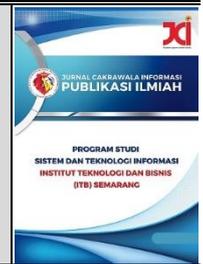




Jurnal Cakrawala Informasi

Journal Homepage: <http://www.itbsemarang.ac.id/sijies/index.php/jci>

e-Mail: jci@itbsemarang.ac.id



Performansi Algoritma C4.5 untuk Prediksi Kelulusan Mahasiswa

Dyah Ardyani Rizqi Azizah Adha ^{1*}

Aulia Noveesa Allanda ²

Diah Ayu Fatmasari ³

Siska Narulita ⁴

^{1, 2, 3, 4} Institut Teknologi dan Bisnis Semarang

INFO ARTIKEL

Histori artikel:

Diterima : 20 November 2023
 Revisi : 01 Desember 2023
 Disetujui : 05 Desember 2023
 Publikasi : 30 Desember 2023

Kata kunci:

Classification
Kelulusan
Data Mining
Decision Tree
C4.5

ABSTRACT

Graduating on time is one of the characteristics of a student's success in obtaining a bachelor's degree. But in reality, students do not always finish within four years. A graduate is a status title achieved by students after the education process is in accordance with the graduation requirements provided by the study program. Graduation on time is included in the components of measuring the quality of higher education institutions. Because it is a criterion in the accreditation process of study programs and institutions by the Higher Education Accreditation Board (BAN PT), Higher education institutions should have and implement procedures to manage and evaluate student graduation standards, which represent higher education's performance in improving quality. Therefore, it is necessary to manage student graduation, namely achieving the student's study period according to the scheduled length of study. In this research, the data mining algorithm used is C4.5 to predict student graduation. This algorithm is included in the classification with a numerical category. In this study, the C4.5 algorithm predicted graduation with an accuracy value of 85.96%.

ABSTRAK

Lulus tepat waktu adalah salah satu ciri keberhasilan mahasiswa dalam meraih gelar sarjana. Namun pada kenyataannya, mahasiswa tidak selalu menyelesaikan dalam waktu empat tahun. Lulusan merupakan gelar status yang dicapai mahasiswa setelah proses pendidikan sesuai dengan persyaratan kelulusan yang diberikan oleh program studi. Kelulusan tepat waktu termasuk dalam komponen pengukuran mutu lembaga pendidikan tinggi. Karena merupakan kriteria dalam proses akreditasi program studi dan

institusi oleh Badan Akreditasi Perguruan Tinggi (BAN PT). Perguruan tinggi sebaiknya memiliki serta menerapkan prosedur untuk mengelola dan mengevaluasi standar kelulusan mahasiswa yang mempresentasikan kinerja perguruan tinggi dalam peningkatan mutu. Oleh karena itu, perlu dilakukan untuk mengelola kelulusan mahasiswa, yaitu tercapainya masa studi mahasiswa sesuai dengan lama studi yang telah terjadwal. Pada penelitian ini, algoritma *data mining* yang digunakan adalah C4.5 untuk memprediksi kelulusan mahasiswa. Algoritma ini termasuk dalam klasifikasi dengan kategori numerik. Pada penelitian ini algoritma C4.5 memprediksi kelulusan dengan nilai *accuracy* 85,96%.

PENDAHULUAN

Dalam memperoleh gelar sarjana, kelulusan tepat waktu merupakan indikator terpenting [1]. Namun pada kenyataannya, tidak semua mahasiswa bisa menyelesaikan tepat waktu atau selama empat tahun. Badan Akreditasi Nasional Perguruan Tinggi (BANPT) dalam hal ini merupakan lembaga yang mendapatkan tugas untuk mengevaluasi perguruan tinggi dengan cara mengakreditasi perguruan tinggi tersebut. BANPT memiliki standar dalam penilaiannya salah satu standarnya yakni, standar tiga untuk mahasiswa dan lulusan, menyelesaikan kuliah tepat waktu termasuk dalam penilaian akreditasi [1]. Jika tidak seimbang antara tingkat kelulusan dan kemahasiswaan akan mempengaruhi penilaian akreditasi program sarjana dan institusi. Karena indikator penilaian pada lulusan perguruan tinggi salah satunya ialah perguruan tinggi memiliki angka efisien edukasi yang ideal [2]. Apabila mahasiswa dapat menyelesaikan studi di perguruan tinggi kurang dari atau sama dengan empat tahun maka mahasiswa tersebut dinyatakan lulus tepat waktu, dan sedangkan apabila mahasiswa tersebut

menyelesaikan studinya lebih dari empat tahun, maka mahasiswa tersebut dinyatakan tidak lulus tepat waktu atau terlambat [3].

Dengan begitu Perguruan Tinggi Negeri (PTN) ataupun Perguruan Tinggi Swasta (PTS) bersaing begitu ketat untuk memberikan performanya terbaiknya dalam mencetak lulusan mahasiswa berkualitas [4]. Sudah sewajarnya pihak PTN maupun PTS memberikan upaya terbaik dalam meningkatkan kualitas dan memberikan pendidikan terbaik kepada mahasiswa [5].

Lulusan merupakan gelar status yang dicapai mahasiswa setelah proses pendidikan sesuai dengan persyaratan kelulusan yang diberikan oleh program studi [6]. Lulusan yang bermutu tinggi memiliki ciri berkompotensi akademik termasuk *hard skills* dan *soft skills* sebagaimana dinyatakan dalam sasaran mutu serta dibuktikan dengan kinerja lulusan tersebut di masyarakat sesuai dengan profesi serta bidang ilmu merupakan indikator keluaran langsung dari proses pendidikan yang dilakukan oleh program studi [7]. Program studi yang berkualitas memiliki sistem pengelolaan lulusan yang baik sehingga dapat menjadikannya sebagai *human capital* bagi program studi yang bersangkutan [7]. Perguruan tinggi sebaiknya memiliki serta menerapkan prosedur untuk mengelola dan mengevaluasi standar tersebut yakni kelulusan mahasiswa yang mempresentasikan kinerja perguruan tinggi dalam peningkatan mutu. Oleh karena itu perlu dilakukan untuk mengelola kelulusan mahasiswa yaitu tercapainya masa studi mahasiswa sesuai dengan lama studi yang telah terjadwal [7]. Namun beberapa perguruan tinggi mengalami kendala bagaimana mahasiswa bisa mencapai masa studi tepat waktu sesuai dengan jadwal yang telah

ditetapkan oleh perguruan tinggi [8]. Mahasiswa yang lulus tepat waktu belum mencapai 50% dalam tujuh tahun terakhir [3]. Sebagai upaya mengelola kelulusan mahasiswa agar dapat lulus secara terjadwal dilakukan prediksi terhadap kelulusan mahasiswa menggunakan *data mining*.

Data mining adalah proses analisa suatu kumpulan data untuk menemukan pengetahuan yang tidak terduga sebelumnya serta meringkas data dengan menggunakan cara yang berbeda dari sebelumnya, sehingga dapat dimengerti serta memberikan manfaat bagi pemilik data [9]. Menurut Yuli (2017) metode-metode atau algoritma, dalam *data mining* memiliki banyak variasi. Untuk melakukan prediksi kelulusan mahasiswa digunakan metode *data mining* prediksi algoritma C4.5 atau *decision tree*. Menurut Asmaul (2021), algoritma *decision tree* adalah representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut dan simpul daun ditandai dengan kelas-kelas yang berbeda [10]. Algoritma *decision tree* C4.5 merupakan sebuah algoritma yang memiliki ciri khas digunakan untuk pengelompokan data numerik dan kategori [3]. Algoritma *decision tree* sebenarnya telah banyak digunakan untuk melakukan prediksi kelulusan. Diantaranya penelitian yang dilakukan oleh Tuhfatul dan Deni (2023) tentang “Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 pada UIN Syarif Hidayatullah Jakarta”. Dalam penelitian ini, peneliti mengembangkan model prediksi dengan menggunakan algoritma C4.5 untuk memprediksi kelulusan tepat waktu, penelitian ini memperoleh *accuracy* 75,52%, *precision* 75,52%, dan *recall* 75,52% .

Penelitian berikutnya dilakukan oleh Swardika & Suniantara (2019) tentang “Analisis C4.5 pada Klasifikasi CART Ketidaktepatan Waktu Kelulusan Mahasiswa Universitas Terbuka”. Pada penelitian ini menggunakan metode *classification and regression tree* (CART) dan C4.5 [11]. Dari penelitian tersebut didapatkan C4.5 mampu meningkatkan akurasi klasifikasi ketidaktepatan waktu kelulusan mahasiswa mencapai konvergen dengan prediksi klasifikasi mencapai 93,23%.

Menurut Ma’ruf et al., (2021) mengenai tentang “Penerapan Model Prediksi Menggunakan Algoritma C4.5 untuk Prediksi Kelulusan Siswa SMK Wahidin”. Penelitian ini menggunakan metode pohon keputusan atau algoritma C4.5. Dengan hasil akurasi dari prediksi kelulusan menggunakan *automatic* memiliki *accuracy* 82,19%, prediksi kelulusan menggunakan *linier sampling* memiliki *accuracy* 93,15%, prediksi kelulusan menggunakan *shuffled sampling* memiliki *accuracy* 84,93%, dan pada prediksi kelulusan menggunakan *stratified sampling* memiliki *accuracy* 82,91% [12].

Penelitian selanjutnya tentang “Metode Prediksi Kelulusan Mahasiswa Menggunakan *Decision Tree* C4.5 dan *Software Weka*” yang dilakukan oleh Isnan & Muanas (2021). Dalam penelitian ini dibangun sebuah model untuk memprediksi status kelulusan mahasiswa Institut Bisnis dan Informatika Kesatuan menggunakan algoritma C4.5. Penelitian ini menggunakan IPK Mahasiswa semester 1 hingga semester 4, untuk mahasiswa tahun 2013 sampai 2016 untuk membangun model prediksi kelulusan mahasiswa. Model prediksi ini menghasilkan akurasi sebesar 73%, hasil yang kurang baik [13]. Hasil ini

kemungkinan disebabkan oleh proporsi data yang digunakan tidak seimbang.

Menurut Gaol et al., (2021) tentang “Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi dengan Menggunakan Algoritma C4.5”. Dalam penelitian ini menggunakan algoritma C4.5. Data yang digunakan sebagai *data training* merupakan alumni prodi sistem informasi STIKOM Tunas Bangsa serta data mahasiswa tingkat akhir prodi sistem informasi STIKOM Tunas Bangsa digunakan sebagai *data testing*. Penerapan algoritma C4.5 ditemukan *accuracy* sebesar 90% dengan *precision* 91,38%, dan *recall* 98,15% pada prediksi kelulusan mahasiswa tingkat akhir prodi sistem informasi di STIKOM Tunas Bangsa Pematangsiantar [14]. Pada penelitian ini memaparkan bahwa atribut IPK merupakan atribut yang berpengaruh paling tinggi dalam menentukan kelulusan mahasiswa. Adanya penerapan *data mining* dengan menggunakan algoritma C4.5 dalam melakukan prediksi kelulusan mahasiswa STIKOM Tunas Bangsa Pematangsiantar prodi sistem informasi dapat membantu pihak institusi dalam mengambil keputusan dalam meningkatkan jumlah lulusan mahasiswa prodi sistem informasi di STIKOM Tunas Bangsa Pematangsiantar [14].

Berdasarkan penelitian terdahulu, maka dipilih algoritma C4.5 untuk prediksi kelulusan mahasiswa. Diharapkan dengan penggunaan algoritma C4.5 dapat dimanfaatkan untuk kebijakan perguruan tinggi untuk prediksi kelulusan mahasiswa dapat menghasilkan akurasi yang lebih baik dari penelitian-penelitian sebelumnya.

TINJAUAN PUSTAKA

A. Data Mining

Definisi *data mining* adalah melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui, dari suatu data [15]. Kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola dan hubungan dalam set data berukuran besar [16]. *Extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data* [17].

Data mining merupakan proses untuk menemukan pola (*pattern*) dari suatu data. Pola (*pattern*) yang ditemukan harus memiliki arti atau mengandung informasi penting [15]. *Data mining* adalah proses analisis dari *dataset* (berukuran besar) untuk menemukan suatu hubungan tak terduga serta meringkas data tersebut dengan cara baru yang dapat dimengerti dan bermanfaat bagi pemilik data [7]. Menurut Larose (2006), *data mining* merupakan proses analisis dari sekumpulan data untuk menemukan pengetahuan yang tidak terduga dan merangkum data dengan menggunakan cara yang berbeda dari sebelumnya, sehingga dapat dipahami dan berguna bagi pemilik data [9]. Kata *mining* itu sendiri berasal dari kata *mine* dari bahasa Inggris yang artinya menambang sumber daya yang tersembunyi [18]. Definisi *data mining* lainnya adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis atau pembelajaran berbasis induksi (*induction based learning*), yaitu proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari [19]. *Knowledge Discovery in Databases* (KDD) adalah penerapan metode *scientific* pada

data mining [19]. Dalam konteks ini, *data mining* merupakan satu langkah dari proses KDD [19].

B. Classification (Klasifikasi)

Klasifikasi adalah proses menempatkan objek tertentu dalam satu set kategori, berdasarkan sifat masing-masing objek [20]. Proses klasifikasi didasarkan pada empat komponen [21]:

1. Class

Class adalah variabel tidak bebas (*dependent*) yang merupakan variabel kategorik yang merepresentasikan label objek setelah diklasifikasi.

2. Prediktor

Prediktor merupakan variabel bebas (*independent*) yang direpresentasikan oleh karakteristik data (atribut) yang menjadi dasar dari klasifikasi.

3. Dataset training

Dataset training merupakan kumpulan data yang berisi komponen *class* (label) dan prediktor (atribut) yang digunakan untuk *training* model untuk mengenali *class* berdasarkan variabel prediktor (atribut) yang tersedia.

4. Dataset testing

Dataset testing berisi data baru yang akan diklasifikasikan menggunakan model yang telah dibangun atau ditemukan, dengan demikian akurasi dari model klasifikasi (performansi model) tersebut dapat dievaluasi.

C. Algoritma Decision Tree (C4.5)

Algoritma *decision tree* (C4.5) adalah algoritma klasifikasi yang memiliki struktur seperti pohon terdiri atas *node* internal (akar) dan *node*

cabang. *Node* internal (akar) merupakan atribut dan sebuah *node* cabang mewakili *class* [22]. Algoritma *decision tree* (C4.5) merupakan algoritma klasifikasi yang mengklasifikasikan sampel data secara *top-down* mulai dari simpul akar dan bergerak sesuai dengan hasil pengujian dari *node* internal sampai *node* cabang dicapai dan *class* ditetapkan [23]. Tahapan-tahapan algoritma *decision tree* (C4.5):

1. Memilih atribut sebagai akar yang didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan rumus:

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

S : himpunan kasus

A : atribut

n : jumlah atribut A

|S_i| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

Nilai entropi dapat dihitung dengan persamaan:

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

S : himpunan (*dataset*) kasus

n : banyaknya partisi

p_i : probabilitas yang didapat dari jumlah kasus pada partisi ke-i dibagi total kasus

2. Membuat cabang untuk setiap nilai didalam akar tersebut.
3. Membagi atribut dalam cabang.

4. Ulangi proses untuk setiap atribut sampai semua atribut pada cabang memiliki *class* yang sama.

D. Confusion Matrix

Confusion matrix merupakan metode yang digunakan untuk mengukur atau melakukan perhitungan akurasi pada konsep *data mining*. *Confusion matrix* terdiri dari data uji yang diprediksi benar dan tidak benar suatu data oleh model klasifikasi. Berikut ini tabel confusion matrix untuk pengklasifikasian ke dalam dua kelas:

Tabel 1. *Confusion Matrix*

	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive Instance</i>	<i>Number of True Positive Instance (TP)</i>	<i>Number of False Negative Instance (FN)</i>
<i>Actual Negative Instance</i>	<i>Number of False Positive Instance (FP)</i>	<i>Number of True Negative Instance (TN)</i>

Berdasarkan tabel *confusion matrix* di atas:

1. *True positive* (TP), merupakan jumlah data positif yang diklasifikasikan sebagai nilai positif.
2. *False positive* (FP), merupakan jumlah data negatif yang diklasifikasikan sebagai nilai positif.
3. *False negative* (FN), merupakan jumlah data positif yang diklasifikasikan sebagai nilai positif.
4. *True negative* (TN), merupakan jumlah data positif yang diklasifikasikan sebagai nilai negatif.

Confusion matrix digunakan untuk memperoleh nilai *precision*, *recall*, dan *accuracy*. Nilai *confusion matrix* biasanya ditunjukkan dalam satuan persen (%) [24]. Nilai yang dihasilkan

melalui metode *confusion matrix* adalah berupa evaluasi sebagai berikut:

1. *Accuracy* (akurasi)

Merupakan persentase jumlah data yang dilakukan pada klasifikasi atau prediksi secara benar oleh algoritma, dengan rumus sebagai berikut:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

2. *Precision* (presisi)

Nilai ketepatan dari metode yang digunakan dalam klasifikasi. Nilai tersebut menunjukkan banyaknya data yang dapat terklasifikasi di kelas yang benar dalam beberapa pengujian. Adapun rumusnya sebagai berikut:

$$Precision = \frac{TP}{(TP + FP)}$$

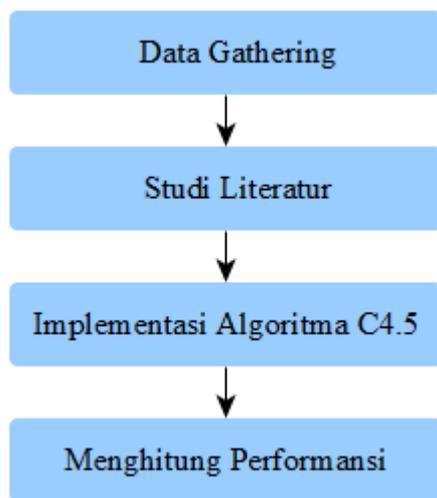
3. *Recall*

Nilai yang dapat mengukur hasil berapa persentase data yang terklasifikasi dengan benar, adapun rumus yang digunakan untuk menghitung nilai *recall* adalah:

$$Recall = \frac{TP}{(TP + FN)}$$

METODE PENELITIAN

Langkah-langkah yang dilakukan pada penelitian ini dapat dilihat pada gambar berikut:



Gambar 1. Kerangka Pikir Penelitian

A. Data Gathering

Pengumpulan, pengukuran, dan analisis berbagai tipe informasi menggunakan teknik berstandar dengan tujuan utama untuk pengumpulan informasi dan data valid sebanyak-banyaknya yang kemudian dianalisis untuk membuat sebuah keputusan yang krusial [25]. Dalam penelitian ini, *dataset* yang digunakan diperoleh dari <https://www.kaggle.com/datasets>.

B. Studi Literatur

Studi literatur merupakan suatu kegiatan yang dilakukan untuk mencari bahan penelitian dengan menggunakan metode pengumpulan data pustaka, membaca, mencatat, dan mengelola data. Studi literatur atau studi pustaka merupakan cara yang digunakan untuk menyelesaikan suatu masalah dengan menelusuri sumber-sumber sebelumnya. Teknik ini dilakukan oleh peneliti untuk menentukan ide, gagasan, atau referensi untuk mendapatkan landasan teori, kerangka berpikir, dan dapat menentukan prediksi sementara yang biasa disebut dengan hipotesis penelitian.

C. Implementasi Algoritma C4.5

Pada tahap ini dilakukan implementasi metode *data mining classification* menggunakan algoritma *decision tree* atau C4.5. Dalam penelitian ini, digunakan *software RapidMiner* sebagai alat bantu dalam proses pengolahan *dataset*.

D. Menghitung Performansi

Untuk mengukur performansi atau kinerja dari model yang dihasilkan oleh algoritma C4.5 digunakan metode evaluasi *confusion matrix*. Dimana dilakukan pengukuran sejauh mana model dapat mengklasifikasikan data dengan benar.

PEMBAHASAN DAN HASIL

A. Data Gathering

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data yang diperoleh melalui pihak lain atau secara tidak langsung [26]. Sumber data sekunder dalam penelitian ini yaitu *dataset* kelulusan mahasiswa. Dalam penelitian ini diambil sampel sebanyak 380 data. Sampel *dataset* yang telah dimasukkan ke dalam *tool Rapidminer* dapat dilihat pada gambar berikut ini:

Row No.	STATUS KEL.	NAMA	JENIS KEL.	STATUS MA.	UMUR	STATUS NIK.	IPS 1	IPS 2	IPS
1	TERLAMBAT	ANIK WICHAYA	PEREMPUNIAN	BEKERJA	28	BELUM MEN.	2.700	2.800	3.200
2	TERLAMBAT	DWI HESTYNI	PEREMPUNIAN	MAHASISWA	32	BELUM MEN.	3	3.300	3.14
3	TERLAMBAT	MURYA ARIE	PEREMPUNIAN	BEKERJA	29	BELUM MEN.	3.500	3.300	3.70
4	TERLAMBAT	NANNI SUSAL	PEREMPUNIAN	MAHASISWA	27	BELUM MEN.	3.170	3.410	3.61
5	TERLAMBAT	RIFKA ISTIQO	PEREMPUNIAN	BEKERJA	29	BELUM MEN.	2.900	2.890	3.30
6	TERLAMBAT	SUKHARYONO	LAKU - LAKU	BEKERJA	27	BELUM MEN.	2.950	2.820	3.00
7	TEPAT	FARUKHATUN	PEREMPUNIAN	MAHASISWA	28	BELUM MEN.	2.700	3.140	2.60
8	TERLAMBAT	FIFI SUNALISA	PEREMPUNIAN	MAHASISWA	27	BELUM MEN.	2.820	2.890	2.30
9	TERLAMBAT	HENDRIK M.	PEREMPUNIAN	BEKERJA	25	MEINGAH	3.600	3.540	3.50
10	TERLAMBAT	IRANI AGUNG	PEREMPUNIAN	BEKERJA	28	BELUM MEN.	2.710	2.550	1.77
11	TERLAMBAT	IRANI SANTO	PEREMPUNIAN	BEKERJA	27	BELUM MEN.	3.140	3.460	3.40
12	TERLAMBAT	IRFAN ENO	PEREMPUNIAN	BEKERJA	32	BELUM MEN.	2.670	2.300	1.57
13	TERLAMBAT	IRWAN HAMBALI	PEREMPUNIAN	BEKERJA	26	BELUM MEN.	2.570	2.820	2.20
14	TERLAMBAT	M SYAFULLAH	PEREMPUNIAN	BEKERJA	31	BELUM MEN.	2.710	3	2.60

Gambar 2. *Dataset* Kelulusan Mahasiswa

B. Studi Literatur

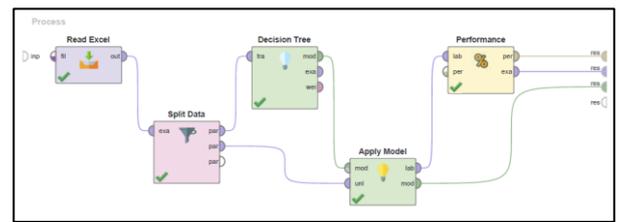
Beberapa penelitian terdahulu yang dijadikan referensi oleh penulis, antara lain

penelitian yang dilakukan Hasibuan dan Mahdiana (2023) yang dalam penelitiannya menyimpulkan bahwa prediksi kelulusan mahasiswa tepat waktu menggunakan algoritma C4.5 menghasilkan akurasi sebesar 75,45% [3]. Penelitian lainnya dilakukan oleh Suwardika dan Suniantara (2019), yang dalam penelitiannya tentang klasifikasi ketidaktepatan waktu kelulusan mahasiswa di Universitas Terbuka, mendapatkan hasil prediksi dengan nilai akurasi sebesar 93,23% [11]. Rusbandi et al., (2021), dalam penelitiannya menggunakan algoritma C4.5 untuk memprediksi kelulusan siswa SMK Wahidin, mendapatkan nilai akurasi sebesar 82,91% [12]. Mulia dan Muanas (2021) dalam penelitiannya menyimpulkan bahwa algoritma klasifikasi C4.5 dapat diterapkan dalam prediksi kelulusan mahasiswa dengan tingkat akurasi sebesar 73% [13]. Dan penelitian lainnya oleh Gaol et al., (2021), yang dalam hasil penelitiannya menyebutkan bahwa hasil akurasi yang diperoleh dari penerapan algoritma C4.5 dalam memprediksi kelulusan mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi, menghasilkan nilai akurasi sebesar 90%, presisi 91,38%, dan *recall* 98,15% [14].

C. Implementasi Algoritma C4.5

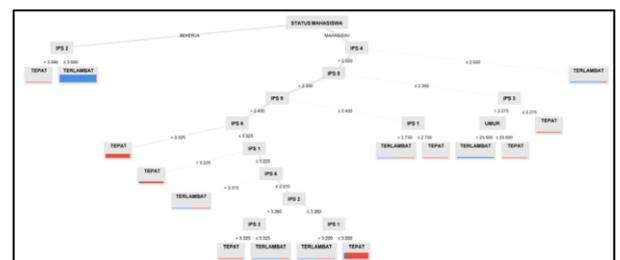
Sesuai dengan yang telah dijelaskan sebelumnya bahwa dalam penelitian ini menggunakan algoritma *classification* C4.5 untuk memprediksi kelulusan mahasiswa. Pada penelitian yang dilakukan ini, dilakukan pembagian *data training* dan *data testing* sebesar 90%:10%, 80%:20%, 70%:30%, dan 60%:40%. Berdasarkan masing-masing pembagian tersebut akan dihitung akurasinya. Adapun proses implementasi *data mining* dengan metode

classification menggunakan algoritma C4.5 pada *RapidMiner* ditunjukkan sebagai berikut:



Gambar 3. Proses *Classification* dengan Algoritma C4.5 pada *RapidMiner*

Sebagai langkah awal, *dataset* dimasukkan ke dalam *RapidMiner* menggunakan komponen *Read Excel* untuk membaca *dataset* yang bertipe xls. Langkah selanjutnya menambahkan komponen *split data* untuk membagi *data training* dan *data testing*. Selanjutnya menambahkan operator algoritma C4.5 dan komponen *apply model* untuk menjalankan *data testing* pada algoritma C4.5. Terakhir, tambahkan komponen *performance* untuk menghitung nilai akurasi. Kemudian tekan *run* untuk menjalankan proses. Visualisasi hasil *classification* ditunjukkan pada gambar pohon keputusan atau *decision tree* di bawah ini:



Gambar 4. Visualisasi *Decision Tree* Hasil *Classification*

D. Menghitung Performansi

Pada penelitian yang dilakukan, untuk mengevaluasi model yang dihasilkan oleh algoritma C4.5 menggunakan *dataset* kelulusan mahasiswa digunakan *confusion matrix*. Performansi algoritma C4.5 dihitung sesuai

perbandingan *data training* dan *data testing* pada proses implementasi. Dari perbandingan tersebut, diperoleh nilai akurasi algoritma C4.5 sebagai berikut:

Tabel 2. Hasil Akurasi Algoritma C4.5

Perbandingan <i>Data Testing</i> dan <i>Data Training</i>	Nilai Akurasi
90% : 10%	65,79%
80% : 20%	80,26%
70% : 30%	85,96%
60% : 40%	85,43%

KESIMPULAN

Berdasarkan pembahasan dan hasil dari penelitian yang telah selesai dilakukan, dapat diambil kesimpulan bahwa pada perbandingan *data training* dan *data testing* sebesar 70%:30% yang diimplementasikan pada algoritma C4.5 dengan *dataset* kelulusan mahasiswa, diperoleh nilai akurasi tertinggi, yaitu sebesar 85,96%.

DAFTAR PUSTAKA

- [1] Badan Akreditasi Nasional Perguruan Tinggi, *Akreditasi Program Studi: Pedoman Penilaian Badan Akreditasi Nasional Perguruan Tinggi*. Indonesia, 2019.
- [2] Badan Akreditasi Nasional Perguruan Tinggi, *Akreditasi Institusi Perguruan Tinggi Departemen Pendidikan Nasional*. Indonesia, 2007.
- [3] T. H. Hasibuan and D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 pada UIN Syarif Hidayatullah Jakarta," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 6, no. 1, pp. 61–74, 2023.
- [4] A. Budiyantra, Irwansyah, E. Prengki, P. A. Pratama, and N. Wiliani, "Komparasi Algoritma Decision Tree, Naive Bayes, dan K-Nearest Neighbor untuk Memprediksi Mahasiswa Lulus Tepat Waktu," *JITK J. Ilmu Pengetah. dan Teknol. Komput.*, vol. 5, no. 2, pp. 265–269, 2020.
- [5] S. A. Nulhaqim, D. H. Heryadi, R. Pancasilawan, and M. Ferdryansyah, "Peranan Perguruan Tinggi dalam Meningkatkan Kualitas Pendidikan di Indonesia untuk Menghadapi ASEAN Community 2015, Studi Kasus: Universitas Indonesia, Universitas Padjadjaran, Institut Teknologi Bandung," *SHARE Soc. Work J.*, vol. 6, no. 2, pp. 197–219, 2016.
- [6] Badan Akreditasi Nasional Perguruan Tinggi, *Akreditasi Program Studi Diploma*. Indonesia, 2010.
- [7] S. Rahmatullah and E. Utami, "Prediksi Tingkat Kelulusan Tepat Waktu dengan Metode Naive Bayes dan K-Nearest Neighbor," *J. Inf. dan Komput.*, vol. 7, no. 1, pp. 7–16, 2019.
- [8] M. Windarti, "Prediksi Masa Studi Mahasiswa Menggunakan Kombinasi Algoritma Bayesian Network dan K-Nearest Neighbors," 2016.
- [9] D. T. Larose, *Data Mining Methods and Models*. New Jersey: John Wiley & Sons, Inc, 2006.
- [10] A. H. Nasrullah, "Implementasi Algoritma Decision Tree untuk Klasifikasi Produk Laris," *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 45–51, 2021.
- [11] G. Suwardika and I. K. P. Suniantara, "Analisis Random Forest pada Klasifikasi CART Ketidaktepatan Waktu Kelulusan Mahasiswa Universitas Terbuka," *Barekeng J. Ilmu Mat. dan Terap.*, vol. 13, no. 3, pp. 179–186, 2019.
- [12] Rusbandi, F. A. Ma'ruf, A. Pratama, I. Sholihin, A. R. Rinaldi, and Faturrohman, "Penerapan Model Prediksi Menggunakan Algoritma C.45 untuk Prediksi Kelulusan Siswa SMK Wahidin," *J. Data Sci. Inform.*, vol. 1, no. 1, pp. 1–20, 2021.
- [13] I. Mulia and Muanas, "Model Prediksi Kelulusan Mahasiswa Menggunakan Decision Tree C4.5 dan Software Weka," *J. Anal. Sist. Pendidik. Tinggi Indones.*, vol. 5, no. 1, pp. 57–64, 2021.
- [14] L. Y. L. Gaol, M. Safii, and D. Suhendro, "Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5," *Brahmana J. Penerapan Kecerdasan Buatan*, vol. 2, no. 2, pp. 97–106, 2021.
- [15] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Third Edition*. Elsevier Inc., 2011.
- [16] B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [17] J. Han, M. Kamber, and J. Pei, *Data Mining, Concepts and Techniques*, Third Edit. Massachusetts: Morgan Kaufmann Publishers, 2012.

- [18] Dennis Aprilla C, Donny Aji Baskoro, L. Ambarwati, and I. W. S. Wicaksana, *Belajar Data Mining dengan Rapid Miner*. Jakarta: Open Content Model, 2013.
- [19] F. Riandari and A. Simangunsong, “Penerapan Algoritma C4.5 untuk Mengukur Tingkat Kepuasan Mahasiswa,” *J. Mantik Penusa*, vol. 3, no. 2, pp. 1–7, 2019.
- [20] R. A. Fisher, “The Use of Multiple Measurements in Taxonomic Problems,” *Annu. Eugen.*, vol. 7, pp. 179–188, 1936, doi: 10.1111/j.1469-1809.1936.tb02137.x.
- [21] F. Gorunescu, *Data Mining Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer, 2011.
- [22] T. W. Liao and E. Triantaphyllou, *Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications*. World Scientific, 2007.
- [23] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
- [24] B. Firmanto, H. Soekotjo, and H. S. Dachlan, “Perbandingan Kinerja Algoritma Promethee dan TOPSIS untuk Pemilihan Guru Teladan,” *J. Penelit. Pendidik. IPA*, vol. 2, no. 1, 2016.
- [25] M. Aulia, “Internet Data Collection,” Surakarta: Universitas Sebelas Maret, 2019.
- [26] S. Narulita, P. Prihati, and A. Priyambodo, “Analisis dan Komparasi Algoritma Klasifikasi untuk Prediksi Kerugian Tower Provider Akibat Penalti yang Diberikan oleh Operator Telekomunikasi karena Keterlambatan Penyelesaian Pekerjaan oleh Tower Provider,” *Cakrawala Inf.*, vol. 2, no. 2, pp. 1–14, 2022, [Online]. Available: <https://itbsemarang.ac.id/sijies/index.php/jci/article/view/233>.