



# Prediksi Kelulusan Mahasiswa Strata 1 (S1) Menggunakan Metode C5.0 di Program Studi Ilmu Komputer

Mubarak Ba'ayesh\*, Ilka Zufria

Sains dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan  
Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara, Indonesia

Email: <sup>1,\*</sup>mubarakbaayeshhh@gmail.com, <sup>2</sup>ilkazufria@uinsu.ac.id

Email Penulis Korespondensi: mubarakbaayeshhh@gmail.com

Submitted: 03/10/2024; Accepted: 29/10/2024; Published: 31/10/2024

**Abstrak**—Kelulusan mahasiswa tepat waktu merupakan salah satu indikator penting keberhasilan Program Studi di perguruan tinggi. Namun, pada Program Studi Ilmu Komputer UIN Sumatera Utara, jumlah mahasiswa yang lulus tepat waktu tergolong rendah. Penelitian ini bertujuan memprediksi kelulusan mahasiswa menggunakan algoritma C5.0, bagian dari pohon keputusan (Decision Tree), untuk memberikan gambaran mahasiswa yang lulus tepat waktu dan terlambat. Data yang digunakan meliputi nilai Indeks Prestasi (IP) semester 1 hingga 4, jumlah SKS, lama tugas akhir, masa studi, dan jalur masuk. Dari 100 data yang dianalisis, model diuji dengan perbandingan data 70:30. Hasil evaluasi menunjukkan bahwa model prediksi dengan algoritma C5.0 mencapai akurasi 100%, serta nilai precision, recall, dan f1-score sebesar 1.00 untuk kedua kelas, yaitu "Tepat Waktu" dan "Terlambat." Penelitian ini membuktikan bahwa algoritma C5.0 mampu memprediksi kelulusan mahasiswa secara akurat dan dapat membantu universitas dalam mengembangkan strategi peningkatan kelulusan tepat waktu.

**Kata Kunci:** Data Mining; Decision Tree; C5.0; Prediksi; Kelulusan Mahasiswa

**Abstract**—Timely graduation is a crucial indicator of the success of study programs in higher education. However, in the Computer Science Study Program at UIN Sumatera Utara, the number of students graduating on time is relatively low. This study aims to predict student graduation using the C5.0 algorithm, a part of Decision Tree, to classify students who graduate on time and those who graduate late. The data used includes GPA scores from semesters 1 to 4, total credits (SKS), final project duration, study period, and entry pathway. From 100 data samples, the model was tested using a 70:30 data split ratio. The evaluation results showed that the prediction model using the C5.0 algorithm achieved 100% accuracy, with Precision, recall, and f1-score values of 1.00 for both classes, namely "On Time" and "Late." This research demonstrates that the C5.0 algorithm can accurately predict student graduation and assist the university in developing strategies to improve timely graduation rates.

**Keywords:** Data Mining; Decision Tree; C5.0; Prediction; Student Graduation

## 1. PENDAHULUAN

Salah satu aspek penting dalam evaluasi keberhasilan penyelenggara Program Studi di suatu perguruan tinggi ialah mahasiswa. Dalam dunia Pendidikan, seperti ruang lingkup perguruan tinggi, mahasiswa selalu dituntut untuk memiliki kinerja akademik yang maksimal sebagai bahan untuk melakukan evaluasi. Sedangkan keterampilan mahasiswa untuk mengembangkan ilmu merupakan aspek penting dalam menentukan kinerja akademik mahasiswa itu sendiri [1].

Fakultas Sains dan Teknologi telah berdiri sejak 29 Desember 2015 yang di resmikan oleh Menteri Agama Republik Indonesia. Sampai saat ini Fakultas Sains dan Teknologi atau biasa disebut SAINTEK masih memiliki 5 Program Studi yaitu, Ilmu Komputer, Sistem Informasi, Matematika, Biologi dan Fisika. Salah satu Program Studi yang di jadikan studi kasus untuk penelitian kali ini adalah Program Studi Ilmu Komputer. Sebelumnya, sejak awal tahun 2000 an upaya untuk melakukan proses dan prosedur formal untuk merubah status IAIN SU menjadi UIN SU sudah di mulai secara intensif, dan terwujud pada Oktober 2014 IAIN SU resmi beralih status kelembagaannya menjadi UIN SU. Upaya tersebut juga mendapat dukungan dari berbagai belah pihak, selain itu dukungan nyata untuk alih status menjadi UIN SU juga telah mendapat bantuan komitmen bantuan dana pembiayaan pembangunan dan pengembangan kampus dari Islamic Development Bank (IsDB) dan Government of Indonesia (GoI).

Pada awal terbentuk Program Studi Strata 1 (S1) Ilmu Komputer ini menggunakan fasilitas gedung perpustakaan dan kemudian beralih ke gedung H. Anif serta gedung Pascasarjana yang berupa ruanagan kuliah, ruangan Program Studi dan ruangan administrasi. Dan kini Program Studi Ilmu Komputer menjadi salah satu pilihan favorit di Universitas Islam Negeri Sumatra Utara Medan, dengan banyaknya calon mahasiswa yang mendaftar untuk masuk dan menjadi mahasiswa di Program Studi tersebut. Oleh karna itu Program Studi Ilmu Komputer Sains dan Teknologi UIN Sumatera Utara seharusnya memiliki tindakan khusus untuk menyeimbangkan antara mahasiswa masuk dan mahasiswa keluar, dan salah satu cara untuk mengatasi permasalahan tersebut adalah mengontrol ketepatan waktu kelulusan mahasiswa [2].

Dalam website [ilkomp.ac.uinsu.id](http://ilkomp.ac.uinsu.id) jumlah total mahasiswa di Prodi Ilmu Komputer UIN SU dari Angkatan 2015 sampai dengan 2022 sudah mencapai angka 1.626 mahasiswa, ini memperkuat bahwa tidak selarasnya jumlah mahasiswa masuk dan mahasiswa lulus. Dan angka kelulusan di Program Studi ini tergolong sangat kecil, terutama untuk mahasiswa yang sudah tidak memiliki beban mata kuliah selain tugas akhir. Kemalasan menjadi sebuah faktor yang membuat masalah ini menjadi berkepanjangan dan mungkin tidak bisa diselesaikan. Selain itu



Program Studi Ilmu Komputer juga menetapkan persyaratan untuk mencapai kelulusan, yaitu [3]: 1) Telah menyelesaikan perkuliahan yang jumlah SKSnya telah ditentukan termasuk, Kerja Praktik, Kuliah Kerja Nyata (KKN), serta Seminar Proposal (SEMPRO) dan Skripsi; 2) Tidak memiliki nilai E sama sekali dari perkuliahan yang sudah diambil; 3) Nilai D tidak boleh lebih dari 7 SKS; 4) Pada mata kuliah Metodologi Penelitian nilai minimal C.

Dari 4 point yang menjadi persyaratan untuk mencapai kelulusan diatas termasuk syarat yang umum dan pastinya tidak begitu rumit, karena hampir seluruh perguruan tinggi menggunakan persyaratan yang sama, dan menurut penulis UIN SU merupakan salah satu dari berbagai Perguruan Tinggi Negeri di Sumatera Utara yang termasuk mudah untuk mendapatkan nilai dari para dosennya, dengan kata lain, untuk mencapai kelulusan tidak terlalu sulit, selagi mahasiswa mau serta mampu mengikuti prosedur yang dijalankan oleh Program Studi Ilmu Komputer sendiri.

Merujuk pada Standar Nasional Perguruan Tinggi Nomor 44 tahun 2015, salah satu kewajiban Universitas adalah mengontrol lama studi yang ditempuh mahasiswa yaitu maksimal 7 tahun akademik untuk program sarjana. Permasalahan mengenai tingkat kelulusan yang tepat waktu di Program Studi Ilmu Komputer memang cukup rendah. Fakta lapangan juga mengatakn bahwa rata-rata mahasiswa Program Studi Ilmu Komputer lulus dengan waktu yang kurang memuaskan yaitu diatas 4,5 tahun atau biasa disebut melewati 9 semester dan telah dipastikan lulus tidak tepat waktu.

Bagi pihak Program Studi mungkin ini menjadi perhatian lebih untuk dibenahi kedepannya. Maka dari itu, dilakukan penelitian untuk memprediksi kelulusan dari masing-masing mahasiswa. Program Studi Ilmu Komputer sendiri memiliki data yang banyak untuk kelulusan mahasiswa di jenjang Strata 1 (S1), seharusnya sebuah proses dapat dilakukan untuk menemukan informasi dengan data-data yang telah terkumpul. Namun pihak Program Studi masih belum melakukan penggalian informasi mengenai prediksi kelulusan mahasiswa Program Studi Ilmu Komputer UIN Sumatera Utara dan hal tersebut mungkin menjadi hambatan yang di fokuskan.

Hal tersebut dikarenakan belum adanya penggalian data yang akan di proses. Proses alternatif yang digunakan untuk menuntaskan masalah tersebut adalah data mining, dan akhirnya akan terbentuk sebuah informasi baru dengan cara memproses data-data yang telah di klasifikasikan sehingga prediksi mengenai kelulusan akan di selesaikan [4].

Data Mining sendiri adalah salah satu cabang ilmu komputer yang mampu melakukan prediksi, karena manajemen yang dapat melakukan prediksi untuk kelulusan mahasiswa. Data Mining adalah teknik untuk menemukan pola-pola yang ada dalam data sebagai sebuah alat untuk membantu menjelaskan data tersebut dan membuat prakiraan dari data itu [5].

Pada kasus ini ilmu dari Data Mining digunakan sebagai bahan pertimbangan dalam mengambil sebuah keputusan yang lebih lanjut tentang apa saja faktor-faktor yang mempengaruhi kelulusan khususnya faktor dalam data induk mahasiswa [6]. Data mining adalah kumpulan teknik yang digunakan untuk menemukan pola-pola tersembunyi dalam data yang sudah dikumpulkan. Dalam data mining, terdapat istilah lain yang memiliki makna serupa, yaitu Knowledge Discovery in Database (KDD). Keduanya memiliki tujuan yang sama, yaitu memanfaatkan data yang ada dalam basis data dan mengolahnya untuk memperoleh informasi baru yang bermanfaat [7].

Berdasarkan masalah yang dijelaskan sebelumnya, C5.0 merupakan metode yang akan dilakukan pada penelitian ini, yaitu merancang sebuah sistem untuk mengolah data mahasiswa Program Studi Ilmu Komputer. C5.0 merupakan salah satu bagian dari Decision Tree. Dibuatnya sistem ini untuk memprediksi mahasiswa yang lulus tepat waktu atau lulus dengan terlambat sehingga menjadi evaluasi bagi para mahasiswa dan pihak Universitas. Decision Tree atau pohon keputusan adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Pohon keputusan juga dapat digunakan untuk mengestimasi nilai dari tabe continue. Banyak algoritma yang dapat di pakai dalam pembentukan pohon keputusan, antara lain ID3, Classification and Regression Tree (CART), C4.5 dan C5 [8][9].

Pada penelitian yang dilakukan oleh Puspitaningrum, dkk, tingkat akurasi aplikasi tersebut berada diatas 70% dengan menggunakan 12 data training dan 24 data testing. Menurut Sobri, dalam jurnal yang berjudul "Penerapan Algoritma C5.0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Mata kuliah Arsitektur Sistem Komputer" dijelaskan bahwa proses prediksi dilakukan berdasarkan dengan klasifikasi algoritma C5.0 dengan beberapa atribut, dan mendapati akurasi yang tinggi sebesar 93,33%.

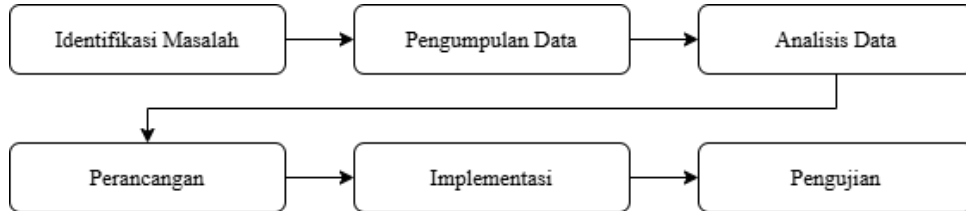
Pada penelitian lainnya yang dilakukan oleh Fitrihanah, dkk menunjukkan hasil akurasi dari algoritma C5.0 cukup baik yakni 93.75% dibanding dengan algoritma Naive Bayes yang hanya sebesar 81.25%. Kemudian pada penelitian lain, pada proses evaluasi yang dilakukan menggunakan confusion matrix memperoleh tingkat akurasi sebesar 91% dan akurasi excellent classification sebesar 0.989 berdasarkan kurva ROC [10].

Selanjutnya, pada penelitian yang dilakukan oleh Zebua, dkk yang melakukan perbandingan algoritma C5.0 dengan regression linear untuk memprediksi kelulusan mahasiswa memperoleh R2 score dari algoritma C5.0 mencapai 93.72% pada data pengujian. Sedangkan hasil R2 score dari regression linear hanya mencapai 40.3%. Penelitian ini menggunakan algoritma C5.0 untuk memprediksi kelulusan mahasiswa berdasarkan nilai indeks prestasi (IP) semester 1 sampai semester 4, jumlah satuan kredit semester (SKS), lama tugas akhir, total masa studi, dan jalur masuk [11].

## 2. METODOLOGI PENELITIAN

### 2.1 Kerangka Penelitian

Pada tahapan ini yang dilakukan adalah mengidentifikasi masalah yang ada dan berfungsi sebagai bahan untuk memahami masalah yang akan diteliti. Dan pada kasus yang akan diteliti adalah bagaimana membangun sebuah sistem yang dapat melakukan prediksi kelulusan mahasiswa di Program Studi Ilmu Komputer Universitas Islam Negeri Sumatera Utara Medan menggunakan metode Decision Tree C5.0 [12].



**Gambar 1.** Kerangka Penelitian

Berdasarkan gambar 1 dapat diketahui tahapan awalan penelitian adalah Mengidentifikasi Masalah secara jelas sehingga masalah dapat dipecahkan secara sistematis. Kedua, Pengumpulan Data kelulusan mahasiswa yang dilakukan pada Program Studi Ilmu Komputer Sains dan Teknologi UIN Sumatera Utara. Ketiga, Analisis Data sebagai tahap pre-processing yang mana data dengan nilai tidak sesuai atau kosong akan dihapus dan hanya menyisakan data yang siap untuk diproses. Keempat, Perancangan adalah tahapan yang bertujuan untuk merancang alur proses dari program yang akan dibuat dengan bentuk flowchart. Flowchart digunakan untuk memvisualisasikan tahapan penyelesaian masalah secara sederhana, terstruktur, rapi, dan jelas dengan menggunakan simbol-simbol yang telah ditetapkan [13]. Kelima, tahap Implementasi yang akan dilakukan dengan membuat model klasifikasi menggunakan script python pada editor google colab. Keenam, tahapan Pengujian akan dilakukan dengan menguji model berdasarkan tingkat akurasi.

a. Identifikasi Masalah

Pada tahapan ini yang dilakukan adalah mengidentifikasi masalah yang ada dan berfungsi sebagai bahan untuk memahami masalah yang akan diteliti. Dan pada kasus yang akan diteliti adalah berupa prediksi kelulusan mahasiswa di Program Studi Ilmu Komputer Universitas Islam Negeri Sumatera Utara Medan menggunakan metode Decision Tree C5.0. Prediksi merupakan sebuah proses yang dilakukan untuk memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi dimasa yang akan mendatang berdasarkan data yang dimiliki dimasa lalu dan juga masa kini [14]. Dalam kegiatan pengambilan keputusan yang ada dalam sistem informasi akademik tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk menggali dan mendapatkan informasi yang baru dari data yang ada. Data mining merupakan suatu proses untuk menemukan pola – pola didalam data, dengan menggunakan beberapa Teknik dimana proses tersebut dilakukan secara otomatis ataupun semi otomatis serta pola – pola yang ditemukan harus bermanfaat [15].

b. Pengumpulan Data

Adapun data atau kebutuhan informasi yang dibutuhkan untuk penelitian ini adalah dataset dengan parameter utama yakni nilai indeks prestasi (IP) semester 1–4, jumlah satuan kredit semester (SKS), lama tugas akhir (semester), masa studi (tahun), dan jalur masuk. Penelitian ini dilakukan di Program Studi Ilmu Komputer UIN Sumatera Utara Medan. Berdasarkan hasil pengumpulan data, diperoleh dataset dari 100 alumni mahasiswa. Pada tabel 1 dibawah ini merupakan 15 dari 100 dataset hasil dari tahap pengumpulan data yang diperoleh dari riset yang telah dilakukan.

**Tabel 1.** Dataset Kelulusan Mahasiswa

No.	NIM	IPS 1	IPS 2	IPS 3	IPS 4	SKS 4	Lama TA	Masa Studi	Jalur Masuk	Kelulusan
1	701182172	3.00	2.50	2.89	2.45	77	2	5	SBMPTN	TL
2	701183220	3.37	3.25	3.23	3.41	83	2	4.5	Mandiri	TL
3	701182143	3.21	3.45	3.68	3.26	84	2	4.5	SBMPTN	TL
4	701181046	3.05	2.65	3.40	2.77	81	5	5.5	SNMPTN	TL
5	701182140	3.42	3.90	3.87	3.54	86	2	4.5	SBMPTN	TL
6	701183242	3.11	3.25	3.59	3.08	85	4	5.5	Mandiri	TL
7	701183248	3.05	3.20	3.14	3.00	83	4	5.5	Mandiri	TL
8	701182122	3.42	3.20	3.27	3.00	83	4	5.5	SBMPTN	TL
9	701183234	3.16	3.05	3.29	2.95	81	3	5	Mandiri	TL
10	701183251	3.58	3.35	3.64	3.54	85	3	5	Mandiri	TL
11	701183258	3.84	4.00	3.87	3.50	86	2	4.5	Mandiri	TL

No.	NIM	IPS 1	IPS 2	IPS 3	IPS 4	SKS 4	Lama TA	Masa Studi	Jalur Masuk	Kelulusan
12	701182124	3.68	3.30	3.68	3.61	84	2	4.5	SBMPTN	TL
13	701182097	3.37	3.45	3.55	3.54	85	3	5	SBMPTN	TL
14	701183244	3.26	3.10	3.64	3.25	85	3	5	Mandiri	TL
15	701182113	3.79	3.70	3.74	3.42	86	5	6	SBMPTN	TL

c. Analisis Data

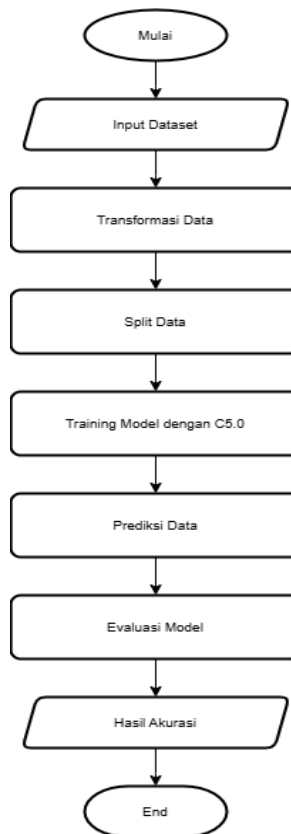
Analisis data ini bertujuan untuk membuat pohon keputusan dan menentukan mahasiswa yang mendapati lulus tepat waktu dan mahasiswa yang lulus tidak tepat waktu atau terlambat dengan menggunakan beberapa variabel yang akan menjadi bahan pertimbangan dari prediksi ini terlihat pada tabel 2 berikut.

**Tabel 2.** Tabel Atribut Data

No.	Nama Variable	Atribut
1.	Identitas Mahasiswa	NIM
2.	Nilai	IPS1, IPS2, IPS3, IPS4
3.	Jumlah Kredit	SKS Semester 4
4.	Periode Perkuliahan	Lama TA dan Masa Studi
5.	Jalur Masuk	Jalur Masuk

d. Perancangan

Untuk melakukan sebuah penelitian kiranya harus sudah mempunyai sebuah rancangan yang nantinya akan dikembangkan kedalam penelitian. Berikut merupakan flowchart yang menggambarkan alur proses sistem dalam melakukan prediksi kelulusan mahasiswa.



**Gambar 2.** Flowchart Sistem

Pada gambar 2 menunjukkan gambar dari flowchart sistem yang dimulai dengan memasukkan dataset kelulusan mahasiswa, data yang dimasukkan pada tahap input dipastikan merupakan data yang telah dinormalisasi. Selanjutnya dilanjutkan dengan tahapan transformasi data, yang mana data di-input akan diubah menjadi data berbentuk kategorikal. Kemudian masuk ke tahap split data, dimana pada tahap ini akan dilakukan pembagian dataset dengan persentase tertentu menjadi data training dan data testing. Dalam hal ini, tingkat presisi suatu prediksi sangat dipengaruhi oleh jumlah dataset [16]. Semakin banyak jumlah dataset, maka hasil tingkat presisi suatu prediksi semakin tinggi. Selanjutnya masuk ke tahapan training model menggunakan decision tree dengan algoritme C5.0. Tahapan training model akan menghasilkan model pohon keputusan yang akan digunakan pada tahapan prediksi. Selanjutnya melakukan prediksi atau pengujian data berdasarkan



data testing yang sudah dibagi pada tahap split data. Proses terakhir adalah melakukan evaluasi dari model yang dihasilkan dari tahap training. Tahapan evaluasi yang dilakukan menggunakan confusion matrix akan menghasil nilai yang berguna untuk mengukur akurasi, recall, dan presisi dari model [17]. Semakin tinggi nilai akurasi dari suatu model maka akan semakin baik model tersebut untuk digunakan sebagai model prediksi dalam konteks penelitian ini adalah kelulusan mahasiswa pada Program Studi Ilmu Komputer UIN Sumatera Utara Medan.

e. Implementasi

Algoritma C5.0 merupakan algoritma data mining dan terkhususnya diterapkan pada teknik decision tree, algoritma C5.0 juga merupakan penyempurnaan algoritma sebelumnya yaitu ID3 dan C4.5 yang dibentuk oleh Ross Quinlan pada tahun 1987 [18]. Selain itu dalam hal akurasi dan memori algoritma ini juga dinilai lebih baik dari algoritma sebelumnya, serta dapat menangani atribut yang bernilai diskrit maupun kontinu. Kelebihan inilah yang membuat algoritma C5.0 lebih unggul dari algoritma sebelumnya. Pemilihan atribut diproses menggunakan information gain. Selanjutnya dipilih atribut yang memiliki nilai gain tertinggi untuk dijadikan root node. Berikut merupakan persamaan untuk menghitung entropy atribut, yaitu [19]:

$$(S_1, S_2, \dots, S_m) = \sum_{i=0}^m P_i \log_2 P_i \tag{1}$$

Keterangan dimana S adalah himpunan kasus, m adalah jumlah atribut class dan pi adalah proporsi dari S1 terhadap S. Sementara untuk menghitung pada nilai gain dapat dilihat pada persamaan berikut [20].

$$E(A) = \sum_{j=1}^y \frac{S_{1j} + \dots + S_{mj}}{S_{1j} + \dots + S_{mj}} \tag{2}$$

$$\sum_{j=1}^y \frac{S_{1j} + \dots + S_{mj}}{S_{1j} + \dots + S_{mj}} \tag{3}$$

merupakan jumlah dari subset "j" yang dibagi dengan jumlah sampel pada S, maka untuk menemukan nilai gain, selanjutnya digunakan formula. Maka nilai gain dapat dihitung menggunakan rumus yang dapat dilihat pada persamaan dibawah ini.

$$Gain A = (S_1, S_2, \dots, S_m) - E(A) \tag{4}$$

Keterangan, A adalah atribut, S adalah himpunan kasus dan S1 adalah jumlah sampel.

f. Pengujian

Pada bagian penulis menggunakan 100 mahasiswa yang diperoleh dari Program Studi Ilmu Komputer Fakultas Sains dan Teknologi Universitas Islam Negeri Sumatera Utara Medan. Dan harapannya dari penelitian untuk prediksi kelulusan mahasiswa akan mencapai akurasi yang memuaskan ataupun berada di angka 80% - 100%, agar kedepannya algoritme C5.0 dapat digunakan untuk permasalahan atau penelitian terkait prediksi (Fathurrohman, 2022).

### 3. HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan tentang hasil penelitian dan hasil dari pembuatan serta serta pembahasannya. Pada penelitian ini menjelaskan tentang prediksi kelulusan mahasiswa dengan menerapkan algoritma C5.0. Pembahasan bertujuan untuk menjelaskan bagaimana penelitian ini menjawab dari semua rumusan masalah dan tujuan dari sistem yang akan dibuat.

#### 3.1 Hasil

a. Dataset

Adapun data atau kebutuhan informasi yang dibutuhkan untuk penelitian ini adalah dataset dengan parameter utama yakni nilai indeks prestasi (IP) semester 1 – 4, jumlah satuan kredit semester (SKS), lama tugas akhir (semester), masa studi (tahun), dan jalur masuk. Penelitian ini dilakukan di Program Studi Ilmu Komputer UIN Sumatera Utara Medan. Berdasarkan hasil pengumpulan data, diperoleh dataset dari 100 alumni mahasiswa. Pada tabel 2 berikut merupakan 15 dari 100 dataset hasil dari tahap pengumpulan data yang diperoleh dari riset yang telah dilakukan.

**Tabel 2.** Dataset Kelulusan Mahasiswa

No.	NIM	IPS 1	IPS 2	IPS 3	IPS 4	SKS 4	Lama TA	Masa Studi	Jalur Masuk	Kelulusan
1	701182172	3.00	2.50	2.89	2.45	77	2	5	SBMPTN	TL
2	701183220	3.37	3.25	3.23	3.41	83	2	4.5	Mandiri	TL
3	701182143	3.21	3.45	3.68	3.26	84	2	4.5	SBMPTN	TL
4	701181046	3.05	2.65	3.40	2.77	81	5	5.5	SNMPTN	TL
5	701182140	3.42	3.90	3.87	3.54	86	2	4.5	SBMPTN	TL
6	701183242	3.11	3.25	3.59	3.08	85	4	5.5	Mandiri	TL

No.	NIM	IPS 1	IPS 2	IPS 3	IPS 4	SKS 4	Lama TA	Masa Studi	Jalur Masuk	Kelulusan
7	701183248	3.05	3.20	3.14	3.00	83	4	5.5	Mandiri	TL
8	701182122	3.42	3.20	3.27	3.00	83	4	5.5	SBMPTN	TL
9	701183234	3.16	3.05	3.29	2.95	81	3	5	Mandiri	TL
10	701183251	3.58	3.35	3.64	3.54	85	3	5	Mandiri	TL
11	701183258	3.84	4.00	3.87	3.50	86	2	4.5	Mandiri	TL
12	701182124	3.68	3.30	3.68	3.61	84	2	4.5	SBMPTN	TL
13	701182097	3.37	3.45	3.55	3.54	85	3	5	SBMPTN	TL
14	701183244	3.26	3.10	3.64	3.25	85	3	5	Mandiri	TL
15	701182113	3.79	3.70	3.74	3.42	86	5	6	SBMPTN	TL

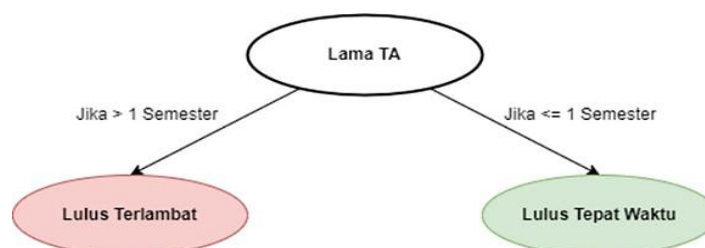
Berdasarkan dataset pada tabel 2, maka akan diuraikan penjelasannya sebagai berikut.

1. IPS merupakan singkatan Indeks Prestasi Semester yang pada tabel tersebut diperoleh dari semester 1 sampai semester 4 yang ditunjukkan dengan singkatan IPS 1, IPS 2, IPS 3, dan IPS 4.
  2. SKS merupakan singkatan satuan kredit semester, yang pada tabel tersebut diperoleh dari total jumlah SKS mahasiswa pada semester 4.
  3. Lama TA merupakan lama masa pengerjaan tugas akhir mahasiswa yang dihitung dalam satuan semester.
  4. Masa Studi merupakan masa studi yang ditempuh mahasiswa sampai memperoleh gelar sarjana komputer yang dihitung dalam satuan tahun.
  5. Jalur Masuk merupakan jalur masuknya mahasiswa dalam penerimaan kampus, yang mana terbagi menjadi Seleksi Bersama Mahasiswa Perguruan Tinggi Negeri (SBMPTN), Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) atau melalui jalur Mandiri.
- b. Penerapan Algoritma C5.0  
 Dengan menerapkan algoritma C5.0 pada data kelulusan mahasiswa maka dilakukan perhitungan manual untuk membandingkan hasil perancangan sistem prediksi kelulusan mahasiswa dengan hasil perhitungan menggunakan algoritme C5.0.

**Tabel 3.** Dataset Kelulusan Mahasiswa

Atribut	Kriteria	Jlh Kasus	Terlambat	Tepat W.	Entropy	Gain	Split Info	G. Ratio
IPS 1	$\geq 3.00$	99	72	27	0.8454	0.0046	0.0808	0.0565
	$< 3.00$	1	1	0	0			
IPS 2	$\geq 3.00$	97	70	27	0.8532	0.0139	0.1944	0.0713
	$< 3.00$	3	3	0	0			
IPS 3	$\geq 3.00$	99	72	27	0.8454	0.0046	0.0808	0.0565
	$< 3.00$	1	1	0	0			
IPS 4	$\geq 3.00$	94	67	27	0.8651	0.0283	0.3274	0.0862
	$< 3.00$	6	6	0	0			
SKS 4	$\geq 80$	98	71	27	0.8493	0.0092	0.1414	0.0649
	$< 80$	2	2	0	0			
Lama TA (semester)	$\leq 1$	27	0	27	0	0.8415	0.8415	1
	$> 1$	73	73	0	0			
Masa Std. (tahun)	$\leq 4$	27	0	27	0	0.8415	0.8415	1
	$> 4$	73	73	0	0			
Jalur Masuk	Mandiri	31	25	6	0.7088			
	SBMPTN	40	30	10	0.8113	0.5749	1.5705	0.366
	SNMPTN	29	18	11	0.9576			

Berdasarkan hasil perhitungan pada tabel 3, diperoleh nilai tertinggi pada parameter Lama TA dan Masa Studi dengan nilai gain ratio adalah 1. Parameter dengan nilai gain ratio tertinggi akan dijadikan prioritas dalam struktur pohon Keputusan. Dengan begitu Lama TA atau Masa Studi dapat dijadikan sebagai akar (root) dari struktur pohon keputusan. Berdasarkan hasil perhitungan tersebut maka dapat digambarkan struktur pohon keputusan sebagai berikut.



**Gambar 3.** Pohon Keputusan

### 3.2 Pembahasan

Berdasarkan analisis yang telah dilakukan pada pembahasan sebelumnya, maka hasil analisis akan ditampilkan pada platform Google Colab.

#### a. Input Data

Tahap awal dari sistem analisis kelulusan mahasiswa dengan algoritma C5.0 adalah melakukan penginputan dataset dengan ekstensi .xlsx. dataset yang di-input memiliki atribut id, ips1, ips2, ips3, ips4, sks4, lama\_ta, masa\_studi, jalur\_masuk dan label

```
from google.colab import files

# Baca dataset
file_path = 'dataset.xlsx'
data = pd.read_excel(file_path)
data.head(5)
```

	id	NIM	ips1	ips2	ips3	ips4	sks4	Lama TA	Masa Studi	jalur_masuk	label
0	1	701182172	3.00	2.50	2.89	2.45	77	2	5.0	SBMPTN	Terlambat
1	2	701183220	3.37	3.25	3.23	3.41	83	2	4.5	Mandiri	Terlambat
2	3	701182143	3.21	3.45	3.68	3.26	84	2	4.5	SBMPTN	Terlambat
3	4	701181046	3.05	2.65	3.40	2.77	81	5	5.5	SNMPTN	Terlambat
4	5	701182140	3.42	3.90	3.87	3.54	86	2	4.5	SBMPTN	Terlambat

**Gambar 4.** Input Data

#### b. Tranformasi Data

Transformasi data merupakan proses mengubah atau memanipulasi data dari bentuk atau format aslinya ke bentuk lain yang lebih sesuai untuk analisis atau pemrosesan lebih lanjut. Tranformasi data dapat dilihat pada gambar 5 dan 6.

```
# 2 = SNMPTN
mapping = {
    'Mandiri': 0,
    'SBMPTN': 1,
    'SNMPTN': 2
}
data['jalur_masuk'] = data['jalur_masuk'].map(mapping)

# Keterangan Lulus (Target)
# 1 = terlambat
# 0 = tepat waktu
#data['label'] = data.apply(lambda row: 1 if row['label'] == 'Terlambat' else 0, axis=1)

data.head(10)
```

	id	NIM	ips1	ips2	ips3	ips4	sks4	Lama TA	Masa Studi	jalur_masuk	label
0	1	701182172	>= 3.00	< 3.00	< 3.00	< 3.00	< 80	2	5.0	1	Terlambat
1	2	701183220	>= 3.00	>= 3.00	>= 3.00	>= 3.00	>= 80	2	4.5	0	Terlambat
2	3	701182143	>= 3.00	>= 3.00	>= 3.00	>= 3.00	>= 80	2	4.5	1	Terlambat

**Gambar 5.** Tranformasi Kategorikal

Setelah data awal selesai disusun dalam bentuk kategorial, maka data kategorial dapat ditranformasi menjadi data numerik agar dapat dapat dihitung. Hasil transformasi numerik dapat dilihat pada gambar 6.

```
# Lama TA (Fitur)
# <= 1 Semester = 1
# > 1 Semester = 0
# data['lama_ta'] = data.apply(lambda row: 1 if row['lama_ta'] == '<= 1 Semester' else 0, axis=1)

# Masa Studi (Fitur)
# <= 4 Tahun = 1
# > 4 Tahun = 0
# data['masa_studi'] = data.apply(lambda row: 1 if row['masa_studi'] == '<= 4 Tahun' else 0, axis=1)

# Memisahkan fitur dan target
x = data.drop(columns=['id', 'label']) # Fitur
y = data['label'] # Target

pd.options.display.max_rows = 200;
data = pd.DataFrame(x)
data.head(25)
```

	NIM	ips1	ips2	ips3	ips4	sks4	Lama TA	Masa Studi	jalur_masuk
0	701182172	1	0	0	0	0	2	5.0	1
1	701183220	1	1	1	1	1	2	4.5	0
2	701182143	1	1	1	1	1	2	4.5	1
3	701181046	1	0	1	0	1	5	5.5	2

**Gambar 6.** Tranformasi Numerik

c. Split Data

Pada tahap ini akan dilakukan pembagian data atau split data menjadi dua bagian, yakni data training dan data testing. Tujuan dari tahap ini adalah untuk memisahkan sebagian data sebagai data training yang akan digunakan untuk melatih model, sementara sisanya digunakan sebagai data testing untuk mengevaluasi performa model. Dengan melakukan split data, kita dapat memastikan bahwa model yang dibangun tidak hanya bekerja dengan baik pada data yang sudah dilihatnya (data training), tetapi juga mampu melakukan generalisasi dengan baik pada data baru atau yang belum pernah dilihat sebelumnya (data testing). Rasio yang digunakan adalah 70:30 dengan 70% data akan digunakan sebagai data training dan 30% data akan digunakan sebagai data testing.

```
# Membagi data menjadi data pelatihan dan pengujian
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=1)
x_train_data = pd.DataFrame(X_train)
x_train_data.head(100)
```

0	701102112	1	0	0	0	2	5.0	1
89	701193237	1	1	1	1	3	5.0	0
57	701203023	1	1	1	1	1	4.0	0
22	701191100	1	1	1	1	1	4.0	2
61	701202100	1	1	1	1	1	4.0	1
63	701183250	1	1	1	1	1	4.0	0
7	701182122	1	1	1	1	4	5.5	1

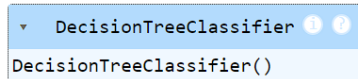
Gambar 7. Split Data

d. Training Model C5.0

Pada tahap ini, akan dilakukan pelatihan data menggunakan algoritma decision tree dengan berdasarkan atribut ips1, ips2, ips3, ips4, sks4, lama\_ta, masa\_studi, dan jalur masuk. Pelatihan data dapat dilihat pada gambar 8.

### 5: Melatih Model C5.0

```
] clf = DecisionTreeClassifier()
  clf.fit(X_train, y_train)
```



Gambar 8. Training Model

e. Prediksi atau Pengujian

Pada tahap ini, setelah model dilatih maka pada tahap ini akan dibuat prediksi menggunakan data pengujian. Tahap membuat prediksi dapat dilihat pada gambar 9.

```
y_pred = clf.predict(X_test)
print(y_pred)
```

```
['Terlambat' 'Terlambat' 'Tepat Waktu' 'Terlambat' 'Terlambat' 'Terlambat'
 'Tepat Waktu' 'Terlambat' 'Terlambat' 'Terlambat' 'Terlambat'
 'Tepat Waktu' 'Terlambat' 'Terlambat' 'Tepat Waktu' 'Terlambat'
 'Tepat Waktu' 'Terlambat' 'Terlambat' 'Terlambat' 'Terlambat'
 'Tepat Waktu' 'Tepat Waktu' 'Terlambat' 'Terlambat' 'Tepat Waktu'
 'Terlambat' 'Tepat Waktu' 'Terlambat' 'Terlambat']
```

Gambar 9. Prediksi atau Pengujian

f. Perhitungan Akurasi

Pada tahap ini, akurasi model akan dihitung untuk mengukur seberapa baik model dapat memprediksi label yang benar pada data pengujian. Akurasi dihitung dengan membandingkan hasil prediksi yang dihasilkan oleh model dengan label sebenarnya pada data pengujian. Akurasi yang tinggi menunjukkan bahwa model memiliki performa yang baik dalam memprediksi data yang belum pernah dilihat sebelumnya. Hasil akurasi model dapat dilihat pada gambar 10.

```
accuracy = accuracy_score(y_test, y_pred)
print(f'Akurasi: {accuracy * 100:.2f}%')
```

Akurasi: 100.00%

```
# Menampilkan laporan klasifikasi
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
Tepat Waktu	1.00	1.00	1.00	9
Terlambat	1.00	1.00	1.00	21
accuracy			1.00	30
macro avg	1.00	1.00	1.00	30
weighted avg	1.00	1.00	1.00	30

**Gambar 10.** Hasil Akurasi C5.0

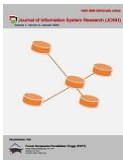
Berdasarkan hasil dari evaluasi model pada gambar 10, dapat diperoleh kesimpulan bahwa dari total 30 data yang digunakan sebagai data prediksi memperoleh akurasi sebesar 100 %, yang menunjukkan bahwa semua prediksi pada data pengujian benar. Metrik Evaluasi lainnya yakni precision, recall, dan f1-score untuk kedua kelas juga menunjukkan nilai 1.00, yang menegaskan bahwa model bekerja dengan sangat baik untuk kedua kelas. Tidak ada kesalahan klasifikasi yang dilakukan oleh model. Dari 30 data yang diprediksi dengan benar, terdapat 9 data yang termasuk dalam kelas “Tepat Waktu” dan 21 data yang termasuk dalam kelas “Terlambat”.

## 4. KESIMPULAN

Dalam penelitian ini, algoritma C5.0 digunakan sebagai model prediksi kelulusan mahasiswa. Algoritma ini diterapkan pada data kelulusan yang mencakup parameter seperti nilai indeks prestasi (IP) dari semester 1 hingga 4, jumlah satuan kredit semester (SKS), lama tugas akhir, masa studi, dan jalur masuk. Penelitian dilakukan dengan menggunakan dataset berjumlah 100 data. Pengujian model dilakukan dengan rasio perbandingan data 70:30. Dengan menggunakan algoritma C5.0, model yang dihasilkan berhasil menghasilkan prediksi yang memuaskan dalam membantu mengidentifikasi mahasiswa yang berpotensi lulus tepat waktu atau lulus tidak tepat waktu (terlambat). Berdasarkan evaluasi yang dilakukan, algoritma C5.0 yang diterapkan pada data kelulusan mahasiswa Program Studi Ilmu Komputer UIN Sumatera Utara berhasil mencapai akurasi 100%. Dengan nilai Precision, recall, dan f1-score yang masing-masing sebesar 1.00 menunjukkan bahwa algoritma ini mampu memisahkan data dengan sangat baik antara kelas “Tepat Waktu” dan “Terlambat” tanpa kesalahan klasifikasi. Algoritma ini tidak hanya membantu dalam memprediksi hasil, tetapi juga memungkinkan pengambilan keputusan yang lebih baik bagi pihak universitas dalam merancang strategi untuk meningkatkan kelulusan tepat waktu di masa yang akan datang.

## REFERENCES

- [1] R. Sulastini, A. P. B. Pandiangan, and R. Supu, “Pembentukan Brand Image Pada Program Studi Pendidikan Guru Madrasah Ibtidaiyah (PGMI) STAI Sangatta Kutai Timur,” *al-Afkar, J. Islam. Stud.*, vol. 6, no. 2, pp. 146–165, 2023.
- [2] W. Wiranda, “Perancangan Augmented Reality Tata Lokasi Gedung Dan Ruangannya Pada Kampus I Uinsu Medan Berbasis Android.” Universitas Islam Negeri Sumatera Utara, 2021.
- [3] A. Armansyah and R. K. Ramli, “Model Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes,” *Edumatic J. Pendidik. Inform.*, vol. 6, no. 1, pp. 1–10, 2022, doi: 10.29408/edumatic.v6i1.4789.
- [4] F. A. Sianturi, P. M. Hasugian, A. Simangunsong, and B. Nadeak, *DATA MINING: Teori dan Aplikasi Weka*, vol. 1. IOCS Publisher, 2019.
- [5] F. D. Pratama, I. Zufria, and T. Triase, “Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Penerima Program Indonesia Pintar,” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 1, pp. 77–84, 2022, doi: 10.36341/rabit.v7i1.2217.
- [6] M. Windarti, H. J. Prasetyo, and R. S. Lutfiyani, “ALGORITMA APRIORI UNTUK MENGUKUR KORELASI JURUSAN SEKOLAH DAN NILAI MATA KULIAH KONSENTRASI TERHADAP TINGKAT KELULUSAN MAHASISWA,” *Lap. Has. Penelit.*, 2021.
- [7] N. Nurahman and J. Susanto, “Klasterisasi Data Penerima Bantuan Langsung Tunai Menggunakan Algoritma K-Means,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, pp. 461–470, 2023.
- [8] M. D. Kahfi, F. R. Umbara, and H. Ashaury, “Prediksi Pengangguran Menggunakan Decision Tree Dengan Algoritma C5.0 Pada Data Penduduk Kecamatan Caringin Kabupaten Bogor,” *Informatics Digit. Expert*, vol. 4, no. 2, pp. 75–80, 2022.
- [9] T. R. Matondang, Y. Ramadhan Nasution, Armansyah, and M. Furqan, “Penerapan Algoritma C4.5 Pada Klasifikasi Status Gizi Balita,” *J. Fasilkom*, vol. 14, no. 1, pp. 216–225, 2024, doi: 10.37859/jf.v14i1.6941.
- [10] S. Suprpto, F. F. JT, and E. Edora, “Prediksi Pengangkatan Karyawan Dengan Metode Klasifikasi Algoritma C5.0



- (Studi Kasus CV. T-Pico Jaya Mandiri),” *J. SIGMA*, vol. 13, no. 1, pp. 35–40, 2022.
- [11] R. Rahim et al., “C4.5 classification data mining for inventory control,” *Int. J. Eng. Technol.*, vol. 7, no. March, pp. 68–72, 2018, doi: 10.14419/ijet.v7i2.3.12618.
- [12] M. Sari et al., “Metodologi penelitian,” *Glob. Eksek. Teknol.*, 2022.
- [13] W. FANI, “PENGEMBANGAN SISTEM INFORMASI PERPUSTAKAAN BERBASIS DESKTOP DI MTS SWASTA ISLAMIAH (YPI) PONTIANAK.” IKIP PGRI PONTIANAK, 2023.
- [14] M. S. Priadana and D. Sunarsi, *Metode penelitian kuantitatif*. Pascal Books, 2021.
- [15] D. Rusdianto and L. Zaelani, “Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Mengetahui Pola Peminjaman Buku di Perpustakaan Universitas Bale Bandung,” *J-SIKA| J. Sist. Inf. Karya Anak Bangsa*, vol. 2, no. 02, pp. 1–10, 2020.
- [16] P. H. Artanti, “Penerapan Neural Network dengan optimasi Ant Colony Optimization dan Backpropagation untuk membangun model prediksi diabetes tahap awal.” Universitas Islam Negeri Maulana Malik Ibrahim, 2023.
- [17] I. G. T. Isa, F. Elfaladonna, and I. Ariyanti, *Buku Ajar Sistem Pendukung Keputusan*. Penerbit NEM, 2022.
- [18] A. Apriyadi, M. R. Lubis, and B. E. Damanik, “Penerapan Algoritma C5.0 Dalam Menentukan Tingkat Pemahaman Mahasiswa Terhadap Pembelajaran Daring,” *Komputa J. Ilm. Komput. dan Inform.*, vol. 11, no. 1, pp. 11–20, 2022, doi: 10.34010/komputa.v11i1.7386.
- [19] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, “Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4. 5,” in *PRISMA, Prosiding Seminar Nasional Matematika*, 2020, pp. 64–71.
- [20] D. Damayanti, “Implementasi Algoritma C4. 5 Prediksi Produksi Komoditas Tanaman Perkebunan Berdasarkan Luas Lahan,” *Tin Terap. Inform. Nusant.*, vol. 2, no. 10, pp. 571–579, 2022.
- [21] M. A. FATHURROHMAN, “Penentuan Strategi Pengelolaan Coffee Shop di Yogyakarta dengan Mengidentifikasi Perilaku dan Karakteristik Konsumen Menggunakan Metode Association Rules dan Clustering (Studi Kasus Pada Mahasiswa Yogyakarta),” 2022.