



Analisis Perbandingan Kinerja Decision Tree C4.5 dan ID3 dalam Klasifikasi Kemiskinan Masyarakat

Rahma Garwati, Muhamat Maariful Hudha, Abd. Charis Fauzan*

Fakultas Ilmu Eksakta, Ilmu Komputer, Universitas Nahdlatul Ulama Blitar, Blitar

Jl. Masjid No.22, Kauman, Kec. Kepanjenkidul, Kota Blitar, Jawa Timur, Indonesia

Email: ¹rahmagarwati2002@gmail.com, ²hudha.maariful@gmail.com, ^{3,*}abdcharis@unublitar.ac.id

Email Penulis Korespondensi: abdcharis@unublitar.ac.id

Submitted: 09/07/2025; Accepted: 22/07/2025; Published: 25/07/2025

Abstrak—Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma pohon keputusan, yaitu ID3 dan C4.5, dalam klasifikasi status kemiskinan masyarakat berdasarkan data survei rumah tangga tahun 2024. Metode yang digunakan adalah pendekatan kuantitatif eksperimental dengan tahapan seleksi atribut, pembersihan data, transformasi, pelatihan model, dan evaluasi performa. Algoritma ID3 diimplementasikan secara manual berdasarkan prinsip entropy dan information gain, sedangkan C4.5 diimplementasikan menggunakan pustaka scikit-learn dengan pengaturan parameter seperti batas kedalaman pohon dan jumlah minimum sampel untuk percabangan, guna menghindari overfitting. Evaluasi dilakukan dengan membagi data menjadi 70% pelatihan dan 30% pengujian. Kinerja kedua model dianalisis menggunakan metrik akurasi, precision, recall, F1-score, dan confusion matrix. Hasil penelitian menunjukkan bahwa kedua algoritma mampu mengklasifikasikan data kemiskinan dengan baik, namun algoritma ID3 menunjukkan performa yang lebih tinggi pada seluruh metrik evaluasi. Meski demikian, C4.5 tetap menunjukkan kestabilan model dan struktur pohon yang lebih sederhana. Temuan ini menunjukkan bahwa pilihan algoritma terbaik sangat dipengaruhi oleh struktur data dan kebutuhan analisis. Penelitian ini diharapkan menjadi kontribusi penting dalam penerapan data mining untuk pemetaan kemiskinan dan pengambilan kebijakan sosial berbasis data secara lebih akurat dan efisien.

Kata Kunci: Pohon Keputusan; Algoritma ID3; Algoritma C4.5; Klasifikasi Kemiskinan; Data Mining

Abstract—This study aims to compare the performance of two decision tree algorithms, namely ID3 and C4.5, in classifying poverty status using household survey data from 2024. A quantitative experimental approach was employed, involving stages such as feature selection, data cleaning, transformation, model training, and performance evaluation. The ID3 algorithm was implemented manually based on entropy and information gain, while the C4.5 algorithm was implemented using the scikit-learn library with parameter adjustments, including limiting tree depth and minimum samples per split to reduce the risk of overfitting. The dataset was divided into 70% training and 30% testing data. Performance evaluation used metrics including accuracy, precision, recall, F1-score, and confusion matrix analysis. Results indicate that both algorithms can effectively classify poverty status, but the ID3 algorithm outperforms C4.5 in all evaluation metrics. However, C4.5 provides better model stability and a more simplified tree structure. These findings suggest that the choice of algorithm should be tailored to the dataset characteristics and analytical needs. This study contributes to the application of data mining techniques in supporting poverty analysis and data-driven policy decisions, ensuring a more accurate and targeted distribution of social assistance.

Keywords: Decision Tree; ID3 Algorithm; C4.5 Algorithm; Poverty Classification; Data Mining

1. PENDAHULUAN

Tantangan kemiskinan masih menjadi perhatian utama yang serius karena pendapatan, pendidikan, kesehatan, akses layanan, kondisi geografis, dan lingkungan merupakan faktor-faktor yang saling terkait dan secara bersama-sama memengaruhi tingkat kemiskinan secara menyeluruh di suatu daerah [1][2]. Berdasarkan informasi dari Badan Pusat Statistik (BPS), pada bulan Maret 2023 jumlah masyarakat yang tergolong miskin di Indonesia tercatat sebanyak 25,90 juta orang atau sekitar 9,36% dari total penduduk[3]. Ini menunjukkan bahwa isu kesejahteraan masyarakat masih menjadi persoalan nyata yang memerlukan perhatian serius dari berbagai pihak. Ketimpangan ekonomi ini bukan hanya soal kekurangan pendapatan, tetapi merupakan kondisi multidimensional yang mencerminkan keterbatasan akses terhadap kebutuhan dasar hidup, seperti makanan, tempat tinggal layak, pendidikan yang memadai, layanan kesehatan, dan peluang kerja.

Selain itu, faktor-faktor lain seperti kondisi rumah, pengeluaran per individu, status pekerjaan, jumlah tanggungan keluarga, serta berbagai aspek sosial ekonomi lainnya juga turut mempengaruhi tingkat kemiskinan. Seluruh faktor tersebut memiliki keterkaitan satu sama lain dan apabila tidak ditangani dengan baik, dapat memperparah situasi kemiskinan yang telah terjadi. Penanganan kemiskinan pun memerlukan pendekatan yang komprehensif dan berkelanjutan, termasuk upaya untuk memahami kondisi masyarakat secara menyeluruh dan membangun sistem yang mampu mendeteksi serta memetakan kondisi tersebut secara akurat. Hal ini menjadi penting mengingat kebijakan penanganan kemiskinan sering kali disusun berdasarkan data yang tersedia, sehingga kualitas data yang akurat dan representatif menjadi landasan bagi pengambilan keputusan yang efektif [4].

Namun, selain kompleksitas faktor penyebab tersebut, masalah klasifikasi kemiskinan yang kurang akurat juga menjadi kendala utama yang sering kali kurang diperhatikan. Kesalahan dalam mengidentifikasi kelompok miskin menyebabkan ketidaktepatan dalam penyaluran bantuan sosial, sehingga sebagian masyarakat yang benar-benar membutuhkan bantuan justru terlewatkan, sementara kelompok yang tidak miskin malah menerima bantuan tersebut[5]. Hal ini menimbulkan ketidakadilan dalam masyarakat serta berisiko menurunkan kepercayaan publik terhadap pemerintah maupun lembaga penyalur bantuan. Ketidaktepatan tersebut sering kali disebabkan oleh



sistem klasifikasi manual yang masih bersifat subjektif, tidak terstandarisasi, atau kurang mempertimbangkan banyaknya variabel relevan.

Kondisi ini memperburuk ketimpangan sosial dan menghambat upaya pengentasan kemiskinan secara efektif. Oleh karena itu, diperlukan perbaikan dalam metode atau sistem klasifikasi agar distribusi bantuan dapat tepat sasaran dan memberikan dampak maksimal bagi masyarakat yang membutuhkan. Salah satu pendekatan yang dapat digunakan adalah pemanfaatan teknologi dalam bidang analisis data, khususnya melalui metode data mining.

Untuk mengatasi permasalahan tersebut, penggunaan algoritma dalam data mining dapat dimanfaatkan untuk pengolahan data kemiskinan secara lebih efektif. Data mining memungkinkan analisis terhadap kumpulan data besar secara sistematis dan terstruktur, untuk menemukan pola atau informasi tersembunyi yang sebelumnya sulit diidentifikasi secara manual. Dalam konteks ini, algoritma Decision Tree C4.5 dipilih karena dikenal mampu menghasilkan hasil klasifikasi dengan akurasi yang tinggi[6]. Metode ini bekerja dengan membangun model pohon keputusan berdasarkan atribut yang relevan, sehingga memungkinkan identifikasi pola dari data yang kompleks[7]. Selain itu, algoritma ini mampu mengelompokkan data ke dalam kategori spesifik, seperti tingkat pengeluaran, kepemilikan aset, atau kondisi hunian, yang sangat relevan untuk analisis kemiskinan.

Dengan hasil analisis yang terstruktur, pengambil kebijakan dapat menentukan prioritas dan strategi yang lebih efektif dalam program pengentasan kemiskinan. Algoritma C4.5 juga mampu menangani atribut numerik maupun kategorikal, serta mengatasi data dengan nilai kosong (missing values)[8], menjadikannya salah satu metode yang adaptif terhadap data lapangan yang sering kali tidak sempurna.

Selain itu, penelitian ini juga memanfaatkan algoritma klasifikasi data mining ID3 yang dikenal luas dan memiliki tingkat akurasi yang tinggi. Algoritma ID3 mampu membangun pohon keputusan secara sederhana dan mudah dipahami, sehingga sangat membantu dalam menganalisis data sekaligus membuat prediksi dengan lebih efektif[9]. ID3 menggunakan ukuran gain informasi untuk memilih atribut terbaik dalam setiap percabangan pohon keputusan, sehingga mampu menghasilkan struktur pohon yang efisien dan cukup optimal. Algoritma ini akan digunakan sebagai pembanding terhadap algoritma Decision Tree C4.5. Diharapkan, perbandingan antara kedua algoritma tersebut dapat memberikan gambaran yang jelas mengenai kelebihan dan kekurangan masing-masing dalam mengolah data dan menghasilkan prediksi yang akurat.

Beberapa penelitian sebelumnya telah menerapkan algoritma ID3 untuk berbagai kebutuhan klasifikasi. Salah satunya menganalisis penggunaan algoritma ID3 dan C4.5 dalam mengukur tingkat kepuasan konsumen pada sebuah perusahaan, di mana data diperoleh dari kuesioner pelanggan untuk mengevaluasi efektivitas pelayanan[10]. Penelitian lainnya membandingkan kedua algoritma tersebut dalam memprediksi ketepatan waktu kelulusan mahasiswa berdasarkan data akademik, sebagai upaya menilai keberhasilan program pendidikan[11]. Selain itu, algoritma ID3 dan C4.5 juga digunakan untuk menentukan kelayakan penerima Bantuan Pangan Non Tunai (BPNT) dengan memanfaatkan data sosial ekonomi masyarakat[12]. Studi-studi ini menunjukkan bahwa kedua algoritma memiliki potensi dalam pengambilan keputusan berbasis data, tergantung pada konteks dan kebutuhan masing-masing.

Selain penggunaan ganda ID3 dan C4.5, beberapa penelitian juga membuktikan efektivitas masing-masing algoritma dalam konteks bantuan sosial secara terpisah. Implementasi ID3 dilakukan untuk menentukan jumlah dana bantuan perbaikan rumah di Bappeda dengan mempertimbangkan berbagai kriteria rumah tangga, menghasilkan perangkat yang mampu memberikan keputusan otomatis[13]. Sementara itu, C4.5 diterapkan dalam penentuan penerima bantuan pemerintah daerah, menghasilkan sistem yang meningkatkan akurasi distribusi dan membantu perangkat desa dalam pengambilan keputusan[14]. Penelitian lain di Desa Keramas, Gianyar, menggunakan algoritma C4.5 untuk mengklasifikasikan kelayakan penerima Bantuan Sosial Tunai (BST), dan hasilnya menunjukkan bahwa C4.5 memiliki akurasi lebih tinggi dibandingkan K-NN dan Naïve Bayes, memperkuat posisinya sebagai algoritma yang andal dalam klasifikasi sosial ekonomi[15].

Meskipun beberapa penelitian terdahulu telah membandingkan kinerja algoritma ID3 dan C4.5, fokus kajiannya umumnya berada pada sektor pendidikan, layanan konsumen, atau bantuan sosial secara umum. Sementara itu, kajian yang secara spesifik membahas kemiskinan sebagai objek utama klasifikasi masih sangat terbatas. Bahkan, studi yang mengangkat isu kemiskinan cenderung hanya menggunakan salah satu dari kedua algoritma tersebut, tanpa melakukan perbandingan menyeluruh terhadap performa masing-masing dalam konteks identifikasi masyarakat miskin. Oleh karena itu, diperlukan penelitian yang secara langsung membandingkan algoritma ID3 dan C4.5 dalam klasifikasi kemiskinan agar diperoleh pemahaman yang lebih mendalam mengenai efektivitas kedua metode dalam menangani isu sosial yang kompleks ini.

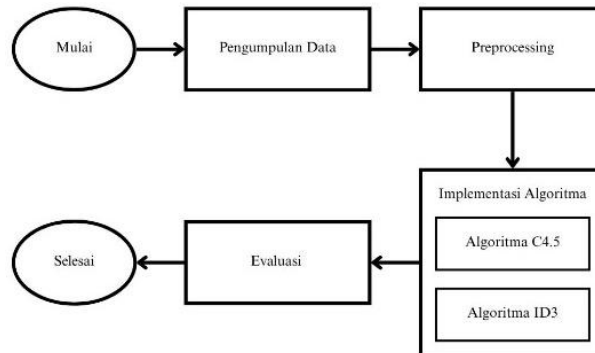
2. METODOLOGI PENELITIAN

2.1 Objek Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental yang bertujuan untuk menganalisis dan membandingkan kinerja dua algoritma pohon keputusan, yaitu ID3 dan C4.5, dalam mengklasifikasikan status kemiskinan masyarakat. Pendekatan ini sejalan dengan studi [16] yang menerapkan decision tree untuk analisis data sosio-ekonomi pedesaan.

2.1. Alur Penelitian

Proses pada metode ini terdiri dari serangkaian tahapan yang terstruktur dan saling berkaitan, dimulai dari pengumpulan data hingga evaluasi model, yang secara keseluruhan dapat dilihat secara lebih jelas dan sistematis pada Gambar 1 sebagai representasi visual dari alur pelaksanaan penelitian.



Gambar 1. Alur Penelitian

2.2.1 Pengumpulan Data

Sesuai dengan yang ditunjukkan pada Gambar 1, penelitian ini menggunakan data sekunder yang diperoleh dari hasil survei rumah tangga tahun 2024 yang dilaksanakan oleh instansi terkait. Data ini mencakup berbagai variabel sosial ekonomi rumah tangga, seperti kepemilikan aset, kondisi tempat tinggal, sumber air, sumber penerangan, dan pengeluaran individu per bulan yang menjadi dasar utama dalam proses klasifikasi status kemiskinan masyarakat melalui metode pohon keputusan.

2.2.2 Preprocessing

Sebelum pemodelan, dilakukan tahap seleksi data, yaitu memilih data dan atribut yang relevan untuk analisis klasifikasi kemiskinan. Selanjutnya, dilakukan pembersihan data untuk menangani nilai kosong dan memastikan kualitas data agar siap digunakan dalam proses pemodelan[17]. Tahap berikutnya adalah transformasi data dengan mengubah data kategorikal menjadi format numerik menggunakan teknik encoding [18][19]. Setelah itu, data dibagi menjadi data latih dan data uji dengan proporsi 70% dan 30%[20]. Proses ini bertujuan untuk membangun dan menguji model klasifikasi menggunakan algoritma Decision Tree C4.5 dan ID3 secara efektif dan akurat.

2.2.3 Implementasi Algoritma

Proses data mining dalam penelitian ini menggunakan dua algoritma, yaitu ID3 dan C4.5, yang pada dasarnya merupakan algoritma pohon keputusan untuk membentuk model klasifikasi[21]. Kedua algoritma ini dipilih karena sering digunakan dalam klasifikasi data sosial ekonomi. Algoritma ID3 memilih atribut terbaik berdasarkan perhitungan information gain yang dihitung dari entropy[22]. Entropy merupakan ukuran ketidakpastian dalam suatu himpunan data dan dirumuskan sebagai:

$$\text{Entropy}(S) = - \sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

Dimana p_i adalah proporsi data pada kelas ke- i dari total data. Semakin homogen data, maka nilai entropy akan semakin mendekati nol. Setelah itu, dilakukan perhitungan information gain dari setiap atribut yang tersedia, yaitu selisih antara entropy awal (sebelum pemisahan) dan rata-rata entropy setelah data dibagi berdasarkan nilai atribut tertentu. Rumus information gain yang digunakan adalah:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|s_v|}{|s|} \cdot \text{Entropy}(s_v) \quad (2)$$

Dengan s_v adalah subset data untuk nilai v pada atribut A . Atribut yang memiliki nilai gain tertinggi akan dipilih sebagai simpul utama (root node) atau menjadi cabang pohon keputusan pada tahap selanjutnya. Proses ini dilakukan secara rekursif, setelah atribut terbaik dipilih, data dibagi ke dalam subset sesuai dengan nilai dari atribut tersebut, dan algoritma mengulang proses serupa pada masing-masing subset. Rekursi akan berhenti ketika salah satu dari tiga kondisi berikut tercapai: pertama, seluruh data dalam subset memiliki label kelas yang sama (homogen); kedua, tidak ada lagi atribut yang tersedia untuk digunakan sebagai pembagi; dan ketiga, subset data kosong.

Sedangkan algoritma C4.5 menerapkan mekanisme pruning untuk menyederhanakan pohon keputusan, sehingga mampu mengurangi risiko overfitting dan menghasilkan model klasifikasi yang lebih akurat serta mudah dipahami oleh pengambil kebijakan. [23]. Implementasi algoritma ID3 dilakukan secara manual sesuai prinsip dasar algoritma, sementara C4.5 diimplementasikan menggunakan pustaka scikit-learn dengan pengaturan parameter yang disesuaikan untuk meningkatkan performa model.



2.2.4 Evaluasi

Evaluasi kinerja kedua algoritma pohon keputusan, ID3 dan C4.5, dilakukan dengan menggunakan metrik evaluasi standar, yaitu akurasi, precision, recall, dan F1-score. Pengukuran ini bertujuan untuk menilai kemampuan masing-masing model dalam mengklasifikasikan status kemiskinan masyarakat secara tepat [19].

Sebelum menghitung metrik evaluasi tersebut, terlebih dahulu digunakan confusion matrix untuk memberikan gambaran lebih rinci mengenai prediksi yang benar dan salah pada setiap kelas[24].

Tabel 1. Confusion Matrix

Table with 3 columns: Actual (Aktual Positive, Aktual Negative), Predicted Positive (True Positive(TP), False Negative (FN)), and Predicted Negative (False Positive(FP), True Negative(TN)).

Setelah memperoleh nilai dari confusion matrix, dilakukan perhitungan menggunakan rumus-rumus berikut:

Accuracy = (TP+TN) / (TP+FP+FN+TN) (3)

Precision = TP / (TP+FP) (4)

Recall = TP / (TP+FN) (5)

F1 - score = 2 * (Precision * Recall) / (Precision + Recall) (6)

Proses evaluasi ini menggunakan data uji yang telah dipisahkan sebelumnya untuk memastikan validitas hasil. Pendekatan evaluasi ini sesuai dengan praktik umum dalam penelitian klasifikasi dan didukung oleh berbagai studi yang menunjukkan bahwa metrik-metrik tersebut efektif dalam mengukur performa algoritma decision tree pada data sosial ekonomi.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi dataset

Dataset yang digunakan dalam penelitian ini merupakan data sekunder hasil survei rumah tangga tahun 2024 yang berjumlah 340 entri. Awalnya, dataset terdiri dari 31 atribut, namun setelah dilakukan seleksi, hanya 15 atribut yang digunakan.

Tabel 2. Dataset Asli

Table with 9 columns: NO KK, Nik, Nama Kepala, Kategori Kesejahteraan, Sumber Penerangan, Sumber Air Minum, Fasilitas Bab, and Pengeluaran Individu Perbulan. It shows data for two individuals with different living conditions.

NO KK	Nik	Nama Kepala	Kategori Kesejahteraan	...	Sumber Penerangan	Sumber Air Minum	Fasilitas Bab	Pengeluaran Individu Perbulan
3505***	3505***	Nama 142	SANGAT MISKIN	...	LISTRIK iPLN: 900 WATT	MATA AIR	TIDAK ADA	Rp 250.001,- s.d 500.000i,-
3505***	3505***	Nama 143	MISKIN	...	LISTRIK PLN: 450 WATT ATAU TANPA METERAN	SUMUR	SENDIRI	Rp 1.000.001,- s.d 1.2i50.000,-
3505***	3505***	Nama 144	MISKIN	...	LISTRIK PLN: 450 WATT ATAU TANPA METERAN	SUMUR	SENDIRI	Rpi 501.001,- s.d 750.000,-
...
3505***	3505***	Nama 339	MISKIN	...	LISTRIK PLN: 450 WATT ATAU TANPA METERAN	SUMUR	SENDIRI	Rp 501.001,- s.d 750.000,-

Dataset mentah yang ditampilkan pada Tabel 2 selanjutnya melalui tahap seleksi, preprocessing, dan transformasi sebelum digunakan dalam proses pelatihan dan pengujian model. Proses preprocessing mencakup pengisian nilai kosong dengan modus, encoding data kategorikal, serta normalisasi pada data numerik seperti pada Tabel 3.

Tabel 3. Dataset Preprocessing

Nama Kepala	Label	Lantai	Atap	Bangunan	...	Sumber Penerangan	Sumber Air Minum	Fasilitas Bab
Nama 1	0	0	0	1	...	3	0	1
...
Nama 142	1	0	0	1	...	3	3	2
Nama 143	0	0	0	1	...	2	4	1
Nama 144	0	0	0	1	...	2	4	1
...
Nama 339	0	0	0	1	...	2	4	1

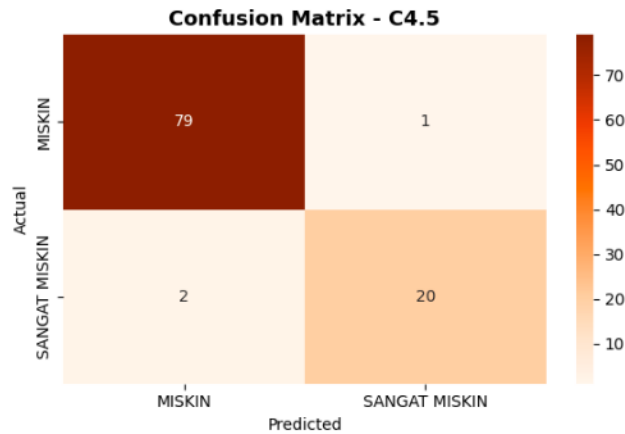
Dataset yang tertera pada Tabel 3 kemudian dibagi dengan proporsi 70% sebagai data latih dan 30% sebagai data uji untuk memastikan model dapat belajar secara optimal sekaligus diuji kemampuannya dalam menggeneralisasi data baru .

3.2 Hasil Evaluasi Model

Sebelum masuk ke dalam pembahasan mendalam mengenai perbandingan performa antara algoritma ID3 dan C4.5, langkah awal yang dilakukan adalah mengevaluasi hasil klasifikasi dari masing-masing model secara terpisah. Evaluasi ini dilakukan berdasarkan hasil confusion matrix yang dihasilkan oleh setiap model, serta penghitungan sejumlah metrik evaluasi penting lainnya seperti akurasi, precision, recall, dan f1-score. Metrik-metrik ini digunakan untuk memberikan gambaran menyeluruh mengenai seberapa baik model mampu mengenali pola dalam data dan mengklasifikasikan entri secara tepat. Pendekatan ini penting untuk memastikan bahwa penilaian performa dilakukan secara objektif dan terukur, sebelum kesimpulan akhir mengenai keunggulan salah satu algoritma ditetapkan.

3.3.1. Hasil Evaluasi Algoritma C4.5

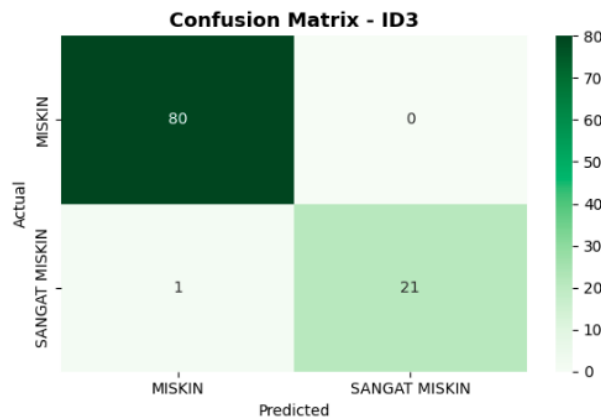
Gambar 2 menunjukkan confusion matrix hasil klasifikasi menggunakan algoritma C4.5 terhadap dua kategori, yaitu "Miskin" dan "Sangat Miskin". Berdasarkan gambar tersebut, model berhasil mengklasifikasikan 79 data kategori miskin dengan benar, sedangkan 1 data kategori miskin salah diklasifikasikan sebagai sangat miskin. Untuk kategori sangat miskin, terdapat 20 data yang berhasil diklasifikasikan dengan tepat, sementara 2 data lainnya keliru diklasifikasikan sebagai miskin. Dari total 102 data yang diuji, sebanyak 99 data berhasil diklasifikasikan dengan benar dan hanya 3 data yang salah, sehingga menghasilkan tingkat akurasi sebesar 97,06%. Hasil ini menunjukkan bahwa algoritma C4.5 memiliki performa klasifikasi yang sangat baik dalam mengidentifikasi kondisi sosial ekonomi berdasarkan data yang digunakan..



Gambar 2. Confusion Matrix - C4.5

3.3.2. Hasil Evaluasi Algoritma ID3

Gambar 3 memperlihatkan confusion matrix hasil klasifikasi yang dilakukan oleh algoritma ID3 terhadap dua kelas, yaitu "Miskin" dan "Sangat Miskin". Dari matrix tersebut dapat dilihat bahwa model ID3 berhasil mengklasifikasikan seluruh 80 data yang benar-benar berstatus miskin dengan akurat ke dalam kelas yang sesuai, tanpa adanya kesalahan klasifikasi ke kategori sangat miskin. Sementara itu, dari total data yang tergolong sangat miskin, sebanyak 21 data berhasil diklasifikasikan dengan tepat, sedangkan hanya 1 data yang salah diklasifikasikan sebagai miskin. Dengan demikian, total data yang diklasifikasikan secara benar oleh model adalah 80 untuk kategori miskin dan 21 untuk kategori sangat miskin, sehingga keseluruhan jumlah klasifikasi yang tepat mencapai 101 dari total 102 data. Kesalahan klasifikasi hanya terjadi sebanyak 1 kasus saja. Berdasarkan hasil tersebut, model ID3 menunjukkan performa klasifikasi yang sangat tinggi dengan tingkat akurasi sebesar 99,02%. Ini menunjukkan bahwa ID3 mampu melakukan pemetaan terhadap kondisi sosial ekonomi dalam dataset secara sangat efektif, bahkan lebih baik daripada algoritma C4.5 pada percobaan ini.



Gambar 3. Confusion Matrix - ID3

3.3.3. Tabel Perbandingan Metrik Evaluasi C4.5 dan ID3

Perbandingan performa kedua algoritma ditampilkan pada Tabel 4. Algoritma ID3 menunjukkan keunggulan pada seluruh metrik evaluasi, termasuk akurasi, precision, recall, dan f1-score, dibandingkan C4.5.

Tabel 4. Evaluasi

Algoritma	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ID3	99.019608	99.031711	99.019608	99.011253
C4.5	97.058824	97.036345	97.058824	97.033758

Tabel 4 menunjukkan hasil evaluasi kinerja dari dua algoritma pohon keputusan, yaitu ID3 dan C4.5, dalam melakukan klasifikasi terhadap status kemiskinan masyarakat. Evaluasi dilakukan menggunakan empat metrik utama, yaitu Accuracy, Precision, Recall, dan F1-Score, yang masing-masing dinyatakan dalam bentuk persentase (%).

Pada algoritma ID3, nilai Accuracy mencapai 99.02%, menunjukkan bahwa model mampu mengklasifikasikan data dengan benar pada hampir seluruh sampel dalam data pengujian. Nilai Precision sebesar 99.03% mengindikasikan bahwa dari seluruh prediksi positif yang dilakukan oleh model ID3, hampir semuanya



benar. Nilai Recall juga sangat tinggi, yaitu 99.02%, yang berarti model berhasil menemukan sebagian besar data yang benar-benar termasuk dalam kelas positif. Kombinasi antara Precision dan Recall yang tinggi ini menghasilkan nilai F1-Score sebesar 99.01%, mencerminkan kinerja yang sangat seimbang dan andal dari algoritma ID3.

Sementara itu, algoritma C4.5 juga menunjukkan performa yang sangat baik, meskipun sedikit lebih rendah dibandingkan ID3. Nilai Accuracy yang dicapai adalah 97.06%, dengan Precision sebesar 97.04%, Recall sebesar 97.06%, dan F1-Score sebesar 97.03%. Hal ini menunjukkan bahwa C4.5 juga memiliki kemampuan klasifikasi yang tinggi, tetapi masih sedikit tertinggal dibandingkan ID3 dalam konteks dataset yang digunakan dalam penelitian ini.

3.3 Analisis Perbandingan Algoritma

Berdasarkan hasil evaluasi, kedua algoritma pohon keputusan menunjukkan performa yang baik dalam klasifikasi status kemiskinan. Namun, ID3 tampil lebih unggul dalam seluruh metrik evaluasi, dengan akurasi 99,02% dibandingkan 97,06% pada C4.5. Keunggulan ID3 didukung oleh pendekatan pemilihan atribut berdasarkan information gain yang sederhana namun efektif. Struktur pohon yang dihasilkan cenderung lebih sesuai dengan karakteristik data yang telah dipreproses, sehingga menghasilkan pemisahan kelas yang lebih tajam. Di sisi lain, meskipun C4.5 memiliki fitur seperti kemampuan menangani atribut numerik dan teknik pruning untuk menghindari overfitting, hasilnya dalam konteks dataset ini tetap berada di bawah ID3. Hal ini menunjukkan bahwa efektivitas algoritma sangat bergantung pada kualitas dan struktur data yang digunakan. Preprocessing yang baik, seperti pengisian nilai kosong, encoding data kategorikal, dan pembagian data secara proporsional, terbukti memberikan kontribusi besar terhadap akurasi model, terutama pada ID3. Data yang bersih dan konsisten memungkinkan algoritma mengenali pola dengan lebih mudah dan menghasilkan prediksi yang akurat. Dalam konteks aplikasi nyata, hasil klasifikasi ini sangat relevan bagi lembaga sosial atau pemerintah dalam menyalurkan bantuan kepada kelompok miskin secara tepat sasaran. Dengan tingkat akurasi yang tinggi, model ID3 mampu meminimalkan kesalahan dalam pengambilan keputusan dan memastikan bahwa bantuan diterima oleh pihak yang benar-benar membutuhkan. Penelitian ini membuktikan bahwa ID3, meskipun secara teknis lebih sederhana dari C4.5, justru lebih optimal untuk dataset ini, sehingga layak dipertimbangkan dalam sistem klasifikasi sosial ekonomi berbasis data. Secara keseluruhan, ID3 tidak hanya unggul dalam performa evaluasi, tetapi juga memberikan nilai praktis dalam mendukung kebijakan distribusi bantuan berbasis data yang lebih adil dan akurat.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan secara komprehensif, dapat disimpulkan bahwa kedua algoritma klasifikasi, yaitu ID3 dan C4.5, memiliki kapabilitas yang mumpuni dalam membentuk model klasifikasi yang efektif untuk mengidentifikasi status kemiskinan masyarakat berdasarkan data survei rumah tangga tahun 2024. Meskipun keduanya berbasis pada pohon keputusan, ID3 menggunakan pendekatan pemilihan atribut berbasis perhitungan information gain secara manual, sedangkan C4.5 mengandalkan pustaka scikit-learn dengan parameter yang dioptimalkan untuk menghindari overfitting. Hasil evaluasi menunjukkan bahwa ID3 unggul dalam semua metrik, dengan akurasi mencapai 99,02%, sedangkan C4.5 tetap menunjukkan performa baik dengan akurasi 97,06% serta menghasilkan struktur pohon yang lebih ringkas dan mudah diinterpretasikan. Perbedaan hasil ini menegaskan bahwa efektivitas suatu algoritma sangat bergantung pada karakteristik data, kualitas preprocessing, dan kompleksitas model yang dibentuk. Namun, penelitian ini memiliki keterbatasan, yaitu hanya menggunakan satu jenis dataset dalam skala lokal, sehingga hasil dan kesimpulannya belum tentu dapat digeneralisasi ke populasi atau wilayah lain dengan karakteristik yang berbeda. Selain itu, pendekatan evaluasi masih terbatas pada metrik kuantitatif tanpa mempertimbangkan aspek interpretabilitas model secara mendalam dalam konteks kebijakan publik. Oleh karena itu, penelitian lanjutan disarankan untuk menguji algoritma ini pada beragam jenis data dan mempertimbangkan integrasi dengan pendekatan kualitatif guna memperkuat validitas hasilnya dalam konteks penerapan nyata.

REFERENCES

- [1] E. Nurliana, B. Irawan, and A. Bahtiar, "Implementasi Data Mining Algoritma K-Means Untuk Klasifikasi Penduduk Miskin Berdasarkan Tingkat Kemiskinan Di Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 1, pp. 1116–1122, 2024, doi: 10.36040/jati.v8i1.8883.*
- [2] Sundari, D. S. Ayuni, and R. S. Prahara, "Analisis Kondisi Sosial Ekonomi Dan Tingkat Pendidikan Masyarakat Desa Talok, Kecamatan Dlanggu, Kabupaten Mojokerto," *J. Agama, Sos. dan Budaya, vol. 6, no. 4, pp. 773–788, 2023.*
- [3] <https://www.bps.go.id>, "Profil Kemiskinan di Indonesia Maret 2023," <https://www.bps.go.id>. Accessed: May 30, 2025. [Online]. Available: <https://www.bps.go.id/id/pressrelease/2023/07/17/2016/profil-kemiskinan-di-indonesia-maret-2023.html>
- [4] W. H. Riyanto, *Model Kelembagaan Daerah dalam Penanganan Kemiskinan*. UMMPress, 2025. [Online]. Available: <https://books.google.co.id/books?id=U5JdEQAAQBAJ>
- [5] Gracenda Febina Br Purba, Dicky M.C. Sinulingga, Josua Togatorop, and Lokot Muda Harahap, "Peran Program Bantuan Sosial dalam Pengentasan Kemiskinan : Evaluasi Dari Berbagai Penelitian," *J. Publ. Ilmu Manaj., vol. 4, no. 1, pp. 108–*



- 117, 2025, doi: 10.55606/jupiman.v4i1.4956.
- [6] J. Bramanda, “Klasifikasi Masyarakat Penerima Bantuan Sosial dari Pemerintah dengan Metode Algoritma C4.5,” *J. Komput. Antart.*, vol. 3, no. 1, pp. 34–41, 2025, doi: 10.70052/jka.v3i1.234.
- [7] T. Novianti, S. A. Mandati, and E. K. Andana, “Peningkatan Evaluasi Risiko Kredit Menggunakan Decision Tree C 4.5,” *J. Manuf. Ind. Eng. Technol.*, vol. 2, no. 2, pp. 1–9, 2023, doi: 10.30651/mine-tech.v2i2.21749.
- [8] A. Fakhri, M. A. Hamzami, M. R. Hadiano, and N. I. S. Alifah, “Perbandingan Akurasi Algoritma C4.5 dan K-NN Untuk Prediksi Kelulusan Mahasiswa Penerima Beasiswa,” *J. Komput. Antart.*, vol. 3, no. 1, pp. 18–25, 2025, doi: 10.70052/jka.v3i1.623.
- [9] F. Ferdina, N. Satyahadewi, and D. Kusnandar, “Penerapan Algoritma Iterative Dichotomiser 3 (Id3) Dalam Klasifikasi Faktor Risiko Penyakit Diabetes Melitus,” *Var. J. Stat. Its Appl.*, vol. 5, no. 2, pp. 139–146, 2023, doi: 10.30598/variancevol5iss2page139-146.
- [10] Alvy Muhalim, D. Pratama, and N. Rahaningsih, “Analisa Perbandingan Metode Algoritma Iterative Dichotomiser 3 Dan Algoritma C.45 Dalam Pengukuran Kepuasan Konsumen,” *Kopertip J. Ilm. Manaj. Inform. dan Komput.*, vol. 6, no. 3, pp. 71–75, Oct. 2022, doi: 10.32485/kopertip.v6i3.176.
- [11] T. Faizah, “Perbandingan Algoritma C4.5 Dan Id3 Untuk Prediksi Ketepatan Waktu Lulus Mahasiswa,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 980–990, 2021, doi: 10.35957/jatisi.v8i2.593.
- [12] M. N. Sholikhah, D. Rahmalia, and M. S. Pradana, “Penerapan Algoritma ID3 dan Algoritma C4.5 Untuk Klasifikasi Penerima BPNT,” *Unisda J. Math. Comput. Sci.*, vol. 9, no. 2, pp. 21–28, 2023, doi: 10.52166/ujmc.v9i2.6111.
- [13] T. Kapri, M. Nasir, and E. P. Agustini, “Implementasi Algoritma Iterative Dichotomiser 3 (ID3) Untuk Penentuan Jumlah Dana Bantuan Perbaikan Rumah Di Bappeda,” *J. Comput. Inf. Syst. Ampera*, vol. 2, no. 1, pp. 58–67, 2021, doi: 10.51519/journalcisa.v2i1.70.
- [14] R. Girsang, E. F. Ginting, and M. Hutasuhut, “Penerapan Algoritma C4.5 Pada Penentuan Penerima Program Bantuan Pemerintah Daerah,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 4, p. 449, 2022, doi: 10.53513/jursi.v1i4.5727.
- [15] N. W. Oktha Pratiwi, N. Widya Utami, and I. Gede Juliana Eka Putra, “Klasifikasi Penentuan Penerima Bantuan Sosial Tunai (BST) Menggunakan Algoritma C4.5 Di Desa Keramas, Gianyar, Bali,” *J. Inform. Teknol. dan Sains*, vol. 4, no. 3, pp. 101–107, 2022, doi: 10.51401/jinteks.v4i3.1667.
- [16] M. Han and I. L. Najord, “A Typical Model Evaluation System for Rural Vocational Education Against Poverty is Based on a Decision Tree Mining Algorithm,” *Inform.*, vol. 48, no. 9, pp. 37–52, 2024, doi: 10.31449/inf.v48i9.5670.
- [17] S. Danil, N. Rahaningsih, R. D. Dana, and M., “Peningkatan Klasifikasi Kemiskinan Indonesia Menggunakan Metode Decision Tree,” *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, pp. 829–835, 2025, doi: 10.23960/jitet.v13i2.6336.
- [18] C. Herdian, A. Kamila, and I. G. Agung Musa Budidarma, “Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi,” *Technol. J. Ilm.*, vol. 15, no. 1, p. 93, 2024, doi: 10.31602/tji.v15i1.13457.
- [19] V. No, Z. A. Mukharyahya, Y. P. Astuti, and O. N. Cahyani, “Edumatic : Jurnal Pendidikan Informatika Perbandingan Naive Bayes dan Support Vector Machine dalam Klasifikasi Tingkat Kemiskinan di Indonesia,” vol. 9, no. 1, pp. 119–128, 2025, doi: 10.29408/edumatic.v9i1.29512.
- [20] W. A. Firmansyah, U. Hayati, and Y. Arie Wijaya, “Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.
- [21] D. Ruswanti, D. Susilo, and R. Riani, “Implementasi CRISP-DM pada Data Mining untuk Melakukan Prediksi Pendapatan dengan Algoritma C.45,” *Go Infotech J. Ilm. STMIK AUB*, vol. 30, no. 1, pp. 111–121, 2024, doi: 10.36309/goi.v30i1.266.
- [22] J. T. M. A. Nazanah and M. I. Jambak, “Pemanfaatan Algoritma Decision Tree ID3 Bagi Manajemen Bimbel Untuk Menentukan Faktor Kelulusan Pada Sekolah Kedinasan,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, pp. 915–924, 2023, doi: 10.30865/klik.v3i6.791.
- [23] N. Qisthi, D. Kasoni, L. Liesnaningsih, and N. Heriyani, “Penerapan Data Mining Untuk Prediksi Stunting Pada Balita Menggunakan Algoritma C4.5,” *Insa. Pembang. Sist. Inf. dan Komput.*, vol. 12, no. 2, pp. 18–25, 2024, doi: 10.58217/ipsikom.v12i2.314.
- [24] R. Ubaidillah Fahmi, A. Anjani Arifiyanti, and T. Luhur Indayanti Sugata, “Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Midi Kriing Menggunakan Support Vector Machine (Svm),” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 3, pp. 4831–4839, 2025, doi: 10.36040/jati.v9i3.13783.
- [25] N. Rumbia et al., “Perbandingan Metode KNN dan Naive Bayes untuk Klasifikasi Kelulusan Mahasiswa Pada Mata Kuliah Probat,” *Jurnal PTI (Jurnal Pendidikan Teknologi Informasi)*, vol. 12, pp. 7–13, 2025, doi: 10.35134/jpti.v12i1.228.