

Algoritma *Adaboost* pada Metode *Decision Tree* untuk Klasifikasi Kelulusan Mahasiswa

Yuveinsiana Crismayella¹, Neva Satyahadewi², Hendra Perdana^{3*}

^{1,2,3}Program Studi Statistika, Fakultas MIPA, Universitas Tanjungpura, Pontianak 78124, Indonesia

*Corresponding author. Email: hendra.perdana@math.untan.ac.id

ABSTRAK

Perguruan tinggi menyediakan pendidikan tinggi sebagai tolok ukur kualitas pendidikan dan mengevaluasi silabus pendidikan tinggi. Tingkat kelulusan dan kapasitas pendaftaran sangat penting untuk penilaian kelulusan dan pengambilan keputusan. Sayangnya, beberapa mahasiswa jurusan statistika tidak dapat menyelesaikan studinya tepat waktu sehingga berdampak pada akreditasi prodi. Untuk itu perlu dikaji karakteristik mahasiswa yang berhasil dan gagal menyelesaikan studinya tepat waktu dengan menggunakan metode klasifikasi data mining yaitu Algoritma C5.0. Dalam penelitian ini, algoritma *Adaboost* dan Algoritma C5.0 digunakan untuk mengklasifikasikan tingkat kelulusan secara akurat. Data kelulusan Program Studi Statistika Universitas Tanjungpura Angkatan 1 tahun ajaran 2021/2022 sampai angkatan II tahun ajaran 2022/2023 menjadi pertimbangan dalam penelitian ini. Pertama, diukur nilai entropi, nilai gain, dan rasio gain. Setelah itu, masing-masing data diberi bobot yang sama, dan dilakukan iterasi sebanyak 100 kali. Analisis menggunakan Algoritma C5.0 menunjukkan Akreditasi Sekolah sebagai variabel dengan gain ratio tertinggi, yang menunjukkan bahwa Akreditasi Sekolah memiliki pengaruh paling menentukan terhadap tingkat kelulusan dengan persentase akurasi 70%. Persentase ini kemudian meningkat menjadi 82,14% setelah dilakukan *boosting* menggunakan algoritma *Adaboost*. Algoritma *Adaboost* dinilai baik dalam meningkatkan akurasi algoritma C5.0. Hasil penelitian ini dapat memberikan wawasan bagi perguruan tinggi dalam merancang kebijakan peningkatan kelulusan tepat waktu berdasarkan faktor-faktor yang mempengaruhi kelulusan mahasiswa.

Kata Kunci:

Masa Studi; Algoritma C5.0; *Adaptive Boosting*; Algoritma *Adaboost*

ABSTRACT

Colleges provide higher education as the benchmark of education quality and evaluate higher education syllabi. Graduation rates and enrollment capacity are essential for graduation assessment and decision-making. Unfortunately, some students majoring in statistics failed to finish their studies on time, impacting the accreditation of the study program. It is necessary to examine the characteristics of students who managed and failed to complete their studies on time using the data mining classification method, namely Algorithm C5.0. In this study, *Adaboost* algorithm and Algorithm C5.0 was employed to classify graduation rates accurately. Graduation data of the Statistics Study Program of Universitas Tanjungpura Batch 1 of 2021/2022 to Batch II of 2022/2023 School years were regarded in this study. First, the entropy, gain, and gain ratio values were measured. After that, each data was given equal weight, and iteration was performed 100 times. The analysis using Algorithm C5.0 showed School Accreditation as the variable with the highest gain ratio, indicating that School Accreditation has the most decisive influence on graduation rates with an accuracy percentage of 70%. This percentage then

increased to 82.14% after the boosting using the Adaboost algorithm. Adaboost Algorithm is regarded as good in improving the accuracy of algorithm C5.0. The results of this study can provide insight for colleges in designing policies to increase on-time graduation based on the factors that influence student graduation.

Keywords:

Study Period; C5.0 Algorithm; Adaptive Boosting; Adaboost Algorithm

Style Sitasi:

Y. Crismayella, N. Satyahadewi, and H. Perdana, "Algoritma Adaboost pada Metode Decision Tree untuk Klasifikasi Kelulusan Mahasiswa", *Jambura J. Math.*, vol. 5, No. 2, pp. 278–288, 2023, doi: <https://doi.org/10.34312/jjom.v5i2.18790>

1. Pendahuluan

Universitas merupakan tempat pendidikan akademik bagi mahasiswa [1]. Tingkat kelulusan mahasiswa selama menempuh pendidikan menjadi penentu kualitas suatu perguruan tinggi. Mahasiswa menjadi tolak ukur yang digunakan untuk menilai kualitas dan mengevaluasi rencana pembelajaran perguruan tinggi. Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Nomor 44 tahun 2015 menyatakan bahwa masa studi maksimal adalah tujuh tahun, dengan IPK di atas 2,0 dan SKS minimal untuk program sarjana adalah 144 [2].

Universitas Tanjungpura (Untan) merupakan universitas negeri yang terletak di Pontianak, Kalimantan Barat, dimana Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) terdapat program studi Statistika. Mahasiswa program sarjana FMIPA Untan dinyatakan lulus tepat waktu jika dapat menyelesaikan studinya dalam waktu maksimal empat tahun dengan beban minimal 144 sks. Mahasiswa yang berisiko lulus tidak tepat waktu dapat dianalisis pada evaluasi keberhasilan mahasiswa pada akhir empat semester pertama. Pada Program Studi Statistika masih banyak ditemukan mahasiswa yang lulus tidak tepat waktu sehingga akan berpengaruh terhadap penilaian akreditasi. Oleh karena itu, karakteristik mahasiswa yang menyelesaikan kuliah tepat waktu maupun tidak dalam penentuan kelulusan mahasiswa dapat dianalisis menggunakan teknik dalam *data mining*.

Data mining merupakan proses pemanfaatan teknik statistik dalam mencari informasi atau pola dari berbagai *database* yang besar dengan cara mengidentifikasi dan mengekstrak pola yang berguna serta menarik. *Data mining* adalah bagian dari *Knowledge Discovery in Database* (KDD) [3]. Ketepatan dalam memilih algoritma atau metode sangat mempengaruhi keseluruhan tujuan dan proses KDD. Menurut Mustafa, *et al.* [4], *data mining* memiliki enam fungsi yaitu deskripsi, klasifikasi, pengelompokan, prediksi, estimasi, dan asosiasi. Metode klasifikasi *decision tree* sering digunakan dalam *data mining*. Dalam *data mining*, Algoritma C5.0 secara khusus digunakan pada *decision tree* [5]. *Decision tree* dapat mengatasi masalah *overfitting* pada jaringan saraf, memilih atribut yang tepat, mengatasi atribut kontinu, mengatasi nilai atribut pada data pelatihan yang hilang, dan meningkatkan efisiensi komputasi. Namun, derajat ketidakseimbangan kelas yang tinggi menjadi kelemahan metode *Decision tree*.

Distribusi kelas yang tidak seimbang dicirikan dengan memiliki lebih banyak kasus daripada beberapa kelas yang lainnya. Masalah ketidakseimbangan merupakan masalah dimana terdapat perbedaan antara kelas positif dan negatif secara signifikan [6]. Ketidakseimbangan kelas mempengaruhi kinerja klasifikasi, sehingga diperlukan

metode yang dapat mengatasinya sehingga dapat mencapai dan meningkatkan kinerja klasifikasi *Decision Tree* yang lebih baik. *Adaptive Boosting (Adaboost)* adalah algoritma *supervised* dalam *data mining* yang digunakan dalam membangun model klasifikasi dan dapat menangani ketidakseimbangan kelas. Algoritma *Boosting* memberikan bobot pada distribusi *training data* dalam setiap iterasi dengan nilai yang berbeda.

Penelitian terkait penentuan kelulusan mahasiswa telah dilakukan di Universitas Pamulang tahun 2019 [7]. Penelitian terhadap 5.610 objek ini dilakukan dengan mengimplementasikan *bagging* dan pada metode klasifikasi *Gradient Boosted Trees (GBT)*. Atribut yang digunakan adalah Usia Mahasiswa, Prodi, Jenis Kelamin, IPS semester satu hingga tujuh, Jenis reguler, dan Status Kelulusan. Berdasarkan hasil penelitian diperoleh nilai akurasi 80,57% dan AUC 0,858 pada *split* 90:10 dengan kategori klasifikasi baik. Sehingga terbukti bahwa pengimplementasian *SMOTE* dan *bagging* dapat mengatasi masalah ketidakseimbangan kelas dalam penentuan kelulusan mahasiswa dan meningkatkan kinerja klasifikasi model *GBT*. Selain itu, metode *Adaboost* juga diimplementasikan kasus prediksi kinerja perusahaan dengan membandingkan antara metode *K-Nearest Neighbor* dan *Adaboost* [8], kasus klasifikasi penerima program bantuan sosial dengan bantuan metode *Decision Tree* [9], kasus *imbalance dataset* diabetes dengan bantuan algoritma *K-Nearest Neighbor (KNN)* untuk klasifikasi [10] dan kasus klasifikasi status gizi buruk balita dengan bantuan algoritma *C4.5* [11].

Penelitian terkait klasifikasi kelulusan mahasiswa dengan menggunakan Algoritma *Adaboost* dan metode *Decision Tree* Algoritma *C5.0* belum pernah dilakukan. Sehingga penelitian tentang penerapan Algoritma *Adaboost* pada Metode *Decision Tree* dalam klasifikasi kelulusan mahasiswa Program Studi Statistika Untan menarik untuk dilakukan. Penelitian ini bertujuan menerapkan Algoritma *Adaboost* pada metode *Decision Tree* dan menentukan nilai akurasinya dalam klasifikasi kelulusan mahasiswa.

2. Metode

Proses klasifikasi dilakukan menggunakan Algoritma *C5.0* dan dilakukan *boosting* menggunakan Algoritma *Adaboost*. Kedua metode ini diterapkan pada data kelulusan mahasiswa Program Studi Statistika Universitas Tanjungpura dengan jumlah data 140 yang terdiri dari sepuluh atribut independen dan satu atribut dependen. Dalam penelitian ini, *software R* digunakan untuk menganalisis data. Terdapat tiga tahap dalam analisis data, yaitu praproses data, pemodelan, dan evaluasi model. Pada tahap praproses data, data mentah diubah menjadi data yang siap untuk dianalisis. Pada tahap pemodelan, model dibangun dan diperoleh hasil prediksi, sedangkan tahap evaluasi model digunakan untuk mengukur kinerja model. Berikut langkah-langkah analisis data:

1. Praproses Data
 - (a) Mempersiapkan data kelulusan mahasiswa Program Studi Statistika Universitas Tanjungpura
 - (b) Melakukan kategorisasi pada atribut dependen dan independen
2. Pemodelan
 - (a) Membangun model klasifikasi berupa pohon klasifikasi Algoritma *C5.0*
 - (b) Melakukan proses *boosting* menggunakan Algoritma *Adaboost* dengan jumlah iterasi maksimum adalah 100.
3. Evaluasi model *Confusion matrix* digunakan untuk mengevaluasi model. *Confusion*

matrix adalah tabulasi silang antara informasi kelas *negative* dan *positive* yang terdapat pada kelas prediksi dan kelas sebenarnya [12]. Kelas positif dalam penelitian ini adalah mahasiswa dengan status kelulusan tepat waktu. Sedangkan kelas negatif adalah mahasiswa dengan status kelulusan tidak tepat waktu. Berdasarkan *confusion matrix*, dilakukan perhitungan nilai akurasi. Nilai akurasi dapat diklasifikasikan menjadi lima, yaitu < 60,00% (klasifikasi gagal), 60,01% – 70,00% (klasifikasi buruk), 70,01% – 80,00% (klasifikasi cukup), 80,01% – 90,00% (klasifikasi bagus), dan 90,01% – 100,00% (klasifikasi sangat bagus).

2.1. Algoritma C5.0

Data yang telah melewati tahap *preprocessing*, dieksplorasi menggunakan *decision tree*. C5.0 merupakan algoritma dalam *data mining* yang digunakan untuk membangun *decision tree*. C5.0 adalah pengembangan dari algoritma ID3 dan C4.5. Bila dibandingkan dengan algoritma klasifikasi lain, C5.0 memiliki kelebihan dalam memberikan waktu eksekusi yang lebih singkat dan tingkat akurasi terbaik [13]. Algoritma ini dimulai dengan menjadikan semua data akar dan atribut yang dipilih dari pohon klasifikasi menjadi pembagi untuk sampel tersebut. Berikut tahapan dalam Algoritma C5.0:

1. Perhitungan nilai *entropy*

$$Entropy(S) = - \sum_{i=1}^N P_i \log_2 P_i \tag{1}$$

dimana

S : Kumpulan data

N : Banyaknya kelas dari variabel dependen

P_i : Perbandingan banyaknya data kelas ke- i dari kumpulan data.

2. Menghitung *split info* setiap kelas pada variabel dan *gain* menggunakan persamaan (2) dan (3):

$$SplitInfo(S, A_i) = - \sum_{i=1}^k \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \tag{2}$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropy(S) \tag{3}$$

dengan

S : Jumlah seluruh sampel dalam himpunan data

A_i : Variabel independen ke- i

$|S_i|$: Jumlah sampel untuk kategori ke- i , dan

k : Banyaknya kategori pada variabel independen A.

3. Dari setiap atribut dihitung nilai *gain rationya*. *Node parent* dipilih dari atribut dengan nilai *gain ratio* tertinggi dan dibuat cabang dari setiap kategori *node parent*. *Gain ratio* dihitung menggunakan persamaan (4):

$$Gain Ratio = \frac{Gain(S, A)}{SplitInfo(S, A_i)} \tag{4}$$

dengan

- $Gain(S, A)$: Informasi *gain* pada variabel independen A, dan
 $SplitInfo(S, A_i)$: *Split* informasi pada variabel independen A kelas ke- i .

2.2. Algoritma Adaptive Boosting (Adaboost)

Algoritma *Adaboost* adalah varian dari algoritma *boosting* [14]. *Adaboost* menggabungkan sekumpulan pengklasifikasi sederhana (lemah) untuk membuat pengklasifikasi yang kuat. Pada awal klasifikasi, bobot yang sama diberikan kepada setiap sampel dalam Algoritma *Adaboost*. Setelah setiap klasifikasi, terjadi penambahan pada bobot hasil yang salah dan pengurangan pada bobot hasil yang benar. Proses ini akan terus diulangi sampai iterasi maksimum. Langkah-langkah dalam Algoritma *Adaboost* adalah sebagai berikut:

1. Kumpulan himpunan data N sampel dengan dua kelas berlabel $y \in \{0, 1\}$.
2. Dalam himpunan data, bobot awal setiap sampel ditetapkan sama, yaitu

$$w_i^1 = \frac{1}{N} \tag{5}$$

dengan

- w_i^1 = Bobot awal pada iterasi pertama untuk semua sampel, dan
 $i = 1, 2, \dots, N$.

3. Untuk semua iterasi $t = 1, 2, \dots, T$ dengan T adalah iterasi maksimum, lakukan proses berikut:
 - (a) Temukan model prediktif $f(x)$ dari himpunan data. Model prediktif $f(x)$ diterapkan ke himpunan data. Jika pengklasifikasian sampel salah, maka "salah = 1" dan jika "benar = 0".
 - (b) Penjumlahan kesalahan pembobotan dari semua sampel dihitung dengan persamaan (6) sebagai berikut:

$$error^t = \sum_{i=1}^N (w_i^t \times error_i^t). \tag{6}$$

- (c) Indeks kepercayaan dari *learner* $f(x)$ dihitung sebagai dengan persamaan (7) sebagai berikut:

$$\alpha^t = \frac{1}{2} \ln \left(\frac{1 - error^t}{error^t} \right). \tag{7}$$

Dalam hal ini, indeks kepercayaan dari model prediktif $f(x)$ bergantung pada kesalahan pembobotan.

- (d) Memperbarui bobot untuk semua *training* sampel asli mengikuti persamaan (8):

$$w_i^{t+1} = w_i^t \times \begin{cases} \exp(-\alpha^t) & \text{untuk menurunkan bobot} \\ \exp(\alpha^t) & \text{untuk menaikkan bobot.} \end{cases} \tag{8}$$

Bila sampel diklasifikasikan benar, maka bobotnya akan menurun, sedangkan jika sampel diklasifikasikan salah, maka bobotnya meningkat.

- (e) Selanjutnya, dilakukan normalisasi bobot dengan persamaan (9):

$$w_i^{*t} = \frac{w_i^t}{\sum_{i=1}^N w_i^t} \tag{9}$$

sehingga,

$$\sum_i^N w_i^{*t+1} = 1.$$

- (f) Jika $error < 0,5$, atau $t < T$, ulangi langkah 1-5; jika tidak maka proses berhenti.
- (g) Model prediksi $f^t(x)$ dengan $t = 1, 2, \dots, T$ diperoleh setelah T iterasi. Peramalan akhir untuk kasus j ditentukan oleh peramalan model T yang digabungkan dengan menggunakan pendekatan *voting*, sesuai dengan persamaan (10):

$$y_j = \text{sign} \sum_{t=1}^T \alpha^t f^t(x) \tag{10}$$

dengan sign = tanda yang digunakan untuk melakukan *voting*.

2.3. Evaluasi Model Klasifikasi

Confusion matrix digunakan untuk mengevaluasi kinerja model klasifikasi. Tabel 1 adalah *confusion matrix* yang menunjukkan *true positive* (TP), *false positive* (FP), *true negative* (TN), dan *false negative* (FN). TP merupakan klasifikasi kasus positif yang hasil prediksinya adalah positif, FN merupakan klasifikasi kasus positif yang hasil prediksinya adalah negatif, FP merupakan klasifikasi kelas negatif yang hasil prediksinya adalah positif dan TN merupakan klasifikasi kasus negatif yang hasil prediksinya adalah negatif [13]. Evaluasi menggunakan nilai akurasi dilakukan berdasarkan nilai yang terdapat pada *confusion matrix* berdasarkan persamaan (11) dan Tabel 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%. \tag{11}$$

Tabel 1. *Confusion Matrix*

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

Akurasi (*accuracy*) merupakan metrik evaluasi hasil klasifikasi. Kelas mayoritas (*negative*) akan mendominasi jika hanya menggunakan akurasi saat bekerja dengan kumpulan data yang tidak seimbang [15].

3. Hasil dan Pembahasan

3.1. Praproses Data

Data dalam penelitian ini merupakan data kelulusan mahasiswa Program Studi Statistika Untan Periode I Tahun Ajaran 2017/2018 sampai Periode II Tahun Ajaran 2022/2023. Jumlah sampel yang digunakan adalah sebanyak 140 sampel. Atribut dependen yang digunakan adalah status kelulusan mahasiswa (Y), sedangkan atribut independennya tersaji pada Tabel 2. Seluruh atribut dikategorikan ke dalam dua kategori.

Diketahui bahwa jumlah mahasiswa yang tidak lulus tepat waktu adalah 84 orang dan

Tabel 2. Pengklasifikasian atribut dependen dan independen

No	Atribut	Kategori	Jumlah	Persentase
1	Status Kelulusan	1. Tidak Tepat Waktu	84	60,00%
		2. Tepat Waktu	56	40,00%
2	Jenis Kelamin	1. Laki-laki	31	22,10%
		2. Perempuan	109	77,90%
3	IPK Semester 1	1. < 3	52	37,10%
		2. ≥ 3	88	62,90%
4	IPK Semester 2	1. < 3	61	43,60%
		2. ≥ 3	79	56,40%
5	IPK Semester 3	1. < 3	68	48,60%
		2. ≥ 3	72	51,40%
6	IPK Semester 4	1. < 3	65	46,40%
		2. ≥ 3	75	53,60%
7	Daerah Asal Domisili	1. Kabupaten	100	71,40%
		2. Kota	40	28,60%
8	Akreditasi Sekolah	1. A	31	22,10%
		2. Selain A	109	77,90%
9	Jalur Masuk	1. SNMPTN	84	60,00%
		2. Selain SNMPTN	56	40,00%
10	Beasiswa	1. Tidak	78	55,70%
		2. Ya	62	44,30%
11	Status Kelulusan Tes TUTEPE Pertama	1. Lulus	12	8,60%
		2. Tidak	123	87,90%

56 orang tepat waktu. Jumlah laki-laki adalah sebanyak 31 orang dan perempuan 109 orang. Atribut IPK menggunakan IPK dari semester 1 hingga semester 4 dan dikategorikan menjadi dua yaitu < 3 dan ≥ 3 . Untuk atribut Daerah Asal Domisili dikategorikan menjadi dua yaitu kabupaten dan kota dimana alumni tersebut berasal saat masih kuliah. Atribut Akreditasi Sekolah dikategorikan menjadi dua yaitu A dan selain A.

3.2. Pemodelan

Bagian ini menampilkan hasil pemodelan dengan menggunakan Algoritma C5.0 dan algoritma *Adaboost*.

3.2.1. Pemodelan menggunakan Algoritma C5.0

Pemodelan Algoritma C5.0 diawali dengan menghitung nilai *entropy*. Setelah itu dilanjutkan dengan menghitung nilai *gain* dan *gain ratio* dari sepuluh atribut yang digunakan untuk membangun pohon klasifikasi. *Node parent* ditentukan dari atribut yang memiliki nilai *gain ratio* paling tinggi. Dari *node parent* tersebut, dibuat cabang dari kategori *node parent*. Lalu diperiksa apakah masih terdapat atribut yang tersisa. Jika kondisi masih, maka ulangi proses dari menghitung nilai *entropy*. Jika kondisi tidak, maka dilanjutkan dengan membangun *Decision Tree* Algoritma C5.0 dan melakukan interpretasi hasil dari *Decision Tree* yang telah dibangun serta perhitungan nilai akurasi berdasarkan *confusion matrix*. Hasil perhitungan nilai *entropy*, *gain*, dan *gain ratio* disajikan pada Tabel 3.

Berdasarkan Tabel 3 dapat diketahui bahwa dari sepuluh atribut independen yang

Tabel 3. Perhitungan nilai *entropy*, *gain*, dan *gain ratio*

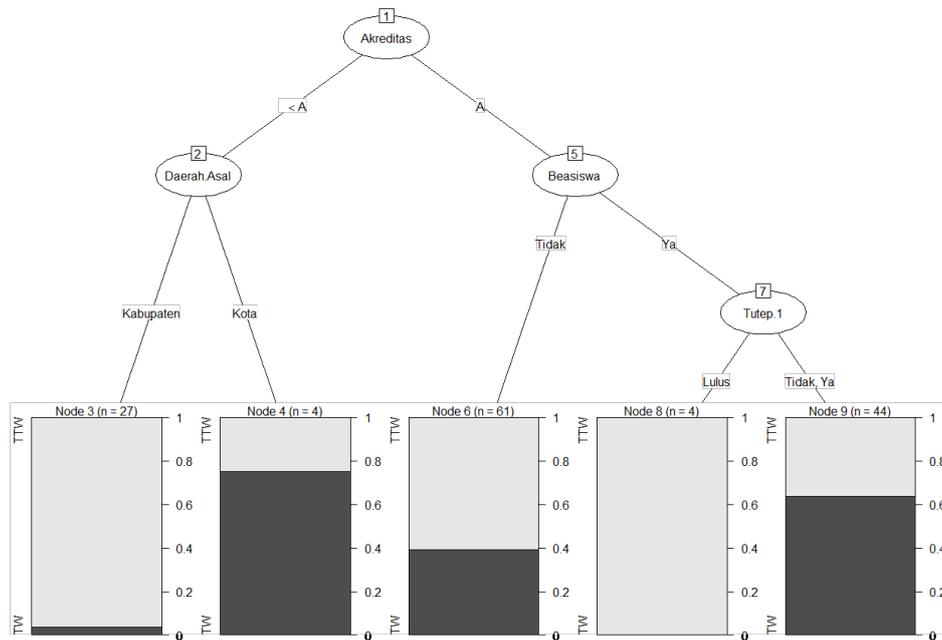
Atribut	Kategori	Entropy	Gain	Gain Ratio
Jenis Kelamin	1. Laki-laki	0,96124	0,00042	0,00022
	2. Perempuan	0,97602		
IPK 1	1. < 3	0,95593	0,00122	0,00063
	2. ≥ 3	0,98038		
IPK 2	1. < 3	0,92312	0,01087	0,00567
	2. ≥ 3	0,99498		
IPK 3	1. < 3	0,90770	0,01552	0,00815
	2. ≥ 3	0,99679		
IPK 4	1. < 3	0,94268	0,01189	0,00612
	2. ≥ 3	1,00000		
Daerah Asal	1. Kabupaten	0,96290	0,00014	0,00007
	2. Kota	0,97306		
Akreditasi	1. A	0,99848	0,07072	0,04553
	2. Selain A	0,55478		
Jalur Masuk	1. SNMPTN	0,99632	0,00825	0,00426
	2. Selain SNMPTN	0,94029		
Beasiswa	1. Tidak	0,99700	0,01095	0,00568
	2. Ya	0,93059		
TUTEK 1	1. Lulus	0,81128	0,04141	0,02313
	2. Tidak	0,97887		

digunakan dalam penelitian ini, atribut akreditasi sekolah memiliki nilai *gain ratio* yaitu 0,04553. Artinya, atribut akreditasi sekolah memiliki pengaruh paling besar dalam klasifikasi kelulusan mahasiswa, sedangkan atribut IPK semester 1 (X_2), IPK semester 2 (X_3), IPK semester 3 (X_4), IPK semester 4 (X_5), dan Jalur Masuk (X_8) tidak terlalu mempengaruhi hasil prediksi dalam klasifikasi. Untuk membentuk Algoritma C5.0, diperlukan perhitungan *entropy*, *gain*, dan *gain ratio* menggunakan Persamaan (1), (3), dan (4), menghasilkan bagan *Decision Tree* seperti pada Gambar 1.

Berdasarkan Gambar 1 ditarik kesimpulan:

1. Jika mahasiswa tersebut berasal dari sekolah dengan akreditasi selain A dan daerah asalnya adalah kabupaten, maka mahasiswa tersebut diklasifikasikan status kelulusannya tidak tepat waktu.
2. Jika mahasiswa tersebut berasal dari sekolah dengan akreditasi selain A dan daerah asalnya adalah kota, maka mahasiswa tersebut diklasifikasikan status kelulusannya tepat waktu.
3. Jika mahasiswa tersebut berasal dari sekolah dengan akreditasi A dan tidak pernah menerima beasiswa saat kuliah, maka mahasiswa tersebut diklasifikasikan status kelulusannya tidak tepat waktu.
4. Jika mahasiswa tersebut berasal dari sekolah dengan akreditasi A, pernah menerima beasiswa saat kuliah dan lulus TUTEK yang pertama, maka mahasiswa tersebut diklasifikasikan status kelulusannya tidak tepat waktu.
5. Jika mahasiswa tersebut berasal dari sekolah dengan akreditasi A, pernah menerima beasiswa saat kuliah dan tidak lulus TUTEK yang pertama, maka mahasiswa tersebut diklasifikasikan status kelulusannya tepat waktu.

Setelah *decision tree* Algoritma C5.0 dibangun, maka dilakukan evaluasi model menggunakan *confusion matrix*. Data awal diklasifikasikan berdasarkan *decision tree*



Gambar 1. Decision Tree Algoritma C5.0

Algoritma C5.0 yang telah dibangun. Diperoleh bahwa data awal yang benar diklasifikasikan sebagai Tidak Tepat Waktu adalah 67 orang dan yang benar diklasifikasikan sebagai Tepat Waktu adalah 31 orang. Sedangkan data awal dengan status kelulusan Tidak Tepat Waktu dan diklasifikasikan sebagai Tepat Waktu adalah 17 orang. Data awal dengan status kelulusan Tepat Waktu dan diklasifikasikan sebagai Tidak Tepat Waktu adalah 25 orang. Hasil selengkapnya ditunjukkan pada Tabel 4.

Tabel 4. Confusion matrix Algoritma C5.0

		Actual	
		Tidak Tepat Waktu	Tepat Waktu
Predicted	Tidak Tepat Waktu	67	25
	Tepat Waktu	17	31

Berdasarkan hasil dari evaluasi model klasifikasi Algoritma C5.0 menggunakan *confusion matrix* pada Tabel 4, diperoleh nilai akurasi menggunakan Persamaan (11) sebesar 70%. Selanjutnya, dilakukan pemodelan dengan menggunakan Algoritma *Adaboost*.

3.2.2. Pemodelan menggunakan Algoritma *Adaboost*

Nilai inisialisasi bobot dari data pada iterasi pertama menggunakan Persamaan (5) dengan iterasi maksimalnya 100 adalah sebesar 0,007143. Ditemukan data yang tidak sesuai dengan kelas aslinya pada iterasi pertama adalah 42 data. Maka selanjutnya adalah menghitung kesalahan data penelitian menggunakan Persamaan (6) diperoleh nilai kesalahan data untuk iterasi pertama adalah sebesar 0,300006. Setelah menghitung kesalahan pada data penelitian, langkah selanjutnya adalah menetapkan bobot data menggunakan Persamaan (7). Diperoleh nilai bobot datanya adalah sebesar 0,423635. Menggunakan Persamaan (8), dilakukan *update* bobot data. Tabel 5 adalah hasil *update*

bobot pada iterasi pertama dimana TTW adalah status kelulusan Tidak Tepat Waktu, sedangkan TW adalah status kelulusan Tepat Waktu. Diperoleh bobot data yang status awalnya TTW dan benar diklasifikasikan sebagai TTW maupun status awalnya TW dan benar diklasifikasikan sebagai TW adalah sebesar 0,00468. Bobot data yang status awalnya TTW dan diklasifikasikan sebagai TW maupun sebaliknya adalah sebesar 0,010911..

Tabel 5. Hasil *Update* Bobot

No	Status	Klasifikasi	Bobot awal	<i>Update</i> bobot
1	TTW	TTW	0,0071	0,00468
2	TTW	TTW	0,0071	0,00468
3	TTW	TTW	0,0071	0,00468
4	TTW	TW	0,0071	0,01091
...
137	TW	TTW	0,0071	0,01091
138	TTW	TTW	0,0071	0,00468
139	TTW	TTW	0,0071	0,00468
140	TW	TW	0,0071	0,00468

Setelah dilakukan *update* pada bobot data, maka langkah selanjutnya adalah melakukan normalisasi bobot menggunakan Persamaan (9). Diperoleh hasil normalisasi bobot untuk data yang diklafifikasikan benar adalah sebesar 0,00510 dan bobot untuk data yang salah klasifikasi adalah sebesar 0,01190. Perhitungan terus diulangi hingga nilai *error* $\geq 0,5$ dan iterasi telah mencapai iterasi maksimum. Setelah itu proses dapat berhenti. Selanjutnya adalah melakukan evaluasi model dari Algoritma *Adaboost* menggunakan *confusion matrix*. Data klasifikasi Algoritma C5.0 diklasifikasikan kembali berdasarkan hasil *boosting* menggunakan Algoritma *Adaboost*. Diperoleh bahwa data yang benar diklasifikasikan sebagai Tidak Tepat Waktu adalah 74 orang dan yang benar diklasifikasikan sebagai Tepat Waktu adalah 41 orang. Sedangkan data dengan status kelulusan Tidak Tepat Waktu dan diklasifikasikan sebagai Tepat Waktu adalah 10 orang. Data awal dengan status kelulusan Tepat Waktu dan diklasifikasikan sebagai Tidak Tepat Waktu adalah 15 orang. Selanjutnya adalah melakukan evaluasi model dari Algoritma *Adaboost* menggunakan *confusion matrix* yang hasilnya disajikan pada Tabel 6.

Tabel 6. *Confusion matrix* Algoritma *Adaboost*

		<i>Actual</i>	
		Tidak Tepat Waktu	Tepat Waktu
<i>Predicted</i>	Tidak Tepat Waktu	74	15
	Tepat Waktu	10	41

Berdasarkan hasil dari evaluasi model klasifikasi Algoritma C5.0 setelah di *boosting* menggunakan Algoritma *Adaboost* menggunakan *confusion matrix* pada Tabel 6, diperoleh nilai akurasi menggunakan Persamaan (10) sebesar 82,14%. Artinya, terjadi peningkatan akurasi dari hasil yang diperoleh dengan algoritma C5.0. Nilai akurasi meningkat sebesar 12,14% setelah di lakukan *boosting* dengan menggunakan algoritma *Adaboost*. Nilai akurasi ini tergolong bagus karena dapat memprediksi hasil klasifikasi untuk kedua kelas ke kategori tidak tepat waktu dan tepat waktu berdasarkan *dataset* yang sudah ada dan terbukti bahwa Algoritma *Adaboost* dapat meningkatkan kinerja Algoritma C5.0.

4. Kesimpulan

Dari penelitian ini dapat disimpulkan bahwa penerapan Algoritma *Adaboost* pada Metode *Decision Tree* dalam klasifikasi kelulusan mahasiswa Program Studi Statistika Untan tergolong bagus dalam meningkatkan akurasi Algoritma C5.0. Algoritma C5.0 mampu mengklasifikasikan dataset ke dalam kategori tepat waktu dan tidak tepat waktu, sedangkan Algoritma *Adaboost* terbukti mampu meningkatkan kinerja dari Algoritma C5.0 dan dari hasil pengujian 140 *dataset* kelulusan mahasiswa Program Studi Statistika Untan. Dengan menggunakan sepuluh atribut, diperoleh akurasi sebesar 70% dengan Algoritma C5.0. Kemudian setelah dilakukan *boosting* menggunakan Algoritma *Adaboost*, akurasi meningkat sebesar 12,14% menjadi 82,14%. Hasil penelitian ini diharapkan dapat memberikan insight bagi perguruan tinggi dalam merancang kebijakan untuk meningkatkan kelulusan tepat waktu berdasarkan faktor-faktor yang mempengaruhi kelulusan mahasiswa.

Referensi

- [1] M. Kadafi, "Metode naïve bayes classifier (nbc) untuk memprediksi tingkat kelulusan mahasiswa tepat waktu," *JSI: Jurnal Sistem Informasi (E-Journal)*, vol. 12, no. 2, pp. 8–17, 2020, doi: 10.36706/jsi.v12i2.12179.
- [2] H. Hartatik, "Optimasi model prediksi kelulusan mahasiswa menggunakan algoritma naive bayes," *Indonesian Journal of Applied Informatics*, vol. 5, no. 1, p. 32, 2021, doi: 10.20961/ijai.v5i1.44379.
- [3] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, "Klasifikasi dengan pohon keputusan berbasis algoritme c4.5." Universitas negeri Semarang, 2020, pp. 64–71.
- [4] M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, "Implementasi data mining untuk evaluasi kinerja akademik mahasiswa menggunakan algoritma naive bayes classifier," *Creative Information Technology Journal*, vol. 4, no. 2, pp. 151–162, 2018, doi: 10.24076/citec.2017v4i2.106.
- [5] F. N. Umma, B. Warsito, and D. A. I. Maruddani, "Klasifikasi status kemiskinan rumah tangga dengan algoritma c5.0 di kabupaten pemalang," *Jurnal Gaussian*, vol. 10, no. 2, pp. 221–229, 2021, doi: 10.14710/j.gauss.v10i2.29934.
- [6] Y. Pristyanto, "Penerapan metode ensemble untuk meningkatkan kinerja algoritme klasifikasi pada imbalanced dataset," *Jurnal Teknoinfo*, vol. 13, no. 1, pp. 11–16, 2019, doi: 10.33365/jti.v13i1.184.
- [7] A. Bisri and R. Rachmatika, "Integrasi gradient boosted trees dengan smote dan bagging untuk deteksi kelulusan mahasiswa," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 8, no. 4, pp. 309–314, 2019, doi: 10.22146/jnteti.v8i4.529.
- [8] A. I. Prianti, R. Santoso, and A. R. Hakim, "Perbandingan metode k-nearest neighbor dan adaptive boosting pada kasus klasifikasi multi kelas," *Jurnal Gaussian*, vol. 9, no. 3, pp. 346–354, 2020, doi: 10.14710/j.gauss.v9i3.28924.
- [9] L. Qadrini, A. Seppewali, and A. Aina, "Decision tree dan adaboost pada klasifikasi penerima program bantuan sosial," *J. Inovasi Penelitian*, vol. 2, no. 7, pp. 1959–1966, 2021.
- [10] N. Novianti, M. Zarlis, and P. Sihombing, "Penerapan algoritma adaboost untuk peningkatan kinerja klasifikasi data mining pada imbalance dataset diabetes," *Jurnal Media Informatika Budidarma*, vol. 6, no. 2, pp. 1200–1206, 2022, doi: 10.30865/mib.v6i2.4017.
- [11] W. Wahyudi, "Optimasi klasifikasi status gizi balita berdasarkan indeks antropometri menggunakan algoritma c4.5 adaboost classification," *Jurnal Ilmiah Komputerisasi Akuntansi*, vol. 12, no. 2, pp. 45–51, 2019.
- [12] I. A. Rahmi, F. M. Afendi, and A. Kurnia, "Metode adaboost dan random forest untuk prediksi peserta jkn-kis yang menunggak," *Jambura Journal of Mathematics*, vol. 5, no. 1, pp. 83–94, 2023, doi: 10.34312/jjom.v5i1.15869.
- [13] T. Tanti, P. Sirait, and A. Andri, "Optimalisasi kinerja klasifikasi melalui seleksi fitur dan

adaboost dalam penanganan ketidakseimbangan kelas," *Jurnal Media Informatika Budidarma*, vol. 5, no. 4, pp. 1377–1385, 2021, doi: 10.30865/mib.v5i4.3280.

- [14] A. N. Rais, "Integrasi smote dan ensemble adaboost untuk mengatasi imbalance class pada data bank direct marketing," *Jurnal Informatika*, vol. 6, no. 2, pp. 278–285, 2019, doi: 10.31311/ji.v6i2.6186.
- [15] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, p. 27, 2019, doi: 10.1186/s40537-019-0192-5.



This article is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). Editorial of JJoM: Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B.J. Habibie, Moutong, Tilongkabila, Kabupaten Bone Bolango, Provinsi Gorontalo 96554, Indonesia.