

Pemanfaatan Algoritma C4.5 untuk Klasifikasi Siswa yang Layak di Ajukan dalam Memperoleh PIP

Utilization of the C4.5 Algorithm for Classifying Students Who Are Eligible to Apply for Obtaining PIP

Rizky Nazwa Fazira¹, Anjar Wanto², Indra Gunawan³
^{1,2,3} STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Article Info

Genesis Artikel:

Diterima, 22 Juni 2023

Direvisi, 29 Juni 2023

Disetujui, 30 Juni 2023

Kata Kunci:

Algoritma C4.5

Klasifikasi

Siswa

Bantuan Pendidikan

PIP

ABSTRAK

Penelitian ini bertujuan untuk mengoptimalkan proses seleksi siswa yang layak mendapatkan Program Indonesia Pintar (PIP) menggunakan algoritma C4.5. PIP merupakan program bantuan sosial yang ditujukan untuk siswa-siswa yang memenuhi kriteria tertentu. Algoritma C4.5 dipilih karena mampu menghasilkan model keputusan yang dapat digunakan untuk mengklasifikasikan siswa berdasarkan berbagai faktor penilaian. Studi ini melibatkan pengumpulan data siswa berdasarkan kriteria PIP yang telah ditentukan. Data tersebut kemudian digunakan sebagai input untuk melatih model menggunakan algoritma C4.5. Proses pelatihan melibatkan identifikasi pola dan hubungan antar variabel yang signifikan dalam menentukan kelayakan siswa untuk PIP. Hasil dari penelitian ini menghasilkan keputusan dalam penerima PIP (Program Indonesia Pintar) dengan klasifikasi penghasilan orang tua rendah yang berhak mendapatkan bantuan. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan efisiensi dan akurasi dalam proses seleksi siswa penerima PIP. Penggunaan algoritma C4.5 diharapkan dapat menghasilkan keputusan yang lebih obyektif dan dapat mengidentifikasi dengan lebih tepat siswa-siswa yang memenuhi syarat untuk mendapatkan bantuan tersebut. Selain itu, penelitian ini juga dapat memberikan wawasan baru terkait implementasi algoritma dalam konteks kebijakan pendidikan dan bantuan sosial.

ABSTRACT

This research aims to optimize the selection process for eligible students for the Smart Indonesia Program (PIP) using the C4.5 algorithm. PIP is a social assistance program aimed at students who meet specific criteria. The C4.5 algorithm was chosen because it can produce a decision model that can be used to classify students based on various assessment factors. This study involved collecting student data based on predetermined PIP criteria. This data is then used as input to train the model using the C4.5 algorithm. The training involves identifying patterns and relationships between significant variables in determining a student's eligibility for PIP. This research resulted in a decision regarding PIP (Smart Indonesia Program) recipients with a low parental income classification who were entitled to assistance. It is hoped that the results of this research can contribute to increasing efficiency and accuracy in the selection process for PIP recipient students. Using the C4.5 algorithm is expected to produce more objective decisions and identify students who qualify for assistance more precisely. Apart from that, this research can also provide new insights regarding implementing algorithms in the context of education policy and social assistance.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Korespondensi:

Rizky Nazwa Fazira

Program Studi Teknik Informatika,

STIKOM Tunas Bangsa, Pematangsiantar, Indonesia,

Email: fazirahnazwa@gmail.com

1. PENDAHULUAN

Pendidikan merupakan aspek krusial dalam pembangunan suatu negara, dan pemerintah memiliki tanggung jawab untuk memastikan akses dan pemerataan pendidikan bagi seluruh masyarakat. Dalam upaya mewujudkan tujuan tersebut, berbagai program bantuan sosial diperkenalkan untuk mendukung kelompok-kelompok tertentu, termasuk di antaranya Program Indonesia Pintar (PIP). PIP merupakan inisiatif pemerintah yang bertujuan memberikan bantuan keuangan kepada siswa-siswa yang memenuhi kriteria tertentu, sehingga mereka dapat melanjutkan pendidikan tanpa hambatan ekonomi. Meskipun PIP memiliki dampak positif yang signifikan, namun tantangan yang dihadapi dalam pelaksanaannya adalah proses seleksi penerima bantuan. Menilai dan memilih siswa-siswa yang benar-benar membutuhkan bantuan merupakan tugas yang kompleks, mengingat variasi dan dinamika kondisi sosial ekonomi setiap individu. Oleh karena itu, perlu adanya pendekatan yang lebih sistematis dan obyektif dalam menentukan siswa-siswa yang berhak mendapatkan PIP. Ada banyak algoritma kecerdasan buatan di dalam domain komputasi yang dapat digunakan untuk menyelesaikan situasi yang rumit, seperti data mining [1]–[5], pendukung keputusan [6]–[11], maupun *neural network* [12]–[18].

Terkait topik penelitian yang dilakukan, maka penelitian ini bertujuan untuk mengkaji potensi pemanfaatan algoritma C4.5 dalam konteks klasifikasi siswa yang layak diusulkan untuk mendapatkan PIP. Algoritma C4.5, yang dikembangkan oleh Ross Quinlan, merupakan salah satu algoritma pohon keputusan yang umum digunakan dalam data mining dan *machine learning*. Keunggulan utama algoritma ini terletak pada kemampuannya untuk menghasilkan model keputusan yang dapat diinterpretasi dengan baik.

Pemilihan algoritma C4.5 sebagai fokus penelitian didorong oleh keinginan untuk meningkatkan efisiensi dan akurasi dalam proses seleksi penerima PIP, dengan mengintegrasikan teknologi kecerdasan buatan dalam pengambilan keputusan, diharapkan dapat tercipta suatu sistem yang lebih obyektif dan dapat mengidentifikasi dengan lebih tepat siswa-siswa yang membutuhkan bantuan tersebut. Oleh karena itu, penelitian ini tidak hanya berfokus pada aplikasi teknologi, tetapi juga membahas implikasi kebijakan pendidikan dan bantuan sosial yang dapat dihasilkan dari pemanfaatan algoritma C4.5. Sebelum memasuki pembahasan lebih lanjut, perlu ditekankan bahwa penelitian ini tidak bermaksud menggantikan peran serta manusia dalam proses pengambilan keputusan. Sebaliknya, penelitian ini mengusulkan integrasi antara kecerdasan buatan dan kebijakan manusiawi untuk menciptakan sistem yang lebih efisien dan adil. Pemahaman mendalam tentang kondisi sosial dan konteks pendidikan tetap menjadi bagian integral dari proses seleksi, dan algoritma C4.5 dianggap sebagai alat bantu untuk meningkatkan obyektivitas dalam pengambilan keputusan. Selain itu, penelitian ini diharapkan dapat memberikan pandangan baru terkait implementasi teknologi dalam konteks kebijakan pendidikan dan bantuan sosial. Kemajuan teknologi telah membuka peluang baru untuk mengatasi tantangan kompleks dalam berbagai bidang, termasuk pendidikan. Dengan menggunakan pendekatan yang inovatif, seperti pemanfaatan algoritma C4.5, kita dapat merancang solusi yang lebih adaptif dan efektif.

Berbagai penelitian terdahulu tentang pemanfaatan algoritma C4.5 yang melatarbelakangi dilakukan nya penelitian ini diantaranya: Penelitian untuk melakukan klasifikasi calon penerima beasiswa dengan menggunakan algoritma C4.5. Beberapa indikator penelitian yang digunakan diantaranya Pendapatan orang tua, Jumlah Tanggungan, Nilai Ujian, dan Nilai Tes. Hasil penelitian ini memperoleh akurasi mencapai 85% dan digambarkan berdasarkan pada *Decision Tree* (pohon keputusan) [19]. Penelitian untuk mengklasifikasi Penerima Beasiswa pada Universitas Hamzanwadi menggunakan algoritma Naive Bayes. Variabel penelitian terdiri dari tujuh atribut, yaitu: Nama, Status DTKS, prestasi, pekerjaan orang tua, jumlah penghasilan orang tua, kepemilikan rumah, dan jumlah tanggungan keluarga. Pengujian dilakukan dengan k-fold cross validation, dan hasil akurasi tertinggi diperoleh dari k-fold 4, mencapai 91,43%, dengan AUC sebesar 0,996% dan diklasifikasikan sebagai excellent. Oleh karena itu, dapat disimpulkan bahwa algoritma Naive Bayes sangat efektif untuk melakukan seleksi kelulusan pemberkasan beasiswa bidik misi di Universitas Hamzanwadi [20]. Penelitian berikutnya dilakukan dengan memanfaatkan algoritma *Decision Tree* C4.5 untuk membuat model penentuan penerima beasiswa Program Indonesia Pintar (PIP) pada SMA Negeri 3 Timang Gajah, Kabupaten Bener Meriah, Aceh. Berdasarkan hasil pengukuran performa model *Decision Tree* yang telah dilakukan diperoleh hasil akurasi Layak PIP dan tidak Layak PIP sebesar 72% artinya, Precision atau siswa yang Layak PIP dari keseluruhan yang diprediksi Layak PIP sebesar 85 %, Recall atau siswa yang Layak PIP dibandingkan dengan keseluruhan siswa yang sebenarnya Layak PIP sebesar 69% dan Specificity atau siswa yang diprediksi tidak Layak PIP dibandingkan dengan keseluruhan siswa yang sebenarnya tidak Layak PIP sebesar 78% [21]. Selanjutnya penelitian untuk menganalisis Algoritma C.45 terhadap penentuan rekomendasi penerima beasiswa pada SMP Swasta Methodist-8 Medan. Penelitian ini menggunakan data nama siswa, kelas, penghasilan orang tua, tanggungan orang tua, dan nilai rata-rata rapor. Berdasarkan evaluasi yang telah dilakukan dalam pengujian ini diperoleh nilai akurasi 93.33 %, hal ini bisa menjadi bukti bahwa algoritma C4.5 cukup akurat untuk memberikan rekomendasi beasiswa [22].

Pada konteks globalisasi dan perkembangan teknologi informasi, pemanfaatan algoritma C4.5 tidak hanya relevan untuk meningkatkan proses seleksi penerima PIP di Indonesia, tetapi juga dapat menjadi inspirasi bagi negara-negara lain yang menghadapi tantangan serupa. Implikasi dan temuan penelitian ini diharapkan dapat memberikan kontribusi positif dalam pengembangan kebijakan pendidikan dan bantuan sosial di berbagai konteks nasional maupun internasional. Oleh karena itu, pendahuluan ini menggarisbawahi urgensi penelitian ini dalam menghadirkan solusi inovatif untuk meningkatkan efisiensi dan keadilan dalam pendistribusian bantuan pendidikan. Dengan fokus pada pemanfaatan algoritma C4.5, penelitian ini diharapkan dapat membuka wawasan baru dalam memahami bagaimana kecerdasan buatan dapat diintegrasikan dalam konteks kebijakan sosial dan pendidikan. Selanjutnya, pembahasan selanjutnya akan merinci metodologi penelitian, kerangka teoritis, serta ekspektasi dari hasil penelitian ini.

2. METODE PENELITIAN

2.1. Data Penelitian

Penelitian ini menggunakan data siswa pada SMA Negeri 2 Pematangsiantar. Pengumpulan data dilakukan dengan teknik observasi ke bagian tata usaha sekolah tersebut, yang menyimpan database siswa. Data siswa yang digunakan untuk penelitian dapat dilihat pada Tabel 1 berikut ini:

Tabel 1. Data Siswa SMA Negeri 2 Pematangsiantar

No	Jenis Tinggal	Alat Transportasi	Penerima KPS	Pekerjaan Orang tua	Penghasilan Orang tua	Penerima KIP
1	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Tidak
2	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 500,000 - Rp. 999,999	Ya
3	Bersama orang tua	Kendaraan pribadi	Terima	Wiraswasta	Rp. 2,000,000 - Rp. 4,999,999	Tidak
4	Bersama orang tua	Jalan kaki	Terima	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Tidak
5	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 500,000 - Rp. 999,999	Tidak
6	Bersama orang tua	Angkutan Umum	Tidak	Wiraswasta	Rp. 2,000,000 - Rp. 4,999,999	Tidak
7	Bersama orang tua	Angkutan Umum	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Ya
8	Bersama orang tua	Jalan kaki	Terima	Wiraswasta	Rp. 500,000 - Rp. 999,999	Ya
9	Bersama orang tua	Kendaraan pribadi	Tidak	Wiraswasta	Rp. 2,000,000 - Rp. 4,999,999	Tidak
10	Bersama orang tua	Jalan kaki	Tidak	Tidak bekerja	Tidak Berpenghasilan	Tidak
11	Bersama orang tua	Jalan kaki	Tidak	PNS	Rp. 2,000,000 - Rp. 4,999,999	Tidak
12	Bersama orang tua	Kendaraan pribadi	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Tidak
13	Bersama orang tua	Kendaraan pribadi	Tidak	Buruh	Rp. 500,000 - Rp. 999,999	Tidak
14	Bersama orang tua	Angkutan Umum	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Ya
15	Bersama orang tua	Kendaraan pribadi	Tidak	Karyawan	Rp. 1,000,000 - Rp. 1,999,999	Tidak
16	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 500,000 - Rp. 999,999	Ya
17	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Tidak
18	Bersama orang tua	Angkutan Umum	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Tidak
19	Bersama orang tua	Jalan kaki	Tidak	Karyawan	Rp. 1,000,000 - Rp. 1,999,999	Ya
20	Bersama orang tua	Kendaraan pribadi	Tidak	PNS	Rp. 2,000,000 - Rp. 4,999,999	Tidak
21	Bersama orang tua	Angkutan Umum	Tidak	Wiraswasta	Rp. 2,000,000 - Rp. 4,999,999	Tidak
22	Bersama orang tua	Jalan kaki	Tidak	Buruh	Rp. 500,000 - Rp. 999,999	Tidak
23	Bersama orang tua	Angkutan Umum	Tidak	Wiraswasta	Rp. 500,000 - Rp. 999,999	Ya
24	Bersama orang tua	Angkutan Umum	Terima	Wiraswasta	Rp. 500,000 - Rp. 999,999	Ya
25	Bersama orang tua	Kendaraan pribadi	Tidak	Lainnya	Rp. 2,000,000 - Rp. 4,999,999	Tidak
26	Bersama orang tua	Angkutan Umum	Tidak	Lainnya	Rp. 2,000,000 - Rp. 4,999,999	Tidak
27	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 2,000,000 - Rp. 4,999,999	Tidak
28	Bersama orang tua	Angkutan Umum	Tidak	Wiraswasta	Rp. 500,000 - Rp. 999,999	Ya
29	Bersama orang tua	Kendaraan pribadi	Tidak	Buruh	Rp. 500,000 - Rp. 999,999	Tidak
30	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 500,000 - Rp. 999,999	Ya
31	Bersama orang tua	Jalan kaki	Tidak	Karyawan	Rp. 1,000,000 - Rp. 1,999,999	Ya
32	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Ya
33	Bersama orang tua	Angkutan Umum	Tidak	Wiraswasta	Rp. 500,000 - Rp. 999,999	Tidak
34	Bersama orang tua	Jalan kaki	Terima	Karyawan	Rp. 1,000,000 - Rp. 1,999,999	Tidak
35	Bersama orang tua	Kendaraan pribadi	Tidak	Karyawan	Rp. 2,000,000 - Rp. 4,999,999	Tidak
36	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Tidak
37	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 500,000 - Rp. 999,999	Tidak
38	Bersama orang tua	Jalan kaki	Tidak	Tidak bekerja	Tidak Berpenghasilan	Ya
39	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Tidak
40	Bersama orang tua	Jalan kaki	Terima	Tidak bekerja	Tidak Berpenghasilan	Ya
41	Bersama orang tua	Kendaraan pribadi	Tidak	PNS	Rp. 2,000,000 - Rp. 4,999,999	Tidak
42	Bersama orang tua	Kendaraan pribadi	Terima	Wiraswasta	Rp. 500,000 - Rp. 999,999	Ya
43	Bersama orang tua	Jalan kaki	Tidak	Tidak bekerja	Tidak Berpenghasilan	Tidak
44	Kost	Angkutan Umum	Tidak	Tidak bekerja	Tidak Berpenghasilan	Tidak
45	Bersama orang tua	Jalan kaki	Tidak	Karyawan	Rp. 1,000,000 - Rp. 1,999,999	Tidak
46	Bersama orang tua	Angkutan Umum	Tidak	PNS	Rp. 2,000,000 - Rp. 4,999,999	Tidak
47	Bersama orang tua	Angkutan Umum	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Tidak
48	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Ya
49	Bersama orang tua	Jalan kaki	Tidak	Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	Tidak
...	...	...	...	...	...	...
100	Bersama orang tua	Angkutan Umum	Tidak	PNS	Rp. 2,000,000 - Rp. 4,999,999	Tidak

2.2. Algoritma C4.5

Pada algoritma C4.5, pohon keputusan dibentuk berdasarkan kriteria pengambilan keputusan. Metode pohon keputusan adalah metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mentransformasikan fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan. Aturan dapat dengan mudah dimengerti dengan Bahasa Alamiah. Mereka juga dapat diungkapkan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari catatan dalam kategori tertentu. Secara umum, algoritma C4.5 membangun pohon keputusan dengan memilih atribut sebagai akar, membuat cabang untuk setiap nilai kasus di cabang tersebut, dan proses ini akan diulang untuk setiap cabang sampai kasus di cabang memiliki kelas yang sama. Pemilihan atribut didasarkan pada nilai gain tertinggi, menggunakan persamaan (1).

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

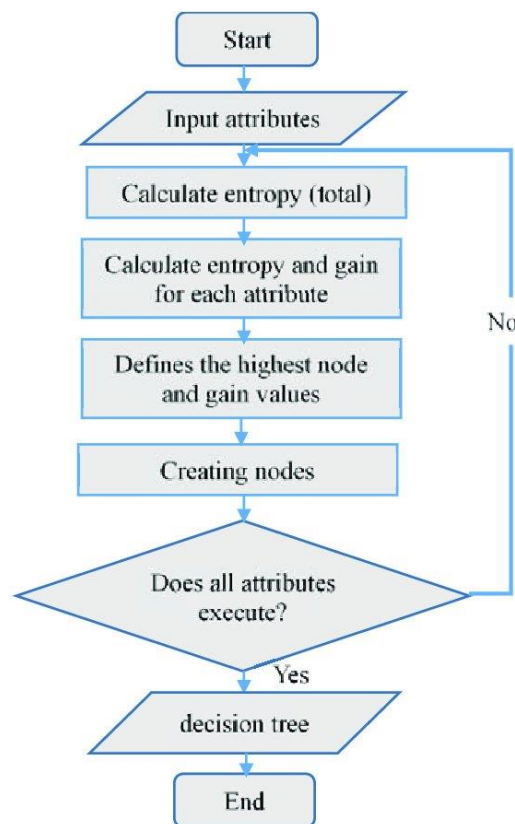
di mana S adalah kasus himpunan, A adalah atribut, N adalah jumlah partisi atribut A, $|S_i|$ adalah jumlah kasus pada partisi ke- i , dan $|S|$ adalah jumlah kasus dalam S. Nilai entropi dihitung sebelum mendapatkan nilai gain. Entropi digunakan untuk menentukan seberapa informatif suatu atribut dalam menghasilkan atribut. Formula dasar entropi seperti pada persamaan (2).

$$Entropy(S) = \sum_{i=1}^n \frac{|S_i|}{|S|} - \pi * \log_2 \pi \quad (2)$$

di mana S adalah kasus himpunan, A adalah fitur, N adalah jumlah partisi S, π adalah proporsi dari S_i ke S.

2.3. Desain Penelitian

Tujuan utama desain adalah memberikan deskripsi desain yang akan dibangun, serta memahami alur informasi dan proses dalam sistem. Gambar 1 menentukan tahapan-tahapan yang akan dilakukan dalam desain sistem. Proses perhitungan dilakukan dengan metode algoritma C4.5, untuk mendapatkan nilai entropi dan nilai gain, yang nantinya akan digunakan untuk membuat pohon keputusan dengan node dan cabang.



Gambar 1. Desain Penelitian Menggunakan Algoritma C4.5

3. HASIL DAN PEMBAHASAN

3.1. Perhitungan Algoritma C4.5

Proses perhitungan algoritma C4.5 pada penelitian ini terdapat 3 langkah yang harus dilakukan sebagai berikut:

1. Melakukan pre-processing data, dimana data mentah pada Tabel 1 harus diubah terlebih dahulu sebelum dilakukan perhitungan algoritma C4.5.
2. Menghitung jumlah kasus. Jumlah kasus untuk keputusan Ya, dan jumlah kasus untuk keputusan Tidak.
3. Menghitung Entropy dari semua kasus dan kasus yang dibagi berdasarkan kelas atribut. Selanjutnya perhitungan Gain untuk masing-masing atribut. Proses perhitungan dan pengujian algoritma C4.5 dilakukan dengan tools MS. Excel dan Rapid Miner.

Selanjutnya dilakukan preprocessing data pada dataset penelitian (Tabel 1), dimana untuk penghasilan orang tua diubah menjadi 3 kelompok, yang dapat dilihat pada Tabel 2 berikut.

Tabel 2. Kelompok Penghasilan Orang Tua

Jenis Penghasilan	Kelompok
Tidak berpenghasilan Kurang dari Rp 500.000	Rendah
Rp 500.000-999.99	Sedang
Rp 1.000.000-1.999.999	
Rp 2.000.000-4.999.999	Tinggi
Rp 5.000.000-20.000.000	

Setelah dilakukan pre-processing data dengan mengganti penghasilan orang tua menjadi 3 kelompok, maka dapat dijelaskan bahwa: Tidak berpenghasilan - kurang dari Rp 500.000 termasuk kelompok Rendah, penghasilan Rp 500.000 - Rp 1.999.999 termasuk kelompok sedang, dan untuk penghasilan Rp 2.000.000 - 20.000.000 termasuk kelompok tinggi.

Menghitung Entropy total :

$$= -(66/100) * \log_2 (66/100) - (27/100) * \log_2 (27/100)$$

$$= 0,905666512$$

Menghitung Entropy dan Gain Jenis Tinggal (N1) :

Entropy (Bersama orang tua):

$$= -(71/99) * \log_2 (71/99) - (28/99) * \log_2 (28/99)$$

$$= 0,859275981$$

Entropy (Kost):

$$= -(1/1) * \log_2 (1/1) - (0/1) * \log_2 (0/1)$$

$$= 0$$

Gain Jenis Tinggal:

$$= 0,905666512 - ((1/100) * 0 + (99/100) * 0,859275981)$$

$$= 0,054983291$$

Menghitung Entropy dan Gain Alat Transportasi (N1) :

Entropy (Jalan Kaki):

$$= -(28/49) * \log_2 (28/49) - (19/49) * \log_2 (19/49)$$

$$= 0,991322492$$

Entropy (Kendaraan Pribadi):

$$= -(19/21) * \log_2 (19/21) - (1/21) * \log_2 (1/21)$$

$$= 0,339796462$$

Entropy (Angkutan Umum):

$$= -(22/30) * \log_2 (22/30) - (7/30) * \log_2 (7/30)$$

$$= 0,81802824$$

Gain Alat Transportasi:

$$= 0,905666512 - ((30/100) * 0,81802824 + (21/100) * 0,339796462 + (49/100) * 0,991322492)$$

$$= 0,054983291$$

Menghitung Entropy dan Gain Pekerjaan Orang Tua (N1) :

Entropy (Buruh):

$$= -(1/1) * \log_2 (1/1) - (0/1) * \log_2 (0/1)$$

$$= 0$$

Entropy (Karyawan):

$$= -(7/9) * \log_2 (7/9) - (2/9) * \log_2 (2/9)$$

$$= 0,764204507$$

Entropy (PNS):

$$= -(10/10) * \log_2 (10/10) - (0/10) * \log_2 (0/10)$$

$$= 0$$

Entropy (Wiraswasta):

$$= -(45/68) * \log_2 (45/68) - (23/68) * \log_2 (23/68)$$

$$= 0,923119984$$

Entropy (Tidak Bekerja):

$$= -(4/6) * \log_2 (4/6) - (2/6) * \log_2 (2/6)$$

$$= 0,918295834$$

Entropy (Lainnya):

$$= -(2/2) * \log_2 (2/2) - (0/2) * \log_2 (0/2)$$

$$= 0$$

Gain Alat Transportasi:

$$= 0,905666512 - ((2/100) * 0 + (6/100) * 0,918295834 + (68/100) * 0,991322492 + (10/100) * 0 + (9/100) * 0,764204507 + (1/100) * 0)$$

$$= 0,154068768$$

Menghitung Entropy dan Gain Penghasilan Orang Tua (N1) :

Entropy (Tinggi):

$$= -(23/23) * \log_2 (23/23) - (0/23) * \log_2 (0/23)$$

$$= 0$$

Entropy (Sedang):

$$= -(45/70) * \log_2 (45/70) - (25/70) * \log_2 (25/70)$$

$$= 0,940285959$$

Entropy (Rendah):

$$= -(4/7) * \log_2 (4/7) - (3/7) * \log_2 (3/7)$$

$$= 0,985228136$$

Gain Penghasilan Orang Tua:

$$= 0,905666512 - ((7/100) * 0,985228136 + (70/100) * 0,940285959) + (23/100) * 0$$

$$= \mathbf{0,178500372}$$

Menghitung Entropy dan Gain Penerima KPS (N1) :

Entropy (Terima):

$$= -(3/9) * \log_2 (3/9) - (6/9) * \log_2 (6/9)$$

$$= 0,918295834$$

Entropy (Tidak):

$$= -(69/91) * \log_2 (69/91) - (22/91) * \log_2 (22/91)$$

$$= 0,797951968$$

Gain Penerima KPS:

$$= 0,905666512 - ((91/100) * 0,797951968 + (9/100) * 0,918295834)$$

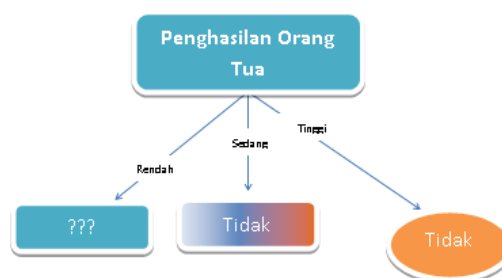
$$= 0,096883596$$

Untuk perhitungan node 1 *entropy* dan *gain* yang telah selesai dapat dilihat pada Tabel 3 berikut ini.

Tabel 3. Perhitungan Node 1

		Jumlah Kasus (S)	Ya (S ₁)	Tidak (S ₂)	Entropy	Gain
node 1		100	27	66	0,905666512	
Jenis tinggal	Bersama orang tua	99	28	71	0,859275981	0,054983291
	Kost	1	0	1	0	
	Jalan kaki	49	19	28	0,991322492	
Alat transportasi	Kendaraan pribadi	21	1	19	0,339796462	0,103152762
	Angkutan Umum	30	7	22	0,81802824	
	Buruh	5	0	5	0	
Pekerjaan Orang tua	Karyawan	9	2	7	0,764204507	0,154068768
	PNS	10	0	10	0	
	Wiraswasta	68	23	45	0,923119984	
	Tidak bekerja	6	2	4	0,918295834	
	Lainnya	2	0	2	0	
Penghasilan orang tua	Tinggi	23	0	23	0	0,178500372
	Sedang	70	25	45	0,940285959	
	Rendah	7	3	4	0,985228136	
Penerima KPS	Terima	9	6	3	0,918295834	0,096883596
	Tidak	91	22	69	0,797951968	

Pada Tabel 3 diperoleh nilai atribut dengan nilai *gain* terbesar yaitu Penghasilan Orang Tua dengan nilai sebesar 0,178500372. Untuk kategori dengan penghasilan orang tua tinggi menghasilkan keputusan Tidak. Untuk kategori penghasilan orang tua sedang menghasilkan 25 keputusan Ya dan 45 menghasilkan keputusan Tidak. Pada kategori penghasilan orang tua rendah menghasilkan 3 keputusan Ya dan 4 keputusan Tidak sehingga dilakukan perhitungan lebih lanjut. Pada perhitungan *node 1* menghasilkan pohon keputusan yang dapat dilihat pada Gambar 2 berikut.



Gambar 2. Pohon Keputusan Node 1

Pada Gambar 2 dapat dilihat bahwa atribut Penghasilan Orang Tua digunakan sebagai *node* akar. Sub atribut Tinggi sudah mendapatkan keputusan Tidak. Untuk sub atribut Sedang mendapat keputusan Tidak. Pada sub atribut Rendah belum

mendapatkan keputusan, maka perlu dilakukan perhitungan lebih lanjut untuk menentukan *node* akar lebih lanjut. Untung tabel perhitungan selanjutnya dapat dilihat pada Tabel 4 berikut.

Tabel 4. Perhitungan Node 2.1

	Rendah	Penghasilan Orang tua (S)	Ya (S1)	Tidak (S2)	Entropy	Gain
node 2.1		7	3	4	0,985228136	
Jenis tinggal	Bersama orang tua	6	3	3	1	0,128085279
	Kost	1	0	1	0	
	Jalan kaki	5	3	2	0,970950594	
Alat transportasi	Kendaraan pribadi	1	0	1	0	0,291691997
	Angkutan Umum	1	0	1	0	
	Buruh	0	0	0	0	
Pekerjaan Orang tua	Karyawan	0	0	0	0	0,198117421
	PNS	0	0	0	0	
	Wiraswasta	1	1	0	0	
	Tidak bekerja	6	2	4	0,918295834	
	Lainnya	0	0	0	0	
Penerima KPS	Terima	2	2	0	0	0,469565211
	Tidak	5	1	4	0,721928095	

Pada Tabel 4 diperoleh nilai atribut dengan nilai *gain* terbesar yaitu Penerima KPS dengan nilai sebesar 0,469565211. Untuk kategori dengan penghasilan orang tua tinggi menghasilkan keputusan Tidak. Untuk kategori penghasilan orang tua sedang menghasilkan 25 keputusan Ya dan 45 menghasilkan keputusan Tidak. Pada kategori penghasilan orang tua rendah menghasilkan 3 keputusan Ya dan 4 keputusan Tidak sehingga dilakukan perhitungan lebih lanjut. Pada perhitungan *node* 2 menghasilkan pohon keputusan yang dapat dilihat pada Gambar 3 berikut.



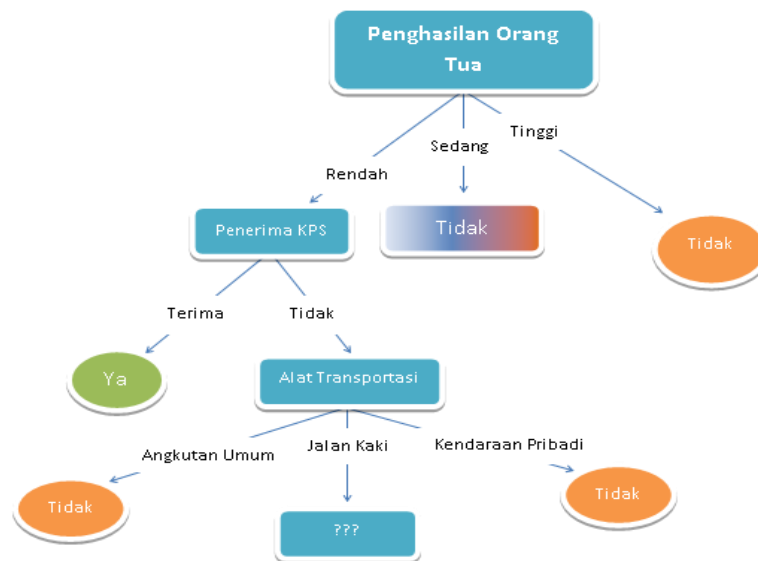
Gambar 3. Pohon Keputusan Node 2.1

Pada Gambar 3 dapat dilihat bahwa sub atribut Rendah sudah mendapatkan keputusan yaitu penerima KPS. Untuk sub atribut Terima mendapat keputusan Ya. Pada sub atribut Tidak belum mendapatkan keputusan, Maka perlu dilakukan perhitungan lebih lanjut untuk menentukan *node* akar lebih lanjut. Untuk tabel perhitungan selanjutnya dapat dilihat pada Tabel 5 berikut.

Tabel 5. Perhitungan Node 2.2

	Rendah	Penghasilan Orang tua (S)	Ya (S1)	Tidak (S2)	Entropy	Gain
node 2.2		5	1	4	0,721928095	
Jenis tinggal	Bersama orang tua	4	1	3	0,811278124	0,072905595
	Kost	1	0	1	0	
	Jalan kaki	3	1	2	0,918295834	
Alat transportasi	Kendaraan pribadi	1	0	1	0	0,170950594
	Angkutan Umum	1	0	1	0	
	Buruh	0	0	0	0	
Pekerjaan Orang tua	Karyawan	0	0	0	0	0
	PNS	0	0	0	0	
	Wiraswasta	0	0	0	0	
	Tidak bekerja	5	1	4	0,721928095	
	Lainnya	0	0	0	0	

Pada tabel 5 diperoleh nilai atribut dengan nilai *gain* terbesar yaitu Alat Transportasi dengan nilai sebesar 0,170950594. Untuk kategori dengan Alat Transportasi kendaraan pribadi menghasilkan keputusan 1 Tidak. Untuk kategori Alat Transportasi angkutan umum menghasilkan 1 keputusan Tidak. Sedangkan Pada kategori Alat Transportasi jalan kaki menghasilkan 1 keputusan Ya dan 2 keputusan Tidak sehingga dilakukan perhitungan lebih lanjut. Pada perhitungan *node* 2.2 menghasilkan pohon keputusan yang dapat dilihat pada Gambar 4 berikut.



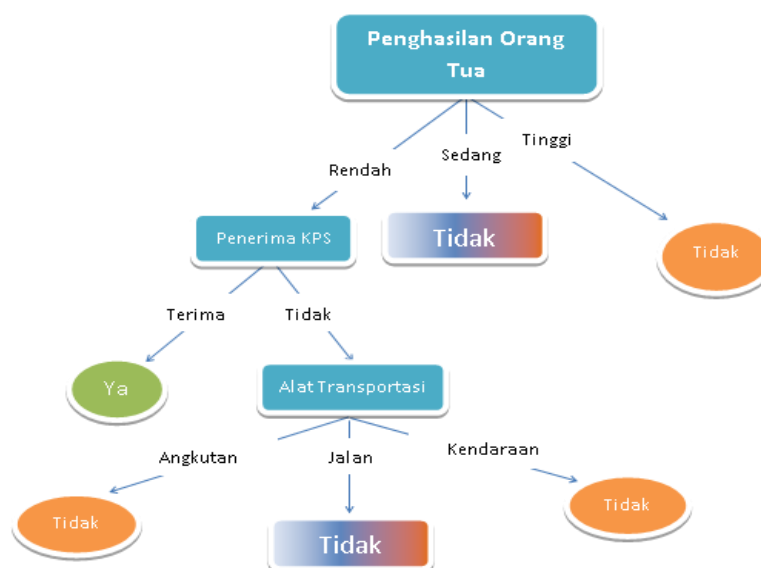
Gambar 4. Pohon Keputusan Node 2.2

Pada Gambar 4 dapat dilihat bahwa sub atribut Tidak sudah mendapatkan keputusan yaitu Alat *Transportasi*. Untuk sub atribut Angkutan Umum mendapat keputusan Tidak. Pada sub atribut Kendaraan Pribadi mendapat keputusan Tidak. Pada sub atribut Jalan Kaki belum mendapatkan keputusan, Maka perlu dilakukan perhitungan lebih lanjut untuk menentukan node akar lebih lanjut. Untung tabel perhitungan selanjutnya dapat dilihat pada Tabel 6 berikut.

Tabel 6. Perhitungan Node 2.3

	Rendah	Penghasilan Orang tua (S)	Ya (S1)	Tidak (S2)	Entropy	Gain
node 2.3		3	1	2	0,918295834	
Jenis tinggal	Bersama orang tua	3	1	2	0,918295834	0
	Kost	0	0	0	0	
	Buruh	0	0	0	0	
	Karyawan	0	0	0	0	
Pekerjaan Orang tua	PNS	0	0	0	0	0,918295834
	Wiraswasta	0	0	0	0	
	Tidak bekerja	3	1	2	0,918295834	
	Lainnya	0	0	0	0	

Pada Tabel 6 dapat dijelaskan bahwa hasil dari perhitungan sub atribut Jalan Kaki tidak memperoleh *gain* terbesar melainkan menghasilkan nilai 0 pada setiap *gain* Jenis Tinggal dan Pekerjaan Orang Tua sehingga tidak perlu dilakukan perhitungan lebih lanjut karena menghasilkan nilai 0. Pada perhitungan *node* 2.3 menghasilkan pohon keputusan sebagai berikut. Hasil akhir pohon keputusan dapat dilihat pada Gambar 5 berikut.

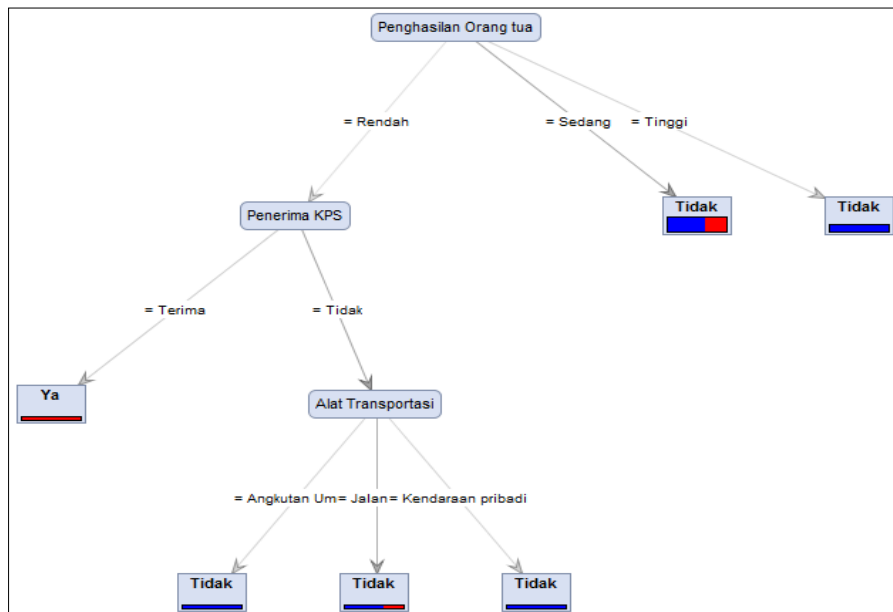


Gambar 5. Pohon Keputusan Node 2.3 (Hasil Akhir Pohon Keputusan)

3.2. Pengujian dengan Rapid Miner

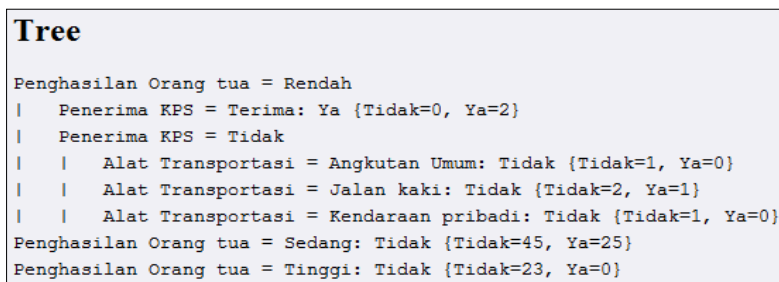
Langkah-langkah yang dilakukan untuk menguji perhitungan manual dengan aplikasi Rapid Miner antara lain:

- Langkah 1 : Membuka software Rapid Miner
- Langkah 2 : Pilih New Process untuk membuat proses baru dan melakukan pengujian data
- Langkah 3 : Menggunakan operator Read Excel untuk meng-input data yang akan diuji. Untuk membantu operator kita dapat mencarinya dengan mengetikkan “nama_operator” yang dicari pada kolom [filter] di bagian Operators. Selanjutnya drag and drop operator Read Excel ke lembar kerja main process
- Langkah 4 : Proses meng-input data pada excel dengan cara meng-klik tombol Import Configuration Wizard
- Langkah 5 : Selanjutnya pilih file name data yang akan digunakan, lalu klik Next. Lakukan import data.
- Langkah 6 : Data yang sudah berhasil di *import* kemudian diolah kedalam *Decision Tree* dengan mencari di kolom filter *Decision Tree* lalu drag and drop operator *Decision Tree* kedalam lembar main process kemudian hubungkan semua. Hasil akhir pohon keputusan dengan menggunakan aplikasi Rapid Miner dapat dilihat pada gambar 6 berikut.



Gambar 6. Hasil Akhir Pohon Keputusan dengan Rapid Miner

Gambar 6 merupakan pohon keputusan yang dihasilkan pada Rapid Miner dengan aturan atau *Rule* yang dapat dilihat pada text view pada Gambar 7 berikut.



Gambar 7. Hasil Rule Decision Tree

Hasil proses yang dilakukan peneliti pada perhitungan Algoritma C4.5 serta RapidMiner diperoleh dari hasil yang sama serta cocok. Sehingga pengujian dengan RapidMiner bisa dikatakan sukses serta bisa mendapatkan pohon keputusan pada permasalahan siswa yang layak untuk diajukan sebagai penerima PIP di SMA Negeri 2 Pematangsiantar.

4. KESIMPULAN

Berdasarkan hasil penelitian, pemanfaatan algoritma C4.5 dalam klasifikasi siswa yang layak diajukan untuk memperoleh Program Indonesia Pintar (PIP) menunjukkan efektivitas. Algoritma C4.5 dapat mengidentifikasi kriteria dan variabel yang signifikan dalam menentukan kelompok siswa yang berhak mendapatkan bantuan PIP. Berdasarkan penggunaan algoritma ini, seleksi siswa penerima PIP dapat dilakukan secara lebih objektif dan efisien. Kesimpulan ini dapat mendukung pengambilan keputusan yang tepat dalam alokasi bantuan pendidikan kepada siswa yang membutuhkan, memastikan bahwa bantuan

tersebut tepat sasaran. Implementasi algoritma C4.5 dalam proses klasifikasi siswa dapat menjadi landasan untuk perbaikan dan optimalisasi sistem penyaluran bantuan pendidikan di berbagai lembaga pendidikan. Sehingga dengan demikian, penelitian ini memberikan kontribusi positif terhadap pengembangan metode seleksi penerima bantuan pendidikan yang lebih efisien dan akurat.

REFERENSI

- [1] N. Arminarahmah, A. D. GS, G. W. Bhawika, M. P. Dewi, and A. Wanto, 'Mapping the Spread of Covid-19 in Asia Using Data Mining X-Means Algorithms', *IOP Conference Series: Materials Science and Engineering*, vol. 1071, no. 1, p. 012018, 2021, doi: 10.1088/1757-899x/1071/1/012018.
- [2] T. H. Sinaga, A. Wanto, I. Gunawan, S. Sumarno, and Z. M. Nasution, 'Implementation of Data Mining Using C4.5 Algorithm on Customer Satisfaction in Tirta Lihou PDAM', *Journal of Computer Networks, Architecture, and High-Performance Computing*, vol. 3, no. 1, pp. 9–20, 2021, doi: 10.47709/cnahpc.v3i1.923.
- [3] A. Pradipta, D. Hartama, A. Wanto, S. Saifullah, and J. Jalaluddin, 'The Application of Data Mining in Determining Timely Graduation Using the C45 Algorithm', *IJISTECH (International Journal of Information System & Technology)*, vol. 3, no. 1, pp. 31–36, 2019, doi: 10.30645/ijistech.v3i1.30.
- [4] N. A. Febriyati, A. D. Gs, and A. Wanto, 'GRDP Growth Rate Clustering in Surabaya City uses the K- Means Algorithm', *International Journal of Information System & Technology*, vol. 3, no. 2, pp. 276–283, 2020, doi: 10.30645/ijistech.v3i2.60.
- [5] I. Parlina *et al.*, 'Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category', in *Journal of Physics: Conference Series*, Aug. 2019, p. 012031. doi: 10.1088/1742-6596/1255/1/012031.
- [6] S. R. Ningsih, R. Wulansari, D. Hartama, A. P. Windarto, and A. Wanto, 'Analysis of PROMETHEE II Method on Selection of Lecturer Community Service Grant Proposals', in *Journal of Physics: Conference Series*, Aug. 2019, p. 012004. doi: 10.1088/1742-6596/1255/1/012004.
- [7] P. Alkhairi, L. P. Purba, A. Eryzha, A. P. Windarto, and A. Wanto, 'The Analysis of the ELECTREE II Algorithm in Determining the Doubts of the Community Doing Business Online', in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Sep. 2019, p. 012010. doi: 10.1088/1742-6596/1255/1/012010.
- [8] R. Watrianthos, W. A. Ritonga, A. Rengganis, A. Wanto, and M. Isa Indrawan, 'Implementation of PROMETHEE-GAIA Method for Lecturer Performance Evaluation', *Journal of Physics: Conference Series*, vol. 1933, no. 1, p. 012067, 2021, doi: 10.1088/1742-6596/1933/1/012067.
- [9] S. Sundari, A. Wanto, Saifullah, and I. Gunawan, 'Sistem Pendukung Keputusan Dengan Menggunakan Metode Electre Dalam Merekomendasikan Dosen Berprestasi Bidang Ilmu Komputer (Study Kasus di AMIK & STIKOM Tunas Bangsa)', in *Seminar Nasional Multi Disiplin Ilmu*, 2017, pp. 1–6. doi: 10.17605/OSF.IO/4TWG6.
- [10] M. Widyasuti, A. Wanto, D. Hartama, and E. Purwanto, 'Rekomendasi Penjualan Aksesoris Handphone Menggunakan Metode Analytical Hierarchy Process (AHP)', *Konferensi Nasional Teknologi Informasi dan Komputer (KOMIK)*, vol. 1, no. 1, pp. 27–32, 2017.
- [11] K. Fatmawati *et al.*, 'Analysis of Promethee II Method in the Selection of the Best Formula for Infants Under Three Years', *Journal of Physics: Conference Series*, vol. 1255, no. 1, p. 012009, Aug. 2019, doi: 10.1088/1742-6596/1255/1/012009.
- [12] I. A. R. Simbolon, F. Yatussa'ada, and A. Wanto, 'Penerapan Algoritma Backpropagation dalam Memprediksi Persentase Penduduk Buta Huruf di Indonesia', *Jurnal Informatika Upgris*, vol. 4, no. 2, pp. 163–169, 2018, doi: 10.26877/jiu.v4i2.2423.
- [13] W. Saputra, J. T. Hardinata, and A. Wanto, 'Resilient method in determining the best architectural model for predicting open unemployment in Indonesia', *IOP Conference Series: Materials Science and Engineering*, vol. 725, no. 1, p. 012115, Jan. 2020, doi: 10.1088/1757-899X/725/1/012115.
- [14] E. Hartato, D. Sitorus, and A. Wanto, 'Analisis Jaringan Saraf Tiruan Untuk Prediksi Luas Panen Biofarmaka Di Indonesia', *SemanTIK*, vol. 4, no. 1, pp. 49–56, 2018, doi: 10.55679/semantik.v4i1.4201.
- [15] B. K. Sihotang and A. Wanto, 'Analisis Jaringan Syaraf Tiruan Dalam Memprediksi Jumlah Tamu Pada Hotel Non Bintang', *Jurnal Teknologi Informasi Techno*, vol. 17, no. 4, pp. 333–346, 2018, doi: 10.33633/tc.v17i4.1762.
- [16] I. S. Purba and A. Wanto, 'Prediksi Jumlah Nilai Impor Sumatera Utara Menurut Negara Asal Menggunakan Algoritma Backpropagation', *Jurnal Teknologi Informasi Techno*, vol. 17, no. 3, pp. 302–311, 2018, doi: 10.33633/tc.v17i3.1769.
- [17] A. Wanto, S. Defit, and A. P. Windarto, 'Algoritma Fungsi Pelatihan pada Machine Learning berbasis ANN untuk Peramalan Fenomena Bencana', *RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, pp. 254–264, 2021, doi: 10.29207/resti.v5i2.3031.
- [18] A. Wanto and J. T. Hardinata, 'Estimations of Indonesian poor people as poverty reduction efforts facing industrial revolution 4.0', *IOP Conference Series: Materials Science and Engineering*, vol. 725, no. 1, p. 012114, Jan. 2020, doi: 10.1088/1757-899X/725/1/012114.
- [19] Yadi Yadi, 'Implementation Algorithm C4.5 Classification Of Prospective Scholarship Recipients', *Jurnal SimanteC*, vol. 11, no. 1, pp. 27–32, 2022, doi: 10.21107/simantec.v11i1.14289.
- [20] N. I. Nurhidayati, Y. Yahya, F. Fathurrahman, L. . Samsu, and W. Amnia, 'Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi)', *Infotek: Jurnal Informatika dan Teknologi*, vol. 6, no. 1, pp. 177–188, 2023, doi: 10.29408/jit.v6i1.7529.
- [21] 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 Elektro dan Informatika*, vol. 5, no. 1, pp. 96–109, 2022, doi: 10.55542/jurtie.v5i1.452.
- [22] S. P. Sipayung, T. P. Sihalo, A. Purba, and J. R. Tarigan, 'Analisa Algoritma C.45 Terhadap penentuan Rekomendasi Penerima Beasiswa SMP Swasta Methodist-8 Medan', *LOFIAN: Jurnal Teknologi Informasi dan Komunikasi*, vol. 3, no. 1, pp. 20–24, 2023, doi: 10.58918/lofian.v3i1.217.