

# PERBANDINGAN ALGORITMA CART DAN C4.5 UNTUK KLASIFIKASI KEPADATAN PENDUDUK SUMATERA UTARA BERDASARKAN FAKTOR SOSIAL EKONOMI DAN KESEHATAN

Dito Nurbiat Moko<sup>1\*</sup>, Regita Karine Mashitawati<sup>2</sup>, Widhi Nugrahaeni Banov Putri<sup>3</sup>,  
Muhammad Riefky<sup>4</sup>

<sup>1,2,3,4</sup>D-IV Studi Demografi dan Pencatatan Sipil, Universitas Sebelas Maret, Kota Surakarta,  
Provinsi Jawa Tengah, Indonesia

Corresponding e-mail: [ditolongiditonm@student.uns.ac.id](mailto:ditolongiditonm@student.uns.ac.id)

Copyright © 2026 The Author



This is an open access article

Under the Creative Commons Attribution Share Alike 4.0 International License

DOI: [10.53866/jimi.v6i1.1001](https://doi.org/10.53866/jimi.v6i1.1001)

## Abstrak

Penduduk merupakan faktor utama dalam menentukan dinamika pembangunan wilayah, khususnya terkait kepadatan yang tidak merata antar kabupaten/kota. Provinsi Sumatera Utara pada tahun 2023 masih menghadapi ketimpangan kepadatan penduduk yang signifikan, di mana konsentrasi penduduk tinggi terjadi di kota besar seperti Medan, sementara wilayah lain relatif rendah. Penelitian ini bertujuan untuk membandingkan kinerja algoritma CART dan C4.5 dalam mengklasifikasikan tingkat kepadatan penduduk kabupaten/kota di Provinsi Sumatera Utara tahun 2023 berdasarkan enam variabel sosial ekonomi dan kesehatan: jumlah penduduk, angka kelahiran, pengeluaran per kapita, Indeks Pembangunan Manusia (IPM), akses air bersih, dan laju pertumbuhan penduduk. Data diperoleh dari Badan Pusat Statistik (BPS) Sumatera Utara. Hasil menunjukkan bahwa kedua algoritma menghasilkan tingkat akurasi yang sama, yaitu 88,89% dengan proporsi data latih 75% dan data uji 25%. IPM muncul sebagai variabel dengan kontribusi terbesar dalam menentukan kategori kepadatan penduduk, diikuti oleh angka kelahiran, laju pertumbuhan, dan akses air bersih. Temuan ini menegaskan bahwa peningkatan IPM dan akses air bersih merupakan prioritas utama bagi pemerintah daerah dalam merumuskan kebijakan pemerataan kependudukan dan pembangunan wilayah di Sumatera Utara. Hasil penelitian tersebut menjadi dasar bagi Pemerintah Provinsi Sumatera Utara dalam merumuskan strategi kebijakan yang berfokus pada peningkatan angka Indeks Pembangunan Manusia (IPM) dan akses air bersih untuk kualitas hidup masyarakat, sekaligus angka kelahiran dan laju pertumbuhan penduduk untuk pemerataan kependudukan di Provinsi Sumatera Utara.

**Kata Kunci:** C4.5, CART, Kepadatan Penduduk, Klasifikasi, Sumatera Utara

## *Comparison of Cart and C4.5 Algorithms for Population Density Classification in North Sumatra Based on Socio Economic and Health Factors*

### *Abstract*

*Population is a crucial factor in shaping regional development dynamics, particularly regarding uneven population density across districts and cities. In 2023, North Sumatra Province still faces significant disparities, with a high concentration in major cities such as Medan, while other regions remain relatively low. This study aims to compare the performance of the CART and C4.5 algorithms in classifying population density levels across districts and cities in North Sumatra Province in 2023, based on six socioeconomic and health-related variables: total population, birth rate, per capita expenditure, Human Development Index (HDI), access to clean water, and population growth rate. The data were obtained from the Central Bureau of Statistics (BPS) of North Sumatra. The results show that both algorithms achieved the same accuracy level*

of 88.89%, using a data split of 75% for training and 25% for testing. The HDI emerged as the most contribution variable in determining population density categories, followed by birth rate, population growth rate, and access to clean water. These findings highlight that improving HDI and access to clean water should be key priorities for local governments in formulating policies aimed at promoting population distribution equity and regional development in North Sumatra. These findings provide a foundation for the Provincial Government to prioritize policies that improve HDI and access to clean water to enhance quality of life, while also addressing birth rate and population growth to promote a more balanced population distribution across North Sumatra.

**Keywords:** C4.5, CART, Classification, North Sumatra, Population Density

## 1. Pendahuluan

Secara demografis, Indonesia merupakan salah satu negara dengan jumlah penduduk terbesar di dunia dan masih mengalami laju pertumbuhan yang relatif tinggi. Kondisi ini menjadikan kepadatan penduduk sebagai isu strategis, khususnya di wilayah perkotaan dan daerah penyangga, karena berdampak langsung pada tekanan infrastruktur, layanan publik, kualitas hidup, ketersediaan lahan, serta perencanaan pembangunan jangka panjang (Maharani et al., 2024). Ketimpangan distribusi penduduk antar wilayah turut memperparah kondisi tersebut dengan menciptakan beban pembangunan yang tidak merata. Oleh karena itu, pemahaman pola kepadatan penduduk secara komprehensif memerlukan analisis berbagai indikator utama, seperti jumlah penduduk, angka kelahiran hidup, pengeluaran per kapita, Indeks Pembangunan Manusia (IPM), ketersediaan air bersih, dan laju pertumbuhan penduduk.

Di Provinsi Sumatera Utara, perbedaan karakteristik kependudukan antar wilayah terlihat cukup mencolok. Kota Medan tercatat sebagai daerah dengan jumlah penduduk tertinggi sebesar 2.474.200 jiwa, diikuti Kabupaten Deli Serdang sebanyak 2.018.200 jiwa dan Kabupaten Langkat dengan 1.066.700 jiwa. Sebaliknya, Kabupaten Nias dan Pakpak Bharat masing-masing hanya memiliki 153.700 jiwa dan 54.100 jiwa, menunjukkan wilayah dengan populasi relatif rendah (BPS Sumatera Utara, 2023). Selain jumlah penduduk, variasi angka kelahiran antar wilayah juga berkontribusi signifikan terhadap dinamika pertumbuhan dan komposisi penduduk, sehingga indikator-indikator tersebut menjadi aspek penting dalam analisis kepadatan penduduk di Sumatera Utara. Sebagai provinsi dengan jumlah penduduk terbesar keempat di Indonesia, Sumatera Utara memiliki tingkat kepadatan yang beragam antara wilayah perkotaan dan daerah terpencil. Urbanisasi yang intensif, mobilitas penduduk antar wilayah, serta pertumbuhan penduduk alami memperkuat ketimpangan distribusi penduduk. Dampaknya, sejumlah kabupaten/kota mengalami tekanan tinggi terhadap fasilitas umum dan layanan dasar, sementara wilayah lain justru menghadapi keterbatasan aktivitas ekonomi dan lambatnya pembangunan infrastruktur akibat minimnya jumlah penduduk (Asmara et al., 2024). Fenomena ini menunjukkan bahwa kepadatan penduduk tidak hanya berkaitan dengan jumlah penduduk semata, tetapi juga berimplikasi langsung pada kualitas pembangunan dan pemerataan kesejahteraan antar wilayah.

Provinsi Sumatera Utara dipilih sebagai lokasi penelitian karena memiliki jumlah penduduk yang besar dengan tingkat kepadatan yang bervariasi antarwilayah pada tahun 2023, sehingga menimbulkan tantangan tersendiri dalam pemerataan pembangunan, infrastruktur, dan pelayanan publik (Siagian, 2021). Untuk mengklasifikasikan kabupaten/kota berdasarkan kepadatan penduduk, penelitian ini menggunakan dua algoritma *decision tree*, yaitu *Classification and Regression Tree* (CART) dan C4.5. Algoritma CART membentuk pohon keputusan biner dengan indeks gini sebagai kriteria pemisahan data serta menerapkan proses *pruning* untuk mengurangi risiko *overfitting* (Daniya et al., 2020). Sementara itu, algoritma C4.5 merupakan pengembangan dari ID3 yang menggunakan *information gain ratio*, mampu menangani atribut kontinu dan data dengan nilai hilang, serta menghasilkan struktur pohon yang lebih fleksibel (Mardiansyah et al., 2021). Penggunaan kedua metode ini diharapkan mampu menghasilkan klasifikasi kepadatan penduduk yang akurat, mudah diinterpretasikan, dan relevan sebagai dasar perumusan kebijakan pembangunan wilayah yang berkelanjutan (Shamrat et al., 2022).

Penelitian mengenai kepadatan penduduk dan penerapan metode klasifikasi telah banyak dilakukan dengan beragam fokus dan pendekatan. Dita dan Legowo (2022) mengkaji pengaruh kepadatan penduduk terhadap kemiskinan dan degradasi lingkungan di Indonesia. Pendekatan spasial berbasis Sistem Informasi Geografis digunakan oleh Azzahra et al. (2025) dalam menganalisis kepadatan penduduk melalui indikator

jumlah penduduk dan kelahiran pada tingkat desa di Medan Sinembah, Limau Manis, dan Ujung Serda. Sementara itu, Irham dan Putri (2023) memanfaatkan variabel jumlah penduduk dan luas wilayah untuk mengkaji keterkaitan kepadatan penduduk dengan Indeks Pembangunan Manusia di Provinsi Lampung. Dari sisi metodologi algoritma CART dan C4.5 telah banyak diaplikasikan pada berbagai domain. Avcı (2025) menunjukkan bahwa CART memiliki performa klasifikasi yang baik dengan akurasi 0.9334 dalam memprediksi indikator kinerja rumah sakit publik, meskipun masih berada di bawah ID3. Temuan serupa disampaikan oleh Lestari et al. (2024) yang menyatakan bahwa CART memiliki tingkat akurasi dan presisi yang tinggi dalam klasifikasi risiko stroke, meskipun nilai *recall* relatif lebih rendah dibandingkan algoritma lain. Di sisi lain, C4.5 terbukti efektif dalam membangun model klasifikasi pada berbagai konteks, seperti prediksi penjualan (Ruliansyah & Betty, 2024) dan klasifikasi prestasi atlet (Kuswanto et al., 2024), dengan kemampuan mengidentifikasi variabel-variabel kunci yang berpengaruh terhadap hasil klasifikasi.

Berdasarkan telaah tersebut, terdapat kesenjangan penelitian yang perlu diatasi khususnya terkait minimnya studi komparatif yang mengintegrasikan algoritma CART dan C4.5 dalam klasifikasi kepadatan penduduk dengan mempertimbangkan karakteristik wilayah serta potensi ketidakseimbangan data (Widyawati et al., 2023). Selain itu, pemanfaatan data kependudukan lokal yang kontekstual sebagai dasar perumusan kebijakan regional masih relatif terbatas. Penelitian ini berfokus pada penerapan dan perbandingan algoritma CART dan C4.5 untuk mengklasifikasikan kepadatan penduduk di Provinsi Sumatera Utara tahun 2023 menggunakan enam variabel indikator kependudukan. Penelitian ini bertujuan untuk mengidentifikasi model klasifikasi terbaik berdasarkan tingkat akurasi tertinggi, sekaligus menentukan variabel prioritas yang berperan dominan dalam pembentukan kategori kepadatan penduduk. Kontribusi penelitian ini tidak hanya terletak pada penguatan kajian akademik melalui analisis komparatif dua algoritma klasifikasi yang populer, tetapi juga pada penyediaan hasil visualisasi model yang aplikatif sebagai dasar pendukung pengambilan kebijakan pembangunan dan pemerataan layanan publik berbasis data kependudukan di tingkat Provinsi Sumatera Utara. Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah Bagaimana kinerja dan perbandingan algoritma CART dan C4.5 dalam mengklasifikasikan tingkat kepadatan penduduk kabupaten/kota di Provinsi Sumatera Utara tahun 2023? Serta Variabel sosial ekonomi dan kesehatan apa yang memiliki kontribusi terbesar dalam mengklasifikasikan kepadatan penduduk kabupaten/kota di Provinsi Sumatera Utara?

## 2. Metode Penelitian

### 2.1. Objek, waktu dan Tempat

Objek studi ini adalah data kepadatan penduduk di kabupaten/kota Provinsi Sumatera Utara tahun 2023 yang mengacu pada publikasi Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Penelitian difokuskan pada perbandingan kinerja algoritma CART dan C4.5 dalam klasifikasi wilayah berdasarkan kepadatan penduduk. Waktu penelitian dilaksanakan pada tahun 2025 dengan menggunakan data sekunder tahun 2023. Tempat penelitian adalah Provinsi Sumatera Utara dengan analisis seluruh kabupaten/kota.

### 2.2. Teknik Pengumpulan Data

Metode pengumpulan data dalam studi ini memanfaatkan data sekunder yang dihimpun melalui *website* Badan Pusat Statistik Provinsi Sumatera Utara (<https://sumut.bps.go.id/id>) tahun 2023. Penelitian ini memanfaatkan data yang mencakup jumlah penduduk, luas wilayah serta angka kelahiran hidup. Rincian masing-masing variabel yang dianalisis dalam penelitian ini dapat dilihat pada Tabel 1.

**Tabel 1.** Variabel Penelitian

Variabel	Keterangan	Satuan
Y	Kepadatan Penduduk	Jiwa/km <sup>2</sup>
X <sub>1</sub>	Jumlah Penduduk	Jiwa
X <sub>2</sub>	Angka Kelahiran	Jiwa
X <sub>3</sub>	Pengeluaran Per Kapita	Ribu Rupiah
X <sub>4</sub>	IPM	-
X <sub>5</sub>	Air Bersih	Persen (%)
X <sub>6</sub>	Laju Pertumbuhan	Persen (%)

Pemilihan variabel dalam studi ini didasarkan pada pertimbangan bahwa dinilai mampu merepresentasikan aspek-aspek yang berkontribusi terhadap kepadatan penduduk di Provinsi Sumatera Utara. Variabel dependen (Y) adalah kepadatan penduduk yang dijadikan dasar dalam membedakan suatu wilayah apakah termasuk kategori padat atau tidak padat. Sementara itu, variabel independen terdiri dari enam indikator, yaitu jumlah penduduk ( $X_1$ ) yang menunjukkan total populasi pada suatu wilayah, angka kelahiran ( $X_2$ ) yang mencerminkan pertumbuhan alami penduduk, pengeluaran per kapita ( $X_3$ ) sebagai ukuran tingkat kesejahteraan ekonomi, Indeks Pembangunan Manusia atau IPM ( $X_4$ ) yang menggambarkan kualitas hidup masyarakat melalui aspek pendidikan, kesehatan, dan ekonomi, ketersediaan air bersih ( $X_5$ ) yang berkaitan dengan akses layanan dasar, serta laju pertumbuhan penduduk ( $X_6$ ) yang merefleksikan kecepatan peningkatan jumlah penduduk akibat kelahiran maupun migrasi. Seluruh variabel tersebut dipandang relevan untuk dianalisis menggunakan algoritma CART dan C4.5, sehingga dapat membentuk model klasifikasi yang lebih akurat serta mudah diinterpretasikan dalam konteks kepadatan penduduk.

### 2.3. Teknik Analisis Data

Metode dalam studi ini menerapkan algoritma CART dan C4.5 dengan bantuan software Google Colab. Langkah-langkah analisis data yang diterapkan dalam riset ini dapat dilihat sebagai berikut.

1. Mengumpulkan data, di mana peneliti memperoleh dataset dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara tahun 2023.
2. Mendeskripsikan karakteristik data yang terkumpul menggunakan statistika deskriptif yang memuat nilai rata-rata, minimum, maksimum, standar deviasi, serta visualisasi pie chart untuk persentase klasifikasi kepadatan penduduk tinggi dan rendah.
3. Evaluasi kinerja model algoritma CART dan C4.5.
4. Mendeskripsikan hasil visualisasi berdasarkan model algoritma terbaik
5. Memberikan kesimpulan dan saran.

Pendekatan CART (Classification and Regression Trees) dan C4.5 adalah dua pendekatan dalam pembelajaran algoritma decision tree untuk membangun model pohon keputusan. Berdasarkan prosesnya dapat dilakukan perbandingan antara algoritma CART dan C4.5, dengan mengawali pengujian lalu melihat hasil serta membandingkan kedua metode.

#### 2.3.1 Algoritma CART (Classification and Regression Trees)

Algoritma CART dikenal sebagai salah satu pendekatan analisis yang digunakan untuk membangun model dalam bentuk pohon keputusan. CART memiliki dua pendekatan utama, yaitu classification tree dan regression tree (Putri & Handayani, 2024). Jika variabel dependen yang digunakan bersifat kategorikal, maka model yang terbentuk berupa pohon klasifikasi. Sebaliknya, jika variabel dependen ditampilkan sebagai data kontinu atau numerik, maka hasil yang diperoleh adalah pohon regresi. Berikut Penjelasan tahapan-tahapan dalam penerapan algoritma CART.

##### a. Pemilihan atribut

Dalam algoritma pohon keputusan, pemilihan atribut pemilah dilakukan menggunakan gini index, yaitu ukuran yang digunakan untuk menilai tingkat ketidakhomogenan data pada suatu simpul. Atribut dengan nilai gini index paling rendah akan dipilih karena mampu menghasilkan pemisahan data yang lebih murni. Proses pemilahan kemudian membentuk simpul baru yang akan terus dibagi secara rekursif hingga mencapai terminal nodes.

##### b. Penentuan Simpul Terminal

Simpul terminal ditetapkan apabila jumlah data pada node tersebut berada di bawah ambang minimum ( $N_{\min}$ ) atau ketika pohon telah mencapai kedalaman maksimal yang ditentukan.

##### c. Penanda Kelas Label

Penentuan label kelas pada simpul terminal dilakukan dengan mengacu pada kelas yang memiliki jumlah observasi paling dominan.

#### 2.3.2. Algoritma C4.5

Algoritma C4.5 dikenal sebagai salah satu pendekatan pohon keputusan yang berperan dalam proses klasifikasi dataset. Algoritma ini mampu menangani atribut numerik maupun kategorikal, serta membentuk model berdasarkan hubungan antara atribut dan kelas data. Proses konstruksi pohon keputusan dilakukan secara rekursif dengan memilih atribut yang memiliki efektivitas tertinggi dalam memisahkan data.

Pemilihan atribut tersebut didasarkan pada perhitungan *information gain* yang telah dinormalisasi, sehingga mampu meminimalkan kecenderungan bias terhadap atribut dengan jumlah kategori yang banyak (Agustin & Defit, 2024). Konsep berikut akan menjelaskan bagaimana tahapan algoritma C4.5 dapat diterapkan.

a. Perhitungan *Entropy*

*Entropy* merupakan ukuran ketidakpastian atau keragaman data. Nilai *entropy* tinggi menunjukkan data tersebar merata di banyak kelas, sedangkan nilai rendah menunjukkan dominasi oleh satu atau beberapa kelas saja (Rahmadeyan & Mustakim, 2023).

b. Perhitungan *Information Gain*

*Information gain* berfungsi untuk menilai sejauh mana tingkat entropi berkurang setelah data di partisi berdasarkan suatu atribut. Atribut dengan nilai *information gain* tertinggi kemudian ditetapkan sebagai akar (*root*), karena dianggap paling optimal dalam membentuk kelompok data yang homogen (Puspita et al., 2022).

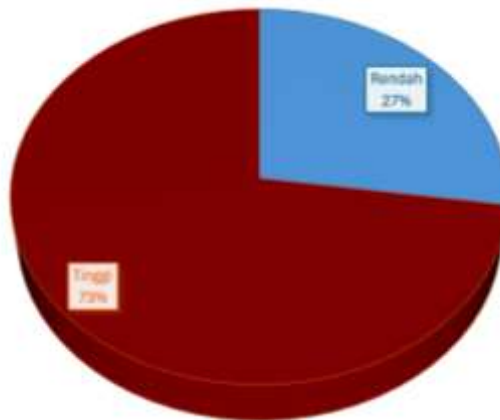
c. Pembentukan Pohon Klasifikasi

Ketika atribut dengan nilai *information gain* tertinggi telah ditentukan sebagai akar (*root*), data kemudian dipisahkan ke dalam cabang sesuai dengan nilai atribut tersebut. Selanjutnya, proses perhitungan *information gain* dan pemisahan data dilakukan kembali secara rekursif pada setiap cabang, hingga seluruh data terklasifikasi atau tidak terdapat lagi atribut yang dapat dijadikan pemisah.

### 3. Hasil dan Pembahasan

#### 3.1. Hasil Penelitian

Kepadatan penduduk dibedakan menjadi dua kategori utama yaitu rendah dan tinggi. Suatu kabupaten/kota digolongkan memiliki kepadatan penduduk tinggi apabila jumlah penduduknya melebihi 500 jiwa per km<sup>2</sup>, sedangkan kabupaten/kota dengan kepadatan penduduk kurang dari atau sama dengan 500 jiwa per km<sup>2</sup> termasuk kategori rendah. Berdasarkan klasifikasi tersebut, diketahui bahwa sebanyak 24 kabupaten/kota berada pada kategori kepadatan tinggi, sedangkan 8 kabupaten/kota lainnya termasuk dalam kategori kepadatan rendah. Hasil klasifikasi tersebut dapat dilihat pada Gambar 1.



**Gambar 1.** Klasifikasi Kepadatan Penduduk

Berdasarkan Gambar 1 menunjukkan visualisasi klasifikasi kepadatan penduduk, bahwa proporsi kabupaten/kota yang dikategorikan memiliki kepadatan penduduk tinggi sebesar 27%, sedangkan 73% sisanya merupakan kabupaten/kota dengan kepadatan penduduk rendah. Hal ini menunjukkan bahwa mayoritas kabupaten dan kota dalam lingkup provinsi tersebut masih menunjukkan tingkat konsentrasi penduduk yang rendah dan cenderung terkonsentrasi pada kawasan tertentu. Tingginya proporsi kabupaten/kota dengan kepadatan rendah mengindikasikan adanya ketimpangan distribusi penduduk yang berpotensi menjadi kendala dalam mendukung efektivitas pelayanan publik maupun kegiatan ekonomi di kawasan tersebut. Hasil visualisasi klasifikasi kepadatan penduduk ini dapat menjadi dasar penting dalam perumusan kebijakan pemerataan pembangunan dengan fokus utama pada peningkatan aksesibilitas, penyediaan infrastruktur, serta pemberian insentif ekonomi di kabupaten/kota yang tergolong berkepadatan

rendah. Selain itu, hasil visualisasi kepadatan penduduk ini dapat dijadikan acuan dalam menentukan arah kebijakan investasi, pengembangan wilayah perdesaan, serta strategi peningkatan kualitas hidup masyarakat secara lebih merata.

Analisis deskriptif dimanfaatkan untuk menyajikan ringkasan umum terkait karakteristik setiap variabel sebagai indikator kepadatan penduduk di setiap kabupaten maupun kota dalam lingkup Provinsi Sumatera Utara pada tahun 2023 dengan hasilnya disajikan pada Tabel 2.

**Tabel 2.** Statistik Deskriptif

	<i>Minimum</i>	<i>Maximum</i>	<i>Mean</i>	<i>Std. Deviation</i>
Jumlah Penduduk	55172	2474166	4662618	5222188
Angka Kelahiran	16	42563	5231.42	8358.50
Pengeluaran Per Kapita	6382	15674	10910.96	2100.48
IPM	64.68	82.61	74.01	4.20
Air Bersih	51.33	100	85.15	14.01
Laju Pertumbuhan	0.58	2.18	1.57	0.40

Berdasarkan Tabel 2 didapatkan karakteristik dari keenam variabel berdasarkan kepadatan penduduk terhadap 33 kabupaten/kota di Provinsi Sumatera Utara tahun 2023, indikator jumlah penduduk memiliki rata-rata sebesar 466261.81 jiwa dengan batas bawah 55172 jiwa dan batas atas 2474166 jiwa serta keragaman data sebesar 522218.79 jiwa. Variabel angka kelahiran menunjukkan rata-rata sebesar 5231.42 jiwa dengan nilai minimum 16 jiwa dan maksimum 42563 jiwa, sedangkan keragaman data sebesar 8358.50 jiwa. Variabel pengeluaran per kapita memiliki rata-rata 10910.96 jiwa berada di dalam rentang antara 6382 jiwa hingga 15674 jiwa serta keragaman data sebesar 2100.48 jiwa. Variabel Indeks Pembangunan Manusia (IPM) memiliki rata-rata sebesar 74.01 berada di dalam rentang batas bawah 64.68 dan batas atas 82.61 serta keragaman data sebesar 4.20. Variabel air bersih menunjukkan rata-rata 85.15% yang berada di dalam rentang nilai minimum 51.33% dan maksimum 100.00% serta keragaman data sebesar 14.01%. Terakhir, variabel laju pertumbuhan penduduk memiliki rata-rata 1.57% berada di dalam rentang nilai minimum 0.58% dan maksimum 2.18% serta keragaman data sebesar 0.40%.

### 3.2. Pembahasan

#### Evaluasi Kinerja Model Algoritma

Evaluasi kinerja algoritma *decision tree*, yakni CART (*Classification and Regression Trees*) dan C4.5, dilakukan sebagai upaya untuk menilai tingkat akurasi, konsistensi, serta efektivitas kedua metode dalam proses klasifikasi data berdasarkan variabel penelitian. Melalui evaluasi ini, diperoleh gambaran yang lebih komprehensif mengenai signifikansi perbedaan performa di antara kedua algoritma, sehingga dapat diidentifikasi metode yang memiliki tingkat akurasi prediksi lebih tinggi dan relevan untuk digunakan pada penelitian ini.

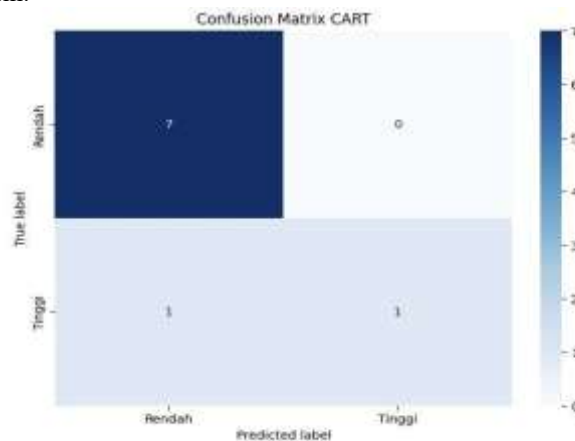
**Tabel 3.** Evaluasi Kinerja Algoritma CART dan C4.5

Partisi	Metode	Akurasi	Presisi	Sensitivitas	Spesifisitas
55% data latih:	CART	80.00%	90.91%	83.34%	66.67%
45% data uji	C4.5	80.00%	90.91%	83.34%	66.67%
60% data latih:	CART	78.58%	90.91%	83.34%	50.00%
40% data uji	C4.5	78.58%	90.91%	83.34%	50.00%
65% data latih:	CART	83.34%	90.00%	90.00%	50.00%
35% data uji	C4.5	83.34%	90.00%	90.00%	50.00%
70% data latih:	CART	80.00%	87.50%	87.50%	50.00%
30% data uji	C4.5	80.00%	87.50%	87.50%	50.00%
75% data latih:	CART	88.89%	87.50%	100.00%	50.00%
25% data uji	C4.5	88.89%	87.50%	100.00%	50.00%
80% data latih:	CART	85.72%	83.34%	100.00%	50.00%
20% data uji	C4.5	85.72%	83.34%	100.00%	50.00%

Berdasarkan Tabel 3, hasil pengujian klasifikasi dengan menggunakan metode algoritma CART (*Classification and Regression Trees*) dan C4.5 pada berbagai proporsi data latih dan data uji menunjukkan bahwa kedua metode menghasilkan performa yang serupa pada setiap skenario. Tingkat akurasi yang diperoleh berada pada rentang 78,58% - 88,89%, dengan capaian tertinggi sebesar 88,89% ketika data dibagi ke dalam proporsi 75% data latih dan 25% data uji. Berdasarkan konfigurasi tersebut, nilai presisi tercatat sebesar 87,5%, sensitivitas mencapai 100%, dan spesifisitas berada pada angka 50%. Berdasarkan hasil pengujian klasifikasi pada tabel, akurasi 88,89% dipandang sebagai representasi terbaik dalam penelitian ini karena menunjukkan kinerja klasifikasi yang paling optimal untuk membedakan tingkat kepadatan penduduk.

### 3.2.1 Algoritma CART (*Classification and Regression Trees*)

CART (*Classification and Regression Trees*) digunakan untuk mengklasifikasikan kepadatan penduduk kabupaten/kota di Provinsi Sumatera Utara berdasarkan tinggi atau rendah, dengan mempertimbangkan 6 variabel utama, yaitu jumlah penduduk, kelahiran, pengeluaran per kapita, Indeks Pembangunan Manusia (IPM), air bersih, dan laju pertumbuhan penduduk dengan pembagian data 25% untuk data uji dan 75% data latih.



**Gambar 2.** *Confusion Matrix* Algoritma CART

Evaluasi *confusion matrix* pada algoritma CART berdasarkan Gambar 2 sebagai berikut:

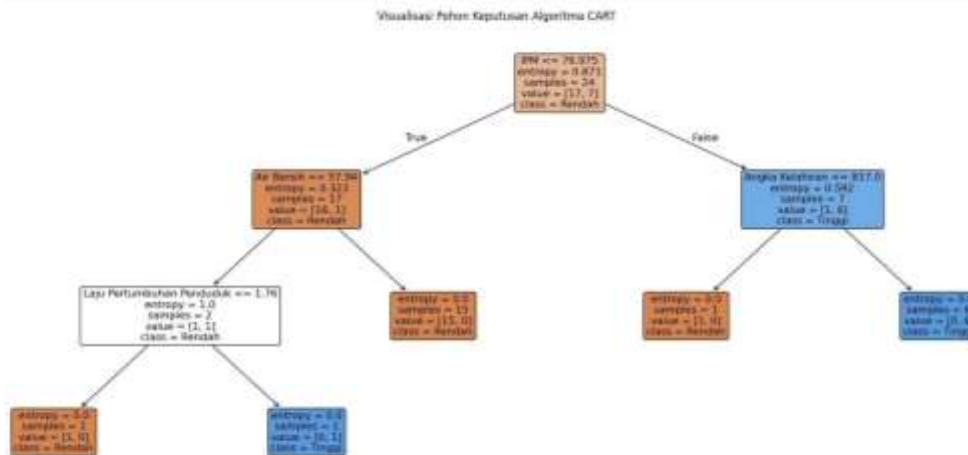
Akurasi	: $(TP + TN) / \text{Total} = (7+1)/9 = 88,89\%$
Presisi	: $TP / (TP + FP) = 7/(7+1) = 87,50\%$
Sensitivitas	: $TP / (TP + FN) = 7/(7+0) = 100,00\%$
Spesifisitas	: $TN / (TN + FP) = 1/(1+1) = 50,00\%$

Berdasarkan evaluasi model algoritma CART pada Gambar 2 terkait status kepadatan penduduk di kabupaten/kota, diperoleh tingkat akurasi sebesar 88,89%, yang berarti dari total 9 kabupaten/kota, sebanyak 8 kabupaten/kota berhasil diprediksi dengan benar sesuai kondisi sebenarnya. Hal ini menunjukkan bahwa secara umum model memiliki kemampuan klasifikasi yang sangat baik.

Nilai presisi untuk kategori “Tinggi” sebesar 100% yang berarti seluruh kabupaten/kota yang diprediksi memiliki kepadatan penduduk tinggi benar-benar berada pada kondisi tersebut. Sementara itu, nilai sensitivitas (*recall*) untuk kategori “Tinggi” hanya sebesar 50% yang berarti model hanya mampu mengenali setengah dari kabupaten/kota yang benar-benar memiliki kepadatan penduduk tinggi. Namun di sisi lain, nilai spesifisitas mencapai 100%, sehingga model sangat efektif dalam mengidentifikasi kabupaten/kota dengan kepadatan penduduk rendah. Semua kabupaten/kota yang sebenarnya berkepadatan rendah berhasil diprediksi dengan tepat oleh model.

Secara keseluruhan, meskipun algoritma CART menunjukkan akurasi dan spesifisitas yang tinggi, kelemahan utamanya terletak pada ketidakmampuan model untuk mendeteksi sebagian kabupaten/kota yang sebenarnya berada pada kategori kepadatan penduduk tinggi.

Visualisasi CART (*Classification and Regression Trees*) dengan menggunakan aplikasi *Google Colab* disajikan sebagai Gambar 3 berikut.



**Gambar 3.** Visualisasi Pohon Keputusan Algoritma CART

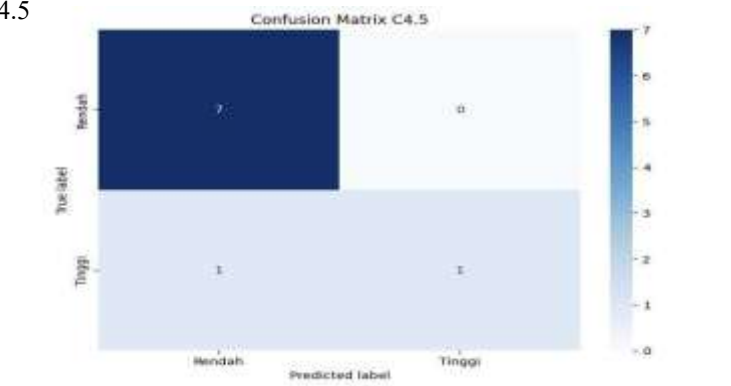
Struktur pohon keputusan berdasarkan Gambar 3 diperoleh sebagai berikut.

- $IPM \leq 76,795$ , Air Bersih  $\leq 57,94\%$ , Laju Pertumbuhan  $\leq 1,76\%$ ; maka wilayah memiliki kepadatan penduduk rendah
- $IPM \leq 76,795$ , Air Bersih  $\leq 57,94\%$ , Laju Pertumbuhan  $> 1,76\%$ ; maka wilayah memiliki kepadatan penduduk tinggi
- $IPM \leq 76,795$ , Air Bersih  $> 57,94\%$ ; maka wilayah memiliki kepadatan penduduk rendah
- $IPM > 76,795$ , Kelahiran  $\leq 817$  jiwa; maka wilayah memiliki kepadatan penduduk rendah
- $IPM > 76,795$ , Kelahiran  $> 817$  jiwa; maka wilayah memiliki kepadatan penduduk tinggi

Berdasarkan hasil visualisasi CART (*Classification and Regression Trees*) pada Gambar 3, diketahui bahwa Indeks Pembangunan Manusia (IPM) menjadi akar pohon atau variabel utama yang digunakan model dalam memisahkan kategori kepadatan penduduk tinggi dan rendah pada kabupaten/kota dalam lingkup Provinsi Sumatera Utara. Ambang batas yang digunakan pada IPM sebesar 76,795 yang menunjukkan bahwa wilayah dengan tingkat pembangunan manusia di bawah nilai tersebut rentan dikategorikan memiliki kepadatan penduduk tinggi, terutama jika dipengaruhi oleh faktor lain seperti keterbatasan akses air bersih dan tingginya angka kelahiran.

Meskipun demikian, terdapat dua variabel yaitu jumlah penduduk dan pengeluaran per kapita yang tidak muncul dalam struktur pohon keputusan. Hal ini dapat dijelaskan karena algoritma hanya memilih variabel yang memiliki pengaruh paling kuat dalam membedakan kategori kepadatan penduduk tinggi atau rendah. Dengan demikian, variabel Indeks Pembangunan Manusia (IPM) sudah cukup dominan dalam menjelaskan variasi antar wilayah, yang kemudian diperkuat oleh faktor pendukung lain seperti akses air bersih, laju pertumbuhan penduduk, serta angka kelahiran, sehingga model tidak menggunakan variabel lain yang informasinya dianggap tercukupi.

### 3.2.2 Algoritma C4.5



**Gambar 4.** Confusion Matrix Algoritma C4.5

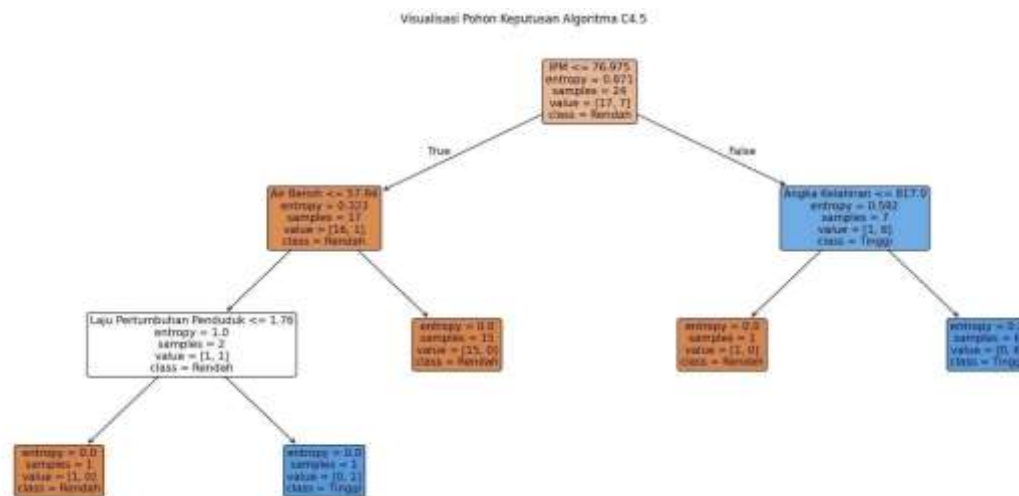
Evaluasi *confusion matrix* pada algoritma C4.5 berdasarkan Gambar 4 sebagai berikut.

Akurasi :  $(TP + TN) / Total = (7+1)/9 = 88.89 \%$   
 Presisi :  $TP / (TP + FP) = 7/(7+1) = 87.50 \%$   
 Sensitivitas :  $TP / (TP + FN) = 7/(7+0) = 100.00 \%$   
 Spesifisitas :  $TN / (TN + FP) = 1/(1+1) = 50.00 \%$

Berdasarkan hasil evaluasi *confusion matrix* pada algoritma C4.5 diperoleh tingkat akurasi sebesar 88,89% yang menunjukkan bahwa secara keseluruhan model mampu memprediksi dengan benar sebanyak 8 dari 9 data kabupaten/kota di Provinsi Sumatera Utara. Hal ini menunjukkan bahwa kinerja klasifikasi model yang cukup baik dalam mengidentifikasi kategori kepadatan penduduk.

Nilai presisi yang diperoleh sebesar 87,50% diartikan dari seluruh kabupaten/kota yang diprediksi memiliki kepadatan penduduk tinggi, sebesar 87,50% di antaranya benar-benar berada pada kondisi tersebut. Dengan demikian, model cukup andal dalam mengurangi kesalahan prediksi positif (*false positive*). Sementara itu, nilai sensitivitas mencapai 100% yang berarti seluruh kabupaten/kota dengan kepadatan penduduk tinggi dapat teridentifikasi secara tepat oleh model tanpa ada satupun yang terlewat. Hal ini menunjukkan bahwa algoritma C4.5 memiliki kemampuan deteksi yang sempurna terhadap kategori tinggi. Namun, nilai spesifisitas hanya sebesar 50% yang menunjukkan model kurang optimal dalam mengenali kabupaten/kota dengan kepadatan penduduk rendah. Dengan kata lain, hanya separuh dari wilayah yang berkepadatan rendah dapat diprediksi secara tepat, sedangkan sisanya masih terklasifikasi salah atau keliru.

Secara keseluruhan, algoritma C4.5 memiliki kinerja yang sangat baik dalam mendeteksi wilayah dengan kepadatan penduduk tinggi, namun masih memiliki keterbatasan dalam membedakan wilayah dengan kepadatan penduduk rendah. Oleh karena itu, meskipun akurasi dan sensitivitas tinggi, nilai spesifisitas yang rendah perlu menjadi perhatian agar model kurang optimal terhadap kategori tertentu.



**Gambar 5.** Visualisasi Pohon Keputusan Algoritma C4.5

Struktur pohon keputusan berdasarkan Gambar 5 diperoleh sebagai berikut.

- $IPM \leq 76,795$ , Air Bersih  $\leq 57,94\%$ , Laju Pertumbuhan  $\leq 1,76\%$ ; maka wilayah memiliki kepadatan penduduk rendah
- $IPM \leq 76,795$ , Air Bersih  $\leq 57,94\%$ , Laju Pertumbuhan  $> 1,76\%$ ; maka wilayah memiliki kepadatan penduduk tinggi
- $IPM \leq 76,795$ , Air Bersih  $> 57,94\%$ ; maka wilayah memiliki kepadatan penduduk rendah
- $IPM > 76,795$ , Kelahiran  $\leq 817$  jiwa; maka wilayah memiliki kepadatan penduduk rendah
- $IPM > 76,795$ , Kelahiran  $> 817$  jiwa; maka wilayah memiliki kepadatan penduduk tinggi

Berdasarkan hasil pohon keputusan menggunakan algoritma C4.5 menunjukkan hasil analisis Indeks Pembangunan Manusia menjadi variabel utama dalam memisahkan kategori kepadatan penduduk tinggi dan rendah kabupaten/kota di Provinsi Sumatera Utara. Ambang batas IPM sebesar 76,795 menunjukkan bahwa

wilayah dengan nilai di bawah angka tersebut cenderung memiliki kepadatan penduduk tinggi, terutama jika disertai keterbatasan air bersih dan tingginya angka kelahiran. Sementara itu, variabel jumlah penduduk dan pengeluaran per kapita tidak muncul dalam model, karena IPM bersama faktor lain seperti air bersih, laju pertumbuhan, dan kelahiran sudah cukup menjelaskan perbedaan kepadatan penduduk antar wilayah.

Algoritma C4.5 menghasilkan analisis model yang sama dengan algoritma CART (*Classification and Regression Trees*), dengan Indeks Pembangunan Manusia sebagai variabel utama dalam memisahkan kategori kepadatan penduduk tinggi dan rendah, dengan diperkuat oleh faktor lain seperti akses air bersih, laju pertumbuhan, dan kelahiran. Selain itu, variabel jumlah penduduk dan pengeluaran per kapita tidak muncul dalam model.

#### Perbandingan Algoritma CART (*Classification and Regression Trees*) dan C4.5

Penelitian ini menerapkan dua metode klasifikasi dalam machine learning untuk mengetahui wilayah kepadatan penduduk kabupaten/kota di Provinsi Sumatera Utara, yaitu dengan algoritma CART (*Classification and Regression Trees*) dan C4.5 dengan tujuan untuk membandingkan kinerja keduanya serta mengidentifikasi metode yang mampu menghasilkan tingkat akurasi klasifikasi paling optimal. Disajikan pada Tabel 4 berikut.

**Tabel 4.** Hasil Akurasi CART Dan C4.5

Metode	Akurasi (%)
CART	88.89%
C4.5	88.89%

Berdasarkan hasil Tabel 4, algoritma CART (*Classification and Regression Trees*) dan C4.5 sama-sama memperoleh tingkat akurasi 88,89%. Capaian ini mengindikasikan bahwa kedua algoritma memiliki kinerja yang sebanding dalam membedakan tinggi dan rendahnya kepadatan penduduk kabupaten/kota di Provinsi Sumatera Utara dengan menggunakan 6 (enam) indikator sosial ekonomi dan demografi. Oleh karena itu, baik CART (*Classification and Regression Trees*) maupun C4.5 dapat dijadikan alternatif metode yang sama-sama layak dalam proses klasifikasi kepadatan penduduk.

#### Visualisasi Fitur Kepentingan



**Gambar 6.** Fitur Kepentingan Pohon Keputusan

Hasil analisis fitur kepentingan dari kedua model terbaik yang diolah melalui *Google Colab* ditunjukkan pada Gambar 6. Berdasarkan hasil tersebut, variabel Indeks Pembangunan Manusia (IPM) memiliki tingkat pengaruh tertinggi dengan nilai sebesar 0.64%, diikuti variabel kelahiran dengan nilai 0.17%, dan akses air bersih 0.08%, serta laju pertumbuhan 0.01%. Sementara itu, variabel jumlah penduduk dan pengeluaran per kapita memperoleh nilai *importance* 0.00%, sehingga tidak memberikan pengaruh atau kontribusi terhadap model. Berdasarkan kerangka *machine learning*, khususnya pada CART (*Classification and Regression Trees*) dan C4.5, nilai *importance* merepresentasikan besarnya peran suatu variabel dalam meningkatkan akurasi klasifikasi. Dengan demikian, hasil menegaskan bahwa Indeks Pembangunan Manusia

(IPM) merupakan variabel dominan dalam menentukan tingkat kepadatan penduduk tinggi dan rendah pada dataset yang dianalisis.

### 3.3. Uraian Hasil Analisis

Hasil klasifikasi dengan algoritma CART dan C4.5 menunjukkan akurasi tertinggi sebesar 88.89% pada proporsi data latih 75% dan data uji 25%. Tingkat akurasi tersebut berarti dari sembilan kabupaten/kota yang dianalisis, delapan di antaranya berhasil diprediksi sesuai kondisi kepadatan penduduk aktual. Pohon keputusan memperlihatkan bahwa Indeks Pembangunan Manusia (IPM) merupakan variabel utama dalam membedakan kategori kepadatan tinggi dan rendah dengan ambang batas 76.79. Temuan ini mengindikasikan bahwa wilayah dengan kualitas pembangunan manusia lebih rendah cenderung masuk kategori padat penduduk terutama ketika diikuti oleh keterbatasan akses air bersih dan tingginya angka kelahiran. Analisis fitur kepentingan memperkuat hasil tersebut karena IPM memiliki nilai importance tertinggi yaitu 0.64% diikuti angka kelahiran 0.17% laju pertumbuhan penduduk 0.10% dan air bersih 0.08%. Sementara itu jumlah penduduk dan pengeluaran per kapita tidak muncul sebagai variabel signifikan sebab kontribusinya telah terwakili oleh variabel IPM, angka kelahiran, laju pertumbuhan penduduk dan persentase air bersih. Dengan demikian penelitian ini menegaskan bahwa dimensi pembangunan manusia dan faktor demografi lebih berpengaruh dalam menentukan kepadatan penduduk dibandingkan ukuran kuantitatif murni seperti jumlah penduduk atau tingkat pengeluaran per kapita.

Hasil dari penelitian ini sejalan dengan teori transisi demografi yang menekankan bahwa pembangunan manusia melalui pendidikan dan kesehatan berperan penting dalam menentukan distribusi serta dinamika penduduk. Hasil yang diperoleh juga konsisten dengan penelitian Maharani et al. (2024) yang menunjukkan bahwa IPM dan angka kelahiran merupakan variabel paling berpengaruh dalam membedakan kepadatan penduduk antar wilayah di Indonesia. Berbeda dengan hipotesis awal yang memperkirakan jumlah penduduk menjadi faktor dominan, analisis justru memperlihatkan bahwa variabel tersebut tidak signifikan dalam model klasifikasi. Kondisi ini kemungkinan dipengaruhi oleh variasi distribusi penduduk yang tinggi sehingga kontribusi jumlah penduduk secara statistik tertutupi oleh kualitas pembangunan manusia. Tidak munculnya pengeluaran per kapita sebagai variabel penting juga dapat dijelaskan oleh faktor sementara seperti bantuan sosial atau fluktuasi harga pangan sehingga indikator ini kurang konsisten dalam membedakan kategori kepadatan. Hasil dari penelitian ini perlu adanya kebijakan pembangunan yang tidak semata-mata difokuskan pada peningkatan pertumbuhan ekonomi, melainkan juga peningkatan kualitas hidup melalui perbaikan pendidikan, kesehatan, dan infrastruktur dasar. Dengan demikian strategi pengendalian kepadatan penduduk di Sumatera Utara lebih tepat diarahkan pada penguatan sumber daya manusia dan pemerataan pembangunan agar konsentrasi penduduk tidak berlebihan di wilayah tertentu sekaligus mengurangi kesenjangan antar wilayah.

Berdasarkan hasil visualisasi model terbaik yaitu CART (*Classification and Regression Trees*) dan C4.5 dapat diidentifikasi bahwa variabel Indeks Pembangunan Manusia (IPM) menjadi faktor prioritas utama yang paling berpengaruh dalam menentukan kategori kepadatan penduduk kabupaten/kota di Provinsi Sumatera Utara. Variabel ini memiliki kontribusi tertinggi dibandingkan variabel lain sehingga menegaskan bahwa kualitas pembangunan manusia mencakup aspek pendidikan, kesehatan, dan standar hidup harus menjadi fokus utama dalam perumusan strategi kebijakan pemerintah Provinsi Sumatera Utara. Selain IPM variabel akses air bersih angka kelahiran dan laju pertumbuhan penduduk juga berkontribusi dalam mempengaruhi hasil klasifikasi meskipun dengan tingkat kontribusi yang lebih rendah.

Pemerintah Provinsi Sumatera Utara perlu merumuskan strategi kebijakan yang lebih komprehensif dan berkelanjutan. Peningkatan kualitas Sumber Daya Manusia (SDM) dapat diwujudkan melalui perluasan akses pendidikan berkualitas, peningkatan kompetensi tenaga kerja, serta program pemberdayaan masyarakat berbasis keterampilan lokal yang mampu menciptakan kemandirian ekonomi. Penyediaan layanan kesehatan reproduksi harus diperkuat melalui peningkatan fasilitas kesehatan di wilayah terpencil, edukasi kesehatan bagi remaja dan keluarga, serta program pengendalian angka kelahiran yang berfokus pada perencanaan keluarga. Pemerataan akses air bersih dapat ditempuh dengan pembangunan infrastruktur penyediaan air bersih berbasis potensi lokal seperti pemanfaatan sumber mata air dan teknologi pengolahan sederhana, sekaligus mendorong partisipasi masyarakat dalam menjaga kualitas lingkungan. Pengendalian laju pertumbuhan penduduk perlu dirancang melalui kebijakan insentif dan regulasi yang mendorong keseimbangan antara pertumbuhan penduduk dan daya dukung wilayah. Seluruh strategi ini harus terintegrasi

dengan perencanaan pembangunan daerah disertai monitoring dan evaluasi berbasis data sehingga kebijakan yang diterapkan tidak hanya menjawab tantangan kependudukan saat ini tetapi juga mampu mewujudkan pembangunan wilayah yang inklusif, merata, dan berkelanjutan.

#### 4. Kesimpulan

Hasil klasifikasi menggunakan algoritma CART dan C4.5 menunjukkan tingkat akurasi tertinggi sebesar 88,89% dengan kinerja yang relatif sebanding. Temuan ini mengindikasikan bahwa kedua algoritma mampu mengklasifikasikan kategori kepadatan penduduk secara efektif. Indeks Pembangunan Manusia (IPM) terbukti sebagai variabel yang paling dominan dalam membedakan kategori kepadatan penduduk, diikuti oleh faktor kelahiran, laju pertumbuhan penduduk, dan ketersediaan air bersih. Sementara itu, variabel jumlah penduduk dan pengeluaran per kapita tidak menunjukkan pengaruh yang signifikan karena informasi yang dikandungnya telah terwakili oleh variabel lain dalam model.

Lebih lanjut, hasil visualisasi dari model terbaik CART dan C4.5 memperkuat temuan bahwa IPM merupakan variabel prioritas utama dengan kontribusi tertinggi dalam menentukan kategori kepadatan penduduk kabupaten/kota di Provinsi Sumatera Utara. Hal ini menunjukkan bahwa perumusan strategi kebijakan oleh Pemerintah Provinsi Sumatera Utara sebaiknya difokuskan pada peningkatan kualitas pembangunan manusia, terutama melalui penguatan sektor pendidikan, kesehatan, dan standar hidup layak. Upaya tersebut perlu didukung oleh pengelolaan variabel lain seperti peningkatan akses air bersih, pengendalian angka kelahiran, serta pengaturan laju pertumbuhan penduduk agar distribusi kepadatan penduduk dapat lebih seimbang.

Penelitian selanjutnya disarankan untuk mengembangkan pendekatan analisis dengan menggunakan metode ensemble, seperti Random Forest atau Gradient Boosting, serta menambahkan variabel prediktor lain, antara lain tingkat urbanisasi, migrasi, pengangguran, dan akses terhadap infrastruktur dasar, guna meningkatkan akurasi dan ketahanan model klasifikasi. Selain itu, Pemerintah Provinsi Sumatera Utara dapat memanfaatkan hasil klasifikasi ini sebagai dasar dalam merumuskan kebijakan pemerataan pembangunan. Strategi pengendalian kepadatan penduduk sebaiknya diarahkan pada peningkatan kualitas hidup masyarakat melalui pendidikan dan kesehatan, mengingat peran signifikan kedua aspek tersebut dalam distribusi penduduk. Di samping itu, intervensi pembangunan perlu difokuskan pada pengurangan konsentrasi penduduk di wilayah perkotaan sekaligus memperkuat daya tarik ekonomi daerah dengan tingkat kepadatan penduduk yang lebih rendah.

#### Bibliografi

- Agustin, R., Sarjon Defit, & Sumijan. (2024). Perbandingan Algoritma CART dan C.4 5 Pada Citra Tandan Buah Sawit Untuk Mengetahui Tingkat Kematangan Dalam Penentuan Harga. *Jurnal KomtekInfo*, 11(4), 263–273. <https://doi.org/10.35134/komtekinfo.v11i4.558>
- Asmara, I. M. A. P., Saimi, S., Putri, F. D. A., & Swandayana, P. G. W. (2024). Analisis Hubungan Kemampuan Pelayanan Kesehatan dengan Rate Rujukan Puskesmas di Wilayah Kerja BPJS Kesehatan Cabang Mataram Tahun 2024. *Bioscientist : Jurnal Ilmiah Biologi*, 12(2), 2072. <https://doi.org/10.33394/bioscientist.v12i2.13233>
- Avcı, K. (2025). Evaluation of Public Hospitals' Performance with Decision Tree Algorithms. *Verimlilik Dergisi*, 59(1), 133–142. <https://doi.org/10.51551/verimlilik.1494277>
- Azzahra, F., Azzarah, R. A., Harahap, M. A., Sianipar, S. E., & Sipahutar, R. A. S. (2025). Integrasi SIG dalam Analisis Kepadatan Penduduk di Desa Limau Manis, Desa Medan Sinembah, dan Desa Ujung Serdang. *JPIG (Jurnal Pendidikan dan Ilmu Geografi)*, 10(1), 84-90.
- BPS Sumatera Utara. (2023). Provinsi Sumatera Utara Dalam Angka 2023. *Badan Pusat Statistik Provinsi Sumatera Utara*. <https://doi.org/10.21067/jpig.v10i1.1039>
- Daniya, T., Geetha, M., & Suresh Kumar, K. (2020). Classification And Regression Trees With Gini Index. *Advances in Mathematics: Scientific Journal*, 9(10), 8237–8247. <https://doi.org/10.37418/amsj.9.10.53>
- Dita, C. Y. E., & Legowo, M. (2022). Analisis kepadatan penduduk yang berpengaruh terhadap kemiskinan dan degradasi lingkungan. In *Prosiding Seminar Nasional Ilmu Ilmu Sosial (SNIIS)* (Vol. 1, pp. 1-12).
- Irham, A. R., & Putri, R. M. (2023). Kepadatan Penduduk terhadap Indeks Pembangunan Manusia di Provinsi Lampung. *Media Komunikasi Geografi*, 24(1), 91–100. <https://doi.org/10.23887/mkg.v24i1.60261>

- Kuswanto, A. D., Badjo, H. N., Kharist, S., Mubarak, M. Z., Saputra, R., & Fitroh, R. M. (2024). Penerapan Algoritma C4. 5 Dalam Klasifikasi Prestasi Atlet: Studi Kasus Pada Daftar Nama Penerima Penghargaan Tahun 2023. *Bridge: Jurnal Publikasi Sistem Informasi dan Telekomunikasi*, 2(3), 103-113. <https://doi.org/10.62951/bridge.v2i3.115>
- Lestari, I., Fitria, D., Syafrandi, & Salma, A. (2024). Comparison Of The C5.0 Algorithm And The CART Algorithm In Stroke Classification. *UNP Journal of Statistics and Data Science*, 2(1), 90–98. <https://doi.org/10.24036/ujsds/vol2-iss1/144>
- Maharani, A., Putri, A. E., & Wulandari, S. P. (2024). Pengaruh Kepadatan Penduduk Terhadap Kualitas Hidup Masyarakat Di Indonesia Tahun 2023 Menggunakan Metode MANOVA. *Multi Proximity: Jurnal Statistika Universitas Jambi*, 3(2), 68–69. <https://doi.org/10.22437/multiproximity.v3i2.38890>
- Mardiansyah, H., Zarlis, M., & Sitompul, O. S. (2021). Analysis Of C4.5 Algorithm Of Water Quality Dataset. *Journal of Physics: Conference Series*, 1898(1), 1–6. <https://doi.org/10.1088/1742-6596/1898/1/012002>
- Puspita, D., Aminah, S., & Arif, A. (2022). Prediction System For Credit Eligibility Using C4.5 Algorithm. *Journal of Informatics and Telecommunication Engineering*, 6(1), 148–156. <http://ojs.uma.ac.id/index.php/jite>
- Putri, K. A., & Handayani, H. H. (2024). Perbandingan Hasil Klasifikasi Tutupan Lahan Menggunakan Metode Artificial Neural Network (ANN), Support Vector Machine (SVM), dan Random Forest (RF) Dengan Bahasa Pemrograman R. *Jurnal Geodesy and Geomatics*, 19(2), 349–360. <https://doi.org/10.12962/geoid.v19i2.1145>
- Rahmadiyah, A., & Mustakim. (2023). Seleksi Fitur Pada Supervised Learning: Klasifikasi Prestasi Belajar Mahasiswa Saat Dan Pasca Pandemi COVID-19. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 9(1), 21–32. <https://doi.org/10.25077/teknosi.v9i1.2023.21-32>
- Ruliansyah, M. J., & Betty, M. (2024). Penerapan Metode C4.5 Dalam Prediksi Penjualan Tim Bev 1 Pada PT. Surya Pangan Sejahtera Bekasi Jawa Barat. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi (JIMIK)*, 5(2), 1269–1278. <https://doi.org/10.35870/jimik.v5i2.664>
- Shamrat, F. M. J. M., Ranjan, R., Hasib, K. M., Yadav, A., & Siddique, A. H. (2022). Performance Evaluation Among ID3, C4.5, And CART Decision Tree Algorithm. *Lecture Notes in Networks and Systems*, 317, 127–142.
- Siagian, Nilasari. (2021). Population Distribution And Density Changing Pattern In Sumatera Utara. *Jurnal Geografi*, 13(2), 225–234. <https://doi.org/10.24114/jg.v13i2.25428>
- Widyawati, D., Faradibah, A., & Belluano, P. L. L. (2023). Comparison Analysis Of Classification Model Performance In Lung Cancer Prediction Using Decision Tree, Naive Bayes, And Support Vector Machine. *Indonesian Journal of Data and Science*, 4(2), 80–89. <https://doi.org/10.56705/ijodas.v4i2.76>