

Klasifikasi Kemiskinan Kabupaten dan Kota di Indonesia Berdasarkan Indikator Sosial Ekonomi Menggunakan Algoritma C4.5

Lorencia Anugrah Suseno¹, Tio Basami Hutapea², Arga Kurniawan³

^{1,2,3}Program Studi Sistem Informasi, Universitas Bina Sarana Informatika, Bekasi, Jawa Barat
Email: ¹lorenciaanugrah2005@gmail.com, ²tiobasami70@gmail.com, ³argakurniawan140@gmail.com

Abstrak

Kemiskinan masih menjadi permasalahan sosial utama di Indonesia sehingga analisis faktor penyebabnya sangat diperlukan untuk membantu pemerintah dalam mengambil keputusan yang tepat. Penelitian ini bertujuan untuk melakukan klasifikasi tingkat kemiskinan menggunakan algoritma C4.5 dengan memanfaatkan dataset yang berisi berbagai indikator sosial-ekonomi seperti tingkat pendidikan, pengangguran, dan pengeluaran per kapita. Tahapan penelitian dimulai dari pengumpulan dan pembersihan data, transformasi, pembagian dataset menjadi data latih dan uji, serta pembangunan model klasifikasi. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan tingkat kemiskinan secara efektif dan menghasilkan pola pohon keputusan yang jelas. Berdasarkan model yang terbentuk, variabel pengeluaran per kapita teridentifikasi sebagai faktor dominan yang paling mempengaruhi status kemiskinan di suatu wilayah. Model ini diharapkan dapat menjadi dasar dalam penyusunan kebijakan yang lebih tepat sasaran untuk mengurangi tingkat kemiskinan di Indonesia.

Kata Kunci: algoritma c4.5, data mining, klasifikasi tingkat kemiskinan, pengeluaran per kapita, pohon keputusan.

Classification of Poverty Levels in Indonesian Regencies and Cities Based on Socio Economic Indicators Using the C4.5 Algorithm

Abstract

Poverty remains major social issue in Indonesia, so an analysis of its causal factors is necessary to assist the government in making informed decisions. This study aims to classify poverty levels using the C4.5 algorithm by utilizing a dataset containing various socio-economic indicators such as education level, unemployment, and per capita expenditure. The research stages began with data collection and cleaning, transformation, splitting the dataset into training and testing data, and building the classification model. The results show that the C4.5 algorithm is capable of classifying poverty levels effectively and producing clear decision tree patterns. Based on the generated model, the per capita expenditure variable was identified as the dominant factor most influencing poverty status in a region. This model is expected to serve as a basis for formulating more targeted policies to reduce poverty levels in Indonesia.

Keywords: *c4.5 algorithm, data mining, decision tree, per capita expenditure, poverty level classification.*

I. PENDAHULUAN

Kemiskinan adalah salah satu persoalan serius yang dihadapi oleh banyak negara yang sedang berkembang, termasuk Indonesia. Berdasarkan data dari BPS, kemiskinan dapat didefinisikan sebagai ketidakmampuan individu untuk memenuhi kebutuhan dasar primer dan sekunder yang merupakan standar minimal. Akibat dari kondisi kemiskinan ini antara lain menciptakan situasi di mana banyak anak tidak memperoleh pendidikan yang memadai, kesulitan dalam membiayai kesehatan, tabungan yang sangat terbatas, akses terhadap layanan publik yang tidak terpenuhi, rendahnya perlindungan sosial bagi keluarga, serta tingginya laju urbanisasi menuju kota [1].

Sebagai Negara berkembang, Indonesia turut menghadapi permasalahan kemiskinan, dengan populasi yang menempati urutan keempat terbanyak di dunia, yaitu 286 juta jiwa atau 3,47% dari seluruh populasi global, Indonesia dihadapkan pada berbagai persoalan. Tingginya angka kemiskinan di Indonesia dipicu oleh sejumlah faktor, seperti pertumbuhan ekonomi yang lambat, inflasi, tingkat pengangguran yang tinggi, dan kondisi pendidikan yang kurang memadai [1].

Dalam dekade terakhir, penerapan Machine Learning untuk analisis sosio-ekonomi telah berkembang pesat dengan berbagai pendekatan algoritma. Penelitian terdahulu telah berupaya memetakan masalah kemiskinan menggunakan metode statistik konvensional maupun komputasi cerdas. Amsat et al [1] menggunakan Regresi Logistik Biner untuk

mengukur kedalaman kemiskinan antar provinsi. Meskipun metode ini efektif untuk melihat probabilitas hubungan antar variabel, Regresi Logistik memiliki keterbatasan dalam memvisualisasikan aturan Keputusan yang kompleks secara hierarkis, sehingga sulit digunakan untuk menentukan ambang batas (threshold) yang spesifik bagi pengambil kebijakan awam.

Selain pendekatan statistik, algoritma klasifikasi berbasis probabilitas seperti Naïve Bayes juga sering digunakan. Namun, studi komparasi yang dilakukan oleh Martin dan Nilawati [2] serta Adawiyah et al. [3] menunjukkan bahwa meskipun Naïve Bayes bekerja cepat pada dataset besar, algoritma ini memiliki kelemahan asumsi independensi antar fitur yang sering kali tidak terpenuhi dalam data sosial-ekonomi nyata. Selain itu, model probabilitas cenderung bersifat abstrak dalam menjelaskan kausalitas (sebab-akibat) langsung kepada pemangku kepentingan (stakeholders).

Berbeda dengan metode-metode tersebut, algoritma C4.5 menawarkan pendekatan *White Box* yang transparan. Santosa et al. [4] membuktikan keunggulan ini saat menerapkan C4.5 untuk penentuan penerima Bantuan Langsung Tunai (BLT), di mana pohon keputusan yang dihasilkan mampu memberikan justifikasi logis “Jika-Maka” (*If-Then-Rules*) yang mudah diaudit. Hal ini diperkuat oleh Hutasuhut et al. [5] yang menemukan bahwa C4.5 memiliki kemampuan penanganan missing values dan data betipe campuran (nominal dan numerik) yang lebih baik dibandingkan algoritma ID3 pendahulunya. Kemampuan interpretabilitas (*interpretability*) inilah yang menjadi alasan utama pemilihan metode C4.5 dalam penelitian ini. Kemampuan ini juga dikonfirmasi oleh Lestari dan Agustiansyah [6] yang memanfaatkan C4.5 untuk merekomendasikan bantuan sosial berdasarkan kluster data penduduk miskin. Lebih lanjut, penelitian terbaru oleh Saputra et al. [7] menunjukkan bahwa penggunaan algoritma C4.5 pada data program bantuan pemerintah (seperti PIP) terbukti efektif dalam meningkatkan ketepatan sasaran penerima manfaat.

Dalam konteks pengentasan kemiskinan, pemerintah tidak hanya membutuhkan prediksi akurasi tinggi, tetapi juga penjelasan mengapa suatu wilayah diklasifikasikan miskin. Metode *Black Box* seperti Neural Network atau *Support Vector Machine* (SVM), meskipun mungkin memiliki akurasi tinggi, sering kali gagal memberikan transparansi aturan yang dibutuhkan untuk landasan hukum kebijakan publik. Oleh karena itu, Algoritma C4.5 dipilih sebagai metode utama dalam penelitian ini karena kemampuannya menyeimbangkan akurasi prediksi dengan kejelasan struktur keputusan, yang sangat krusial untuk menghasilkan rekomendasi kebijakan yang tepat sasaran dan akuntabel [8],[9].

Penelitian ini berfokus pada penerapan algoritma C4.5 untuk mengklasifikasikan tingkat kemiskinan di tingkat kabupaten / kota. Tujuannya adalah menghasilkan model klasifikasi yang tidak hanya objektif dan akurat, tetapi juga memberikan pemahaman eksplisit mengenai aturan logis yang memisahkan status kemiskinan suatu wilayah.

II. METODOLOGI PENELITIAN

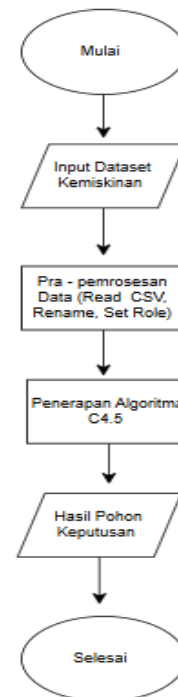
A. Pendekatan Penelitian

Penelitian ini menerapkan pendekatan kuantitatif melalui teknik *Data Mining* dengan memanfaatkan algoritma C4.5 untuk mengklasifikasikan status kemiskinan di Indonesia. Pemilihan metode ini didasarkan pada kemampuannya dalam mengungkap pola hubungan antar variabel sosio - ekonomi, serta menghasilkan model prediksi yang strukturnya mudah dipahami (*interpretable*). Sebagaimana dijelaskan oleh Mardi [9], *Data Mining* merupakan komponen krusial dalam tahapan *Knowledge Discovery in Database* (KDD) untuk mengekstrasi informasi berharga dari kumpulan data.

Algoritma klasifikasi yang digunakan adalah C4.5 yang bekerja dengan cara membangun struktur pohon keputusan (*Decision Tree*) Merujuk pada Putra dan Wijaya [10], algoritma ini menyeleksi atribut pemecah terbaik berdasarkan perhitungan nilai *Gain* dan *Gain Ratio* dari data berlabel. Model pohon keputusan yang terbentuk selanjutnya divalidasi menggunakan data uji untuk mengukur tingkat akurasi dan kemampuan generalisasinya.

B. Tahapan Penelitian

Kerangka kerja penelitian ini terbagi menjadi tiga fase utama, yaitu pengumpulan data, pra-pemrosesan (preprocessing), dan pembentukan model klasifikasi. Alur lengkap dari tahapan penelitian ini diilustrasikan dalam diagram alir (*flowchart*) pada Gambar 1.



Gambar 1. Diagram Alir Tahapan Penelitian

Berdasarkan Gambar 1, alur penelitian dimulai dengan proses input dataset kemiskinan yang diperoleh dari repositori publik. Data mentah tersebut kemudian masuk ke tahap pra – pemrosesan yang meliputi pembersihan data, penyesuaian tipe atribut, dan pelabelan kelas target. Setelah data siap,

proses dilanjutkan dengan pembagian data (*split data*) dan penerapan algoritma C4.5 untuk membangun model. Model yang terbentuk (Pohon Keputusan) kemudian dievaluasi kinerjanya sebelum penelitian dinyatakan selesai. Tahapan sistematis ini dirancang untuk memastikan bahwa hasil klasifikasi dapat dipertanggungjawabkan secara ilmiah.

C. Pengumpulan dan Pra-pemrosesan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari repositori publik Kaggle. Dataset berjudul “Klasifikasi Tingkat Kemiskinan di Indonesia” ini memiliki total 514 data (*records*) yang merepresentasikan seluruh Kabupaten / Kota di Indonesia, dengan 13 atribut yang menyertainya. File data disimpan dalam format CSV (*Comma Separated Values*) dan mencakup aspek ekonomi, Pendidikan, Kesehatan, dan infrastruktur dasar. Rincian atribut yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Deskripsi Variabel Dataset

No	Atribut	Tipe Data	Keterangan
1	Provinsi	Nominal	Nama Provinsi
2	Kab/Kota	Nominal	Nama Kabupaten/Kota
3	Persentase Penduduk Miskin	Numerik	Persentase penduduk di bawah garis kemiskinan
4	Rata – rata Lama Sekolah	Numerik	Rata – rata tahun sekolah penduduk usia 15+
5	Pengeluaran per Kapita	Numerik	Rata – rata pengeluaran setahun (Ribuan Rupiah)
6	Indeks Pembangunan Manusia	Numerik	Indikator kualitas hidup (Kesehatan&Pendidikan)
7	Angka Harapan Hidup	Numerik	Rata – rata perkiraan usia hidup
8	Akses Sanitasi Layak	Numerik	Persentase rumah tangga dengan sanitasi layak
9	Akses Air Minum Layak	Numerik	Persentase rumah tangga dengan air bersih
10	Tingkat Pengangguran	Numerik	Persentase angkatan kerja yang menganggur
11	Tingkat Partisipasi Kerja	Numerik	Persentase penduduk usia kerja yang bekerja
12	PDRB Konstan	Numerik	Nilai tambah bruto ekonomi wilayah
13	Klasifikasi Kemiskinan	Nominal	Label target (Miskin/Tidak Miskin)

Proses pengolahan data dilakukan dengan bantuan perangkat lunak RapidMiner melalui serangkaian tahapan berikut:

1) Akuisisi Data (Read CSV): Menyiapkan kumpulan data berjudul “Klasifikasi Tingkat Kemiskinan di Indonesia” yang mencakup beragam indikator sosial-ekonomi, antara lain tingkat pendidikan, pengeluaran per individu, akses terhadap sanitasi layak, akses air minum yang aman, tingkat pengangguran, serta persentase penduduk yang hidup dalam kemiskinan (P0). Kumpulan data awal disimpan dalam format CSV. proses pemanggilan data

dalam Rapidminer dilakukan oleh operator CSV yang secara otomatis membaca file berformat comma-separated values dan mengubahnya menjadi example set untuk keperluan pemrosesan lebih lanjut.

- 2) Pembersihan Data (*Data Cleaning*): Sebelum dilakukan pemodelan, dataset melalui tahap validasi kaulitas data (*data quality assessment*) untuk memastikan integritasnya. Pemeriksaan dilakukan terhadap statistik deskriptif data untuk mendeteksi kemungkinan adanya nilai yang hilang (*missing values*), noise, atau duplikasi. Berdasarkan hasil pemeriksaan, dataset terkonfirmasi dalam kondisi lengkap dan bersih, sehingga dapat langsung digunakan untuk proses klasifikasi tanpa memerlukan Teknik imputasi atau penyaringan lanjutan. Hal ini memastikan bahwa perhitungan nilai *entropy* dan *information gain* tidak bias akibat data yang cacat.
- 3) Penyesuaian Atribut (*Rename*): melakukan pengubahan nama atribut agar lebih sederhana dan lebih mudah dipahami selama analisis. Tahapan ini dikerjakan menggunakan operator rename. Dalam penelitian ini, atribut “pengeluaran per Kapita Disesuaikan (Ribu/Rupiah/Orang/Tahun)” diubah namanya menjadi “Pengeluaran perkapita”.
- 4) Konversi Tipe Data (Numerical to Real): memastikan bahwa tipe data untuk atribut Pengeluaran perkapita sesuai dengan kebutuhan algoritma yang akan diaplikasikan.
- 5) Generate Attributes: Pembuatan atribut kelas dengan menggunakan operator Generate Attributes di Rapidminer. Atribut kelas status_kemiskinan dibuat berdasarkan nilai kemiskinan (GK) yang ditetapkan oleh BPS. Nilai GK yang sebesar Rp.609.160 per kapita per bulan diubah menjadi satuan Ribu Rupiah per Orang per Tahun, sehingga ambang batas yang diperoleh adalah 7.411. Nilai ini dipergunakan sebagai dasar klasifikasi ke dalam kategori “Miskin” dan “Tidak Miskin”.
- 6) Set Role: menentukan peran atribut melalui operator Set Role. Di tahap ini, atribut status_kemiskinan ditentukan sebagai label, sementara atribut yang lain berperan sebagai atribut prediktor.
- 7) Split Data: Dataset kemudian dipecah menjadi dua bagian menggunakan operator Split data untuk tujuan pelatihan dan pengujian model. Pembagian dilakukan dengan perbandingan 70% untuk data pelatihan dan 30% untuk data pengujian. Langkah ini bertujuan untuk menilai kemampuan model klasifikasi terhadap data yang belum pernah dilatih sebelumnya.
- 8) Decision Tree: memilih atribut terbaik sebagai simpul berdasarkan nilai gain ratio, sehingga dapat membentuk struktur pohon keputusan yang efisien untuk mengklasifikasikan status kemiskinan dalam kategori “Miskin” dan “Tidak Miskin”
- 9) Apply Model: menerapkan model yang telah dilatih pada data pelatihan ke data pengujian, menghasilkan prediksi kelas untuk setiap data yang diuji. Dalam tahap ini, atribut prediksi yang berupa status_kemiskinan (Miskin atau Tidak Miskin) ditambahkan secara otomatis ke dalam hasil dataset.

10) Performance: mengevaluasi model klasifikasi yang telah dibuat. Proses evaluasi digunakan menggunakan Performance (Classification) di Rapidminer. Operator ini digunakan untuk mengukur kinerja model Decision Tree dalam mengklasifikasikan status kemiskinan berdasarkan data pengujian.

III. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan model klasifikasi kemiskinan yang terbentuk melalui dua tahapan strategis, yaitu penentuan label (*labeling*) berdasarkan standar ekonomi nasional dan pembentukan model prediktif menggunakan algoritma C4.5.

A. Penentuan Label Kelas (*Labeling*)

Tahap awal berfokus pada pembentukan variabel target biner untuk memisahkan kategori “Miskin” dan “Tidak Miskin”. Acuan utama yang digunakan adalah Garis Kemiskinan (GK) yang ditetapkan oleh Badan Pusat Statistik (BPS). Berdasarkan standar pengeluaran harian sebesar Rp 20.305 per kapita, nilai tersebut dikonversi ke dalam satuan tahunan (20.305×365 hari) sehingga diperoleh ambang batas sebesar Rp 7.411.325 per kapita per tahun.

Mengingat atribut Pengeluaran per Kapita Disesuaikan dalam dataset menggunakan satuan Ribu Rupiah, maka nilai ambang batas tersebut disesuaikan (dibagi 1.000) menjadi 7.411,325. Proses pelabelan dilakukan secara deterministik menggunakan operator Generate Attributes pada RapidMiner dengan fungsi ekspresi:

If (Pengeluaran_perkapita < 7411.325, "Miskin", "Tidak Miskin")

Secara sistematis, aturan penentuan status kemiskinan (*S*) untuk setiap wilayah (*i*) berdasarkan pengeluaran per kapita (*X*) dirumuskan sebagai berikut:

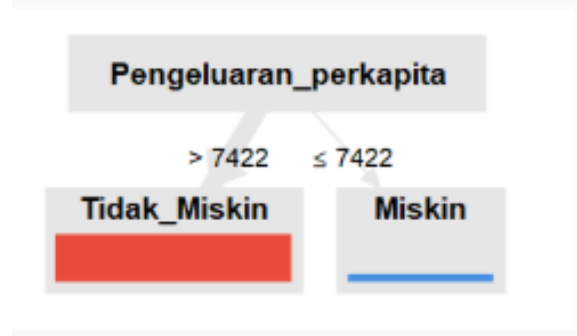
$$S_i = \begin{cases} \text{"Miskin"}, & \text{jika } X_i < 7.411,325 \\ \text{"Tidak Miskin"}, & \text{jika } X_i \geq 7.411,325 \end{cases}$$

Pendekatan ini memastikan bahwa setiap record data memiliki label kelas yang objektif dan valid berdasarkan standar ekonomi nasional sebelum diproses lebih lanjut oleh algoritma C4.5.

B. Analisis Model Pohon Keputusan

Penerapan algoritma C4.5 pada dataset yang telah berlabel menghasilkan struktur pohon keputusan (*Decision Tree*) yang memetakan hubungan antara indikator sosial-ekonomi dengan status kemiskinan. Visualisasi model yang berbentuk dapat dilihat pada Gambar 2.

Berdasarkan Gambar 2, algoritma C4.5 menempatkan variabel Pengeluaran per Kapita sebagai simpul akar (*root node*). Hal ini mengindikasikan bahwa dalam perhitungan *Information Gain*, variabel pengeluaran memiliki nilai tertinggi dibandingkan variabel lain seperti pendidikan atau sanitasi, menjadikannya determinan paling dominan dalam klasifikasi ini.



Gambar 2. Struktur Pohon Keputusan Hasil Algoritma C4.5

Lebih lanjut, pohon keputusan tersebut dapat diterjemahkan menjadi aturan logika “Jika – Maka” (*If-Then-Rules*) yang mendetail sebagai berikut:

1. Aturan 1 (Kategori Miskin):

- Logika: Jika (*IF*) nilai Pengeluaran per Kapitan kurang dari atau sama dengan 7.422 (dalam ribuan rupiah)
- Keputusan: MAKA (*THEN*) wilayah tersebut diklasifikasikan sebagai “Miskin”.
- Analisis: Aturan ini menunjukkan bahwa ambang batas kritis kemiskinan terdeteksi secara otomatis oleh algoritma pada nilai tersebut. Wilayah dengan daya beli di bawah angka ini dianggap tidak memenuhi standar kebutuhan dasar minimum.

2. Aturan 2 (Kategori Tidak Miskin):

- Logika: Jika (*IF*) nilai Pengeluaran per Kapita lebih besar dari 7.422
- Keputusan: MAKA (*THEN*) wilayah tersebut diklasifikasikan sebagai “Tidak Miskin”
- Analisis: Aturan ini menegaskan bahwa variabel pengeluaran adalah prediktor tunggal yang paling dominan. Tidak adanya percabangan (*branching*) lebih lanjut ke atribut lain menunjukkan bahwa variabel ekonomi makro memiliki *Information Gain* yang jauh lebih superior dibandingkan variabel sosial lainnya.

Struktur aturan yang sederhana namun tegas ini membuktikan bahwa Algoritma C4.5 mampu menghasilkan model yang sangat *interpretable*. Bagi pembuat kebijakan, aturan ini memberikan parameter numerik yang jelas untuk menentukan prioritas intervensi pengentasan kemiskinan.

C. Evaluasi Kinerja Model

Evaluasi model dilakukan untuk menilai seberapa baik algoritma C4.5 dapat mengklasifikasikan data uji dengan akurat. Proses evaluasi ini memanfaatkan Confusion Matrix yang diperoleh melalui operator kinerja (Klasifikasi) di Rapidminer. Hasil dari evaluasi ditampilkan pada Gambar 3, yang memperlihatkan perbandingan antara kelas sebenarnya dan kelas yang di prediksi.

	true Miskin	true Tidak_Miskin	class precision
pred. Miskin	41	0	100.00%
pred. Tidak_Miskin	0	319	100.00%
class recall	100.00%	100.00%	

Gambar 3. Confusion Matrix Hasil Evaluasi Algoritma C4.5

Mengacu pada Gambar 3, evaluasi menunjukkan bahwa model tersebut mampu mengklasifikasikan semua data uji dengan presisi. Sebanyak 41 data dengan kelas sebenarnya miskin terprediksi dengan akurat sebagai miskin, sementara 319 data dengan kelas sebenarnya tidak miskin juga teridentifikasi dengan tepat sebagai miskin. Tidak ada kesalahan dalam klasifikasi yang terjadi, baik dalam bentuk false positive maupun false negative.

Untuk mengukur kinerja model klasifikasi, digunakan beberapa metrik evaluasi yang dihitung berdasarkan Confusion Matrix, yaitu akurasi, precision, dan recall. Perhitungan masing-masing metrik dirumuskan sebagai berikut:

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Presisi} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

Keterangan:

1. TP (True Positive) : Data miskin yang diprediksi miskin
2. TN (True Negative) : Data tidak miskin yang diprediksi tidak miskin
3. FP (False Positive) : Data tidak miskin yang diprediksi miskin
4. FN (False Negative) : Data miskin yang diprediksi tidak miskin

Berdasarkan hasil evaluasi metrik pada data yang diuji 30%, model klasifikasi yang menggunakan algoritma C4.5 menunjukkan tingkat akurasi, presisi, dan recall sebesar 100%. Ini mengindikasikan bahwa model tersebut dapat mengklasifikasikan semua data pengujian dengan tingkat ketelitian yang sangat tinggi.

Tingginya kinerja model ini dipengaruhi oleh pemisahan kelas yang jelas berdasarkan nilai pengeluaran per kapita, sehingga pola data di setiap kelas cenderung seragam. Dengan demikian, penerapan metode pembagian data 70% untuk data latih dan 30% untuk data uji berhasil memberikan gambaran objektif mengenai performa model pada data yang tidak pernah dipakai dalam pelatihan sebelumnya.

D. Pembahasan dan Analisis Aturan

Struktur pohon keputusan memperlihatkan pola pemisahan yang tegas (*clear split*). Algoritma berhasil mengidentifikasi ambang batas numerik yang optimal untuk memisahkan kedua kelas secara presisi. Hasil ini sejalan dengan temuan Santoso dkk. [4] yang menyatakan bahwa algoritma C4.5 sangat efektif dalam menangani data sosial-

ekonomi karena kemampuannya menghasilkan aturan (*rules*) yang mudah diinterpretasikan oleh manusia.

Berbeda dengan metode statistik murni yang seringkali menghasilkan model “kotak hitam” (*black box*), model pohon keputusan ini memberikan transparansi logika. Pengambil kebijakan dapat melihat secara langsung bahwa jika pengeluaran per kapita berada di bawah ambang batas tertentu, maka wilayah tersebut dipastikan masuk kategori miskin tanpa perlu melihat atribut lain secara mendalam. Stabilitas model ini menegaskan bahwa pendekatan *Data Mining C4.5* dapat diandalkan sebagai alat bantu pendukung keputusan (*Decision Support System*) bagi pemerintah daerah.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Data Mining C4.5 untuk mengklasifikasikan status kemiskinan tingkat kabupaten / kota di Indonesia berdasarkan indikator sosial - ekonomi. Berdasarkan hasil pengujian dan analisis model, dapat disimpulkan bahwa algoritma C4.5 mampu membentuk struktur pohon keputusan yang efektif dalam memetakan karakteristik kemiskinan secara objektif. Temuan utama dari penelitian ini menunjukkan bahwa variabel Pengeluaran per Kapita merupakan faktor penentu paling dominan (root node) dengan nilai Information Gain tertinggi dibandingkan indikator lainnya.

Secara spesifik, model pohon keputusan yang dihasilkan menetapkan nilai ambang batas (*threshold*) pemisahan yang tegas pada angka Rp 7.411.325 per kapita per tahun. Wilayah dengan rata - rata pengeluaran di bawah angka tersebut secara otomatis diklasifikasikan ke dalam kategori “Miskin”, sedangkan yang berada di atasnya masuk kategori “Tidak Miskin”. Hal ini membuktikan kekuatan algoritma C4.5 dalam menghasilkan aturan (*rules*) yang transparan dan mudah diinterpretasikan oleh pengambil kebijakan. Meskipun demikian, penelitian ini masih terbatas pada penggunaan variabel makro ekonomi. Oleh karena itu, penelitian selanjutnya disarankan untuk mengintegrasikan variabel multidimensi yang lebih kompleks, seperti akses infrastruktur digital atau indeks kesehatan lingkungan, guna meningkatkan granularitas dan akurasi prediksi model di masa mendatang.

REFERENSI

- [1] P. : Amsset, I. Batusangkar, I. B. Press, N. Putu, N. Hendayanti, and M. Nurhidayati, “SAINSTEK : JURNAL SAINS DAN TEKNOLOGI Regresi Logistik Biner dalam Penentuan Ketepatan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi-Provinsi di Indonesia Article History,” 2020.
- [2] M. Martin and L. Nilawati, “Comparison of C4.5 and Naïve Bayes Algorithms for Assessment of Public Complaints Services,” *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 5, no. 1, pp. 101–111, Jul. 2021, doi: 10.31289/jite.v5i1.5292.
- [3] P. Ramadani, R. Fadillah, Q. Adawiyah, and B. Restu Al Ghazali, “JURNAL MEDIA INFORMATIKA [JUMIN]

- Perbandingan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kelayakan Program Keluarga Harapan,” 2019.
- [4] G. A. D. Santosa, B. Santosa, L. P. Sumirat, and L. P. Sumirat, “Sistem Pendukung Keputusan Penentuan Penerima BLT-DD Menggunakan Algoritma C4.5,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 1177–1188, Jul. 2024, doi: 10.57152/malcom.v4i3.1579.
- [5] R. Girsang, E. Fahmi Ginting, M. Hutasuhut, S. Informasi, and S. Triguna Dharma, “Penerapan Algoritma C4.5 Pada Penentuan Penerima Program Bantuan Pemerintah Daerah,” 2022, [Online]. Available: <https://ojs.trigunadharna.ac.id/index.php/jsi>
- [6] S. Lestari and S. Agustiansyah, “Implementasi Data Mining Clustering Data Penduduk Miskin Menggunakan Metode Algoritma C4.5 Untuk Merekomendasikan Bantuan Sosial Pada RT. 05/01 Kelurahan Jati Mekar Kecamatan Jati Asih Kota Bekasi,” *Jurnal Teknik Elektro dan Komputasi (ELKOM)*, vol. 5, no. 1, pp. 95–104, Mar. 2023, doi: 10.32528/elkom.v5i1.8339.
- [7] S. Saputra, N. Suarna, I. Ali, and D. Solihudin, “Journal of Artificial Intelligence and Engineering Applications Improving the Classification Model of Smart Indonesia Program Recipients in Koorwilbidikcam Sumber Using the C4.5 Algorithm,” 2025. [Online]. Available: <https://ioinformatic.org/>
- [8] S. F. Damanik, A. Wanto, and I. Gunawan, “Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok.” [Online]. Available: <https://ejournal.catuspata.com/index.php/jkdn/index>
- [9] Y. Mardi, “Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Yuli Mardi,” 2017.
- [10] I. Clement, O. Putra, and H. Wijaya, “PENERAPAN ALGORITMA C4.5 DALAM DATA MINING UNTUK IDENTIFIKASI FAKTOR RISIKO STROKE PADA DATASET MEDIS.”