

Implementasi Machine Learning Metode K-Nearest Neighbor Untuk Memprediksi Status Kelulusan Mahasiswa Program Studi Informatika Universitas Dehasen Bengkulu

¹Al Kausar Ah, ²Devi Sartika, ³Ockhy Jey Fhiter

¹ Mahasiswa, Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Dehasen Bengkulu
Alamat : alkausarstack@gmail.com

^{2,3} Dosen Tetap, Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Dehasen Bengkulu
Jl. Meranti Raya No. 32 Kota Bengkulu 38228 Telp. (0736) 22027, 26957 Fax. (0736) 341139;
e-mail: devisartika@unived.ac.id Ockhy@unived.ac.id

Received: Juni 2026, Revised: Agustus 2026, Accepted: Oktober 2026)

Abstract-Student graduation rates are a key performance indicator that's really crucial for the quality and reputation of a university. However, study delays and the risk of dropping out are still serious challenges for the Informatics Study Program at Dehasen University Bengkulu. This research focuses on tackling these issues by developing an early prediction system for student graduation status to help support more proactive academic decision-making. This research uses an Educational Data Mining (EDM) approach with the Knowledge Discovery in Databases (KDD) framework to process student academic datasets. It implements the K-Nearest Neighbor (KNN) algorithm. Based on the testing results with the hyperparameter $K=5$, the KNN model achieved an accuracy rate of 50%, which indicates that the decision boundary between students who pass and those who don't cannot be significantly separated by the numeric academic variables currently used. This finding suggests that further development is needed by adding latent variables or sociological data, as well as more specific feature selection techniques to better capture student behavior patterns and improve the model's performance in the future.

Keywords: Data Mining, K-Nearest Neighbor, Graduation Prediction, Machine Learning

Intisari-Tingkat kelulusan mahasiswa merupakan indikator kinerja utama yang sangat krusial bagi kualitas dan reputasi perguruan tinggi, namun keterlambatan studi dan risiko putus kuliah masih menjadi tantangan serius bagi Program Studi Informatika Universitas Dehasen Bengkulu. Penelitian ini difokuskan untuk mengatasi permasalahan tersebut dengan mengembangkan sistem prediksi status kelulusan mahasiswa secara dini guna mendukung pengambilan kebijakan akademik yang lebih proaktif. Penelitian ini menggunakan pendekatan Educational Data Mining (EDM) dengan kerangka kerja Knowledge Discovery in Databases (KDD) untuk memproses dataset akademik mahasiswa. Implementasi algoritma K-Nearest Neighbor (KNN). Berdasarkan hasil pengujian dengan hiperparameter $K=5$, model KNN menghasilkan tingkat akurasi sebesar 50%, yang mengindikasikan bahwa batas keputusan (decision boundary) antara mahasiswa lulus dan tidak lulus belum dapat dipisahkan secara signifikan oleh variabel akademik numerik yang digunakan saat ini. Temuan ini menyimpulkan bahwa diperlukan pengembangan lebih lanjut dengan menambahkan variabel laten atau data sosiologis, serta teknik seleksi fitur yang lebih spesifik untuk menangkap pola perilaku mahasiswa secara lebih akurat dan meningkatkan performa model di masa mendatang

Kata Kunci: Data Mining, K-Nearest Neighbor, Prediksi Kelulusan, Machine Learning

I.PENDAHULUAN

Perkembangan teknologi informasi telah membawa transformasi signifikan dalam pengelolaan data di sektor pendidikan. Institusi pendidikan tinggi dituntut untuk terus meningkatkan kualitas tata kelola dan hasil akademik guna menghasilkan sumber daya manusia yang unggul dan kompetitif. Salah satu indikator performa utama dalam penilaian kualitas sebuah perguruan tinggi maupun program studi adalah tingkat kelulusan mahasiswa yang tepat waktu. Keberhasilan mahasiswa dalam menyelesaikan studi tidak hanya memberikan dampak positif bagi individu, tetapi juga berpengaruh signifikan terhadap akreditasi dan reputasi lembaga di mata masyarakat (Moonallika et al., 2020). Integrasi sistem informasi akademik menghasilkan jumlah data yang besar atau *big data*, yang jika dikelola dengan teknik Data Mining dan *Machine Learning* dapat memberikan pengetahuan yang bermanfaat bagi institusi. Namun, pada kenyataannya, fenomena mahasiswa yang mengalami keterlambatan lulus hingga risiko putus studi (*drop out*) masih menjadi tantangan serius bagi banyak universitas. Tingginya angka kegagalan ini mencerminkan perlunya sistem evaluasi dini untuk mendeteksi potensi hambatan akademik sejak awal masa perkuliahan. Untuk menjawab tantangan tersebut, pemanfaatan *Educational Data Mining* (EDM) kini menjadi solusi krusial dalam mengekstraksi pola-pola tersembunyi dari data historis mahasiswa untuk menghasilkan prediksi performa yang akurat (Tumbilung & Efendi, 2025). Berdasarkan dari penelitian terdahulu menunjukkan terdapat tiga algoritma klasifikasi yang paling dominan digunakan dalam domain pendidikan, yaitu Naive Bayes (NB), K-Nearest Neighbor (KNN), dan *Decision Tree* (DT). Hasil kajian menunjukkan bahwa algoritma KNN dan *Decision Tree* cenderung memberikan tingkat akurasi yang tinggi ketika menggunakan atribut akademik seperti Indeks Prestasi Kumulatif (IPK), nilai ujian, dan tingkat kehadiran. Sementara itu, Naive Bayes sering dipilih karena efisiensi komputasinya, terutama pada dataset dengan karakteristik kategorikal (Agustina et al., 2025). Penelitian komparatif menunjukkan bahwa efektivitas algoritma ini bervariasi tergantung pada karakteristik data. Sebagai contoh, penggunaan *Decision Tree* mampu mencapai akurasi hingga 100% dalam memprediksi risiko *drop out*, sedangkan algoritma KNN

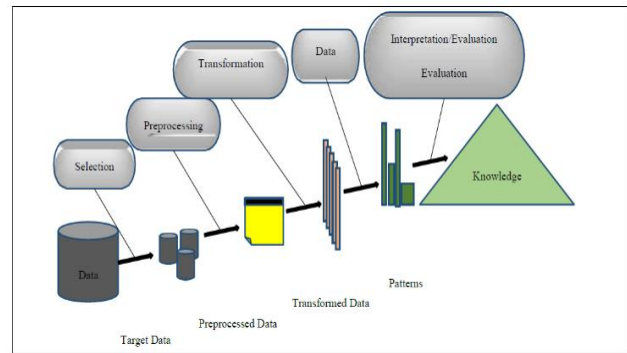
mencatatkan akurasi sebesar 93,06% dalam mengklasifikasikan status kelulusan mahasiswa Teknik Informatika. Di sisi lain, metode *Naive Bayes* terbukti efektif dalam memprediksi ketepatan waktu lulus dengan akurasi 80% pada studi kasus kampus baru yang jumlah lulusannya masih terbatas (Tumbilung & Efendi, 2025). Selain pada level kelulusan program studi, prediksi juga dilakukan secara lebih spesifik pada tingkat mata kuliah kunci, seperti Sistem Manajemen Basis Data. Implementasi KNN pada level ini mampu memberikan gambaran awal mengenai mahasiswa yang berpotensi tidak lulus dengan akurasi mencapai 90%. Informasi ini sangat penting bagi dosen pengampu untuk melakukan langkah antisipatif atau perbaikan pola pengajaran agar target capaian pembelajaran dapat tuntas tercapai (Ahmad et al., 2024). Meskipun model prediktif telah banyak dikembangkan, performa algoritma sangat dipengaruhi oleh kualitas praproses data dan pemilihan atribut yang relevan. Sebagian besar penelitian terdahulu masih memiliki keterbatasan pada volume dataset yang kecil atau fokus pada variabel akademik standar tanpa melibatkan faktor eksternal secara luas. Oleh karena itu, penelitian ini bertujuan untuk mengoptimalkan penggunaan algoritma K-Nearest Neighbor dengan mengintegrasikan atribut yang lebih komprehensif guna meningkatkan akurasi prediksi di Progran Studi Informatika Universitas Dehasen Bengkulu.

TINJAUAN PUSTAKA

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Wulandari et al., 2024). Dalam konteks pendidikan tinggi, data mining berperan penting dalam mengolah data akademik mahasiswa yang relevan bagi pengelola institusi (Syam et al., 2024). Data mining merupakan pendekatan sistematis yang memanfaatkan algoritma komputasi untuk mengolah data berskala besar sehingga menghasilkan informasi strategis yang mendukung perencanaan dan evaluasi kebijakan (Rahmat et al. 2021).

III. METODOLOGI PENELITIAN

Metode penelitian yang diadopsi adalah kerangka kerja *Knowledge Discovery in Databases* (KDD). KDD dipilih karena merepresentasikan alur ilmiah yang komprehensif dalam menambang data, mulai dari identifikasi masalah hingga presentasi pengetahuan. Tahapan komputasi utama dalam kerangka kerja ini difokuskan pada penerapan algoritma *Machine Learning*, secara spesifik menggunakan metode klasifikasi *K-Nearest Neighbor* (KNN) untuk membentuk model prediktif.



Gambar 1 kerangka kerja Knowledge Discovery in Databases (KDD)

Keterangan:

1. *Data Selection* (Pemilihan Data)
Dataset yang digunakan adalah dataset yang akademik mahasiswa program studi Informatika Fakultas Ilmu Komputer Universitas Dehasen Bengkulu.
2. Pembersihan data dan proses data (*cleaning and processing*)
Proses ini digunakan buat melakukan pembersihan data seperti miss value, duplikat data dan lain-lain
3. Transformasi data (*transformation*)
4. Proses *transformation* diperlukan untuk mendapatkan hasil yang lebih akurat dan mengurangi waktu komputasi terutama dalam masalah dengan skala besar.
5. Data mining (modelling data)
6. Data-data yang telah diseleksi dan ditransformasi ditambang dengan berbagai teknik kemudian dapat dilakukan proses modeling data dengan algoritma KNN. Pada Proses data mining diperlukan untuk mencari pola atau informasi menarik dalam data yang telah dipilih dengan melakukan training dan testing..
7. Evaluasi pola dan presentasi pengetahuan (*knowledge extraction*)

Tahap ini merupakan pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya, penggambaran visualisasi dan penyajian pengetahuan yang digunakan untuk memperoleh pengetahuan dengan merepresentasikan pengetahuan dalam bentuk yang mudah dipahami oleh pengguna

IV. HASIL DAN PEMBAHASAN

Hasil dan Implementasi

Implementasi dari model *Machine Learning* dengan metode *K-Nearest Neighbor* (KNN) dalam memprediksi status kelulusan mahasiswa diproses di lingkungan *Google Colab* untuk memastikan integritas komputasi dan reproduksibilitas hasil. Pembahasan difokuskan pada bagaimana data akademik diproses menjadi informasi yang bermakna melalui algoritma KNN, serta bagaimana parameter-parameter model dioptimalkan untuk mendapatkan

tingkat akurasi yang representatif. Melalui tahapan implementasi ini, diharapkan diperoleh gambaran objektif mengenai kemampuan model dalam mengklasifikasikan status kelulusan berdasarkan variabel-variabel akademik yang tersedia.

Pra-pemrosesan Data (Data Pre-processing)

Tahap awal penelitian melibatkan pemahaman mendalam terhadap dataset *dataset_kelulusan_mahasiswa.csv*. Dataset ini memiliki dimensi yang cukup besar, yakni 100.000 baris dan 10 kolom. Berdasarkan hasil observasi data, ditemukan beberapa nilai yang hilang (*missing values*) pada fitur 'Jumlah Cuti Akademik' dan 'Jumlah Semester'. Proses pembersihan data dilakukan untuk memastikan integritas input sebelum masuk ke tahap pelatihan model.

Untuk mengakomodasi algoritma KNN yang berbasis perhitungan jarak matematis, dilakukan dua transformasi data utama yaitu :

Pengodean Kategorikal (*Categorical Encoding*) :

Variabel seperti 'Pekerjaan Sambil Kuliah' (dengan nilai 'Ya'/'Tidak') dan 'Kategori Kehadiran' dikonversi menjadi format numerik. Variabel 'Pekerjaan Sambil Kuliah' dipetakan ke dalam nilai biner (1 untuk 'Ya', 0 untuk 'Tidak'), sedangkan 'Kategori Kehadiran' menggunakan *Label Encoder* untuk transformasi ke nilai numerik

```
# Encoding Kategorikal
le = LabelEncoder()
df['Pekerjaan Sambil Kuliah'] = df['Pekerjaan Sambil Kuliah'].map({'Ya': 1, 'Tidak': 0})
df['Kategori Kehadiran'] = le.fit_transform(df['Kategori Kehadiran'])

# Mendefinisikan X (Fitur) dan y (Target)
X = df.drop('Status Kelulusan', axis=1)
y = df['Status Kelulusan']

# Normalisasi (Min-Max Scaling)
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)
```

Gambar 2. Encoding Categorical

Normalisasi (*Min-Max Scaling*) :

Mengingat algoritma KNN sangat sensitif terhadap skala fitur, proses normalisasi menggunakan *Min-Max Scaling* diterapkan untuk merentangkan nilai fitur ke dalam skala [0, 1]. Hal ini krusial agar fitur dengan nilai nominal besar tidak mendominasi perhitungan *Euclidean Distance* terhadap fitur dengan nilai yang lebih kecil

```
# Encoding Kategorikal
le = LabelEncoder()
df['Pekerjaan Sambil Kuliah'] = df['Pekerjaan Sambil Kuliah'].map({'Ya': 1, 'Tidak': 0})
df['Kategori Kehadiran'] = le.fit_transform(df['Kategori Kehadiran'])

# Mendefinisikan X (Fitur) dan y (Target)
X = df.drop('Status Kelulusan', axis=1)
y = df['Status Kelulusan']

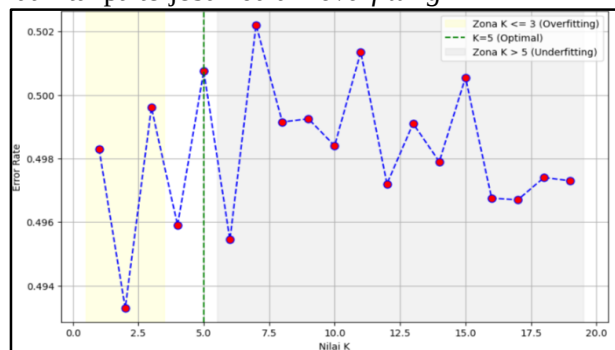
# Normalisasi (Min-Max Scaling)
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)
```

Gambar 3. Normalisasi (Min-Max Scaling)

Pemodelan dan Optimasi Parameter

Tahap pemodelan merupakan inti dari penelitian ini dalam merepresentasikan pola kelulusan mahasiswa

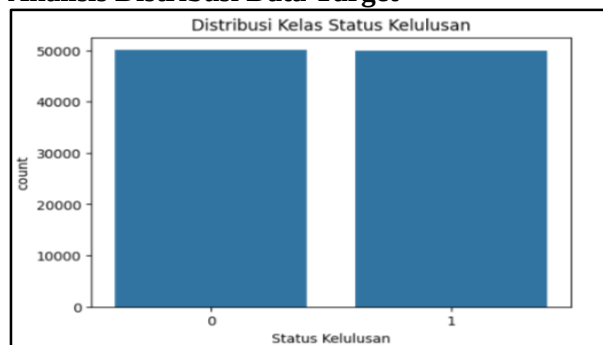
melalui algoritma *K-Nearest Neighbor* (KNN). Proses ini diawali dengan pembagian dataset menjadi dua bagian utama, yaitu data latih (*training set*) sebesar 80% dan data uji (*testing set*) sebesar 20%. Pembagian ini dilakukan dengan metode *hold-out validation* untuk memastikan bahwa model dilatih dengan porsi data yang cukup besar guna menangkap pola karakteristik mahasiswa, sekaligus menyediakan data yang independen untuk mengevaluasi kemampuan generalisasi model terhadap data baru (*unseen data*). Algoritma KNN dipilih karena sifatnya yang non-parametrik, sehingga tidak memerlukan asumsi distribusi data yang ketat. Dalam proses pemodelan, penentuan *hyperparameter* K (jumlah tetangga terdekat) menjadi langkah krusial. Nilai K yang dipilih akan secara langsung mempengaruhi *decision boundary* model; nilai K yang terlalu rendah cenderung menyebabkan *overfitting* (terlalu sensitif terhadap *noise*), sementara nilai K yang terlalu tinggi dapat memicu *underfitting* (kehilangan detail pola lokal karena terlalu banyak tetangga yang dirata-ratakan). Hasil proses pengujian nilai K dapat dilihat pada gambar 4.3, Visualisasi *Error Rate vs. K Value* menunjukkan fluktuasi nilai kesalahan, di mana titik K=5 dipilih sebagai hiperparameter optimal. Pemilihan ini didasarkan pada stabilitas *error* yang dihasilkan, yang dinilai cukup mampu menyeimbangkan sensitivitas model terhadap data latih tanpa terjebak dalam *overfitting*



Gambar 4. Visualisasi Grafik Nilai K

Pembahasan

Analisis Distribusi Data Target

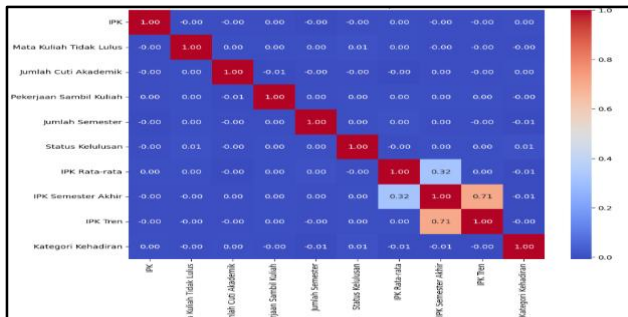


Gambar 5. Grafik Distribusi Target

Berdasarkan visualisasi pada Gambar, distribusi status kelulusan mahasiswa menunjukkan proporsi yang relatif seimbang (*balanced dataset*) antara

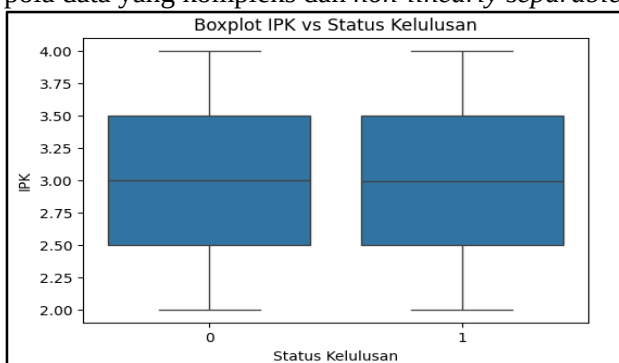
kelompok mahasiswa yang lulus dan tidak lulus. Dalam konteks pemodelan data, keseimbangan kelas ini merupakan atribut positif yang krusial. Ketiadaan dominasi kelas mayoritas meminimalisir risiko bias *classifier*, sehingga model KNN dapat melakukan kalkulasi probabilitas kelas secara objektif. Kondisi ini memberikan justifikasi metodologis bahwa perlakuan khusus terhadap data seperti *oversampling* atau *undersampling* tidak diperlukan, karena dataset telah merepresentasikan skenario klasifikasi biner yang adil bagi kedua kelompok

Analisis Korelasi dan Karakteristik Fitur



Gambar 6. Grafik Korelasi dan Karakteristik Fitur

Hasil *Correlation Heatmap* sebagaimana disajikan pada Gambar 4.5 menunjukkan nilai koefisien korelasi yang rendah antara variabel prediktor dengan status kelulusan. Temuan ini mengindikasikan bahwa tidak terdapat ketergantungan linear yang kuat antara variabel individu—seperti jumlah cuti atau kehadiran—terhadap variabel target. Hasil analisis ini juga diperkuat pada grafik *Boxplot* pada Gambar 4.5 yang memperlihatkan tumpang tindih (*overlap*) nilai IPK pada kedua kelompok mahasiswa. Secara teoretis, tumpang tindih ini menyiratkan bahwa batas pemisah (*decision boundary*) antara mahasiswa yang lulus dan tidak lulus tidak dapat didefinisikan secara linear. Oleh karena itu, penggunaan algoritma KNN menjadi sangat relevan dalam penelitian ini. Sifat KNN yang non-parametrik memungkinkannya untuk mengabaikan asumsi distribusi linear dan berfokus pada pengelompokan berdasarkan kedekatan *Euclidean distance* dalam ruang fitur, yang merupakan pendekatan yang lebih adaptif terhadap pola data yang kompleks dan *non-linearly separable*

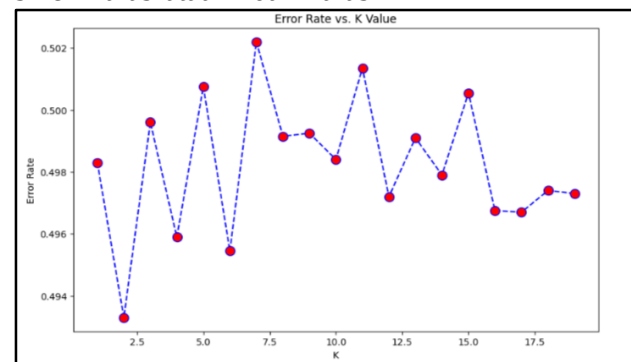


Gambar 7. Grafik Boxplot IPK vs Status Kelulusan

Optimasi Parameter K dan Stabilitas Model

Algoritma KNN mengklasifikasikan status kelulusan mahasiswa berdasarkan kedekatan profil akademik dengan tetangga terdekatnya. Dengan menggunakan $K=5$, model menghitung jarak *Euclidean* dari data uji terhadap 5 tetangga terdekat di ruang fitur. Hasil pengujian menunjukkan bahwa karakteristik dataset yang memiliki 100.000 entri memberikan tantangan tersendiri bagi kompleksitas komputasi KNN.

Penggunaan $K=5$ dalam penelitian ini didasarkan pada hasil observasi kurva *Error Rate* yang cenderung stabil pada angka tersebut. Secara teoretis, pemilihan nilai K ganjil berfungsi untuk menghindari kondisi *tie* (suara seimbang) saat proses *voting* kelas mayoritas, yang dalam kasus ini adalah klasifikasi biner 'Lulus' atau 'Tidak Lulus'

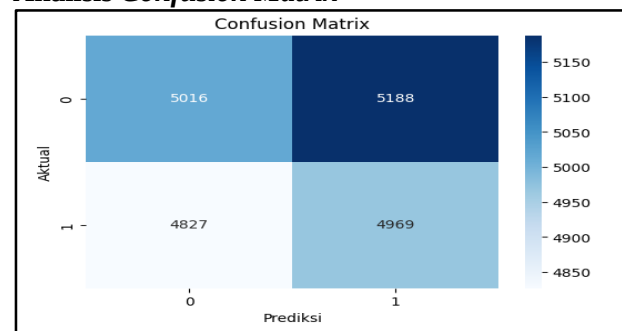


Gambar 8. Error Rate KNN

Evaluasi dan Pengujian

Tahap evaluasi dilakukan untuk mengukur kemampuan model KNN dalam melakukan prediksi status kelulusan mahasiswa *K-Nearest Neighbors* (KNN) berdasarkan data pengujian yang digunakan. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* yang diperoleh dari *classification report*

Analisis Confusion Matrix



Gambar 9. Hasil Evaluasi Confusion Matrix

Berdasarkan gambar 4.8 menunjukkan pola distribusi kesalahan yang hampir merata pada kedua kelas. Tingginya angka *False Positive* dan *False Negative* mengindikasikan bahwa model mengalami kesulitan dalam memisahkan profil akademik mahasiswa yang lulus dengan yang tidak lulus secara tegas. Ketidakmampuan model untuk membedakan kedua kelas ini secara signifikan menunjukkan adanya kemiripan pola (*feature overlap*) yang tinggi di antara

kedua kelompok mahasiswa tersebut. Dengan kata lain, ruang fitur yang digunakan belum mampu memisahkan batas keputusan (*decision boundary*) secara optimal.

Analisis Metrik Performa (Precision, Recall, dan F1-Score)

Selain *Confusion Matrix*, dilakukan pengukuran kinerja berdasarkan metrik *Precision*, *Recall*, dan *F1-Score* untuk mendapatkan perspektif yang lebih mendalam. Berdasarkan hasil klasifikasi pada **Tabel [Nomor Tabel]**, model menghasilkan nilai akurasi sebesar 0,49 atau 49%

Confusion Matrix:				
[[5016 5188]				
[4827 4969]]				
Classification Report:				
	precision	recall	f1-score	support
0	0.51	0.49	0.50	10204
1	0.49	0.51	0.50	9796
accuracy			0.50	20000
macro avg	0.50	0.50	0.50	20000
weighted avg	0.50	0.50	0.50	20000
Akurasi: 0.49925				

Gambar 10. Nilai Akurasi, Presisi dan Recall

Nilai akurasi yang berada di kisaran 50% ini menempatkan model pada tingkat *baseline* atau setara dengan probabilitas tebakan acak (*random guessing*) dalam klasifikasi biner. Analisis metrik menunjukkan nilai *Precision* dan *Recall* yang seimbang pada kedua kelas (0,49 – 0,51), yang membuktikan bahwa model tidak memiliki bias terhadap kelas tertentu. Hal ini selaras dengan kondisi dataset yang terdistribusi seimbang; namun, secara performa teknis, angka ini menegaskan bahwa algoritma KNN dengan parameter $K=5$ belum berhasil menemukan pola diskriminatif yang kuat dalam variabel prediktor yang tersedia.

V. PENUTUP

A. Kesimpulan

Berdasarkan hasil implementasi algoritma *K-Nearest Neighbor* (KNN) dan analisis data yang telah dilakukan dalam penelitian ini, dapat ditarik beberapa kesimpulan sebagai berikut:

Penelitian ini berhasil membangun kerangka kerja sistematis dalam memprediksi status kelulusan mahasiswa melalui tahapan pra-pemrosesan data (seperti *encoding* dan *scaling*), penentuan hiperparameter, hingga evaluasi model. Implementasi algoritma KNN telah dilakukan secara teknis sesuai dengan kaidah *machine learning*, di mana pemilihan nilai $K=5$ terbukti sebagai titik optimal yang memberikan keseimbangan terbaik antara *bias* dan *variance*. Hasil analisis korelasi dan distribusi data menunjukkan bahwa variabel akademik yang digunakan dalam dataset memiliki tingkat *overlap* yang tinggi dan tidak menunjukkan hubungan linear yang kuat terhadap status kelulusan. Hal ini mengonfirmasi bahwa status kelulusan mahasiswa

merupakan fenomena yang kompleks dan tidak dapat dipisahkan secara sederhana berdasarkan variabel numerik dasar saja. Berdasarkan hasil pengujian final dengan *Confusion Matrix*, model KNN menghasilkan tingkat akurasi sebesar 50%. Angka ini mencerminkan batas kemampuan model dalam mengklasifikasikan kelas berdasarkan informasi yang tersedia. Hasil evaluasi tersebut membuktikan bahwa untuk dataset ini, model berada pada level klasifikasi acak, yang mengindikasikan bahwa terdapat keterbatasan informasi (*informational constraints*) pada variabel prediktor saat ini dalam menangkap pola perilaku mahasiswa secara presisi.

B. Saran

Berdasarkan keterbatasan yang ditemukan selama proses penelitian, berikut adalah beberapa saran untuk pengembangan dan peningkatan kualitas penelitian selanjutnya:

Menambahkan variabel atau kriteria untuk melakukan *feature engineering* dengan menambahkan variabel laten atau data kualitatif-sosiologis. Faktor-faktor seperti kondisi ekonomi, dukungan sosial, motivasi intrinsik, serta keterlibatan mahasiswa dalam kegiatan organisasi kampus berpotensi memberikan kontribusi yang lebih besar dalam mendeskripsikan profil mahasiswa yang lulus dan tidak. Melakukan analisis lebih mendalam terkait penanganan *outlier* dan data pencilan sebelum dilakukan pemodelan. Selain itu, eksplorasi terhadap teknik seleksi fitur (*feature selection*) dapat diterapkan untuk mengidentifikasi variabel mana yang sebenarnya memiliki kontribusi paling relevan, guna mengurangi *noise* dan meningkatkan fokus model dalam mempelajari pola yang sebenarnya. Model yang dikembangkan nantinya diintegrasikan ke dalam Sistem Pendukung Keputusan (SPK) yang mampu memberikan rekomendasi intervensi akademik bagi mahasiswa yang teridentifikasi memiliki risiko 'tidak lulus' berdasarkan pola yang ditemukan oleh model.

DAFTAR PUSTAKA

- [1] Adawiyah, Q. (2024). Penerapan Algoritma K-Means Clustering untuk Mengelompokkan Rekomendasi Jenis KB Berbasis Machine Learning (Studi Kasus di Puskesmas Pegang Baru). Universitas Putra Indonesia YPTK Padang.
- [2] Agustina, F. D., Arif, M., & Ahmad, S. (2025). Systematic Literature Review atas Kinerja Algoritma KNN, Naïve Bayes, dan Decision Tree pada Berbagai Studi Prediksi dan Klasifikasi. *Jurnal Jawara Sistem Informasi*, 3(1).
- [3] Ahmad, N., Hafizh, S., & Sulthanah, R.

- (2024). Prediksi Kelulusan Mata Kuliah Mahasiswa Teknologi Informasi Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Manajemen Informatika (JAMIKA)*, 14(2), 135–149. <https://doi.org/10.34010/jamika.v14i2.12454>
- [4] Alfarizi, M. R. sirfatullah, Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning. *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)*, 2(1), 1–6.
- [5] Andriani, S., Nazir, A., Candra, R. M., Syafria, F., & Afrianty, I. (2023). Implementasi Algoritma K-Nearest Neighbor Untuk Menentukan Klasifikasi Kelulusan Mahasiswa Teknik Informatika. *Journal of Computer System and Informatics (JoSYC)*, 4(4), 922–930. <https://doi.org/10.47065/josyc.v4i4.3914>
- [6] Fahmi, M. N. (2023). Implementasi Mechine Learning menggunakan Python Library : Scikit-Learn (Supervised dan Unsupervised Learning). *Sains Data Jurnal Studi Matematika Dan Teknologi*, 1(2), 87–96. <https://doi.org/10.52620/sainsdata.v1i2.31>
- [7] Faisal, M., Utami, W. S., & Parmica, S. (2023). Implementasi Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Indeks Kemiskinan. *Journal Sensi*, 9(1), 11–23. <https://doi.org/10.33050/sensi.v9i1.2616>
- [8] Febryana, S., Wulandari, I. A., & Prabowo, D. (2024). Rancang Bangun Aplikasi Pengolahan Data Inventory Barang Berbasis Desktop Pada Cv. Hamim Group. *Jurnal Mahasiswa Sistem Informasi (JMSI)*, 6(1), 24–35. <https://doi.org/10.24127/jmsi.v6i1.7553>
- [9] Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM model menggunakan metode decision tree dengan algoritma CART untuk prediksi curah hujan berpotensi banjir. *Journal of Applied Informatics and Computing (JAIC)*, 5(2), 103. <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [10] Iriane, R., & Nurfaizah. (2023). Penerapan Data Mining Untuk Prediksi Penjualan Produk Pangan Hewan Menggunakan Metode K-Nearest Neighbor. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 3(5), 509–515. <https://djournals.com/klik>
- [11] Irsabmalfi, R., & Rahmadewi, R. (2025). Deteksi Anatomi Gigi Menggunakan Arsitektur YOLOV8 (You Only Look Once) Untuk Pengenalan Jenis-Jenis Gigi. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4870–4877.
- [2] Moonallika, P. S. C., Fredlina, K. Q., & Sudiatmika, I. B. K. (2020). Penerapan Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus STMIK Primakara). *Progresif: Jurnal Ilmiah Komputer*, 16(1), 47–56. <http://ojs.stmik-banjarbaru.ac.id/index.php/progresif/article/view/427>
- [13] Nahjan, M. R., Heryana, N., & Voutama, A. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101–104.
- [14] Naibaho, A. F. A., & Zahra, A. (2023). Prediksi Kelulusan Siswa Sekolah Menengah Pertama Menggunakan Machine Learning. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3). <https://doi.org/10.23960/jitet.v11i3.3056>
- [15] Naldy, E. T., & Andri, A. (2021). Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN. *Jurnal Nasional Ilmu Komputer*, 2(2), 89–101. <https://doi.org/10.47747/jurnalnik.v2i2.525>
- [16] Noiija, H. C., Wemaf, P. A., Nurdianty, O. A., Sohilit, W., Haumahu, S., Yusuf, H., Tomagola, K., Wa Rindi La Ana, & Salaiswa, T. S. (2023). Perancangan Sistem Informasi Akuntansi Untuk Analisis Siklus Pendapatan Pada Orantata Celullar Menggunakan DFD Dan Flowchart. *Jurnal Bisnis Dan Manajemen (JURBISMAN)*, 1(2), 577–592. <https://doi.org/10.61930/jurbisman.v1i2.188>
- [17] Octarina, S., Puspita, F. M., Yuliza, E., & Indrawati. (2025). Pendampingan

- Penggunaan Google Colab Pada Pembelajaran Python Dan Machine Learning Bagi Dosen Matematika DI PALEMBANG. *Jurnal Terpadu*, 6(1), 56–66.
- [18] Rahmat, A., Hardi, H., Syam, F. A., Zamzami, Z., Febriadi, B., & Windarto, A. P. (2021). Utilization of the field of data mining in mapping the area of the Human Development Index (HDI) in Indonesia. *Journal of Physics: Conference Series*, 1783(1). <https://doi.org/10.1088/1742-6596/1783/1/012035>
- [19] Santoso, P., Abijono, H., & Anggreini, N. L. (2021). Algoritma Supervised Learning Dan Unsupervised Learning Dalam Pengolahan Data. *Jurnal Teknologi Terapan: G-Tech*, 4(2), 315–318. <https://doi.org/10.33379/gtech.v4i2.635>
- [20] Sartika, D., & Jumadi, J. (2023). Implementasi Data Mining Untuk Pengelompokan Wilayah Pelanggaran Lalu Lintas Menggunakan Metode K-Means Pada Polres Bengkulu. *Journal of Science and Social Research*, 4307(2), 345–349. <http://jurnal.goretanpena.com/index.php/JSSR>
- [21] Surbakti, N. M., Angelyca, Talia, A., Perangin-Angin, C. B., Nainggolan, D. O., Friskauly, N. D., & Tumorang, S. R. B. (2024). Penggunaan Bahasa Pemrograman Python dalam Pembelajaran Kalkulus Fungsi Dua Variabel. *Algoritma : Jurnal Matematika, Ilmu Pengetahuan Alam, Kebumihan Dan Angkasa*, 2(3), 98–107.
- [22] Syam, S., Tokoro, Y., Judijanto, L., Garonga, M., Sinaga, F. M., Umar, N., Handika, I. P. S., Iin, J. N., Apriyanto, A., & Sitanggang, A. T. (2024). *Data Mining: Teori dan Penerapannya dalam Berbagai Bidang*. PT. Sonpedia Publishing Indonesia.
- [23] Tumbilung, C. F., & Efendi, R. (2025). Prediksi Risiko Drop Out Mahasiswa Menggunakan Model Machine Learning Berbasis Data Akademik (Studi Kasus : Universitas XYZ). *JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 16(2), 176–183. <https://ejurnal.provisi.ac.id/index.php/JTIKP/article/view/1257/878>
- [24] Wardhana, R. G., Wang, G., & Sibuea, F. (2023). PENERAPAN MACHINE LEARNING DALAM PREDIKSI TINGKAT KASUS PENYAKIT DI INDONESIA. *Journal of Information System Management (JOISM)*, 5(1).
- [25] Widodo, A. P., & Setiawan, A. (2025). Systematic Literature Review : Peran Machine Learning dalam Manajemen Sumber Daya Manusia. *Optimal: Jurnal Ekonomi Dan Manajemen*, 5, 650–665. <https://doi.org/10.55606/optimal.v5i4.8089>
- [26] Wiguna, D. P., Juliana, L., & Ginting, S. E. (2025). Aplikasi Google Colab Berbasis Python dalam Menerapkan Teori Pohon dengan Algoritma Random Forest Classifier. *Scripta Technica : Journal of Engineering and Applied Technology*, 1(2), 261–270.
- [27] Wulandari, G., Sasmita, S., & Rahmadayanti, F. (2024). Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Obat. *J. Tek. Inf. Dan Komput*, 7(1), 144–150.
- [28] Yoliad, D. N. (2023). Data mining Dalam Analisis Tingkat Penjualan Barang Elektronik Menggunakan Algoritma K-means. *Insearch (Information Syst. Res. J., Vol. 3, No. 1*.