

Klasifikasi Algoritma C4.5 untuk Penentuan Penerima Program Indonesia Pintar pada MIS Al-Khoirot

C4.5 Algorithm Classification for Determining Smart Indonesia Program Recipients at MIS Al-Khoirot

Weni Ratna Sari Oktapia Ningse¹, S Sumarno², Zulaini Masruro Nasution³

^{1,2,3} STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Article Info

Genesis Artikel:

Diterima, 3 Maret 2022

Direvisi, 13 Maret 2022

Disetujui, 14 Maret 2022

Kata Kunci:

PIP

Data Mining

Algoritma C4.5

Klasifikasi

MIS Al-Khoirot

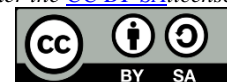
ABSTRAK

Tujuan penelitian adalah untuk membantu pihak sekolah dalam menyeleksi data siswa sebagai penerima PIP (Program Indonesia Pintar) menjadi lebih objektif dan praktis serta dapat membantu dalam meningkatkan ketepatan sasaran dari penerima bantuan dana PIP tersebut. Pada penelitian ini menggunakan teknik Data Mining dengan menggunakan algoritma C4.5. Sumber data penelitian yang digunakan di peroleh dari observasi dan wawancara kepada pihak sekolah MIS Al-Khoirot Tambun Nabolon Pematang Siantar. Adapun variabel penelitian yang digunakan adalah pekerjaan orang tua, penghasilan orang tua, pemegang KKS (Kartu Keluarga Sejahtera), pemegang SKTM (Surat Keterangan Miskin). Pada penelitian ini alternatif yang digunakan sebagai sampel adalah data siswa MIS Al-Khoirot. Hasil dari penelitian ini didapatkan bahwa atribut yang paling dominan adalah pemegang SKTM dengan gain sebesar 0,833764907, selain itu penelitian ini menghasilkan 8 (delapan) rules dengan tingkat akurasi sebesar 98,00%. Berdasarkan hal ini dapat disimpulkan bahwa algoritma C4.5 dapat digunakan untuk klasifikasi Penentuan Penerima Program Indonesia Pintar pada MIS Al-Khoirot.

ABSTRACT

The purpose of the research is to assist the school in selecting student data as recipients of the PIP (Smart Indonesia Program) to be more objective and practical and to assist in increasing the accuracy of the targeting of the recipients of the PIP funds. In this study using Data Mining techniques using the C4.5 algorithm. The source of the research data used was obtained from observations and interviews with the MIS Al-Khoirot Tambun Nabolon Pematang Siantar school. The research variables used were parents' occupations, parents' income, KKS (Prosperous Family Card) holders, SKTM holders (Poor Certificate). In this study, the alternative used as a sample is the data of MIS Al-Khoirot students. The results of this study found that the most dominant attribute was the SKTM holder with a gain of 0.833764907, besides that this study produced 8 (eight) rules with an accuracy rate of 98.00%. Based on this, it can be concluded that the C4.5 algorithm can be used for the classification of the Determination of Smart Indonesia Program Recipients at MIS Al-Khoirot.

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



Penulis Korespondensi:

Weni Ratna Sari Oktapia Ningse,
Program Studi Teknik Informatika,
STIKOM Tunas Bangsa, Pematangsiantar, indonesia
Email: weni.oktapia21@gmail.com

1. PENDAHULUAN

Program Indonesia Pintar merupakan salah satu cara yang dilakukan oleh pemerintah negara Indonesia untuk meminimalisir angka putus sekolah seperti hal nya yang tertuang dalam instruksi Presiden Republik Indonesia Nomor 7 tahun 2014 mengenai pelaksanaan program keluarga produktif melalui Program Indonesia Pintar (PIP) yang bertujuan untuk mencegah siswa/i dari kemungkinan untuk putus sekolah/drop out (Kemendikbud, 2018). Namun dalam pelaksanaannya, masih banyak Penerima dana PIP yang masih tidak tepat sasaran. Hal ini dapat dilihat berdasarkan hasil survey yang dilakukan oleh Indonesia *Corruption Watch* pada tahun 2018 menyimpulkan bahwa data yang digunakan untuk PIP masih belum akurat dikarenakan hampir setengah warga miskin yaitu sekitar 41,9% tidak terdaftar sebagai peserta PIP (ICW, 2018). Beberapa hasil dari evaluasi dan studi berlanjut terhadap pelaksanaan PIP menunjukkan kelemahan dari program ini, yaitu terkait ketepatan penetapan sasaran penerima dana PIP dimana di temukan masih banyaknya rumah tangga tidak miskin yang menerima bantuan dana PIP dan jumlah beasiswa yang kurang memadai.

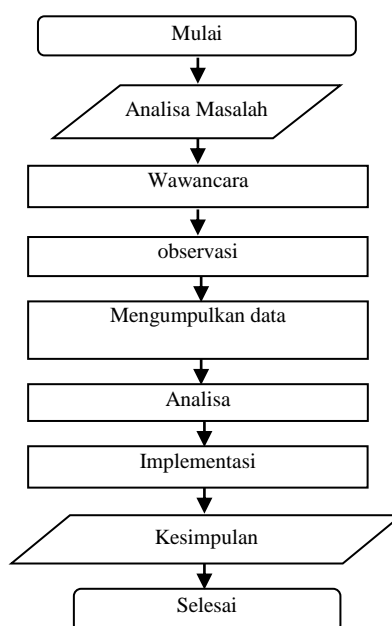
Sasaran penerima bantuan dana PIP masih lemah dimana ditemukan banyak penerima bantuan dana PIP yang bukan berasal dari keluarga/rumah tangga miskin dan banyak siswa dari keluarga /rumah tangga miskin tidak menerima manfaat bantuan dana PIP, serta kendala waktu dalam menentukan penerima PIP di MIS Al-Khoirot yaitu waktu yang dibutuhkan dari mulai proses seleksi data, proses seleksi dilakukan dengan cara mengamati data-data siswa mulai proses akan diseleksi siapa yang dapat menerima bantuan berdasarkan kriteria yang sudah ditentukan oleh pemerintah. Sehingga proses penyeleksian dalam penentuan Penerima PIP tidak terjadi secara sistematis dan terkadang tidak tepat sasaran. Banyak cabang ilmu komputer yang dapat menyelesaikan permasalahan diantaranya: Sistem Pendukung Keputusan [1]–[7], Jaringan Saraf Tiruan [8]–[15], maupun Data mining [16]–[25]. Dalam merekomendasikan penerima bantuan dana PIP tersebut, jadi memerlukan waktu yang cukup lama dan sering terjadi kesalahan. Oleh karena itu, agar proses rekomendasi penerima bantuan dana Program Indonesia Pintar (PIP) di MIS Al-Khoirot menjadi lebih objektif dan praktis, sebaiknya dilakukan secara komputerisasi, yaitu dengan mengembangkan suatu sistem yang mengimplementasikan metode yang dapat mempermudah proses dalam pengambilan keputusan serta dapat membantu dalam meningkatkan ketepatan sasaran dari penerima bantuan dana PIP tersebut. Beberapa penelitian terdahulu yang dijadikan penulis bahan refrensi untuk melakukan penelitian dalam bidang *Data mining* [26]–[28].

Penelitian ini di lakukan pada sekolah MIS Al-Khoirot Tambun Nabolon Pematangsiantar yang sekarang ini mempunyai siswa sebanyak 198 orang yang terbagi dalam 6 kelas. Penelitian ini bertujuan untuk menganalisis penerima Program Indonesia Pintar (PIP) dengan memanfaatkan teknik *data mining*, menggunakan algoritma C4.5, sehingga membantu pihak sekolah dalam mencari solusi dan dapat mengetahui tingkat kelayakan siswa yang akan mendapatkan PIP.

2. METODE PENELITIAN

2.1. Rancangan Penelitian

Rancangan penelitian merupakan beberapa rancangan atau alur dari Analisis Klasifikasi C4.5 pada kelayakan data siswa sebagai penerima Program Indonesia Pintar (PIP) MIS Al-Khoirot Jl. Basuh Tangan yang dimulai dari pengambilan data hingga pengolahan data yang akan dibuat sehingga outputnya lebih jelas, adapun rancangan ini dapat dilihat dalam rancangan flowchart sebagai berikut pada gambar 1 berikut.



Gambar 1. Rancangan Penelitian

3. HASIL DAN ANALISIS

3.1. Analisis Data

Dalam penelitian ini data yang digunakan akan diolah dari hasil observasi yang diberikan oleh operator MIS Al-Khoirot yang di ambil dari database pendidikan yang sudah ada dengan jumlah siswa sebanyak 150 siswa. Penulis melakukan observasi langsung kepada pihak sekolah dengan jumlah 50 sampel data. Berikut tabel pengambilan data siswa MIS Al-Khoirot yang digunakan.

Tabel 1. Pengambilan Sampel Data Siswa

Siswa	Jumlah Siswa MIS Al-Khoirot	Data Sampel Yang di Gunakan
MIS Al-Khoirot	150 siswa	50 siswa
Jumlah	150	50

Tabel 2. Data Penelitian Normalisasi

No	Pekerjaan Orang Tua	Penghasilan Orang Tua	Pemegang KKS	Pemegang SKTM	Keterangan
1	2	2	1	1	1
2	2	2	2	2	1
3	2	2	2	1	1
4	2	2	1	2	1
5	2	2	2	1	1
6	2	2	2	2	2
7	3	2	1	2	1
8	2	2	2	1	1
9	2	2	2	2	2
...
91	2	2	2	2	2
92	3	2	1	2	1
93	1	1	2	2	2
94	2	3	1	2	1
95	3	2	2	1	1
96	1	1	2	2	2
97	2	3	1	2	1
98	3	2	2	1	1
99	3	3	1	2	1
100	2	2	2	2	2

Data yang digunakan terdiri dari :

1. Penghasilan orang tua (C1)
2. Pekerjaan orang tua (C2)
3. Pemegang KKS (C3)
4. Pemegang SKTM (C4)

Setiap variabel terdiri dari kriteria yang telah ditentukan oleh pihak sekolah sebagai syarat penerima PIP. Data yang digunakan merupakan jenis data *deskriptif*. Hasil dari atribut yang didapatkan dari wawancara dapat dijelaskan pada tabel berikut :

Tabel 3. Penjelasan Atribut

No	Atribut	Penjelasan
1	Penghasilan orang tua	1. Tinggi (>3 s/d 5jt) 2. Sedang (2 jt s/d 3 jt) 3. Rendah (1 jt s/d 500 Ribu)
2	Pekerjaan orang tua	1. Peg.Negeri (TNI, Polisi, Guru, Dosen, Dokter, Bidan, Perawat) 2. Peg.Swasta non.formal (Pengusaha/wiraswasta, pedagang, petani, peternak, nelayan, buruh, sopir, seniman/pelukis, artis/sejenisnya) 3. Peg.swasta formal (Pegawai swasta, pensiunan, pilot/ pramugari)
3	Pemegang KKS	1. Ya 2. Tidak
4	Pemegang SKTM	1. Ya 2. Tidak
5	Keterangan	1. Ya 2. Tidak

3.2. Pengolahan Data Menggunakan Algoritma C4.5

Perhitungan algoritma C4.5 dimulai dengan memilih atribut akar terlebih dahulu dengan mencari jumlah kasus keseluruhan, jumlah kasus keputusan menerima PIP dan tidak menerima PIP. Menghitung *entropy* dari semua kasus yang dibagi berdasarkan pekerjaan orang tua (C1), penghasilan orang tua (C2), pemegang kks (kartu keluarga sejahtera) (C3), pemegang sktm (surat ketetapan miskin) (C4). Setelah itu dilakukan perhitungan gain untuk masing-masing atribut. Proses pertama algoritma C4.5 adalah menentukan nilai *entropy*. Langkah pertama, tentukan *entropy* total kasus terlebih dahulu. Rumus untuk mencari *entropy* dari data siswa tersebut yaitu :

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

Keterangan :

S : Himpunan Kasus

pi : Proporsi dari Si terhadap S

n : Jumlah Partisi S

Maka :

$$Entropy(Total) = \left(-\frac{\text{sum}(ya)}{\text{total}} * \log_2 \left(\frac{\text{sum}(ya)}{\text{total}} \right) \right) + \left(-\frac{\text{sum}(tidak)}{\text{total}} * \log_2 \left(\frac{\text{sum}(tidak)}{\text{total}} \right) \right)$$

$$Entropy(Total) = \left(-\frac{35}{50} * \log_2 \left(\frac{35}{50} \right) \right) - \left(-\frac{15}{50} * \log_2 \left(\frac{15}{50} \right) \right) = 0,881290899$$

Tabel 4. Entropy Total Kasus

Total Kasus	Jumlah (Ya)	Jumlah (Tidak)	Entropy
50	35	15	0,881290899

3.3. Perhitungan Mencari Gain

Kemudian menghitung *Entropy* dari semua kasus yang terbagi berdasarkan atribut “penghasilan orang tua”, “Pekerjaan orang tua”, “alat transportasi dari rumah ke sekolah”, “status penerima KKS/KPS” seperti berikut.

$$Entropy(S) = Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Entropy (pekerjaan orang tua, peg.negeri)

$$= \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3} \right) \right) + \left(-\frac{3}{3} * \log_2 \left(\frac{3}{3} \right) \right) = 0$$

Entropy (pekerjaan orang tua, peg.swasta non formal)

$$= \left(-\frac{25}{34} * \log_2 \left(\frac{25}{34} \right) \right) + \left(-\frac{9}{34} * \log_2 \left(\frac{9}{34} \right) \right) = 0,833764907$$

Entropy (pekerjaan orang tua, peg.swasta formal)

$$= \left(-\frac{9}{13} * \log_2 \left(\frac{9}{13} \right) \right) + \left(-\frac{4}{13} * \log_2 \left(\frac{4}{13} \right) \right) = 0,89049164$$

Entropy (penghasilan orang tua, tinggi)

$$= \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5} \right) \right) + \left(-\frac{4}{5} * \log_2 \left(\frac{4}{5} \right) \right) = 0,721928095$$

Entropy (penghasilan orang tua, sedang)

$$= \left(-\frac{24}{34} * \log_2 \left(\frac{24}{34} \right) \right) + \left(-\frac{10}{34} * \log_2 \left(\frac{10}{34} \right) \right) = 0,873981048$$

Entropy (penghasilan orang tua, rendah)

$$= \left(-\frac{9}{11} * \log_2 \left(\frac{9}{11} \right) \right) + \left(-\frac{2}{11} * \log_2 \left(\frac{2}{11} \right) \right) = 0,684038436$$

Entropy (pemegang kks, ya)

$$= \left(-\frac{15}{15} * \log_2 \left(\frac{15}{15} \right) \right) + \left(-\frac{0}{15} * \log_2 \left(\frac{0}{15} \right) \right) = 0$$

Entropy (pemegang kks, tidak)

$$= \left(-\frac{19}{35} * \log_2 \left(\frac{19}{35} \right) \right) + \left(-\frac{16}{35} * \log_2 \left(\frac{16}{35} \right) \right) = 0,994693795$$

Entropy (pemegang sktm, ya)

$$= \left(-\frac{18}{18} * \log_2 \left(\frac{18}{18} \right) \right) + \left(-\frac{0}{18} * \log_2 \left(\frac{0}{18} \right) \right) = 0$$

Entropy (pemegang sktm, tidak)

$$= \left(-\frac{16}{32} * \log_2 \left(\frac{16}{32} \right) \right) + \left(-\frac{16}{32} * \log_2 \left(\frac{16}{32} \right) \right) = 1$$

Setelah mendapatkan hasil dari semua *entropy* langkah selanjutnya ialah menghitung gain dari setiap atribut.

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

Gain (total, pekerjaan orang tua)

$$= 0,881290899 - \left(\frac{3}{50} * 0 \right) + \left(\frac{34}{50} * 0,833764907 \right) + \left(\frac{13}{50} * 0,89049164 \right) \\ = 0,098343699$$

Gain (total, penghasilan orang tua)

$$= 0,881290899 - \left(\frac{5}{50} * 0,721928095 \right) + \left(\frac{34}{50} * 0,873981048 \right) + \\ \left(\frac{11}{50} * 0,684038436 \right) = 0,065818108$$

Gain (total, pemegang kks)

$$= 0,881290899 - \left(\frac{15}{50} * 0 \right) + \left(\frac{35}{50} * 0,994693795 \right) = 0,005777324$$

Gain (total, pemegang sktm)

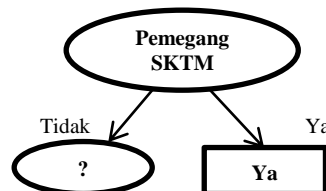
$$= 0,881290899 - \left(\frac{18}{50} * 0 \right) + \left(\frac{35}{50} * 0,994693795 \right) = 0,801715396$$

Berikut ini hasil dari perhitungan di atas :

Tabel 5. Hasil Perhitungan *Node 1*

Atribut	Jml Kasus	Ya (S1)	Tidak (S2)	Entropy	Gain
TOTAL	50	35	15	0,881290899	
Pekerjaan Orang Tua					0,098343699
Peg.negeri	3	0	3	0	
Peg.swasta non formal	34	25	9	0,833764907	
Peg.swasta formal	13	9	4	0,89049164	
Penghasilan Orang Tua					0,065818108
Tinggi	5	1	4	0,721928095	
Sedang	34	24	10	0,873981048	
Rendah	11	9	2	0,684038436	
Pemegang KKS					0,801715396
Ya	15	15	0	0	
Tidak	35	19	16	0,994693795	
Pemegang SKTM					0,833764907
Ya	18	18	0	0	
Tidak	32	16	16	1	

Pada tabel di atas dapat kita lihat bahwa atribut pemegang sktm memiliki *again* tertinggi yaitu 0,833764907. Maka pemegang sktm menjadi *node* akar. Pemegang sktm memiliki 3 nilai yaitu pekerjaan orang tua, penghasilan orang tua dan pemegang kks. Pada baik dan buruk telah mengklasifikasi kasus menjadi 1 yaitu keputusan “ya”, sedangkan pemegang sktm masih diperlukan perhitungan lagi karena masih memiliki hasil antara “ya” dan “tidak”, perhitungan tersebut dilakukan untuk menentukan *node* akar selanjutnya, maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut.



Gambar 2. Pohon Keputusan 1

Langkah selanjutnya menghitung *node* akar pemegang sktm, caranya sebagai berikut : Pilih *node* akar pemegang sktm kemudian hitung jumlah kasus, jumlah kasus keputusan “ya”, jumlah kasus keputusan “tidak” dan *entropy* pada atribut pekerjaan orang tua, penghasilan orang tua, dan pemegang kks seperti dibawah.

Entropy (pemegang sktm = tidak, pekerjaan orang tua : peg.negeri)

$$= \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) - \left(-\frac{3}{3} * \log_2 \left(\frac{3}{3}\right)\right) = 0$$

Entropy (pemegang sktm = tidak, pekerjaan orang tua : peg.swasta non formal)

$$= \left(-\frac{12}{22} * \log_2 \left(\frac{12}{22}\right)\right) - \left(-\frac{10}{22} * \log_2 \left(\frac{10}{22}\right)\right) = 0,994030211$$

Entropy (pemegang sktm = tidak, pekerjaan orang tua : peg.swasta formal)

$$= \left(-\frac{4}{7} * \log_2 \left(\frac{4}{7}\right)\right) - \left(-\frac{3}{7} * \log_2 \left(\frac{3}{7}\right)\right) = 0,985228136$$

Entropy (pemegang sktm = tidak, penghasilan orang tua: tinggi)

$$= \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right) - \left(-\frac{4}{5} * \log_2 \left(\frac{4}{5}\right)\right) = 0,721928095$$

Entropy (pemegang sktm = tidak, penghasilan orang tua: sedang)

$$= \left(-\frac{10}{20} * \log_2 \left(\frac{10}{20}\right)\right) - \left(-\frac{10}{20} * \log_2 \left(\frac{10}{20}\right)\right) = 1$$

Entropy (pemegang sktm = tidak, penghasilan orang tua: rendah)

$$= \left(-\frac{5}{7} * \log_2 \left(\frac{5}{7}\right)\right) - \left(-\frac{2}{7} * \log_2 \left(\frac{2}{7}\right)\right) = 0,863120569$$

Entropy (pemegang sktm = tidak, pemegang kks: ya)

$$= \left(-\frac{15}{15} * \log_2 \left(\frac{15}{15}\right)\right) - \left(-\frac{0}{15} * \log_2 \left(\frac{0}{15}\right)\right) = 0$$

Entropy (pemegang sktm = tidak, pemegang kks: tidak)

$$= \left(-\frac{1}{17} * \log_2 \left(\frac{1}{17}\right)\right) - \left(-\frac{16}{17} * \log_2 \left(\frac{16}{17}\right)\right) = 0,985228136$$

Gain (pemegang sktm = tidak, pekerjaan orang tua)

$$= 1 - \left(\frac{2}{32} * 0\right) + \left(\frac{22}{32} * 0,994030211\right) + \left(\frac{7}{32} * 0,985228136\right) = 0,270639803$$

Gain (pemegang sktm = tidak, penghasilan orang tua)

$$= 1 - \left(\frac{5}{32} * 0,721928095\right) + \left(\frac{20}{32} * 1\right) + \left(\frac{7}{32} * 0,863120569\right) = 0,736804493$$

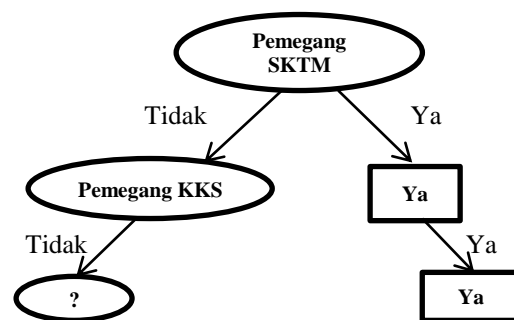
Gain (pemegang sktm = tidak, pemegang kks)

$$= 1 - \left(\frac{15}{32} * 0 \right) + \left(\frac{17}{32} * 0,322756959 \right) = 0,985228136$$

Tabel 6. Hasil Perhitungan *Node 2*

Atribut	Jml kasus	Ya (S1)	Tidak (S2)	Entropy	Gain
Pemegang sktm = tidak	32	16	16	1	0,270639803
Pekerjaan Orang Tua					
Peg.negeri	3	0	3	0	
Peg.swasta non formal	22	12	10	0,994030211	0,736804493
Peg.swasta formal	7	4	3	0,985228136	
Penghasilan Orang Tua					
Tinggi	5	1	4	0,721928095	0,985228136
Sedang	20	10	10	1	
Rendah	7	5	2	0,863120569	
Pemegang kks					
Ya	15	15	0	0	
Tidak	17	1	16	0,322756959	

Pada tabel diatas *gain* tertinggi pada pemegang sktm = tidak adalah pemegang kks = 0,985228136 maka menghasilkan pohon keputusan sebagai berikut :



Gambar 3. Pohon Keputusan 2

Langkah selanjutnya menghitung *node* akar pemegang kks = tidak, caranya sebagai berikut. Pilih *node* akar pemegang kks = tidak kemudian hitung jumlah kasus, jumlah kasus keputusan “ya”, jumlah kasus keputusan “tidak” dan *entropy* pada atribut pemegang kks = tidak, pekerjaan orang tua dan penghasilan orang tua seperti berikut.

Entropy (pemegang kks = tidak, pekerjaan orang tua; peg.negeri)

$$= \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3} \right) \right) - \left(-\frac{3}{3} * \log_2 \left(\frac{3}{3} \right) \right) = 0$$

Entropy (pemegang kks = tidak, pekerjaan orang tua; peg.swasta non formal)

$$= \left(-\frac{1}{10} * \log_2 \left(\frac{1}{10} \right) \right) - \left(-\frac{9}{10} * \log_2 \left(\frac{9}{10} \right) \right) = 0,468995594$$

Entropy (pemegang kks = tidak, pekerjaan orang tua; peg.swasta formal)

$$= \left(-\frac{0}{4} * \log_2 \left(\frac{0}{4} \right) \right) - \left(-\frac{4}{4} * \log_2 \left(\frac{4}{4} \right) \right) = 0$$

Entropy (pemegang kks = tidak, penghasilan orang tua: tinggi)

$$= \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5} \right) \right) - \left(-\frac{4}{5} * \log_2 \left(\frac{4}{5} \right) \right) = 0,721928095$$

Entropy (pemegang kks = tidak, penghasilan orang tua: sedang)

$$= \left(-\frac{0}{10} * \log_2 \left(\frac{0}{10} \right) \right) - \left(-\frac{10}{10} * \log_2 \left(\frac{10}{10} \right) \right) = 0$$

Entropy (pemegang kks = tidak, penghasilan orang tua: rendah)

$$= \left(-\frac{0}{2} * \log_2 \left(\frac{0}{2} \right) \right) - \left(-\frac{2}{2} * \log_2 \left(\frac{2}{2} \right) \right) = 0$$

Gain (pemegang kks = tidak, pekerjaan orang tua)

$$= 0,322756959 - \left(\frac{3}{17} * 0 \right) + \left(\frac{10}{17} * 0,468995594 \right) + \left(\frac{4}{17} * 0 \right) = 0,046877198$$

Gain (pemegang kks = tidak, penghasilan orang tua)

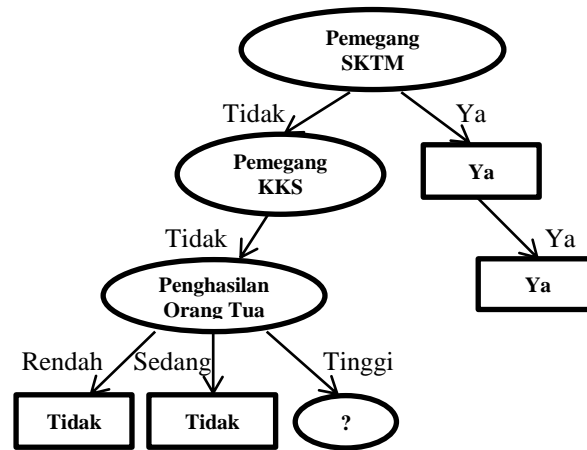
$$= 0,322756959 - \left(\frac{5}{17} * 0,721928095 \right) + \left(\frac{10}{17} * 0 \right) + \left(\frac{2}{17} * 0 \right) = 0,902410119$$

Tabel 7. Hasil Perhitungan *Node 3*

Atribut	Jml kasus	Ya (S1)	Tidak (S2)	Entropy	Gain
Pemegang kks = tidak	17	1	16	0,322756959	0,046877198
Pekerjaan Orang Tua					
Peg.negeri	3	0	3	0	
Peg.swasta non formal	10	1	9	0,468995594	

Atribut	Jml kasus	Ya	Tidak	Entropy	Gain
Peg.swasta formal	4	0	4	0	0,902410119
Penghasilan Orang Tua					
Tinggi	5	1	4	0,721928095	
Sedang	10	0	10	0	
Rendah	2	0	2	0	

Pada tabel diatas *gain* tertinggi pada pemegang kks = tidak adalah penghasilan orang tua = 0,902410119 maka menghasilkan pohon keputusan sebagai berikut :



Gambar 4. Pohon Keputusan 3

Langkah selanjutnya menghitung *node* akar tinggi caranya sebagai berikut: Pilih *node* akar tinggi pada penghasilan orang tua, karena tinggi masih bernilai “ya” dan “tidak” maka diperlukan perhitungan kembali.

Entropy (penghasilan orang tua = tinggi, pekerjaan orang tua : peg.negeri)

$$= \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) - \left(-\frac{3}{3} * \log_2 \left(\frac{3}{3}\right)\right) = 0$$

Entropy (penghasilan orang tua = tinggi, pekerjaan orang tua : peg.swasta non formal)

$$= \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2}\right)\right) - \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2}\right)\right) = 1$$

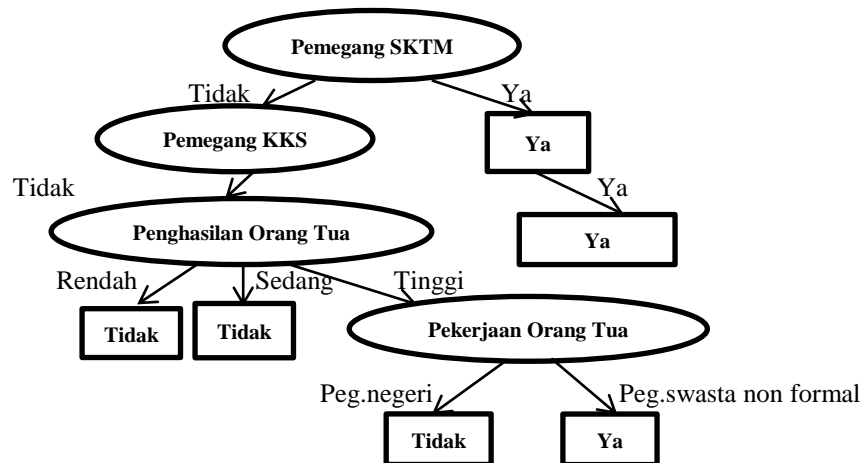
Gain (penghasilan orang tua = tinggi, pekerjaan orang tua)

$$= 0,721928095 - \left(\frac{3}{5} * 0\right) + \left(\frac{2}{5} * 1\right) = 0,321928095$$

Tabel 8. Hasil Perhitungan *Node* 4

Atribut	Jml kasus	Ya	Tidak	Entropy	Gain
Penghasilan Orang Tua = Tinggi	5	(S1) 1	(S2) 4	0,721928095	0,321928095
Pekerjaan Orang Tua					
Peg.negeri	3	0	3	0	
Peg.swasta non formal	2	1	1	1	

Pada tabel diatas *gain* tertinggi pada penghasilan orang tua=tinggi diketahui *gain* tertinggi = 0,321928095 maka menghasilkan pohon keputusan sebagai berikut.

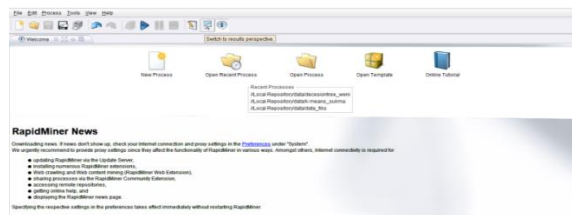


Gambar 5. Pohon Keputusan 4

3.4. Proses Pengujian Dengan *Rapidminer*

Berikut ini merupakan langkah-langkah penerapan algoritma C4.5 pada *RapidMiner* 5.3.

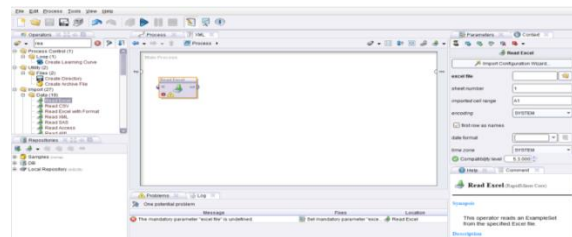
1. Jalankan aplikasi *Rapidminer* 5.3. lalu pilih *new process*



Gambar 6. Tampilan Utama *Rapidminer* 5.3

Gambar diatas adalah tampilan awal *Rapidminer Studio* 5.3, tahap awal ketika ingin memulai mengkoneksikan database ke *Rapidminer Studio* 5.3 terlebih dahulu ialah dengan membuat proses baru.

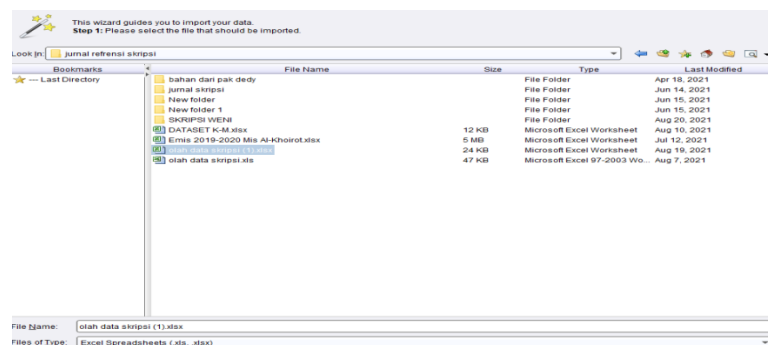
2. Pilih operator *Read Excel* di *Main Process* untuk memasukkan data berbentuk *excel*.



Gambar 7. Tampilan *Import Data Excel*

Gambar diatas adalah tampilan *panel* proses, langkah selanjutnya buka *Panel operators* cari *Read Excel*, karena tipe file yang digunakan dalam penelitian ini menggunakan format *Excel*, kemudian *drag* dan *drop* *read excel* tersebut kedalam panel proses.

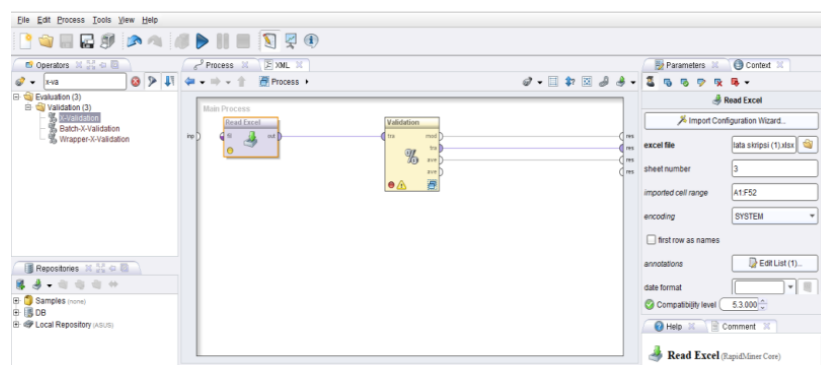
3. *Importing Data*



Gambar 8. *Importing Data*

Setelah data dipilih kemudian klik *next*, lalu seleksi data apa saja yang akan digunakan, pada penelitian ini data yang digunakan hanya penghasilan orang tua, pekerjaan orang tua, alat transportasi dari rumah ke sekolah dan status penerima KKS/KPS.

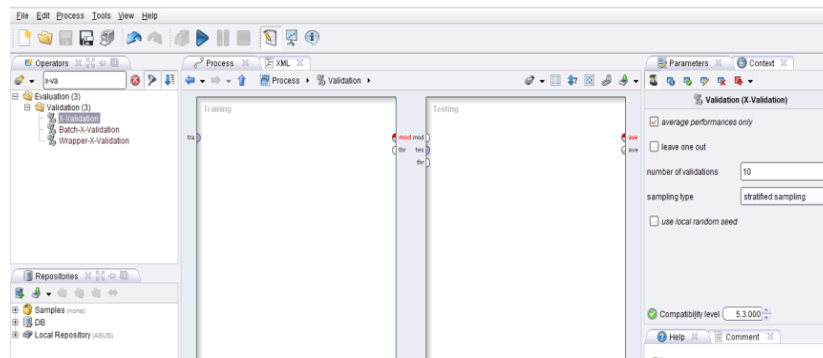
4. Menghubungkan *port read excel* dengan *port validation*



Gambar 9. Penghubungan *Port Read Excel* Dengan *Port Validation* Data

Langkah selanjutnya untuk melakukan *testing* ialah dengan *drag* dan *drop split validation* pada menu *operators* kedalam panel *process*, setelah *split validation* muncul maka lakukan penghubungan antara *read excel* dengan *split validation* seperti gambar di atas.

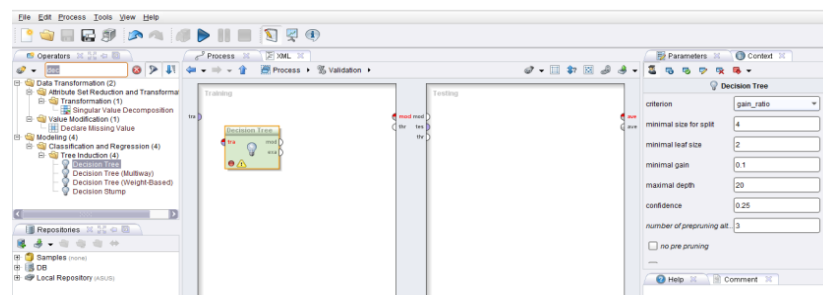
5. Tampilan panel *process validation*



Gambar 10. Panel *Process Validation*

Kemudian *double klik* pada *validation* untuk membuka tampilan sub *process training* dan *testing*.

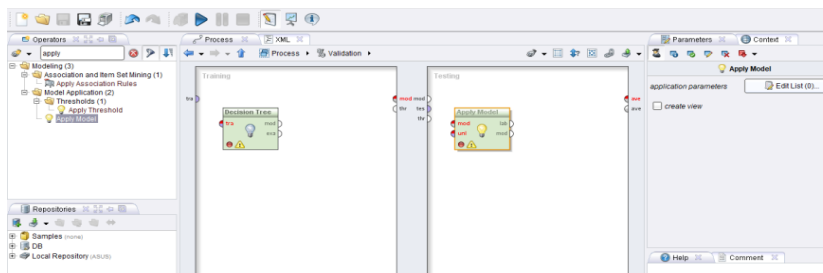
6. Tampilan *operation decision tree* pada tabel *training*



Gambar 11. Tampilan *Operators Decision Tree* Pada Tabel *Training*

Pada tabel *training* lakukan *drag* dan *drop* algoritma yang akan dilakukan yaitu *decision tree*.

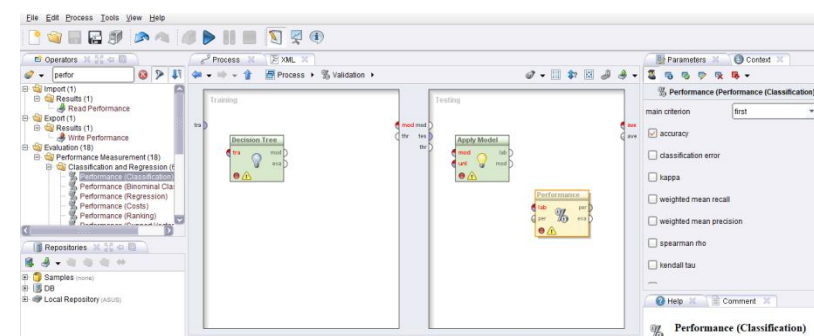
7. Tampilan *operators apply model* pada tabel *testing*



Gambar 12. Tampilan *Operators Apply Model* Pada Tabel *Testing*

Pada tabel *testing* lakukan *drag* dan *drop* *apply model*. *Apply model* berfungsi untuk mempelajari informasi *exampleset* yang telah dilatih dan digunakan untuk prediksi menggunakan model ini.

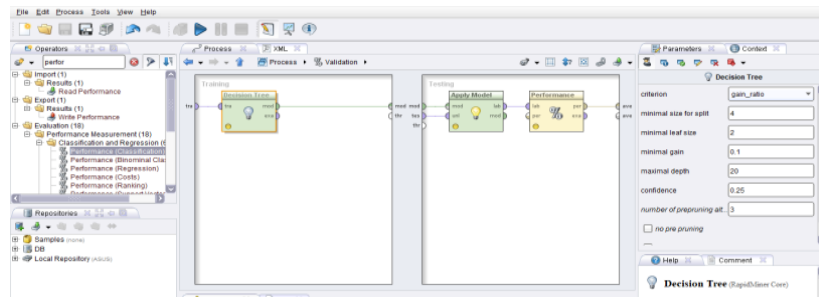
8. Tampilan *operators performance* pada tabel *testing*



Gambar 13. Tampilan *Operators Performance* Pada Tabel *Testing*

Kemudian lakukan *drag and drop performance clasification* pada *table testing*. *Performance* digunakan untuk evaluasi statistik dari kinerja klasifikasi dan memberikan daftar nilai kriteria dari kinerja klasifikasi tersebut.

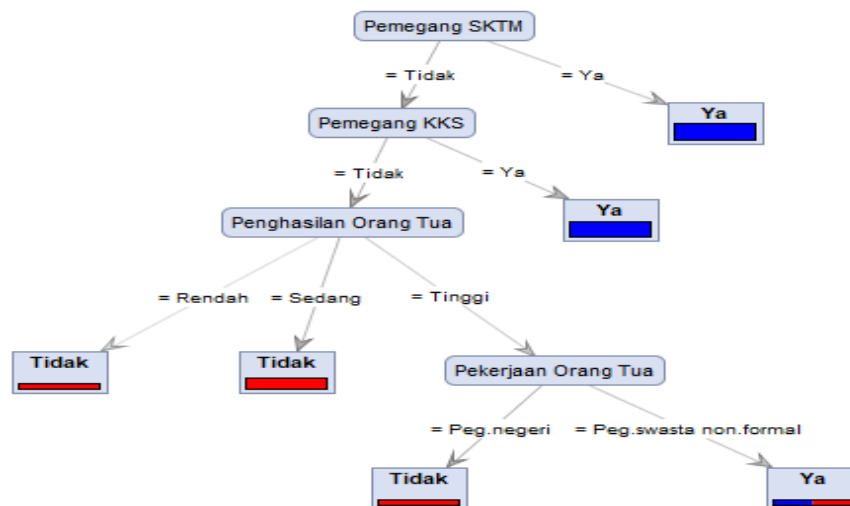
9. Tampilan menghubungkan *port decision tree*, *apply model* dan *performance*



Gambar 14. Tampilan Menghubungkan *Port Decision Tree*, *Apply Model* Dan *Performance*

Selanjutnya hubungkan port-port dari operator *decision tree*, operator *Apply Model* dan operator *performance* seperti gambar diatas, kemudian klik icon Run pada toolbar untuk menampilkan hasil.

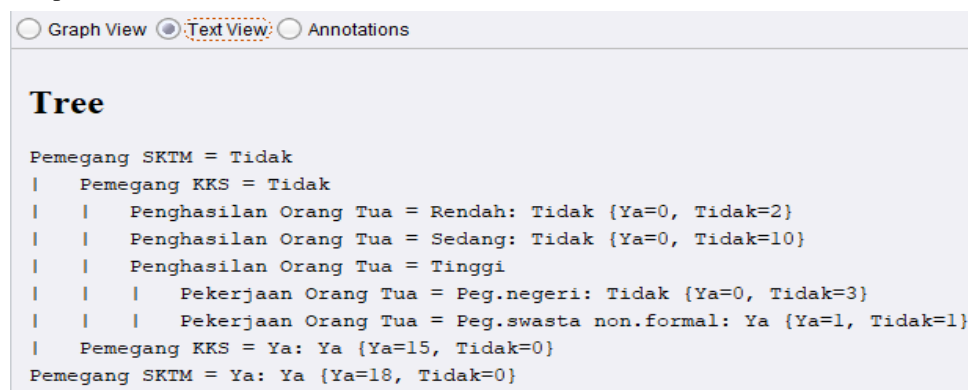
10. Tampilan *decision tree*



Gambar 15. Hasil *Decision Tree*

Setelah dilakukan perhitungan dan pengujian data pada masing-masing atribut dengan algoritma C4.5, maka akan didapatkan pola pohon keputusan akhir.

11. Tampilan Deskripsi *Decision Tree*



Gambar 16. Deskripsi *Decision Tree*

Gambar di atas menunjukkan hasil deskripsi secara lengkap dari pohon keputusan (*decision tree*) yang telah terbentuk dengan menggunakan algoritma C4.5. Dari hasil deskripsi juga menunjukkan bahwa penggunaan data mining algoritma C4.5 baik digunakan dalam proses menggali data (*data mining proses*) untuk menarik beberapa kesimpulan yang divisualisasikan dengan pohon keputusan (*decision tree*). Berikut ini *rules* yang dihasilkan dari pohon keputusan tersebut.

Tabel 9. *Rule* yang Dihasilkan

No	Rules	Keputusan
1.	Jika bukan pemegang sktm maka	Tidak
2.	Jika bukan pemegang kks maka	Tidak
3.	Jika penghasilan orang tua rendah dan bukan penerima sktm dan kks maka	Tidak
4.	Jika pekerjaan orang tua tinggi maka	Tidak
5.	Jika pekerjaan orang tua peg.swasta non formal dan memiliki kks maka	Ya
6.	Jika pemegang kks maka	Ya
7.	Jika pemegang sktm maka	Ya
8.	Jika penghasilan orang tua tinggi, pekerjaan peg.swasta non.formal dan tidak mempunyai kks/sktm maka	Tidak

12. Tingkat Akurasi

	true Ya	true Tidak	class precision
pred. Ya	0	1	100.00%
pred. Tidak	1	16	94.12%
class recall	97.06%	100.00%	

Gambar 17. Tingkat Akurasi

Berdasarkan pengolahan data menggunakan *software rapidminer* didapat nilai akurasi sistem sebesar 98.00% artinya bahwa *rule* yang dihasilkan tingkat kebenarannya mendekati 100%. Dimana model yang telah dibentuk diuji tingkat akurasinya dengan memasukkan atau uji yang berasal dari data *training* dengan menggunakan *split validation* pada aplikasi *rapidminer* 5.3 untuk menguji tingkat akurasi

4. KESIMPULAN

Berdasarkan seluruh hasil tahapan penelitian yang telah dilakukan pada Analisis Data Siswa Dalam Penentuan Penerima PIP (Program Indonesia Pintar) di MIS Al-Khoirrot dengan algoritma C4.5 dapat disimpulkan sebagai berikut: Permasalahan menentukan kelayakan data siswa dalam penentu penerima PIP dapat diselesaikan menggunakan teknik data mining, yaitu dengan algoritma C4.5 menghasilkan 8 (delapan) rules dan tingkat akurasi yang dihasilkan oleh metode tersebut adalah 98,00%. Dengan adanya penerapan data mining algoritma C4.5 diharapkan mampu memberi solusi dalam menentukan kelayakan data siswa dalam penentuan penerima Program Indonesia Pintar di MIS Al-Khoirrot.

REFERENSI

- [1] A. Wanto and E. Kurniawan, "Seleksi Penerimaan Asisten Laboratorium Menggunakan Algoritma AHP Pada AMIK-STIKOM Tunas Bangsa Pematangsiantar," *Jurnal Informatika dan Komputer (JIKO)*, vol. 3, no. 1, pp. 11–18, 2018.
- [2] V. V. Sianipar, A. Wanto, and M. Safii, "Decision Support System for Determination of Village Fund Allocation Using AHP Method," *The IJICS (International Journal of Informatics and Computer Science) ISSN*, vol. 4, no. 1, pp. 20–28, 2020.
- [3] P. P. P. A. N. . F. I. R.H Zer, Masitha, A. P. Windarto, and A. Wanto, "Analysis of the ELECTRE Method on the Selection of Student Creativity Program Proposals," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, Aug. 2019.
- [4] D. M. Sinaga, R. R. R. Alfah, A. P. Windarto, and A. Wanto, "Jurnal Sains dan Informatika," *Jurnal Sains dan Informatika*, vol. 5, no. 2, pp. 129–135, 2018.
- [5] D. R. Sari, N. Rofiqo, D. Hartama, A. P. Windarto, and A. Wanto, "Analysis of the Factors Causing Lazy Students to Study Using the ELECTRE II Algorithm," *Journal of Physics: Conference Series*, vol. 1255, no. 1, p. 012007, Aug. 2019.
- [6] T. Imandasari, A. Wanto, and A. P. Windarto, "Analisis Pengambilan Keputusan Dalam Menentukan Mahasiswa PKL Menggunakan Metode PROMETHEE," *Jurnal Riset Komputer (JURIKOM)*, vol. 5, no. 3, pp. 234–239, 2018.
- [7] R. Watrionthos, 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.
- [8] G. W. Bhawika *et al.*, "Implementation of ANN for Predicting the Percentage of Illiteracy in Indonesia by Age Group," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [9] A. Wanto *et al.*, "Analysis of the Backpropagation Algorithm in Viewing Import Value Development Levels Based on Main Country of Origin," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.

- [10] E. Siregar, H. Mawengkang, E. B. Nababan, and A. Wanto, "Analysis of Backpropagation Method with Sigmoid Bipolar and Linear Function in Prediction of Population Growth," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [11] M. K. Z. Sormin, P. Sihombing, A. Amalia, A. Wanto, D. Hartama, and D. M. Chan, "Predictions of World Population Life Expectancy Using Cyclical Order Weight / Bias," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [12] A. Wanto *et al.*, "Analysis of the Accuracy Batch Training Method in Viewing Indonesian Fisheries Cultivation Company Development," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [13] Y. Andriani, H. Silitonga, and A. Wanto, "Analisis Jaringan Syaraf Tiruan untuk prediksi volume ekspor dan impor migas di Indonesia," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 4, no. 1, pp. 30–40, 2018.
- [14] W. Saputra, J. T. Hardinata, and A. Wanto, "Implementation of Resilient Methods to Predict Open Unemployment in Indonesia According to Higher Education Completed," *JITE (Journal of Informatics and Telecommunication Engineering)*, vol. 3, no. 1, pp. 163–174, 2019.
- [15] N. L. W. S. R. Ginantra *et al.*, "Performance One-step secant Training Method for Forecasting Cases," *Journal of Physics: Conference Series*, vol. 1933, no. 1, pp. 1–8, 2021.
- [16] 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 Conf. Series: Materials Science and Engineering*, vol. 1071, no. 012018, pp. 1–7, 2021.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, "COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm," *Journal of Physics: Conference Series*, vol. 1783, no. 1, p. 012027, 2021.
- [21] 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*, 2019, vol. 1255, no. 1, p. 012031.
- [22] M. A. Hanafiah, A. Wanto, and P. B. Indonesia, "Implementation of Data Mining Algorithms for Grouping Poverty Lines by District/City in North Sumatra," *International Journal of Information System & Technology*, vol. 3, no. 2, pp. 315–322, 2020.
- [23] I. S. Damanik, A. P. Windarto, A. Wanto, Poningsih, S. R. Andani, and W. Saputra, "Decision Tree Optimization in C4.5 Algorithm Using Genetic Algorithm," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
- [24] A. Wanto *et al.*, *Data Mining : Algoritma dan Implementasi*. Yayasan Kita Menulis, 2020.
- [25] D. Hartama, A. Perdana Windarto, and A. Wanto, "The Application of Data Mining in Determining Patterns of Interest of High School Graduates," *Journal of Physics: Conference Series*, vol. 1339, no. 1, p. 012042, Dec. 2019.
- [26] H. J. Jajam and N. I. P. Astuti, "PENERAPAN ALGORITMA C4.5 UNTUK PENENTUAN PENERIMA BEASISWA (Studi Kasus: SDN Karawang Kulon III) 1," *Ilmu Komputer dan Teknologi Informasi*, vol. 3, no. 1, pp. 25–29, 2018.
- [27] R. Rismayanti, "Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa Di Stt Harapan Medan," *Jurnal Media Infotama*, vol. 12, no. 2, pp. 116–120, 2016.
- [28] D. Ardiansyah and W. Walim, "Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner," *Jurnal Inkofar*, vol. 1, no. 2, pp. 5–12, 2018.
- [29] R. Z. Nainggolan, K. Ibnutama, and M. G. Suryanata, "Implementasi Data Mining Dengan Metode Regresi Linier Berganda Dalam Estimasi Mahasiswa Baru Pada Sekolah Tinggi Agama Islam Raudhatul Akmal BatangKuis," *Jurnal CyberTech*, vol. 1, no. 1, pp. 13–20, 2021.