

Model Prediksi Keberhasilan Mahasiswa Pada Mata Kuliah Rekayasa Perangkat Lunak Menggunakan Algoritma C5.0

Aprilia Sulistyohati^{1*}, Fauzan Natsir²

^{1,2} Teknik Informatika, Universitas Indraprasta PGRI

¹aprilias6891@gmail.com, ²fauzan.natsir@gmail.com

*Penulis Korespondensi

ABSTRAK

Kemajuan teknologi di era Revolusi Industri 5.0 mendorong perguruan tinggi untuk memanfaatkan teknologi dalam meningkatkan kualitas pendidikan. Pada pendidikan tinggi, keberhasilan mahasiswa dalam menyelesaikan mata kuliah inti menjadi salah satu indikator utama dalam pencapaian akademik. Mata kuliah Rekayasa Perangkat Lunak (RPL) merupakan salah satu mata kuliah utama dalam program studi Teknik Informatika yang memiliki peran dalam membekali mahasiswa dengan kemampuan analisis, desain, dan pengembangan perangkat lunak. Data mining adalah proses eksplorasi dan analisis data berskala besar untuk mengungkap pola, hubungan, atau informasi tersembunyi. Salah satu metode penting dalam data mining adalah klasifikasi, yang dapat dilakukan menggunakan algoritma C5.0, dimana algoritma ini bekerja dengan membangun pohon keputusan berdasarkan atribut dalam dataset, kemudian menggunakan pohon tersebut untuk memprediksi hasil. Pada penelitian ini penulis menggunakan algoritma C5.0 untuk menganalisis data histori akademik mahasiswa, seperti keaktifan, nilai tugas, UTS, dan UAS. Atribut tersebut digunakan untuk memprediksi keberhasilan mahasiswa pada mata kuliah RPL. Penelitian ini melibatkan tahapan studi literatur, pengumpulan data, *pre-processing*, perhitungan *entropy*, dan gain untuk menentukan atribut dengan nilai tertinggi sebagai akar pohon keputusan. Hasilnya menunjukkan bahwa atribut tugas memiliki nilai gain tertinggi dan menjadi akar pohon keputusan. Model yang dihasilkan mencapai akurasi 88,33% pada data testing, ini membuktikan efektivitas algoritma C5.0 dalam memprediksi keberhasilan mahasiswa pada matakuliah RPL. Penelitian ini memberikan kontribusi pada pengembangan model prediksi berbasis data untuk mendukung pengambilan keputusan akademik yang lebih proaktif dan efisien guna meningkatkan keberhasilan akademik mahasiswa.

Kata kunci: Algoritma C5.0, prediksi, kelulusan, *data mining*

ABSTRACT

Technological advances in the era of the Industrial Revolution 5.0 encourages universities to use technology in improving the quality of education. In higher education, student success in completing core courses is one of the main indicators of academic achievement. Software Engineering (RPL) course is one of the core courses of the informatics engineering study program which has a role in equipping students with the ability to analyze, design, and develop software. Datamining is the process of exploring and analyzing data on a large scale to find patterns, relationships, or hidden information. One of the main techniques in data mining is classification. One of the classification processes can be done using the C5.0 algorithm, where this algorithm works by building a decision tree based on attributes in the dataset, then using the tree to predict the results. In this research, the author applies the C5.0 algorithm to process student academic history data, such as activeness, assignment grades, UTS, and UAS. These attributes are used to predict student graduation in RPL courses. This research includes literature study, data collection, pre-processing, entropy calculation, and gain to determine the attribute with the highest value as the root of the decision tree. The results of this study show that the assignment attribute has the highest gain value and becomes the root of the decision tree. The resulting model achieved 88.33% accuracy on the testing data, this proves the effectiveness of the C5.0 algorithm in predicting student graduation in RPL courses. This research contributes to the development of data-based prediction models to support more proactive and efficient academic decision making to improve student academic success.

Keywords: C5.0 algorithm, prediction, graduation, datamining

1. PENDAHULUAN

Kemajuan teknologi saat ini telah memberikan dampak besar pada berbagai aspek kehidupan, termasuk di sektor pendidikan tinggi. Di era Revolusi Industri 5.0, teknologi tidak hanya sarana pendukung, tetapi juga dapat berkontribusi dalam menciptakan pembelajaran yang lebih fleksibel dan berpusat pada kebutuhan individu [1]. Perguruan tinggi, sebagai institusi yang bertanggung jawab dalam mencetak generasi berkompeten, dituntut untuk memanfaatkan teknologi secara maksimal dalam proses pendidikan. Implementasi teknologi ini memungkinkan perguruan tinggi untuk mengidentifikasi mahasiswa yang berisiko gagal lebih dini, memahami pola belajar

mereka, dan memberikan intervensi yang sesuai. Hal ini tidak hanya meningkatkan peluang keberhasilan mahasiswa, tetapi juga membantu institusi dalam memenuhi tuntutan pendidikan modern.

Dalam pendidikan tinggi, keberhasilan mahasiswa dalam menyelesaikan mata kuliah inti menjadi salah satu indikator utama dalam pencapaian akademik. Di bidang pendidikan, teknologi ini dapat dimanfaatkan untuk meningkatkan efisiensi dan akurasi dalam berbagai aspek, termasuk prediksi keberhasilan mahasiswa. Salah satu tantangan utama di pendidikan tinggi adalah memastikan bahwa mahasiswa dapat menyelesaikan program studi mereka dengan hasil yang memuaskan. Mata kuliah Rekayasa Perangkat Lunak (RPL) merupakan salah satu mata kuliah utama dari program studi teknik informatika yang memiliki peran strategis dalam membekali mahasiswa dengan kemampuan analisis, desain, dan pengembangan perangkat lunak. Namun, tingkat kesulitan materi dan keterbatasan pemahaman sering kali menjadi tantangan yang memengaruhi tingkat keberhasilan mahasiswa. Banyak mahasiswa yang mengalami kendala dalam menyelesaikan mata kuliah ini, baik karena aspek teknis maupun non-teknis, seperti kurangnya manajemen waktu atau motivasi belajar. Tingkat kesulitan mata kuliah ini membutuhkan pendekatan yang lebih inovatif untuk memastikan bahwa mahasiswa tidak hanya mampu lulus, tetapi juga memahami materi secara mendalam.

Untuk mengatasi permasalahan ini, diperlukan pendekatan berbasis teknologi yang mampu memberikan prediksi akurat terkait keberhasilan mahasiswa pada mata kuliah RPL. Pendekatan prediktif ini tidak hanya membantu mengidentifikasi mahasiswa yang berisiko gagal lebih dini, tetapi juga membuka kesempatan untuk melakukan intervensi yang sesuai dan efektif [2], [3]. Dengan pendekatan ini, pihak institusi dapat memperoleh wawasan mendalam tentang faktor-faktor utama yang memengaruhi keberhasilan mahasiswa.

Berbagai penelitian serupa telah dilakukan oleh peneliti sebelumnya untuk memprediksi kinerja akademik mahasiswa. Hasil penelitian tersebut menunjukkan bahwa algoritma *decision tree* mampu memberikan prediksi yang akurat terkait keberhasilan mahasiswa berdasarkan data historis mereka [4]–[7]. Selain itu, algoritma C5.0 juga terbukti efektif untuk memprediksi kemungkinan mahasiswa mengalami kegagalan atau dropout dalam pendidikan tinggi [8]. Temuan ini relevan untuk memahami bagaimana teknik prediktif dapat digunakan dalam memantau keberhasilan mahasiswa.

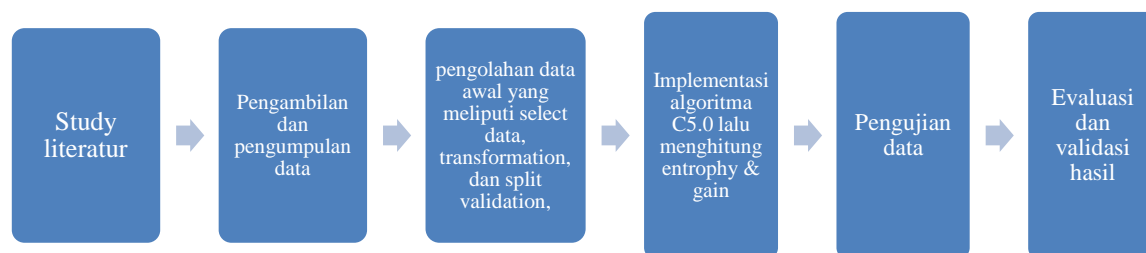
Oleh karena itu, penelitian ini memanfaatkan algoritma C5.0 untuk membantu perguruan tinggi mengolah data historis akademik mahasiswa secara lebih efektif [9][10], seperti nilai ujian, kehadiran, partisipasi dalam tugas, hingga aktivitas pembelajaran untuk memprediksi keberhasilan mahasiswa pada mata kuliah Rekayasa Perangkat Lunak (RPL). Algoritma C5.0, sebagai salah satu algoritma *decision tree* yang efisien dan akurat, menawarkan kemampuan analisis yang handal dalam memproses data akademik yang kompleks. Algoritma ini mampu membangun model prediksi yang tidak hanya akurat tetapi juga mudah diinterpretasikan oleh pengambil keputusan, seperti dosen maupun pihak administrasi akademik. Implementasi algoritma C5.0 dalam memprediksi keberhasilan RPL tidak hanya berfungsi sebagai alat analisis, tetapi juga menjadi dasar dalam merancang strategi pembelajaran yang lebih terarah dan berbasis data untuk meningkatkan peluang keberhasilan mahasiswa dalam menyelesaikan mata kuliah RPL. Hal ini sejalan dengan tujuan menciptakan lingkungan pendidikan yang proaktif dan adaptif terhadap kebutuhan individu mahasiswa.

2. METODOLOGI PENELITIAN

Data mining merupakan proses eksplorasi dan analisis data berskala besar untuk mengungkap pola, hubungan, atau informasi tersembunyi [11][9][10]. Proses ini menggunakan berbagai metode dan algoritma untuk melakukan tugas seperti klasifikasi, prediksi, segmentasi, atau asosiasi. Salah satu metode utama dalam data mining adalah klasifikasi, yang bertujuan memprediksi kategori atau kelas dari data baru berdasarkan pola yang telah diidentifikasi dari data historis. Dalam konteks pendidikan, data mining sering digunakan untuk mengolah data histori akademik mahasiswa seperti nilai ujian, tingkat kehadiran, dan aktivitas pembelajaran untuk menemukan pola yang dapat memprediksi keberhasilan mahasiswa. Teknik ini membantu institusi pendidikan mengambil keputusan berbasis data untuk meningkatkan kualitas pendidikan.

Klasifikasi dapat dilakukan menggunakan algoritma C5.0, yang merupakan pengembangan dari algoritma C4.5 dan banyak digunakan dalam proses klasifikasi data mining. Algoritma ini bekerja dengan membangun pohon keputusan berdasarkan atribut dalam dataset, lalu menggunakan pohon tersebut untuk melakukan prediksi. Algoritma C5.0 menentukan atribut yang paling berpengaruh terhadap hasil prediksi. Pohon keputusan yang dihasilkan terdiri dari node dan cabang, di mana setiap node merepresentasikan atribut, sementara cabang menunjukkan nilai atribut tersebut. Keputusan diambil berdasarkan nilai gain tertinggi dari hasil perhitungan semua atribut.

Adapun tahapan-tahapan dalam penelitian ini akan penulis jabarkan secara terperinci. Penulis menggunakan dua jenis data primer dan sekunder dalam pengumpulan datanya. Data primer diambil dari objek penelitian yaitu data histori akademik mahasiswa yang mengambil matakuliah rekayasa perangkat lunak (RPL). Secara umum, langkah-langkah yang dilakukan oleh penulis untuk menyelesaikan penelitian ini adalah sebagai berikut:



Gambar 1. Langkah-langkah penelitian

Studi literatur dilakukan melalui *review* jurnal internasional maupun nasional tentang topik sejenis dengan penelitian ini. Tahapan selanjutnya meliputi pengambilan dan pengumpulan data, diikuti oleh tahap *pre-processing* untuk membersihkan data dengan menghilangkan *missing value*, data yang tidak konsisten, serta data yang rusak. Proses awal pengolahan data mencakup pemilihan data (*select data*), transformasi data, dan *split validation*. Pada tahap *select data*, digunakan data yang sudah tersedia dari nilai mata kuliah Rekayasa Perangkat Lunak, mencakup keaktifan di kelas, nilai tugas, UTS, dan UAS. Setelah atribut dipilih, proses dilanjutkan dengan transformasi data untuk mengubah data mentah menjadi format yang lebih cocok untuk analisis. Kemudian dilakukan *split validation* dengan membagi data menjadi dua bagian, yaitu *data training* (80%) dan *data testing* (20%). Langkah-langkah penggunaan algoritma C5.0 sebagai berikut: [14]–[16]

- 1) Membuat Sistem Keputusan: Tentukan atribut kondisi dan keputusan, kemudian susun sistem pengambilan keputusan berdasarkan objek yang ada.
- 2) Menghitung Jumlah Data: Tentukan jumlah data berdasarkan atribut dan kondisi tertentu.
- 3) Pemilihan Node: Pilih atribut dengan nilai gain tertinggi sebagai node dan buat cabang untuk setiap anggota node.
- 4) Periksa Entropi: Periksa nilai entropi pada setiap node; jika entropi nol, tentukan daun yang terbentuk. Jika tidak, lanjutkan ke langkah berikutnya.
- 5) Iterasi Proses: Jika nilai entropi lebih besar dari nol, ulangi proses dengan *node* baru sampai semua anggota node memiliki entropi nol.

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini berasal dari histori mahasiswa/i pada mata kuliah Rekayasa Perangkat Lunak. Data sampel ini yang digunakan untuk memprediksi keberhasilan mahasiswa/i berdasarkan nilai mata kuliah rekayasa perangkat lunak. Pada proses pengumpulan sample didapatkan hasil 172 record data yang akan dilanjutkan ke proses berikutnya. Data tersebut kemudian dibagi untuk keperluan pemodelan, dengan 80% atau sebanyak 137 data digunakan sebagai data training, dan 20% sisanya, yaitu 35 data digunakan sebagai data testing. Data tersebut dikelompokkan ke dalam beberapa kriteria berdasarkan atributnya sebagai berikut:

Tabel 1. Sample data penelitian

Nama	Keaktifan	Tugas	UTS	UAS	Kelulusan
Siswa 1	3	65	88	97	Lulus
Siswa 2	1	79	78	66	Lulus
Siswa 3	2	71	63	99	Lulus
Siswa 60	1	73	64	93	Lulus
Siswa 61	1	84	80	57	Lulus
Siswa 62	3	66	86	68	Lulus
Siswa 63	2	55	53	95	Tidak Lulus
Siswa 64	1	80	87	99	Lulus
Siswa 65	3	58	90	91	Lulus
Siswa 106	2	85	62	69	Lulus
Siswa 107	3	50	80	80	Lulus
Siswa 108	3	62	86	96	Lulus
Siswa 109	1	56	95	53	Tidak Lulus
Siswa 110	2	71	97	83	Lulus
Siswa 125	1	78	69	100	Lulus
Siswa 126	2	98	77	67	Lulus
Siswa 127	3	72	65	84	Lulus

Nama	Keaktifan	Tugas	UTS	UAS	Kelulusan
Siswa 136	1	58	86	80	Lulus
Siswa 137	3	84	88	80	Lulus
Siswa 148	3	53	77	64	Tidak Lulus
Siswa 149	3	61	57	63	Tidak Lulus
Siswa 150	3	93	64	79	Lulus
Siswa 151	3	91	75	70	Lulus
Siswa 152	2	81	55	92	Lulus
Siswa 153	1	60	85	70	Lulus
Siswa 154	3	80	70	72	Lulus
Siswa 155	3	64	63	65	Tidak Lulus
...
Siswa 170	2	60	85	56	Tidak Lulus
Siswa 171	3	63	73	84	Lulus
Siswa 172	3	92	88	71	Lulus

Pada tabel 1 menyajikan data nilai mahasiswa pada mata kuliah rekayasa perangkat lunak. Selanjutnya, dilakukan proses transformasi data untuk mengubah data mentah menjadi format yang lebih cocok untuk analisis lanjutan, sehingga dapat meningkatkan performa model yang digunakan. Adapun data hasil transformasi sebagai berikut:

Tabel 2. *Sample data penelitian*

Nama	Keaktifan	Tugas	UTS	UAS	Hasil
Siswa 1	Baik	Cukup	Baik	Baik	Lulus
Siswa 2	Kurang	Cukup	Cukup	Cukup	Lulus
Siswa 3	Cukup	Cukup	Cukup	Baik	Lulus
Siswa 60	Kurang	Cukup	Cukup	Baik	Lulus
Siswa 61	Kurang	Cukup	Cukup	Kurang	Lulus
Siswa 62	Baik	Cukup	Baik	Cukup	Lulus
Siswa 63	Cukup	Kurang	Kurang	Baik	Tidak Lulus
Siswa 64	Kurang	Cukup	Baik	Baik	Lulus
Siswa 65	Baik	Kurang	Baik	Baik	Lulus
Siswa 106	Cukup	Rajin	Cukup	Cukup	Lulus
Siswa 107	Baik	Kurang	Cukup	Cukup	Lulus
Siswa 108	Baik	Kurang	Baik	Baik	Lulus
Siswa 109	Kurang	Kurang	Baik	Kurang	Tidak Lulus
Siswa 110	Cukup	Cukup	Baik	Cukup	Lulus
Siswa 125	Kurang	Cukup	Cukup	Baik	Lulus
Siswa 126	Cukup	Rajin	Cukup	Cukup	Lulus
Siswa 127	Baik	Cukup	Cukup	Cukup	Lulus
Siswa 136	Kurang	Kurang	Baik	Cukup	Lulus
Siswa 137	Baik	Cukup	Baik	Cukup	Lulus
Siswa 148	Baik	Kurang	Cukup	Cukup	Tidak Lulus
Siswa 149	Baik	Kurang	Kurang	Cukup	Tidak Lulus
Siswa 150	Baik	Rajin	Cukup	Cukup	Lulus
Siswa 151	Baik	Rajin	Cukup	Cukup	Lulus
Siswa 152	Cukup	Cukup	Kurang	Baik	Lulus
Siswa 153	Kurang	Kurang	Baik	Cukup	Lulus
Siswa 154	Baik	Cukup	Cukup	Cukup	Lulus
Siswa 155	Baik	Kurang	Cukup	Cukup	Tidak Lulus
...
Siswa 170	Cukup	Kurang	Baik	Kurang	Tidak Lulus
Siswa 171	Baik	Kurang	Cukup	Cukup	Lulus
Siswa 172	Baik	Rajin	Baik	Cukup	Lulus

Dari tabel 2 di atas merupakan data hasil transformasi yang sudah siap untuk proses klasifikasi dengan menggunakan algoritma C5.0. Berdasarkan atribut keaktifan di kelas, tugas, UTS dan UAS penulis menghitung jumlah kasus dengan hasil lulus dan tidak lulus. Proses berikutnya dilakukan perhitungan *entropy* dan *gain* dari masing-masing variabel. Tahapan dalam mengimplementasikan algoritma C5.0 dengan menggunakan data training sebagai berikut :

1. Siapkan data training yang berjumlah 137 data
2. Hitung jumlah mahasiswa yang lulus dan tidak lulus dari masing-masing atribut
3. Hitung nilai entropy total, dimana mahasiswa lulus sebanyak 125 orang dan mahasiswa yang tidak lulus sejumlah 12 orang.

Perhitungan nilai entropy dapat dilihat pada rumus berikut : [17]–[19]

$$Entropy(S) = \sum_{i=0}^n - p_i * \log_2 p_i \quad (1)$$

Keterangan :

S = himpunan kasus
A = atribut
n = jumlah partisi
p_i = proporsi dari S_i terhadap S

Dari rumus entropy di atas didapatkan hasil :

$$entropy(S) = -\left(\frac{94}{137}\right) * \log_2\left(\frac{94}{137}\right) + \left(-\left(\frac{43}{137}\right) * \log_2\left(\frac{43}{137}\right)\right) = 0.898$$

4. Hitung nilai gain dari masing-masing atribut dengan menggunakan rumus berikut ini : [17], [19], [20]

$$Gain(S, A) = Entropy(S) - \sum_{i=0}^n \frac{|S_i|}{|S|} * entropy(S) \quad (2)$$

Keterangan :

S = himpunan kasus
A = fitur
n = jumlah partisi atribut A
|S_i| = Jumlah kasus pada partisi ke-i
|S| = Jumlah kasus dalam S

Penulis melakukan perhitungan nilai *gain* untuk setiap atribut dan menentukan atribut dengan nilai gain tertinggi. Atribut dengan nilai gain tertinggi ini akan digunakan sebagai akar (root) pohon keputusan. Berdasarkan rumus yang digunakan, penulis menghitung nilai gain untuk atribut keaktifan di kelas sebagai berikut:

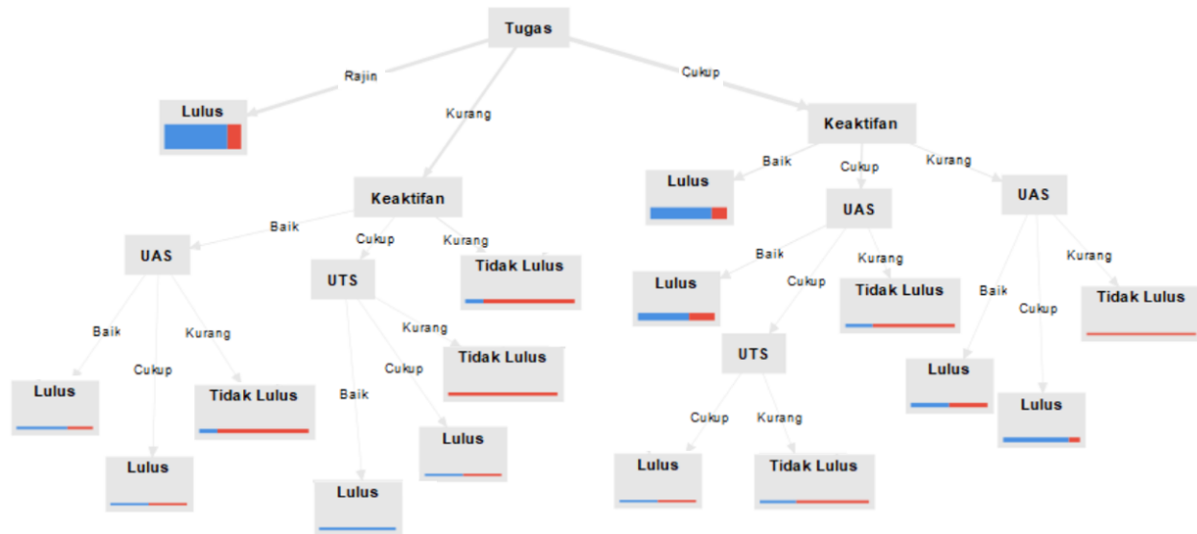
$$Gain(S, A) = 0.898 - \left(\frac{47}{137} * 0.903\right) + \left(\frac{45}{137} * 0.956\right) + \left(\frac{45}{137} * 0.802\right) = 0.010$$

Hasil perhitungan entropy dan gain ditunjukkan pada tabel 3.

Tabel 3. Perhitungan entropy dan gain

Total	Atribut	Jumlah kasus	Lulus	Tidak Lulus	Entropi	Gain
		137	94	43	0.898	
Keaktifan	Baik	47	32	15	0.903	0.010
	Cukup	45	28	17	0.956	
	Kurang	45	34	11	0.802	
Tugas	Rajin	56	36	20	0.940	0.034
	Cukup	45	37	8	0.675	
	Kurang	36	21	15	0.980	
UTS	Baik	61	43	18	0.875	0.016
	Cukup	35	20	15	0.985	
	Kurang	41	31	10	0.801	
UAS	Baik	54	35	19	0.936	0.003
	Cukup	45	32	13	0.867	
	Kurang	38	27	11	0.868	

Berdasarkan perhitungan yang terdapat pada tabel 3 di atas, atribut tugas memperoleh nilai tertinggi, yaitu 0.034, sehingga atribut tugas dipilih sebagai root pohon keputusan, seperti yang terlihat pada gambar 2 sebagai berikut:



Gambar 2. Pohon keputusan algoritma C5.0

Tujuan utama dari penelitian dengan menggunakan *decision tree*/algoritma C5.0 yaitu menghasilkan pola/rule [21], [22]. Dari output pohon keputusan pada gambar 2, maka didapatkan rules yang terbentuk pada gambar 3 di bawah ini:

```
Tugas = Rajin: Lulus {Lulus = 37, Tidak Lulus = 8}
Tugas = Kurang:
| Keaktifan = Baik
| | UAS = Baik: Lulus {Lulus = 5, Tidak Lulus = 1}
| | UAS = Cukup: Lulus {Lulus = 3, Tidak Lulus = 1}
| | UAS = Kurang: Tidak Lulus {Lulus = 2, Tidak Lulus = 1}
| Keaktifan = Cukup
| | UTS = Baik: Lulus {Lulus = 4, Tidak Lulus = 0}
| | UTS = Cukup: Lulus {Lulus = 1, Tidak Lulus = 1}
| | UTS = Kurang: Tidak Lulus {Lulus = 1, Tidak Lulus = 5}
| Keaktifan = Kurang: Tidak Lulus {Lulus = 1, Tidak Lulus = 13}
Tugas = Cukup:
| Keaktifan = Baik: Lulus {Lulus = 16, Tidak Lulus = 4}
| Keaktifan = Cukup
| | UAS = Baik: Lulus {Lulus = 8, Tidak Lulus = 4}
| | UAS = Cukup
| | | UTS = Cukup: Lulus {Lulus = 3, Tidak Lulus = 1}
| | | UTS = Kurang: Tidak Lulus {Lulus = 0, Tidak Lulus = 2}
| | UAS = Kurang: Tidak Lulus {Lulus = 1, Tidak Lulus = 3}
| Keaktifan = Kurang
| | UAS = Baik: Lulus {Lulus = 3, Tidak Lulus = 1}
| | UAS = Cukup: Lulus {Lulus = 6, Tidak Lulus = 1}
| | UAS = Kurang: Tidak Lulus {Lulus = 0, Tidak Lulus = 2}
```

Gambar 3. Rules hasil algoritma C5.0

Setelah diperoleh output model *decision tree*, langkah selanjutnya yaitu menghitung akurasi dari model yang dibuat. Pada *decision tree* digunakan confusion matrix yang memuat nilai *accuracy*, *class recall*, dan *class precision*. Berikut ini merupakan hasil *confusion matrix* dari *decision tree* yang telah dibuat:

accuracy: 88.33% +/- 15.32% (micro average: 88.57%)

	true Lulus	true Tidak Lulus	class precision
pred. Lulus	28	3	90.32%
pred. Tidak Lulus	1	3	75.00%
class recall	96.55%	50.00%	

Gambar 4. Hasil confusion matrix

Dari gambar 4 di atas, dapat diketahui bahwa model *decision tree* pada penelitian ini memiliki tingkat akurasi 88.33%. Hal ini menunjukkan bahwa dalam data yang diolah pada penelitian ini 88.33% keputusannya dapat digunakan secara akurat.

4. KESIMPULAN

Berdasarkan pembahasan di atas, maka dapat disimpulkan bahwa Algoritma C5.0 efektif digunakan untuk memprediksi tingkat keberhasilan mahasiswa pada mata kuliah RPL dengan mempertimbangkan atribut keaktifan di kelas, nilai tugas, UTS, dan UAS. Nilai *gain* tertinggi didapatkan dari atribut tugas, dimana atribut tersebut digunakan sebagai *root* dari pohon keputusan. Penelitian ini menggunakan 172 data sampel yang terdiri dari 137 data *training* dan 35 data *testing*. Dimana pengujian ini menghasilkan tingkat akurasi sebesar 70% untuk data *training* dan memiliki tingkat akurasi 88.33% untuk data *testing*.

DAFTAR PUSTAKA

- [1] Y. Supriya *et al.*, "Industry 5.0 in Smart Education: Concepts, Applications, Challenges, Opportunities, and Future Directions," *IEEE Access*, vol. 12, pp. 81938–81967, 2024, doi: 10.1109/ACCESS.2024.3401473.
- [2] M. Sharma, S., & Sharma, "Application of Machine Learning Algorithms in Education: A Review," *J. Comput. Appl.*, vol. 10, no. 4, pp. 22–29, 2018.
- [3] M. Khan, A., & Khan, "Prediction of Student Performance Using C5.0 Algorithm," *Int. J. Comput. Appl.*, vol. 179, no. 10, pp. 27–33, 2018.
- [4] X. Wang, Y., & Li, "Prediction of Student Performance in Higher Education Using Classification Algorithms," *Int. J. Inf. Educ. Technol.*, vol. 6, no. 7, pp. 506–510, 2016.
- [5] I. Villarrasa-Sapiña, X. García-Massó, E. Liébana, and G. Monfort Torres, "Academic achievement prediction in secondary education by decision tree analysis," *Educ. XXI*, vol. 27, no. 1, pp. 253–279, Jan. 2024, doi: 10.5944/educxxi.33351.
- [6] M. Pasic, A. Vates, F. Ferizbegovic, H. Bajric, and M. Pasic, "Machine learning prediction and analysis of students' academic performance," *Period. Eng. Nat. Sci.*, vol. 11, no. 5, p. 27, Oct. 2023, doi: 10.21533/pen.v11i5.3793.
- [7] K. Nahar, B. I. Shova, T. Ria, H. B. Rashid, and A. H. M. S. Islam, "Mining educational data to predict students performance," *Educ. Inf. Technol.*, vol. 26, no. 5, pp. 6051–6067, Sep. 2021, doi: 10.1007/s10639-021-10575-3.
- [8] A. Kumar, P., & Rathi, "Predicting Student Dropout in Higher Education Using Machine Learning Algorithms," *Int. J. Comput. Sci. Inf. Secur.*, vol. 17, no. 8, pp. 1–6, 2019.
- [9] Zeindri Saputra, "analisis dan perancangan aplikasi untuk klasifikasi menggunakan KNN," 2011. [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI
- [10] M. Sathe and A. Adamuthe, "Comparative Study of Supervised Algorithms for Prediction of Students' Performance," *Int. J. Mod. Educ. Comput. Sci.*, 2021, doi: 10.5815/IJMECS.2021.01.01.
- [11] P. C. Mandal, I. Mukherjee, G. Paul, and B. N. Chatterji, "Digital image steganography: A literature survey," *Inf. Sci. (N.Y.)*, vol. 609, pp. 1451–1488, Sep. 2022, doi: 10.1016/j.ins.2022.07.120.
- [12] J. C.-W. Lin, N. Sahoo, G. Srivastava, and W. Ding, "Introduction to the Special Issue on Pattern-Driven Mining, Analytics, and Prediction for Decision Making, Part II," *ACM Trans. Manag. Inf. Syst.*, vol. 13, no. 3, pp. 1–3, Sep. 2022, doi: 10.1145/3512468.
- [13] J. Olufemi Ogunleye, "The Concept of Data Mining," 2022. doi: 10.5772/intechopen.99417.
- [14] D. Dalbergio, M. N. Hayati, and Y. N. Nasution, "Klasifikasi Lama Studi Mahasiswa Menggunakan Metode C5.0 pada Studi Kasus Data Keberhasilan Mahasiswa Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Mulawarman," *Semin. Nas. Mat. Stat. dan Apl.*, vol. 1, no. 1, pp. 36–42, 2019.
- [15] A. C. Wijaya, N. A. Hasibuan, and P. Ramadhani, "Implementasi Algoritma C5 . 0 Dalam Klasifikasi Pendapatan Masyarakat (Studi Kasus : Kelurahan Mesjid Kecamatan Medan Kota)," *J. Inform. dan Teknol.*, vol. 13, pp. 192–198, 2018.
- [16] M. Pardede, E. Buulolo, and E. Ndruru, "Implementasi Algoritma C5.0 Pada Keberhasilan Peserta Ujian Kemahiran Berbahasa Indonesia (Ukbi) Pada Balai Bahasa Sumatera Utara," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komput.*, vol. 3, no. 1, pp. 64–72, 2019.
- [17] M. S. Sungkar and M. T. Qurohman, "Penerapan Algoritma C5.0 Untuk Prediksi Keberhasilan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1166, 2021, doi: 10.30865/mib.v5i3.3116.
- [18] J. Sinta and A. S. Sinaga, "Data Mining Algoritma C4.5 Pada Klasifikasi Kredit Koperasi Simpan Pinjam," *J. Ilmu Elektro Komput. dan Inform.*, vol. 4, no. 2, pp. 144–154, 2018.
- [19] R. Pratiwi, M. N. Hayati, and S. Prangga, "Perbandingan Klasifikasi Algoritma C5.0 Dengan Classification and Regression Tree (Studi Kasus : Data Sosial Kepala Keluarga Masyarakat Desa Teluk Baru Kecamatan Muara Ancalong Tahun 2019)," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 14, no. 2, pp. 273–284, 2020, doi: 10.30598/barekengvol14iss2pp273-284.
- [20] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *INOVTEK Polbeng - Seri Inform.*, vol. 1, no. 2, p. 126, 2016, doi: 10.35314/isi.v1i2.131.
- [21] I. H. Witten and E. Frank, *Data Mining Practical Machine Learning Tools and Techniques*. 2017. [Online]. Available: <http://books.google.com/books?hl=en&lr=&id=QTnOcZJzUoC&oi=fnd&pg=PR17&dq=Data+Mining+Practical+Machine+Learning+Tools+and+Techniques&ots=3gpDdrWiOc&sig=TZS7G8l1eXSa2SpAvfD6aBoJ2lw>

- [22] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, pp. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838.