
**KLASIFIKASI KETEPATAN WAKTU PENYELESAIAN STUDI BERDASARKAN
RIWAYAT AKADEMI MAHASISWA MENGGUNAKAN ALGORITMA C4.5**

Eunike Logo Buke¹, Tri Ana Setyarini², Petrus Katemba³

^{1,2,3}STIKOM Uyelindo Kupang

Email: nikelogobuke@gmail.com¹, setyarinitriana@gmail.com², petruskatemba@gmail.com³

Abstrak: Ketepatan waktu penyelesaian Studi merupakan salah satu indikator utama keberhasilan studi mahasiswa sekaligus penentu kualitas akreditasi institusi pendidikan tinggi. Kondisi di Program Studi Teknik Informatika STIKOM Uyelindo Kupang menunjukkan adanya ketidakseimbangan antara jumlah mahasiswa masuk dan lulusan tepat waktu yang seringkali dipengaruhi oleh akumulasi riwayat akademik semester sebelumnya. Untuk memetakan pola tersebut, dibutuhkan sistem yang mampu mengklasifikasi ketepatan waktu penyelesaian Studi berdasarkan data historis mahasiswa secara objektif. Algoritma C4.5 memiliki kemampuan dalam membangun model pohon keputusan *decision tree* yang transparan serta handal dalam menangani data bertipe numerik maupun kategorikal secara simultan. Data yang digunakan berupa rekam jejak akademik dari 300 mahasiswa angkatan 2019 hingga 2021 yang meliputi variabel IPK, jumlah SKS, dan mata kuliah prasyarat. Hasil penelitian diharapkan dapat memberikan model prediksi dini bagi program Studi untuk mengidentifikasi mahasiswa yang berisiko terlambat, sekaligus menjadi dasar dalam perumusan kebijakan intervensi akademik yang lebih efektif.

Kata Kunci: Algoritma C4.5, Data Mining, Klasifikasi, Riwayat Akademik, Studi.

Abstract: Timeliness of study completion is one of the main indicators of student success and also determines the quality of accreditation of higher education institutions. Conditions in the Informatics Engineering Study Program at STIKOM Uyelindo Kupang show an imbalance between the number of students entering and graduating on time, which is often influenced by the accumulation of academic history in previous semesters. To map this pattern, a system is needed that is able to classify the timeliness of study completion based on historical student data objectively. The C4.5 algorithm has the ability to build a transparent decision tree model and is reliable in handling both numeric and categorical data simultaneously. The data used are academic track records of 300 students from the 2019 to 2021 intakes, which include variables such as GPA, number of credits, and prerequisite courses. The results of this study are expected to provide an early prediction model for study programs to identify students at risk of being late, as well as serve as a basis for formulating more effective academic intervention policies.

Keywords: C4.5 Algorithm, Data Mining, Classification, Academic History, Study.

PENDAHULUAN

Studi atau Skripsi merupakan karya ilmiah wajib yang menjadi tahap final bagi mahasiswa untuk meraih gelar sarjana. Waktu yang dibutuhkan mahasiswa untuk menyelesaikan skripsi sangat bervariasi, mulai dari satu semester, dua semester, hingga lebih, yang pada akhirnya menjadi faktor penentu utama ketepatan waktu Studi. Kemampuan mahasiswa untuk menyelesaikan skripsi dalam rentang waktu ideal akan berdampak langsung pada ketepatan masa Studi secara keseluruhan. Bagi sebuah program Studi, tingginya angka kelulusan tepat waktu merupakan salah satu tolak ukur kualitas dan menjadi komponen penting dalam penilaian akreditasi oleh BAN-PT. Keterlambatan kelulusan ini seringkali berpuncak pada fase tugas akhir, namun akarnya berasal dari pola perjalanan akademik mahasiswa di semester-semester sebelumnya. Oleh karena itu, diperlukan sebuah analisis mendalam terhadap data historis untuk mengidentifikasi faktor-faktor penentu dari riwayat akademik yang dapat memprediksi potensi keterlambatan tersebut.

Pada Program Studi Teknik Informatika Stara Satu STIKOM Uyelindo Kupang, salah satu akar masalah yang teridentifikasi adalah adanya ketidakseimbangan antara jumlah mahasiswa yang diterima per angkatan dengan jumlah mahasiswa yang berhasil lulus tepat waktu. Untuk menemukan pola tersebut, penelitian ini akan berfokus pada analisis data kuantitatif dan objektif yang bersumber langsung dari riwayat akademik mahasiswa. Keterlambatan Studi bukanlah masalah yang muncul tiba-tiba, melainkan akumulasi dari berbagai catatan akademik. Variabel-variabel kunci yang akan dianalisis meliputi Indeks Prestasi Kumulatif (IPK) sebagai indikator performa umum, jumlah Satuan Kredit Semester (SKS) yang berhasil diselesaikan sebagai ukuran progres Studi, total mata kuliah yang harus diulang sebagai sinyal kesulitan belajar, serta nilai pada mata kuliah fondasi seperti Metodologi Penelitian dan Rekayasa Perangkat Lunak yang esensial untuk tugas akhir. Selain itu, data administratif seperti status pernah mengambil cuti akademik juga akan disertakan karena pengaruhnya yang langsung terhadap durasi Studi.

Untuk menjawab kebutuhan tersebut, penelitian ini akan memanfaatkan teknik data mining dengan metode klasifikasi, yaitu sebuah proses untuk memetakan data ke dalam kelas-kelas yang telah ditentukan. Menganalisis variabel-variabel riwayat akademik yang telah diidentifikasi, penelitian ini menerapkan pendekatan data mining dengan metode klasifikasi. Sejalan dengan urgensi masalah yang dikemukakan oleh (Manik, et.al., 2024) mengenai

dampak efisiensi kelulusan terhadap kualitas institusi, maka diperlukan sebuah solusi berbasis data. Oleh karena itu, penelitian ini akan mengolah 300 data riwayat akademik mahasiswa angkatan 2019, 2020, dan 2021 menggunakan Algoritma C4.5. Pemilihan metode ini didasarkan pada relevansinya dengan penelitian utama oleh (Kamil dan Cholil, 2020), yang telah membuktikan bahwa Algoritma C4.5 efektif dalam mengolah atribut-atribut akademik untuk menemukan pola-pola penentu. Algoritma ini dipilih karena kemampuannya dalam membangun model pohon keputusan (*decision tree*) yang dapat memetakan secara jelas variabel mana dari riwayat akademik yang menjadi faktor paling berpengaruh dalam memprediksi keterlambatan penyelesaian tugas akhir. Menurut referensi tersebut, algoritma ini sangat relevan karena kemampuannya yang andal dalam mengolah tipe data yang beragam secara simultan, baik data numerik maupun data kategorikal tanpa memerlukan transformasi yang kompleks. Selain itu, model pohon keputusan yang dihasilkannya bersifat transparan, sehingga tidak hanya memberikan prediksi, tetapi juga mengungkapkan hierarki kepentingan dari setiap variabel, menunjukkan faktor mana yang menjadi penentu paling dominan dalam keterlambatan penyelesaian tugas akhir.

Hasil akhir dari penelitian ini sebuah sistem aplikasi, sebuah model klasifikasi prediktif yang kaya akan wawasan. Solusi yang ditawarkan adalah pengetahuan berbasis data yang dapat mengungkap faktor-faktor kunci dan profil mahasiswa yang berisiko mengalami keterlambatan. Model pohon keputusan yang terbentuk akan berfungsi sebagai pedoman identifikasi atau dasar pengambilan keputusan bagi Program Studi. Dengan pedoman ini, pihak program Studi dapat secara proaktif mengenali mahasiswa yang memiliki profil risiko tinggi berdasarkan catatan riwayat akademik mereka. Dengan demikian, penelitian ini memberikan landasan ilmiah untuk merumuskan kebijakan intervensi yang lebih efektif seperti bimbingan akademik yang terfokus guna meningkatkan angka kelulusan tepat waktu dan mendukung mutu program Studi.

TINJAUAN PUSTAKA

Penelitian terdahulu yang dilakukan oleh Orpa et.al., (2019) membahas tentang prediksi masa Studi mahasiswa menggunakan algoritma C4.5. Tujuan penelitian ini adalah untuk memprediksi apakah mahasiswa akan lulus cepat, tepat waktu, terlambat, atau drop out berdasarkan data riwayat akademik dari semester 1 hingga semester 4. Variabel yang digunakan

meliputi Indeks Prestasi Semester (IPS), Indeks Prestasi Kumulatif (IPK), dan jumlah SKS yang telah lulus, yang kemudian diolah menggunakan teknik data mining untuk membangun model klasifikasi. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu menghasilkan model prediksi dengan tingkat akurasi yang baik, sehingga efektif dalam memetakan pola masa Studi mahasiswa. Persamaan penelitian ini dengan penelitian yang sedang dilakukan terletak pada pemanfaatan atribut riwayat akademik (IPK dan SKS) sebagai dasar pembangunan model prediksi, sedangkan perbedaannya terdapat pada tujuan analisis.

Penelitian dilakukan oleh Suriani (2023) membahas penerapan algoritma C4.5 untuk menentukan masa Studi mahasiswa berdasarkan data akademik seperti IPK, SKS, semester, dan nilai tugas akhir. Penelitian ini menggunakan kerangka KDD (*Knowledge Discovery in Database*) sebagai acuan metodologis untuk memastikan proses data mining dilakukan secara terstruktur, dimulai dari *data selection*, *data cleansing*, *data transformation*, hingga tahap *data mining* menggunakan algoritma C4.5. Tahapan KDD yang digunakan meliputi pemilihan atribut penting, pembersihan data dari nilai yang hilang atau tidak konsisten, melakukan transformasi agar data sesuai dengan format pemodelan, hingga mengekstraksi pola-pola klasifikasi yang berguna melalui C4.5. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu menghasilkan model prediksi kategori masa Studi mahasiswa dengan akurasi yang lebih baik dibandingkan evaluasi manual pihak akademik, sehingga efektif digunakan untuk menilai ketepatan waktu kelulusan mahasiswa. Dengan demikian, penelitian ini tidak hanya menunjukkan efektivitas algoritma C4.5, tetapi juga menegaskan bahwa penggunaan tahapan KDD membantu menghasilkan model prediksi yang lebih akurat dan terarah.

1) Ketepatan Studi

Ketepatan waktu Studi, yang secara ideal ditargetkan selesai dalam 8 semester atau 4 tahun, merupakan indikator penting dalam menilai keberhasilan mahasiswa sekaligus efisiensi penyelenggaraan program Studi, termasuk di STIKOM Uyelindo Kupang. Dalam proses penyelesaian Studi, Studi atau skripsi memegang peranan penting sebagai karya ilmiah wajib dan tahap akhir yang menentukan kelulusan. Variasi waktu yang dibutuhkan mahasiswa untuk menyelesaikan skripsi menyebabkan fase ini sering menjadi titik krusial yang memengaruhi apakah mahasiswa dapat lulus tepat waktu atau mengalami keterlambatan. Namun, keterlambatan yang tampak pada saat pengerjaan skripsi sejatinya merupakan akumulasi dari perjalanan akademik mahasiswa di semester-semester sebelumnya, seperti capaian SKS, nilai

mata kuliah, IPK, masa cuti, maupun pengulangan mata kuliah. Akumulasi faktor-faktor tersebut berpotensi menghambat kelancaran proses penyelesaian Studidan pada akhirnya berdampak pada ketepatan waktu kelulusan mahasiswa serta penilaian kualitas dan akreditasi institusi.

2) Machine Learning dan Data Mining

Machine Learning adalah sebuah tipe dari Kecerdasan Buatan (AI) yang berfokus pada pengembangan program komputer yang dapat mengajarkan dirinya sendiri untuk belajar tanpa diprogram secara eksplisit (Suyanto, 2020). Proses kerja *Machine Learning* sangat mirip dengan Data Mining, di mana sistem mencari data untuk menemukan pola dan menggunakan pembelajaran tersebut untuk meningkatkan kemampuan prediksinya akan mendeteksi pola dalam data dan menyesuaikan tindakan yang sesuai. Untuk menguji algoritmanya, *Machine Learning* menggunakan dua set data terpisah: set pelatihan (*training set*), yang digunakan untuk mengajari program, dan set pengujian (*test set*), yang tidak diberikan variabel target, berfungsi untuk mengukur keakuratan prediksi dari model yang telah dilatih (Siregar dan Mubarak, 2023).

Data Mining adalah suatu disiplin ilmu yang mempelajari metode untuk mengekstraksi pengetahuan atau menemukan pola-pola menarik, tersembunyi, dan berpotensi bermanfaat dari kumpulan data yang sangat besar. Tujuannya adalah untuk mengubah data mentah menjadi informasi dan pengetahuan yang dapat ditindaklanjuti untuk pengambilan keputusan (Febrianto, et.al., 2025).

3) Text Classification

Klasifikasi (*Classification*) adalah tugas dalam pemrosesan teks yang bertujuan untuk memilih label kelas yang paling tepat untuk suatu input teks. Sebagai fondasi dalam banyak aplikasi, proses ini memerlukan proses *training*, terutama dalam konteks *supervised learning*, di mana *classifier* dilatih menggunakan training corpora berlabel yang benar untuk membangun model yang akurat. Hampir semua metode klasifikasi teks utama, seperti *Decision Tree*, *kNN*, *Naïve Bayes*, dan *SVM*, beroperasi dalam kerangka *supervised learning* ini, memungkinkan sistem untuk mengkategorikan dokumen secara otomatis. (Siregar dan Mubarak, 2023).

4) Metode K-Folds Cross Validation

K-fold cross validation merupakan salah satu dari teknik yang difungsikan untuk memilah data menjadi train data serta test data. Teknik ini banyak diterapkan peneliti karena didapati mengurangi bias yang didapatkan didalam pengambilan sebuah sampel. K yaitu besar angka pemilahan data yang digunakan untuk pembagian train dan test. metode 5-Fold Cross Validation sebagai teknik validasi untuk mengukur kinerja model klasifikasi. Metode ini membagi dataset menjadi lima bagian (fold) yang berukuran sama. Pada setiap iterasi, satu fold digunakan sebagai data uji (test set) dan empat fold lainnya digunakan sebagai data latih (training set). Dengan jumlah data sebanyak 300, setiap fold berisi 60 data uji dan 240 data latih. Pemilihan 5-Fold dilakukan karena ukuran dataset yang sedang, memberikan keseimbangan antara stabilitas evaluasi dan efisiensi komputasi. Tabel berikut adalah ilustrasi pembagian data dengan metode K-Fold cross validation (Dengen et al., 2020).

5) Algoritma C4.5

Algoritma C4.5 adalah algoritma pohon keputusan (*Decision Tree*) yang dikembangkan oleh John Ross Quinlan. Algoritma ini merupakan pengembangan dari pendahulunya, Algoritma ID3. Perbedaan utama antara keduanya adalah, jika ID3 menggunakan *Information Gain*, maka C4.5 menggunakan *Gain Ratio* untuk memilih atribut pemilah terbaik (*the best split attribute*) agar hasil penentuannya tidak bias. Algoritma C4.5 sangat serbaguna karena dapat menangani data yang bernilai kategorikal maupun numerik (*kontinu*). Khusus untuk data numerik, C4.5 akan membangun nilai-nilai batas (*thresholds*) untuk mendiskritisasi data menjadi sejumlah interval, sehingga nilai-nilai tersebut dapat diperlakukan sebagai nilai kategorikal. Selain itu, C4.5 juga memiliki kemampuan penting untuk menangani nilai atribut yang hilang ditandai dengan "?" dan mengatasi masalah overfitting pada pohon keputusan yang dihasilkan. C4.5 melakukannya dengan menerapkan pemangkasan pohon (*pruning*) pasca-induksi, yaitu dengan membuang cabang-cabang yang overfit dan menggantinya dengan simpul-simpul daun yang dianggap lebih umum dan representatif.

6) Evaluasi Model Klasifikasi

Klasifikasi adalah salah satu tugas utama dalam data mining untuk membangun model yang dapat memetakan suatu item data ke dalam kelas yang telah ditentukan "Tepat Waktu" dan

"Terlambat". Untuk mengetahui seberapa baik performa model, dilakukan proses evaluasi menggunakan Confusion Matrix (Hasibuan & Mahdiana, 2023). Confusion Matrix adalah tabel yang merangkum hasil prediksi dengan membandingkannya dengan kelas aktual. Tabel ini berisi empat komponen utama:

Tabel 2. 1 Confusion Matrix 1

Kelas Aktual	Kelas Hasil Prediksi		Jumlah
	Ya (+)	Tidak (-)	
Ya (+)	TP	FN	P
Tidak (-)	FP	TN	N
Jumlah	P'	N'	P + N

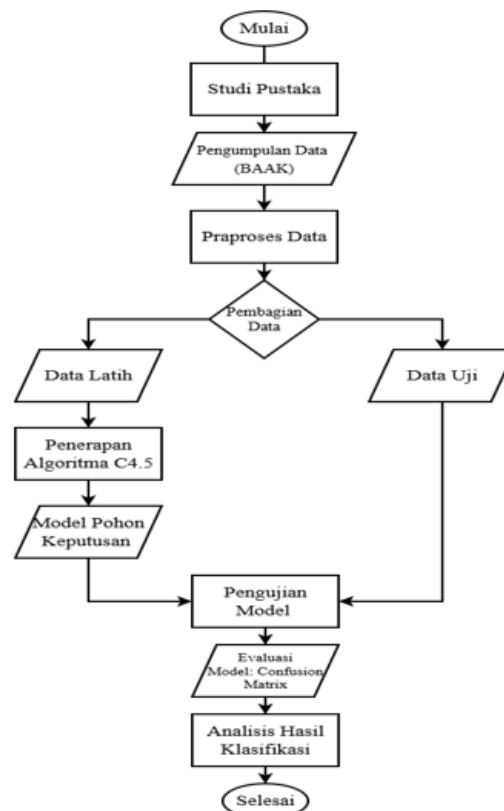
1. *True Positives* (TP): Jumlah data positif ("Tepat Waktu") yang diprediksi dengan benar sebagai positif.
2. *True Negatives* (TN): Jumlah data negatif ("Terlambat") yang diprediksi dengan benar sebagai negatif.
3. *False Positives* (FP): Jumlah data negatif ("Terlambat") yang salah diprediksi sebagai positif ("Tepat Waktu").
4. *False Negatives* (FN): Jumlah data positif ("Tepat Waktu") yang salah diprediksi sebagai negatif ("Terlambat").

7) Python

Python, pada tahun 1990 telah menjadi bahasa pemrograman dominan dalam komunitas ilmiah, terutama di bidang sains dan teknik. Kepopuleran ini didukung oleh keberadaan pustaka andal seperti SciPy dan NumPy yang mengoptimalkan operasi vektor dan matriks untuk keperluan Data Mining. Pada versi 3.5, penelitian ini menetapkan penggunaan Python versi 3.11.12 demi stabilitas dan pembaruan fitur. Guna mengimplementasikan prinsip Algoritma C4.5 pada klasifikasi data numerik dan kategorikal, penelitian ini mengintegrasikan tiga pustaka utama. Pertama, Scikit-learn (sklearn) sebagai mesin pemodelan inti yang menerapkan Decision Tree Classifier berbasis entropi sebagai adaptasi modern C4.5. Pandas memegang peran praproses data, khususnya untuk manipulasi data tabular dan konversi (*encoding*)

variabel kategorikal menjadi format numerik. NumPy sebagai landasan komputasi numerik yang efisien untuk memastikan kesiapan data sebelum diproses oleh model

METODE PENELITIAN



Berdasarkan alur flowchart penelitian pada Gambar 3.1, maka dapat dijelaskan sebagai berikut:

1. Studi Pustaka

Dilakukan dengan mencari literatur atau sumber-sumber pustaka pendukung. Proses ini bertujuan mengumpulkan landasan teori yang memadai dari jurnal, buku, dan penelitian sebelumnya yang relevan dengan topik data mining, klasifikasi ketepatan waktu Studi, dan penerapan algoritma C4.5

2. Pengumpulan Data Riwayat Akademik BAAK

Data utama penelitian ini adalah data sekunder berupa riwayat akademik mahasiswa STIKOM Uyelindo Kupang angkatan 2019-2020. Data ini diperoleh melalui teknik Studi

dokumentasi dari Biro Administrasi Akademik dan Kemahasiswaan (BAAK) dan menjadi dasar untuk analisis serta pembentukan model klasifikasi. Selain itu, pada tahap ini variabel penelitian dibedakan menjadi:

- a. Variabel Primer (digunakan sebagai input utama model C4.5):
 - i. Indeks Prestasi Kumulatif (IPK)
 - ii. Jumlah SKS Lulus
 - iii. Jumlah Mata Kuliah Mengulang
 - iv. Nilai Mata Kuliah Kunci
 - v. Indeks Prestasi Semester (IPS) Terakhir
 - vi. Lama Masa Studi (Semester)
 - vii. Jumlah SKS Tidak Lulus
- b. Variabel Sekunder (pendukung, tidak masuk langsung ke perhitungan C4.5)
 - i. Status Cuti Akademik
 - ii. Status KIP Kuliah

Pembagian ini dilakukan agar model lebih fokus pada variabel akademik yang benar-benar memengaruhi ketepatan waktu kelulusan.

3. Praproses Data

Pembersihan dan mempersiapkan data mentah riwayat akademik agar siap diproses oleh model. Langkah-langkah yang dilakukan antara lain:

- a. *Data Cleaning* dilakukan untuk membersihkan data dari elemen-elemen yang tidak konsisten atau tidak lengkap (*missing values*), seperti data IPK atau SKS yang kosong, serta memastikan tidak ada data duplikat.
- b. Seleksi Atribut dilakukan pemilihan atribut yang dianggap paling relevan dan berpengaruh terhadap klasifikasi kelulusan Mengulang, Cuti Akademik. Atribut yang tidak relevan akan dihapus.
- c. Transformasi Data mengubah data ke dalam format yang sesuai untuk algoritma C4.5. Ini termasuk mendefinisikan atribut label target secara jelas, yaitu membagi status kelulusan menjadi dua kelas "Tepat Waktu" dan "Terlambat".

4. Penerapan Algoritma C4.5

Perhitungan Gain Ratio data yang telah melalui tahap praproses dimasukkan ke dalam model. Algoritma C4.5 digunakan untuk melakukan proses klasifikasi dengan membangun pohon keputusan. Algoritma ini bekerja dengan menghitung nilai Gain Ratio tertinggi dari semua atribut untuk menentukan atribut mana yang menjadi *root node* dan simpul percabangan selanjutnya. Dataset riwayat akademik mahasiswa dibagi menjadi dua subset data yang terpisah, yaitu:

- a. Data Latih (*Training Set*): Bagian data yang digunakan sebagai input untuk proses pembelajaran algoritma C4.5 guna membangun model pohon keputusan.
- b. Data Uji (*Testing Set*): Bagian data yang dipisahkan dan tidak dilibatkan dalam proses pembentukan model. Data ini disimpan khusus untuk tahap pengujian guna mengukur keakuratan model dalam memprediksi data baru.

5. Model Pohon Keputusan (*Decision Tree*)

Hasil dari proses penerapan C4.5 adalah sebuah model klasifikasi yang berbentuk Pohon Keputusan. Model ini merepresentasikan sekumpulan aturan (*rules*) yang memetakan pola-pola dari riwayat akademik yang paling berpengaruh terhadap status kelulusan.

6. Pengujian Model (*K-Folds Cross Validation*)

Pengujian dilakukan untuk menilai kinerja dan validitas model pohon keputusan yang telah terbentuk. Metode *K-Folds Cross Validation* digunakan untuk membagi data secara adil. Evaluasi Model (*Confusion Matrix, Akurasi, Presisi, Recall*) Hasil prediksi dari seluruh proses pengujian dibandingkan dengan data aktual menggunakan Confusion Matrix. Dari matriks ini, akan dihitung nilai Akurasi, Presisi, dan Recall sebagai ukuran kuantitatif performa model

7. Hasil Klasifikasi dan Analisis

Melakukan analisis terhadap hasil evaluasi. Nilai Akurasi, Presisi, dan Recall akan ditinjau untuk menentukan efektivitas model C4.5. Rules yang dihasilkan dari Pohon Keputusan juga dianalisis untuk menemukan pola faktor penentu kelulusan mahasiswa.

HASIL DAN PEMBAHASAN

1. Implementasi Sistem

Aplikasi ini dibuat menggunakan bahasa pemrograman Python dengan tampilan antarmuka berbasis GUI (Graphical User Interface) menggunakan library Tkinter.

Sistem ini bertujuan membantu pengguna dalam memprediksi apakah mahasiswa dapat menyelesaikan studi tepat waktu atau tidak berdasarkan jumlah SKS yang telah lulus.

a. Implementasi Perangkat Lunak

Dalam proses pembangunan sistem digunakan beberapa perangkat lunak pendukung sebagai berikut:

Tabel 4. 1 Implementasi Perangkat Lunak

No	Perangkat lunak	Fungsi
1	Python	Bahasa pemrograman utama
2	<i>Tkinter</i>	Membuat tampilan antarmuka aplikasi
3	Pandas	Membaca dan mengolah data Excel
4	<i>Scikit-Learn</i>	Implementasi algoritma C4.5
5	Microsoft Excel	Penyimpanan dataset
6	Visual Studio Code	Editor pemrograman

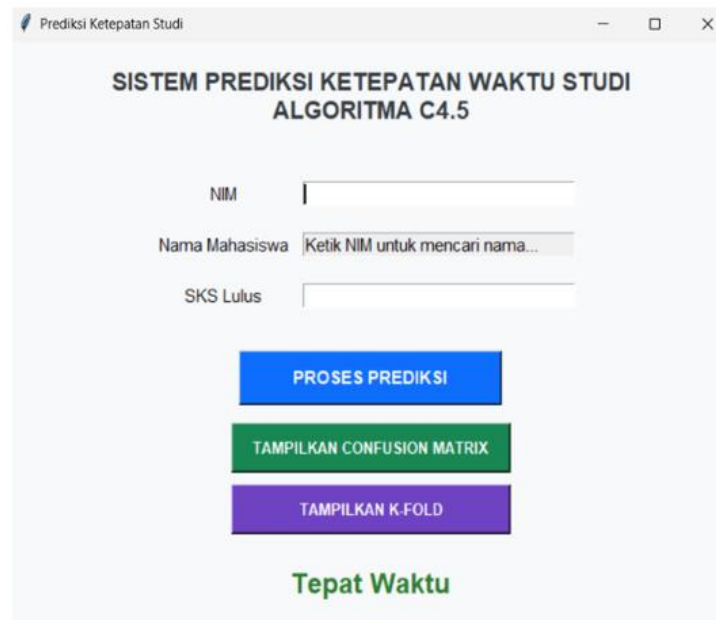
b. Implementasi Tampilan Sistem

Tampilan sistem dibuat menggunakan Tkinter agar aplikasi dapat dijalankan dalam bentuk desktop. Desain tampilan dibuat sederhana namun tetap menarik sehingga mudah digunakan oleh pengguna.

Pada halaman utama aplikasi terdapat beberapa bagian utama, yaitu:

1. Judul aplikasi.
2. Input Nomor Induk Mahasiswa (NIM).
3. Kolom nama mahasiswa otomatis.
4. Input jumlah SKS lulus.
5. Tombol proses prediksi.

6. Hasil prediksi status kelulusan.



Gambar 4. 1 Tampilan Sistem

2. Pengujian Sistem

Pengujian sistem dilakukan untuk memastikan bahwa seluruh fungsi aplikasi dapat berjalan dengan baik sesuai dengan tujuan penelitian. Pengujian dilakukan menggunakan metode *Black Box Testing*.

a. Pengujian *Black Box*

Pengujian *Black Box* dilakukan dengan menguji fungsi-fungsi utama sistem tanpa melihat kode program.

Tabel 4. 2 Hasil Pengujian Sistem

No	Skenario Pengujian	Hasil yang Diharapkan	Hasil
1	Input NIM valid	Nama mahasiswa tampil otomatis	Berhasil
2	Input NIM tidak valid	Sistem menampilkan pesan mahasiswa tidak ditemukan	Berhasil

3	Input SKS berupa angka	Sistem memproses prediksi	Berhasil
4	Input SKS berupa huruf	Sistem menampilkan pesan error	Berhasil
5	Tombol prediksi ditekan	Hasil klasifikasi tampil	Berhasil

b. Pengujian Hasil Prediksi

Pengujian hasil prediksi dilakukan menggunakan beberapa data mahasiswa untuk mengetahui kemampuan sistem dalam melakukan klasifikasi. Menggunakan *confusion matrix*, penelitian ini juga menggunakan metode *K-Fold Cross Validation* untuk mengukur kestabilan model dalam melakukan prediksi. Metode *K-Fold* bekerja dengan membagi data menjadi beberapa bagian (*fold*). Pada penelitian ini digunakan nilai $K = 5$, sehingga data dibagi menjadi 5 bagian. Setiap bagian akan digunakan secara bergantian sebagai data uji dan data lainnya digunakan sebagai data latih. oleh algoritma C4.5. Pada proses pengujian, data dibagi menjadi data latih dan data uji menggunakan metode `train_test_split` dengan pembagian 80% data latih dan 20% data uji. Selanjutnya sistem melakukan proses prediksi terhadap data uji dan menghasilkan nilai akurasi serta *confusion matrix*.

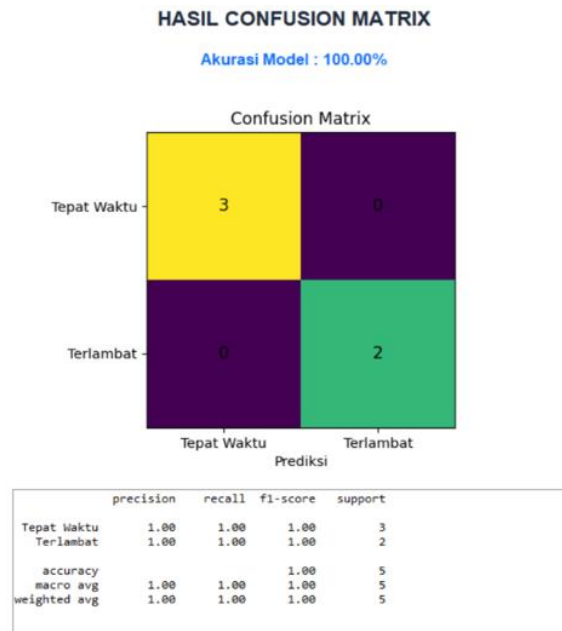
HASIL K-FOLD CROSS VALIDATION

```
HASIL AKURASI TIAP FOLD

Fold 1 : 100.00%
Fold 2 : 100.00%
Fold 3 : 100.00%
Fold 4 : 100.00%
Fold 5 : 100.00%

Rata-rata Akurasi : 100.00%
```

c. Pengujian Menggunakan Confusion Matrix



Pengujian Confusion Matrix dilakukan untuk mengetahui tingkat ketepatan hasil klasifikasi yang dihasilkan oleh sistem. Pengujian ini membandingkan antara data aktual dengan hasil prediksi yang dihasilkan

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa sistem prediksi ketepatan waktu studi mahasiswa menggunakan algoritma C4.5 berhasil dibangun menggunakan bahasa pemrograman Python dengan tampilan GUI berbasis *Tkinter*. Sistem mampu membaca data mahasiswa dari file Excel, menampilkan nama mahasiswa secara otomatis berdasarkan NIM, serta melakukan prediksi status penyelesaian studi berdasarkan jumlah SKS lulus mahasiswa. Hasil pengujian menunjukkan bahwa seluruh fungsi sistem dapat berjalan dengan baik sesuai dengan kebutuhan pengguna. Dengan adanya sistem ini, proses prediksi ketepatan waktu studi mahasiswa dapat dilakukan dengan lebih cepat dan membantu pihak program studi dalam melakukan analisis data mahasiswa.

Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan sistem selanjutnya, yaitu:

1. Menambahkan atribut lain seperti IPK, jumlah mata kuliah mengulang, dan tingkat kehadiran mahasiswa agar hasil prediksi lebih akurat.
2. Mengembangkan sistem berbasis web agar dapat diakses secara online.
3. Menambahkan fitur penyimpanan hasil prediksi ke dalam database.
4. Mengembangkan tampilan antarmuka agar lebih menarik dan interaktif.
5. Menambahkan fitur login pengguna untuk meningkatkan keamanan sistem.

Menggunakan database seperti MySQL atau SQLite agar pengelolaan data menjadi lebih baik.

DAFTAR PUSTAKA

- Rifki Febrianto, Kusuma Wijaya, Elkin Rilvani. 2025. Penerapan algoritma C4.5 untuk prediksi kelulusan mahasiswa berdasarkan data akademik. 3(8). [internet] [diakses 11 November 2025]. Tersedia pada:
<https://jurnal.mediaakademik.com/index.php/jma/article/view/2692>
- Anjani, A., dan Yamasari, Y. 2023. Klasifikasi Tingkat Stres Mahasiswa Menggunakan Metode Berbasis Tree. *Journal of Informatics and Computer Science*, 05, 83–89. [internet] [diakses 11 November 2025] Tersedia pada:
<https://ejournal.unesa.ac.id/index.php/jinacs/article/view/54859>
- Atala Panhareni, G., Gita Anugrah, I., Kunci, K., Penyelesaian Skripsi, W., dan Naïve Bayes, W. 2022. Menggunakan Metode Weighted Naïve Bayes (Studi Kasus: Program Studi Teknik Informatika Universitas Muhammadiyah Gresik). *INDEXIA: Informatic and Computational Intelligent Journal*, 4(1), 33–46. [internet] [diakses 11 November 2025]. Tersedia pada: <https://journal.umg.ac.id/index.php/indexia/article/view/3589>
- Dengen, C. N., Kusriani, K., dan Luthfi, E. T. 2020. Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *Sisfotenika*, 10(1), 1. [internet] [diakses 24 Oktober 2025]. Tersedia pada:
<https://sisfotenika.stmikpontianak.ac.id/index.php/ST/article/view/484>

- Faizah, T., dan Jananto, A. 2021. Perbandingan Algoritma C4 . 5 Dan Id3 Untuk Prediksi. 8(2), 980–990. [internet] [diakses 11 November 2025] Tersedia pada: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/593>
- Fitri, S., Nurjanah, N., dan Astuti, W. 2018. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa (Studi Kasus: Umtas). *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 9(1), 633–640. [internet] [diakses 11 November 2025]. Tersedia pada: <https://jurnal.umk.ac.id/index.php/simet/article/view/2002>
- Hasibuan, T. H., dan Mahdiana, D. 2023. Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta. *Skanika*, 6(1), 61–74. [internet] [diakses 23 Oktober 2025] Tersedia pada: <https://jom.fti.budiluhur.ac.id/SKANIKA/article/view/2976>
- Kamil, M., dan Cholil, W. 2020. Analisis Perbandingan Algoritma C4.5 dan Naive Bayes pada Lulusan Tepat Waktu Mahasiswa di Universitas Islam Negeri Raden Fatah Palembang. *Jurnal Informatika*, 7(2), 97–106. [internet] [diakses 23 Oktober 2025]. Tersedia pada: <https://ejournal.bsi.ac.id/ejournal/index.php/ji/article/view/7723>
- Mk, C. R. P., Sumarauw, S. J. A., dan Regar, V. E. 2024. Data Mining Model Klasifikasi Untuk Ketepatan Waktu Kelulusan Mahasiswa. *Dharmas Education Journal (DE_Journal)*, 5(1), 555–562. [internet] [diakses 24 Oktober 2025] Tersedia pada: https://ejournal.undhari.ac.id/index.php/de_journal/article/view/1349
- Paskalis, E., Orpa, K., dan Ripanti, E. F. 2019. Model Prediksi Awal Masa Studi Mahasiswa. 7(4), 272–278. [internet] [diakses 24 Oktober 2025] Tersedia pada: <https://jurnal.untan.ac.id/index.php/justin/article/view/33163>
- Putra, Y. 2022. Perbandingan Penerapan Algoritma C4.5, Naïve Bayes, dan K-Nearest Neighbor (K-NN) dalam Klasifikasi Kelulusan Mahasiswa. [internet] [diakses 11 November 2025]. Tersedia pada: <https://repository.unilak.ac.id/2841/>
- Salsabillah, T., Tanti, L., dan Info, A. 2025. Model prediksi kelulusan tepat waktu. 6(2), 680–693. [internet] [diakses 11 November 2025]. Tersedia pada: <https://jurnal.dharmawangsa.ac.id/index.php/djtechno>
- Suriani, U. 2023. Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5. *Journalcisa*, 3(2), 55–66. [internet] [diakses

11 November 2025]. Tersedia pada: <https://journal-computing.org/index.php/journal-cisa/article/view/393>

Suyanto. 2018. Machine Learning. Bandung (ID): Informatika.

Budiharto, W. 2018. Pemrograman Python untuk Ilmu Komputer dan Teknik. Yogyakarta (ID): Andi.