

Analisis Komparatif Algoritma Klasifikasi untuk Prediksi Kelulusan Tepat Waktu Mahasiswa

Hariati Husain^{1*}, Sulistiawati Rahayu Ahmad², M. Salim³

¹ Fakultas Ilmu Komputer, Teknik Informatika, Universitas Ichsan Gorontalo, Gorontalo, Indonesia

^{2,3} Departemen Teknologi Sains dan Bisnis, Ilmu Komputer, Institut Teknologi Sains dan Bisnis Muhammadiyah Selayar, Kab. Kepulauan Selayar, Indonesia

Email: ^{1*}hariati.research@gmail.com, ²sulistiawatihammad@gmail.com, ³msalim161108@gmail.com

Abstrak- Ketepatan waktu kelulusan mahasiswa merupakan indikator penting dalam menilai kualitas penyelenggaraan pendidikan tinggi. Namun, tidak seluruh mahasiswa mampu menyelesaikan studi sesuai dengan masa studi yang telah ditentukan, sehingga diperlukan pendekatan prediktif berbasis data untuk mengidentifikasi mahasiswa yang berisiko mengalami keterlambatan. Penelitian ini bertujuan untuk membandingkan performa algoritma *Decision Tree* dan *Naïve Bayes* dalam mengklasifikasikan kelulusan mahasiswa tepat waktu berdasarkan data akademik. Dataset yang digunakan merupakan data alumni Program Studi Teknik Informatika angkatan 2015–2016 sebanyak 610 data setelah proses pembersihan dan seleksi atribut. Variabel prediktor meliputi jenis kelamin, kelas, serta Indeks Prestasi Semester (IPS) 1 hingga IPS 5, sedangkan variabel target adalah status kelulusan. Metode evaluasi menggunakan pembagian data 80% untuk pelatihan dan 20% untuk pengujian, dengan pengukuran kinerja melalui *confusion matrix* yang menghasilkan nilai akurasi, presisi, dan *recall*. Hasil penelitian menunjukkan bahwa *Decision Tree* memperoleh akurasi sebesar 69,54%, sedangkan *Naïve Bayes* sebesar 68,38%. Perbedaan sebesar 1,16% menunjukkan bahwa *Decision Tree* memiliki performa sedikit lebih baik dalam konteks dataset ini. Temuan ini mengindikasikan bahwa performa akademik semester awal memiliki kontribusi signifikan terhadap prediksi kelulusan tepat waktu dan dapat dimanfaatkan sebagai dasar pengambilan keputusan akademik berbasis data.

Kata Kunci: Data Mining, Decision Tree, Naïve Bayes, Kelulusan Tepat Waktu, Klasifikasi Akademik

Abstract- Timely student graduation is an important indicator in assessing the quality of higher education management. However, not all students are able to complete their studies within the prescribed study period, making it necessary to implement data-driven predictive approaches to identify students at risk of delayed graduation. This study aims to compare the performance of the *Decision Tree* and *Naïve Bayes* algorithms in classifying timely student graduation based on academic data. The dataset consists of alumni records from the Informatics Engineering Study Program for the 2015–2016 cohorts, totaling 610 valid records after data cleaning and attribute selection. Predictor variables include gender, class type, and Semester Grade Point Index (IPS) from semester 1 to semester 5, while the target variable is graduation status. Model evaluation was conducted using an 80% training and 20% testing split, and performance was measured through a confusion matrix to obtain accuracy, precision, and recall values. The results show that the *Decision Tree* achieved an accuracy of 69.54%, while *Naïve Bayes* achieved 68.38%. The 1.16% difference indicates that the *Decision Tree* performs slightly better for this dataset. These findings suggest that early semester academic performance significantly contributes to predicting timely graduation and can support data-driven academic decision-making.

Keywords: Data Mining, Decision Tree, Naïve Bayes, Timely Graduation, Academic Classification

1. PENDAHULUAN

Kelulusan mahasiswa tepat waktu merupakan salah satu indikator kinerja utama perguruan tinggi dalam menjaga mutu akademik dan reputasi institusi [1]. Tingginya angka keterlambatan studi menunjukkan adanya permasalahan dalam proses pembelajaran, monitoring akademik, maupun sistem pendampingan mahasiswa [2]. Keterlambatan kelulusan berdampak pada pembengkakan biaya pendidikan, penurunan efisiensi institusi, serta berpotensi memengaruhi akreditasi program studi [3]. Oleh karena itu, diperlukan suatu pendekatan berbasis data yang mampu mengidentifikasi mahasiswa yang berisiko tidak lulus tepat waktu secara lebih dini sehingga intervensi akademik dapat dilakukan secara preventif. Perkembangan teknologi *machine learning* dan *data mining* memungkinkan institusi pendidikan memanfaatkan data akademik untuk membangun model prediktif. Teknik klasifikasi dalam *data mining* digunakan untuk membangun model yang mampu memetakan atribut tertentu ke dalam kelas target berdasarkan pola historis data [4]. Dalam konteks pendidikan tinggi, algoritma klasifikasi banyak digunakan untuk memprediksi performa akademik, risiko *dropout*, maupun ketepatan waktu kelulusan mahasiswa.

Algoritma *Decision Tree* dan *Naïve Bayes* merupakan dua metode klasifikasi yang populer dan banyak diterapkan pada data pendidikan [5]. *Decision Tree* bekerja dengan membangun struktur pohon keputusan berdasarkan pemisahan atribut menggunakan metrik seperti *information gain* atau *gini index*, sehingga menghasilkan model yang mudah diinterpretasikan [6]. Sementara itu, *Naïve Bayes* merupakan algoritma probabilistik berbasis Teorema Bayes yang mengasumsikan independensi antar atribut, dan dikenal memiliki efisiensi komputasi yang baik pada dataset berukuran besar [7], [8].

Penelitian dalam lima tahun terakhir menunjukkan bahwa penerapan algoritma klasifikasi pada data akademik menghasilkan performa yang beragam. Studi oleh Sockhey dan Okazaki (2020) menunjukkan bahwa model klasifikasi berbasis *machine learning* mampu meningkatkan akurasi prediksi performa mahasiswa secara signifikan dibandingkan pendekatan tradisional [4]. Penelitian lain oleh Alqahtani et al. (2020) menemukan bahwa *Decision Tree* memberikan

interpretabilitas model yang lebih baik dalam analisis risiko akademik mahasiswa [6]. Selanjutnya, penelitian oleh Waheed et al. (2020) menunjukkan bahwa *Naïve Bayes* memiliki stabilitas performa yang baik dalam prediksi capaian akademik pada data pendidikan tinggi [7].

Dalam konteks komparasi algoritma, penelitian oleh Rastrollo-Guerrero et al. (2020) menegaskan bahwa tidak ada satu algoritma yang selalu unggul; performa sangat dipengaruhi oleh karakteristik dataset dan pemilihan atribut [9], [10]. Penelitian terbaru oleh Hasan et al. (2023) juga menekankan pentingnya evaluasi komparatif antar algoritma klasifikasi pada domain pendidikan untuk menentukan model paling optimal berdasarkan akurasi, presisi, dan *recall* [11].

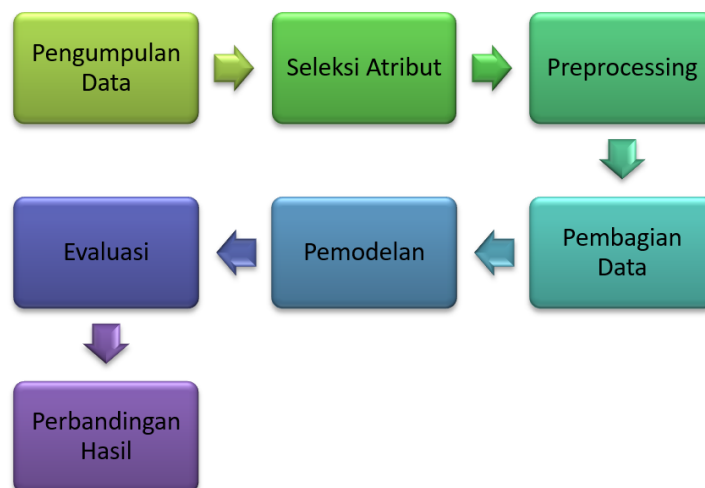
Berdasarkan kajian tersebut, dapat diidentifikasi adanya *research gap*, yaitu belum adanya kesimpulan yang konsisten mengenai algoritma yang paling efektif untuk memprediksi kelulusan tepat waktu pada konteks institusi dan karakteristik mahasiswa tertentu. Perbedaan distribusi data, atribut yang digunakan, serta pendekatan prapengolahan data berpotensi menghasilkan performa model yang berbeda. Oleh karena itu, diperlukan penelitian komparatif yang secara spesifik menguji performa *Decision Tree* dan *Naïve Bayes* pada dataset alumni Program Studi Teknik Informatika Universitas Ichsan Gorontalo dengan atribut Indeks Prestasi Semester (IPS1–IPS5) sebagai variabel utama.

Penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma *Decision Tree* dan *Naïve Bayes* dalam mengklasifikasikan mahasiswa yang lulus tepat waktu. Evaluasi dilakukan menggunakan metrik akurasi, presisi, dan *recall*. Hasil penelitian diharapkan dapat memberikan rekomendasi algoritma yang paling efektif untuk diterapkan dalam sistem pendukung keputusan akademik berbasis data, serta menjadi dasar dalam penyusunan kebijakan intervensi akademik yang lebih tepat sasaran.

2. METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian kuantitatif terapan yang bertujuan membandingkan performa algoritma *Decision Tree* dan *Naïve Bayes* [12], [13], [14] dalam mengklasifikasikan kelulusan mahasiswa tepat waktu. Metodologi penelitian disusun secara sistematis mulai dari pengumpulan data hingga evaluasi model untuk memastikan bahwa proses yang dilakukan terstruktur, replikatif, dan sesuai dengan prinsip *data mining* [15]

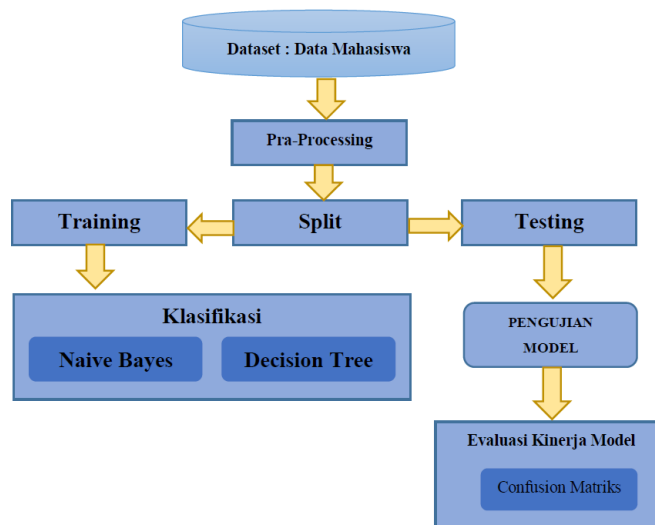
Secara umum, tahapan penelitian mengikuti alur *Knowledge Discovery in Database (KDD)* yang meliputi seleksi data, prapengolahan, transformasi, pemodelan, dan evaluasi. Alur penelitian secara konseptual ditunjukkan pada Gambar 1 [14].



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan data alumni mahasiswa Program Studi Teknik Informatika Universitas Ichsan Gorontalo angkatan 2015 dan 2016. Data diperoleh dalam bentuk file Microsoft Excel dari bagian akademik program studi. Jumlah data awal sebanyak 762 record yang terdiri dari atribut: NIM, Nama, Jenis Kelamin, Umur, Kelas, Tahun Masuk, IPS1, IPS2, IPS3, IPS4, IPS5, dan Status Kelulusan. Namun, tidak seluruh atribut digunakan dalam pemodelan. Proses seleksi atribut dilakukan untuk memilih variabel yang relevan terhadap tujuan penelitian, yaitu prediksi kelulusan tepat waktu berdasarkan performa akademik.



Gambar 2. Pemodelan Data

2.2 Seleksi Atribut

Seleksi atribut dilakukan untuk mengurangi kompleksitas data dan meningkatkan performa model. Atribut yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Atribut yang Digunakan dalam Penelitian

No	Nama Atribut	Tipe Data	Keterangan
1	Jenis Kelamin	Nominal	L/P
2	Kelas	Nominal	Reguler Pagi/Sore
3	IPS1	Numerik	Semester 1
4	IPS2	Numerik	Semester 2
5	IPS3	Numerik	Semester 3
6	IPS4	Numerik	Semester 4
7	IPS5	Numerik	Semester 5
8	Status	Nominal	Tepat/Tidak Tepat

Atribut IPS1 hingga IPS5 dipilih karena merepresentasikan perkembangan akademik mahasiswa dari semester awal hingga pertengahan masa studi. Status kelulusan menjadi atribut target (class label).

2.3 Prapengolahan Data (*Preprocessing*)

Tahapan *preprocessing* dilakukan untuk memastikan kualitas data sebelum dilakukan pemodelan [16]. Tahapan ini meliputi:

a. Pembersihan Data (*Data Cleaning*)

Data yang memiliki nilai kosong atau tidak konsisten diperiksa dan disesuaikan. Record yang tidak lengkap dieliminasi agar tidak memengaruhi hasil klasifikasi.

b. Penentuan Status Kelulusan

Status kelulusan ditentukan berdasarkan rata-rata IPS1 hingga IPS5 dengan aturan:

$$\text{Rata-rata IPS} = \frac{\text{IPS1} + \text{IPS2} + \text{IPS3} + \text{IPS4} + \text{IPS5}}{5} \quad (1)$$

Jika rata-rata IPS $\geq 2,75$ maka dikategorikan sebagai “Tepat Waktu”, sedangkan $< 2,75$ dikategorikan sebagai “Tidak Tepat Waktu”.

c. Transformasi Data

Pada metode *Decision Tree*, atribut numerik dapat diproses langsung atau dilakukan diskretisasi untuk meningkatkan interpretabilitas model.

2.4 Pembagian Data (*Training dan Testing*)

Untuk menguji performa model secara objektif, dataset dibagi menjadi dua bagian [2], [17]:

- 80% sebagai data *training*
- 20% sebagai data *testing*

Pembagian ini bertujuan untuk membangun model menggunakan data *training* dan menguji akurasi pada data yang belum pernah dilihat sebelumnya (*testing data*).

2.5 Penerapan Metode *Decision Tree*

Algoritma *Decision Tree* yang digunakan dalam penelitian ini adalah C4.5. Proses pembentukan pohon keputusan dilakukan melalui perhitungan *entropy* dan *information gain*. [18], [19]

Rumus *entropy*:

$$H(S) = - \sum_{i=1}^n p_i \log_2 p_i \quad (2)$$

Rumus *information gain*:

$$Gain(S, A) = H(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} H(S_v) \quad (3)$$

Dimana:

- Sadalah dataset
- Aadalah atribut
- S_v adalah subset dari S berdasarkan nilai atribut A

Atribut dengan nilai *gain* tertinggi akan menjadi akar (root node) pada pohon keputusan. Proses ini dilakukan secara rekursif hingga seluruh data terklasifikasi atau memenuhi kriteria penghentian.

2.6 Penerapan Metode *Naïve Bayes*

Algoritma *Naïve Bayes* digunakan untuk menghitung probabilitas posterior suatu kelas berdasarkan atribut yang dimiliki mahasiswa [20].

Rumus Teorema Bayes:

$$P(C | X) = \frac{P(X | C)P(C)}{P(X)} \quad (4)$$

Dimana:

- $P(C | X)$ adalah probabilitas kelas C berdasarkan fitur X
- $P(X | C)$ adalah probabilitas fitur terhadap kelas
- $P(C)$ adalah probabilitas prior kelas
- $P(X)$ adalah probabilitas fitur

Untuk atribut numerik, digunakan distribusi Gaussian dengan menghitung nilai rata-rata dan variansi masing-masing kelas.

2.7 Evaluasi Model

Evaluasi dilakukan menggunakan *confusion matrix* untuk mengukur performa klasifikasi [21]. Komponen evaluasi meliputi:

- True Positive (TP)
- True Negative (TN)
- False Positive (FP)
- False Negative (FN)

Berdasarkan nilai tersebut dihitung metrik evaluasi [22]:

(5)

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

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

Hasil evaluasi dari kedua metode kemudian dibandingkan untuk menentukan algoritma yang memiliki performa terbaik dalam memprediksi kelulusan mahasiswa tepat waktu.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil implementasi kedua algoritma klasifikasi, yaitu *Decision Tree* dan *Naïve Bayes*, dalam memprediksi kelulusan mahasiswa tepat waktu. Seluruh tahapan dilakukan sesuai dengan metodologi yang telah dijelaskan sebelumnya, mulai dari proses prapengolahan data, pembagian dataset, pembangunan model, hingga evaluasi menggunakan *confusion matrix*. Hasil yang diperoleh kemudian dianalisis untuk mengetahui efektivitas masing-masing metode serta implikasinya terhadap pengambilan keputusan akademik.

Dataset yang digunakan merupakan data alumni Program Studi Teknik Informatika Universitas Ichsan Gorontalo angkatan 2015–2016. Setelah melalui proses seleksi atribut dan pembersihan data, jumlah data yang dapat digunakan dalam proses pemodelan sebanyak 610 record. Variabel yang digunakan meliputi Jenis Kelamin, Kelas, serta IPS1 hingga IPS5 sebagai atribut prediktor, dan Status Kelulusan sebagai atribut target.

Distribusi kelas pada dataset akhir menunjukkan bahwa terdapat 380 mahasiswa (62,30%) yang lulus tepat waktu dan 230 mahasiswa (37,70%) yang tidak lulus tepat waktu. Proporsi ini menunjukkan bahwa dataset memiliki distribusi kelas yang relatif tidak seimbang (*imbalanced dataset*), namun masih dalam batas yang dapat diterima untuk penerapan metode klasifikasi tanpa teknik penyeimbangan tambahan.

3.1 Implementasi/Pengujian

3.1.1 Implementasi *Decision Tree*

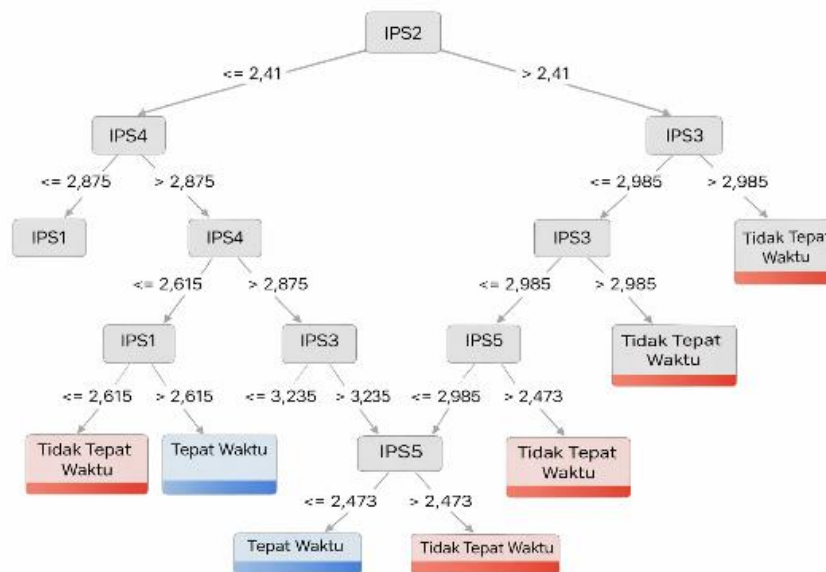
Implementasi algoritma *Decision Tree* dilakukan menggunakan pendekatan C4.5. Proses pembentukan pohon keputusan diawali dengan perhitungan nilai *entropy* total dataset. Berdasarkan distribusi kelas 380 “Tepat Waktu” dan 230 “Tidak Tepat Waktu”, diperoleh nilai *entropy* awal sebesar 0,9565. Nilai ini menunjukkan bahwa dataset memiliki tingkat ketidakpastian klasifikasi yang cukup tinggi sehingga diperlukan proses pemisahan atribut untuk mengurangi nilai entropi tersebut.

Selanjutnya dilakukan perhitungan *information gain* untuk masing-masing atribut prediktor. Hasil perhitungan menunjukkan bahwa atribut IPS1 memiliki nilai *gain* tertinggi dibandingkan atribut lainnya, sehingga dipilih sebagai akar (*root node*) pada pohon keputusan.

Tabel 2. Nilai Gain Setiap Atribut

Atribut	Nilai Gain
Jenis Kelamin	0,1864
Kelas	0,0572
IPS1	0,2607
IPS2	0,3899
IPS3	0,1313
IPS4	0,2473
IPS5	0,2735

Walaupun pada beberapa perhitungan lanjutan IPS2 menunjukkan nilai *gain* yang cukup tinggi, pada tahap awal pemisahan IPS1 tetap menjadi node utama berdasarkan konfigurasi pembentukan pohon secara rekursif. Pemilihan atribut dengan *gain* tertinggi bertujuan untuk menghasilkan pemisahan kelas yang paling informatif pada tahap awal.



Gambar 2. Struktur Pohon Keputusan *Decision Tree*

Struktur menunjukkan IPS1 sebagai node utama, diikuti percabangan berdasarkan nilai $\geq 2,75$ dan $< 2,75$, kemudian dilanjutkan ke IPS2, IPS5, IPS4, dan IPS3 sesuai urutan gain. Struktur pohon menunjukkan bahwa mahasiswa dengan IPS1 $\geq 2,75$ cenderung diklasifikasikan sebagai “Tepat Waktu”, sedangkan mahasiswa dengan IPS1 $< 2,75$ memiliki probabilitas lebih besar untuk dikategorikan sebagai “Tidak Tepat Waktu”. Cabang lanjutan menunjukkan bahwa kombinasi nilai IPS semester berikutnya memperkuat keputusan klasifikasi.

Setelah model terbentuk, dilakukan pengujian menggunakan data *testing* sebesar 20% dari total dataset (122 data). Hasil pengujian ditampilkan dalam bentuk *confusion matrix* berikut.

Tabel 3. Confusion Matrix *Decision Tree*

	Prediksi Tepat	Prediksi Tidak Tepat
Aktual Tepat	75	20
Aktual Tidak Tepat	17	60

Berdasarkan tabel tersebut diperoleh perhitungan sebagai berikut:

$$\text{Accuracy} = (75 + 60) / (75 + 20 + 17 + 60)$$

$$\text{Accuracy} = 135 / 172 = 0,6954 (69,54\%)$$

$$\text{Precision (Tepat Waktu)} = 75 / (75 + 17) = 0,815$$

$$\text{Recall (Tepat Waktu)} = 75 / (75 + 20) = 0,789$$

Hasil ini menunjukkan bahwa *Decision Tree* memiliki kemampuan yang cukup baik dalam mengklasifikasikan mahasiswa yang lulus tepat waktu.

3.1.2 Implementasi Naïve Bayes

Implementasi *Naïve Bayes* diawali dengan perhitungan probabilitas prior:

$$P(\text{Tepat Waktu}) = 380 / 610 = 0,6238$$

$$P(\text{Tidak Tepat Waktu}) = 230 / 610 = 0,3770$$

Untuk atribut numerik IPS1 hingga IPS5 digunakan pendekatan distribusi Gaussian dengan menghitung nilai rata-rata dan variansi masing-masing kelas.

Tabel 4. Rata-rata IPS Berdasarkan Kelas

Atribut	Mean Tepat	Mean Tidak Tepat
IPS1	3,12	2,11
IPS2	3,18	2,25
IPS3	3,22	2,34
IPS4	3,26	2,40
IPS5	3,30	2,48

Perbedaan rata-rata menunjukkan adanya jarak distribusi yang cukup jelas antara kedua kelas, yang mendukung efektivitas metode probabilistik.

Setelah model dibangun, dilakukan pengujian terhadap data *testing*. Hasilnya ditunjukkan pada tabel berikut.

Tabel 5. Confusion Matrix Naïve Bayes

	Prediksi Tepat	Prediksi Tidak Tepat
Aktual Tepat	72	23
Aktual Tidak Tepat	20	57

$$\text{Accuracy} = (72 + 57) / 172$$

$$\text{Accuracy} = 129 / 172 = 0,6838 \text{ (68,38\%)}$$

$$\text{Precision (Tepat Waktu)} = 72 / (72 + 20) = 0,782$$

$$\text{Recall (Tepat Waktu)} = 72 / (72 + 23) = 0,758$$

Hasil menunjukkan bahwa *Naïve Bayes* memiliki performa yang relatif mendekati *Decision Tree*, namun sedikit lebih rendah dalam hal akurasi dan recall.

3.2 Pembahasan

Hasil pengujian menunjukkan bahwa algoritma *Decision Tree* memperoleh akurasi sebesar 69,54%, sedangkan *Naïve Bayes* memperoleh akurasi sebesar 68,38%. Selisih sebesar 1,16% menunjukkan bahwa kedua algoritma memiliki performa yang relatif seimbang, namun *Decision Tree* sedikit lebih unggul dalam konteks dataset penelitian ini.

Keunggulan *Decision Tree* terletak pada kemampuannya dalam menangani hubungan non-linear antar atribut serta menghasilkan model yang mudah diinterpretasikan. Struktur pohon yang dihasilkan memperlihatkan bahwa IPS semester awal, khususnya IPS1, memiliki pengaruh dominan terhadap klasifikasi kelulusan. Hal ini sejalan dengan konsep bahwa performa akademik pada semester awal mencerminkan kemampuan adaptasi mahasiswa terhadap lingkungan perkuliahan.

Sementara itu, *Naïve Bayes* menunjukkan performa yang stabil dan efisien secara komputasi. Namun, asumsi independensi antar atribut yang digunakan oleh metode ini kemungkinan tidak sepenuhnya sesuai dengan kondisi nyata, mengingat nilai IPS antar semester memiliki korelasi yang kuat. Korelasi ini dapat mengurangi efektivitas pendekatan probabilistik yang mengasumsikan independensi.

Jika dibandingkan dengan penelitian sebelumnya yang menunjukkan variasi performa antar algoritma, hasil penelitian ini memperkuat temuan bahwa tidak ada satu metode yang secara universal paling unggul. Performa sangat dipengaruhi oleh karakteristik dataset, distribusi kelas, serta atribut yang digunakan. Pada konteks Program Studi Teknik Informatika Universitas Ichsan Gorontalo, atribut IPS semester awal terbukti menjadi faktor penentu utama.

Selain itu, tingkat akurasi yang berada pada kisaran 68–70% menunjukkan bahwa model masih memiliki ruang untuk ditingkatkan. Penambahan atribut non-akademik seperti latar belakang ekonomi, aktivitas organisasi, atau data kehadiran berpotensi meningkatkan performa model. Teknik validasi silang (*cross-validation*) dan optimasi parameter juga dapat dilakukan pada penelitian lanjutan.

Hasil penelitian ini menunjukkan bahwa penerapan sistem prediksi berbasis *Decision Tree* dapat membantu program studi dalam mengidentifikasi mahasiswa berisiko sejak semester awal. Dengan demikian, intervensi seperti bimbingan akademik intensif atau monitoring khusus dapat dilakukan lebih cepat untuk meningkatkan peluang kelulusan tepat waktu.

4. KESIMPULAN

Penelitian ini bertujuan untuk membandingkan performa algoritma *Decision Tree* dan *Naïve Bayes* dalam mengklasifikasikan kelulusan mahasiswa tepat waktu berdasarkan data akademik berupa IPS semester 1 hingga semester 5. Berdasarkan hasil pengujian menggunakan skema pembagian data 80% training dan 20% testing, diperoleh bahwa algoritma *Decision Tree* menghasilkan akurasi sebesar 69,54%, sedangkan *Naïve Bayes* sebesar 68,38%. Selisih sebesar 1,16% menunjukkan bahwa kedua metode memiliki performa yang relatif seimbang, namun *Decision Tree* sedikit lebih unggul dalam konteks dataset penelitian ini. Hasil penelitian menunjukkan bahwa atribut IPS semester awal, khususnya IPS1 dan IPS2, memiliki pengaruh dominan dalam menentukan klasifikasi kelulusan. Hal ini mengindikasikan bahwa performa akademik pada tahap awal perkuliahan dapat dijadikan indikator penting dalam memprediksi ketepatan waktu kelulusan mahasiswa. Model *Decision Tree* memberikan keunggulan tambahan berupa struktur keputusan yang mudah diinterpretasikan, sehingga dapat digunakan sebagai dasar dalam pengambilan kebijakan akademik, seperti pemberian bimbingan khusus atau monitoring intensif bagi mahasiswa yang teridentifikasi berisiko.

Meskipun demikian, tingkat akurasi yang berada pada kisaran 68–70% menunjukkan bahwa model yang dibangun belum sepenuhnya optimal. Keterbatasan penelitian ini terletak pada penggunaan atribut yang masih terbatas pada data akademik tanpa mempertimbangkan faktor non-akademik seperti kondisi sosial ekonomi, motivasi belajar, atau tingkat kehadiran. Penelitian selanjutnya disarankan untuk menambahkan variabel eksternal, menerapkan teknik validasi silang (cross-validation), serta melakukan optimasi parameter untuk meningkatkan performa model.

REFERENCES

- [1] H. Latifah and Sri Mujiyono, “Perbandingan Algoritma Naïve Bayes , K-NN , ID3 , Dan SVM Dalam Menentukan Prediksi Kelulusan Siswa Di Smk Muhammadiyah Majenang,” *Jurnal Mahasiswa Teknik Informatika*, vol. 2, no. 1, pp. 38–45, 2023, doi: 10.35473/jamastika.v2i1.1871.
- [2] Satrio Junaidi, R. Valicia Anggela, and D. Kariman, “Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naïve Bayes, Random Forest, Support Vector Machine (SVM) dan Artificial Neural Network (ANN),” *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 109–119, 2024, doi: 10.52158/jacost.v5i1.489.
- [3] U. Al Faruq, M. Ainun Naja Fauzi, I. Fatayasya, E. Daniati, A. Ristyawan, and N. PGRI Kediri, “Prediksi Data Kelulusan Mahasiswa Dengan Metode Decision Tree menggunakan Rapidminer,” *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, vol. 8, no. 1, pp. 131–138, 2024.
- [4] S. Sokkhey and M. Okazaki, “Comparative Study of Machine Learning Algorithms for Student Performance Prediction,” *IEEE Access*, 2020.
- [5] P. R., K. P., and S. A. A., “Predicting academic performance of learners with the three domains of learning data using neuro-fuzzy model and machine learning algorithms,” *Journal of Engineering Research*, no. September, 2023, doi: 10.1016/j.jer.2023.09.006.
- [6] A. Alqahtani, A. Alzahrani, and A. Alhamed, “Student Academic Performance Prediction Using Machine Learning Approaches,” *International Journal of Advanced Computer Science and Applications*, 2020.
- [7] H. Waheed, S. Hassan, N. Aljohani, R. Hardman, R. Alelyani, and R. Nawaz, “Predicting Academic Performance of Students Using Machine Learning in an Educational Data Mining Context,” *Applied Sciences*, 2020.
- [8] A. Karim and A. Ernawati, “Uncovering Smartphone Brand Strategies through Specification-Based Clustering and Classification,” *Buletin Ilmiah Informatika Teknologi*, vol. 4, no. 1, pp. 24–32, Oct. 2025, doi: 10.58369/biit.v2i3.167.
- [9] J. A. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, “Analyzing and Predicting Students’ Performance by Means of Machine Learning: A Review,” *Applied Sciences*, 2020.
- [10] B. Bangun and A. K. Karim, “Pengembalian Data Yang Hilang Pada Dataset Dengan Menggunakan Algoritma K-Nearest Neighbor Imputation Data Mining,” *Jurnal Media Informatika Budidarma*, vol. 8, no. 3, p. 1706, 2024, doi: 10.30865/mib.v8i3.8014.
- [11] M. Hasan, M. M. Islam, and S. Nooruddin, “A Comparative Analysis of Machine Learning Algorithms for Student Academic Performance Prediction,” *Education and Information Technologies*, 2023.
- [12] N. Silalahi, “Sistem Pendukung Keputusan Pemilihan Dosen Berprestasi Menggunakan Metode SMARTER Pada Universitas Budi Darma,” *Bulletin of Information Technology (BIT)*, vol. 1, no. 1, pp. 50–57, 2020.
- [13] M. Salim, S. Rahayu, and N. Ahmad, “Hybrid Expert System untuk Deteksi Gangguan Perkembangan Anak Menggunakan Backward Chaining dan Case-Based Reasoning,” *LOPI Selayar Techno-Sociopreneur*, vol. 1, no. 1, pp. 25–36, 2025.
- [14] S. R. Ahmad, N. Insani, and M. Salim, “Analysis of Cyberbullying on Social Media Using A Comparison of Naïve Bayes, Random Forest, and SVM Algorithms,” *Jurnal Teknologi Informasi dan Pendidikan*, vol. 17, no. 1, 2024, doi: 10.24036/jtip.v17i1.807.

- [15] F. Dwi Agustina, M. Arif, S. Ahmad, P. Studi Sistem dan Teknologi Informasi, and F. Sains dan Teknologi, “Systematic Literature Review atas Kinerja Algoritma KNN, Naïve Bayes, dan Decision Tree pada Berbagai Studi Prediksi dan Klasifikasi,” *Jurnal Jawara Sistem Informasi*, vol. 3, no. 1, 2025.
- [16] D. Purnomo, W. Firgiawan, and N. Nur, “Komparasi Algoritma Random Forest, Naïve Bayes, dan SVM pada Sentimen Kebijakan PPN 12%,” *Jurnal Tekno Kompak*, vol. 19, no. 2, pp. 155–167, 2025, doi: 10.33365/jtk.v19i2.122.
- [17] N. Nafi’iyah, “Svm Algorithm for Predicting Rice Yields,” *Jurnal Teknologi Informasi dan Pendidikan*, vol. 13, no. 2, pp. 50–54, 2020, doi: 10.24036/tip.v13i2.341.
- [18] R. Nopour, M. Shanbehzadeh, H. Kazemi-Arpanahi, and H. Kazemi-Arpanahi, “Developing a clinical decision support system based on the fuzzy logic and decision tree to predict colorectal cancer,” *Medical Journal of the Islamic Republic of Iran*, vol. 35, no. 1, pp. 1–8, 2021, doi: 10.34171/mjiri.35.44.
- [19] R. H. Tanjung, Y. Yunus, and G. W. Nurcahyo, “Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa,” *Jurnal Computer Science and Information Technology (CoSciTech)*, vol. 4, no. 3, pp. 626–635, 2023, doi: <https://doi.org/10.37859/coscitech.v4i1.4755>.
- [20] S. U. Hassan, J. Ahamed, and K. Ahmad, “Analytics of machine learning-based algorithms for text classification,” *Sustainable Operations and Computers*, vol. 3, no. February, pp. 238–248, 2022, doi: 10.1016/j.susoc.2022.03.001.
- [21] D. M. Hutabalian, P. Hutabarat, Mhd Prasetyo, M. A. Irnanda, N. D. P. Dalimunthe, and R. Rosnelly, “Klasifikasi Tingkat Kedisiplinan Siswa Menggunakan Algoritma Machine Learning: Decision Tree, KNN, dan Naive Bayes,” *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, vol. 4, no. 3, pp. 1987–1992, 2026, doi: 10.62712/juktisi.v4i3.788.
- [22] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotaе, “Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset,” *Expert Systems with Applications*, vol. 212, no. August 2022, p. 118715, 2023, doi: 10.1016/j.eswa.2022.118715.