

PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI KELAYAKAN PENERIMA PROGRAM INDONESIA PINTAR (PIP) DI SD NEGERI 13 JONGKONG

Ayang Kristina¹, Sinta Rukiastiandari²

^{1,2} Program Studi Teknik Informatika, Universitas Bina Sarana Informatika
 Jl. Keramat Raya, Jakarta – DKI Jakarta, Indonesia
 Email: ¹ayangkristina2015@gmail.com, ²sinta.sru@bsi.ac.id

ABSTRAK

Pendidikan merupakan fondasi utama dalam pembangunan sumber daya manusia, namun keterbatasan finansial sering menjadi hambatan bagi keluarga kurang mampu dalam mengakses pendidikan. Untuk menjawab masalah tersebut, pemerintah meluncurkan Program Indonesia Pintar (PIP) sebagai langkah strategis dalam menekan angka putus sekolah. Akan tetapi, di SD Negeri 13 Jongkong proses seleksi penerima PIP masih dilakukan secara manual sehingga rawan subjektivitas dan ketidaktepatan sasaran. Penelitian ini bertujuan menerapkan algoritma C4.5 sebagai sistem klasifikasi kelayakan penerima PIP agar seleksi lebih objektif dan akurat. Metode yang digunakan adalah pendekatan kuantitatif deskriptif dengan teknik data mining terhadap 70 data siswa yang diperoleh dari dokumentasi sekolah dan ekspor Dapodik. Atribut penelitian meliputi pekerjaan dan penghasilan orang tua, alat transportasi, status bantuan sosial, serta jumlah tanggungan anak. Analisis menggunakan algoritma C4.5 pada perangkat lunak RapidMiner dengan validasi melalui confusion matrix. Hasil penelitian menunjukkan bahwa atribut Penghasilan Orang Tua terpilih sebagai akar pohon keputusan karena memiliki nilai gain tertinggi, sedangkan evaluasi performa model menghasilkan akurasi 95,71%, precision 96,08%, recall 98,00%, dan F1-Score 97,03%. Temuan ini membuktikan bahwa algoritma C4.5 efektif digunakan sebagai sistem pendukung keputusan untuk meningkatkan ketepatan seleksi penerima PIP, serta dapat direkomendasikan untuk diimplementasikan lebih luas pada sekolah lain guna meminimalisasi bias dalam penentuan penerima bantuan.

Kata kunci: algoritma C4.5, data mining, klasifikasi, Program Indonesia Pintar, seleksi penerima bantuan

ABSTRACT

Education is the fundamental foundation for human resource development, yet financial constraints often hinder underprivileged families from accessing it. To address this issue, the government launched the Program Indonesia Pintar (PIP) as a strategic effort to reduce school dropout rates. However, at SD Negeri 13 Jongkong, the selection process for PIP recipients is still conducted manually, making it prone to subjectivity and misallocation. This study aims to apply the C4.5 algorithm as a classification system for determining PIP eligibility in order to ensure a more objective and accurate selection. The research employed a descriptive quantitative approach with data mining techniques on 70 student records obtained from school documentation and Dapodik exports. The research attributes include parents' occupation and income, transportation, social assistance status, and number of dependents. Analysis was conducted using the C4.5 algorithm in RapidMiner software, validated through a confusion matrix. The results show that the attribute Parents' Income was selected as the root node due to its highest gain value, while the performance evaluation achieved 95.71% accuracy, 96.08% precision, 98.00% recall, and a 97.03% F1-Score. These findings demonstrate that the C4.5 algorithm is effective as a decision support system to improve the accuracy of PIP recipient selection and can be recommended for broader implementation in other schools to minimize bias in beneficiary determination.

Keywords: C4.5 algorithm, classification, data mining, Indonesia Smart Program, scholarship recipient selection

1. PENDAHULUAN

Pendidikan memegang peranan penting dalam membentuk kualitas sumber daya manusia dan mendorong kemajuan bangsa. Namun, pemerataan akses pendidikan berkualitas masih menjadi tantangan di Indonesia, khususnya di daerah tertinggal, terdepan, dan terluar (3T) yang menghadapi rendahnya kualitas infrastruktur, minimnya tenaga pendidik, serta kondisi sosial ekonomi yang memprihatinkan sehingga akses pendidikan belum merata [1]. Kondisi ini berdampak pada tingginya risiko putus sekolah, terutama bagi peserta didik dari keluarga kurang mampu.

Sebagai intervensi, pemerintah meluncurkan Program Indonesia Pintar (PIP) untuk membantu peserta didik dari keluarga miskin atau rentan miskin agar dapat tetap mengenyam pendidikan. PIP diatur melalui petunjuk teknis penyaluran dengan kelayakan penerima berdasarkan integrasi data Data Pokok Pendidikan (Dapodik), Data



Terpadu Kesejahteraan Sosial (DTKS), dan Kartu Indonesia Pintar (KIP) [2], [3]. Secara konseptual, mekanisme ini dirancang untuk memastikan penyaluran bantuan tepat sasaran sesuai kriteria resmi. Di berbagai daerah, tantangan terbesar dalam penentuan penerima PIP adalah ketersediaan dan keakuratan data. Meskipun Dapodik telah menjadi basis utama pendataan, keterbatasan sumber daya manusia (SDM) di sekolah, terutama di wilayah pedesaan, membuat proses analisis dan verifikasi data sebelum pengimputan tidak dilakukan secara optimal. Akibatnya, pengisian data kerap dilakukan terburu-buru atau sekadar melengkapi administrasi tanpa memastikan kebenaran dan kelengkapan informasi. Cara konvensional ini tidak hanya kurang efisien, tetapi juga berisiko menghasilkan data kurang akurat sehingga seleksi penerima bantuan berpotensi tidak tepat sasaran [4].

Permasalahan tersebut juga ditemukan di SD Negeri 13 Jongkong, sebuah sekolah dasar di Dusun Jongkong, Desa Nanga Raku, Kecamatan Sayan, Kabupaten Melawi, Provinsi Kalimantan Barat. Meskipun jumlah siswa aktif tercatat kurang dari 80 pada tahun ajaran 2024/2025, data wawancara dan dokumentasi menunjukkan bahwa proses seleksi PIP masih belum optimal. Hal ini ditandai dengan adanya beberapa siswa dari keluarga kurang mampu yang belum masuk daftar calon penerima, sementara ada pula siswa dengan kondisi ekonomi relatif lebih baik yang justru terdata sebagai calon penerima. Keterbatasan guru dan tenaga administrasi juga menjadi kendala, karena sekolah tidak memiliki tenaga administrasi khusus sehingga pengisian Dapodik sepenuhnya dilakukan oleh guru yang merangkap tugas mengajar sekaligus administrasi. Kondisi ini berimplikasi pada ketidaklengkapan data siswa di Dapodik, terutama pada instrumen penting seperti pekerjaan orang tua, tingkat pendidikan, jumlah tanggungan, kepemilikan aset, kepemilikan KIP, serta kepesertaan dalam DTKS. Selain itu, hasil observasi menunjukkan bahwa penentuan calon penerima PIP masih dilakukan secara manual dan cenderung berdasarkan perkiraan subjektif tanpa verifikasi lapangan yang sistematis. Kondisi tersebut memperlihatkan bahwa tanpa adanya dukungan sistem yang lebih objektif, proses seleksi penerima bantuan rawan dipengaruhi faktor subjektifitas dan keterbatasan sumber daya. Akibatnya, tujuan utama PIP untuk menjamin akses pendidikan yang adil bagi anak-anak dari keluarga tidak mampu dapat terganggu. Oleh karena itu, dibutuhkan pendekatan berbasis teknologi yang mampu membantu sekolah dalam mengelola data dengan lebih cepat, tepat, dan transparan. Untuk menjawab persoalan tersebut, penelitian ini tidak hanya menganalisis ketepatan pengisian data pada Dapodik, tetapi juga melakukan verifikasi langsung ke lapangan guna memastikan kesesuaian data siswa dengan kondisi riil [2], [3], [5]–[11].

Ketidaklengkapan dan ketidakakuratan data ini dapat menyebabkan siswa yang berhak justru tidak diusulkan, sementara siswa yang kurang memenuhi kriteria tetap masuk dalam daftar penerima. Dampaknya, efektivitas PIP menurun dan dapat menimbulkan ketidakpuasan di masyarakat. Perkembangan teknologi analisis data menawarkan solusi yang lebih cepat, objektif, dan akurat dibandingkan metode konvensional. Salah satunya adalah algoritma data mining C4.5, yang membangun pohon keputusan dengan perhitungan information gain dan entropy untuk mengidentifikasi atribut paling berpengaruh dalam menentukan kelayakan penerima bantuan [12]. Beberapa penelitian terdahulu mendukung hal ini, misalnya menunjukkan bahwa algoritma C4.5 mampu mencapai akurasi hingga 96% dalam memprediksi kelayakan penerima PIP berbasis data Dapodik [10]. Penelitian lain melaporkan bahwa algoritma *Naïve Bayes* dapat mencapai akurasi sempurna 100% dibandingkan *Decision Tree* (C4.5) yang hanya mencapai 64,42% dalam klasifikasi penerima PIP [2]. Sebaliknya, penelitian lain menemukan bahwa algoritma C4.5 justru lebih unggul dengan akurasi 91,25% dan AUC 0,930 dibandingkan *Naïve Bayes* dengan akurasi 87,11% dan AUC 0,923 pada seleksi penerima PKH [3]. Perbedaan hasil tersebut menunjukkan bahwa efektivitas algoritma sangat dipengaruhi oleh konteks data dan domain penelitian, sehingga pengujian pada kasus spesifik di SD Negeri 13 Jongkong menjadi penting untuk menemukan model terbaik.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan menerapkan algoritma C4.5 untuk mengklasifikasikan kelayakan penerima PIP di SD Negeri 13 Jongkong agar proses seleksi lebih cepat, transparan, dan akurat [13], [14]. Adapun kontribusi penelitian ini adalah memberikan sistem pendukung keputusan yang dapat membantu sekolah dalam meningkatkan kualitas data Dapodik, mengurangi subjektivitas dalam penentuan penerima bantuan, serta menjadi rujukan bagi pengembangan mekanisme seleksi PIP di sekolah lain [5], [8], [15].

2. METODE PENELITIAN

Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode *data mining* [16]. Proses analisis dilakukan menggunakan algoritma C4.5 untuk mengklasifikasikan tingkat kelayakan penerima Program Indonesia Pintar (PIP) di SD Negeri 13 Jongkong. Teknik dan tahapan penelitian ini dilaksanakan melalui beberapa metode yang diuraikan sebagai berikut:

1. Studi Pustaka
Melakukan kajian literatur terkait algoritma C4.5, teknik validasi model, serta penerapan sistem pendukung keputusan di bidang pendidikan dasar. Sumber literatur diperoleh dari jurnal ilmiah terbitan lima tahun terakhir, buku akademik, dan dokumen resmi Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
2. Identifikasi Masalah
Menggalikan informasi melalui observasi lapangan dan wawancara dengan kepala sekolah serta operator Data Pokok Pendidikan (Dapodik) untuk memahami prosedur pengusulan PIP yang berlaku.
3. Pengumpulan Data
Menghimpun data primer berupa dokumen internal sekolah dan data ekspor dari sistem Dapodik.



4. Pra-Pemrosesan Data
Melaksanakan proses pembersihan data, transformasi, dan seleksi atribut sesuai tahapan *data mining* agar data siap diolah menggunakan algoritma C4.5.
5. Penerapan Algoritma
Membangun model klasifikasi dengan menggunakan perangkat lunak RapidMiner Studio 9.10 melalui perhitungan *entropy*, *information gain*, *gain ratio*, serta penerapan pruning untuk menghasilkan pohon keputusan yang optimal [17].
6. Evaluasi Model
Melakukan evaluasi kinerja model dengan *confusion matrix* yang mencakup ukuran akurasi, presisi, *recall*, dan *F1-score*.
7. Kesimpulan dan Saran
Menarik kesimpulan berdasarkan hasil analisis model klasifikasi serta memberikan rekomendasi perbaikan guna meningkatkan objektivitas dan efektivitas proses seleksi penerima PIP pada periode selanjutnya.

Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui tiga teknik utama, yaitu wawancara, observasi, dan dokumentasi. Teknik yang dilakukan diuraikan sebagai berikut:

1. Wawancara
Wawancara dilaksanakan secara langsung dan terstruktur dengan kepala sekolah serta operator Data Pokok Pendidikan (Dapodik) untuk memperoleh informasi mengenai prosedur dan kendala dalam pengusulan Program Indonesia Pintar (PIP). Hasil wawancara menunjukkan bahwa proses pengusulan masih menghadapi berbagai hambatan, di antaranya keterbatasan tenaga administrasi, ketidaklengkapan pengisian data siswa pada Dapodik, serta belum adanya mekanisme verifikasi lapangan yang sistematis dalam menentukan calon penerima.
2. Observasi (Pengamatan)
Observasi dilakukan dengan cara mengamati secara sistematis situasi di lapangan, termasuk mekanisme administrasi sekolah, kondisi sosial ekonomi siswa, serta proses verifikasi calon penerima PIP. Hasil observasi menunjukkan bahwa administrasi sekolah masih terbatas pada pencatatan manual dan pengisian Dapodik yang belum lengkap. Beberapa data penting, seperti pekerjaan orang tua dan jumlah tanggungan keluarga, sering kali kosong atau tidak akurat. Selain itu, kondisi sosial ekonomi sebagian besar siswa menunjukkan kebutuhan mendesak akan bantuan, namun proses verifikasi calon penerima masih bersifat perkiraan subjektif tanpa acuan instrumen yang jelas.
3. Dokumentasi
Dokumentasi digunakan untuk menghimpun data dari arsip internal sekolah dan data ekspor Dapodik yang berisi informasi detail mengenai identitas siswa, kondisi keluarga, serta status bantuan sosial.
Ketiga teknik ini saling melengkapi: wawancara memberikan pemahaman mendalam dari perspektif pelaksana, observasi memberikan gambaran faktual terkait praktik di lapangan, dan dokumentasi memastikan tersedianya data kuantitatif yang akurat. Dengan demikian, kombinasi ketiga metode ini diharapkan mampu menghasilkan data yang komprehensif, valid, dan dapat mendukung analisis klasifikasi menggunakan algoritma C4.5 secara lebih tepat sasaran. Selanjutnya, dataset yang digunakan dalam penelitian ini terdiri atas 70 entri siswa SD Negeri 13 Jongkong, Provinsi Kalimantan Barat, yang mencakup peserta didik dari kelas I hingga VI. Data tersebut bersumber dari Dapodik serta dokumen internal sekolah, dengan atribut penelitian yang mencerminkan kondisi siswa dan keluarga sebagai dasar penentuan kelayakan penerima Program Indonesia Pintar. Adapun atribut yang digunakan dalam penelitian ini adalah sebagai berikut:

Tabel 1. Atribut Penelitian

(Sumber: Kementerian Pendidikan dan Kebudayaan Republik Indonesia, 2020; Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2024)

No	Atribut	Tipe Data	Keterangan
1	Pekerjaan Orangtua	Kategorikal	Buruh, Guru, Karyawan Swasta, Petani, PNS, Tukang, Wiraswasta
2	Penghasilan orangtua	Kategorikal	Tinggi, sedang, rendah
3	Alata Transportasi	Kategorikal	Dalam rupiah perbulan
4	Jumlah Tanggungan	Numerik	Jumlah anak dalam keluarga
5	Status Bantuan Sosial	Kategorikal	Ya/Tidak (PKH, KKS, BPNT, KIP)
6	Status PIP	Kategorikal	Usulan Sekolah : Ya/Tidak

Atribut penelitian pada Tabel 1 tidak hanya ditentukan berdasarkan kondisi empiris di sekolah, tetapi juga dibandingkan dengan kriteria resmi sebagaimana diatur dalam Persesjen Kemendikbudristek Nomor 19 Tahun 2024 tentang Petunjuk Pelaksanaan Program Indonesia Pintar Pendidikan Dasar dan Menengah dan Permendikbud No. 10 Tahun 2020 tentang Program Indonesia Pintar. Dengan demikian variabel penelitian seperti pekerjaan orang tua, tingkat penghasilan, status bantuan sosial, kepemilikan Aset, dan kepemilikan KIP selaras dengan indikator kelayakan yang berlaku secara nasional [18], [19].

Penerapan Algoritma C4.5

Algoritma C4.5 adalah salah satu teknik klasifikasi berbasis pohon keputusan yang bekerja dengan membentuk model melalui pemilihan atribut secara bertahap, berdasarkan besarnya informasi yang diperoleh dari setiap atribut [20], [21]. Tahapan prosesnya mencakup:

1. Menghitung *entropy* dan *information gain* untuk setiap atribut.

Rumus *Entropy*:

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

Rumus *Information Gain*:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|Sv|}{|S|} \cdot Entropy(Sv) \dots\dots\dots(2)$$

2. Memilih atribut dengan gain ration tertinggi sebagai root node.
3. Melakukan pembentukan pohon keputusan hingga semua data terklasifikasi.
4. Menerapkan *post-pruning* dengan metode *Reduced Error Pruning* untuk menghilangkan cabang yang tidak relevan, sehingga model menjadi lebih sederhana dan menghindari overfitting.

Metrik Evaluasi

Kinerja model klasifikasi dalam penelitian ini diukur menggunakan empat metrik evaluasi utama, yaitu *akurasi*, *presisi*, *recall*, dan *F1-Score*. Perhitungan keempat metrik tersebut didasarkan pada nilai *True Positive* (TP), *True Negatif* (TN), *False Positif* (FP), maka *False Negatif* (FN) dari hasil klasifikasi. Berikut rumus metrik evaluasinya:

Rumus Akurasi:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \dots\dots\dots(3)$$

Rumus *Precision*:

$$precision = \frac{TP}{TP + FP} \dots\dots\dots(4)$$

Rumus *Recall*:

$$recall = \frac{TP}{TP + FN} \dots\dots\dots(5)$$

Rumus *F1-Score*:

$$F1 - Score = 2 \times \frac{presisi \times recall}{presisi + recall} \dots\dots\dots(6)$$

Akurasi pada Persamaan (3) menunjukkan proporsi klasifikasi yang benar terhadap seluruh data. Presisi (4) digunakan untuk mengukur ketepatan klasifikasi kelas positif. *Recall* (5) menggambarkan kemampuan model menemukan data positif secara benar. Sementara itu, *F1-Score* (6) merupakan rata-rata harmonis dari presisi dan recall, sehingga memberikan ukuran kinerja yang lebih seimbang.

3. HASIL DAN PEMBAHASAN

Hasil Klasifikasi

Penelitian ini dilaksanakan di SD Negeri 13 Jongkong sebagai lokasi utama pengambilan data, dengan tujuan mengevaluasi sekaligus meningkatkan ketepatan penyaluran Program Indonesia Pintar (PIP). Data diperoleh melalui wawancara dengan pihak sekolah, observasi kondisi sosial ekonomi siswa, serta dokumentasi data penerima dan non-penerima PIP pada semester genap tahun ajaran 2024/2025, termasuk impor data Dapodik yang diunduh pada 29 April 2025. Informasi yang dihimpun mencakup sejumlah atribut penentu kelayakan, seperti pekerjaan dan penghasilan orang tua, jumlah tanggungan anak, alat transportasi yang digunakan, serta status bantuan yang pernah diterima. Seluruh data kemudian diolah dan diklasifikasikan menggunakan algoritma C4.5. Dataset penelitian saat ini terdiri dari 70 siswa dari kelas I hingga VI, dengan 50 siswa termasuk kategori Layak (Ya) dan 20 siswa termasuk kategori Tidak Layak (Tidak). Identitas siswa tetap dijaga kerahasiaannya melalui penggunaan inisial pada data yang dianalisis. Berikut Tabel 2 menampilkan data internal SDN 13 Jongkong.

Tabel 2. Data Internal SDN 13 Jongkong
(Sumber: SDN Jongkong, 2025)

No	Nama	Pekerjaan Orang Tua	Penghasilan Orang Tua	Alat Transportasi	Status Bantuan	Tanggungan Anak
1	Siswa 1	Wiraswasta	Rp. 1.000.000 - Rp. 1.999.999	Mobil	Tidak	1
2	Siswa 2	PNS	Rp. 2.00.000 - Rp. 4.999.999	Mobil	Tidak	3



3	Siswa 3	Petani	Rp. 500.000 - Rp. 999.999	Motor	Ya	2
4	Siswa 4	Tukang	Rp. 1.000.000 - Rp. 1.999.999	Motor	Tidak	2
...
...
69	Siswa 69	Guru	Rp. 1.500.000 - Rp. 2.000.000	Jalan kaki	Ya	2
70	Siswa 70	Petani	Rp. 2.000.000 - Rp. 3.000.000	Motor	Ya	3

Tahap *Pre-Processing*

Tahap seleksi atribut bertujuan untuk mengidentifikasi variabel-variabel yang relevan dan berpengaruh dalam menentukan kelayakan siswa sebagai penerima Program Indonesia Pintar (PIP). Atribut yang dipilih tidak hanya berdasarkan asumsi peneliti, tetapi juga dianalisis menggunakan metode *information gain* pada algoritma C4.5 untuk mengetahui tingkat keterkaitan setiap atribut terhadap label klasifikasi [22]. Dengan cara ini dapat dipastikan bahwa atribut yang digunakan memang signifikan dalam memengaruhi hasil klasifikasi. Selanjutnya, dilakukan transformasi data untuk menyesuaikan format dengan kebutuhan algoritma C4.5, yakni dengan mengonversi data numerik menjadi bentuk kategorikal agar dapat diolah secara optimal. Berikut Transformasi yang dilakukan meliputi:

Tabel 3. Kelompok Penghasilan Orang Tua
(Sumber: Kemendikbudristek: Panduan Dapodik, 2024)

Jenis Penghasilan	Kelompok
Tidak berpenghasilan – kurang dari Rp. 500.000	Sangat Rendah
Rp. 500.001 – Rp. 1.499.999	Rendah
Rp. 1.500.000 – Rp. 3.499.999	Sedang
Rp. 3.500.000 – Rp. 20.000.000	Tinggi

Tabel 3 menunjukkan pengelompokan jenis penghasilan orang tua ke dalam empat kategori, yaitu sangat rendah, rendah, sedang, dan tinggi. Kategori sangat rendah mencakup kondisi tidak berpenghasilan hingga kurang dari Rp. 500.000 per bulan. Kategori rendah berada pada rentang Rp. 500.001 sampai dengan Rp. 1.499.999 per bulan. Selanjutnya, kategori sedang mencakup penghasilan antara Rp. 1.500.000 hingga Rp. 3.499.999 per bulan, yang umumnya menggambarkan keluarga dengan kondisi ekonomi menengah. Sementara itu, kategori tinggi mencakup penghasilan mulai dari Rp. 3.500.000 hingga Rp. 20.000.000 per bulan, yang menunjukkan kondisi ekonomi relatif lebih stabil. Pengelompokan ini dilakukan agar variabel penghasilan dapat dianalisis secara kategorikal dalam proses klasifikasi menggunakan algoritma C4.5 [23].

Tabel 4. Kelompok Jumlah Tanggungan Anak
(Sumber: Penelitian Mandiri, 2025)

Jumlah Tanggungan Anak	Kelompok
1-2	Sedikit
2-3	Sedang
3-5	Banyak

Tabel 4 menjelaskan pengelompokan jumlah tanggungan anak dalam keluarga ke dalam tiga kategori, yaitu sedikit, sedang, dan banyak. Kategori sedikit mencakup keluarga dengan 1–2 anak tanggungan, kategori sedang mencakup 2–3 anak, sedangkan kategori banyak mencakup 3–5 anak tanggungan. Pengelompokan ini bertujuan untuk menyederhanakan variabel jumlah tanggungan sehingga lebih mudah dianalisis dalam proses klasifikasi menggunakan algoritma C4.5. Selain itu, jumlah tanggungan anak menjadi indikator penting dalam menilai kondisi sosial ekonomi keluarga, karena semakin banyak anak yang ditanggung, semakin besar pula beban ekonomi yang harus dipenuhi oleh orang tua.

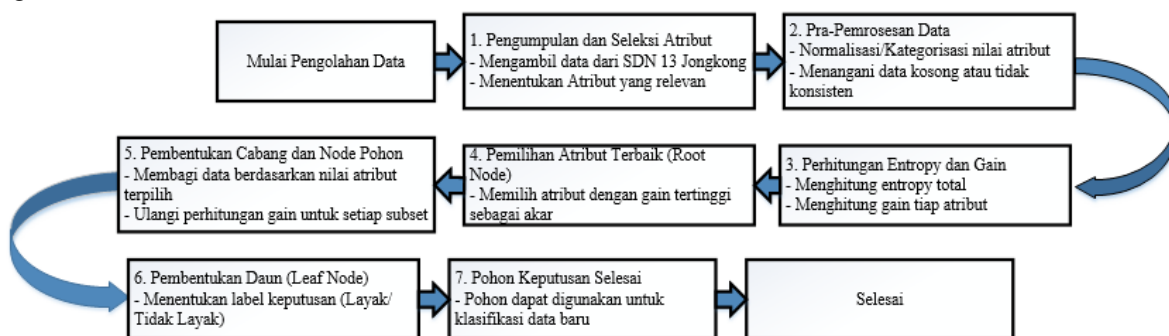
Tabel 5. Hasil *Pre-Processing* Data
(Sumber: Penelitian Mandiri, 2025)

Pekerjaan Orang Tua	Penghasilan Orang Tua	Alat Transportasi	Status Bantuan	Tanggungan Anak	LayaK PIP
Wiraswasta	Sedang	Mobil	Tidak	Sedikit	Tidak
PNS	Tinggi	Mobil	Tidak	Sedang	Tidak
Petani	Rendah	Motor	Ya	Sedikit	Ya
Tukang	Sedang	Motor	Tidak	Sedikit	Ya
Petani	Sedang	Motor	Tidak	Sedikit	Ya
...
...
PNS	Sedang	Motor	Ya	Sedang	Ya
Guru	Rendah	Jalan kaki	Ya	Sedikit	Ya
Petani	Rendah	Motor	Ya	Sedang	Ya

Tabel 5 menampilkan hasil pra-pemrosesan (*pre-processing*) data yang telah dikonversi ke dalam bentuk kategorikal agar sesuai dengan kebutuhan analisis menggunakan algoritma C4.5. Atribut yang digunakan meliputi pekerjaan orang tua, penghasilan orang tua, alat transportasi, status bantuan, serta jumlah tanggungan anak, dengan label akhir berupa kelayakan penerima Program Indonesia Pintar (Layak PIP). Sebagai contoh, siswa dengan orang tua berprofesi sebagai petani dengan penghasilan rendah, menggunakan alat transportasi motor, memiliki status bantuan “Ya”, serta jumlah tanggungan sedikit, dikategorikan sebagai “Layak PIP”. Sebaliknya, siswa dengan orang tua berpenghasilan tinggi dan berprofesi sebagai PNS, meskipun memiliki transportasi memadai, justru masuk kategori “Tidak Layak PIP” karena tidak memenuhi kriteria sasaran program. Proses ini memastikan setiap data siswa terstandarisasi sesuai indikator penelitian sehingga siap digunakan dalam tahap klasifikasi untuk menghasilkan model prediksi yang akurat.

Implementasi Algoritma C4.5 Perhitungan *Entropy* dan *Information Gain*

Perhitungan *entropy* dan *information gain* merupakan tahap penting dalam pembentukan pohon keputusan menggunakan algoritma C4.5. Tahapan ini bertujuan untuk mengukur tingkat ketidakpastian suatu atribut serta menentukan atribut mana yang paling informatif dalam proses klasifikasi. Dengan pendekatan ini, pohon keputusan dapat dibangun secara sistematis sehingga menghasilkan model klasifikasi yang akurat dan mampu mengidentifikasi kelayakan penerima bantuan secara objektif. Proses perhitungan tersebut dapat dijelaskan melalui diagram berikut:



Gambar 1. Proses *Decision Tree*
(Sumber: Penelitian Mandiri, 2025)

Diagram di atas menggambarkan alur penerapan algoritma C4.5 mulai dari pengumpulan hingga terbentuknya pohon keputusan. Proses dimulai dengan pengumpulan dan seleksi atribut dari SDN 13 Jongkong, dilanjutkan dengan pra-pemrosesan berupa normalisasi, kategorisasi, serta penanganan data kosong agar konsisten. Selanjutnya, dilakukan perhitungan *entropy* dan *information gain* untuk menentukan atribut dengan nilai gain tertinggi yang dipilih sebagai akar pohon. Proses ini berlanjut pada pembentukan cabang dan node dengan membagi data sesuai nilai atribut terpilih, kemudian menentukan label keputusan pada daun pohon (layak/tidak layak). Setelah pohon selesai terbentuk, model klasifikasi dapat digunakan untuk mengevaluasi maupun memprediksi kelayakan penerima bantuan pada data baru.

1. Perhitungan Entropy Total

Dataset penelitian ini terdiri dari 70 siswa SD Negeri 13 Jongkong pada semester genap tahun ajaran 2024/2025, dengan rincian 50 siswa termasuk dalam kategori Layak (Ya) dan 20 siswa termasuk dalam kategori Tidak Layak (Tidak). Meskipun jumlah data belum mencakup seluruh kemungkinan kombinasi kategori atribut ($7 \times 4 \times 3 \times 2 \times 3 = 504$ kombinasi), keberadaan 70 entri dianggap memadai untuk menggambarkan kondisi faktual pada sekolah yang menjadi lokasi penelitian. Hal ini sejalan dengan tujuan penelitian yang berfokus pada studi kasus di SD Negeri 13 Jongkong, sehingga kualitas analisis lebih dititikberatkan pada kesesuaian dengan realitas lapangan dibandingkan pada kelengkapan variasi kombinasi data. Dengan demikian, keterbatasan jumlah data tidak mengurangi validitas penelitian, melainkan memberikan gambaran yang kontekstual mengenai pola kelayakan penerima Program Indonesia Pintar (PIP) di lingkungan sekolah tersebut. Selanjutnya, data yang terkumpul digunakan sebagai dasar perhitungan *entropy* awal dan menjadi landasan pembentukan pohon keputusan melalui algoritma C4.5. Berikut perhitungan nilai *entropy* total sebagai berikut:

Entropy Total:

$$p_{\text{layak}} = \frac{50}{70} = 0,7142857$$

$$p_{\text{tidak layak}} = \frac{20}{70} = 0,2857143$$

$$\log_2(0,7142857) = \frac{\ln(0,7142857)}{\ln(2)} = \frac{-0,336472}{0,693147} = -0,485427$$

$$\log_2(0,2857143) = \frac{\ln(0,2857143)}{\ln(2)} = \frac{-1,252763}{0,693147} = -1,807355$$

$$\begin{aligned}
 Entropy(S) &= -(0,7142857 \cdot (-0,485427) + 0,2857143 \cdot (-1,807355)) \\
 &= -(-0,346734 + (-0,516387)) \\
 &= -(-0,863121) \\
 &= (0,863121)
 \end{aligned}$$

2. Perhitungan Nilai *Entropy* dan *Gain* Setiap Atribut
 a. Atribut Pekerjaan Orang Tua

Tabel 6. Atribut Pekerjaan Orang Tua
 (Sumber: Penelitian Mandiri, 2025)

Atribut	Kategori	Total	Layak	Tidak Layak
Pekerjaan Orang Tua	Buruh	5	1	4
	Guru	13	12	1
	Karyawan Swasta	4	3	1
	Petani	29	26	3
	PNS	7	4	3
	Tukang	3	1	2
	Wiraswasta	9	3	6

Berikut perhitungan *Entropy* masing-masing kategori:

Entropy Kategori Buruh:

$$\begin{aligned}
 p_{layak} &= \frac{1}{5} = 0,2 \\
 p_{tidak layak} &= \frac{4}{5} = 0,8 \\
 \log_2(0,2) &= \frac{\ln(0,2)}{\ln(2)} = \frac{-1,609438}{0,693147} = -2,321928 \\
 \log_2(0,8) &= \frac{\ln(0,8)}{\ln(2)} = \frac{-0,223144}{0,693147} = -0,321928 \\
 Entropy(S) &= -((0,2 \cdot -2,321928) + (0,6 \cdot -0,321928)) \\
 &= -(-0,464386 + -0,257542) \\
 &= 0,721928
 \end{aligned}$$

Entropy Kategori Guru:

$$\begin{aligned}
 p_{layak} &= \frac{12}{13} = 0,923077 \\
 p_{tidak layak} &= \frac{1}{13} = 0,076923 \\
 \log_2(0,923077) &= \frac{\ln(0,9230772)}{\ln(2)} = \frac{-1,07925}{0,693147} = -0,114333 \\
 \log_2(0,076923) &= \frac{\ln(0,0769238)}{\ln(2)} = \frac{-2,564949}{0,693147} = -3,700440 \\
 Entropy(S) &= -((0,923077 \cdot -0,114333) + (0,076923 \cdot -3,700440)) \\
 &= -(-0,105568 + (-0,284650)) \\
 &= 0,390218
 \end{aligned}$$

Entropy Kategori Karyawan Swasta:

$$\begin{aligned}
 p_{layak} &= \frac{3}{4} = 0,75 \\
 p_{tidak layak} &= \frac{1}{4} = 0,25 \\
 \log_2(0,75) &= \frac{\ln(0,75)}{\ln(2)} = \frac{-0,28768}{0,693147} = -0,41504 \\
 \log_2(0,25) &= \frac{\ln(0,25)}{\ln(2)} = \frac{-1,38629}{0,693147} = -2 \\
 Entropy(S) &= -((0,75 \cdot -0,41504) + (0,25 \cdot -2)) \\
 &= -(-0,31128 + (-0,5)) \\
 &= 0,81128
 \end{aligned}$$

Entropy Kategori Petani:

$$\begin{aligned}
 p_{layak} &= \frac{26}{29} = 0,89655 \\
 p_{tidak layak} &= \frac{3}{29} = 0,10345 \\
 \log_2(0,89655) &= \frac{\ln(0,89655)}{\ln(2)} = \frac{-0,10906}{0,693147} = -0,15734
 \end{aligned}$$

$$\log_2(0,10345) = \frac{\ln(0,10345)}{\ln(2)} = \frac{-2,26870}{0,693147} = -3,27310$$

$$\text{Entropy}(S) = -((0,89655 \cdot -0,15734) + (0,10345 \cdot -3,27310))$$

$$= -(-0,14106 + (-0,33857))$$

$$= 0,47963$$

Entropy Kategori PNS:

$$p_{\text{layak}} = \frac{4}{7} = 0,5714$$

$$p_{\text{tidak layak}} = \frac{3}{7} = 0,4286$$

$$\log_2(0,5714) = \frac{\ln(0,5714)}{\ln(2)} = \frac{-0,5596}{0,693147} = -0,8074$$

$$\log_2(0,4286) = \frac{\ln(0,4286)}{\ln(2)} = \frac{-0,8473}{0,693147} = -1,2224$$

$$\text{Entropy}(S) = -((0,5714 \cdot -0,8074) + (0,4286 \cdot -1,2224))$$

$$= -(-0,4614 + (-0,5239))$$

$$= 0,9853$$

Entropy Kategori Tukang:

$$p_{\text{layak}} = \frac{1}{3} = 0,3333$$

$$p_{\text{tidak layak}} = \frac{2}{3} = 0,6667$$

$$\log_2(0,3333) = \frac{\ln(0,3333)}{\ln(2)} = \frac{-1,0986}{0,693147} = -1,5850$$

$$\log_2(0,6667) = \frac{\ln(0,6667)}{\ln(2)} = \frac{-0,4055}{0,693147} = -0,5850$$

$$\text{Entropy}(S) = -((0,3333 \cdot -1,5850) + (0,6667 \cdot -0,5850))$$

$$= -(-0,5283 + (-0,3900))$$

$$= 0,9183$$

Entropy Kategori Wiraswasta:

$$p_{\text{layak}} = \frac{3}{9} = 0,3333$$

$$p_{\text{tidak layak}} = \frac{6}{9} = 0,6667$$

$$\log_2(0,3333) = \frac{\ln(0,3333)}{\ln(2)} = \frac{-1,0986}{0,693147} = -1,5850$$

$$\log_2(0,6667) = \frac{\ln(0,6667)}{\ln(2)} = \frac{-0,4055}{0,693147} = -0,5850$$

$$\text{Entropy}(S) = -((0,3333 \cdot -1,5850) + (0,6667 \cdot -0,5850))$$

$$= -(-0,5283 + (-0,3900))$$

$$= 0,9183$$

Gain untuk Atribut Pekerjaan Orang Tua:

$$= 0,86311 - \left(\frac{5}{70} \times 0,7219 + \frac{13}{70} \times 0,3912 + \frac{4}{70} \times 0,8113 + \frac{29}{70} \times 0,4798 + \frac{7}{70} \times 0,9852 + \frac{3}{70} \times 0,9183 \right)$$

$$+ \frac{9}{70} \times 0,9183$$

$$\text{Gain} = 0,8631 - 0,6254 = 0,2377$$

3. Penentuan Node Ke-1 Pohon Keputusan

Tabel 7. Perhitungan Node Ke-i
(Sumber: Penelitian Mandiri, 2025)

Atribut	Kategori	Total	Ya	Tidak	Entropy	Gain
i	Total	70	50	20	0,8631206	
Pekerjaan Orang Tua	Buruh	5	1	4	0,7219281	0,237803498
	Guru	13	12	1	0,3912436	
	Karyawan Swasta	4	3	1	0,8112781	
	Petani	29	26	3	0,479832024	
	PNS	7	4	3	0,985228136	
	Tukang	3	1	2	0,918295834	
	Wiraswasta	9	3	6	0,918295834	
		70				



					0,286306788
Penghasilan Orang Tua	Rendah	25	22	3	0,529361
	Sangat Rendah	1	1	0	0
	Sedang	35	27	8	0,775513
	Tinggi	9	0	9	0
		70			
					0,24580195
Alat Transportasi	Jalan kaki	20	16	4	0,7219281
	Mobil	15	3	12	0,7219281
	Motor	35	31	4	0,5127091
		70			
					0,061924387
Status Bantuan	Ya	47	38	9	0,7045767
	Tidak	23	12	11	0,998636
		70			
					0,040067616
Tanggungans Anak	Sedikit	42	28	14	0,9182958
	Sedang	23	17	6	0,8280557
	Banyak	5	5	0	0
		70			

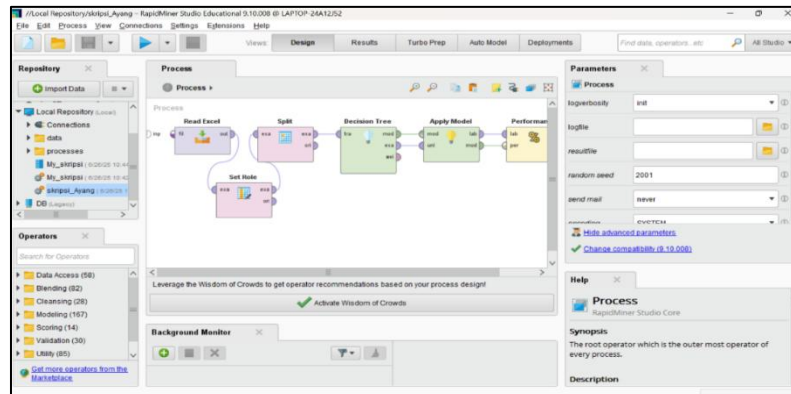
Hasil perhitungan *entropy* dan *information gain* menunjukkan bahwa atribut Penghasilan Orang Tua memiliki nilai *gain* tertinggi, yaitu 0,2863, sehingga dipilih sebagai akar pohon keputusan karena paling berpengaruh dalam mengurangi ketidakpastian klasifikasi. Atribut Alat Transportasi dan Pekerjaan Orang Tua menyusul dengan nilai *gain* masing-masing 0,2458 dan 0,2378, yang tetap berperan penting sebagai cabang dalam struktur pohon keputusan. Sementara itu, atribut Status Bantuan (0,0619) dan Jumlah Tanggungan Anak (0,0401) menunjukkan kontribusi yang relatif kecil. Dengan demikian, dapat disimpulkan bahwa faktor ekonomi berupa penghasilan orang tua merupakan variabel dominan dalam proses klasifikasi, sedangkan faktor bantuan sosial dan jumlah tanggungan anak memiliki pengaruh yang lebih rendah. Tabel 8 berikut menyajikan perhitungan lanjutan pada Atribut terpilih yaitu Penghasilan Orang Tua.

Tabel 8. Perhitungan Atribut Ke-ii (Penghasilan Orang Tua)
(Sumber: Penelitian Mandiri, 2025)

Atribut	Kategori	Total	Ya	Tidak	Entropy	Gain	
i.ii (Penghasilan Orang Tua)	Total	25	22	3	0,5293609		
						0,081339883	
Pekerjaan Orang Tua	Buruh	2	1	1	1		
	Guru	6	5	1	0,650022422		
	Karyawan Swasta	0	0	0	0		
	Petani	15	14	1	0,353359335		
	PNS	1	1	0	0		
	Tukang	0	0	0	0		
	Wiraswasta	1	1	0	0		
		25				0,020493851	
Rendah	Alat Transportasi	Jalan Kaki	14	13	1	0,3712323	
		Mobil	0	0	0	0	
		Motor	11	9	2	0,6840384	
		25					
Status Bantuan	Ya	19	18	1	0,2974722		
	Tidak	6	4	2	0,9182958		
		25				0,082890956	
Tanggungans Anak	Sedikit	14	11	3	0,7495953		
	Sedang	6	6	0	0		
	Banyak	5	5	0	0		
		25				0,109587521	

Tabel 8 menampilkan perhitungan lanjutan pada node ke-ii dengan atribut Penghasilan Orang Tua. Pada node ini, terdapat total 25 data dengan distribusi 22 siswa layak dan 3 siswa tidak layak, menghasilkan nilai *entropy* sebesar 0,5294. Selanjutnya, dilakukan evaluasi terhadap atribut turunan, yaitu Pekerjaan Orang Tua, Alat Transportasi, Status Bantuan, dan Jumlah Tanggungan Anak. Hasil perhitungan menunjukkan bahwa atribut Tanggungan Anak memberikan nilai *gain* tertinggi yaitu 0,1096, diikuti oleh Status Bantuan (0,0829), Pekerjaan Orang Tua (0,0813), dan Alat Transportasi (0,0205). Temuan ini mengindikasikan bahwa pada kelompok dengan

pra-pemrosesan kemudian diimpor, diolah, dan dianalisis untuk menghasilkan pohon keputusan sesuai dengan hasil perhitungan entropy dan gain sebelumnya. Dengan demikian, penggunaan RapidMiner tidak hanya mempermudah proses implementasi, tetapi juga memastikan bahwa model klasifikasi yang dibangun lebih sistematis, terstruktur, dan dapat divalidasi secara langsung melalui fitur evaluasi yang tersedia. Berikut langkah-langkah implementasi di Rapidminer:



Gambar 3. Pengolahan Algoritma C4.5
(Sumber: Penelitian Mandiri, 2025)

Alur proses implementasi pada RapidMiner ditunjukkan pada Gambar 3, yang terdiri dari beberapa tahapan utama. Pertama, data penelitian diimpor menggunakan operator *Read Excel* untuk memasukkan *dataset* yang sudah dipra-pemrosesan [24]. Selanjutnya, operator *Set Role* digunakan untuk menentukan atribut target atau label klasifikasi, yaitu kelayakan penerima PIP. *Dataset* kemudian dibagi menjadi data latih dan data uji melalui operator *Split Data*, sehingga model dapat dievaluasi secara lebih objektif. Pada tahap berikutnya, operator *Decision Tree* diterapkan untuk membangun model klasifikasi berbasis algoritma C4.5. Model yang terbentuk kemudian diuji menggunakan operator *Apply Model*, yang mengaplikasikan pohon keputusan terhadap data uji. Terakhir, operator *Performance* digunakan untuk mengevaluasi kinerja model melalui metrik evaluasi seperti akurasi, *presisi*, *recall*, dan *F1-score*. Dengan alur ini, RapidMiner secara sistematis memfasilitasi proses pembentukan dan validasi model klasifikasi.

Tree

```

Penghasilan Orang Tua = Rendah
| Tanggungan Anak = Banyak: Ya {Tidak=0, Ya=5}
| Tanggungan Anak = Sedang: Ya {Tidak=0, Ya=6}
| Tanggungan Anak = Sedikit
| | Pekerjaan Orang Tua = Buruh
| | | Alat Transportasi = Jalan kaki: Ya {Tidak=0, Ya=1}
| | | | Alat Transportasi = Motor: Tidak {Tidak=1, Ya=0}
| | | Pekerjaan Orang Tua = Guru
| | | | Alat Transportasi = Jalan kaki: Ya {Tidak=0, Ya=1}
| | | | | Alat Transportasi = Motor: Ya {Tidak=1, Ya=2}
| | | | Pekerjaan Orang Tua = PNS: Ya {Tidak=0, Ya=1}
| | | | Pekerjaan Orang Tua = Petani
| | | | | Alat Transportasi = Jalan kaki: Tidak {Tidak=1, Ya=1}
| | | | | | Alat Transportasi = Motor: Ya {Tidak=0, Ya=4}
| | | | | Pekerjaan Orang Tua = Wiraswasta: Ya {Tidak=0, Ya=1}
| | | | | Pekerjaan Orang Tua = Sangat Rendah: Ya {Tidak=0, Ya=1}
Penghasilan Orang Tua = Sangat Rendah: Ya {Tidak=0, Ya=1}
Penghasilan Orang Tua = Sedang
| Pekerjaan Orang Tua = Buruh: Tidak {Tidak=3, Ya=0}
| Pekerjaan Orang Tua = Guru: Ya {Tidak=0, Ya=7}
| Pekerjaan Orang Tua = Karyawan Swasta: Ya {Tidak=0, Ya=3}
| Pekerjaan Orang Tua = PNS: Ya {Tidak=0, Ya=3}
| Pekerjaan Orang Tua = Petani
| | Alat Transportasi = Jalan kaki: Ya {Tidak=0, Ya=1}
| | | Alat Transportasi = Mobil: Tidak {Tidak=1, Ya=0}
| | | | Alat Transportasi = Motor
| | | | | Status Bantuan = Tidak: Ya {Tidak=1, Ya=4}
| | | | | | Status Bantuan = Ya: Ya {Tidak=0, Ya=6}
| | | | | Pekerjaan Orang Tua = Tukang

```

Gambar 4. Rule Hasil *Decision Tree*
(Sumber: Penelitian Mandiri, 2025)

Berdasarkan hasil pembentukan pohon keputusan dengan algoritma C4.5, atribut Penghasilan Orang Tua terpilih sebagai akar pohon karena memiliki nilai gain tertinggi, sehingga menjadi faktor utama dalam klasifikasi kelayakan penerima bantuan PIP. Dari atribut ini terbentuk empat cabang utama, yaitu Rendah, Sangat Rendah, Sedang, dan Tinggi. Pada cabang Rendah, proses klasifikasi masih dilanjutkan ke atribut Tanggungan Anak. Jika jumlah tanggungan tergolong Banyak atau Sedang, maka siswa langsung dikategorikan Layak, sedangkan pada kategori Sedikit keputusan ditentukan lebih lanjut oleh Pekerjaan Orang Tua. Pada pekerjaan seperti PNS dan Wiraswasta, keputusan langsung Layak, sementara pekerjaan Buruh, Guru, dan Petani masih ditelusuri melalui atribut Alat Transportasi, yang menghasilkan variasi keputusan Layak maupun Tidak Layak sesuai jenis transportasi yang digunakan. Pada cabang Sangat Rendah, keputusan langsung jatuh pada kategori Layak, sementara pada cabang Tinggi seluruh data langsung diklasifikasikan sebagai Tidak Layak. Adapun cabang

Sedang mengarahkan ke atribut Pekerjaan Orang Tua, di mana pekerjaan Buruh langsung dikategorikan Tidak Layak, sementara Guru, Karyawan Swasta, dan PNS secara konsisten menghasilkan Layak. Pada kategori Petani, keputusan masih bercabang ke atribut Alat Transportasi yang selanjutnya diperhalus oleh Status Bantuan, namun pada akhirnya tetap dominan mengarah ke keputusan Layak. Sementara itu, pekerjaan Tukang dan Wiraswasta juga melibatkan atribut transportasi sebagai turunan, namun hasil akhirnya tetap cenderung Layak. Dengan demikian, pohon keputusan ini memperlihatkan bahwa meskipun faktor utama penentu klasifikasi adalah Penghasilan Orang Tua, atribut lain seperti Tanggungan Anak, Pekerjaan Orang Tua, Alat Transportasi, dan Status Bantuan tetap berkontribusi penting dalam memperjelas jalur klasifikasi hingga mencapai simpul daun berupa keputusan akhir Layak atau Tidak Layak.

Evaluasi Model

Berdasarkan hasil evaluasi model klasifikasi dengan menggunakan beberapa metrik, diperoleh performa yang sangat baik dalam menentukan kelayakan penerima bantuan PIP. Nilai akurasi mencapai 95,71%, menunjukkan bahwa sebagian besar klasifikasi model sesuai dengan data aktual. Hal ini tergambar dari *confusion matrix*, di mana terdapat 49 data “Layak” dan 18 data “Tidak Layak” yang berhasil diklasifikasikan dengan benar. Namun, masih terdapat 1 data “Tidak Layak” yang salah terklasifikasi sebagai “Layak” (*false positive*) dan 2 data “Layak” yang terklasifikasi keliru sebagai “Tidak Layak” (*false negative*). Dari sisi *precision*, model menghasilkan nilai 96,08% untuk kelas positif “Layak”, yang berarti hampir seluruh klasifikasi “Layak” tepat, meskipun masih ada sedikit kesalahan. Pada metrik *recall*, model menunjukkan sensitivitas yang sangat tinggi dengan nilai 98,00%, yang mengindikasikan bahwa hampir semua siswa yang benar-benar “Layak” dapat dikenali dengan baik, hanya 2 data yang terlewat. Keseimbangan antara *precision* dan *recall* tercermin pada nilai *F1-Score* sebesar 97,03%, yang menegaskan bahwa model tidak hanya mampu mengklasifikasikan mayoritas siswa yang layak, tetapi juga tetap selektif dalam menghindari kesalahan klasifikasi. Secara keseluruhan, kombinasi nilai akurasi, *precision*, *recall*, dan *F1-Score* ini menunjukkan bahwa algoritma C4.5 yang digunakan mampu memberikan klasifikasi yang handal, konsisten, dan akurat dalam mendukung keputusan penentuan kelayakan penerima bantuan.

Analisis Kesalahan Klasifikasi

Berdasarkan hasil klasifikasi kelayakan penerima Program Indonesia Pintar (PIP) menggunakan algoritma *Decision Tree* pada RapidMiner, diperoleh total 70 data siswa yang dibandingkan antara nilai aktual dengan hasil klasifikasi model. Dari jumlah tersebut, sebanyak 67 siswa berhasil diklasifikasikan dengan benar, sedangkan 3 data mengalami kesalahan klasifikasi, sehingga menghasilkan tingkat akurasi sebesar 95,71%. Tingkat keyakinan (*confidence*) model pada sebagian besar kasus sangat tinggi, ditunjukkan oleh nilai *confidence* 1.0 pada banyak data, yang menandakan keputusan klasifikasi cukup kuat. Namun, terdapat pula beberapa kasus dengan nilai *confidence* rendah (seperti 0,25; 0,333; dan 0,5) yang berkontribusi pada kesalahan klasifikasi. Kesalahan ini meliputi adanya siswa yang sebenarnya “Layak” namun terklasifikasi sebagai “Tidak Layak” (*false negative*), serta siswa yang seharusnya “Tidak Layak” tetapi masuk ke kategori “Layak” (*false positive*). Kondisi ini memperlihatkan bahwa meskipun model menunjukkan performa klasifikasi yang sangat baik secara keseluruhan, masih ada sebagian kecil data dengan tingkat keyakinan rendah yang berpotensi menimbulkan kesalahan klasifikasi. Berikut tabel 9. Hasil klasifikasi kelayakan.

Tabel 9. Hasil Klasifikasi Kelayakan
(Sumber: Penelitian Mandiri, 2025)

No	Kelayakan	Prediction	Confidence (Tidak)	Confidence (Ya)	Alat Transportasi	Pekerjaan Orang Tua	Penghasilan Orang Tua	Status Bantuan	Tanggungan Anak
1	Tidak	Tidak	1.0	0.0	Mobil	Wiraswasta	Sedang	Tidak	Sedikit
2	Tidak	Tidak	1.0	0.0	Mobil	PNS	Tinggi	Tidak	Sedang
3	Ya	Ya	0.0	1.0	Motor	Petani	Rendah	Ya	Sedikit
4	Ya	Ya	0.0	1.0	Motor	Tukang	Sedang	Tidak	Sedikit
5	Ya	Ya	0.2	0.8	Motor	Petani	Sedang	Tidak	Sedikit
...
...
69	Ya	Ya	0.0	1.0	Jalan kaki	Guru	Rendah	Ya	Sedikit
70	Ya	Ya	0.0	1.0	Motor	Petani	Sedang	Ya	Sedang

4. SIMPULAN

Berdasarkan analisis terhadap 70 data siswa SD Negeri 13 Jongkong menggunakan algoritma C4.5, penelitian ini menghasilkan model klasifikasi kelayakan penerima Program Indonesia Pintar (PIP) dengan kinerja sangat baik, ditunjukkan oleh akurasi 95,71%, *precision* 96,08%, *recall* 98,00%, dan *F1-Score* 97,03%, dengan distribusi data 50 siswa layak dan 20 tidak layak serta *confusion matrix* yang mencatat 49 *true positive*, 18 *true negative*, 2

false positive, dan 1 *false negative*. Atribut Penghasilan Orang Tua terpilih sebagai akar pohon keputusan karena memiliki gain tertinggi, diikuti atribut lain seperti Tanggungan Anak, Pekerjaan Orang Tua, Alat Transportasi, dan Status Bantuan, yang menunjukkan bahwa faktor ekonomi keluarga paling dominan dalam memengaruhi kelayakan, meskipun temuan ini masih perlu divalidasi dengan kondisi lapangan untuk memastikan kesesuaiannya dengan realitas sosial. Analisis kesalahan mengungkapkan adanya tiga data yang salah klasifikasi, umumnya disebabkan kombinasi atribut yang kontradiktif sehingga menghasilkan tingkat keyakinan rendah, dan hal ini menjadi dasar bahwa penelitian mendatang perlu melibatkan data yang lebih luas, memperkaya variabel sosial-ekonomi yang digunakan, serta mengeksplorasi algoritma lain agar hasil klasifikasi lebih konsisten dan kesalahan klasifikasi dapat diminimalkan, khususnya pada kasus siswa layak yang berpotensi tidak teridentifikasi.

DAFTAR PUSTAKA

- [1] S. Maulido, P. Karmijah, and V. Rahmi, "Upaya Meningkatkan Pendidikan Masyarakat Di Daerah Terpencil," *Jurnal Publikasi Ilmu Pendidikan, Pembelajaran dan Ilmu Sosial*, vol. 2, no. 1, pp. 198–208, 2023, doi: 10.61132/sadewa.v2i1.488.
- [2] A. Amalia, A. Irma Purnamasari, and I. Ali, "Implementasi Algoritma C4.5 Dan Naïve Bayes Dalam Pengambilan Keputusan Untuk Program Indonesia Pintar (Pip) Di Sekolah Dasar Negeri 04 Majalangu," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 2, pp. 1889–1896, 2024, doi: 10.36040/jati.v8i2.8311.
- [3] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Jurnal Sistem Informasi*, vol. 9, no. 1, pp. 103–115, 2020, doi: 10.32520/stmsi.v9i1.596.
- [4] N. W. O. Pratiwi, N. W. Utami, and I. H. J. E. Putra, "Klasifikasi Penentuan Penerima Bantuan Sosial Tunai (BST) Menggunakan Algoritma C4.5 Di Desa Keramas, Gianyar, Bali," *Jurnal Informatika, Teknologi dan Sains*, vol. 4, no. 3, pp. 101–107, 2022, doi: 10.51401/jinteks.v4i3.1667.
- [5] S. Yunita and V. N. Alaeyda, "Penerapan Algoritma C4.5 Untuk Prediksi Penerimaan Beasiswa di SD 4 Pelangsian," *Journal Innovative Creative and Information Technology*, vol. 8, no. 2, pp. 181–193, 2022, doi: 10.33050/icit.v8i2.2408.
- [6] B. F. Tanyu, A. Abbaspour, Y. Alimohammadlou, and G. Tecuci, "Landslide susceptibility analyses using Random Forest, C4.5, and C5.0 with balanced and unbalanced datasets," *Journal Catena*, vol. 2, no. 3, pp. 1–14, 2021, doi: 10.1016/j.catena.2021.105355.
- [7] B. A. C. Permana, R. Ahmad, H. Bahtiar, A. Sudianto, and I. Gunawan, "Classification of diabetes disease using decision tree algorithm (C4.5)," *Journal of Physics: Conference Series*, vol. 1, no. 1, pp. 1–8, 2021, doi: 10.1088/1742-6596/1869/1/012082.
- [8] N. Y. L. Gaol, "Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5," *Jurnal Informasi & Teknologi*, vol. 2, no. 1, pp. 23–29, 2020, doi: 10.37034/jidt.v2i1.22.
- [9] P. Fikria, R. K. Dinata, and Y. Afrillia, "Classification Of Outpatient Visit Status Walking at Dr. Zubir Mahmud Hospital Using Algoritma C4.5," *International Journal of Engineering, Science and Information Technology*, vol. 5, no. 3, pp. 1–11, 2025, doi: 10.52088/ijesty.v5i3.865.
- [10] N. Aprilyani, I. Zulfa, and H. Syahputra, "Penerapan Algoritma Decision Tree C4.5 Untuk Model Penentuan Penerima Beasiswa Program Indonesia Pintar (PIP) Studi Kasus Sma Negeri 3 Timang Gajah," *Jurnal Teknik Informatika dan Elektro*, vol. 5, no. 1, pp. 96–109, 2022, doi: 10.55542/jurtie.v5i1.452.
- [11] S. Saputra, N. Suarna, I. Ali, and D. Solihudin, "Improving the Classification Model of Smart Indonesia Program Recipients in Koorwilbidikcam Sumber Using the C4.5 Algorithm," *Journal of Artificial Intelligence and Engineering Applications*, vol. 4, no. 3, pp. 1734–1738, 2025, doi: 10.59934/jaiea.v4i3.1000.
- [12] B. Kanwal, R. S. Shoukat, S. U. Rehman, M. Kundi, T. AlSaedi, and A. Alahmadi, "A New Framework for Scholarship Predictor Using a Machine Learning Approach," *Intelligent Automation & Soft Computing*, vol. 39, no. 5, pp. 829–854, 2024, doi: 10.32604/iasc.2024.054645.
- [13] M. A. A. Lobo and A. C. Talakua, "Klasifikasi Data Penerima Beasiswa Menggunakan Algoritma C4.5," *Jurnal Penelitian Inovatif*, vol. 4, no. 2, pp. 575–582, 2024, doi: 10.54082/jupin.355.
- [14] D. M. A. Sanjaya and A. I. I. P. N. W. Utami, "Penerapan Data Mining untuk Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Algoritma C4.5: Studi Kasus STMIK Primakara," *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, vol. 6, no. 1, pp. 84–97, 2022, doi: 10.22437/jiituj.v6i1.19600.
- [15] W. A. Firmansyach, U. Hayati, and Y. Arie Wijaya, "Analisa Terjadinya Overfitting dan Underfitting pada Algoritma Naive Bayes dan Decision Tree dengan Teknik Cross Validation," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.
- [16] S. Susliansyah, S. Y. Wargiyo, H. Sumarno, H. Priyono, and L. Maulida, "Rancangan Aplikasi Algoritma C4.5 pada Stunting Balita Menggunakan Bahasa Python," *Jurnal Manajemen Informatika Komputer*, vol. 9, no. 1, pp. 254–265, 2025, doi: 10.33395/remik.v9i1.14426.
- [17] M. Anita, I. G. D. Yulianti, and S. V. Pasaribu, "Klasifikasi Faktor Risiko Penyakit Jantung Menggunakan Machine," *Jurnal Teknologi Informasi*, vol. 16, no. 1, pp. 68–78, 2025, doi: 10.52972/hoaq.vol16no1.
- [18] Kementerian Pendidikan dan Kebudayaan Republik Indonesia, "Peraturan Menteri Pendidikan dan

- Kebudayaan Republik Indonesia Nomor 10 Tahun 2020 tentang Program Indonesia Pintar,” Jakarta, 2020. [Online]. Available: <https://peraturan.bpk.go.id/Home/Details/163716/permendikbud-no-10-tahun-2020>
- [19] K. P. K. R. dan Teknologi, “Peraturan Sekretaris Jenderal Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 19 Tahun 2024 tentang Petunjuk Pelaksanaan Program Indonesia Pintar Pendidikan Dasar dan Pendidikan Menengah,” Jakarta, 2024.
- [20] L. Rhomaningtias, A. Khairunisa, S. Shella, M. Wara, and K. M. Hindrayani, “Analisis Sentimen Ulasan Aplikasi Smile Indonesia Menggunakan Metode Naive Bayes dan Support Vector Machine (SVM),” *Jurnal Teknologi Informasi*, vol. 16, no. 1, pp. 79–91, 2025, doi: 10.52972/hoaq.vol16no1.
- [21] M. Mardiansa, H. L. Sari, and P. Prahasti, “Penerapan Data Mining untuk Mengetahui Minat Siswa pada Pelajaran Matematika menggunakan Metode K-Means Clustering,” *Jurnal Multidisiplin Dehasen*, vol. 4, no. 2, pp. 693–702, 2023, doi: 10.31328/jointecs.v3i2.787.
- [22] B. Kanwal, R. S. Shoukat, S. U. Rehman, M. Kundi, T. AlSaedi, and A. Alahmadi, “A New Framework for Scholarship Predictor Using a Machine Learning Approach,” *Intelligent Automation & Soft Computing*, vol. 39, no. 5, pp. 829–854, 2024, doi: 10.32604/iasc.2024.054645.
- [23] D. S. S. M.A, *Panduan Pengisian Aplikasi Dapodik Versi 2024*, 2024th ed. Jakarta: Direktorat Jenderal Pendidikan Dasar dan Menengah, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2024. [Online]. Available: https://cdn-dapodik.kemdikbud.go.id/panduan/Panduan_Lengkap_Aplikasi_Dapodik_versi_2024.pdf
- [24] A. F. Rahmawati and B. F. Anbiya, “Pemanfaatan Microsoft Excel untuk Pengolahan Penilaian Sumatif Hasil Belajar Siswa PAI dalam Kurikulum Merdeka,” *Jurnal Harmoni Nusa Bangsa*, vol. 1, no. 2, pp. 126–141, 2024, doi: 10.47256/jhnb.v2i1.481.