

ANALISIS PROGRES AKADEMIK SEBAGAI INDIKATOR KETERLAMBATAN PEMBAYARAN UKT MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (STUDI KASUS UBP KARAWANG)

Muhamad Yazid Nizar Rafif, Bayu Priyatna, Fitria Nurapriani, Baenil Huda

Sistem Informasi, Universitas Buana Perjuangan Karawang

Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Indonesia

si22.muhamadrafif@mhs.ubpkarawang.ac.id

ABSTRAK

Pembayaran Uang Kuliah Tunggal (UKT) merupakan kewajiban mahasiswa yang sangat berpengaruh dalam mendukung kelancaran aktivitas akademik dan stabilitas keuangan di perguruan tinggi. Namun, pada kenyataannya masih terdapat mahasiswa yang mengalami keterlambatan yang dapat dipengaruhi oleh berbagai faktor serta kondisi tertentu. Permasalahan ini menjadi dorongan dalam melakukan analisis berbasis data untuk mengidentifikasi pola keterlambatan secara sistematis. Penelitian ini bertujuan untuk menganalisis progres akademik mahasiswa sebagai indikator keterlambatan pembayaran menggunakan algoritma *K-Nearest Neighbor* (K-NN). Metode yang digunakan adalah pendekatan kuantitatif dengan teknik klasifikasi *data mining* berdasarkan data historis mahasiswa di Universitas Buana Perjuangan Karawang. Dataset yang digunakan berjumlah 4.558 mahasiswa yang terdiri dari angkatan 2021 dan 2022 dengan atribut angkatan, periode semester, dan IPK terakhir. Data diproses melalui tahapan *preprocessing*, pembentukan label keterlambatan berdasarkan progres semester, pembagian data latih dan data uji, pemodelan algoritma KNN, serta evaluasi model menggunakan *confusion matrix* dan *matrix precision, recall*, serta *F1-score* dengan nilai parameter $K = 7$. Hasil penelitian menunjukkan model bahwa model menghasilkan tingkat akurasi sebesar 52,19%, yang menunjukkan bahwa kemampuan klasifikasi masih berada di tingkat sedang. Meskipun demikian, model ini mampu memberikan gambaran awal mengenai pola keterlambatan pembayaran, sehingga dapat dimanfaatkan sebagai dasar dalam pengembangan sistem pendukung keputusan.

Kata kunci: *K-Nearest Neighbor, Data Mining, Keterlambatan UKT*

1. PENDAHULUAN

Pembayaran Uang Kuliah Tunggal (UKT) merupakan kewajiban mahasiswa dalam sistem administrasi keuangan di perguruan tinggi. Ketepatan waktu pembayaran UKT memiliki peran penting dalam mendukung kelancaran proses akademik serta menjaga stabilitas keuangan kampus [1].

Namun, pada kenyataannya masih terdapat mahasiswa yang mengalami keterlambatan dalam membayar UKT. Kondisi ini dapat dipengaruhi oleh beberapa faktor seperti kondisi ekonomi keluarga, manajemen waktu dan sebagainya [2].

Permasalahan yang ditemukan pada kampus swasta di Karawang, mahasiswa yang belum melakukan pembayaran UKT hingga batas waktu yang ditentukan tidak dapat mengakses sistem akademik, termasuk proses dalam pengisian Kartu Rencana Studi (KRS). Kondisi ini menunjukkan bahwa keterlambatan pembayaran UKT tidak hanya berdampak pada aspek administrasi, tetapi juga menghambat aktivitas akademik mahasiswa. Oleh karena itu, diperlukan suatu pendekatan yang mampu menganalisis pola keterlambatan pembayaran secara sistematis berdasarkan data yang tersedia.

Data mahasiswa yang tersimpan pada sistem informasi akademik yang dikola oleh Pusat Data dan Informasi (PUSDATIN) memiliki potensi untuk dimanfaatkan dalam mengidentifikasi keterlambatan pembayaran. Pendekatan *data mining* dapat digunakan

untuk menggali pola tersembunyi pada data historis mahasiswa [3].

Salah satu algoritma klasifikasi yang dapat digunakan untuk menggali pola tersembunyi adalah *K-Nearest Neighbor* (K-NN) [4].

Algoritma tersebut bekerja dengan cara mengukur tingkat kemiripan antar data berdasarkan jarak tertentu melalui atribut seperti angkatan, periode semester, dan IPK terakhir. Melalui pendekatan ini, mahasiswa dapat diklasifikasikan ke dalam kategori tertentu, sehingga dapat diketahui karakteristik data mahasiswa yang cenderung mengalami keterlambatan.

Sejumlah penelitian sebelumnya telah memanfaatkan algoritma *K-Nearest Neighbor* (K-NN) untuk memprediksi keterlambatan pembayaran biaya pendidikan [5].

Hasil penelitian di tingkat perguruan tinggi menunjukkan bahwa penerapan K-NN menghasilkan akurasi sekitar 52% dengan variasi nilai parameter K tertentu (6).

Sementara itu, penelitian yang dilakukan pada lingkungan sekolah menunjukkan performa yang lebih tinggi, dengan nilai akurasi mencapai 90% dengan tambahan metode evaluasi berupa *cross-validation* [6].

Perbedaan hasil tersebut mengindikasikan bahwa kinerja algoritma K-NN sangat dipengaruhi oleh karakteristik dataset, pemilihan atribut, serta metode evaluasi yang digunakan [7].

Selain itu, penelitian yang menyesuaikan model klasifikasi dengan kondisi data lokal yang dikelola langsung oleh institusi masih tergolong terbatas [8].

Berdasarkan uraian tersebut, permasalahan dalam penelitian ini berfokus pada bagaimana menganalisis pola keterlambatan pembayaran UKT berdasarkan progres akademik mahasiswa menggunakan algoritma *K-Nearest Neighbor*, serta faktor-faktor yang memengaruhi terjadinya keterlambatan, dan tingkat akurasi dari model klasifikasi yang dihasilkan [9].

Penelitian ini bertujuan untuk menerapkan algoritma *K-Nearest Neighbor* dalam menganalisis pola keterlambatan pembayaran UKT mahasiswa, mengidentifikasi karakteristik mahasiswa yang berpotensi mengalami keterlambatan, serta mengukur performa model klasifikasi yang dibangun.

Sebagai upaya penyelesaian permasalahan tersebut, penelitian ini menggunakan pendekatan data mining dengan metode klasifikasi algoritma *K-Nearest Neighbor*. Hasil dari penelitian ini diharapkan dapat memberikan gambaran awal mengenai pola keterlambatan pembayaran UKT serta menjadi dasar dalam pengembangan sistem pendukung keputusan berbasis data di perguruan tinggi, khususnya pada pengelolaan administrasi keuangan akademik.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data Mining merupakan sebuah proses menggali pola, hubungan, atau informasi tersembunyi yang berguna dari kumpulan data-data besar. Data-data tersebut berupa seperti data transaksi, data administrasi, atau dataset lainnya [10].

Menurut Putra dan Wijaya (2023), *data mining* berperan penting dalam bidang seperti pendidikan, kesehatan, perbankan, dan pemerintahan karena kemampuannya menemukan pola dari data historis untuk menghasilkan sebuah prediksi analisis masa depan [11].

2.2. Knowledge Discovery in Database (KDD)

KDD adalah proses yang dimulai dari pengumpulan data mentah hingga menghasilkan pengetahuan baru yang dapat digunakan sebagai dasar analisis dan pengambilan keputusan [6]. Proses ini meliputi *data selection*, *data preprocessing*, *data transformation*, *data mining*, dan evaluasi hasil [12].

2.3. Algoritma K-Nearest Neighbor (K-NN)

K-Nearest Neighbor merupakan algoritma *supervised learning* yang digunakan untuk melakukan klasifikasi dengan cara membandingkan data baru terhadap data latih yang sudah ada [13].

Untuk menentukan prediksi, K-NN mencari nilai k terdekat berdasarkan jarak tertentu [14]. Setelah jarak dihitung, K-NN melihat kelas yang paling banyak muncul dari tetangga terdekatnya [15].

Kelas yang paling banyak muncul itulah yang dipilih sebagai hasil prediksi [16].

2.4. Evaluasi Confusion Matrix

Tahap evaluasi yaitu untuk melihat tingkat performa dari pola yang dihasilkan algoritma [17].

Parameter yang digunakan dalam mengevaluasi algoritma adalah *Confusion Matrix*. *Confusion Matrix* merupakan metode evaluasi yang digunakan untuk menilai performa model klasifikasi dengan cara membandingkan hasil prediksi model terhadap data aktual [18].

Dalam *confusion matrix* terdapat empat komponen utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Keempat komponen tersebut merupakan dasar dalam menghitung performa model [19].

2.5. Machine Learning

Machine Learning merupakan salah satu dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan sistem yang mampu belajar dari data tanpa harus diprogram secara eksplisit untuk setiap tugas [20].

Melalui pendekatan, komputer dapat mengenali pola, membuat prediksi, serta mengambil keputusan berdasarkan data yang telah dipelajari sebelumnya [21].

2.6. Penelitian Terdahulu

Penelitian yang dilakukan oleh Muhamad Rosyid dkk, dalam artikel tersebut bertujuan untuk memprediksi keterlambatan pembayaran biaya kuliah mahasiswa dengan menggunakan *K-Nearest Neighbor* (K-NN). Latar belakang penelitian tersebut didasarkan pada pentingnya pengelolaan keuangan pada perguruan tinggi, khususnya dalam mengantisipasi mahasiswa yang berpotensi mengalami keterlambatan pembayaran.

Selanjutnya, Muhamad Rosyid dkk menunjukkan bahwa algoritma K-NN mampu digunakan untuk melakukan klasifikasi keterlambatan pembayaran dengan tingkat akurasi yang cukup baik. Evaluasi model dilakukan menggunakan *confusion matrix* untuk mengukur performa klasifikasi berdasarkan nilai akurasi yang dihasilkan [22].

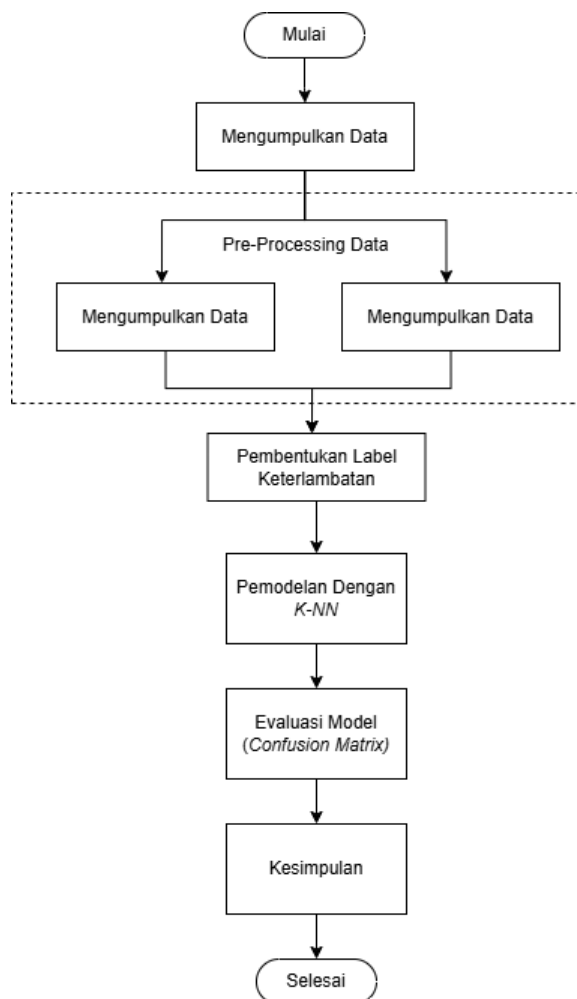
Namun, penelitian ini masih memiliki keterbatasan pada pemilihan atribut sehingga kinerja model kurang efektif. Temuan penelitian menunjukkan bahwa metode K-NN efektif dalam memprediksi keterlambatan pembayaran mahasiswa yaitu dengan pemilihan atribut-atribut yang relevan, sehingga dapat dijadikan sebagai dasar dalam pengambilan keputusan.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif, dengan seluruh proses analisis didasarkan pada pengolahan data numerik dan perhitungan matematis menggunakan algoritma *K-Nearest Neighbor* (K-NN). Pendekatan ini dipilih untuk menghasilkan model klasifikasi yang mampu

mengidentifikasi secara objektif progres akademik sebagai indikator keterlambatan pembayaran UKT mahasiswa.

Pengolahan data dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python dan memanfaatkan platform *Google Colaboratory* (*Google Colab*). Tahapan penelitian meliputi pengumpulan data, pembentukan label keterlambatan, pemodelan *K-Nearest Neighbor*, dan evaluasi model menggunakan *confusion matrix*. Adapun alur penelitian yang dilakukan dapat dilihat pada gambar berikut.



Gambar 1. Alur Prosedur Penelitian

Gambar1. Menunjukkan bahwa penelitian diawali dengan tahap pengumpulan data. Data yang digunakan berupa data historis mahasiswa yang diperoleh dari Pusat Data dan Informasi (PUSDATIN) Universitas Buana Perjuangan Karawang. Data tersebut memuat informasi riwayat mahasiswa selama menjalani proses perkuliahan. Dataset yang digunakan berjumlah 4.558 data mahasiswa, yang terdiri dari mahasiswa angkatan 2021 dan 2022 yang berisikan kolom atribut NIM mahasiswa, angkatan, periode semester, dan IPK terakhir. Berikut ilustrasi data keseluruhan dapat dilihat pada tabel 2.

Tabel 1. Ilustrasi Data Historis Mahasiswa

No	NIM	Angkatan	Periode Semester	IPK Terakhir
1	2141622120101	2021	20242	3.5
2	2141622120102	2021	20242	3.75
3	2141622120103	2021	20242	3.85
4	2141622120104	2021	20242	3.5
5	2141622120105	2021	20242	3.88
...	...	...	...	...
4.556	22416257201121	2022	20252	4
4.557	22416257201122	2022	20252	3.98
4.558	22416257201123	2022	20252	3.75

Setelah data terkumpul tahap berikutnya adalah proses *preprocessing* yang mencakup transformasi dan normalisasi untuk memastikan kualitas data yang digunakan. Data yang telah diproses kemudian dibagi menjadi dua, yaitu data latih dan data uji, guna mendukung proses evaluasi model. Tahapan selanjutnya adalah pemodelan menggunakan algoritma *K-Nearest Neighbor* (K-NN), dimana model tersebut dibangun berdasarkan data latih untuk memprediksi keterlambatan pembayaran. Model yang telah terbentuk kemudian di evaluasi menggunakan *confusion matrix* melalui perhitungan matematis untuk mengukur tingkat akurasi prediksi, sehingga dapat memberikan gambaran yang lebih jelas mengenai kinerja model yang dihasilkan.

3.1. Pre-processing Data

Tahapan *pre-processing data* dilakukan untuk memastikan data yang digunakan memiliki kualitas yang baik sebelum masuk ke proses pemodelan. Proses ini mencakup beberapa langkah, yaitu *data cleaning*, seleksi fitur, serta normalisasi data.

Tahap *data cleaning* dilakukan dengan menghapus data yang tidak lengkap (*missing value*) serta menghilangkan data duplikat agar tidak memengaruhi hasil proses klasifikasi. Selanjutnya, dilakukan proses seleksi fitur dengan menentukan atribut yang relevan terhadap kebutuhan prediksi, yaitu angkatan, periode semester, dan IPK terakhir. Sementara itu, NIM tidak digunakan sebagai atribut karena hanya berfungsi sebagai identitas mahasiswa.

Karena algoritma *K-Nearest Neighbor* sensitif terhadap skala data, maka dilakukan normalisasi menggunakan metode *Min-Max Scaling* dengan rumus:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Normalisasi ini bertujuan untuk menyamakan rentang nilai setiap fitur agar berada pada interval 0-1 sehingga tidak terjadi dominasi atribut tertentu dalam perhitungan jarak.

3.2. Pembentukan Label Keterlambatan

Pada penelitian ini mengalami keterbatasan dalam atribut data yaitu stastus pembayaran, Sehingga diperlukan proses pembentukan label secara mandiri berdasarkan informasi yang tersedia. Proses ini bertujuan untuk mengidentifikasi mahasiswa yang

mengalami kendala dalam progres akademiknya sebagai indikator keterlambatan pembayaran UKT.

Tahap awal yang dilakukan adalah mengekstrak informasi dari atribut periode semester. Atribut ini memiliki format lima digit, contohnya 20242, yang dimana empat digit pertama menunjukkan tahun akademik dan satu digit terakhir menunjukkan periode semester (1 = untuk semester ganjil dan 2 untuk semester genap). Proses ekstraksi menggunakan manipulasi string dalam bahasa pemrograman *python* untuk memisahkan kedua informasi tersebut ke dalam atribut baru.

Berikut merupakan implementasi ekstraksi data dapat dilihat pada gambar 2.

```
# Ambil tahun dari Periode Semester
df['Tahun Periode'] = df['Periode Semester'].astype(str).str[:4].astype(int)

# Ambil semester (1 atau 2)
df['Semester Periode'] = df['Periode Semester'].astype(str).str[-1].astype(int)
```

Gambar 2. Implementasi Ekstraksi Data

Hasil dari proses ini menghasilkan dua atribut baru, yaitu tahun periode dan semester periode, yang akan digunakan pada tahap perhitungan selanjutnya.

Selanjutnya, perhitungan semester keseluruhan dilakukan untuk menentukan posisi semester mahasiswa berdasarkan data terbaru yang tercatat. Perhitungan ini menggunakan rumus berikut:

$$\text{Semester Keseluruhan} = (\text{Tahun Periode} - \text{Angkatan}) \times 2 + \text{Semester Periode} \quad (2)$$

Rumus ini digunakan karena terdapat dua semester dalam satu tahun akademik. Selisih antara tahun periode semester dan tahun angkatan dikalikan dua untuk mendapatkan jumlah semester yang telah diselesaikan, kemudian ditambahkan ke jumlah semester dalam periode tersebut.

Selanjutnya implementasi perhitungan semester keseluruhan dapat dilihat pada gambar 3.

```
df['Semester Keseluruhan'] = ((df['Tahun Periode'] - df['Angkatan']) * 2) + df['Semester Periode']
```

Gambar 3. Implementasi Perhitungan Semester Keseluruhan

Setelah mendapatkan nilai semester keseluruhan, langkah selanjutnya adalah menentukan semester yang tepat, yaitu semester ideal yang seharusnya diselesaikan mahasiswa berdasarkan jangka waktu studinya. Dalam penelitian ini, acuan waktu yang digunakan adalah tahun 2025 semester 2, sesuai dengan periode tahun saat ini.

Langkah terakhir adalah menetapkan label keterlambatan. Mahasiswa dikategorikan sebagai tertinggal (1) jika nilai semester keseluruhan mereka lebih rendah daripada nilai semester seharusnya. Sebaliknya, mahasiswa dikategorikan normal (0) jika mahasiswa tersebut telah mencapai nilai yang diharapkan.

Berikut dibawah ini adalah implementasi label keterlambatan dapat dilihat pada gambar 4.

```
df['Status_Keterlambatan'] = df.apply(
    lambda x: 1 if x['Semester Keseluruhan'] < x['Semester Saat ini'] else 0, axis=1
)
```

Gambar 4. Implementasi Label Keterlambatan

Melalui proses ini, diperoleh atribut baru berupa label status keterlambatan, yang digunakan sebagai variabel target dalam proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* (K-NN). Pendekatan ini memungkinkan identifikasi keterlambatan secara objektif berdasarkan perbandingan antara kemajuan semester keseluruhan dan semester saat ini.

3.3. K-Nearest Neighbor

Dalam penelitian ini, algoritma *K-Nearest Neighbor* (K-NN) digunakan untuk mengelompokkan mahasiswa ke dalam dua kategori, yaitu normal dan tertinggal dalam pembayaran UKT. K-NN merupakan metode klasifikasi yang termasuk dalam *instance-based learning*, di mana penentuan kelas pada data uji dilakukan berdasarkan tingkat kedekatannya terhadap sejumlah data latih.

Pada algoritma K-NN, kedekatan antar data dihitung menggunakan metode *euclidean distance*. Rumus perhitungan jarak antara data uji dan data latih dinyatakan sebagai berikut:

$$(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Dimana :

- $d_{x,y}$: Jarak *euclidean* antara data latih dan uji
- x : Nilai atribut ke- i pada data latih
- y : Nilai atribut ke- i pada data uji
- n : Jumlah data atribut

Dalam penelitian ini, dilakukan pengujian beberapa nilai k untuk menentukan parameter terbaik. Berdasarkan hasil evaluasi model, nilai $K = 7$ dipilih sebagai parameter optimal karena menghasilkan performa klasifikasi terbaik pada data uji.

Sebelum perhitungan jarak dilakukan, seluruh fitur telah melalui proses normalisasi agar setiap atribut memiliki kontribusi yang seimbang dalam perhitungan jarak *euclidean*.

3.4. Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja algoritma *K-Nearest Neighbor* dalam mengklasifikasikan keterlambatan pembayaran UKT mahasiswa. Pada penelitian ini, proses evaluasi dilakukan dengan menggunakan *confusion matrix*.

Confusion matrix merupakan salah satu metode evaluasi yang digunakan untuk melihat jumlah prediksi yang tepat maupun yang keliru pada setiap kelas. Matriks ini memiliki empat komponen utama, yaitu:

- a. *True Positive* (TP): jumlah data terlambat yang diprediksi terlambat
- b. *True Negative* (TN): jumlah data tepat waktu yang diprediksi tepat waktu.

- c. *False Positive* (FP): jumlah data tepat waktu yang diprediksi terlambat.
 - d. *False Negative* (FN): jumlah data terlambat yang diprediksi tepat waktu.
- Struktur *confusion matrix* dapat digambarkan sebagai berikut:

Tabel 2. Struktur *Confusion Matrix*

Aktual		Prediksi	
		Positif	Negatif
	Positif	TP	FP
	Negatif	FN	TN

Dalam penelitian ini, evaluasi dilakukan 20% data uji dari total 4.558 dataset, sehingga kinerja model mencerminkan kemampuan algoritma dalam menggeneralisasi terhadap data yang belum pernah dilatih sebelumnya.

4. HASIL DAN PEMBAHASAN

4.1. Pembentukan Label Keterlambatan

Dataset awal yang didapati dari PUSDATIN tidak memiliki atribut mengenai status pembayaran UKT. Oleh karena itu, penelitian ini dilakukan pembentukan label secara terstruktur berdasarkan progres semester mahasiswa.

Atribut periode semester memiliki format lima digit, contohnya 20211, yang terdiri dari empat digit pertama sebagai tahun akademik dan satu digit terakhir sebagai periode semester (1 = semester ganjil, 2 = semester genap). Dari atribut tersebut dilakukan proses ekstraksi untuk memperoleh dua komponen, yaitu tahun periode dan semester periode.

Semester keseluruhan mahasiswa dihitung berdasarkan selisih antara tahun periode dan tahun angkatan, dengan mempertimbangan bahwa dalam satu tahun akademik terdapat dua semester.

Perhitungan secara matematis dilakukan menggunakan rumus:

$$\text{Semester Keseluruhan} = (\text{Tahun Periode} - \text{Angkatan}) \times 2 + \text{Semester Periode} \quad (4)$$

Selanjutnya dihitung semester yang seharusnya ditempuh mahasiswa berdasarkan tahun akademik 2025 periode semester 2 sebagai acuan waktu penelitian. Label status keterlambatan ditentukan dengan semester keseluruhan.

Dengan begitu mahasiswa dikategorikan sebagai:

- a. 0 (Normal), jika proses semester sesuai atau tidak tertinggal.
- b. 1 (Tertinggal), jika semester saat ini lebih rendah dari semester yang seharusnya.

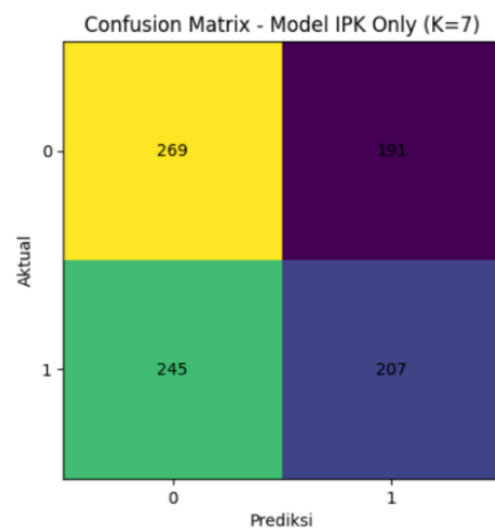
Berdasarkan hasil pengolahan, diperoleh 2.441 mahasiswa dalam kategori tertinggal. Pendekatan ini digunakan sebagai indikator keterlambatan progres akademik yang berkaitan dengan sistem pembayaran UKT, karena pada sistem akademik kampus mahasiswa yang belum melakukan pembayaran hingga batas waktu yang sudah ditentukan tidak dapat

mengakses sistem akademik untuk pengisian KRS maupun layanan lainnya.

4.2. Hasil Pengujian Model KNN

Penelitian ini menggunakan data sebanyak 4.558 mahasiswa yang berasal dari tahun angkatan 2021 dan 2022. Dataset tersebut kemudian dibagi dengan proporsi 80% untuk data latih (3.646 data) dan 20% untuk data uji (912 data). Data uji yang berjumlah 912 ini digunakan untuk evaluasi kinerja model.

Proses pengujian model dilakukan dengan menerapkan algoritma *K-Nearest Neighbor* (K-NN) dengan menggunakan nilai parameter terbaik, yaitu $K = 7$. Hasil dari evaluasi model kemudian disajikan dalam bentuk *confusion matrix* seperti yang ditunjukkan pada gambar berikut.

Gambar 5. Hasil *Confusion Matrix* dengan $K=7$

Berdasarkan gambar 2 menunjukkan hasil klasifikasi model *K-Nearest Neighbor* (K-NN) pada rasio perbandingan 80:20, dimana model berhasil memprediksi dengan benar 269 data pada kategori “normal” dan 207 data pada kategori “tertinggal”, terdapat 245 data yang diklasifikasikan “tertinggal” tetapi diprediksikan “normal” (False Negative) dan terakhir terdapat 191 data yang diklasifikasikan “normal” tetapi di prediksi “tertinggal” (False Positive). Berdasarkan hal tersebut dapat diketahui bahwa model berhasil memprediksi dengan benar 269 mahasiswa “normal” dan 207 mahasiswa “tertinggal”

4.3. Perhitungan Akurasi

Nilai akurasi dihitung menggunakan rumus matematis seperti berikut.

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

$$\text{Accuracy} = \frac{269 + 207}{912}$$

$$\text{Accuracy} = 52,19\% \quad (5)$$

Hasil ini mengindikasikan bahwa model dapat mengidentifikasi sekitar 52% dari data uji dengan benar. Secara keseluruhan, tingkat akurasi ini dapat dikategorikan sebagai menengah dan menandakan bahwa model tersebut belum berhasil membedakan kedua kategori dengan baik.

4.4. Analisis Precision, Recall, dan F1-Score

Selain akurasi, analisis juga dilakukan dengan menggunakan metrik *precision*, *recall*, dan *F1-Score* untuk membeikan gambaran kinerja model secara lebih menyeluruh.

	precision	recall	f1-score	support
0	0.52	0.58	0.55	460
1	0.52	0.46	0.49	452
accuracy			0.52	912
macro avg	0.52	0.52	0.52	912
weighted avg	0.52	0.52	0.52	912

Gambar 6. Hasil Precision, Recall, dan F1-Score

Untuk kategori Normal (0): *Precision* = 0,52, *Recall* = 0,58, *F1-Score* = 0,55. Nilai *recall* yang lebih tinggi pada kategori normal menunjukkan bahwa model cukup efektif dalam mendeteksi mahasiswa yang progres semesternya sesuai. Untuk kelas tertinggal (1): *Precision* = 0,52, *Recall* = 0,46, *F1-Score* = 0,49.

Nilai *recall* yang lebih rendah pada kategori tertinggal menandakan bahwa model masih menghadapi tantangan dalam mengidentifikasi seluruh mahasiswa yang benar-benar mengalami keterlambatan dalam progres semester.

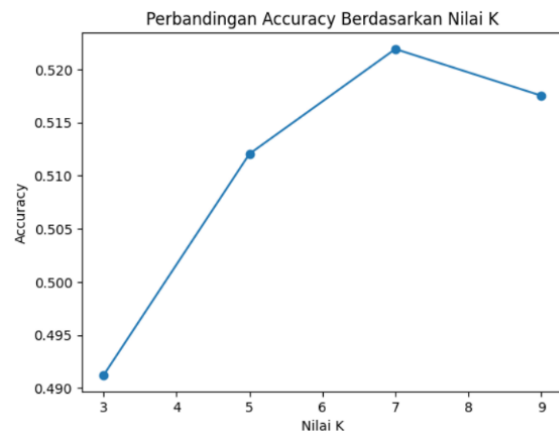
Secara keseluruhan, rata-rata performa model berada pada kisaran 0,52 yang menunjukkan keseimbangan performa masih belum ideal.

4.5. Pembahasan

Hasil penelitian menunjukkan bahwa penerapan algoritma *K-Nearest Neighbor* menggunakan nilai $K=7$ menghasilkan tingkat akurasi sebesar 52,19%. Nilai tersebut mengindikasikan bahwa kemampuan model dalam melakukan klasifikasi masih relatif terbatas, khususnya dalam membedakan antara mahasiswa dengan kategori normal dan yang tertinggal berdasarkan atribut yang digunakan.

Untuk mengetahui pengaruh parameter K terhadap performa model, dilakukan pengujian menggunakan beberapa nilai K , yaitu $K=3$, $K=5$, $K=7$, $K=9$. Hasil perbandingan akurasi berdasarkan nilai K ditunjukkan pada berikut.

Berdasarkan grafik tersebut, dapat dilihat bahwa nilai akurasi model mengalami peningkatan dari $K=3$ hingga mencapai titik tertinggi pada $K=7$, kemudian mengalami sedikit penurunan pada $K=9$. Hal ini menunjukkan bahwa pemilihan nilai K sangat berpengaruh terhadap kinerja algoritma *K-Nearest Neighbor* (K-NN).



Gambar 7. Hasil Perbandingan Accuracy Nilai K

Nilai K yang terlalu kecil cenderung membuat model lebih sensitif terhadap variasi data maupun *noise*, sedangkan nilai K yang terlalu besar dapat menyebabkan model kurang mampu menangkap pola dalam data. Oleh karena itu, nilai $K=7$ dipilih sebagai parameter yang paling optimal karena mampu menghasilkan tingkat akurasi tertinggi pada dataset yang digunakan.

Meskipun demikian, tingkat akurasi yang diperoleh masih berada pada kategori sedang. Hal ini dapat disebabkan oleh keterbatasan atribut yang digunakan dalam penelitian, yaitu hanya meliputi angkatan, periode semester, dan IPK terakhir. Atribut tersebut belum sepenuhnya menggambarkan faktor-faktor yang dapat mempengaruhi keterlambatan progres akademik mahasiswa.

Selain itu, label keterlambatan dalam penelitian ini dibentuk berdasarkan indikator progres semester sebagai pendekatan tidak langsung terhadap keterlambatan pembayaran UKT. Faktor lain seperti kondisi ekonomi mahasiswa, maupun aktivitas di luar perkuliahan juga dapat mempengaruhi keterlambatan progres studi.

5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk menganalisis progres akademik mahasiswa sebagai indikator keterlambatan pembayaran UKT dengan menggunakan algoritma *K-Nearest Neighbor* (K-NN). Berdasarkan data yang digunakan yaitu data historis mahasiswa dari angkatan 2021 dan 2022 sebanyak 4.558 data yang mencakup atribut angkatan, periode semester, dan IPK terakhir.

Berdasarkan hasil pengujian model algoritma *K-Nearest Neighbor* dengan parameter nilai $K=7$ menghasilkan tingkat akurasi sebesar 52,19% pada data uji. Hasil tersebut menunjukkan bahwa kemampuan model dalam melakukan klasifikasi berada pada kategori sedang, Khususnya dalam membedakan mahasiswa yang termasuk kedalam kategori normal dan tertinggal berdasarkan indikator progres semester sebagai acuannya.

Evaluasi menggunakan *confusion matrix* menunjukkan bahwa model mampu mengklasifikasikan

dengan benar sebanyak 269 data mahasiswa pada kategori normal dan 207 mahasiswa pada kategori tertinggal. Namun demikian, masih terdapat kesalahan klasifikasi yang cukup signifikan, khususnya dalam mengidentifikasi mahasiswa yang mengalami keterlambatan progres akademik.

Selain itu, hasil pengukuran menggunakan *precision*, *recall* dan *F1-Score* yang berada pada kisaran 0,52 menunjukkan bahwa performa model belum optimal, terutama dalam mengidentifikasi mahasiswa yang benar-benar mengalami keterlambatan. Hal ini mengindikasikan bahwa atribut yang digunakan dalam penelitian masih belum cukup representatif dalam menggambarkan kondisi sebenarnya.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma K-NN bisa digunakan sebagai pendekatan awal dalam menganalisis pola keterlambatan progres akademik mahasiswa. Namun, untuk memperoleh hasil yang lebih akurat dan optimal, diperlukan penambahan atribut yang lebih relevan serta pengembangan metode yang lebih kompleks pada penelitian selanjutnya.

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya.

Pertama, penelitian selanjutnya dapat menambahkan atribut lain yang lebih relevan dalam memprediksi keterlambatan progres akademik mahasiswa, status pekerjaan, atau riwayat pembayaran UKT secara langsung sehingga model dapat menghasilkan tingkat akurasi yang baik.

Kedua, penelitian selanjutnya dapat mempertimbangkan penggunaan algoritma klasifikasi lain seperti *Decision Tree*, *Random Forest*, maupun *Support Vector Machine* untuk melakukan perbandingan kinerja model, sehingga dapat diperoleh metode yang paling optimal dalam memprediksi keterlambatan progres mahasiswa.

Ketiga, hasil penelitian ini diharapkan dapat menjadi gambaran dalam penerapan model K-NN dan menjadi bahan pertimbangan bagi pihak universitas dalam mengembangkan sistem analisis berbasis data untuk melihat progres akademik mahasiswa secara lebih efektif.

DAFTAR PUSTAKA

- [1] T. A. Y. Siswa and R. P. Wibowo, "Komparasi Metode Seleksi Fitur Dalam Prediksi Keterlambatan Pembayaran Biaya Kuliah," *Teknika*, Vol. 12, No. 1, Pp. 73–82, Mar. 2023,
- [2] M. Ardiansyah, T. Suharto, And A. S. Farid, "Upaya Penanganan Uang Kuliah Tunggal (Ukt) Bermasalah Bagi Mahasiswa Yang Tidak Mampu Pada Perguruan Tinggi," *Jiip - Jurnal Ilmiah Ilmu Pendidikan*, Vol. 5, No. 10, Pp. 4432–4441, Oct. 2022.
- [3] S. Widaningsih and S. Yusuf, "Penerapan Data Mining Untuk Memprediksi Siswa Berprestasi Dengan Menggunakan Algoritma K Nearest Neighbor," *Jurnal Teknik Informatika Dan Sistem Informasi*, Vol. 9, No. 3, 2022.
- [4] U. Nijunnihayah, S. S. Hilabi, F. Nurapriani, And E. Novalia, "Implementasi Algoritma K-Nearest Neighbor Untuk Prediksi Penjualan Alat Kesehatan Pada Media Alkes," *Malcom: Indonesian Journal of Machine Learning and Computer Science*, Vol. 4, No. 2, Pp. 695–701, Apr. 2024.
- [5] R. Rahmadini, Enjel Erika Lorencislubis, Aji Priansyah, Yolanda R.W.N, And Tuti Meutia, "Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Mahasiswa Akuntansi Samudra*, Vol. 4, No. 4, Pp. 223–235, Aug. 2023.
- [6] M. Surur, N. A. Santoso, And B. A. Santoso, "Klasifikasi Keterlambatan Pembayaran Spp Santri Menggunakan Algoritma K-Nearest Neighbor Di Pesantren Al Fajar Tegal," *Riggs: Journal of Artificial Intelligence and Digital Business*, Vol. 4, No. 3, Pp. 873–883, Aug. 2025.
- [7] L. Kesuma, B. Andika, I. Mariami, S. Informasi, And S. Triguna Dharma, "Aplikasi Data Mining Untuk Pengelompokkan Siswa Layak Menerima Bantuan Program Indonesia Pintar Menggunakan Metode K-Nearest Neighbor," Vol. 4, Pp. 692–702, 2025.
- [8] F. M. Delta Maharani, A. Lia Hananto, S. Shofia Hilabi, F. Nur Apriani, A. Hananto, And B. Huda, "Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor," *Metik Jurnal*, Vol. 6, No. 2, Pp. 97–103, Dec. 2022.
- [9] Hera Karmila And Shindy Yuliyatini, "Systematic Literature Review: Metode Normalisasi Data Yang Efektif Pada Deteksi Penyakit Diabetes Menggunakan Machine Learning (Algoritma K-Nearest Neighbors)," *Journal Of Scientech Research And Development*, Vol. 6, No. 1, Pp. 2099–2110, Jun. 2024.
- [10] A. Prasetyo, N. Musyaffa, And R. Sastra, "Implementasi Data Mining Untuk Analisis Data Penjualan Dengan Menggunakan Algoritma Apriori (Studi Kasus Dapoerin's) Program Studi Sistem Informasi [2]," Vol. Viii, No. 2, 2020.
- [11] B. Huda *et al.*, "Analisis Sentimen E-Learning X Terhadap Antarmuka Pengguna Menggunakan Kombinasi Multinomial Naive Bayes Dan Pendekatan Design Thinking," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 4, pp. 895–902, Aug. 2024.
- [12] A. Yandi Saputra and Y. Primadasa, "Saputra, A. Y., & Primadasa, Y. Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Techno. Com*, vol. 17, no. 4, 395-403," Nov.2018

- [13] Nugroho, A., & Amrullah, A. Evaluasi kinerja algoritma k-NN menggunakan k-fold cross validation pada data debitur KSP Galih Manunggal. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, vol. 5, no .2, 294-300. May 2023.
- [14] Tukino, Gunawan, Sutarto, And D. Manongga, "Evaluation of Critical Thinking Disposition in Learning Using E-Learning," *Internal (Information System Journal)*, Vol. 5, No. 2, Pp. 187–195, 2022.
- [15] S. N. Bakri and L. S. Harahap, "Analisis Klasifikasi Algoritma K-Nearest Neighbor (K-NN) Pada Struktur Daerah Di Kota Medan," *Jurnal Ilmu Komputer Dan Sistem Informasi*, Vol. 4, No. 2, Pp. 182–193, May 2025.
- [16] I. Darmayanti, P. Subarkah, L. R. Anunggilarsa, And J. Suhaman, "Prediksi Potensi Siswa Putus Sekolah Akibat Pandemi Covid-19 Menggunakan Algoritme K-Nearest Neighbor," *Jst (Jurnal Sains Dan Teknologi)*, Vol. 10, No. 2, Pp. 230–238, Nov. 2021.
- [17] H. A. Dwi Fasnuari, H. Yuana, And M. T. Chulkamdi, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus," *Antivirus: Jurnal Ilmiah Teknik Informatika*, Vol. 16, No. 2, Pp. 133–142, Oct. 2022.
- [18] Y. Apridiansyah, Z. Padli, Y. Reswan, And H. Witriyono, "Penerapan Metode Yolov5 Untuk Klasifikasi Dan Deteksi Objek Menggunakan Video Non-Real-Time," *Jurnal Processor*, Vol. 20, No. 2, Oct. 2025.
- [19] V. Syafana, S. S. Hilabi, E. Novalia, And B. Huda, "Prediksi Angka Kelahiran Dalam Berbagai Kelompok Umur Ibu Menggunakan Metode K-Nearest Neighbor," *Malcom: Indonesian Journal of Machine Learning And Computer Science*, Vol. 4, No. 3, Pp. 1096–1103, Jul. 2024.
- [20] A. L. Hananto, A. Hananto, B. Huda, A. Y. Rahman, E. Novalia, And B. Priyatna, "Determination Of Training Participants In Community Work Training Centers Using The Naïve Bayes Classifier Algorithm," *Joiv: International Journal On Informatics Visualization*, Vol. 8, No. 3, P. 1162, Sep. 2024.
- [21] U. Nijunniyah, S. S. Hilabi, F. Nurapriani, And E. Novalia, "Implementasi Algoritma K-Nearest Neighbor Untuk Prediksi Penjualan Alat Kesehatan Pada Media Alkes," *Malcom: Indonesian Journal of Machine Learning and Computer Science*, Vol. 4, No. 2, Pp. 695–701, Apr. 2024.
- [22] Akhmad, M. R., & Siswa, T. A. Y. Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi. *Progresif J. Ilm. Komput*, Vol.18, No. 2, 185-192. Aug.2022