

Klasifikasi Status Kemiskinan Rumah Tangga Berdasarkan Karakteristik Demografi dan Hunian Menggunakan Algoritma *Clasifcation and Regression Tree*

Aji Pandu Winata dkk.



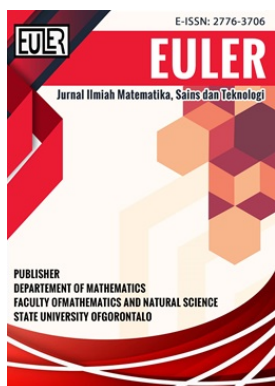
Volume 13, Issue 3, Pages 411–418, Dec. 2025

Diterima 19 Oktober 2025, Direvisi 28 November 2025, Disetujui 1 Desember 2025, Diterbitkan 5 Desember 2025

To Cite this Article : A. P. Winata dkk., “Klasifikasi Status Kemiskinan Rumah Tangga Berdasarkan Karakteristik Demografi dan Hunian Menggunakan Algoritma *Clasifcation and Regression Tree*”, *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 13, no. 3, pp. 411–418, 2025, <https://doi.org/10.37905/euler.v13i3.34952>

© 2025 by author(s)

JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI

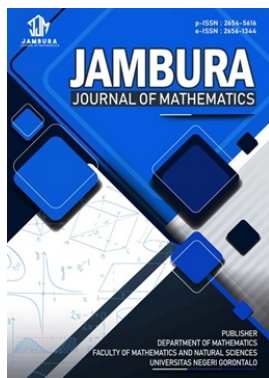


- Homepage : <http://ejournal.ung.ac.id/index.php/euler/index>
- Journal Abbreviation : Euler J. Ilm. Mat. Sains dan Teknol.
- Frequency : Three times a year
- Publication Language : English (preferable), Indonesia
- DOI : <https://doi.org/10.37905/euler>
- Online ISSN : 2776-3706
- License : Creative Commons Attribution-NonCommercial 4.0 International License
- Publisher : Department of Mathematics, Universitas Negeri Gorontalo
- Country : Indonesia
- OAI Address : <http://ejournal.ung.ac.id/index.php/euler/oai>
- Google Scholar ID : QF_r_gAAAAJ
- Email : euler@ung.ac.id

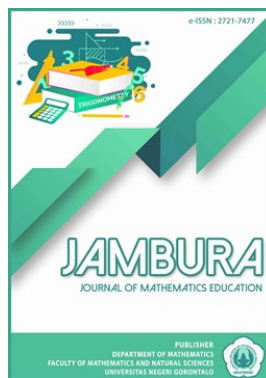
JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



Jambura Journal of Biomathematics



Jambura Journal of Mathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics

Klasifikasi Status Kemiskinan Rumah Tangga Berdasarkan Karakteristik Demografi dan Hunian Menggunakan Algoritma *Classification and Regression Tree*

Aji Pandu Winata¹, Dinda Aprilia¹, Idhia Sriliana^{1,*}, Fitri Aryati², Reny Puspasari²

¹Program Studi Statistika, Universitas Bengkulu, Bengkulu 28122, Indonesia

²Badan Pusat Statistik Provinsi Bengkulu, Bengkulu 38225, Indonesia

ARTICLE HISTORY

Diterima 19 Oktober 2025
Direvisi 28 November 2025
Disetujui 1 Desember 2025
Diterbitkan 5 Desember 2025

KATA KUNCI

Kemiskinan
Rumah Tangga
CART

KEYWORDS

Poverty
Household
CART

ABSTRAK. Kemiskinan merupakan permasalahan multidimensional yang masih menjadi fokus utama pembangunan di Provinsi Bengkulu. Pendekatan berbasis pohon keputusan, khususnya *Classification and Regression Tree (CART)*, dapat digunakan untuk mengklasifikasikan rumah tangga berdasarkan karakteristik demografi dan hunian. Data yang digunakan dalam penelitian ini berasal dari Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2024 dengan variabel respon yang digunakan adalah status rumah tangga dan variabel prediktor, yaitu data yang digunakan berasal dari Survei Sosial Ekonomi Nasional (SUSENAS) 2024 dengan variabel respon berupa status rumah tangga serta enam variabel prediktor yang meliputi tipe daerah, jenis kelamin kepala rumah tangga, umur kepala rumah tangga, jumlah anggota rumah tangga, luas lantai rumah, dan status kepemilikan rumah. Analisis dilakukan melalui tahapan preprocessing data, statistik deskriptif, pembagian data latih, pembangunan model CART, identifikasi variabel yang berpengaruh, dan evaluasi model menggunakan *confusion matrix*. Hasil penelitian menunjukkan bahwa jumlah anggota rumah tangga, luas lantai rumah, dan umur kepala rumah tangga merupakan variabel yang paling berpengaruh dalam membedakan status rumah tangga miskin dan tidak miskin. Evaluasi model menghasilkan *accuracy* sebesar 0,73, *sensitivity* 0,58, dan *specificity* 0,75. Nilai *accuracy* dan *specificity* menunjukkan kinerja klasifikasi yang cukup baik, namun nilai *sensitivity* yang rendah menunjukkan model masih kurang optimal dalam mendeteksi rumah tangga miskin, salah satunya karena ketidakseimbangan proporsi data. Temuan ini menunjukkan bahwa metode CART dapat digunakan untuk analisis kemiskinan di Provinsi Bengkulu, namun perbaikan model diperlukan untuk meningkatkan kemampuan identifikasi rumah tangga miskin.

ABSTRACT. Poverty is a multidimensional issue that remains a central focus of development in Bengkulu Province. Decision-tree-based approaches, particularly the *Classification and Regression Tree (CART)*, can be used to classify households based on demographic and housing characteristics. This study utilizes data from the 2024 National Socio-Economic Survey (SUSENAS), with household poverty status as the response variable and six predictor variables including area type, gender of household head, age of household head, number of household members, floor area of the house, and housing ownership status. The analysis consists of data preprocessing, descriptive statistics, data splitting, CART model construction, identification of influential variables, and model evaluation using a *confusion matrix*. The results show that the number of household members, floor area of the house, and age of the household head are the most influential variables in distinguishing poor and non-poor households. Model evaluation produced an *accuracy* of 0.73, *sensitivity* of 0.58, and *specificity* of 0.75. The *accuracy* and *specificity* values indicate adequate classification performance, while the low *sensitivity* suggests that the model is still less optimal in detecting poor households, partly due to class imbalance in the dataset. These findings indicate that the CART method can be applied to poverty analysis in Bengkulu Province, although further model improvement is needed to enhance its capability in identifying poor households.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. **Editorial of EULER:** Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

1. Pendahuluan

Kemiskinan masih menjadi salah satu permasalahan utama dalam pembangunan ekonomi di Indonesia, termasuk di Provinsi Bengkulu. Berdasarkan data Badan Pusat Statistik (BPS), tercatat pada bulan Maret 2024, persentase penduduk miskin di Provinsi

Bengkulu mencapai 13,56 persen (281,36 ribu jiwa) [1]. Angka ini menjadikan Bengkulu sebagai provinsi dengan tingkat kemiskinan tertinggi kedua di Pulau Sumatera dan menempati peringkat kesepuluh secara nasional.

Secara umum, kemiskinan menggambarkan kondisi ketika individu atau rumah tangga tidak mampu memenuhi kebutuhan dasarnya [2]. Fenomena ini tidak hanya berkaitan dengan ke-

*Penulis Korespondensi.

terbatasan pendapatan, tetapi juga mencakup berbagai dimensi kesejahteraan seperti akses terhadap pendidikan, kesehatan, pekerjaan, dan partisipasi sosial. Dalam pendekatan *capability* yang dikemukakan Amartya Sen, kemiskinan dipahami sebagai keterbatasan kemampuan seseorang untuk menjalani kehidupan yang layak dan bermartabat [3]. BPS mengelompokkan rumah tangga sebagai miskin apabila rata-rata pengeluaran per kapita per bulan berada di bawah garis kemiskinan, sedangkan rumah tangga dengan pengeluaran di atas garis tersebut dikategorikan tidak miskin [1]. Dengan karakteristik sosial ekonomi yang beragam, Provinsi Bengkulu memerlukan strategi pengelolaan pendapatan daerah yang efektif untuk menghadapi tantangan kemiskinan [4]. Analisis kemiskinan pada tingkat rumah tangga menjadi penting untuk memahami faktor-faktor yang memengaruhi kesejahteraan masyarakat secara lebih spesifik.

Pada tingkat rumah tangga, status kemiskinan tidak hanya ditentukan oleh rendahnya pendapatan, tetapi juga dipengaruhi oleh berbagai variabel lain yang berkaitan dengan kondisi demografi dan hunian. Karakteristik demografi, seperti jenis kelamin kepala rumah tangga, usia, jumlah anggota rumah tangga, serta lokasi tempat tinggal, berperan penting dalam menggambarkan kemampuan rumah tangga memenuhi kebutuhannya. Di sisi lain, kondisi hunian, misalnya luas lantai rumah dan status kepemilikan rumah, menjadi indikator nyata kesejahteraan fisik yang membedakan rumah tangga miskin dan tidak miskin [5]. Untuk memahami bagaimana variabel-variabel tersebut berkontribusi dalam membedakan status rumah tangga, diperlukan metode analisis yang mampu mengidentifikasi pola secara jelas dan mudah diinterpretasikan. Salah satu metode yang sering digunakan adalah *Classification and Regression Tree (CART)*.

Metode *CART* adalah salah satu metode klasifikasi berbasis pohon keputusan yang bersifat nonparametrik, sehingga tidak membutuhkan asumsi distribusi tertentu pada data [6]. Prinsip kerja *CART* menggunakan pendekatan *binary recursive partitioning*, yaitu proses pembentukan pohon klasifikasi dengan membagi data secara berulang menjadi dua kelompok yang lebih homogen berdasarkan kriteria pemisah terbaik (*goodness of split criterion*). Keunggulan utama metode ini terletak pada kemampuannya menangani data berukuran besar dengan tipe campuran (numerik dan kategorik), serta menghasilkan model yang mudah dipahami melalui visualisasi struktur pohon keputusan [7]. Dalam analisis sosial ekonomi, termasuk studi kemiskinan, *CART* banyak digunakan untuk melihat variabel-variabel yang paling berpengaruh terhadap status rumah tangga miskin dan tidak miskin. Pendekatan ini mampu menggambarkan hubungan antarvariabel secara nonlinier serta memberikan hasil klasifikasi yang akurat dan mudah diinterpretasikan oleh pengambil kebijakan [8].

Berbagai penelitian terdahulu telah menerapkan metode *CART* untuk menganalisis kemiskinan di berbagai daerah, dan hasilnya menunjukkan bahwa metode ini unggul dibandingkan algoritma pohon klasifikasi lainnya. Penelitian yang dilakukan oleh [6] menggunakan metode *CHAID* dan *CART* untuk mengklasifikasikan status kemiskinan rumah tangga di Kota Semarang berdasarkan data Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2016. Hasil penelitian tersebut menunjukkan bahwa *CART* menghasilkan tingkat akurasi sebesar 95,2%, lebih tinggi dibandingkan *CHAID* dengan akurasi 95,1%, dengan variabel penting seperti bahan bakar utama memasak, penerimaan beras miskin, jumlah

anggota rumah tangga, dan sumber air minum sebagai penentu utama klasifikasi. Selanjutnya, penelitian oleh [8] juga membandingkan performa metode *CART* dengan *Support Vector Machine (SVM)* dalam mengklasifikasikan status kemiskinan rumah tangga di Kabupaten Wonosobo menggunakan data SUSENAS tahun 2018. Meskipun *SVM* menghasilkan akurasi sedikit lebih tinggi, *CART* tetap dianggap lebih unggul dalam aspek interpretasi dan kemudahan penerapan, dengan tingkat akurasi mencapai 89,31% dan variabel signifikan yang meliputi pendidikan kepala rumah tangga, kepemilikan sanitasi, sumber air utama, dan luas lantai rumah. Berdasarkan hasil kedua penelitian tersebut, metode *CART* terbukti mampu memberikan hasil klasifikasi yang baik serta lebih mudah diinterpretasikan dibandingkan metode lain, sehingga sesuai untuk digunakan dalam penelitian ini guna memetakan rumah tangga miskin di Provinsi Bengkulu.

Berdasarkan hasil penelitian sebelumnya, penerapan metode *CART* di Provinsi Bengkulu menjadi penting mengingat tingginya tingkat kemiskinan di wilayah tersebut dan perlunya identifikasi faktor-faktor utama yang memengaruhi status kemiskinan rumah tangga. Dengan mempertimbangkan kondisi tersebut, penelitian ini bertujuan untuk mengidentifikasi variabel-variabel demografi dan hunian yang berpengaruh terhadap status kemiskinan rumah tangga di Provinsi Bengkulu, membangun model klasifikasi status kemiskinan rumah tangga menggunakan algoritma *CART*, serta mengevaluasi performa model yang dihasilkan berdasarkan ukuran *accuracy*, *sensitivity*, dan *specificity*. Hasil penelitian ini diharapkan mampu mengidentifikasi secara jelas variabel-variabel utama yang berkontribusi terhadap kemiskinan rumah tangga di Provinsi Bengkulu. Dengan demikian, harapannya penelitian ini dapat menjadi acuan bagi pemerintah daerah dalam menyusun kebijakan pengentasan kemiskinan berbasis data.

2. Metode

2.1. Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini berasal dari Survei Sosial Ekonomi Nasional (SUSENAS) periode Maret 2024, yang mencakup seluruh sampel rumah tangga di Provinsi Bengkulu sebanyak 5.773 rumah tangga. Dalam penelitian ini, variabel respon yang digunakan adalah status rumah tangga dan variabel prediktor, yaitu tipe daerah, jenis kelamin kepala rumah tangga, umur kepala rumah tangga, jumlah anggota rumah tangga, luas lantai rumah, dan status kepemilikan rumah. Adapun keterangan untuk masing-masing variabel disajikan pada Tabel 1.

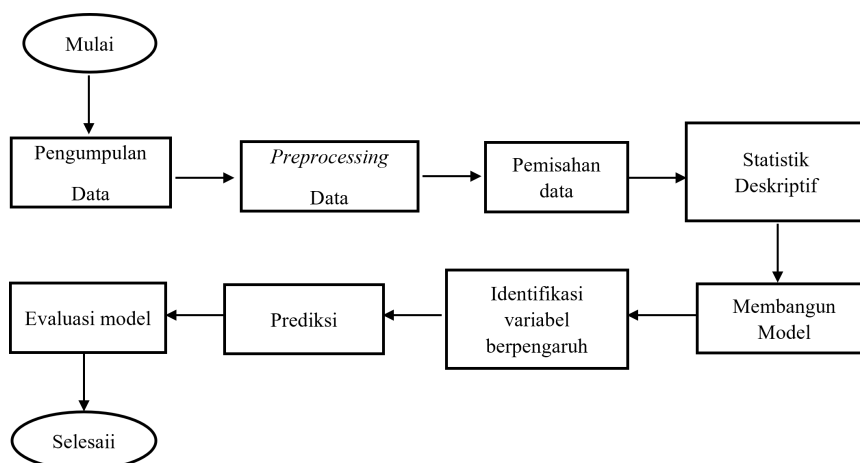
2.2. Prosedur Penelitian

2.2.1. Tahapan Penelitian

Penelitian ini menggunakan bantuan program *R* dalam analisis *Classification and Regression Tree (CART)* untuk mendapatkan variabel-variabel yang berpengaruh dalam memprediksi status rumah tangga miskin di Provinsi Bengkulu. Tahapan penelitian dalam penelitian ini dilakukan melalui beberapa tahapan. Tahap awal yaitu *preprocessing* data untuk memastikan data bersih dan siap dianalisis. Selanjutnya dilakukan statistik deskriptif untuk menggambarkan kondisi rumah tangga di Provinsi Bengkulu. Data kemudian dibagi menjadi data *training* dan data *testing* untuk proses pelatihan dan pengujian model. Model klasifikasi dibangun menggunakan algoritma *CART*, diikuti dengan identifikasi variabel yang berpengaruh terhadap status kemiskinan. Ta-

Tabel 1. Variabel penelitian

Variabel	Simbol	Tipe	Kategori	Keterangan
Status Rumah Tangga	Y	Kategorik	1	Miskin
			2	Tidak miskin
Tipe daerah	X_1	Kategorik	1	Pedesaan
			2	Perkotaan
Jenis kelamin kepala rumah tangga	X_2	Kategorik	1	Laki-laki
			2	Perempuan
Umur kepala rumah tangga	X_3	Numerik	–	–
Jumlah anggota rumah tangga	X_4	Numerik	–	–
Luas lantai rumah	X_5	Numerik	–	–
Status Kepemilikan Rumah	X_6	Kategorik	1	Milik sendiri
			2	Kontrak/sewa
			3	Dinas
			4	Lainnya



Gambar 1. Tahapan penelitian

hap berikutnya adalah prediksi status kemiskinan rumah tangga, dan diakhiri dengan evaluasi model menggunakan ukuran seperti *accuracy*, *sensitivity*, dan *specificity* untuk menilai kemampuan model dalam memprediksi status kemiskinan rumah tangga di Provinsi Bengkulu. Tahapan penelitian dapat dilihat pada Gambar 1.

2.2.2. Classification and Regression Tree (CART)

CART adalah salah satu metode klasifikasi yang berbasis pohon keputusan. Metode ini dapat diterapkan pada *dataset* yang besar, di mana variabel yang digunakan dapat bertipe numerik maupun kategorik [9]. CART memiliki keunggulan dibandingkan metode lainnya, yaitu hasilnya lebih mudah diinterpretasikan, akurat, dan komputasinya cepat diterapkan untuk *dataset* yang observasinya besar, variabel yang banyak, serta kombinasi variabel numerik dan kategorik dapat dianalisis menggunakan pemisahan data secara biner [10]. Algoritma ini bekerja dengan prinsip *binary recursive partitioning*, yaitu membagi data secara berulang ke dalam kelompok yang lebih homogen berdasarkan variabel pemilah (*splitting variable*) yang paling optimal [7].

Algoritma CART telah digunakan secara luas dalam berbagai disiplin ilmu. [11] menjelaskan bahwa CART mampu memilih variabel terbaik secara otomatis dari sejumlah prediktor dan menghasilkan struktur pohon yang mudah diinterpretasikan. Sementara itu, [12] menunjukkan penerapan CART dalam regresi untuk memperkirakan variabel kontinu seperti kadar air tanah berdasarkan faktor meteorologi. Pendekatan ini dapat menangani data yang bersifat nonlinier, baik untuk klasifikasi dengan res-

pon kategorik maupun prediksi dengan respon numerik. Menurut [13], tahapan utama dalam algoritma CART mencakup proses *pre-processing* data, pembentukan model melalui teknik pemilahan biner, serta evaluasi menggunakan *confusion matrix* dan *cross validation* untuk memastikan kestabilan model. Pendekatan ini menekankan pentingnya penentuan variabel pemilah (*splitting variable*) yang optimal untuk menghasilkan simpul-simpul yang homogen dan mengurangi kesalahan klasifikasi pada tahap akhir. Selain itu, [14] juga menegaskan bahwa pemangkasan pohon (*pruning*) menjadi langkah penting untuk menghindari *overfitting* dan menjaga keseimbangan antara kompleksitas model dan kemampuan generalisasinya pada data baru. Sedangkan kekurangan dari metode CART yaitu membutuhkan data pelatihan dalam jumlah besar dan cenderung kurang berkinerja baik pada data mentah [15].

Berikut adalah langkah-langkah dalam pemodelan CART dalam penelitian ini:

1. Pembagian Data
Dataset dibagi menjadi data *training* dan data *testing*.
2. Pemilihan Parameter Optimal
Berbagai kombinasi parameter, seperti nilai *cost*, kedalaman pohon, dan jumlah minimum data pada setiap simpul, diuji untuk mencari pengaturan yang menghasilkan performa model paling baik.
3. Pembentukan Model
Model CART dibangun menggunakan data *training*, di mana pada setiap percabangan algoritma akan memilih variabel

yang paling baik untuk memisahkan kelompok miskin dan tidak miskin.

4. Pemangkasan Pohon (*Pruning*)
Pohon dipangkas untuk mencegah *overfitting* dan memastikan model stabil ketika diterapkan pada data baru.
5. Menentukan Variabel yang Berpengaruh
Algoritma secara otomatis memilih variabel yang paling berpengaruh dengan melihat pola pemisahan terbaik dalam data.
6. Prediksi Status Rumah Tangga
Model akhir digunakan untuk memprediksi apakah sebuah rumah tangga masuk kategori miskin atau tidak pada data *testing*.
7. Evaluasi Model
Hasil prediksi dievaluasi menggunakan *confusion matrix* dengan ukuran *accuracy*, *sensitivity*, dan *specificity*.

2.2.3. Evaluasi Model CART

Sebelum melakukan evaluasi model *CART*, terlebih dahulu dibentuk *confusion matrix*, yaitu tabel hasil klasifikasi yang berfungsi untuk mengevaluasi kinerja suatu model klasifikasi. Melalui *confusion matrix* dapat dihitung berbagai metrik evaluasi. Secara umum, metrik yang diperoleh dari *confusion matrix* terbagi menjadi dua, yaitu nilai akurasi dan tingkat kesalahan (*error*) [16]. Ukuran performa model klasifikasi menggunakan *confusion matrix* [17], disajikan pada **Tabel 2**.

Tabel 2. *Confusion Matrix*

	Actual Positive	Actual Negative
Predict Positive	True Positive (TP)	False Positive (FP)
Predict Negative	False Negative (FN)	True Negative (TN)

Ukuran yang dapat digambarkan menggunakan matriks di atas adalah sebagai berikut:

1. *Accuracy* (Akurasi)
Menunjukkan rasio antara jumlah prediksi yang benar (positif maupun negatif) terhadap total keseluruhan data.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

2. *Recall* (Sensitivitas)
Menunjukkan rasio antara jumlah prediksi benar positif terhadap keseluruhan data yang tergolong benar positif.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

3. *Specificity* (Spesifisitas)
Menunjukkan tingkat ketepatan model dalam memprediksi kelas negatif dibandingkan dengan seluruh data yang sebenarnya termasuk kategori negatif.

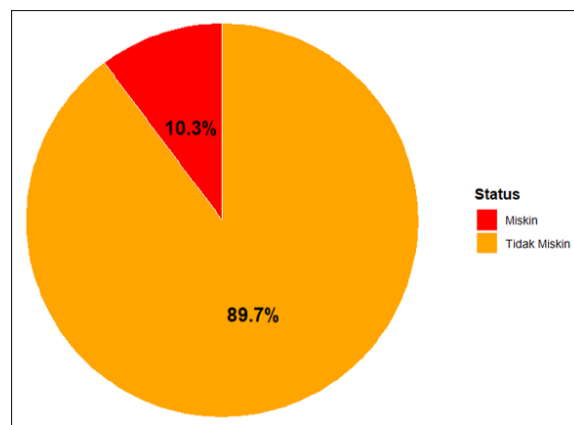
$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

3. Hasil dan Pembahasan

3.1. Statistik Deskriptif Data dan Uji Statistik

Tahapan *preprocessing* data dilakukan dengan proses pemeriksaan *missing value* pada data SUSENAS yang telah diolah. Pada data ini tidak ditemukan sampel rumah tangga yang memiliki

missing value, oleh karena itu dapat dilanjutkan ke tahapan analisis selanjutnya. Pengklasifikasian rumah tangga miskin dan tidak miskin dilakukan dengan membandingkan pengeluaran per kapita dengan garis kemiskinan Provinsi Bengkulu bulan Maret tahun 2024 sebesar Rp 671.095. Jika pengeluaran per kapita < garis kemiskinan, maka rumah tangga dikategorikan miskin dan demikian pula sebaliknya. Sampel terdiri dari 5.773 rumah tangga, dengan 553 rumah tangga terkategori miskin dan 5.180 rumah tangga terkategori tidak miskin. Proporsi rumah tangga miskin dan tidak miskin disajikan pada **Gambar 2**.



Gambar 2. Proporsi status kemiskinan rumah tangga

Berdasarkan **Gambar 2**, proporsi antara rumah tangga yang terkategori miskin jauh lebih kecil dibandingkan dengan proporsi rumah tangga tidak miskin, hal ini sering disebut sebagai kondisi data yang tidak seimbang (*imbalanced data*). Kemudian, untuk melihat apakah variabel prediktor memiliki hubungan dengan variabel respon digunakan uji *t-student* untuk variabel yang bertipe numerik dan uji *chi-square* untuk variabel yang bertipe kategorik. Hasil pengujian masing-masing disajikan pada **Tabel 3** dan **Tabel 4**.

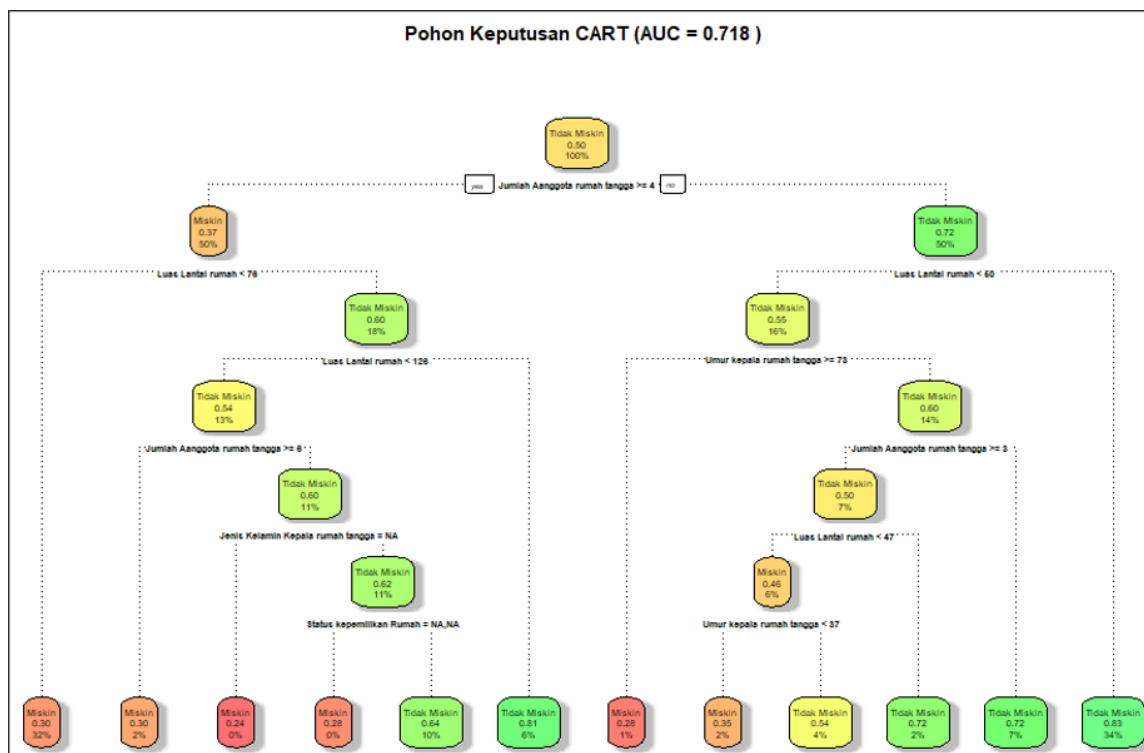
Tabel 3. Uji *Chi-Square* variabel kategorik

Variabel	p-value	Hasil
Tipe daerah (X_1)	0,0011	Signifikan
Jenis kelamin kepala rumah tangga (X_2)	0,0459	Signifikan
Status kepala rumah tangga (X_6)	0,0136	Signifikan

Tabel 4. Uji *t-student* variabel numerik

Variabel	p-value	Hasil
Umur kepala rumah tangga (X_3)	$2,092 \times 10^{-7}$	Signifikan
Jumlah anggota rumah tangga (X_4)	$2,2 \times 10^{-16}$	Signifikan
Luas lantai rumah (X_5)	$2,2 \times 10^{-16}$	Signifikan

Berdasarkan **Tabel 3** dan **Tabel 4**, seluruh variabel prediktor menunjukkan hasil yang signifikan. Hal ini berarti bahwa semua variabel tersebut layak digunakan dalam proses pemodelan menggunakan algoritma *Classification and Regression Tree (CART)*, karena masing-masing berpotensi memberikan kontribusi terhadap hasil klasifikasi status kemiskinan rumah tangga. Dengan demikian, penelitian ini dapat dilanjutkan ke tahap pembangunan model untuk membentuk struktur *decision tree* yang akan digunakan dalam proses klasifikasi selanjutnya.



Gambar 3. Model CART dengan parameter terbaik

3.2. *Pemodelan CART*

Sebelum melakukan pemodelan *Classification and Regression Tree (CART)*, dataset dibagi menjadi dua bagian, yaitu data pelatihan (*training data*) dan data pengujian (*testing data*). Data pelatihan digunakan untuk membangun model pembelajaran, sedangkan data pengujian berfungsi untuk mengevaluasi kinerja model yang dihasilkan. Proses pembagian dataset dilakukan secara acak dengan proporsi 80%:20%. Gambaran pembagian data *training* dan data *testing* disajikan pada Tabel 5.

Tabel 5. Pembagian data untuk model CART

Variabel	Total	Rumah Tangga Miskin	Rumah Tangga Tidak Miskin
Training	4.618	469	4.149
Testing	1.155	124	1.031

Pemodelan *CART* dimulai dengan pencarian nilai parameter optimal, yang terdiri atas *cost complexity*, kedalaman maksimum pohon keputusan, serta jumlah minimum observasi dalam setiap simpul. Kombinasi parameter yang diuji kemudian dibandingkan untuk memperoleh performa klasifikasi paling optimal. Beberapa kombinasi parameter yang menghasilkan performa klasifikasi paling baik disajikan pada Tabel 6.

Tabel 6. Kombinasi parameter optimal model CART

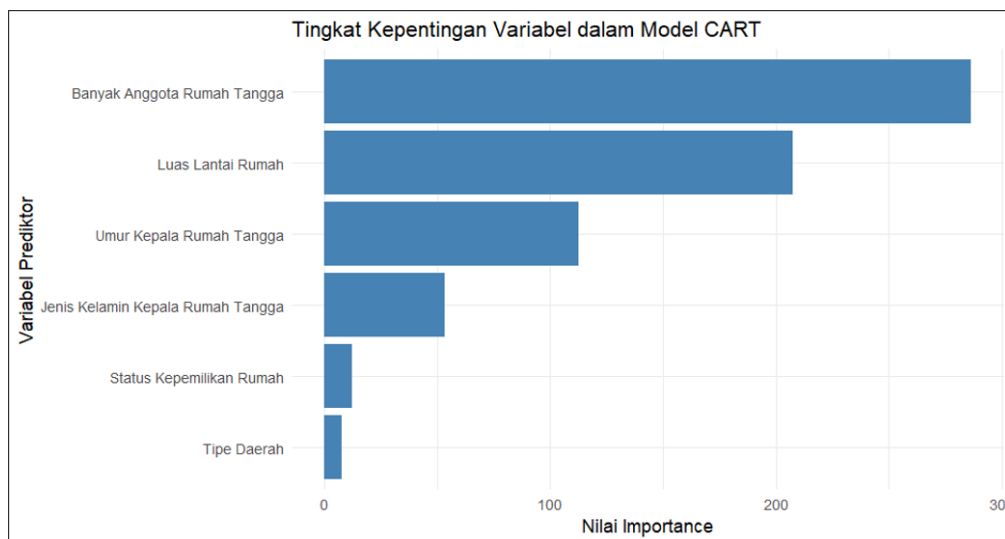
No	Cost Complexity	Kedalaman Pohon	Jumlah Minimum Observasi	AUC
1	0.002	15	8	0.68
2	0.003	20	10	0.72
3	0.004	25	12	0.71
4	0.005	30	13	0.70

Kombinasi parameter terbaik dapat dilihat dari nilai AUC (*Area Under the Curve*) yang merupakan ukuran luas di bawah kurva

ROC (Receiver Operating Characteristic) dan menunjukkan kemampuan model dalam membedakan antara kelas positif dan negatif. Semakin besar nilai AUC (mendekati 1), maka semakin baik kemampuan model dalam melakukan klasifikasi, sedangkan nilai AUC yang mendekati 0,5 menunjukkan bahwa model tidak lebih baik dari prediksi acak. Berdasarkan Tabel 6 kombinasi parameter terbaik yang memberikan nilai AUC terbesar diperoleh pada nilai parameter *cost complexity*, kedalaman pohon, dan jumlah minimum observasi masing-masing sebesar 0,003, 20, dan 10. Berdasarkan kombinasi parameter terbaik tersebut, diperoleh model *CART* yang disajikan pada Gambar 3

Berdasarkan Gambar 3, model *CART* dibangun dengan menggunakan parameter terbaik, yaitu *cost complexity*, kedalaman pohon, dan jumlah minimum observasi berturut-turut sebesar 0,003, 20, dan 10. Hasil pemodelan menunjukkan bahwa jumlah anggota rumah tangga (X_4) merupakan variabel yang paling berpengaruh karena menjadi variabel *split* pertama dalam pembentukan pohon. Rumah tangga dengan jumlah anggota < 4 orang memiliki proporsi tidak miskin sebesar 72%, sedangkan rumah tangga dengan jumlah anggota ≥ 4 orang memiliki proporsi miskin sebesar 37%.

Pada cabang kiri, yaitu jumlah anggota rumah tangga ($X_4 < 4$), percabangan ditentukan oleh luas lantai rumah (X_5). Rumah dengan luas lantai $< 78 \text{ m}^2$ banyak masuk kategori miskin sebesar 30%, sedangkan rumah dengan luas lantai $\geq 78 \text{ m}^2$ cenderung tidak miskin sebesar 60%. Pada cabang kanan, yaitu rumah tangga dengan jumlah anggota ($X_4 \geq 4$), percabangan utama ditentukan oleh luas lantai rumah 60 m^2 . Rumah dengan luas lantai $\geq 60 \text{ m}^2$ memiliki proporsi tidak miskin sebesar 72%, sedangkan rumah dengan luas lantai $< 60 \text{ m}^2$ lebih banyak tergolong miskin (55%). Pada kelompok luas lantai rumah ($X_5 < 60$), percabangan selanjutnya dipengaruhi oleh umur kepala rumah tangga (X_3). Ke-



Gambar 4. Tingkat kepentingan variabel klasifikasi status kemiskinan

Tabel 7. Prediksi kelas miskin

Status Rumah Tangga (Y)	Prediksi Status Rumah Tangga (Y)	Tipe daerah (X_1)	Jenis Kelamin Kepala RT (X_2)	Umur Kepala Rumah Tangga (X_3)	Banyaknya Anggota RT (X_4)	Luas Lantai Rumah (X_5)	Status Kepemilikan Rumah (X_6)
Tidak	Tidak	2	1	37	5	45	1
Tidak	Tidak	2	2	56	4	30	1
Tidak	Tidak	2	1	36	4	36	1
Tidak	Tidak	2	1	47	4	30	1
Tidak	Tidak	2	1	30	4	35	3

Tabel 8. Prediksi kelas tidak miskin

Status Rumah Tangga (Y)	Prediksi Status Rumah Tangga (Y)	Tipe daerah (X_1)	Jenis Kelamin Kepala RT (X_2)	Umur Kepala Rumah Tangga (X_3)	Banyaknya Anggota RT (X_4)	Luas Lantai Rumah (X_5)	Status Kepemilikan Rumah (X_6)
Tidak	Tidak	2	1	59	2	54	1
Tidak	Tidak	1	1	61	2	120	1
Tidak	Tidak	1	1	21	1	36	2
Tidak	Tidak	1	1	37	2	149	1
Tidak	Tidak	1	1	63	2	120	1

pala rumah tangga yang berumur ≥ 73 tahun memiliki proporsi miskin sebesar 28%, sedangkan yang berumur lebih muda umumnya tidak miskin. Tingkat kepentingan variabel prediktor dalam mengklasifikasikan rumah tangga, disajikan pada Gambar 4.

Berdasarkan Gambar 4, dapat diketahui bahwa variabel jumlah anggota rumah tangga (X_4), luas lantai rumah (X_5), dan umur kepala rumah tangga (X_3) secara berturut-turut memiliki kontribusi terbesar dalam klasifikasi rumah tangga di Provinsi Bengkulu. Selanjutnya, dilakukan analisis terhadap hasil klasifikasi yang diperoleh dari model CART menggunakan data testing. Sebagai contoh, disajikan lima observasi pada masing-masing kelas hasil prediksi, yaitu kelas miskin dan tidak miskin, untuk menunjukkan kemampuan model dalam membedakan kedua kategori tersebut, pada Tabel 7 dan Tabel 8.

Berdasarkan Tabel 7 dan Tabel 8, karakteristik rumah tangga yang tergolong dalam kategori miskin umumnya memiliki jumlah anggota rumah tangga (X_4) yang relatif banyak, yaitu antara 4 sampai 6 orang, serta luas lantai rumah (X_5) yang cenderung sempit, berkisar antara 30 m² hingga 45 m². Selain itu, umur kepala rumah tangga (X_3) pada kelompok miskin berada pada rentang usia 30 hingga 56 tahun dan tipe daerah (X_1) berada di

wilayah pedesaan.

Sementara itu, rumah tangga yang diklasifikasikan sebagai tidak miskin memiliki karakteristik yang berbeda. Rumah tangga tidak miskin umumnya memiliki jumlah anggota rumah tangga yang lebih sedikit, yaitu antara satu hingga tiga orang, dengan luas lantai rumah yang lebih besar, berkisar antara 54 m² hingga 149 m². Kepala rumah tangga pada kelompok ini memiliki rentang umur antara 21 hingga 63 tahun, dan sebagian besar tinggal di daerah perkotaan.

3.3. Evaluasi Model CART

Evaluasi model CART dilakukan dengan menggunakan confusion matrix. Confusion matrix digunakan untuk membandingkan hasil klasifikasi model dengan data aktual, sehingga dapat diketahui sejauh mana model CART mampu mengklasifikasikan status kemiskinan rumah tangga secara tepat. Confusion matrix hasil klasifikasi model CART disajikan pada Tabel 9.

Berdasarkan Tabel 9, confusion matrix digunakan untuk mengevaluasi kinerja model klasifikasi dalam membedakan dua kelas, yaitu Rumah Tangga Miskin dan Rumah Tangga Tidak Miskin. Berdasarkan hasil, model menunjukkan kemampuan yang

Tabel 9. Confusion matrix model CART

Klasifikasi	Aktual	
	Rumah Tangga Miskin	Rumah Tangga Tidak Miskin
Predicted Rumah Tangga Miskin	72	260
Predicted Rumah Tangga Tidak Miskin	52	771

bervariasi untuk kedua kelas tersebut. Model berhasil mengidentifikasi 72 kasus Rumah Tangga Miskin dan 771 kasus Rumah Tangga Tidak Miskin secara benar. Namun, masih terdapat sejumlah kesalahan klasifikasi, yakni 52 kasus Rumah Tangga Miskin yang salah diklasifikasikan sebagai Rumah Tangga Tidak Miskin dan 260 kasus Rumah Tangga Tidak Miskin yang salah diprediksi sebagai Rumah Tangga Miskin. Berdasarkan *confusion matrix* tersebut, dapat dihitung beberapa ukuran evaluasi model seperti *accuracy*, *sensitivity*, dan *specificity*. Hasil ukuran evaluasi model disajikan pada Tabel 10.

Tabel 10. Evaluasi model CART

Metrik Evaluasi Klasifikasi	Nilai
<i>Accuracy</i>	0,73
<i>Sensitivity</i>	0,58
<i>Specificity</i>	0,75

Berdasarkan Tabel 10, nilai *accuracy* sebesar 0,73 menunjukkan bahwa model CART yang digunakan sudah cukup baik dalam mengklasifikasikan status kemiskinan rumah tangga di Provinsi Bengkulu. Artinya, sekitar 73% rumah tangga berhasil diprediksi dengan benar, baik kategori miskin maupun tidak miskin. Nilai *specificity* sebesar 0,75 menunjukkan bahwa model cukup akurat dalam mengenali rumah tangga tidak miskin, yaitu sebesar 75%. Sementara itu, nilai *sensitivity* hanya sebesar 0,58, yang berarti model masih kurang baik dalam mengenali rumah tangga miskin karena hanya mampu mengidentifikasi rumah tangga miskin sebesar 58,1% dari total rumah tangga miskin. Ketidakmampuan model ini dapat disebabkan oleh ketidakseimbangan data, di mana proporsi sampel rumah tangga tidak miskin jauh lebih besar dibandingkan rumah tangga miskin, sehingga model lebih mudah memprediksi kelas tidak miskin dibandingkan miskin.

Jika dibandingkan dengan penelitian terdahulu, performa model ini memang masih lebih rendah. Penelitian yang dilakukan oleh Ispriyanti dkk [6] di Kota Semarang memperoleh akurasi 95,20% dengan metode CART, sementara penelitian Nuzula dkk [8] di Wonosobo mendapatkan akurasi 89,31%. Selain akurasi yang lebih tinggi, nilai *sensitivity* pada penelitian-penelitian tersebut juga lebih baik dalam mendeteksi rumah tangga miskin. Perbedaan ini dapat terjadi karena karakteristik sosial ekonomi tiap daerah berbeda, kualitas data tidak sama, dan variabel prediktor yang digunakan juga berbeda.

Dengan demikian, hasil penelitian ini tetap sejalan dengan studi terdahulu bahwa CART dapat digunakan untuk klasifikasi kemiskinan. Namun, performa model sangat dipengaruhi oleh komposisi data dan variabel yang digunakan, terutama dalam meningkatkan *sensitivity* agar model lebih mampu mengenali rumah tangga miskin. Secara keseluruhan, model CART yang dibangun sudah cukup baik, tetapi masih perlu perbaikan pada kemampuan mendeteksi rumah tangga miskin.

4. Kesimpulan

Berdasarkan hasil penelitian menunjukkan bahwa metode *Classification and Regression Tree (CART)* cukup baik digunakan untuk mengklasifikasikan status kemiskinan rumah tangga di Provinsi Bengkulu berdasarkan data SUSENAS tahun 2024. Model yang dibangun menghasilkan nilai *accuracy* sebesar 0,73, *sensitivity* sebesar 0,58, dan *specificity* sebesar 0,75. Hasil ini menunjukkan bahwa model sudah cukup baik secara keseluruhan, meskipun kemampuan dalam mendeteksi rumah tangga miskin masih perlu ditingkatkan. Variabel jumlah anggota rumah tangga (X_4), luas lantai rumah (X_5), dan umur kepala rumah tangga (X_3) menjadi faktor yang paling berpengaruh dalam membedakan rumah tangga miskin dan tidak miskin. Temuan ini menegaskan bahwa karakteristik demografi dan kondisi fisik hunian berperan penting dalam menentukan kesejahteraan rumah tangga. Hasil penelitian ini diharapkan dapat menjadi acuan bagi pemerintah daerah dalam merancang kebijakan penanggulangan kemiskinan yang lebih tepat sasaran dan berbasis data.

Kontribusi Penulis. Aji Pandu Winata: konseptualisasi, metodologi, analisis formal, perangkat lunak, visualisasi, dan penulisan–persiapan draf awal. Dinda Aprilia: kurasi data, investigasi, metodologi, dan penulisan–persiapan draf awal. Idhia Sriliana: supervisi, validasi, penulisan–tinjauan dan penyuntingan. Fitri Aryati: investigasi dan sumber daya. Reny Puspasari: investigasi dan sumber daya. Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

Ucapan Terima Kasih. Penulis mengucapkan terima kasih kepada Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Bengkulu, atas dukungan akademik dan fasilitas yang diberikan selama proses penelitian. Ucapan terima kasih juga disampaikan kepada Badan Pusat Statistik (BPS) Provinsi Bengkulu atas penyediaan data SUSENAS 2024. Penulis turut menghargai kontribusi berbagai pihak yang telah memberikan masukan dan dukungan dalam penyelesaian artikel ini.

Pembiayaan. Penelitian ini tidak menerima pendanaan dari pihak eksternal.

Konflik Kepentingan. Para penulis menyatakan tidak ada konflik kepentingan yang terkait dengan artikel ini.

Ketersediaan Data. Tidak tersedia.

Referensi

- [1] Badan Pusat Statistik Provinsi Bengkulu, "Profil Kemiskinan Provinsi Bengkulu: Maret 2024 (Berita Resmi Statistik No. 41/07/17/Th. XVIII)," BPS Provinsi Bengkulu, Jul. 1, 2024.
- [2] L. Priseptian and W. P. Primandhana, "Analisis Faktor-Faktor yang Mempengaruhi Kemiskinan," *FORUM EKONOMI: Jurnal Ekonomi, Manajemen dan Akuntansi*, vol. 24, no. 1, Jan. 2022, doi: 10.30872/jfor.v24i1.10362.
- [3] M. Adon and G. T. Jeraman, "Masyarakat Madani: Kontribusi Teori Kemiskinan sebagai Deprivasi Kapabilitas dari Amartya Sen dalam Upaya Pemberdayaan Masyarakat Miskin," *Jurnal Masyarakat Madani*, vol. 8, no. 1, 2023.

- [4] Z. A. Pradana and E. Pasaribu, "Pengaruh Pendidikan, PDRB, dan Pendapatan Asli Daerah Terhadap Tingkat Kemiskinan di Provinsi Bengkulu," *Oikos-Nomos Jurnal Kajian Ekonomi dan Bisnis*, vol. 16, no. 2, pp. 55–62, Dec. 2023. doi: [10.37479/jkeb.v16i2.23072](https://doi.org/10.37479/jkeb.v16i2.23072).
- [5] D. Fitriza, A. A. Putra, D. Vionanda, and Zilrahmi, "Classification of Poor Households in West Sumatra Province using Decision Tree Algorithm C4.5," *UNP Journal of Statistics and Data Science*, vol. 2, no. 2, pp. 145–150, May 2024. doi: [10.24036/ujsds/vol2-iss2/157](https://doi.org/10.24036/ujsds/vol2-iss2/157).
- [6] D. Ispriyanti, A. Prahutama, M. Mustafid, and T. Tarno, "Klasifikasi Penerimaan Beras Miskin Di Kota Semarang Menggunakan Algoritma Chisquare Automatic Interaction Detection (CHAID) Dan Classification And Regression Tree (CART)," *MEDIA STATISTIKA*, vol. 12, no. 1, pp. 63–72, Jul. 2019. doi: [10.14710/medstat.12.1.63-72](https://doi.org/10.14710/medstat.12.1.63-72).
- [7] N. I. Prabawati, Widodo, and H. Ajie, "Kinerja Algoritma Classification And Regression Tree (CART) dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta," *PINTER: Jurnal Pendidikan Teknik Informatika dan Komputer*, vol. 3, no. 2, pp. 139–145, Dec. 2019. doi: [10.21009/pinter.3.2.9](https://doi.org/10.21009/pinter.3.2.9).
- [8] L. Nuzula, A. Prahutama, and A. R. Hakim, "Klasifikasi Status Kemiskinan Rumah Tangga dengan Metode Support Vector Machines (SVM) dan Classification and Regression Trees (CART) Menggunakan GUI R (Studi Kasus di Kabupaten Wonosobo Tahun 2018)," *Jurnal Gaussian*, vol. 9, no. 4, pp. 525–534, Dec. 2020. doi: [10.14710/j.gauss.v9i4.29449](https://doi.org/10.14710/j.gauss.v9i4.29449).
- [9] D. A. Wulandari, D. Kusnandar, and Y. Intisari, "Bagging Classification Trees Untuk Klasifikasi Tingkat Kesejahteraan Rumah Tangga Di Kalimantan Barat," *Buletin Ilmiah Math. Stat. dan Terapannya (BIMASTER)*, vol. 8, no. 4, 2019. doi: [10.26418/bbimst.v8i4.36035](https://doi.org/10.26418/bbimst.v8i4.36035).
- [10] H. D. Darmawan, D. Yuniarti, and Y. N. Nasution, "Klasifikasi Lama Masa Studi Mahasiswa Menggunakan Perbandingan Metode Algoritma C4.5 dan Algoritma Classification and Regression Tree," *Jurnal EKSPONENSIAL*, vol. 8, no. 2, 2017.
- [11] Z. Cheng *et al.*, "Evaluation of Classification and Regression Tree (CART) Model in Weight Loss Prediction Following Head and Neck Cancer Radiation Therapy," *Advances in Radiation Oncology*, vol. 3, no. 3, pp. 346–355, Jul. 2018. doi: [10.1016/j.adro.2017.11.006](https://doi.org/10.1016/j.adro.2017.11.006).
- [12] T. H. Wu, P. Y. Chen, C. C. Chen, M. J. Chung, Z. K. Ye, and M. H. Li, "Classification and Regression Tree (CART)-based Estimation of Soil Water Content Based on Meteorological Inputs," *Agricultural Water Management*, vol. 299, Jun. 2024. doi: [10.1016/j.agwat.2024.108869](https://doi.org/10.1016/j.agwat.2024.108869).
- [13] P. Subarkah, "Penerapan Algoritme Klasifikasi Classification And Regression Trees (CART) Untuk Diagnosis Penyakit Diabetes Retinopathy," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 294–301, May 2020. doi: [10.30812/matrik.v19i2.676](https://doi.org/10.30812/matrik.v19i2.676).
- [14] F. S. Aguiar, L. L. Almeida, A. Ruffino-Netto, A. L. Kritski, F. C. Q. Mello, and G. L. Werneck, "Classification and Regression Tree (CART) Model to Predict Pulmonary Tuberculosis in Hospitalized Patients," *BMC Pulmonary Medicine*, vol. 12, Aug. 2012. doi: [10.1186/1471-2466-12-40](https://doi.org/10.1186/1471-2466-12-40).
- [15] U. Hasanah and A. Nurhopipah, "Prediksi Persentase Body Fat Menggunakan Algoritma CART dan M5'," *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 4, pp. 351–363, Feb. 2023. doi: [10.35746/jtim.v4i4.316](https://doi.org/10.35746/jtim.v4i4.316).
- [16] R. R. Adhitya, W. Witanti, and R. Yuniarti, "Perbandingan Metode CART Dan Naïve Bayes Untuk Klasifikasi Customer Churn," *INFOTECH Journal*, vol. 9, no. 2, pp. 307–318, Jul. 2023. doi: [10.31949/infotech.v9i2.5641](https://doi.org/10.31949/infotech.v9i2.5641).
- [17] A. Mahmud, A. Pangestika, A. P. Ramadhanty, G. M. Putra, G. S. N. D. S. Putri, and R. Nooraeni, "Klasifikasi Status Desa/Kelurahan DIY (Yogyakarta) Menggunakan Model Decision Tree," *Engineering, Mathematics and Computer Science (EMACS) Journal*, vol. 3, no. 1, pp. 33–41, Feb. 2021. doi: [10.21512/emasjournal.v3i1.6787](https://doi.org/10.21512/emasjournal.v3i1.6787).