaqin, A. K. (2021). Penerapan Algoritma Naïve Bayes Classifier dalam Memprediksi Status Keberlanjutan Polis Nasabah Asuransi PT.X. *Jurnal Riset Statistika*, 1(2), 117–126. <https://doi.org/10.29313/jrs.v1i2.435>

- BPS Jawa Barat. (2024). *Gini Ratio Jawa Barat Maret 2024*. <https://jabar.bps.go.id/id/pressrelease/2024/07/01/1160/gini-ratio-jawa-barat-maret-2024-tercatat-sebesar-0-421.html>
- Fanggidae, J. J. R., Ekowati, C. K., Nenohai, J. M. H., & Udil, P. A. (2021). Klasifikasi Faktor–faktor yang Mempengaruhi Prestasi Akademik Mahasiswa Pendidikan Matematika FKIP Undana dengan Metode CHAID. *FRAKTAL: JURNAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA*, 2(1), 23–33. <https://doi.org/10.35508/FRACTAL.V2I1.4018>
- Fazrilillah, U. N., & Rohaeni, O. (2024). Penerapan Metode Holt’s Exponential Smoothing Dalam Memprediksi Jumlah Nasabah Kredit. *DataMath: Journal of Statistics and Mathematics*, 2(1), 11–16.
- Gallagher, C. A., Monroe, H. M., & Fish, J. L. (2000). *An Iterative Approach To Classification Analysis*.
- Gandhawangi, S. (2023). *Bansos Tidak Tepat Sasaran, Negara Merugi Ratusan Miliar Rupiah Per Bulan - Kompas.id*. <https://www.kompas.id/baca/humaniora/2023/09/06/bansos-tidak-tepat-sasaran-negara-merugi-ratusan-miliar-per-bulan>
- GoodStats Data. (2024). *Jawa Barat Jadi Provinsi dengan Penduduk Terbanyak 2024*. <https://data.goodstats.id/statistic/jawa-barat-jadi-provinsi-dengan-penduduk-terbanyak-2024-ulqWO>
- Hair Jr, J. F., Babin, B. J., Black, W. C., & Anderson, R. E. (2014). *Multivariate Data Analysis* (7th ed.). www.pearsoned.co.uk
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction*.
- Kass, G. V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 29(2), 119–127. <https://doi.org/10.2307/2986296>
- Mandjarreki, S. (2021). Analisis Sosial Kemiskinan. *Jurnal Mimbar Kesejahteraan Sosial*.
- Saputra, D. (2024). *Chi-Square Automatic Interaction Detector (CHAID) untuk Segementasi Pasar – International Business Management*. <https://bbs.binus.ac.id/ibm/2018/05/chi-square-automatic-interaction-detector-chaid-untuk-segmentasi-pasar/>
- Singgalang. (2024). *Rp 227,43 Miliar Dana Bansos tak Tersalurkan*. <https://sumbar.bpk.go.id/temuan-bpk-rp-22743-miliar-dana-bansos-tak-tersalurkan/>
- Tohari, A., Susanti, Y., Tri, D., & Kusmayadi, A. (2017). Penerapan Metode CHAID (Chi-Squared Automatic Interaction Detection) Dan Exhaustive CHAID Pada Klasifikasi Produksi Jagung di Pulau Jawa.
- Zakiyatis Salmaini, & Suliadi. (2021). SPC (Statistical Process Control) Fase II Diagram Kendali Cusum (Cumulative Sum) Nonparametrik Berdasarkan Statistik Mann-Whitney Pada Data Harga Saham PT NO. *Jurnal Riset Statistika*, 1(2), 83–91. <https://doi.org/10.29313/jrs.v1i2.404>