Prediksi Pola Penggunaan Waktu Mahasiswa Menggunakan Machine Learning

Wildan Ariel Heradi[1*], Aris Wahyu Prasetyo[2], Ericko Budhi Wirayudha[3], Muhammad Amir Yahya[4], Adzriel Bima Putra[5], Maulana Bintang Hayuka[6]

[1]Teknik Informatika
Universitas Duta Bangsa Surakarta
[1*]240103150@mhs.udb.ac.id

[2]Teknik Informatika
Universitas Duta Bangsa Surakarta
[2]240103126@mhs.udb.ac.id

[3]Teknik Informatika
Universitas Duta Bangsa Surakarta
[3]240103130@mhs.udb.ac.id

[4]Teknik Informatika
Universitas Duta Bangsa Surakarta
[4]240103258@mhs.udb.ac.id

[5]Teknik Informatika
Universitas Duta Bangsa Surakarta
[5]240103122@mhs.udb.ac.id

[6]Teknik Informatika
Universitas Duta Bangsa Surakarta
[6]240103136@mhs.udb.ac.id

Abstrak*— Pengelolaan waktu merupakan salah satu faktor yang berpengaruh terhadap produktivitas dan keberhasilan akademik mahasiswa. Pola penggunaan waktu yang kurang seimbang dapat berdampak pada menurunnya efektivitas belajar maupun aktivitas pendukung lainnya. Penelitian ini bertujuan membangun model prediksi pola penggunaan waktu mahasiswa serta membandingkan kinerja algoritma Decision Tree, K-Nearest Neighbor (KNN), dan Naive Bayes dalam proses klasifikasi. Data penelitian diperoleh melalui survei terhadap mahasiswa dan menghasilkan 518 data, yang setelah melalui tahap preprocessing berupa data cleaning, seleksi atribut, dan transformasi data menjadi 516 data valid. Variabel yang digunakan meliputi jam kuliah, jam belajar mandiri, jam santai atau hiburan, jam olahraga, dan jam organisasi, dengan target berupa kategori pola penggunaan waktu mahasiswa. Dataset dibagi menggunakan rasio 80:20 menjadi data latih dan data uji. Evaluasi dilakukan menggunakan Confusion Matrix dengan metrik akurasi sebagai dasar perbandingan kinerja model. Hasil penelitian menunjukkan bahwa algoritma Decision Tree memberikan performa terbaik dengan akurasi sebesar 95,19%, diikuti oleh K-Nearest Neighbor sebesar 88,46% dan Naive Bayes sebesar 77,88%. Temuan tersebut menunjukkan bahwa Decision Tree lebih efektif dalam mengenali pola penggunaan waktu mahasiswa pada dataset yang digunakan sehingga berpotensi dimanfaatkan sebagai pendekatan pendukung dalam analisis perilaku belajar mahasiswa.
Kata kunci— Machine Learning, Decision Tree, K-Nearest Neighbor, Naive Bayes, Pola Penggunaan Waktu.

Abstract*— Time management is one of the factors that influences students' productivity and academic performance. An unbalanced allocation of time may reduce learning effectiveness and limit participation in other beneficial activities. This study aims to develop a predictive model for student time-use patterns and compare the performance of the Decision Tree, K-Nearest Neighbor (KNN), and Naive Bayes algorithms for classification. The dataset was collected through a student survey consisting of 518 records. After preprocessing, including data cleaning, attribute selection, and data transformation, 516 valid records were retained for analysis. The input variables consisted of lecture hours, independent study hours, leisure time, exercise time, and organizational activities, while the target variable represented the category of student time-use patterns. The dataset was divided into training and testing sets using an 80:20 ratio. Model performance was evaluated using a Confusion Matrix with accuracy as the primary evaluation metric. The experimental results indicate that the Decision Tree algorithm achieved the highest accuracy of 95,19%, followed by K-Nearest Neighbor at 88.46% and Naive Bayes at 77.88%. These findings suggest that Decision Tree is more effective in classifying student time-use patterns on the dataset used and has potential to support educational data analysis related to student learning behavior.
Keywords— Machine Learning, Decision Tree, K-Nearest Neighbor, Naive Bayes, Student Time Management.

I. PENDAHULUAN

Perkembangan *Machine Learning* telah dimanfaatkan dalam berbagai bidang, termasuk pendidikan, untuk menganalisis data dan menghasilkan informasi yang dapat mendukung pengambilan keputusan [1], [2]. Dalam lingkungan pendidikan tinggi, pendekatan ini mulai digunakan untuk mengidentifikasi pola perilaku mahasiswa berdasarkan data aktivitas sehari-hari sehingga dapat memberikan gambaran mengenai kebiasaan belajar maupun pengelolaan waktu [1].

Pengelolaan waktu merupakan salah satu aspek yang berperan penting dalam menunjang keberhasilan akademik mahasiswa [1]. Dalam menjalani perkuliahan, mahasiswa dihadapkan pada berbagai aktivitas, seperti mengikuti perkuliahan, belajar mandiri, berorganisasi, berolahraga, hingga melakukan aktivitas hiburan. Pembagian waktu yang kurang seimbang dapat

mempengaruhi produktivitas, kualitas belajar, serta pencapaian akademik [1]. Oleh karena itu, diperlukan pendekatan berbasis data untuk memahami pola penggunaan waktu mahasiswa secara lebih objektif dan terukur [2].

Berbagai penelitian telah memanfaatkan algoritma klasifikasi berbasis *Machine Learning* untuk menganalisis data pada bidang pendidikan baik untuk memprediksi prestasi akademik mahasiswa maupun kinerja siswa berdasarkan data akademik dan perilaku [3], [4], [5]. Namun, penelitian yang secara khusus membandingkan performa beberapa algoritma dalam mengklasifikasikan pola penggunaan waktu mahasiswa masih relatif terbatas. Perbedaan karakteristik setiap algoritma memungkinkan hasil prediksi yang berbeda sehingga diperlukan evaluasi untuk menentukan metode yang paling sesuai dengan karakteristik data yang digunakan [6], [7]. Penelitian-penelitian sebelumnya umumnya berfokus pada prediksi prestasi atau kelulusan akademik, sedangkan kajian yang secara khusus membandingkan performa algoritma klasifikasi untuk pola penggunaan waktu mahasiswa berdasarkan lima dimensi aktivitas harian (jam kuliah, belajar mandiri, hiburan, olahraga, dan organisasi) masih jarang dilakukan. Kebaruan penelitian ini terletak pada penerapan dan perbandingan langsung tiga algoritma, yaitu Decision Tree, K-Nearest Neighbor, dan Naive Bayes, pada dataset survei aktivitas mahasiswa lokal, sehingga dapat memberikan rekomendasi model yang paling sesuai untuk konteks tersebut.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan membangun model klasifikasi pola penggunaan waktu mahasiswa menggunakan algoritma *Decision Tree, K-Nearest Neighbor* (*KNN*), dan *Naive Bayes*. Dataset yang digunakan diperoleh melalui survei aktivitas harian mahasiswa dan diproses melalui tahap *preprocessing* sebelum dilakukan pelatihan model. Selanjutnya, performa masing-masing algoritma dievaluasi menggunakan *Confusion Matrix* untuk menentukan model dengan tingkat akurasi terbaik dalam memprediksi pola penggunaan waktu mahasiswa.

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan metode *Machine Learning* untuk memprediksi pola penggunaan waktu mahasiswa berdasarkan aktivitas sehari-hari [1]. Tahapan penelitian diawali dengan pengumpulan data melalui survei, kemudian dilanjutkan dengan proses *preprocessing* yang meliputi pembersihan data, seleksi atribut, dan transformasi data agar sesuai untuk proses pemodelan. Dataset yang telah diproses selanjutnya dibagi menjadi data latih dan data uji, kemudian digunakan untuk membangun model klasifikasi menggunakan algoritma *Decision Tree*, *K-Nearest Neighbor* (*KNN*), dan *Naive Bayes*. Performa masing-masing model dievaluasi menggunakan *Confusion Matrix* sehingga dapat dilakukan perbandingan untuk menentukan algoritma dengan kinerja terbaik dalam mengklasifikasikan pola penggunaan waktu mahasiswa.

A. Pengumpulan Data

Penelitian ini menggunakan data yang diperoleh melalui penyebaran kuesioner kepada mahasiswa untuk memperoleh informasi mengenai pola penggunaan waktu dalam aktivitas sehari-hari. Kuesioner dirancang untuk mengumpulkan data terkait durasi waktu yang dialokasikan mahasiswa pada beberapa jenis kegiatan, yaitu jam kuliah, jam belajar mandiri, jam santai atau hiburan, jam olahraga, dan jam organisasi. Data yang terkumpul selanjutnya digunakan sebagai dasar dalam proses pembentukan model klasifikasi.

Hasil pengumpulan data menunjukkan bahwa diperoleh sebanyak 518 data responden. Selanjutnya dilakukan proses pemeriksaan awal untuk mengidentifikasi data yang tidak lengkap maupun tidak konsisten. Setelah tahap pembersihan data selesai, diperoleh sebanyak 516 data yang dinyatakan valid dan digunakan pada proses analisis.

Variabel target pada penelitian ini adalah pola penggunaan waktu mahasiswa yang diklasifikasikan ke dalam tiga kategori, yaitu Akademik Dominan, Seimbang, dan Non-Akademik Dominan. Seluruh data yang telah memenuhi kriteria kemudian digunakan sebagai masukan

pada tahapan *preprocessing* dan pembangunan model *Machine Learning*.

B. Preprocessing Data

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses pelatihan model *Machine Learning* [8]. Proses ini bertujuan menghasilkan data yang lebih bersih, konsisten, dan sesuai dengan kebutuhan algoritma klasifikasi sehingga dapat meningkatkan kualitas prediksi yang dihasilkan.

Tahap awal yang dilakukan adalah *data cleaning* untuk mengidentifikasi data yang tidak lengkap (*missing value*), data duplikat, maupun data yang memiliki format penulisan yang tidak konsisten. Berdasarkan hasil pemeriksaan, terdapat dua data yang tidak memenuhi kriteria sehingga dihapus dari dataset. Setelah proses tersebut selesai, jumlah data yang digunakan dalam penelitian menjadi 516 data.

Selanjutnya dilakukan proses seleksi atribut dengan mempertahankan variabel yang memiliki keterkaitan langsung dengan tujuan penelitian. Atribut yang tidak digunakan dalam proses klasifikasi, seperti informasi identitas responden, dihilangkan untuk mengurangi kompleksitas data dan mempermudah proses pemodelan.

Tahap berikutnya adalah transformasi data, yaitu mengubah data ke dalam format numerik agar dapat diproses oleh algoritma *Machine Learning*. Selain itu, dilakukan proses *label encoding* pada variabel target sehingga setiap kategori dapat direpresentasikan dalam bentuk numerik. Hasil dari seluruh tahapan *preprocessing* berupa dataset yang telah siap digunakan pada proses pembagian data, pelatihan model, dan evaluasi.

C. Pembagian Dataset

Dataset yang telah melalui tahap *preprocessing* kemudian dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*) [4]. Pembagian dilakukan menggunakan rasio 80:20, di mana 80% data digunakan untuk melatih model dan 20% sisanya digunakan untuk menguji kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dipelajari sebelumnya.

Berdasarkan rasio pembagian tersebut, diperoleh sebanyak 412 data sebagai data latih (*training data*) dan 104 data sebagai data uji (*testing data*). Data latih digunakan untuk membangun model klasifikasi menggunakan algoritma *Decision Tree*, *K-Nearest Neighbor* (*KNN*), dan *Naive Bayes*, sedangkan data uji digunakan untuk mengevaluasi performa masing-masing model menggunakan *Confusion Matrix*.

D. Pembangunan Model Machine Learning

Pada penelitian ini digunakan tiga algoritma klasifikasi, yaitu *Decision Tree*, *K-Nearest Neighbor* (*KNN*), dan *Naive Bayes* [7]. Ketiga algoritma dipilih karena memiliki karakteristik dan mekanisme klasifikasi yang berbeda sehingga dapat memberikan gambaran mengenai performa masing-masing metode dalam mengklasifikasikan pola penggunaan waktu mahasiswa. Setiap algoritma dibangun menggunakan data latih (*training data*) yang telah melalui tahap *preprocessing*, kemudian dievaluasi menggunakan data uji (*testing data*) untuk mengetahui kemampuan model dalam melakukan klasifikasi terhadap data yang belum pernah diproses sebelumnya.

1. Decision Tree

Decision Tree merupakan algoritma klasifikasi yang membentuk struktur pohon keputusan berdasarkan atribut yang memiliki kemampuan terbaik dalam membedakan setiap kelas data [9]. Pemilihan atribut dilakukan menggunakan nilai *Entropy* dan *Information Gain*. Pada penelitian ini, algoritma *Decision Tree* digunakan untuk mengklasifikasikan pola penggunaan waktu mahasiswa berdasarkan lima variabel aktivitas, yaitu jam kuliah, jam belajar mandiri, jam santai atau hiburan, jam olahraga, dan jam organisasi.

Nilai *Entropy* digunakan untuk mengukur tingkat ketidakpastian data pada setiap kelas dan dihitung menggunakan Persamaan (1).

$$Entropy(S) = -\Sigma\, pi\, log_2(pi) \qquad (1)$$

di mana *pi* merupakan probabilitas kemunculan kelas ke-*i* pada himpunan data *S*.

Selanjutnya, nilai *Information Gain* digunakan untuk menentukan atribut yang memberikan pemisahan data terbaik dalam proses pembentukan pohon keputusan. Perhitungan *Information Gain* ditunjukkan pada Persamaan (2).

$$Gain(S, A) = Entropy(S) - \Sigma \left(\frac{|Sv|}{|S|}\right) Entropy(Sv) \qquad (2)$$

2. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (*KNN*) merupakan algoritma klasifikasi yang menentukan kelas suatu data berdasarkan kedekatan jarak dengan sejumlah data latih (*training data*) yang memiliki karakteristik serupa [6]. Pada penelitian ini, *KNN* digunakan untuk mengklasifikasikan pola penggunaan waktu mahasiswa berdasarkan kemiripan lima variabel aktivitas yang digunakan sebagai fitur.

Perhitungan kedekatan antar data dilakukan menggunakan metode *Euclidean Distance* sebagaimana ditunjukkan pada Persamaan (3).

$$d(x, y) = \sqrt{\Sigma(xi - yi)^2} \qquad (3)$$

di mana *xi* dan *yi* merupakan nilai atribut dari dua data yang dibandingkan, sedangkan n menyatakan jumlah atribut yang digunakan. Data kemudian diklasifikasikan ke dalam kelas yang paling banyak muncul di antara sejumlah tetangga terdekat.

3. Naive Bayes

Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes untuk menentukan kemungkinan suatu data termasuk ke dalam kelas tertentu [7]. Algoritma ini mengasumsikan bahwa setiap atribut bersifat independen sehingga proses klasifikasi dapat dilakukan berdasarkan nilai probabilitas dari masing-masing atribut. Pada penelitian ini, *Naive Bayes* digunakan untuk mengklasifikasikan pola penggunaan waktu mahasiswa berdasarkan lima variabel aktivitas yang digunakan sebagai fitur.

Perhitungan probabilitas pada algoritma *Naive Bayes* dilakukan menggunakan Persamaan (4).

$$\left[P(H|X) = \frac{P(X|H)P(H)}{P(X)}\right]$$

di mana $P(H|X)P(H|X)P(H|X)$ merupakan probabilitas hipotesis terhadap data, $P(X|H)P(X|H)P(X|H)$ merupakan probabilitas data terhadap hipotesis, P(H)P(H)P(H) adalah probabilitas awal hipotesis, dan *P(X)P(X)P(X)* merupakan probabilitas kemunculan data.

E. Evaluasi Model

Evaluasi dilakukan untuk mengukur kemampuan masing-masing model dalam mengklasifikasikan pola penggunaan waktu mahasiswa [10]. Proses pengujian menggunakan data uji (*testing data*) sebanyak 104 data yang tidak digunakan selama proses pelatihan model.

Kinerja setiap model dianalisis menggunakan *Confusion Matrix*, yang menghasilkan empat komponen utama, yaitu *True Positive* (*TP*), *True Negative* (*TN*), *False Positive* (*FP*), dan *False Negative* (*FN*). Berdasarkan nilai tersebut, tingkat akurasi model dihitung menggunakan Persamaan (5).

Nilai akurasi dihitung menggunakan persamaan:

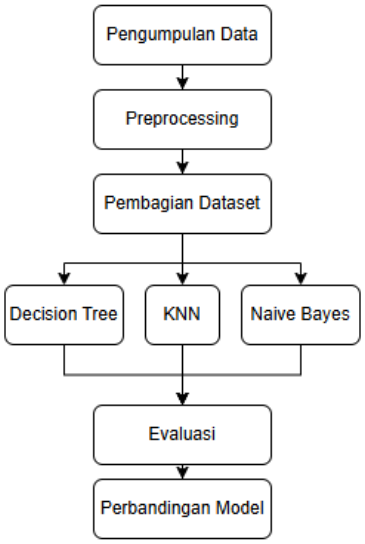
$$[Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

Nilai akurasi yang diperoleh digunakan sebagai dasar untuk membandingkan performa algoritma *Decision Tree*, *K-Nearest Neighbor* (*KNN*), dan *Naive Bayes*, sehingga dapat ditentukan model yang memberikan hasil klasifikasi terbaik pada dataset yang digunakan.

F. Alur Penelitian

Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1. Penelitian dilakukan melalui beberapa tahapan, yaitu pengumpulan data, *preprocessing*, pembagian dataset, pembangunan model menggunakan algoritma *Decision Tree, K-Nearest Neighbor* (*KNN*), dan *Naive Bayes*, serta evaluasi performa model untuk menentukan algoritma dengan kinerja terbaik.

Secara lebih rinci, tahap pengumpulan data dilakukan melalui penyebaran kuesioner daring kepada mahasiswa selama periode tertentu untuk merekam durasi aktivitas harian. Tahap preprocessing mencakup pemeriksaan data yang hilang atau tidak konsisten, seleksi atribut yang relevan dengan tujuan penelitian, serta transformasi data ke bentuk numerik agar dapat diproses oleh algoritma klasifikasi. Setelah data siap, dilakukan pembagian dataset dengan rasio 80:20 untuk membentuk data latih dan data uji. Pada tahap pembangunan model, setiap algoritma dilatih menggunakan data latih dengan parameter default, kemudian diuji menggunakan data uji yang belum pernah dipelajari sebelumnya. Tahap akhir berupa evaluasi performa dilakukan dengan menghitung nilai akurasi, precision, recall, dan F1-score dari Confusion Matrix pada masing-masing model, sehingga dapat diperoleh perbandingan objektif antar algoritma.



Gambar 1. Alur Penelitian

III. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil pengujian model *Machine Learning* yang telah dibangun berdasarkan dataset pola penggunaan waktu mahasiswa. Pembahasan diawali dengan hasil *preprocessing* data, kemudian dilanjutkan dengan hasil pengujian masing-masing algoritma, yaitu *Naive Bayes*, *K-Nearest Neighbor* (*KNN*), dan *Decision Tree*. Selanjutnya dilakukan perbandingan performa model berdasarkan hasil evaluasi untuk menentukan algoritma dengan kinerja terbaik.

A. Hasil Preprocessing Data

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses pelatihan model *Machine Learning*. Proses ini meliputi *data cleaning*, seleksi atribut, dan transformasi data sehingga diperoleh dataset yang lebih konsisten serta sesuai dengan kebutuhan proses klasifikasi.

Berdasarkan hasil *preprocessing*, dari 518 data awal diperoleh sebanyak 516 data yang memenuhi kriteria untuk digunakan pada tahap pemodelan. Selain itu, atribut yang tidak berkaitan dengan proses klasifikasi, seperti *timestamp*, nama universitas, dan program studi, dihapus sehingga hanya tersisa atribut yang digunakan sebagai fitur pada proses pelatihan model. Ringkasan hasil *preprocessing* disajikan pada Tabel I.

Tabel I. Statistik Deskriptif Dataset Hasil Preprocessing

Statistik Deskriptif Variabel Numerik

Variabel	N	Min	Max	Rata rata	Std. Deviasi
Jam kuliah/hari	516	0	4	2,11	1,25
Jam belajar mandiri/hari	516	0	4	1,32	0,79
Jam santai/hiburan/hari	516	0	4	1,56	1,03
Jam olahraga/hari	516	0	4	0,82	0,74
Jam organisasi/hari	516	0	4	1,02	0,89

Distribusi Variabel Target pada Data Final

Kategori	Jumlah	Presentase
Seimbang	238	46,1%
Akademik Dominan	217	42,1%
Non-Akademik Dominan	61	11,8%
Total	516	100%

Berdasarkan Tabel I, seluruh variabel numerik memiliki jumlah data yang sama, yaitu 516 data, yang menunjukkan bahwa dataset telah memenuhi kebutuhan proses pemodelan. Nilai rata-rata tertinggi terdapat pada variabel jam kuliah sebesar 2,11, sedangkan nilai rata-rata terendah terdapat pada variabel jam olahraga sebesar 0,82. Selain itu, distribusi variabel target menunjukkan bahwa kategori Seimbang memiliki jumlah data terbanyak, diikuti oleh Akademik Dominan dan Non-Akademik Dominan.

B. Hasil Data Cleaning

Tahap *data cleaning* dilakukan untuk meningkatkan kualitas dataset sebelum digunakan pada proses pemodelan. Proses ini meliputi identifikasi data yang tidak lengkap (*missing value*), penyeragaman format penulisan, serta penghapusan data yang tidak memenuhi kriteria. Setelah proses *data cleaning* selesai, dataset memiliki struktur yang lebih konsisten dan siap digunakan pada tahap selanjutnya. Perbandingan kondisi dataset sebelum dan sesudah *data cleaning* disajikan pada Tabel II.

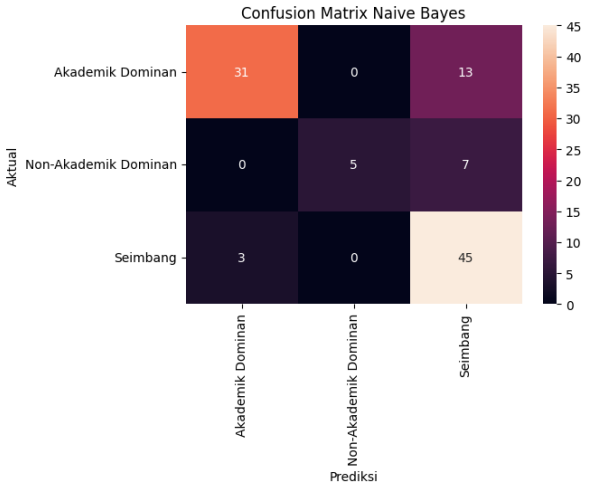
Tabel II. Perbandingan Kondisi Dataset Sebelum dan Sesudah Data Cleaning

Aspek	Sebelum Data Cleaning	Sesudah Data Cleaning
Jumlah data	518	516
Missing value	Ada	Tidak ada
Format jam	Campuran (teks&angka)	Numerik
Atribut	Seluruh atribut hasil survei	Hanya atribut yang digunakan pada pemodelan
Dataset siap pemodelan	Tidak	Ya

Berdasarkan Tabel II, proses *data cleaning* berhasil meningkatkan kualitas dataset melalui penghapusan data yang tidak valid serta penyeragaman format data menjadi bentuk numerik yang konsisten. Hasil tersebut menunjukkan bahwa dataset telah memenuhi kebutuhan untuk digunakan pada proses *Machine Learning*.

C. Hasil Pengujian Naive Bayes

Algoritma *Naive Bayes* digunakan untuk mengklasifikasikan pola penggunaan waktu mahasiswa berdasarkan lima variabel aktivitas yang digunakan sebagai fitur. Hasil pengujian algoritma *Naive Bayes* ditunjukkan melalui *Confusion Matrix* pada Gambar 2 dan hasil evaluasi pada Tabel III.



Gambar 2. Confusion Matrix Naive Bayes

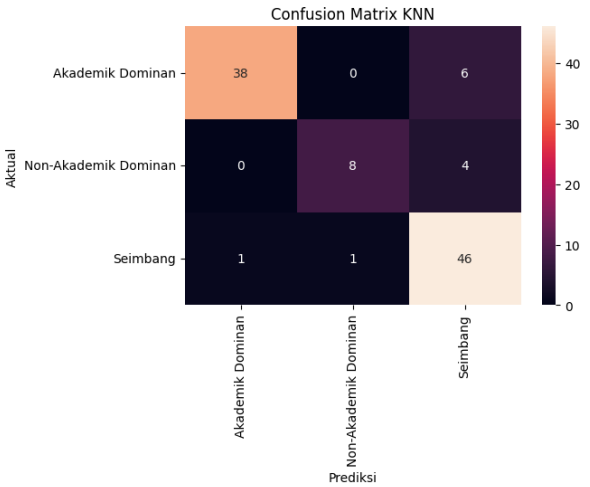
Tabel III. Hasil Performa Naive Bayes

Kelas	Precision	Recall	F1-Score	Support
Akademik Dominan	0.91	0.70	0.78	44
Non-Akademik Dominan	1.00	0.42	0.59	12
Seimbang	0.69	0.94	0.80	48
Accuracy	-	-	0.78	104
Macro Average	0.87	0.69	0.73	104
Weighted Average	0.82	0.78	0.77	104

Berdasarkan hasil pengujian, algoritma *Naive Bayes* memperoleh akurasi sebesar 77,88%. Nilai *F1-score* tertinggi diperoleh pada kelas Seimbang sebesar 0,80, diikuti kelas Akademik Dominan sebesar 0,78, sedangkan kelas Non-Akademik Dominan memperoleh nilai 0,59. Hasil tersebut menunjukkan bahwa model masih mengalami kesulitan dalam mengklasifikasikan kelas Non-Akademik Dominan.

D. Hasil Pengujian KNN

Algoritma *K-Nearest Neighbor* (*KNN*) digunakan untuk mengklasifikasikan pola penggunaan waktu mahasiswa berdasarkan tingkat kemiripan antar data. Hasil pengujian algoritma *KNN* ditunjukkan melalui *Confusion Matrix* pada Gambar 3 dan hasil evaluasi pada Tabel IV.



Gambar 3. Confusion Matrix KNN

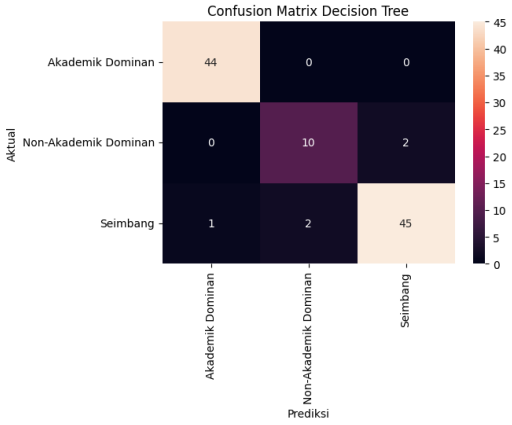
Tabel IV. Hasil Performa K-Nearest Neighbor (K=5)

Kelas	Precision	Recall	F1-Score	Support
Akademik Dominan	0.97	0.86	0.92	44
Non-Akademik Dominan	0.89	0.67	0.76	12
Seimbang	0.82	0.96	0.88	48
Accuracy	-	-	0.88	104
Macro Average	0.89	0.83	0.85	104
Weighted Average	0.89	0.88	0.88	104

Berdasarkan hasil pengujian, algoritma *K-Nearest Neighbor* (*KNN*) memperoleh akurasi sebesar 88,46%. Nilai *F1-score* tertinggi diperoleh pada kelas Akademik Dominan sebesar 0,92, diikuti kelas Seimbang sebesar 0,88, sedangkan kelas Non-Akademik Dominan memperoleh nilai 0,76. Hasil tersebut menunjukkan bahwa algoritma *KNN* mampu mengklasifikasikan sebagian besar data dengan baik pada dataset penelitian.

E. Hasil Pengujian Decision Tree

Algoritma *Decision Tree* digunakan untuk mengklasifikasikan pola penggunaan waktu mahasiswa berdasarkan proses pembentukan pohon keputusan dari atribut-atribut yang digunakan sebagai fitur. Hasil pengujian algoritma *Decision Tree* disajikan melalui *Confusion Matrix* pada Gambar 4 dan hasil evaluasi pada Tabel V.



Gambar 4. Confusion Matrix Decision Tree

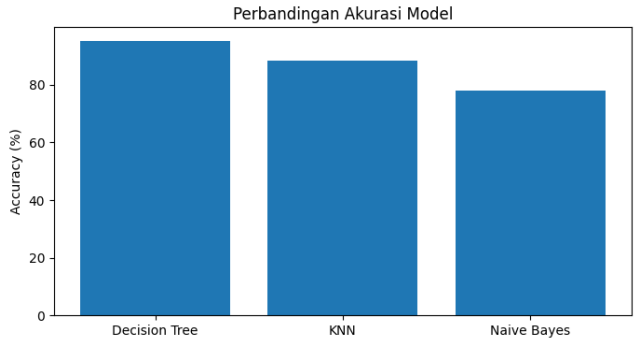
Tabel V. Hasil Performa Decision Tree

Kelas	Precision	Recall	F1-Score	Support
Akademik Dominan	0.98	1.00	0.99	44
Non-Akademik Dominan	0.83	0.83	0.83	12
Seimbang	0.96	0.94	0.95	48
Accuracy	-	-	0.95	104
Macro Average	0.92	0.92	0.92	104
Weighted Average	0.95	0.95	0.95	104

Berdasarkan hasil pengujian, algoritma *Decision Tree* memperoleh akurasi sebesar 95,19%. Nilai *F1-score* tertinggi diperoleh pada kelas Akademik Dominan sebesar 0,99, diikuti kelas Seimbang sebesar 0,95, sedangkan kelas Non-Akademik Dominan memperoleh nilai 0,83. Hasil tersebut menunjukkan bahwa algoritma *Decision Tree* mampu menghasilkan klasifikasi yang akurat pada dataset penelitian.

F. Perbandingan Performa Model

Perbandingan performa model dilakukan untuk mengetahui algoritma yang memberikan hasil klasifikasi terbaik pada dataset pola penggunaan waktu mahasiswa. Perbandingan dilakukan berdasarkan nilai akurasi serta metrik evaluasi yang diperoleh dari masing-masing algoritma. Ringkasan hasil perbandingan performa model disajikan pada Gambar 5 dan Tabel VI.



Gambar 5. Perbandingan antar model

Tabel VI. Perbandingan Hasil Evaluasi Model

Algoritma	Akurasi (%)	Precision	Recall	F1-Score
Naive Bayes	77,88	0,82	0,78	0,77
K-Nearest Neighbor (KNN)	88,46	0,89	0,88	0,88
Decision Tree	95,19	0,95	0,95	0,95

Berdasarkan hasil perbandingan pada Tabel VI, algoritma *Decision Tree* memperoleh performa terbaik dengan akurasi sebesar 95,19%, diikuti oleh *K-Nearest Neighbor* (*KNN*) sebesar 88,46%, sedangkan *Naive Bayes* memperoleh akurasi sebesar 77,88%. Selain memiliki akurasi tertinggi, *Decision Tree* juga menghasilkan nilai *precision*, *recall*, dan *F1-score* yang lebih baik dibandingkan algoritma lainnya. Hasil tersebut menunjukkan bahwa *Decision Tree* merupakan algoritma yang paling sesuai untuk mengklasifikasikan pola penggunaan waktu mahasiswa pada dataset penelitian.

IV. KESIMPULAN

Penelitian ini berhasil membangun model klasifikasi untuk memprediksi pola penggunaan waktu mahasiswa menggunakan algoritma *Naive Bayes*, *K-Nearest Neighbor* (*KNN*), dan *Decision Tree*. Tahapan penelitian meliputi proses *preprocessing* data, pembagian dataset, pembangunan model, serta evaluasi menggunakan *Confusion Matrix* dan metrik evaluasi lainnya.

Berdasarkan hasil pengujian, algoritma *Decision Tree* memberikan performa terbaik dengan akurasi sebesar 95,19%, diikuti oleh *K-Nearest Neighbor* (*KNN*) sebesar 88,46%, dan *Naive Bayes* sebesar 77,88%. Hasil tersebut menunjukkan bahwa *Decision Tree* lebih efektif dalam mengklasifikasikan pola penggunaan waktu mahasiswa pada dataset yang digunakan. Dengan demikian, tujuan penelitian untuk membandingkan performa ketiga algoritma *Machine Learning* dalam memprediksi pola penggunaan waktu mahasiswa telah berhasil dicapai.

Penelitian ini memiliki beberapa keterbatasan. Pertama, data yang digunakan hanya diperoleh dari mahasiswa Universitas Duta Bangsa Surakarta melalui survei mandiri (self-report), sehingga hasil klasifikasi belum tentu dapat digeneralisasi pada populasi mahasiswa di institusi lain. Kedua, jumlah kelas Non-Akademik Dominan pada dataset relatif kecil dibandingkan dua kelas lainnya, yang berkontribusi pada rendahnya nilai recall dan F1-score pada kelas tersebut di seluruh model. Ketiga, penelitian ini belum melakukan tuning hyperparameter secara mendalam pada masing-masing algoritma sehingga performa yang dilaporkan berpotensi masih dapat ditingkatkan. Secara praktis, temuan ini mengimplikasikan bahwa algoritma Decision Tree dapat dipertimbangkan sebagai pendekatan awal untuk membangun sistem pemantauan pola penggunaan waktu mahasiswa berbasis data, misalnya sebagai dasar bagi unit akademik dalam memberikan intervensi atau bimbingan kepada mahasiswa dengan kecenderungan pola non-akademik dominan. Penelitian selanjutnya disarankan untuk memperluas cakupan data, menyeimbangkan jumlah data antar kelas, serta melakukan optimasi parameter model guna meningkatkan akurasi klasifikasi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Penulis juga menyampaikan terima kasih kepada dosen pengampu mata kuliah Kecerdasan Buatan yang telah memberikan arahan, bimbingan, serta dukungan dalam penyusunan penelitian ini.

Selain itu, penulis juga berterima kasih kepada seluruh pihak yang telah membantu dalam proses pengumpulan data, pengolahan data, serta penyusunan laporan ini. Dukungan dan kerja sama yang diberikan sangat berarti dalam kelancaran penyelesaian penelitian ini.

REFERENSI

[1] J. Lu *et al*., "Machine learning analysis of factors affecting college students' academic performance," *Front. Psychol.*, vol. 15, 2024, doi: 10.3389/fpsyg.2024.1447825.

[2] M. Novitri Waruwu *et al*., "IMPLEMENTASI ALGORITMA MACHINE LEARNING UNTUK DETEKSI PERFORMA AKADEMIK MAHASISWA (IMPLEMENTATION OF MACHINE LEARNING ALGORITHM FOR STUDENT ACADEMIC PERFORMANCE DETECTION)."

[3] H. Pratiwi, M. Ibnu Sa, and S. Widya Cipta Dharma, "Strategi Manajemen Pendidikan Berbasis Machine Learning untuk Prediksi Prestasi Siswa," *BEduManageRs Journal Borneo Educational Management and Research Journal*, vol. 6, no. 1, 2025.

[4] Riska Rismaya, Dwi Yuniarto, and David Setiadi, "Penerapan Algoritma Machine Learning dalam Prediksi Prestasi Akademik Mahasiswa," *Router : Jurnal Teknik Informatika dan Terapan*, vol. 3, no. 1, pp. 15–23, Feb. 2025, doi: 10.62951/router.v3i1.389.

[5] A. Florent Amor, M. Rizal Ihsanudin, S. Oka Yulistianto, and D. Direvisi, "Prediksi Kinerja Siswa SMP Berdasarkan Data Akademik dan Perilaku Menggunakan Machine Learning," vol. 18, no. 2, pp. 67–76, 2025, doi: 10.30989/teknomatika.

[6] 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, Jan. 2026, doi: 10.62712/juktisi.v4i3.788.

[7] Z. Amri, K. Kusrini, and K. Kusnawi, "Prediksi Tingkat Kelulusan Mahasiswa menggunakan Algoritma Naïve Bayes, Decision Tree, ANN, KNN, dan SVM," *Edumatic: Jurnal Pendidikan Informatika*, vol. 7, no. 2, pp. 187–196, Dec. 2023, doi: 10.29408/edumatic.v7i2.18620.

[8] T. Gori, A. Sunyoto, and H. Al Fatta, "Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 215–224, Feb. 2024, doi: 10.25126/jtiik.20241118074.

[9] E. F. Wati, L. Indriyani, E. Sunita, and A. D. Kuswanto, "Optimasi Machine Learning dalam Memprediksi Kelulusan Mahasiswa", [Online]. Available: http://jurnal.bsi.ac.id/index.php/reputasi

[10] A. S. Fauzan, A. Irma, P. Sari, and I. Ali, "ANALISIS PERBANDINGAN ALGORITMA DECISIOIN TREE DAN NAÏVE UNTUK MENGEVALUASI PRESTASI BELAJAR SISWA STUDI KASUS: SMK AL-MUSYAWIRIN," 2024.