

KLASIFIKASI STATUS KESEJAHTERAAN MASYARAKAT KABUPATEN KEPULAUAN MENTAWAI DENGAN METODE REGRESI LOGISTIK BINER DAN CLASSIFICATION AND REGRESSION TREE (CART)

Classification of Welfare Status of The Mentawai Islands District Community Using Binary Logistic Regression and Classification and Regression Tree (CART) Methods

Lulu Anata Sanur^{1*}, M. Al Haris², Fatkhurokhman Fauzi³

^{1,2,3}Program Studi Statistika, Universitas Muhammadiyah Semarang
Alamat Jl. Kedungmundu No.18, Kedungmundu, Kec. Tembalang, Kota Semarang, Jawa Tengah, Indonesia

E-mail Correspondence Author: lulu.anataa@gmail.com

Abstrak

Kabupaten Kepulauan Mentawai merupakan salah satu daerah yang saat ini masih berstatus sebagai daerah tertinggal, ditandai dengan minimnya ketersediaan fasilitas sarana prasarana. Pembangunan infrastruktur adalah kunci utama untuk memajukan daerah Kabupaten Kepulauan Mentawai seperti adanya trans daerah sebagai penghubung antar pulau sehingga ekonomi kemasyarakatan akan turut tumbuh. Tujuan penelitian ini untuk mengetahui karakteristik atau variabel yang memiliki pengaruh pada pengkategorian status kesejahteraan rumah tangga ke dalam klasifikasi miskin dan tidak miskin. Klasifikasi adalah suatu teknik statistik yang digunakan untuk mengelompokkan data yang telah tersusun secara sistematis. Ada dua pendekatan berbeda untuk mengklasifikasi objek, yaitu metode parametrik dan metode nonparametrik. Penelitian ini memakai metode regresi logistik biner dan *Classification and Regression Tree* (CART) karena memiliki performansi yang baik, sehingga dalam penelitian ini akan mencoba memperoleh perbandingan nilai akurasi terbaik diantara kedua metode tersebut. Lalu hasilnya akan di evaluasi dengan nilai APER dan nilai akurasi klasifikasi. Data yang digunakan adalah hasil Susenas tahun 2022 sebanyak 326 sampel dengan data *testing* dan data *training* adalah 20% dan 80%. Dari hasil penelitian kedua metode, variabel umur, tingkat pendidikan terakhir, dan kesehatan kepala rumah tangga memiliki pengaruh signifikan terhadap model klasifikasi. Akurasi klasifikasi model regresi logistik biner mencapai 93,94% yang lebih tinggi dibandingkan dengan model klasifikasi CART yang bernilai 89,40%. Oleh karena itu, bisa ditarik kesimpulan bahwa model regresi logistik biner ialah pemilihan terbaik untuk memprediksi faktor kesejahteraan rumah tangga di Kabupaten Kepulauan Mentawai.

Kata Kunci: APER, *Classification and Regression Tree* (CART), Garis Kemiskinan, Perbandingan, Regresi Logistik Biner

Abstract

Mentawai Islands Regency is one of the areas that currently still has the status of a disadvantaged area, characterized by the minimal availability of infrastructure and facilities. Infrastructure development is the main key to advancing the Mentawai Islands Regency area, such as the existence of regional trans as a link between islands so that the community economy will also grow. The aim of this research is to determine the characteristics or variables that have an influence on the categorization of household welfare status into poor and non-poor classifications. Classification is a statistical technique used to group data

that has been arranged systematically. There are two different approaches to classifying objects, namely parametric methods and nonparametric methods. This research uses binary logistic regression and Classification and Regression Tree (CART) methods because they have good performance, so in this research we will try to obtain a comparison of the best accuracy values between the two methods. Then the results will be evaluated with the APER value and classification accuracy value. The data used is the 2022 Susenas results of 326 samples with testing data and training data being 20% and 80%. From the results of research on both methods, the variables age, highest level of education, and health of the head of the household have a significant influence on the classification model. The classification accuracy of the binary logistic regression model reached 93.94% which was higher than the CART classification model which was 89.40%. Therefore, it can be concluded that the binary logistic regression model is the best choice for predicting household welfare factors in the Mentawai Islands Regency.

Keywords: APER, Binary Logistic Regression, Classification and Regression Tree (CART), Comparison, Poverty Line



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

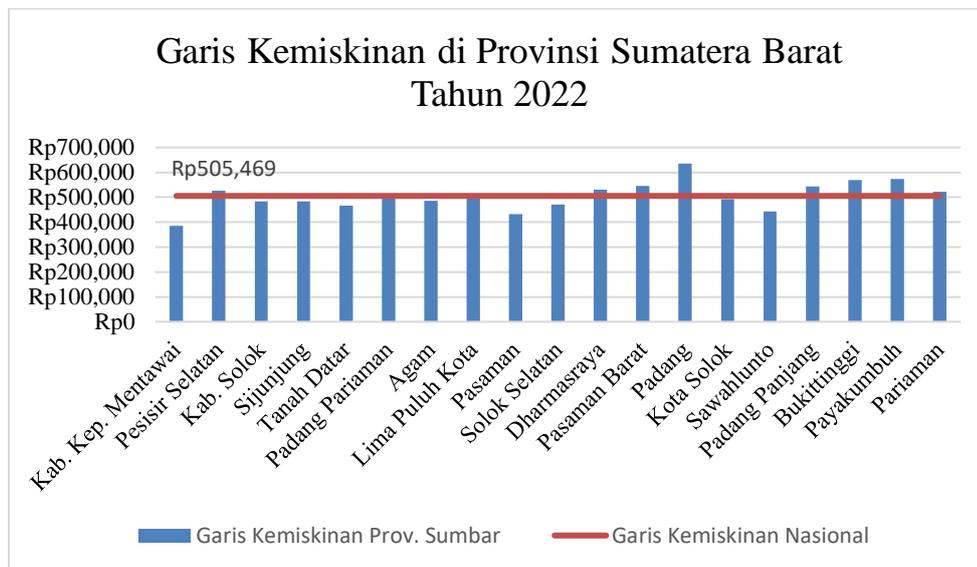
1. PENDAHULUAN

Kesejahteraan lahir batin bagi individu yang menetap di perkotaan atau diperdesaan adalah kehidupan yang diinginkan oleh seluruh orang. Kesejahteraan merupakan kondisi dimana keluarga telah memenuhi kebutuhan dasar, sosial, dan psikologis mereka secara optimal, termasuk kebutuhan yang bersifat pengembangan [1]. Tujuan kesejahteraan adalah untuk memenuhi berbagai tingkat kebutuhan manusia, termasuk kebutuhan primer, sekunder, dan tersier.

Indikator kesejahteraan rumah tangga dinilai berdasarkan faktor terkait dengan kepala rumah tangga, misalkan pada umur, tingkatan pendidikan, tingkat pendapatan, jenis kelamin, jumlah tanggungan, dan status perkawinan, kepemilikan aset dan rumah, kepemilikan hewan ternak, kesehatan keluarga, konsumsi makanan, status pekerjaan, serta akses terhadap pengetahuan informasi, dan keterlibatan politik. Kemiskinan umumnya didefinisikan sebagai ketidakmampuan ekonomi suatu masyarakat untuk mencukupi kebutuhan pangan dan nonpangan, yang diukur berdasarkan pengeluaran [2]. Oleh karena itu, masyarakat miskin ialah masyarakat yang mempunyai keterbatasan finansial serta pengeluaran per kapita per bulan rendah atau berada dibawah Garis Kemiskinan (GK).

Persoalan tentang kesejahteraan penduduk dan pengananan kemiskinan menjadi tugas bagi pemerintah, termasuk pemerintah Kabupaten Kepulauan Mentawai di wilayah Sumatera Barat. Kepulauan Mentawai dibagi menjadi empat pulau yang dihuni, memiliki luas 6.034 km² dengan 87.632 jiwa yang tersebar di 10 kecamatan. Sarana transportasi penduduk disana sebgaiian besar adalah melalui jalur laut dan sumber mata pencarian masyarakatnya adalah berladang, bertani, dan sebagai nelayan. Kabupaten Kepulauan Mentawai saat ini masih memiliki status sebagai salah satu daerah tertinggal. Hal ini ditunjukkan dengan masih kurangnya ketersediaan fasilitas sarana prasarana dan infrastruktur yang tersedia [3]. Menurut Nasrul Abit selaku Wakil Gubernur Sumatera Barat tahun 2019, infrastruktur yang belum sesuai dan tidak tersebar menjadi alasan utama Kabupaten Kepulauan Mentawai di Sumatera Barat masih terjebak dalam kondisi ketertinggalan [4]. Wakil Gubernur Sumatera Barat, Audy Joinaldy menyampaikan pembangunan infrastruktur di Mentawai terus ditingkatkan karena Mentawai sudah banyak dikunjungi dan dikenal wisatawan dunia, sehingga perlu adanya fokus pembangunan terutama pada infrastrukturnya.

Menurut informasi di website Badan Pusat Statistik (BPS), Kabupaten Kepulauan Mentawai menjadi wilayah termiskin di Provinsi Sumatera Barat. Data mengenai GK di Provinsi Sumatera Barat tahun 2022 disajikan pada Gambar 1.



Gambar 1. Grafik Garis Kemiskinan di Provinsi Sumatera Barat

Grafik GK di Kabupaten Kepulauan Mentawai masih berada di bawah standar GK Nasional yang ditetapkan. Berdasarkan laporan BPS, GK Nasional mengalami kenaikan sebesar 5,95% pada September 2022 jika dibandingkan dengan angka GK pada Maret 2022. GK Nasional per kapita naik dari Rp.505.469 per bulan menjadi Rp.535.547. Hal ini berarti warga negara yang tinggal di Indonesia dengan pendapatan per kapita per bulan di bawah Rp.505.469 per kapita tergolong dalam kelompok demografi tidak mampu [5]. Semakin tinggi GK maka angka kemiskinan akan terancam meningkat [6].

Klasifikasi adalah teknik statistik yang untuk mengatur dan mengkategorikan data yang disusun dalam urutan tertentu. Pengklasifikasian objek bisa dilakukan dengan menggunakan dua metodologi yang berbeda, yakni metode parametrik dan metode nonparametrik. Penelitian ini menggunakan pendekatan CART yang merupakan metode nonparametrik. Kedua metode memiliki nilai akurasi yang tinggi sehingga cocok untuk dilakukan perbandingan analisis.

Kondisi kesejahteraan rumah tangga di Provinsi Bali yang diteliti oleh [7] menggunakan penerapan analisis regresi logistik biner. Berdasarkan temuan penelitian, ditemukan berbagai faktor yang mempunyai pengaruh signifikan, yaitu jumlah individu dalam satu keluarga, jenis kelamin kepala keluarga, tingkat pendidikan terakhir, lapangan usaha, dan sektor utama pekerjaan kepala keluarga.

Literatur yang ada mengenai penerapan metode CART untuk menganalisis kasus kesejahteraan rumah tangga adalah penelitian oleh [8]. Dalam penelitian tersebut, ditemukan 47 node terminal pada klasifikasi optimal. Pemilah utama di antara node-node ini adalah jumlah anggota keluarga. Penerapan teknik klasifikasi pohon menghasilkan pohon yang optimal yang mencapai akurasi klasifikasi uji senilai 84,00% ketika diuji pada data *testing*.

Hasil penelitian yang dilakukan oleh [9] yang menggunakan analisis faktor untuk menguji elemen-elemen yang memengaruhi tingkat kesejahteraan masyarakat di Provinsi Jawa Tengah. Ditemukan bahwa faktor pertama mencakup variabel-variabel seperti gaji

minimum, rata-rata pengeluaran dalam keluarga, rasio kemiskinan, dan jumlah penduduk usia produktif. Sementara faktor kedua terdiri dari angka harapan hidup, keluhan kesehatan, angka partisipasi murni, dan angka partisipasi sekolah.

Berdasarkan latar belakang di atas, metode regresi logistik biner dan CART memiliki performansi yang baik, sehingga di penelitian ini akan mencoba untuk memperoleh nilai akurasi terbaik diantara kedua yang dipakai. Maka, peneliti merasa tertarik untuk menetapkan judul penelitian sebagai “Klasifikasi Status Kesejahteraan Masyarakat Kabupaten Kepulauan Mentawai Provinsi Sumatera Barat dengan Metode Regresi Logistik Biner dan *Classification and Regression Tree* (CART)”. Penentuan judul tersebut didasarkan pada konteks permasalahan yang terkait dengan fenomena gap yang muncul dalam penelitian ini.

2. METODE PENELITIAN

2.1. Metode Penelitian

Penelitian ini memakai data sekunder. Artinya informasi atau data yang digunakan tidak dikumpulkan secara langsung oleh peneliti melainkan berasal dari sumber yang sudah ada sebelumnya. Data yang disajikan dari Survei Sosial Ekonomi Nasional (SUSENAS) yang dilakukan oleh BPS di Kabupaten Kepulauan Mentawai pada tahun 2022. Dataset dapat diunduh melalui situs resmi website BPS di www.silastik.bps.go.id. Jumlah sampel yang dipakai ialah sejumlah 326 sampel. Variabel terikat (Y) yang digunakan pada penelitian ini adalah miskin dan tidak miskin. Sedangkan variabel bebas (x) ada 10 variabel, yaitu umur KRT (x_1), jenis kelamin KRT (x_2), status perkawinan KRT (x_3), banyak anggota rumah tangga (x_4), tingkat pendidikan tertinggi KRT (x_5), status pekerjaan utama KRT (x_6), kesehatan KRT 1 bulan terakhir (x_7), status kepemilikan tempat tinggal (x_8), bahan bakar memasak (x_9), dan kepemilikan telepon seluler (x_{10}).

2.2. Langkah Penelitian

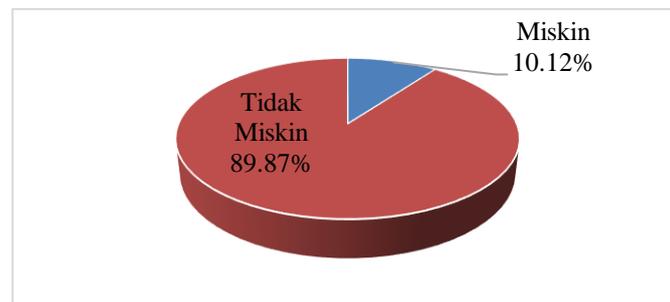
Software yang digunakan adalah R-studio. Langkah penelitian sebagai berikut:

1. Melakukan analisis statistik deskriptif terhadap variabel-variabel yang tercakup dalam penelitian.
2. Data dibagi menjadi dua bagian dengan 80% digunakan untuk data training dan 20% untuk data testing.
3. Pengolahan data analisis regresi logistik biner dengan langkah-langkah sebagai berikut:
 - a. Lakukan uji multikolinearitas untuk mendeteksi adanya multikolinearitas antara variabel bebas.
 - b. Uji estimasi parameter.
 - c. Uji signifikansi parameter dengan uji serentak dan uji parsial.
 - d. Bentuk model dengan memasukkan variabel-variabel yang signifikan.
 - e. Lakukan evaluasi model menggunakan uji kecocokan model untuk memperoleh model terbaik.
 - f. Gunakan nilai odd ratio untuk menginterpretasi hubungan antara variabel bebas dan variabel terikat dalam regresi logistik.
 - g. Perhitungan nilai APER dan akurasi klasifikasi model.
4. Melakukan analisis CART dengan langkah-langkah sebagai berikut:
 - a. Membangun pohon klasifikasi maksimal menggunakan data training.
 - i. Gunakan indeks Gini sebagai kaidah untuk memecah atau memilah pemilah dari variabel bebas.

- ii. Evaluasi hasil pemilahan menggunakan kriteria goodness of split.
 - iii. Tetapkan label kelas pada node terminal.
 - b. Gunakan metode cost complexity minimum untuk melakukan pemangkasan pada pohon klasifikasi.
 - c. Menentukan pohon klasifikasi optimal.
 - d. Ukur ketetapan model pohon klasifikasi optimal dengan menggunakan data training dan uji model dengan data testing untuk mendapatkan angka akurasi klasifikasi.
 - e. Menginterpretasikan hasil model pohon klasifikasi optimal.
 - f. Menghitung nilai APER dan akurasi klasifikasi model.
5. Penentuan model terbaiknya dengan membandingkan hasil dari metode regresi logistik biner dan CART berdasarkan nilai akurasi model klasifikasi dari kedua metode yang memiliki nilai interval kepercayaan yang lebih sempit. ini berisi tentang sumber data, variabel-variabel penelitian, teknik pengambilan sampel, cara pengumpulan data, dan metode analisis data yang diuraikan terperinci.

3. HASIL DAN PEMBAHASAN

Status kesejahteraan rumah tangga dipisahkan ke dalam dua kategori, yaitu kelompok miskin dan tidak miskin. Disajikan gambaran karakteristik rumah tangga di Kabupaten Kepulauan Mentawai dari 10 variabel bebas yang memiliki pengaruh. Informasi mengenai status kesejahteraan rumah tangga di Kabupaten Kepulauan Mentawai disajikan dalam Gambar 2.



Gambar 2. Status Kesejahteraan Rumah Tangga

Analisis tentang kesejahteraan rumah tangga di Kabupaten Kepulauan Mentawai pada tahun 2022 dengan 326 sampel yang diteliti, menunjukkan bahwa mayoritas rumah tangga di Kabupaten Kepulauan Mentawai cenderung tidak miskin atau dapat dikategorikan sejahtera. Dalam klasifikasi tersebut, sebanyak 89,87% atau setara dengan 293 rumah tangga termasuk dalam kategori tidak miskin. Sementara itu, sekitar 10,12% atau 33 rumah tangga masuk dalam kategori miskin. Meskipun sebagian besar masyarakat Kabupaten Kepulauan Mentawai dianggap tidak miskin, sejumlah keluarga mungkin tetap rentan terhadap resiko ekonomi. Faktor-faktor seperti pekerjaan yang tidak stabil atau ketergantungan pada sektor ekonomi tertentu dapat menyebabkan ketidakpastian ekonomi bagi beberapa keluarga.

1. Regresi Logistik Biner

a) Uji Multikolinearitas

Langkah awal pada metode regresi logistik biner adalah melakukan uji asumsi. Uji Asumsi yang dilakukan adalah uji Multikolinearitas dengan

menggunakan nilai *Variance Inflation Error* (VIF). Nilai VIF yang diharapkan adalah kecil dari 10.

Tabel 1. Nilai VIF

(X)	VIF
x_1	1,130
x_2	1,055
x_3	1,136
x_4	1,029
x_5	1,045
x_6	1,080
x_7	1,045
x_8	1,029
x_9	1,030
x_{10}	1,072

Dapat disimpulkan bahwasanya nilai VIF untuk setiap variabel bebas berada dibawah 10, maka tolak H_0 . Artinya, pada data yang dipakai tidak ada gejala terjadinya multikolineaitas antara variabel bebas.

b) Penaksiran Parameter

Pendekatan yang dilakukan adalah menggunakan *Maximum Likelihood Estimation* (MLE)

Tabel 2. Penduga Parameter

Variabel Bebas (X)	Estimasi Parameter β_i	<i>P-value</i>	Keputusan
Konstanta	-0,138	0,911	Terima H_0
x_1	0,026	0,032 *	Tolak H_0
x_2 (2)	-0,295	0,492	Terima H_0
x_3 (2)	-0,259	0,625	Terima H_0
x_3 (3)	13,339	0,993	Terima H_0
x_4	0,006	0,992	Terima H_0
x_5 (2)	-0,004	0,993	Terima H_0
x_5 (3)	2,173	0,043 *	Tolak H_0
x_5 (4)	0,578	0,615	Terima H_0
x_6 (2)	-0,334	0,549	Terima H_0
x_6 (3)	-0,355	0,637	Terima H_0
x_6 (4)	0,409	0,724	Terima H_0
x_6 (5)	0,581	0,620	Terima H_0
x_6 (6)	0,116	0,854	Terima H_0
x_7 (2)	1,096	0,019 *	Tolak H_0
x_8 (2)	15,41	0,992	Terima H_0
x_8 (3)	-0,821	0,382	Terima H_0
x_9 (2)	-0,040	0,964	Terima H_0
x_9 (3)	0,305	0,681	Terima H_0
x_9 (4)	14,11	0,993	Terima H_0
x_{10} (2)	0,172	0,699	Terima H_0

c) Uji Signifikansi Parameter

- Uji Serentak

Pada uji ini berguna untuk mengetahui signifikansi parameter secara keseluruhan terhadap variabel terikat. Uji ini digunakan uji G atau uji *likelihood maximum*.

Tabel 3. Uji Serentak

	$G(-2\text{LogLik})$	db	$\chi^2_{(a,b)}$
Model 1	162,81	21	$\chi^2_{(0,05,21)} = 32,67$
Model 2	167,69	6	$\chi^2_{(0,05,6)} = 12,59$

Dapat disimpulkan bahwasanya model pertama menghasilkan nilai uji rasio *likelihood* atau nilai $G = 162,81 > \chi^2_{(0,05,21)} = 32,67$ dan model kedua menghasilkan nilai $G = 167,69 > \chi^2_{(0,05,6)} = 12,59$. Oleh sebab itu, keputusan yang diambil ialah tolak H_0 . Kesimpulannya, setidaknya satu variabel bebas atau secara serentak yang memberikan pengaruh signifikan terhadap variabel terikat pada status kesejahteraan rumah tangga.

- Uji Parsial

Uji signifikansi parameter secara parsial dilakukan dengan menguji setiap variabel bebas secara individu. Uji parsial ini berguna untuk mengetahui pengaruh dari masing-masing variabel bebas terhadap variabel terikat.

Tabel 4. Uji Parsial

Variabel Bebas (X)	Estimasi Parameter β_i	Wald	P-value	Keputusan
Konstanta	-0,124	-0,202	0,840	Terima H_0
x_1	0,024	2,233	0,025 *	Tolak H_0
x_5 (2)	0,041	0,088	0,929	Terima H_0
x_5 (3)	2,138	2,039	0,041 *	Tolak H_0
x_5 (4)	0,818	0,754	0,450	Terima H_0
x_7 (2)	1,160	2,580	0,009 *	Tolak H_0

d) Model Regresi Logistik Biner

Dari analisis signifikansi parameter secara individu, terlihat bahwasanya uji signifikansi parameter yang hanya menunjukkan hasil signifikan untuk variabel x_1 , x_5 , dan x_7 . Oleh karena itu, nilai koefisien regresi sebagai berikut:

Tabel 5. Koefisien Regresi

Variabel Bebas (X)	Estimasi Parameter β_i	Wald	P-value	Keputusan
Konstanta	-0,124	-0,202	0,840	Terima H_0
x_1	0,024	2,233	0,025 *	Tolak H_0
x_5 (2)	0,041	0,088	0,929	Terima H_0
x_5 (3)	2,138	2,039	0,041 *	Tolak H_0
x_5 (4)	0,818	0,754	0,450	Terima H_0
x_7 (2)	1,160	2,580	0,009 *	Tolak H_0

Model regresi logistik biner dalam bentuk logit sebagai berikut:

$$g(x) = -0,124 + 0,024x_1 + 0,041x_{5(2)} + 2,138x_{5(3)} + 0,818x_{5(4)} + 1,160x_{7(2)}$$

Maka diperoleh model akhir regresi logistik biner:

$$\pi(x_i) = \frac{e^{g(x)}}{1+e^{g(x)}}$$

$$\pi(x_i) = \frac{e^{(-0,124+0,024x_1+0,041x_{5(2)}+2,138x_{5(3)}+0,818x_{5(4)}+1,160x_{7(2)})}}{1 + e^{(-0,124+0,024x_1+0,041x_{5(2)}+2,138x_{5(3)}+0,818x_{5(4)}+1,160x_{7(2)})}}$$

Pada model yang terbentuk, dapat diidentifikasi bahwasanya variabel yang berpengaruh positif terhadap kecenderungan rumah tangga menjadi miskin melibatkan umur KRT (x_1), tingkat pendidikan SMA Sederajat $x_{5(3)}$ dan kesehatan KRT yang tidak punya keluhan penyakit $x_{7(2)}$. Oleh karena itu, dapat diungkapkan bahwa model regresi logistik biner memperoleh variabel atau faktor yang berpengaruh terhadap status kesejahteraan rumah tangga di Kabupaten Kepulauan Mentawai ialah variabel umur KRT, tingkat pendidikan kategori SMA Sederajat, dan kesehatan KRT yang tidak punya keluhan penyakit.

Koefisien yang terkait dengan setiap variabel ($x_1, x_{5(2)}, x_{5(3)}, x_{5(4)}, x_{7(2)}$) mengindikasikan seberapa besar dampak variabel tersebut terhadap log-odds hasil positif. Jika koefisien positif, peningkatan nilai variabel akan meningkatkan log-odds hasil positif. Sedangkan jika koefisien negatif, peningkatan nilai variabel akan meningkatkan log-odds hasil negatif.

e) Uji Kecocokan Model

Pengujian kecocokan model dilakukan dengan tujuan untuk menilai apakah terdapat perbedaan yang signifikan antara hasil observasi yang diperoleh dengan nilai prediksi yang mungkin diperoleh dari model. Uji ini memakai uji Hosmer dan Lemeshow.

Tabel 6. Uji Hosmer dan Lemeshow

\hat{C}	db	$\chi^2_{(0,05,8)}$	P-value
Konstanta	-0,124	-0,202	0,840

Dari hasil uji Hosmer dan Lemeshow, ditemukan nilai \hat{C} bernilai 11,866 lebih kecil dari nilai *chi-square* tabel 15,507 dan *p-value* 0,157 yang lebih besar dari tingkat signifikansi α (0,05), oleh karena itu, tidak dapat menolak H_0 . Maka bisa ditarik kesimpulan bahwasanya tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model, menunjukkan bahwa model yang diperoleh sesuai dengan data pengamatan.

f) Interpretasi Koefisien Regresi Logistik

Koefisien dalam model regresi logistik biner bisa diinterpretasikan memakai skor odds ratio. Odds ratio mencerminkan kecenderungan antara satu kategori dengan kategori lain pada variabel bebas yang bersifat kualitatif.

Tabel 7. Nilai Odds Ratio

Variabel Bebas (X)	Estimasi Parameter β_i	Odds Ratio
Konstanta	-0,124	$OR_1 = e^{0,024} = 1,024$
x_1	0,024	$OR_{5(2)} = e^{0,041} = 1,041$
x_5 (2)	0,041	$OR_{5(3)} = e^{2,138} = 8,482$
x_5 (3)	2,138	$OR_{5(3)} = e^{0,818} = 2,265$
x_5 (4)	0,818	$OR_{7(2)} = e^{1,160} = 3,189$
x_7 (2)	1,160	$OR_1 = e^{0,024} = 1,024$

- Variabel x_1 (Umur KRT)
 Nilai odds ratio untuk variabel umur KRT adalah 1,024. Ini mengindikasikan bahwa rumah tangga dengan umur KRT yang lebih muda memiliki kemungkinan 1,024 kali lebih tinggi untuk menjadi rumah tangga tidak miskin dibandingkan dengan rumah tangga yang memiliki umur KRT yang lebih tua.
 - Variabel $x_{5(2)}$ (Tingkat Pendidikan KRT Kategori SMA Sederajat)
 Odds ratio untuk variabel tingkat pendidikan SMA Sederajat adalah 8,482. Ini menandakan bahwa rumah tangga dengan tingkat pendidikan SMA Sederajat memiliki kecenderungan menjadi rumah tangga yang tidak miskin sebanyak 8,482 kali lebih besar dibandingkan dengan rumah tangga dengan dari tingkat pendidikan SD Sederajat. Hal ini selaras dengan penelitian yang dilakukan oleh (Utomo, 2009).
 - Variabel $x_{7(2)}$ (Kesehatan KRT yang tidak punya keluhan penyakit)
 Odds ratio untuk variabel kesehatan KRT yang tidak punya keluhan penyakit adalah 3,189. Ini menunjukkan bahwa rumah tangga dengan KRT yang tidak punya keluhan penyakit memiliki kecenderungan menjadi rumah tangga tidak miskin sebanyak 3,189 kali lebih besar dibandingkan dengan rumah tangga yang memiliki KRT dengan keluhan penyakit.
- g) Akurasi Klasifikasi Regresi Logistik Biner

Tabel 8. Tingkat Akurasi Klasifikasi pada Model Regresi Logistik Biner

Taksiran Klasifikasi	Hasil Observasi		Total
	Miskin	Tidak Miskin	
Miskin	0	1	1
Tidak Miskin	3	62	65
Total	3	63	66

Nilai akurasi klasifikasi metode regresi logistik biner dihitung menggunakan APER seperti berikut.

$$\text{APER (\%)} = \left(\frac{3+1}{66} \right) \times 100\% = 0,060 \times 100\% = 6,06\%$$

$$\text{Akurasi Klasifikasi} = 100\% - 6,06\% = 93,94\%$$

APER dihitung sebagai rasio antara jumlah pengamatan yang diprediksi salah oleh fungsi klasifikasi dengan n total pengamatan. Berdasarkan perhitungan nilai APER, ditemukan tingkat kesalahan klasifikasi sebesar 6,06%, yang mengartikan bahwa nilai akurasi klasifikasi sebesar 93,94%. Hal ini menunjukkan bahwa model regresi logistik dapat memprediksi dengan tepat pengamatan sebesar 93,94%.

2. Classification and Regression Tree (CART)

a) Pembentukan Pohon Klasifikasi

Dalam pembentukan pohon klasifikasi maksimal ialah mencari pemilah yang mungkin dari masing-masing variabel bebas yang akan menjadi pemilah utama. Metode pemilahan indeks Gini akan selalu memisahkan kelas yang jumlah kelas terpenting atau observasi paling besar terlebih dahulu pada node. Kelas merupakan kategori dari variabel terikat. Penelitian ini terdiri dari dua kelas yaitu kelas miskin dan tidak miskin. Pemilah terbaik adalah pemilah yang menghasilkan nilai penurunan heterogenitas tertinggi yang dikenal sebagai *kriteria goodness of split*.

Pohon klasifikasi maksimal diperoleh melalui serangkaian proses pemilahan biner yang secara berulang diterapkan pada data observasi. Hasilnya yaitu pohon klasifikasi dengan ukuran besar dan tingkat kedalaman yang tinggi yang terdiri dari 51 *node*, dimana

terdapat 1 *node* induk, 24 *node* anak, dan 26 *node* terminal. Pohon maksimal yang terbentuk mencapai tingkat kedalaman sebesar 12.

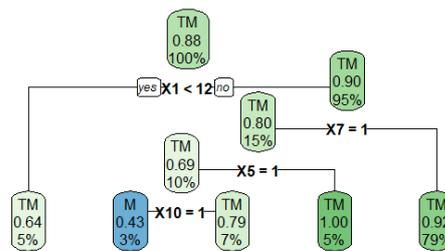
b) Pemangkasan Pohon Klasifikasi Maksimal

Tabel 8. Tabel Cost Complexity Parameter

	CP	N Split	Rel Error	CV Error	SE
1	0,033	0	1,000	1,000	0,171
2	0,025	3	0,900	1,300	0,191
3	0,016	8	0,766	1,566	0,206
4	0,013	17	0,600	1,700	0,213
5	0,011	22	0,533	1,800	0,218
6	0,000	25	0,500	1,866	0,220

Ditemukan nilai CP untuk setiap pemisahan berturut-turut adalah 0,033; 0,025; 0,016; 0,013; 0,011; 0,000. Selain itu, tercatat nilai cross validation error untuk setiap pemisahan berturut-turut adalah 1,000; 1,300; 1,566; 1,700; 1,800; 1,866. Berdasarkan nilai cross validation error yang paling kecil yakni 1,000, maka dipilih nilai CP sebesar 0,033 sebagai nilai yang digunakan untuk pemangkasan pohon.

c) Pohon Klasifikasi Optimal



Gambar 3. Pohon Klasifikasi Optimal

Diketahui *node* terminal yang dihasilkan pada pohon klasifikasi optimal sebanyak 5 *node* terminal, 3 *node* anak, dan 1 *node* induk dengan kedalaman sebanyak 5. Variabel yang sangat menentukan klasifikasi status kesejahteraan rumah tangga di Kabupaten Kepulauan Mentawai ialah umur KRT (x_1), kesehatan KRT (x_7), tingkat pendidikan KRT (x_5), dan kepemilikan telepon seluler (x_{10}).

d) Akurasi Klasifikasi CART

Tabel 8. Tingkat Akurasi Klasifikasi pada Model Regresi Logistik Biner

Taksiran Klasifikasi	Hasil Observasi		Total
	Miskin	Tidak Miskin	
Miskin	0	4	1
Tidak Miskin	3	59	62
Total	3	63	66

Perhitungan nilai akurasi klasifikasi metode CART dilakukan dengan menggunakan nilai APER sebagai berikut:

$$\text{APER} (\%) = \left(\frac{3+4}{66} \right) \times 100\% = 0,106 \times 100\% = 10,60\%$$

$$\text{Akurasi Klasifikasi} = 100\% - 10,60\% = 89,40\%$$

Perhitungan nilai APER dilakukan dengan menghitung proporsi pengamatan yang diprediksi salah oleh fungsi klasifikasi, dengan n sebagai total pengamatan. Hasil

perhitungan APER atau nilai kesalahan klasifikasi pada penggunaan pohon CART menunjukkan angka senilai 10,60%, sehingga diperoleh nilai akurasi klasifikasi bernilai 89,40%. Hal ini mengindikasikan bahwa model CART dapat memprediksi dengan tepat sekitar 89,40% dari total pengamatan.

3. Perbandingan Interval Kepercayaan

Perbandingan hasil dari kedua metode mungkin menunjukkan perbedaan akurasi yang tidak signifikan secara statistik. Untuk memperkuat hasil ini, perbandingan kinerja klasifikasi dapat dilakukan dengan menggunakan uji statistik interval kepercayaan. Interval kepercayaan seringkali digunakan untuk membandingkan kinerja klasifikasi yang berbeda, membantu dalam menentukan metode klasifikasi yang lebih efektif untuk suatu dataset. Perhitungan interval kepercayaan ini didasarkan pada nilai akurasi klasifikasi yang telah dihasilkan sebelumnya.

Interval kepercayaan untuk nilai akurasi klasifikasi pada metode regresi logistik biner dapat dengan nilai akurasi 93,94%, jumlah sampel data *testing* sebanyak 66, dan nilai $Z_{tabel} = 1,96$.

$$\begin{aligned} \text{Interval Kepercayaan} &= \frac{2 \cdot N \cdot acc_{rlb} + Z_{\alpha/2}^2 \pm Z_{\alpha/2} \sqrt{Z_{\alpha/2}^2 + 4 \cdot N \cdot acc_{rlb} - 4 \cdot N \cdot acc_{rlb}^2}}{2(N + Z_{\alpha/2}^2)} \\ &= \frac{2 \cdot 66 \cdot 0,9394 + 1,96^2 \pm 1,96 \sqrt{1,96^2 + 4 \cdot 66 \cdot 0,9394 - 4 \cdot 66 \cdot 0,9394^2}}{2(66 + 1,96^2)} \\ &= \frac{127,8424 \pm 8,514276998}{139,6832} \end{aligned}$$

Interval kepercayaan atau akurasi kotor untuk nilai akurasi klasifikasi regresi logistik biner berkisar antara 0,8542 hingga 0,9761. Oleh karena itu, nilai akurasi klasifikasi yang diperoleh dapat dianggap berada dalam rentang tersebut.

Berikutnya mencari nilai interval kepercayaan untuk metode CART dengan tingkat akurasi sebesar 89,40%.

$$\begin{aligned} \text{Interval Kepercayaan} &= \frac{2 \cdot N \cdot acc_{CART} + Z_{\alpha/2}^2 \pm Z_{\alpha/2} \sqrt{Z_{\alpha/2}^2 + 4 \cdot N \cdot acc_{CART} - 4 \cdot N \cdot acc_{CART}^2}}{2(N + Z_{\alpha/2}^2)} \\ &= \frac{2 \cdot 66 \cdot 0,8940 + 1,96^2 \pm 1,96 \sqrt{1,96^2 + 4 \cdot 66 \cdot 0,8940 - 4 \cdot 66 \cdot 0,8940^2}}{2(66 + 1,96^2)} \\ &= \frac{121,8496 \pm 10,52928637}{139,6832} \end{aligned}$$

Interval kepercayaan atau akurasi kotor untuk nilai akurasi klasifikasi CART berkisar antara 0,7969 hingga 0,9477. Dengan demikian, nilai akurasi klasifikasi yang dihasilkan berada dalam rentang tersebut.

Berdasarkan hasil dari kedua interval tersebut, dapat disimpulkan interval kepercayaan untuk nilai akurasi regresi logistik menunjukkan margin yang lebih kecil dibandingkan dengan interval CART. Hal ini mengindikasikan bahwa interval yang lebih sempit memberikan nilai statistik yang lebih akurat. Oleh karena itu, disimpulkan bahwa model terbaik untuk memprediksi faktor dari status kesejahteraan rumah tangga di Kabupaten Kepulauan Mentawai adalah metode regresi logistik biner.

4. KESIMPULAN

Kabupaten Kepulauan Mentawai saat ini masih memiliki status sebagai salah satu daerah tertinggal. Hal ini ditunjukkan dengan masih kurangnya ketersediaan fasilitas sarana prasarana dan infrastruktur yang tersedia. Menurut Nasrul Abit selaku Wakil Gubernur Sumatera Barat tahun 2019, infrastruktur yang belum sesuai dan tidak tersebar menjadi alasan utama Kabupaten Kepulauan Mentawai di Sumatera Barat masih terjebak dalam kondisi ketertinggalan. Pembangunan infrastruktur menjadi kunci utama untuk

kesejahteraan dan pembangunan wilayah Kabupaten Kepulauan Mentawai seperti adanya transportasi daerah sebagai penghubung antar pulau dan memperlancar akses sehingga perekonomian masyarakat juga semakin berkembang. Dalam pelaksanaan program perencanaan pembangunan di daerah Kabupaten Kepulauan Mentawai masih belum terealisasi sesuai dengan tujuan yang diinginkan. Kondisi infrastruktur jalan di Kabupaten Kepulauan Mentawai saat ini belum memadai karena belum diaspal, sehingga memerlukan perbaikan, perluasan, dan perhatian lebih dari pemerintah daerah. Selain kekhawatiran yang disebutkan di atas, adanya keterbatasan gedung kantor, fasilitas pendidikan, fasilitas kesehatan, dan tenaga sumber manusia yang dapat mengelola semuanya secara efektif.

Berdasarkan penelitian yang telah dilakukan, metode regresi logistik biner memiliki nilai akurasi sebesar 93,94% sedangkan metode CART memiliki nilai akurasi sebesar 89,40%. Oleh karena itu, metode regresi logistik biner lebih baik digunakan pada penelitian ini karena memiliki nilai akurasi yang lebih tinggi, dengan variabel-variabel yang teridentifikasi paling berpengaruh terhadap status kesejahteraan rumah tangga di Kabupaten Kepulauan Mentawai adalah $(x_1, x_{5(2)}, x_{5(3)}, x_{5(4)}, x_{7(2)})$.

DAFTAR PUSTAKA

- [1] Jabar.bkkbn.go.id, "Mulailah Dengan Membangun Kesejahteraan Keluarga," *jabar.bkkbn.go.id*, 2020. <https://jabar.bkkbn.go.id/?p=1580> (accessed Jul. 13, 2022).
- [2] Badan Pusat Statistik, "Kemiskinan dan Ketimpangan," *Badan Pusat Statistik*, 2020. <https://www.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html#subjekViewTab2>
- [3] F. M. Illahi and A. Ariusni, "Pengaruh Fasilitas Rumah Terhadap Kesejahteraan Masyarakat Kabupaten Kepulauan Mentawai," *J. Kaji. Ekon. dan Pembang.*, vol. 2, no. 2, p. 55, 2020, doi: 10.24036/jkep.v2i2.12643.
- [4] sumbar antarnews, "Fakta penyebab Kepulauan Mentawai masih tertinggal," *sumbar antarnews*, 2019. <https://sumbar.antaranews.com/berita/281562/fakta-penyebab-kepulauan-mentawai-masih-tertinggal> (accessed Jul. 19, 2023).
- [5] Cnbcindonesia, "Kategori Miskin di RI: Pengeluaran di Bawah Rp 17.851 / Hari," *cnbcindonesia*, 2023. [https://www.cnbcindonesia.com/news/20230119080431-4-406648/kategori-miskin-di-ri-pengeluaran-di-bawah-rp-17851---hari#:~:text=Jika dihitung dari besaran GK,kapita masuk kategori tidak mampu.](https://www.cnbcindonesia.com/news/20230119080431-4-406648/kategori-miskin-di-ri-pengeluaran-di-bawah-rp-17851---hari#:~:text=Jika%20dihitung%20dari%20besaran%20GK,kapita%20masuk%20kategori%20tidak%20mampu.) (accessed Jul. 14, 2023).
- [6] mediaindonesia, "Angka Kemiskinan Terancam Meningkatkan Akibat Peningkatan Garis Kemiskinan," *media indonesia*, 2023. <https://mediaindonesia.com/ekonomi/585298/angka-kemiskinan-terancam-meningkat-akibat-peningkatan-garis-kemiskinan> (accessed Jul. 14, 2023).
- [7] V. S. Permatasari and L. Yuliana, "Penerapan Regresi Logistik Biner pada Status Kesejahteraan Rumah Tangga di Provinsi Bali Tahun 2020," *Semin. Nas. Off. Stat.*, vol. 2021, no. 1, pp. 497–506, 2021, doi: 10.34123/semnasoffstat.v2021i1.943.
- [8] G. Mahalisa, "Metode Pohon Gabungan Pada Cart Untuk Analisa Kesejahteraan Rumah Tangga Di Banjarmasin," *Technol. J. Ilm.*, vol. 7, no. 1, pp. 15–24, 2016, doi: 10.31602/tji.v7i1.608.
- [9] N. I. D. Gustyarini, "Analisis Faktor Yang Mempengaruhi Kesejahteraan Masyarakat Di Provinsi Jawa Tengah," pp. i–78, 2019.

