JTIF | Jurnal INOVATIF WIRA WACANA Vol. 04 No. 02 September 2025, hlm 20-29 **pISSN** (2961-7316)
eISSN (2962-5998)

Klasifikasi Masa Studi Mahasiswa Program Studi Teknik Informatika Menggunakan Algoritma *K-Nearest Neighbor*

(*Classification of Study Period of Informatics Engineering Study Program Students Using K-Nearest Neighbor Algorithm*)

Anjelina D. L. Nenohaifeto[1], Yustina Rada[2]

[1,2]Program Studi Teknik Informatika, Universitas Kristen Wira Wacana Sumba

E-mail: [1]anjelinanenohaifeto@gmail.com, [2]yustinarada@unkriswina.ac.id,

KEYWORDS:
Study Duration, Data Mining, K-NN, Classification, Student

ABSTRACT

The study period of students is one of the important indicators in assessing the quality of higher education. The Informatics Engineering Study Program at Wira Wacana Christian University of Sumba (UNKRISWINA Sumba) faces challenges in increasing the student acceptance period, with many senior students still having not completed their final assignments. Students have difficulty completing their studies on time due to various factors, such as lack of academic readiness, minimal guidance, and obstacles in accessing supporting resources. This study aims to analyze and manage student academic data using the K-Nearest Neighbor (K-NN) algorithm to identify patterns that have the potential to cause delays in study periods. The process is carried out through the stages of data collection, pre-processing, grouping based on study period, applying the K-NN algorithm, and evaluating model performance using accuracy, precision, recall, and F1-score metrics. The results showed that the best K value varied for each class, with the highest accuracy of 92.31% in the 2020 class when using K=3. This classification model has proven effective in predicting students' study period and can be used as a tool in detecting the risk of study delays. These results are expected to provide insight for study programs to design more targeted and data-based intervention strategies.

KATA KUNCI:
Masa Studi, Data mining, K-NN, Klasifikasi, Mahasiswa

ABSTRAK

Masa studi siswa merupakan salah satu indikator penting dalam menilai kualitas pendidikan tinggi. Program Studi Teknik Informatika di Universitas Kristen Wira Wacana Sumba (UNKRISWINA Sumba) menghadapi tantangan dalam meningkatkan waktu penerimaan mahasiswa, dengan masih banyaknya mahasiswa angkatan atas yang belum menyelesaikan tugas akhir mereka. Mahasiswa mengalami kesulitan menyelesaikan studi tepat waktu karena berbagai faktor, seperti kurangnya kesiapan akademik, minimnya bimbingan, serta kendala dalam mengakses sumber daya pendukung. Penelitian ini bertujuan untuk melakukan analisis dan pengelolaan data akademik mahasiswa dengan memanfaatkan algoritma K-Nearest Neighbor (K-NN) guna mengidentifikasi pola yang berpotensi menyebabkan keterlambatan masa studi. Proses dilakukan melalui tahapan pengumpulan data, pra-pemrosesan, pengelompokan berdasarkan masa studi, penerapan algoritma K-NN, dan evaluasi kinerja model menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa nilai K terbaik berbeda-beda untuk tiap angkatan, dengan akurasi tertinggi sebesar 92,31% pada angkatan 2020 saat menggunakan K=3. Model klasifikasi ini terbukti efektif dalam memprediksi masa studi siswa dan dapat digunakan sebagai alat bantu dalam mendeteksi risiko keterlambatan studi. Hasil ini diharapkan dapat memberikan wawasan bagi pihak program studi untuk merancang strategi intervensi yang lebih tepat sasaran dan berbasis data.

PENDAHULUAN

Pendidikan tinggi memiliki peran penting dalam membentuk kompetensi dan kesiapan individu dalam menghadapi tantangan global. Salah satu indikator utama keberhasilan pendidikan tinggi adalah ketepatan waktu mahasiswa dalam menyelesaikan studi. Ketepatan ini tidak hanya mencerminkan kinerja akademik mahasiswa secara individu, tetapi juga berkontribusi terhadap mutu program studi serta penilaian akreditasi institusi pendidikan tinggi [1]. Berdasarkan ketentuan dari Kementerian Pendidikan dan Kebudayaan Republik Indonesia dalam Permendikbud Nomor 3 Tahun 2020 tentang Standar Nasional Pendidikan Tinggi, masa studi ideal bagi mahasiswa program sarjana (Strata 1) adalah selama 4 tahun atau 8 semester, dengan batas maksimal penyelesaian studi adalah 14 semester. Ketentuan ini dirancang untuk memberikan fleksibilitas akademik tanpa mengorbankan kualitas lulusan dan tetap mendorong efisiensi penyelenggaraan pendidikan [2]. Dalam konteks ini, banyak penelitian dilakukan untuk mengklasifikasikan mahasiswa yang berpotensi lulus tepat waktu atau tidak, sebagai upaya dalam meningkatkan kualitas lulusan dan pengelolaan program studi [1][2]. Fleksibilitas ini mencerminkan pendekatan pendidikan tinggi yang lebih inklusif dan humanistik karena menyadari bahwa setiap mahasiswa memiliki latar belakang, kemampuan belajar, dan tantangan yang berbeda [3].

Di Program Studi Teknik Informatika UNKRISWINA Sumba, ditemukan adanya peningkatan jumlah mahasiswa yang belum menyelesaikan studi tepat waktu, khususnya pada angkatan 2018 hingga 2020. Kondisi ini diperburuk oleh ketimpangan rasio dosen dan mahasiswa, yang berdampak pada efektivitas layanan akademik dan bimbingan. Masalah ini menunjukkan perlunya pendekatan berbasis data untuk memetakan dan memahami pola penyelesaian studi mahasiswa secara objektif.

Salah satu metode yang dapat digunakan adalah data mining, yaitu proses penggalian informasi bermakna dari data besar dengan memanfaatkan teknik statistik, kecerdasan buatan, dan pembelajaran mesin [4]. Dalam konteks pendidikan tinggi, data mining berguna untuk menganalisis data akademik mahasiswa, seperti untuk memprediksi masa studi berdasarkan variabel akademik. Menurut Han dan Kamber (2012), proses data mining mengikuti tahapan KDD (Knowledge Discovery in Databases), yang mencakup pembersihan data, integrasi, seleksi, transformasi, penambangan pola, evaluasi, dan penyajian hasil. Salah satu teknik dalam data mining yang relevan untuk prediksi masa studi adalah klasifikasi, yang bertujuan membangun model untuk mengelompokkan data ke dalam kelas tertentu berdasarkan atribut historis[5].

Dalam penelitian ini digunakan algoritma K-Nearest Neighbor (K-NN), yaitu metode klasifikasi berbasis jarak yang mengelompokkan data berdasarkan kemiripannya dengan data latih. K-NN menghitung jarak antara data uji dan data latih menggunakan rumus Euclidean, lalu menentukan kelas berdasarkan mayoritas tetangga terdekat[6]. Kelebihan algoritma ini terletak pada kesederhanaannya dan tidak memerlukan asumsi distribusi data, meskipun tetap sensitif terhadap nilai K dan skala data, sehingga diperlukan proses pre-processing seperti normalisasi. Tahap pre-processing bertujuan meningkatkan kualitas data dengan membersihkan nilai yang tidak valid, mendeteksi outlier, serta menyamakan skala fitur agar hasil klasifikasi lebih akurat [6][7].

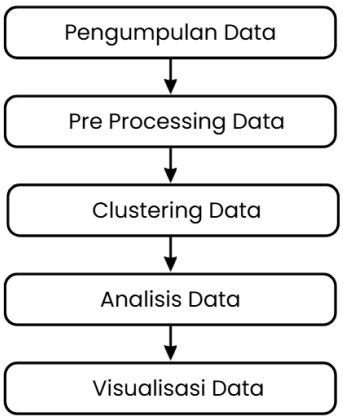
Selain itu, untuk mendukung akurasi klasifikasi, teknik clustering juga digunakan sebagai metode unsupervised learning yang mengelompokkan mahasiswa berdasarkan karakteristik akademik mereka. Clustering membantu memahami struktur data dan mengidentifikasi kelompok mahasiswa yang berpotensi mengalami keterlambatan studi atau risiko dropout [8].

Penelitian ini menggunakan Python dan Google Colab sebagai alat bantu pengembangan model. Python dipilih karena mendukung berbagai pustaka analisis data seperti SciPy, sedangkan Google Colab

memungkinkan eksperimen berbasis cloud dengan akses GPU/TPU tanpa instalasi lokal[9][10]. Kombinasi ini memberikan fleksibilitas tinggi dalam mengolah dan menguji model klasifikasi K-NN.

Dengan demikian, tujuan utama dari penelitian ini adalah untuk mengklasifikasikan masa studi mahasiswa berdasarkan variabel IPK, jumlah SKS, mata kuliah yang diulang, dan masa studi. Hasil klasifikasi diharapkan dapat membantu pihak program studi dalam mengevaluasi dan merancang intervensi akademik yang lebih efektif, guna mendorong ketepatan waktu kelulusan dan meningkatkan mutu pendidikan di lingkungan Program Studi Teknik Informatika UNKRISWINA Sumba.

METODE PENELITIAN



Gambar 1. Alur Penelitian

Alur penelitian ini mencakup serangkaian tahapan yang dimulai dari pengumpulan data, *pre-processing*, *Clustering,* analisis, hingga penyajian hasil. Pengumpulan data dilakukan dengan mengambil data akademik mahasiswa dari Sistem Informasi Akademik UNKRISWINA Sumba. Data yang dikumpulkan meliputi IPK, jumlah SKS yang telah ditempuh, jumlah mata kuliah yang diulang, serta status kelulusan mahasiswa yang diklasifikasikan ke dalam dua kategori, yaitu lulus ≤ 4 tahun (tepat waktu) dan lulus > 4 tahun (terlambat). Untuk memudahkan proses pengolahan lebih lanjut, data ini kemudian dikonversi ke dalam format CSV.

Setelah pengumpulan data, tahap selanjutnya adalah *pre-processing data*. Langkah ini mencakup proses pembersihan (*data cleaning*) untuk menghapus data yang tidak lengkap atau tidak valid, transformasi data agar sesuai untuk analisis, seleksi fitur (*feature selection)* terhadap variabel yang paling berpengaruh, normalisasi data untuk menyamakan skala IPK dan SKS, serta pengkodean kategori status kelulusan ke dalam bentuk numerik (0 untuk ≤ 4 tahun dan 1 untuk > 4 tahun).

Tahap berikutnya adalah pengelompokan data, di mana mahasiswa dikelompokkan ke dalam dua kategori berdasarkan masa studi mereka. Mahasiswa yang menyelesaikan studi dalam waktu ≤ 4 tahun dimasukkan dalam kategori "Lulus Tepat Waktu", sedangkan mahasiswa yang menyelesaikan studi lebih dari 4 tahun dimasukkan ke dalam kategori "Lulus Terlambat".

Setelah data siap, dilakukan analisis menggunakan algoritma K-Nearest Neighbor (K-NN) untuk memprediksi masa studi mahasiswa baru berdasarkan kemiripan data akademik dengan mahasiswa yang telah lulus. Proses analisis dilakukan melalui beberapa tahapan, yaitu:

1. Pemisahan Data: Dataset dibagi menjadi data latih (75%) dan data uji (25%) agar model dapat belajar dan diuji secara seimbang.
2. Menentukan Nilai K: Menentukan jumlah tetangga terdekat yang akan digunakan dalam proses klasifikasi. Nilai K yang optimal dipilih berdasarkan performa terbaik dari evaluasi model.
3. Proses Klasifikasi: Proses ini dilakukan dengan menghitung jarak antar data menggunakan metode Euclidean Distance. Perhitungan jarak bertujuan untuk menentukan tingkat kedekatan (kemiripan) antara data uji dan data latih berdasarkan atribut IPK, jumlah SKS yang telah ditempuh, dan jumlah mata kuliah yang diulang. Rumus *Euclidean Distance* yang digunakan dalam penelitian ini adalah sebagai berikut:

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

Persamaan 1. Rumus Euclidean Distance

4. Penentuan Kelas: Kelas dari mahasiswa pada data uji ditentukan berdasarkan mayoritas dari K tetangga terdekat.
5. Evaluasi Model: Untuk mengukur kinerja model klasifikasi, digunakan metrik akurasi, precision, recall, dan F1-score [6].

Hasil dari klasifikasi ini kemudian disajikan dalam berbagai bentuk visualisasi. Grafik batang digunakan untuk menunjukkan distribusi kelulusan mahasiswa, scatter plot untuk melihat hubungan antar variabel, heatmap untuk mengidentifikasi korelasi antar atribut, serta confusion matrix untuk mengevaluasi performa model K-NN. Visualisasi ini bertujuan untuk memberikan gambaran yang lebih informatif dan mendukung pengambilan keputusan dalam meningkatkan ketepatan waktu kelulusan mahasiswa di UNKRISWINA Sumba.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Dalam penelitian ini, data yang digunakan berasal dari data akademik mahasiswa Program Studi Teknik Informatika Universitas Kristen Wira Wacana Sumba (UNKRISWINA Sumba) dari tiga angkatan, yaitu angkatan 2018, 2019, dan 2020. Jumlah total data yang dikumpulkan adalah sebanyak 225 mahasiswa, dengan rincian sebagai berikut: 74 mahasiswa dari angkatan 2018, 69 mahasiswa dari angkatan 2019, dan 82 mahasiswa dari angkatan 2020. Data yang dikumpulkan mencakup beberapa atribut penting, yaitu Indeks Prestasi Kumulatif (IPK), jumlah Satuan Kredit Semester (SKS) yang telah ditempuh, jumlah mata kuliah yang diulang, serta status kelulusan mahasiswa yang diklasifikasikan ke dalam dua kategori, yakni lulus tepat waktu (≤ 4 tahun) dan lulus terlambat (> 4 tahun).

Pre Processing Data

Data melalui proses pre-processing sebagai berikut:

1. Pembersihan nilai yang tidak valid pada kolom IPK, SKS, dan MK_Ulang.
2. Konversi masa studi menjadi label klasifikasi.
3. Normalisasi fitur numerik menggunakan *Min Max Scaler* agar semua fitur berada dalam skala 0-1.
4. Penghapusan nilai kosong (NaN) yang muncul akibat error parsing atau input tidak valid.

Clustering Data

Clustering dilakukan berdasarkan label klasifikasi, yaitu sebagai berikut :

1. *Cluster* 0: Mahasiswa lulus ≤ 4 tahun
2. *Cluster* 1: Mahasiswa lulus > 4 tahun

Visualisasi scatter plot berdasarkan IPK dan SKS digunakan untuk melihat distribusi masing-masing *Cluster*.

Analisis Data

Pada tahap ini dilakukan penerapan algoritma K-NN untuk melakukan klasifikasi masa studi mahasiswa. Proses analisis ini dibagi dalam beberapa tahapan:

1. Pemisahan data

   Data yang telah dipreprocessing dibagi menjadi dua bagian, yaitu:

   a. Data latih: sebanyak 75% dari total data
   b. Data uji: sebanyak 25% dari total data

   Pemisahan ini bertujuan agar model dapat belajar dari data historis dan diuji akurasinya terhadap data baru. Proses ini dilakukan menggunakan fungsi train_test_split dari library sklearn.model_selection. Sehingga didapatkan pembagian data latih dan data uji, sebagai berikut:

Tabel 1. Pemisahan data uji dan data latih

Angkatan	Data Latih 75%	Data Uji 25%
2018	55 Mahasiswa	19 Mahasiswa
2019	51 Mahasiswa	18 Mahasiswa
2020	61 Mahasiswa	21 Mahasiswa

2. Menentukan nilai K

   Nilai K menentukan berapa jumlah tetangga terdekat yang digunakan dalam klasifikasi. Penentuan K dilakukan dengan mencoba beberapa nilai dan melihat hasil akurasinya.

Gambar 2. Perbadingan Akurasi K-NN Berdasarkan Nilai K

Tabel 2. Penentuan Nilai K

Angkatan	Nilai K Terbaik	Akurasi Tinggi
2018	1	0.8947
2019	9	0.6111
2020	2	0.9048

Berdasarkan hasil pengujian algoritma K-NN terhadap data mahasiswa angkatan 2018, 2019, dan 2020, diperoleh nilai K terbaik yang berbeda-beda pada masing-masing angkatan. Nilai K terbaik ditentukan berdasarkan akurasi tertinggi yang dicapai dari proses klasifikasi terhadap data uji.

Angkatan 2018 memiliki nilai K terbaik pada K=1 dengan akurasi sebesar 0.8947. Ini menunjukkan bahwa menggunakan satu tetangga terdekat menghasilkan prediksi yang paling akurat untuk angkatan ini. Hal ini dapat disebabkan oleh data yang memiliki jarak antar kelas yang cukup tegas, sehingga satu tetangga sudah cukup mewakili label kelas yang benar. Precision dan recall yang tinggi menunjukkan model bekerja baik dalam mengenali kedua kelas.

Angkatan 2019 menunjukkan nilai K terbaik pada K=9 dengan akurasi 0.6111. Meskipun akurasi tidak terlalu tinggi, pemilihan sembilan tetangga terdekat memberikan performa terbaik dibanding nilai K lainnya. Hal ini mengindikasikan bahwa pola distribusi data angkatan ini mungkin kurang konsisten, sehingga membutuhkan lebih banyak tetangga untuk menyeimbangkan prediksi, meskipun trade-off-nya adalah menurunnya recall untuk salah satu kelas.

Angkatan 2020 memperoleh nilai K terbaik pada K=2 dengan akurasi tertinggi mencapai 0.9048. Ini menunjukkan bahwa dengan mempertimbangkan dua tetangga terdekat, model dapat mengklasifikasikan masa studi mahasiswa dengan sangat baik. Hasil confusion matrix juga memperlihatkan performa yang seimbang, dengan recall tinggi pada kelas utama.

Perbedaan nilai K terbaik antar angkatan menunjukkan bahwa karakteristik distribusi data mahasiswa dari tiap angkatan memiliki pola yang berbeda. Oleh karena itu, dalam penerapan KNN, pemilihan nilai K perlu disesuaikan dengan kondisi data masing-masing kelompok untuk memperoleh performa klasifikasi yang optimal.

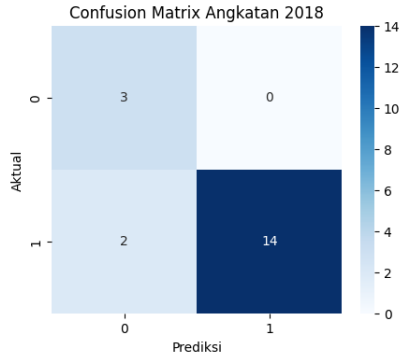
3. Perhitungan jarak dan penentuan kelas

Perhitungan jarak dan penentuan kelas dalam algoritma K-NN dilakukan secara otomatis saat proses pemodelan dan prediksi. Pada tahap ini, model KNN dibentuk dengan menggunakan fungsi KNeighborsClassifier(n_neighbors=k) dari pustaka scikit-learn, di mana parameter k menyatakan jumlah tetangga terdekat yang akan dipertimbangkan dalam proses klasifikasi. Model kemudian dilatih menggunakan data latih (X_train, y_train) melalui perintah model.fit(), dan prediksi dilakukan terhadap data uji (X_test) dengan perintah model.predict().

4. Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan masa studi mahasiswa berdasarkan IPK, jumlah SKS, dan jumlah mata kuliah yang diulang. Evaluasi menggunakan confusion matrix dan classification report, serta dilakukan secara terpisah untuk setiap angkatan guna memahami karakteristik prediksi masing-masing kelompok data.

a. Angkatan 2018



Gambar 3. Confusion Matrix Angkatan 2018

Classification Report:

	precision	recall	f1-score	support
0	0.60	1.00	0.75	3
1	1.00	0.88	0.93	16
accuracy			0.89	19
macro avg	0.80	0.94	0.84	19
weighted avg	0.94	0.89	0.90	19

Gambar 4. Classification report Angkatan 2018

Berdasarkan confusion matrix hasil klasifikasi model K-NN terhadap data mahasiswa angkatan 2018, dari total 19 data uji, sebanyak 3 data dari kelas 0 berhasil diprediksi dengan benar (true negative), dan tidak ada data kelas 0 yang salah diklasifikasikan sebagai kelas 1 (false positive). Sementara itu, untuk kelas 1, sebanyak 14 data berhasil diklasifikasikan dengan benar (true positive), dan hanya 2 data yang salah diklasifikasikan sebagai kelas 0 (false negative).

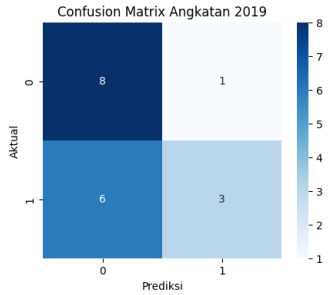
Hal ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali data dari kelas 1, dan juga cukup akurat dalam mengidentifikasi kelas 0. Tidak adanya false positive memperlihatkan bahwa model tidak salah mengklasifikasikan data dari kelas 0 sebagai kelas 1.

Hasil evaluasi ini tercermin dalam metrik classification report, di mana precision untuk kelas 0 mencapai 0.60 dan recall-nya 1.00. Artinya, seluruh data yang diprediksi sebagai kelas 0 memang benar berasal dari kelas tersebut, meskipun masih ada beberapa data kelas 0 yang tidak berhasil teridentifikasi. Untuk kelas 1, precision mencapai 1.00 dan recall sebesar 0.88, menunjukkan bahwa model sangat konsisten dan akurat dalam mengenali data kelas 1.

Akurasi keseluruhan model adalah 0.89 (89%), yang berarti 17 dari 19 data uji berhasil diprediksi dengan tepat. Nilai f1-score untuk kelas 0 adalah 0.75, dan untuk kelas 1 mencapai 0.93. Rata-rata makro untuk precision, recall, dan f1-score masing-masing adalah 0.80, 0.94, dan 0.84. Sementara itu, nilai rata-rata berbobot (weighted average) adalah 0.94, 0.89, dan 0.90.

Secara umum, hasil ini menunjukkan bahwa model K-NN bekerja sangat baik untuk data angkatan 2018, dengan performa yang seimbang dan kecenderungan kuat dalam mengenali data dari kedua kelas dengan tingkat kesalahan yang rendah.

b. Angkatan 2019



Gambar 5. Confusion Matrix Angkatan 2019

Classification Report:

	precision	recall	f1-score	support
0	0.57	0.89	0.70	9
1	0.75	0.33	0.46	9
accuracy			0.61	18
macro avg	0.66	0.61	0.58	18
weighted avg	0.66	0.61	0.58	18

Gambar 6. Classification Report Angkatan 2019

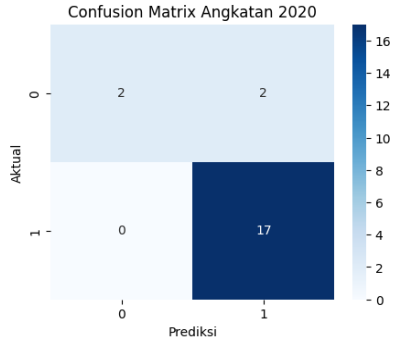
Berdasarkan hasil evaluasi model K-NN terhadap data mahasiswa angkatan 2019, diperoleh akurasi sebesar 0.6111 atau sekitar 61%, yang menunjukkan bahwa 11 dari 18 data uji berhasil diklasifikasikan dengan benar. Confusion matrix menunjukkan bahwa dari 9 data aktual kelas 0, sebanyak 8 data berhasil diprediksi dengan benar sebagai kelas 0 (true negative), dan hanya 1 data yang salah diklasifikasikan sebagai kelas 1 (false positive). Sementara itu, dari 9 data kelas 1, hanya 3 yang berhasil diklasifikasikan dengan benar (true positive), dan 6 data lainnya salah diklasifikasikan sebagai kelas 0 (false negative).

Nilai precision untuk kelas 0 adalah 0.57, artinya sekitar 57% dari prediksi kelas 0 memang benar berasal dari kelas tersebut. Sedangkan precision untuk kelas 1 adalah 0.75, yang menunjukkan bahwa tiga data yang diprediksi sebagai kelas 1 seluruhnya benar. Untuk metrik recall, kelas 0 memiliki nilai sebesar 0.89 karena hampir semua data aktual kelas 0 berhasil dikenali dengan benar, sedangkan kelas 1 memiliki recall sebesar 0.33, menunjukkan bahwa hanya sebagian kecil data kelas 1 yang berhasil diidentifikasi oleh model.

F1-score sebagai gabungan antara precision dan recall memberikan nilai 0.70 untuk kelas 0 dan 0.46 untuk kelas 1. Secara keseluruhan, nilai rata-rata makro untuk precision, recall, dan f1-score masing-masing adalah 0.66, 0.61, dan 0.58. Sedangkan nilai rata-rata berbobot (weighted average) adalah 0.66, 0.61, dan 0.58.

Hasil ini menunjukkan bahwa performa model terhadap data angkatan 2019 masih belum optimal. Model cenderung lebih andal dalam mengenali data kelas 0, tetapi kurang efektif dalam mengklasifikasikan data kelas 1. Ketidakseimbangan ini bisa disebabkan oleh distribusi data atau pola yang kurang konsisten pada kelas 1, sehingga diperlukan pendekatan tambahan atau penyesuaian model untuk meningkatkan performa klasifikasi secara keseluruhan.

c. Angkatan 2020



Gambar 7. Confusion Matrix Angkatan 2020

Classification Report:	precision	recall	f1-score	support
0	1.00	0.50	0.67	4
1	0.89	1.00	0.94	17
accuracy			0.90	21
macro avg	0.95	0.75	0.81	21
weighted avg	0.91	0.90	0.89	21

Gambar 8. Classification Report Angkatan 2020

Berdasarkan hasil evaluasi model K-NN terhadap data mahasiswa angkatan 2020, diperoleh akurasi sebesar 0.9048 atau sekitar 90%, yang berarti bahwa 19 dari 21 data uji berhasil diklasifikasikan dengan benar oleh model. Confusion matrix menunjukkan bahwa dari 4 data aktual kelas 0, sebanyak 2 data berhasil diprediksi dengan benar (true negative), sementara 2 lainnya salah diklasifikasikan sebagai kelas 1 (false positive). Untuk kelas 1, seluruh 17 data berhasil diprediksi dengan tepat (true positive), tanpa adanya kesalahan klasifikasi.

Precision untuk kelas 0 mencapai 1.00, yang berarti semua prediksi terhadap kelas 0 memang benar berasal dari kelas tersebut. Namun, recall-nya hanya 0.50, karena hanya separuh dari total data aktual kelas 0 yang berhasil dikenali oleh model. Sebaliknya, untuk kelas 1, precision sebesar 0.89 dan recall mencapai 1.00 menunjukkan bahwa sebagian besar prediksi kelas 1 akurat, dan tidak ada data aktual kelas 1 yang terlewat.

F1-score untuk kelas 0 adalah 0.67, sedangkan untuk kelas 1 mencapai 0.94, menandakan keseimbangan yang sangat baik antara precision dan recall pada kelas mayoritas. Nilai rata-rata makro untuk precision, recall, dan f1-score masing-masing adalah 0.95, 0.75, dan 0.81. Sedangkan nilai rata-rata berbobot (weighted average) adalah 0.91, 0.90, dan 0.89.

Hasil ini menunjukkan bahwa model memiliki kinerja yang sangat baik dalam mengenali dan mengklasifikasikan data kelas mayoritas (kelas 1), namun performa terhadap kelas minoritas (kelas 0) masih dapat ditingkatkan, khususnya dalam hal recall. Perlu strategi tambahan seperti penyeimbangan data atau penyesuaian parameter untuk meningkatkan sensitivitas terhadap kelas yang jumlahnya lebih sedikit.

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian mengenai klasifikasi masa studi mahasiswa Program Studi Teknik Informatika menggunakan algoritma K-NN, dapat disimpulkan bahwa algoritma ini berhasil diterapkan untuk memprediksi masa studi mahasiswa (≤ 4 tahun atau > 4 tahun) berdasarkan variabel IPK, jumlah SKS yang

telah ditempuh, dan jumlah mata kuliah yang diulang. Proses klasifikasi mencakup tahapan preprocessing, normalisasi, pembagian data, pengujian nilai K, hingga evaluasi model. Hasil menunjukkan bahwa nilai K terbaik bervariasi untuk tiap angkatan: K=1 untuk angkatan 2018 dengan akurasi 69,23%, K=3 untuk angkatan 2019 dengan akurasi 75%, dan K=3 untuk angkatan 2020 dengan akurasi tertinggi sebesar 92,31%. Evaluasi menggunakan confusion matrix dan classification report menunjukkan bahwa K-NN cukup efektif, terutama untuk data angkatan 2020 yang memiliki pola lebih konsisten. Hasil klasifikasi ini berpotensi dimanfaatkan oleh program studi untuk mendeteksi mahasiswa yang berisiko terlambat lulus, sehingga dapat diberikan pendampingan akademik sejak dini.

Penelitian selanjutnya disarankan untuk menambahkan variabel lain seperti keaktifan dalam kegiatan akademik/non-akademik, nilai mata kuliah inti, atau tingkat kehadiran guna meningkatkan akurasi model. Selain itu, pengembangan sistem klasifikasi berbasis machine learning dalam bentuk aplikasi atau sistem informasi yang terintegrasi dengan sistem akademik kampus dapat menjadi solusi praktis dalam pemantauan mahasiswa. Penggunaan algoritma pembanding seperti Decision Tree, Random Forest, atau Support Vector Machine juga disarankan untuk mengevaluasi dan menentukan metode klasifikasi yang paling optimal.

DAFTAR PUSTAKA

[1] Inna AN. 2019 “Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*. 10(2): 2252-4983

[2] W. Agwil, H. Fransiska, and N. Hidayati, “Analisis Ketepatan Waktu Lulus Mahasiswa Dengan Menggunakan Bagging Cart,” *FIBONACCI J. Pendidik. Mat. dan Mat.*, vol. 6, no. 2, p. 155, 2020, doi: 10.24853/fbc.6.2.155-166.

[3] L. M. Huizen, A. Hendrawan, and A. P. R. Pinem, “Analisis Faktor yang Mempengaruhi Lama Studi Mahasiswa dengan Metode SMARTER,” *Aiti*, vol. 19, no. 2, pp. 213–227, 2022, doi: 10.24246/aiti.v19i2.213-227.

[4] Ni PRPI, Igusti AMS, Made S. 2024. Faktor Yang Mempengaruhi Lama Studi Mahasiswa (Studi Kasus: Mahasiswa Program Sarjana Fakultas Mipa Universitas Udayana). *E-Jurnal Matematika.* 13(3): 187–194.

[5] FRJ. Lingling, Siska R, Silvia R. 2025. Pemodelan Lama Masa Studi Mahasiswa dengan Regresi Cox Proportional Hazard (Modeling The Length Of a Student’s Study Period Using Cox Proportional Hazards Regression). *Jurnal Edukasi Matematika*. 1(2): 81–93.

[6] Nursetia W. 2021. Prediksi Kelulusan Mahasiswa Menggunakan K-Nearest Neighbor Berbasis Particle Swarm Optimization. *Jurnal Teknologi Informasi Indonesia*. 6(2): 118–127.

[7] S. Andriani, A. Nazir, R. M. Candra, F. Syafria, and I. Afrianty, “Implementasi Algoritma K-Nearest Neighbor Untuk Menentukan Klasifikasi Kelulusan Mahasiswa Teknik Informatika,” *J. Comput. Syst. Informatics ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak) Vol. 4, No. 4, August 2023, Page 922-930*, vol. 13, no. 1, pp. 104–116, 2023.

[8] N. L. Ratniasih, “Penerapan Algoritma K-Nearest Neighbour (K-Nn) Untuk Penentuan Mahasiswa Berpotensi Drop Out,” *J. Teknol. Inf. dan Komput.*, vol. 5, no. 3, pp. 314–318, 2019, doi: 10.36002/jutik.v5i3.804.

[9] M. R. S. Alfarizi, M. Z. Al-farish, M. Taufiqurrahman, G. Ardiansah, and M. Elgar, “Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning,” *Karya Ilm. Mhs. Bertauhid (KARIMAH TAUHID)*, vol. 2, no. 1, pp. 1–6, 2023.

[10] T. Carneiro, R. V. M. Da Nobrega, T. Nepomuceno, G. Bin Bian, V. H. C. De Albuquerque, and P. P. R. Filho, “Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications,” *IEEE Access*, vol. 6, pp. 61677–61685, 2018, doi: 10.1109/ACCESS.2018.2874767.