

Penerapan Metode CART Dalam Klasifikasi Jurusan Siswa Baru

Destia Arini Hairunnisa^{*}, Cucu Suhery, Rahmi Hidayati

FMIPA, Prodi Rekayasa Sistem Komputer, Universitas Tanjungpura, Pontianak, Indonesia

Email: ^{1,*}destiaaha00@student.untan.ac.id, ²cusuhery@siskom.untan.ac.id, ³rahmihidayati@siskom.untan.ac.id

Correspondence Author Email: destiaaha00@student.untan.ac.id

Submitted: 15/07/2023; Accepted: 31/08/2023; Published: 31/08/2023

Abstrak—Sekolah Menengah Kejuruan Negeri 3 Pontianak merupakan salah satu pendidikan kejuruan yang ada di Kota Pontianak. Setiap tahun ajaran baru SMK Negeri 3 Pontianak menerima sekitar 320 siswa baru. Banyaknya jumlah calon siswa baru tersebut membuat proses penjurusan yang dilakukan oleh pihak sekolah menjadi kurang efisien dan membutuhkan waktu yang lama untuk menentukan jurusan siswa baru. Dengan adanya sistem yang dapat mengklasifikasikan jurusan siswa baru membantu dalam proses penentuan jurusan siswa. Penelitian ini menggunakan metode *Classification and Regression Trees* (CART) untuk proses pengklasifikasian dalam menentukan jurusan bagi siswa baru. Indikator penilaian yang digunakan untuk klasifikasi terdiri dari minat, nilai ujian sekolah (us) mtk, nilai us ipa, nilai us bahasa indonesia, nilai rapor matematika, nilai rapor ipa, nilai rapor ips, nilai rapor bahasa indonesia, dan nilai rapor bahasa inggris. Klasifikasi jurusan pada SMK 3 Pontianak terdiri dari jurusan akuntansi, perkantoran, pemasaran, dan perhotelan. Hasil penelitian ini menunjukkan bahwa Algoritma CART berhasil melakukan klasifikasi dengan baik. Jumlah data yang digunakan sebanyak 320. Data tersebut kemudian dibagi menjadi 224 data *training* dan 96 data *testing*. Algoritma CART menghasilkan pohon keputusan, *rules*, dan jurusan siswa baru yang telah diklasifikasi. Berdasarkan hasil pengujian menggunakan *confusion matrix* diperoleh hasil akurasi sistem sebesar 84,38%.

Kata Kunci: CART; Jurusan; Klasifikasi; Nilai; Siswa

Abstract—SMK Negeri 3 Pontianak is one of the vocational education schools in Pontianak City. Every new academic year SMK Negeri 3 Pontianak accepts around 320 new students. The large number of prospective new students makes the majoring process carried out by the school become less effective and takes a long time to determine majors for new students. With a system that can classify new student majors, it helps in the process of determining student majors. This study use the Classification and Regression Trees (CART) algorithm for classification process in determining majors for new students. The assessment indicators used for classification consist of interest, MTK (US) school exam scores, school exam IPA scores, school exam Indonesian language scores, math report cards, science report cards, social science report cards, Indonesian language report cards, and English report cards. Classification of majors at SMK 3 Pontianak consists of accounting, office, marketing, and hospitality majors. The results of this study indicate that the CART Algorithm succeeds in performing a good classification. The amount of data used is 320 data which is divided into 224 training data and 96 test data. The CART algorithm generates decision trees, rules, and new student majors that have been classified. Based on the test results using the confusion matrix, the system accuracy results are 84.38%.

Keywords: CART; Department; Classification; Value; Student

1. PENDAHULUAN

Pendidikan merupakan usaha sadar dan terencana untuk mewujudkan suasana proses pembelajaran agar peserta didik aktif dalam mengembangkan potensi dirinya[1]. Pendidikan menengah terdiri dari pendidikan menengah umum dan pendidikan menengah kejuruan. Pendidikan kejuruan merupakan pendidikan menengah yang mempersiapkan peserta didik terutama untuk bekerja dalam bidang tertentu[2]. SMK Negeri 3 Pontianak memiliki 4 jurusan, yaitu Jurusan Akuntansi, Perkantoran, Pemasaran, dan Perhotelan. Siswa baru yang akan mendaftar diberikan kesempatan untuk memilih jurusan yang sesuai minat, bakat, dan kemampuannya masing-masing. Namun, banyak calon siswa baru yang belum mengetahui dan belum mengenali minat, bakat dan kemampuannya sehingga terkadang siswa ragu dalam memilih jurusan. Setiap tahun di SMK Negeri 3 Pontianak menerima sekitar 320 siswa baru. Banyaknya jumlah calon siswa baru tersebut membuat proses penjurusan yang dilakukan oleh pihak sekolah menjadi kurang efektif karena proses klasifikasi tersebut masih dilakukan dengan mengumpulkan seluruh berkas siswa baru dan mengecek nilai yang diperoleh oleh setiap siswa. Untuk mempermudah proses penjurusan siswa, diperlukan sebuah sistem yang dapat mengklasifikasikan jurusan siswa sehingga membantu pihak sekolah dalam mengambil keputusan penjurusan dengan tepat dan cepat. Pada penelitian ini klasifikasi penjurusan siswa dilakukan berdasarkan 9 indikator yaitu minat, nilai Ujian Sekolah Matematika (US MTK), nilai Ujian Sekolah Ilmu Pengetahuan Alam (US IPA), nilai Ujian Sekolah Bahasa Indonesia (US BI), nilai Rapor Matematika, nilai Rapor Ilmu Pengetahuan Alam, nilai Rapor Ilmu Pengetahuan Sosial (IPS), nilai Rapor Bahasa Indonesia, dan nilai Rapor Bahasa Inggris saat berada dibangku Sekolah Menengah Pertama (SMP).

Salah satu algoritma yang dapat digunakan untuk proses klasifikasi adalah algoritma *Classification and Regression Trees* (CART). CART merupakan salah satu algoritma dari *Decision Tree* (pohon keputusan). Algoritma CART menggunakan perhitungan *index gini* untuk pembentukan cabang dan perhitungan *gini gain* untuk pembentukan simpul (*node*) yang hasil akhirnya akan membentuk pohon keputusan dan akan menghasilkan klasifikasi[3]. Penelitian sebelumnya yang menggunakan algoritma CART untuk menentukan jurusan yaitu di MAN 1 Inhil. Hasil penelitian tersebut adalah sebuah sistem yang dapat mengklasifikasikan

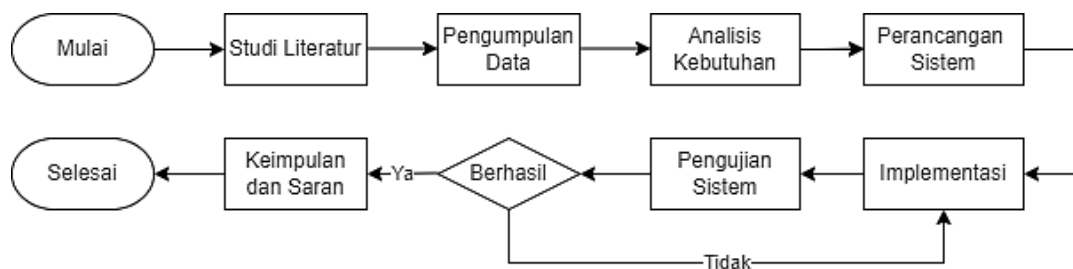
jurusan siswa baru menggunakan data sebanyak 360 data siswa[4]. Penelitian lain yang menggunakan klasifikasi yaitu klasifikasi jurusan siswa baru pada MAN 1 di OKU Timur. Penelitian tersebut menghasilkan hasil klasifikasi berupa tiga jurusan IPA, IPS, dan Agama dengan tingkat akurasi sebesar 88,654% [5]. Penelitian selanjutnya dengan menerapkan Algoritma CART untuk memprediksi penjualan pada produk *fast moving* dan *slow moving*[6]. Penelitian lain yang menggunakan metode CART yaitu pada klasifikasi bekerja dan pengangguran[7]. Penelitian terkait lainnya tentang metode CART digunakan untuk menentukan faktor-faktor yang mempengaruhi pembayaran kredit oleh nasabah. Penelitian tersebut menghasilkan akurasi sebesar 89,1% [8]. Penelitian selanjutnya tentang penerapan metode CART yaitu menganalisis prediksi masa studi mahasiswa baik alumni maupun mahasiswa baru[9]. Penelitian sebelumnya tentang Algoritma CART digunakan untuk klasifikasi penyakit ginjal kronik. Hasil penelitian tersebut mendapat tingkat akurasi sebesar 98,33% [10]. Algoritma CART juga digunakan dalam klasifikasi penentuan kelayakan penerima bantuan yang mendapat tiga keluaran yaitu layak, dipertimbangkan dan tidak layak[11]. Penelitian lain yang menggunakan algoritma CART yaitu pada penentuan bakat dan minat anak[12].

Berdasarkan latar belakang dan penelitian yang telah dilakukan sebelumnya, penelitian ini menerapkan Algoritma CART untuk menentukan klasifikasi jurusan siswa baru di SMK Negeri 3 Pontianak. Sistem yang dibuat berbasis *website* dan digunakan untuk memudahkan pihak sekolah dalam mengklasifikasikan jurusan siswa baru berdasarkan minat siswa, nilai ujian sekolah dan nilai rapor.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam melaksanakan penelitian dengan menerapkan algoritma CART untuk klasifikasi jurusan siswa baru di SMK Negeri 3 Pontianak, terdapat tujuh tahapan yang dilakukan. Tahapan pertama adalah studi literatur untuk mencari informasi serta materi yang dapat digunakan sebagai sumber dalam penelitian ini. Studi literatur yang digunakan dapat berupa jurnal yang berkaitan dengan penelitian, buku, skripsi, artikel, atau data lain yang dapat digunakan untuk mendukung penelitian. Tahapan kedua adalah pengumpulan data, metode pengumpulan data yang digunakan dalam penelitian ini adalah dengan wawancara kepada pihak kepala sekolah SMK Negeri 3 Pontianak. Tahapan ketiga adalah analisis kebutuhan, untuk mendukung kinerja sistem penerapan metode CART dalam klasifikasi jurusan siswa baru. Tahapan ketiga dibagi menjadi dua, yaitu analisis kebutuhan perangkat keras dan perangkat lunak. Tahapan keempat adalah perancangan sistem berupa perancangan berupa perangkat lunak. Tahapan kelima adalah implementasi untuk mengimplementasikan penerapan algoritma CART dalam klasifikasi jurusan siswa baru ke dalam sistem berupa kode program. Tahapan keenam adalah pengujian sistem yang dilakukan menggunakan suatu perangkat lunak dengan tujuan untuk mengetahui hasil akurasi dari sistem yang telah dibuat agar sesuai dengan kegunaannya serta mengetahui kemampuan metode yang digunakan, baik dari *input* maupun *output*. Untuk menghitung akurasi menggunakan *confusion matrix*. Tahapan ketujuh adalah kesimpulan dan saran, setelah tahap pengujian sistem telah berhasil dilakukan, tahap terakhir dalam penelitian yaitu penarikan kesimpulan yang didapatkan dari hasil pengujian serta menjawab pertanyaan yang terdapat dalam permasalahan masalah. Kemudian pemberian saran dilakukan untuk penyelesaian terhadap masalah yang belum diselesaikan dalam penelitian. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.2 Data Mining

Data mining merupakan suatu kegiatan yang berhubungan dalam mengumpulkan data, penggunaan data sebelumnya digunakan untuk menemukan pengetahuan, informasi, dan relasi yang terdapat dalam data yang berjumlah besar. Keluaran yang didapatkan dalam *data mining* ini digunakan untuk sebagai solusi dalam penentuan keputusan[13].

2.3 Klasifikasi

Klasifikasi merupakan proses yang digunakan untuk menemukan model dalam membedakan konsep atau kelas data yang bertujuan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui sebelumnya[14].

2.4 Decision Tree

Decision Tree atau Pohon keputusan adalah salah satu metode dari klasifikasi yang menggunakan struktur data berbentuk seperti pohon. Konsep dari pohon keputusan tersebut ialah mengubah data menjadi pohon keputusan *rules*[15].

2.5 Algoritma Classification And Regression Trees (CART)

CART (*Classification and Regression Trees*) merupakan salah satu algoritma dari teknik pengolahan data pohon keputusan (*Decision Tree*)[16]. Algoritma CART melakukan pemilihan yang dilakukan pada sekelompok data yang terkumpul dalam suatu ruang yang disebut simpul menjadi dua kelompok. Dua kelompok tersebut disebut simpul anak[17]. Data yang digunakan dalam metode CART adalah data *training* (latih) dan data *testing* (uji). Data *training* adalah data yang siap untuk *dimining* yang telah melewati data *preprocessing* sedangkan data *testing* adalah data yang digunakan untuk menguji *rules* klasifikasi yang diperoleh dari data *training*[3]. Langkah-langkah dalam menyelesaikan perhitungan dengan menggunakan Algoritma CART sebagai berikut[18]:

1. Menghitung nilai *Index Gini* total data yang digunakan.
2. Menghitung nilai *Index Gini* pada seluruh indikator yang digunakan, Untuk menghitung *Index Gini* digunakan Persamaan 1.

$$Index\ Gini\ (t) = 1 - \sum_{j=1}^n P^2(j|t) \tag{1}$$

$$Dimana, P^2(j|t) = \frac{n_j(t)}{n(t)}$$

Keterangan:

$P(j|t)$ = probabilitas bersyarat kelas j yang berada pada simpul t

$n_j(t)$ = banyak pengamatan kelas j pada *node* t

$n(t)$ = banyak pengamatan pada *node* t

3. Menghitung nilai *Gini Gain* total indikator Untuk menghitung total indikator digunakan Persamaan 2.

$$Gini\ Gain = \sum_{j=1}^n \frac{|S_j|}{|S|} \times Index\ Gini(t) \tag{2}$$

Keterangan:

$|S_j|$ = jumlah data dalam partisi ke- j

$|S|$ = jumlah total data

4. Indikator yang memiliki nilai *Gini Gain* tertinggi dipilih menjadi *node* akar.
5. Indikator yang memiliki *Gini Gain* lebih rendah dari *node* akar akan dipilih menjadi *node* cabang (*branches*) dan mengulangi lagi langkah ke-1 sampai ke-4, tetapi *Gini Gain* tertinggi pada langkah sebelumnya tidak dihitung kembali.
6. Pada langkah ke-4 sudah memiliki *node* akar, maka indikator selanjutnya yang memiliki *Gain* tertinggi dipilih menjadi cabang (*branches*).
7. Mengulangi langkah ke-1 sampai ke-4 sampai menghasilkan nilai *Gain* = 0 untuk semua indikator yang tersisa.
8. Menghasilkan pohon klasifikasi dan menghasilkan *rules*.

2.6 Confusion Matrix

Pada penelitian ini *confusion matrix* digunakan untuk pengujian data siswa. *Confusion Matrix* merupakan sebuah tabel yang terdiri atas banyaknya baris data uji yang diprediksi benar (*positive*) dan tidak benar (*negative*) oleh model klasifikasi. *Confusion matrix* dengan *matrix* 4x4 disebut juga model klasifikasi multi kelas yang digunakan pada variabel lebih dari dua kelas yang berfungsi untuk menghasilkan kinerja berupa TP, TN, FP, dan FN[19]. Tabel *confusion matrix* 4x4 dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrics 4x4

Confusion Mtarix	Classified				
	0	1	2	3	
Actual	0	+C ₀₀	-C ₀₁	-C ₀₂	-C ₀₃
	1	-C ₁₀	+C ₁₁	-C ₁₂	-C ₁₃
	2	-C ₂₀	-C ₂₁	+C ₂₂	-C ₂₃
	3	-C ₃₀	-C ₃₁	-C ₃₂	+C ₃₃

Berdasarkan tabel *confusin matrix*, maka dapat dihitung beberapa parameter sebagai berikut[20]:

1. Akurasi

Akurasi adalah nilai dari total data uji coba yang telah benar diklasifikasikan. Rumus dari akurasi dapat dilihat pada Persamaan 3.

$$Akurasi = \frac{TP}{TP+FN+FP+TN} \times 100\% \tag{3}$$

2. Precision

Precision adalah data yang diambil sesuai dengan informasi yang dibutuhkan. Rumus dari precision dapat dilihat pada Persamaan 4.

$$Precision = \frac{TP}{FP+TP} \times 100\% \tag{4}$$

3. Recall

Recall adalah data yang berhasil dilakukan terhadap bagian data yang relevan dengan query. Rumus dari recall dapat dilihat pada Persamaan 5.

$$Recall = \frac{TP}{FN+TP} \times 100\% \tag{5}$$

4. Laju Error

Laju error digunakan untuk menghitung kesalahan identifikasi. Rumus dari laju error dapat dilihat pada Persamaan 6.

$$Laju\ Error = \frac{FP+FN}{TP+FN+FP+TN} \times 100\% \tag{6}$$

3. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini sebanyak 320 data siswa. Indikator yang digunakan sebanyak 9 indikator yang terdiri dari minat, nilai us mtk, nilai us ipa, nilai us bahasa indonesia, nilai rapor matematika, nilai rapor ipa, nilai rapor ips, nilai rapor bahasa indonesia, dan nilai rapor bahasa inggris. Klasifikasi jurusan pada penelitian ini terdiri dari Jurusan Akuntansi, Perkantoran, Pemasaran, dan Perhotelan.

3.1 Data Siswa Baru

Data siswa baru diperoleh dengan melakukan wawancara kepada pihak sekolah. Data siswa baru yang digunakan adalah pada tahun ajaran 2022/2023. Data latih dan uji akan disajikan dalam bentuk tabel dengan data pribadi siswa baru yang akan disimbolkan dengan SB1, SB2, SB3 sampai dengan SB320. Data siswa baru tahun ajaran 2022/2023 dapat dilihat pada Tabel 2.

Tabel 2. Data Siswa Baru Tahun Ajaran 2022/2023

No	Nama Siswa	Minat	Nilai US			Nilai Rapor					Jurusan	Hasil CART
			MTK	IPA	B.I	MTK	IPA	IPS	B.I	B.ING		
1	SB1	AK	A	A	A	A	B	A	B	B	AK	AK
2	SB2	AK	A	C	B	A	A	B	B	B	AK	AK
3	SB3	AK	A	B	A	A	A	A	A	A	AK	PM
4	SB4	AK	A	A	C	A	A	B	A	B	AK	AK
5	SB5	AK	B	A	A	A	A	A	B	A	AK	AK
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
316	SB316	AK	A	A	A	B	A	A	A	A	AK	AK
317	SB317	AK	A	A	A	B	A	A	A	A	AK	AK
318	SB318	PK	C	B	C	A	A	A	A	A	PH	PM
319	SB319	AK	B	A	B	A	A	A	A	A	PH	PH
320	SB320	AK	A	A	A	A	A	A	A	A	PH	AK

Keterangan:

AK = Akuntansi

PK = Perkantoran

PM = Pemasaran

PH = Perhotelan

3.2 Hasil Perhitungan Algoritma CART

Pada penelitian ini, menggunakan perhitungan dengan perbandingan data latih dan data uji sebesar 70:30. Untuk mengetahui data latih dan data uji yang digunakan dari 320 data dihitung dengan perbandingan 70:30 dan diperoleh hasil data latih dan data uji sebagai berikut:

$$Jumlah\ data\ latih = 70\% \times 320 = \frac{70}{100} \times 320 = 224\ data$$

$$Jumlah\ data\ uji = 30\% \times 320 = \frac{30}{100} \times 320 = 96\ data$$

Tahapan perhitungan untuk menghasilkan pohon keputusan menggunakan algoritma CART adalah sebagai berikut.

1. Menghitung jumlah data yang secara keseluruhan dan jumlah data dari setiap indikator yang digunakan. Perhitungan jumlah data pada data latih dapat dilihat pada Tabel 3.

Tabel 3. Perhitungan Jumlah Data pada Data Latih

No.	Indikator	Kategori	Jumlah	AK	PK	PM	PH
	Jurusan		224	110	50	50	14
1	Minat Siswa	Akuntansi	135	66	34	27	8
		Perkantoran	58	29	11	15	3
		Pemasaran	22	13	2	5	2
		Perhotelan	9	2	3	3	1
2	US MTK	A	105	55	26	18	6
		B	67	30	17	16	4
		C	52	25	7	16	4
3	US IPA	A	101	52	24	20	5
		B	81	39	15	18	9
		C	42	19	11	12	0
4	US B. Ind	A	113	53	31	22	7
		B	85	46	12	21	6
		C	26	11	7	7	1
5	Rapor MTK	A	143	71	33	29	10
		B	62	30	11	19	2
		C	19	9	6	2	2
6	Rapor IPA	A	168	93	33	29	13
		B	43	13	15	15	0
		C	13	4	2	6	1
7	Rapor IPS	A	183	89	41	39	14
		B	33	15	9	9	0
		C	8	6	0	2	0
8	Rapor B. Indonesia	A	177	88	29	47	13
		B	38	19	15	3	1
		C	9	3	6	0	0
9	Rapor B. Inggris	A	171	71	48	38	14
		B	42	30	2	10	0
		C	11	9	0	2	0

2. Menghitung *Index Gini* menggunakan Persamaan 1. *Index Gini* total menggunakan data sebanyak 224 data sesuai yang tertera pada Tabel 3. Perhitungan *Index Gini* total sebagai berikut.

$$IndexGini(Jurusan) = 1 - \left(\left(\frac{110}{224} \right)^2 + \left(\frac{50}{224} \right)^2 + \left(\frac{50}{224} \right)^2 + \left(\frac{14}{224} \right)^2 \right) = 0,65529339$$

3. Menghitung *Index Gini* pada masing-masing indikator. Perhitungan yang dilakukan adalah perhitungan *Index Gini* dari minat siswa.

$$Index\ Gini\ (Minat\ Siswa,\ Akuntansi) = 1 - \left(\left(\frac{66}{135} \right)^2 + \left(\frac{34}{135} \right)^2 + \left(\frac{27}{135} \right)^2 + \left(\frac{8}{135} \right)^2 \right) = 0,65404666$$

$$Index\ Gini\ (Minat\ Siswa,\ Perkantoran) = 1 - \left(\left(\frac{29}{58} \right)^2 + \left(\frac{11}{58} \right)^2 + \left(\frac{15}{58} \right)^2 + \left(\frac{3}{58} \right)^2 \right) = 0,64447088$$

$$Index\ Gini\ (Minat\ Siswa,\ Pemasaran) = 1 - \left(\left(\frac{13}{22} \right)^2 + \left(\frac{2}{22} \right)^2 + \left(\frac{5}{22} \right)^2 + \left(\frac{2}{22} \right)^2 \right) = 0,58264464$$

$$Index\ Gini\ (Minat\ Siswa,\ Perhotelan) = 1 - \left(\left(\frac{2}{9} \right)^2 + \left(\frac{3}{9} \right)^2 + \left(\frac{3}{9} \right)^2 + \left(\frac{1}{9} \right)^2 \right) = 0,71604938$$

4. Menghitung *Gini Gain* menggunakan Persamaan 2. Perhitungan yang dilakukan adalah perhitungan *Gini Gain* masing-masing indikator.

$$Gini\ Gain\ (Minat) = 0,655293 - \left(\left(\frac{135}{224} \times 0,654046 \right) + \left(\frac{58}{224} \times 0,644470 \right) + \left(\frac{22}{224} \times 0,582644 \right) \right)$$

$$+ \left(\frac{9}{224} \times 0,716049 \right) = 0,008247$$

$$\begin{aligned} \text{Gini Gain (US MTK)} &= 0,655293 - \left(\left(\frac{105}{224} \times 0,631655 \right) + \left(\frac{67}{224} \times 0,674537 \right) + \left(\frac{52}{224} \times 0,650147 \right) \right) \\ &= 0,006518 \end{aligned}$$

$$\begin{aligned} \text{Gini Gain (US IPA)} &= 0,655293 - \left(\left(\frac{101}{224} \times 0,636800 \right) + \left(\frac{81}{224} \times 0,672153 \right) + \left(\frac{42}{224} \times 0,645124 \right) \right) \\ &= 0,004148 \end{aligned}$$

$$\begin{aligned} \text{Gini Gain (US B.I)} &= 0,655293 - \left(\left(\frac{113}{224} \times 0,663011 \right) + \left(\frac{85}{224} \times 0,621176 \right) + \left(\frac{26}{224} \times 0,674556 \right) \right) \\ &= 0,006816 \end{aligned}$$

$$\begin{aligned} \text{Gini Gain (Rapor MTK)} &= 0,65529 - \left(\left(\frac{143}{224} \times 0,654212 \right) + \left(\frac{62}{224} \times 0,63943 \right) + \left(\frac{19}{224} \times 0,653739 \right) \right) \\ &= 0,005210 \end{aligned}$$

$$\begin{aligned} \text{Gini Gain (Rapor IPA)} &= 0,65529 - \left(\left(\frac{168}{224} \times 0,619217 \right) + \left(\frac{43}{224} \times 0,665224 \right) + \left(\frac{13}{224} \times 0,662721 \right) \right) \\ &= 0,024719 \end{aligned}$$

$$\begin{aligned} \text{Gini Gain (Rapor IPS)} &= 0,655293 - \left(\left(\frac{183}{224} \times 0,662008 \right) + \left(\frac{33}{224} \times 0,644628 \right) + \left(\frac{8}{224} \times 0,375 \right) \right) \\ &= 0,006095 \end{aligned}$$

$$\begin{aligned} \text{Gini Gain (Rapor B.I)} &= 0,655293 - \left(\left(\frac{177}{224} \times 0,650068 \right) + \left(\frac{38}{224} \times 0,587257 \right) + \left(\frac{9}{224} \times 0,44444 \right) \right) \\ &= 0,024141 \end{aligned}$$

$$\begin{aligned} \text{Gini Gain (Rapor B. Ing)} &= 0,655293 - \left(\left(\frac{171}{224} \times 0,69272 \right) + \left(\frac{42}{224} \times 0,43083 \right) + \left(\frac{11}{224} \times 0,29752 \right) \right) \\ &= 0,031078 \end{aligned}$$

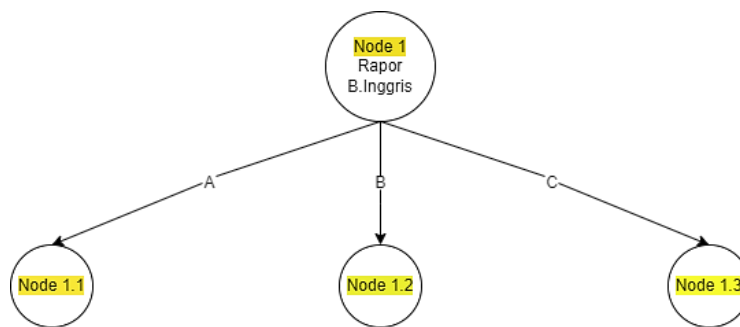
Perhitungan menggunakan Persamaan (1) dan Persamaan (2) terus dilakukan untuk menghasilkan simpul akar yang didapatkan dari nilai *Index Gini* tertinggi. Dari 9 indikator yang digunakan akan menghasilkan *Gini Gain* yang tertinggi. *Gain Gain* tertinggi tersebut akan digunakan untuk menghitung simpul cabang selanjutnya. Jumlah data untuk perhitungan menggunakan algoritma CART dapat dilihat pada Tabel 4.

Tabel 4. Jumlah Data Untuk Perhitungan Menggunakan Algoritma CART

No	Indikator	Kategori	Jumlah	AK	PK	PM	PH	<i>Index Gini</i>	<i>Gini Gain</i>
	Jurusan		224	110	50	50	14	0,65529339	
1	Minat Siswa	Akuntansi	135	66	34	27	8	0,65404666	0,00824771
		Perkantoran	58	29	11	15	3	0,64447088	
		Pemasaran	22	13	2	5	2	0,58264464	
		Perhotelan	9	2	3	3	1	0,71604938	
2	US MTK	A	105	55	26	18	6	0,63165535	0,00651869
		B	67	30	17	16	4	0,67453778	
		C	52	25	7	16	4	0,65014795	
3	US IPA	A	101	52	24	20	5	0,63680032	0,00414824
		B	81	39	15	18	9	0,67215366	
		C	42	19	11	12	0	0,64512473	
4	US Bahasa Indonesia	A	113	53	31	22	7	0,66301198	0,00681654
		B	85	46	12	21	6	0,62117647	
		C	26	11	7	7	1	0,67455621	
5	Rapor MTK	A	143	71	33	29	10	0,65421294	0,00521008
		B	62	30	11	19	2	0,6394381	
		C	19	9	6	2	2	0,65373962	
6	Rapor IPA	A	168	93	33	29	13	0,61921748	0,0247194
		B	43	13	15	15	0	0,66522447	

No	Indikator	Kategori	Jumlah	AK	PK	PM	PH	Index Gini	Gini Gain
7	Rapor IPS	C	13	4	2	6	1	0,66272192	0,00609575
		A	183	89	41	39	14	0,66200845	
		B	33	15	9	9	0	0,64462811	
		C	8	6	0	2	0	0,375	
8	Rapor B. Indonesia	A	177	88	29	47	13	0,65006864	0,02414189
		B	38	19	15	3	1	0,58725763	
		C	9	3	6	0	0	0,44444446	
9	Rapor Bahasa Inggris	A	171	71	48	38	14	0,69272599	0,03107861
		B	42	30	2	10	0	0,43083901	
		C	11	9	0	2	0	0,29752067	

Berdasarkan Tabel 4, indikator yang memiliki nilai *Gini Gain* tertinggi adalah nilai *Gini Gain* pada Rapor B.Ingggris dengan nilai *Gini Gain* sebesar 0,03107861 sehingga indikator tersebut menjadi *node* akar atau *node* 1. Selanjutnya, terdapat cabang untuk *node* 1, yaitu A, B, dan C, sehingga terdapat tiga cabang yang masih melakukan perhitungan kembali untuk menghasilkan klasifikasi jurusan, yaitu *node* 1.1, *node* 1.2 dan *node* 1.3. Pohon keputusan *node* 1 dapat dilihat pada Gambar 2.



Gambar 2. Pohon Keputusan Node 1

Perhitungan *node* 1.1 data nilai A yang terdapat pada indikator Rapor B.Ingggris yang akan menjadi data total menggunakan. Perhitungan *node* 1.2 menggunakan data nilai B yang terdapat pada indikator Rapor B.Ingggris yang akan menjadi data total. Selanjutnya, perhitungan *node* 1.3 menggunakan data nilai C yang terdapat pada indikator Rapor B.Ingggris yang akan menjadi data total. Setelah perhitungan *node* 1.1, 1.2 dan 1.3 dihasilkan, maka akan melakukan perhitungan lagi hingga *node* cabang tersebut hanya menghasilkan *node* terminal. Setelah perhitungan untuk menghasilkan *node* dilakukan, maka akan menghasilkan pohon keputusan. Setelah terbentuk pohon keputusan, maka dapat disimpulkan aturan (*rules*) sebagai berikut.

1. If Nilai Rapor B.Ingggris = A, and Nilai Rapor B.Indonesia = A, and Nilai Rapor IPA = A, and Nilai US MTK = A, and Nilai US B.Indonesia = A, and Nilai Rapor MTK = A, and Nilai US IPA = A, and Nilai Rapor IPS = A, then Hasil = Akuntansi.
2. If Nilai Rapor B.Ingggris = A, and Nilai Rapor B.Indonesia = A, and Nilai Rapor IPA = A, and Nilai US MTK = A, and Nilai US B.Indonesia = A, and Nilai Rapor MTK = A, and Nilai US IPA = A, and Nilai Rapor IPS = B, then Hasil = Akuntansi.
3. If Nilai Rapor B.Ingggris = A, and Nilai Rapor B.Indonesia = A, and Nilai Rapor IPA = A, and Nilai US MTK = A, and Nilai US B.Indonesia = A, and Nilai Rapor MTK = A, and Nilai US IPA = B, then Hasil = Pemasaran.
4. If Nilai Rapor B.Ingggris = A, and Nilai Rapor B.Indonesia = A, and Nilai Rapor IPA = A, and Nilai US MTK = A, and Nilai US B.Indonesia = A, and Nilai Rapor MTK = A, and Nilai US IPA = C, then Hasil = Akuntansi.
5. If Nilai Rapor B.Ingggris = A, and Nilai Rapor B.Indonesia = A, and Nilai Rapor IPA = A, and Nilai US MTK = A, and Nilai US B.Indonesia = A, and Nilai Rapor MTK = B, and Nilai US IPA = A, then Hasil = Akuntansi.
6. If Nilai Rapor B.Ingggris = A, and Nilai Rapor B.Indonesia = A, and Nilai Rapor IPA = A, and Nilai US MTK = A, and Nilai US B.Indonesia = A, and Nilai Rapor MTK = B, and Nilai US IPA = B, then Hasil = Perkantoran.
7. If Nilai Rapor B.Ingggris = A, and Nilai Rapor B.Indonesia = A, and Nilai Rapor IPA = A, and Nilai US MTK = A, and Nilai US B.Indonesia = A, and Nilai Rapor MTK = C, then Hasil = Akuntansi.
8. If Nilai Rapor B.Ingggris = A, and Nilai Rapor B.Indonesia = A, and Nilai Rapor IPA = A, and Nilai US MTK = A, and Nilai US B.Indonesia = B, and Nilai Rapor MTK = A, and Nilai Rapor IPS = A, and Nilai US IPA = A, then Hasil = Akuntansi.
9. If Nilai Rapor B.Ingggris = A, and Nilai Rapor B.Indonesia = A, and Nilai Rapor IPA = A, and Nilai US MTK = A, and Nilai US B.Indonesia = B, and Nilai Rapor MTK = A, and Nilai Rapor IPS = A, and Nilai US IPA = B, then Hasil = Akuntansi.
10. If Nilai Rapor B.Ingggris = A, and Nilai Rapor B.Indonesia = A, and Nilai Rapor IPA = A, and Nilai US MTK = A, and Nilai US B.Indonesia = B, and Nilai Rapor MTK = A, and Nilai Rapor IPS = A, and Nilai US IPA = C, then Hasil = Akuntansi.

125. *If Nilai Rapor B.Inggris = C, and Nilai Rapor IPA = C, then Hasil = Pemasaran*

3.3 Hasil Pengujian Confusion Matrix

Proses pengujian menggunakan *confusion matrix* dengan menggunakan perbandingan hasil manual yang didapat dengan hasil sistem yang dihasilkan oleh Algoritma CART. Pada pengujian menggunakan *confusion matrix* data yang digunakan 96 data. Pengujian ini dilakukan untuk menghasilkan nilai akurasi, *recall*, *precision*, dan laju *error*. Berdasarkan hasil yang diperoleh terdapat beberapa perbedaan hasil manual dengan hasil sistem pada perbandingan 70:30. Berdasarkan data yang pada Tabel 2, maka didapatkan nilai TP, TN, FP, dan FN untuk melakukan perhitungan menggunakan *confusion matrix*. *Confusion matrix* yang digunakan adalah *matrix* 4x4 seperti pada Tabel 5.

Tabel 5. Nilai TP, TN, FP, dan FN

Perbandingan 70:30		Classified			
		Akuntansi	Perkantoran	Pemasaran	Perhotelan
Actual	Akuntansi	45	1	5	0
	Perkantoran	7	19	1	0
	Pemasaran	0	1	12	0
	Perhotelan	0	0	0	5

Berdasarkan hasil dari Tabel 5, maka didapatkan akurasi, *precision*, *recall*, dan laju *error* sebagai berikut:

$$Akurasi = \frac{TP}{TP+FN+FP+TN} \times 100\% = \frac{45+19+12+5}{45+1+5+0+7+19+1+0+0+1+0+0+0+5} \times 100\% = 84,38$$

$$Precision = \frac{TP}{FP+TP} \times 100\%$$

$$Precision_{Akuntansi} = \frac{45}{7+0+0+45} \times 100\% = 86,54\%$$

$$Precision_{Perkantoran} = \frac{19}{1+0+1+19} \times 100\% = 90,47\%$$

$$Precision_{Pemasaran} = \frac{12}{5+1+0+12} \times 100\% = 66,67\%$$

$$Precision_{Perhotelan} = \frac{5}{0+0+0+5} \times 100\% = 100\%$$

$$Precision_{rata-rata} = \frac{(86,54+90,47+66,67+100)\%}{4} = 85,92\%$$

$$Recall = \frac{TP}{FN+TP} \times 100\%$$

$$Recall_{Akuntansi} = \frac{45}{1+5+0+45} \times 100\% = 88,24\%$$

$$Recall_{Perkantoran} = \frac{19}{7+1+0+19} \times 100\% = 70,37\%$$

$$Recall_{Pemasaran} = \frac{12}{0+1+0+12} \times 100\% = 92,31\%$$

$$Recall_{Perhotelan} = \frac{5}{0+0+0+5} \times 100\% = 100\%$$

$$Recall_{rata-rata} = \frac{(88,24+70,37+92,31+100)\%}{4} = 87,73\%$$

$$Laju Error = \frac{FP+FN}{TP+FN+FP+TN} \times 100\%$$

$$Laju Error = \frac{1+5+0+7+1+0+1+0+0+0+0+0}{45+1+5+0+7+19+1+0+0+1+0+0+0+5} \times 100\% = 15,62\%$$

Pengujian menggunakan *confusion matrix* tidak hanya menggunakan perbandingan 70:30 tetapi juga menggunakan perbandingan 50:50, 60:40, 80:20 dan 90:10. Data siswa baru menghasilkan hasil dari pada sistem dari *rules* yang telah diperoleh. Berikut hasil pengujian data uji dari ketiga perbandingan tersebut pada Tabel 6.

Tabel 6. Hasil Pengujian menggunakan *Confusion Matrix*

No.	Perbandingan	Data Latih	Data Uji	Akurasi (%)	Precision (%)	Recall (%)	Laju Error (%)
1	50:50	160	160	43,14%	22,51%	41,31%	56,86%
2	60:40	192	128	53,33%	38,91%	41,62%	46,67%
3	70:30	224	96	84,38%	85,92%	87,73%	15,62%

No.	Perbandingan	Data Latih	Data Uji	Akurasi (%)	Precision (%)	Recall (%)	Laju Error (%)
4	80:20	256	64	84,13%	86,24%	84,7%	15,87%
5	90:10	288	32	70,97%	75%	69,44%	29,03%

4. KESIMPULAN

Berdasarkan hasil dari keseluruhan penelitian yang telah dilakukan, diperoleh hasil klasifikasi jurusan siswa baru menggunakan 320 data siswa baru tahun ajaran 2022/2023 dengan Algoritma CART, yaitu terdapat 153 siswa yang diklasifikasikan secara benar ke jurusan akuntansi, 62 siswa diklasifikasikan secara benar ke jurusan perkantoran, 12 siswa diklasifikasikan secara benar ke jurusan pemasaran, dan 52 siswa diklasifikasikan secara benar ke jurusan perhotelan. Terdapat 41 data yang hasilnya tidak sesuai dengan jurusan. Berdasarkan hasil pengujian menggunakan *Confusion Matrix* dengan tiga perbandingan data latih dan data uji diperoleh hasil akurasi yang berbeda, yaitu pada perbandingan 70:30 diperoleh hasil akurasi sebesar 84,38%. Pada perbandingan 80:20 diperoleh hasil akurasi sebesar 82,81%. Kemudian, untuk perbandingan 90:10 diperoleh hasil akurasi sebesar 70,97%. Dari hasil tersebut, dapat disimpulkan bahwa Algoritma CART memiliki hasil akurasi tertinggi pada pengujian menggunakan perbandingan 70:30 dengan jumlah 224 data *training* dan 96 data *testing* yaitu sebesar 84,38%.

REFERENCES

- [1] D. Pristiwanti, B. Badariah, S. Hidayat, dan R. S. Dewi, "Pengertian Pendidikan," *J. Pendidik. Dan Konseling*, vol. 4, no. 6, hal. 1707–1715, 2022.
- [2] B. Hadi, "Fenomena Learning Loss pada Pendidikan Sekolah Menengah Kejuruan di Indonesia," *Edudikara J. Pendidik. dan Pembelajaran*, vol. 6, no. 4, hal. 290–296, 2022, doi: 10.32585/edudikara.v6i4.262.
- [3] R. Pratiwi, "Perbandingan Klasifikasi Algoritma C5.0 Dan Classification And Regression Tree (Studi Kasus: Data Sosial Kepala Keluarga Masyarakat Desa Teluk Baru Kecamatan Muara Ancalong Tahun 2019)," *Ilmu Mat. dan Terap.*, vol. 14, No.2, hal. 267–278, 2020.
- [4] S. Monalisa dan F. Hadi, "Penerapan Algoritma CART Dalam Menentukan Jurusan Siswa di MAN 1 Inhil," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, hal. 387–394, 2020, doi: 10.32736/sisfokom.v9i3.932.
- [5] F. Haris dan Kurniati, "Penerapan Algoritma Classification and Regression Tree (Cart) Untuk Klasifikasi Jurusan Siswa Baru Man 1 Oku Timur," hal. 331–339, 2020, [Daring]. Tersedia pada: http://repository.umrah.ac.id/219/1/JURNAL_ZUMROTUL_MUTIAH.pdf
- [6] F. D. Savitri, "Penerapan Metode Cart Dalam Memprediksi Penjualan Produk Fast Moving Dan Slow Moving," vol. 1, no. 4, hal. 119–125, 2022.
- [7] P. Metode, P. Bekerja, K. Subang, I. Nabila, I. M. Sumertajaya, dan M. Raharjo, "Penerapan Metode CART pada Pengklasifikasian Bekerja dan Pengangguran di Kabupaten Subang *," vol. 11, no. 2, hal. 120–129, 2022.
- [8] I. W. Misshuari dan Chairunisah, "Penerapan Metode Classification and Regression Trees (Cart) Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Pembayaran Kredit Oleh Nasabah Di Pt Bprs Gebu Prima Medan," *Karismatika*, vol. 6, no. 3, hal. 10–20, 2020.
- [9] S. Kasus, D. Alumni, F. Teknik, dan U. Bengkulu, "Analisis Prediksi Masa Studi Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritme Cart (Classification And Regression Trees)," vol. 10, no. 1, 2022.
- [10] F. GUSTRIANDINI, "Implementasi Klasifikasi Penyakit Ginjal Kronik Menggunakan Decision Tree Algoritma Classification And Regression Trees (Cart)," 2023.
- [11] A. Aribowo, R. Kuswandhie, dan Y. Primadasa, "Implementasi Algoritma CART Dalam Penentuan Kelayakan Penerima Bantuan PKH Di Desa Ngadirejo Implementation of the CART Algorithm in Determining the," vol. 7, no. 1, hal. 40–51, 2021.
- [12] N. November dan E. R. D. Gui, "Sistem Cerdas Untuk Penentuan Pohon Keputusan Bakat dan Minat Anak Menggunakan Algoritma CART," vol. 1, no. 3, hal. 737–751, 2022.
- [13] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Yogyakarta: CV BUDI UTAMA, 2020.
- [14] A. Srirahayu dan L. S. Pribadie, "Review Paper Data Mining Klasifikasi Data Mining," vol. 14, no. April, 2023.
- [15] C. Pada, D. Karyawan, dan B. Promosi, "ANALISIS PERBANDINGAN ALGORITMA DECISION TREE C4 . 5 DAN," vol. 17, hal. 567–582, 2023.
- [16] I. Fida *dkk.*, "Implementasi Algoritma Cart Dalam Klasifikasi Penyakit Diabetes," hal. 1–8, 2023.
- [17] F. R. Z. B. R. B. Putra, *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia, 2023. [Daring]. Tersedia pada: [DATA_MINING_Algoritma_dan_Penerapannya/zLHGEEAAQBAJ?hl=id&gbpv=0](https://www.sonpedia.com/Data_Mining_Algoritma_dan_Penerapannya/zLHGEEAAQBAJ?hl=id&gbpv=0)

- [18] R. Metivianis, D. E. Ratnawati, dan B. Rahayudi, “Analisis Sentimen pengguna Twitter terhadap Vaksinasi Sinovac dan AstraZeneca menggunakan Algoritma CART,” vol. 6, no. 4, hal. 1913–1920, 2022.
- [19] T. Untuk, M. Mengetahui, dan D. S. Suparno, “Pengenalan Pola Untuk Mengetahui Jumlah Target Pengunjung Mall Berdasarkan Usia , Gender , Pendapatan Pertahun , Target Pasar Menggunakan Metode EDA , K-Means , Hierarchial Clustering , Confusion Matrix,” vol. 3, no. 2, hal. 61–69, 2021.
- [20] A. . Lerner, *The 2x2 Matrix*. Springer Cham, 2022. [Daring]. Tersedia pada: <https://doi.org/10.1007/978-3-030-74920-0>