JTIK
Jurnal Teknologi Informasi dan Komunikasi

JTIK (Jurnal Teknologi Informasi dan Komunikasi)
p-ISSN : 2087-0868 (media cetak); e-ISSN : 2598-9707 (media online)
Vol. 16, No. 2, September 2025
http://ejurnal.provisi.ac.id/index.php/JTIKP

PREDIKSI RISIKO DROP OUT MAHASISWA MENGGUNAKAN MODEL MACHINE LEARNING BERBASIS DATA AKADEMIK (Studi Kasus : Universitas XYZ)

Chresto Friedrich Tumbilung[1], Rissal Efendi[2]
[1,2] Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana Salatiga

Article Info

Article history:
Received June 15, 2025
Revised June 29, 2025
Accepted August 29, 2025

Keywords:
Higher Education,
Dropout Prediction,
Machine Learning,
Decision Tree,
Naive Bayes,
K-Nearest Neighbor,
Academic Data.

ABSTRACT

Higher education is crucial for developing competitive human resources, yet the issue of student dropout (DO) remains a significant challenge for institutions. This study aims to develop a predictive model for identifying students at risk of dropout using machine learning techniques. By analyzing academic data, including Grade Point Averages (GPA), course loads, attendance rates, and failure rates, the research employs three machine learning algorithms: Decision Tree, Naive Bayes, and K-Nearest Neighbor (KNN). The results indicate that the Decision Tree model outperforms the others, achieving a perfect accuracy of 100% in classifying students as either "Graduated" or "Dropout." Naive Bayes also shows strong performance with 95% accuracy, particularly excelling in identifying actual dropout cases. Conversely, KNN exhibits the lowest effectiveness. The findings suggest that implementing the Decision Tree model can significantly enhance early detection and intervention strategies for at-risk students, ultimately improving academic management and student retention rates.

Corresponding Author:
Chresto Friedrich Tumbilung,
Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana Salatiga
Jl. Dr. O. Notohamidjojo No.1 - 10, Blotongan, Kec. Sidorejo, Kota Salatiga
Email: cftumbilung05@gmail.com

1. PENDAHULUAN

Pendidikan tinggi merupakan fondasi utama dalam mencetak sumber daya manusia yang unggul dan mampu bersaing di era globalisasi [1]. Perguruan tinggi tidak hanya dituntut untuk memberikan proses pembelajaran yang berkualitas, tetapi juga menjamin keberlangsungan studi mahasiswa hingga mencapai kelulusan. Namun demikian, dalam praktiknya, fenomena mahasiswa yang mengalami putus studi atau drop out (DO) masih menjadi persoalan serius di berbagai institusi pendidikan tinggi, baik negeri maupun swasta. Tingginya angka drop out bukan hanya merugikan mahasiswa secara individu, tetapi juga mencerminkan rendahnya efektivitas sistem akademik dan manajemen institusi, serta berimplikasi pada akreditasi, efisiensi anggaran, dan reputasi lembaga pendidikan [2].

Berdasarkan berbagai studi dan laporan internal perguruan tinggi, terdapat sejumlah faktor yang sering dikaitkan dengan potensi drop out mahasiswa, antara lain rendahnya indeks prestasi kumulatif (IPK), ketidakteraturan dalam pengambilan SKS, tingkat kehadiran yang buruk, masalah ekonomi, hingga faktor psikologis dan sosial. Kendati demikian, sebagian besar identifikasi terhadap mahasiswa berisiko DO selama ini masih bersifat manual dan dilakukan setelah risiko tersebut menjadi nyata. Tidak adanya sistem prediktif yang sistematis membuat intervensi menjadi terlambat dan kurang efektif dalam mencegah kasus drop out [3].

Di sisi lain, seiring berkembangnya teknologi informasi dan sistem akademik digital, data mahasiswa telah tersimpan dalam sistem informasi manajemen akademik (SIMA) dengan cukup rapi. Data ini mencakup histori akademik mahasiswa, data kehadiran, pola pengambilan mata kuliah, hingga data administrasi pembayaran [4]. Ironisnya, data tersebut masih jarang dimanfaatkan secara maksimal untuk kepentingan prediktif yang bersifat preventif. Padahal, kemajuan teknologi kecerdasan buatan (AI), khususnya cabang machine learning, membuka peluang besar untuk membangun sistem yang mampu melakukan prediksi terhadap mahasiswa yang berpotensi mengalami drop out secara lebih akurat dan berbasis data [5].

Drop out (DO) merupakan kondisi di mana mahasiswa menghentikan proses studinya sebelum menyelesaikan program pendidikan yang ditempuh. DO bisa terjadi karena berbagai faktor, baik yang bersifat internal seperti motivasi belajar, kondisi keuangan, kondisi psikologis, maupun eksternal seperti lingkungan sosial, kualitas pengajaran, dan sistem akademik [13]. Tingginya tingkat DO dapat merugikan institusi pendidikan karena menunjukkan ketidakefektifan dalam mempertahankan mahasiswa hingga lulus serta berdampak terhadap akreditasi dan reputasi kampus.Dalam konteks akademik, beberapa indikator yang sering dikaitkan dengan risiko DO meliputi: (a) Nilai akademik yang rendah atau fluktuatif, (b) jumlah pengambilan SKS yang tidak konsisten, (c) tingkat kehadiran yang rendah, (d) ketidaktuntasan dalam mata kuliah tertentu, dan (e) durasi studi yang melebihi masa studi ideal adalah indikator yang sering dikaitkan dengan risiko drop out (DO) pada mahasiswa.

Data akademik mencakup informasi yang berkaitan dengan aktivitas belajar mahasiswa, seperti: (a) Nilai IPK dan IPS per semester, (b) jumlah SKS yang diambil dan lulus, (c) jumlah mata kuliah yang tidak lulus (E/T), (d) frekuensi pengulangan mata kuliah, (e) durasi studi tiap semester, dan (f) kehadiran atau absensi kelas adalah indikator penting yang dapat mempengaruhi kinerja akademik mahasiswa. Data tersebut dapat dijadikan sebagai fitur dalam membangun model machine learning untuk mengklasifikasikan apakah seorang mahasiswa berada dalam kategori berisiko tinggi mengalami DO [18].

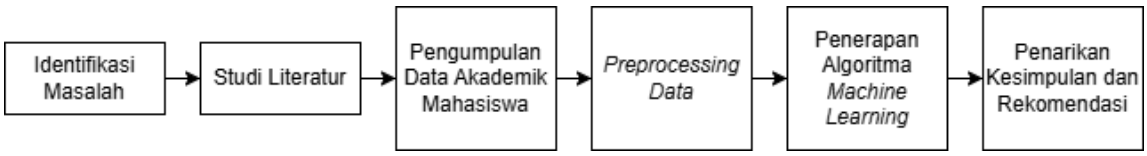
Machine learning adalah metode analitik cerdas yang dapat mempelajari pola dari data historis untuk menghasilkan model prediktif yang adaptif dan dapat diperbarui seiring bertambahnya data. Berbagai algoritma seperti Decision Tree, Naive Bayes, dan K-Nearest Neighbors memungkinkan sistem untuk mengenali pola-pola akademik yang secara statistik berhubungan dengan kemungkinan terjadinya DO. Dengan pendekatan ini, institusi pendidikan dapat mengambil langkah proaktif dengan memberikan perhatian khusus, pendampingan akademik, maupun konseling psikologis kepada mahasiswa yang terdeteksi berada dalam kelompok risiko tinggi [6].

Namun demikian, pengembangan model prediktif drop out tidaklah sederhana. Tantangan utama dalam penelitian ini adalah bagaimana mengidentifikasi atribut-atribut akademik yang paling relevan dan signifikan, memilih model machine learning yang tepat dengan mempertimbangkan keterbatasan data dan kompleksitas pemrosesan, serta mengevaluasi performa model dalam hal akurasi, presisi, dan sensitivitas. Selain itu, aspek interpretabilitas juga menjadi pertimbangan penting, mengingat hasil dari model ini akan digunakan untuk pengambilan keputusan oleh pihak akademik yang belum tentu berlatar belakang teknis [7].

Berdasarkan latar belakang tersebut, penelitian ini difokuskan pada pengembangan model prediksi risiko drop out mahasiswa berbasis machine learning dengan memanfaatkan data akademik yang tersedia. Diharapkan, model ini dapat memberikan kontribusi nyata dalam meningkatkan kualitas manajemen akademik dan memberikan solusi yang lebih berbasis data dalam menangani isu DO secara dini dan terarah. Penelitian ini bertujuan untuk merancang dan mengimplementasikan model prediksi risiko drop out mahasiswa berbasis data akademik, untuk menganalisis dan membandingkan performa algoritma machine learning sederhana dalam memprediksi drop out mahasiswa, dan untuk memberikan rekomendasi sistem pendeteksian dini mahasiswa berisiko drop out bagi institusi pendidikan.

2. METODE PENELITIAN

Jenis penelitian yang digunakan dalam studi ini adalah penelitian kuantitatif dengan pendekatan eksperimental [19]. Penelitian ini bertujuan untuk menguji efektivitas model machine learning dalam memprediksi risiko drop out mahasiswa berdasarkan data akademik historis. Pendekatan ini digunakan untuk mendapatkan hasil prediksi yang dapat diukur secara statistik. Alur metode penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Metode Penelitian

Alur metode penelitian yang digunakan oleh peneliti dalam studi ini dimulai dengan identifikasi masalah, yang bertujuan untuk mengenali permasalahan utama dalam lingkungan akademik, khususnya mengenai mahasiswa yang berisiko mengalami drop out (DO). Permasalahan ini dianalisis berdasarkan fenomena yang ada serta kebutuhan institusi pendidikan dalam mendeteksi dini risiko DO. Setelah masalah diidentifikasi, peneliti melakukan studi literatur untuk memperkuat dasar teori dan mengetahui pendekatan serta metode yang relevan dari penelitian sebelumnya. Literatur yang dikaji mencakup teori dropout, model machine learning dalam pendidikan, serta studi-studi terdahulu yang membahas topik serupa.

Selanjutnya, data dikumpulkan dari sistem akademik perguruan tinggi (SIAKAD) dan wawancara pihak terkait. Data yang diperoleh meliputi IPK, IPS, jumlah SKS yang diambil dan lulus, mata kuliah yang gagal, riwayat pengulangan mata kuliah, serta kehadiran mahasiswa. Setelah pengumpulan data, tahap berikutnya adalah data preprocessing, di mana data yang telah dikumpulkan diproses agar siap digunakan dalam pemodelan machine learning. Proses ini mencakup pembersihan data (handling missing values), normalisasi data numerik, serta pengkodean data kategorikal.

Tahap selanjutnya adalah penerapan algoritma machine learning, di mana beberapa algoritma yang dipilih, seperti Decision Tree, Naive Bayes, dan K-Nearest Neighbor (KNN), diterapkan untuk membangun model prediksi terhadap data yang telah diproses. Akhirnya, peneliti menyusun kesimpulan dari hasil penelitian, termasuk efektivitas model prediksi yang dibangun dan memberikan rekomendasi praktis bagi pihak kampus untuk mengidentifikasi dan menangani mahasiswa yang berisiko DO secara dini.

Penelitian ini dilakukan di Universitas XYZ, dengan waktu pelaksanaan dari bulan berdurasi 3 bulan. Pengumpulan data dilakukan dengan metode dokumentasi dan ekstraksi data dari sistem akademik universitas (SIAKAD) dan hasil wawancara [22]. Sumber data terdiri dari data primer diperoleh melalui wawancara dengan dosen wali akademik dan pihak Biro Administrasi Akademik (BAA) untuk mengetahui faktor-faktor umum penyebab DO. Data sekunder berupa data akademik mahasiswa dari semester 1 hingga semester terakhir, yang meliputi: (a) Indeks Prestasi Kumulatif (IPK), (b) Indeks Prestasi Semester (IPS), (c) jumlah SKS yang diambil dan lulus, (d) jumlah mata kuliah yang gagal (nilai E atau tidak lulus), (e) riwayat pengambilan ulang mata kuliah, (f) kehadiran atau absensi, dan (g) status aktif/non-aktif mahasiswa. Data tersebut mencerminkan kinerja akademik mahasiswa dari semester 1 hingga semester terakhir.

Teknik analisis data dalam penelitian ini meliputi beberapa langkah penting. Pertama, (1) pre-processing data dilakukan untuk mempersiapkan data agar siap untuk analisis lebih lanjut. Selanjutnya, (2) pembagian dataset dilakukan untuk memisahkan data menjadi data latih dan data uji. Kemudian, (3) pemodelan machine learning diterapkan untuk membangun model prediksi berdasarkan data yang telah diproses. Akhirnya, (4) evaluasi model dilakukan untuk menilai kinerja dan efektivitas model yang telah dibangun.

3. HASIL DAN PEMBAHASAN

3.1. Deskripsi Data

Penelitian ini menggunakan data akademik mahasiswa sebagai basis data dari Universitas XYZ sebuah perguruan tinggi swasta di Indonesia. Dataset terdiri dari 100 entri mahasiswa dengan atribut yang dijelaskan pada Tabel 1.

Tabel 1. Atribut Entri Mahasiswa

Nama Atribut	Tipe Data	Keterangan
IPS_semester_1	*float*	Indeks Prestasi Semester 1
IPS_semester_2	*float*	Indeks Prestasi Semester 2
SKS_total_lulus	*integer*	Jumlah total SKS yang telah diluluskan
matakuliah_gagal	*integer*	Jumlah mata kuliah dengan nilai E
kehadiran_ratarata	*float (%)*	Rata-rata kehadiran tiap semester (%)
status_mahasiswa	*category*	Label target: 'Lulus', '*Dropout'*

Atribut diatas digunakan sebagai acuan dalam menentukan prediksi mahasiswa apakah berdampak *dropout* atau lulus.

3.2. Implementasi dan Pelatihan Model Algoritma

Untuk mengklasifikasikan status mahasiswa (Lulus atau *Dropout*), digunakan tiga algoritma sederhana yaitu *Decision Tree, Naive Bayes, dan K-Nearest Neighbor* (KNN), yang dikenal sederhana namun cukup kuat dalam menangani klasifikasi berbasis data kategorikal dan numerik.

Tabel 2. Preprocessing Data

```
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import pandas as pd
import numpy as np

# Simulasi Data
np.random.seed(42)
data = pd.DataFrame({
   'ips1': np.round(np.random.uniform(1.5, 4.0, 500), 2),
   'ips2': np.round(np.random.uniform(1.5, 4.0, 500), 2),
   'sks_lulus': np.random.randint(20, 140, 500),
   'mk_gagal': np.random.randint(0, 6, 500),
   'kehadiran': np.round(np.random.uniform(60, 100, 500), 2)
})

# Label berdasarkan aturan
data['status'] = np.where((data['ips1'] < 2.0) | (data['ips2'] < 2.0) | (data['mk_gagal'] > 3)
(data['kehadiran'] < 75), 'Dropout', 'Lulus')

# Split Data
X = data[['ips1', 'ips2', 'sks_lulus', 'mk_gagal', 'kehadiran']]
y = data['status']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Pelatihan Model
model = DecisionTreeClassifier(max_depth=5, random_state=42
```

Tabel 2 menunjukan hasil kodingan dari *preprocessing* data menggunakan *python* sebagai *tools* dari pengimplementasian proses *preprocessing* data.

Tabel 3. Evaluasi Model Algoritma

```
# Prediksi
y_pred = model.predict(X_test)
# Evaluasi
print("Akurasi:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))
```

Tabel 3 menunjukan hasil kodingan dari evaluasi model, evaluasi model dilakukan untuk melihat seberapa baik model dalam mengklasifikasikan data mahasiswa.

Tabel 4. Kode Prediksi Mahasiswa Lulus

```
contoh_mahasiswa = pd.DataFrame({
   'ips1': [2.6],
   'ips2': [2.9],
   'sks_lulus': [65],
   'mk_gagal': [1],
   'kehadiran': [85.0]
})
hasil_prediksi = model.predict(contoh_mahasiswa)
print("Status Prediksi:", hasil_prediksi[0])
```

Tabel 4 menunjukan hasil kodingan dari prediksi mahasiswa lulus, bagian kode ini berfungsi sebagai notifikasi atau tanda yang menunjukan bahwa mahasiswa dengan atribut serta nilai yang telah diatur dapat diprediksikan lulus melalui model algoritma yang telah dibuat.

Tabel 5. Hasil Evaluasi Model dan Kinerja

Model Algoritma	***Accuracy***	***Precision***		***Recall***		***F1 Score***	
		Lulus	**DO**	**Lulus**	**DO**	**Lulus**	**DO**
Decision Tree	100%	100%	100%	100%	100%	100%	100%
Naive Bayes	95%	100%	93,7%	80,8%	100%	89.4%	96,7%
K-Nearest Neighbor (KNN)	73%	48,4%	84,1%	57,7%	78,4%	52,6%	81,1%

Tabel 5 menunjukan hasil evaluasi model yang telah digunakan dan telah diproses, hasil menunjukan model Keakuratan sempurna dari model ini mengindikasikan bahwa *Decision Tree* mampu menangkap pola-pola pada data secara optimal. Hal ini juga dapat terjadi karena model kemungkinan *overfitting* terhadap data *training*, terlebih karena data yang digunakan hanya beberapa atribut sehingga terlalu teratur atau tidak mencerminkan kompleksitas dunia nyata. Meski begitu, ini menunjukkan bahwa *Decision Tree* sangat efisien jika fitur-fitur yang digunakan cukup informatif dengan menunjukan hasil prediksi sempurna, hasil dari evaluasi *Decision Tree* mendapatkan nilai Dropout Prediksi: Dropout (74), Lulus Prediksi: Lulus (26) dengan ini tidak terdapat data yang salah.

Model *Naive Bayes* menunjukkan kemampuan klasifikasi yang baik dengan kekuatan utama pada prediksi kelas *Dropout*. Meskipun sedikit keliru dalam mengklasifikasikan 5 mahasiswa yang sebenarnya

Lulus, model ini masih menunjukkan potensi untuk digunakan dalam skenario nyata dengan data yang lebih kompleks. Karena asumsi independensi antar fitur, performa *Naive Bayes* dapat menurun jika fitur-fitur dalam dataset saling berkorelasi. hasil dari evaluasi *Naive Bayes* mendapatkan nilai Dropout Prediksi: Dropout (74), Lulus Prediksi: Lulus (21) dengan ini terdapat 5 data pada lulus prediksi yang salah.

Model *K-Nearest Neighbor* (KNN) menunjukkan performa prediksi yang kurang stabil. Meskipun masih cukup baik dalam mengklasifikasikan mahasiswa yang *Dropout*, model ini kesulitan membedakan mahasiswa yang seharusnya *Lulus*. Hal ini kemungkinan besar disebabkan oleh sensitivitas algoritma ini terhadap skala dan distribusi data, serta pemilihan parameter k yang belum dioptimalkan lebih lanjut. Model ini mungkin memerlukan normalisasi data dan tuning parameter agar hasil lebih optimal. hasil dari evaluasi Model *K-Nearest Neighbor* (KNN) mendapatkan nilai Dropout Prediksi: Dropout (58), Lulus Prediksi: Lulus (15) dengan ini terdapat 16 data pada lulus prediksi yang salah dan 11 data pada *dropout* prediksi yang salah.

3.3. Pembahasan dan Rekomendasi Model Algoritma

Berdasarkan hasil dan *testing* yang telah dilakukan, analisis komparatif terhadap tiga algoritma klasifikasi, yaitu *Decision Tree, Naive Bayes, dan K-Nearest Neighbor* (KNN), memperlihatkan perbedaan signifikan dalam kapabilitas prediktifnya. Evaluasi kinerja model didasarkan pada metrik standar klasifikasi, meliputi akurasi (*Accuracy*), presisi (*Precision*), recall (*Recall*), dan skor F1 (*F1 Score*). Secara spesifik, metrik *Precision, Recall, dan F1 Score* diukur untuk mengestimasi kemampuan model dalam mengklasifikasikan secara tepat dua kategori outcome, yaitu "Lulus" dan "DO"

Observasi terhadap hasil evaluasi mengungkapkan bahwa algoritma *Decision Tree* menunjukkan performa optimal di seluruh metrik evaluasi. Model ini mencapai akurasi sempurna sebesar 100%, mengindikasikan kemampuan klasifikasi yang absolut tanpa adanya kesalahan prediksi. Lebih lanjut, nilai Precision, Recall, dan F1 Score untuk kedua kelas target ("Lulus" dan "DO") juga mencapai nilai maksimum (100%). Hal ini mengimplikasikan bahwa model *Decision Tree* memiliki kemampuan yang sangat baik dalam mengidentifikasi secara tepat semua instans dari kedua kelas, dengan tingkat positif palsu (*false positive*) dan negatif palsu (*false negative*) yang nihil.

Algoritma *Naive Bayes*, di sisi lain, memperlihatkan kinerja yang sangat kompeten meskipun sedikit di bawah superioritas *Decision Tree*. Akurasi yang dicapai adalah 95%, dengan nilai *Precision*, *Recall*, dan F1 *Score* yang secara umum tinggi untuk kedua kategori. Temuan menarik adalah nilai *Recall* untuk kelas "DO" yang mencapai 100%, mengindikasikan bahwa model ini mampu mengidentifikasi seluruh instans aktual dari kategori tersebut. Kendati demikian, nilai Precision untuk kelas "DO" tercatat sedikit lebih rendah (93,7%), yang mengisyaratkan adanya sejumlah kecil kasus di mana model memprediksi suatu instans sebagai "DO" padahal faktanya termasuk dalam kategori "Lulus".

Sebaliknya, algoritma *K-Nearest Neighbor* (KNN) menunjukkan kinerja yang relatif inferior dibandingkan dengan dua model lainnya. Akurasi yang diperoleh hanya sebesar 73%, dan nilai *Precision*, *Recall*, serta *F1 Score* untuk kedua kelas secara konsisten lebih rendah. Secara khusus, nilai *Precision* untuk kelas "Lulus" sangat rendah (48,4%), mengindikasikan proporsi instans yang cukup besar yang diprediksi sebagai "Lulus" namun sebenarnya termasuk dalam kategori "DO". Meskipun nilai Recall untuk kelas "DO" tergolong moderat (78,4%), secara keseluruhan, kapabilitas klasifikasi model KNN dalam membedakan kedua kategori outcome kurang optimal.

Dalam menentukan model dengan kinerja terbaik dan memberikan rekomendasi yang valid, perlu dipertimbangkan tujuan spesifik dari tugas klasifikasi ini. Apabila prioritas utama adalah mencapai tingkat akurasi global yang maksimal dengan minimisasi kesalahan klasifikasi secara keseluruhan, maka model Decision Tree secara eksplisit merupakan pilihan yang paling sesuai mengingat performanya yang sempurna di seluruh metrik evaluasi. Akan tetapi, dalam konteks aplikasi tertentu, penekanan yang berbeda mungkin diberikan pada metrik kinerja yang spesifik. Sebagai contoh, apabila implikasi dari kegagalan mengidentifikasi instans "DO" (*false negative*) memiliki konsekuensi yang lebih serius, maka model Naive Bayes dengan nilai Recall 100% untuk kelas "DO" dapat menjadi pertimbangan yang relevan, meskipun dengan sedikit kompromi pada nilai Precision untuk kelas tersebut. Mempertimbangkan hasil evaluasi secara komprehensif, dan tanpa adanya informasi tambahan mengenai implikasi asimetris dari kesalahan klasifikasi antar kelas, model Decision Tree direkomendasikan sebagai model dengan kinerja terbaik dan paling menjanjikan untuk tugas

klasifikasi ini, dikarenakan kemampuannya dalam menghasilkan prediksi yang akurat dan seimbang untuk kedua kategori outcome yang diamati.

4. KESIMPULAN

Berdasarkan analisis komparatif terhadap tiga algoritma klasifikasi yang meliputi Decision Tree, Naive Bayes, dan K-Nearest Neighbor (KNN), dapat disimpulkan bahwa model Decision Tree menunjukkan kinerja klasifikasi yang paling superior dalam konteks penelitian ini. Hasil evaluasi yang konsisten menunjukkan bahwa Decision Tree mencapai akurasi, presisi, recall, dan skor F1 sebesar 100% untuk kedua kelas target, yaitu "Lulus" dan "DO". Capaian ini mengindikasikan kemampuan model dalam melakukan klasifikasi tanpa adanya kesalahan prediksi, baik dalam mengidentifikasi instans positif maupun negatif. Model Naive Bayes menunjukkan kinerja yang sangat baik sebagai alternatif kedua, dengan akurasi sebesar 95% dan nilai metrik lainnya yang tinggi. Keunggulan spesifik model ini terletak pada kemampuannya dalam mengidentifikasi seluruh instans kelas "DO" (recall 100%), meskipun dengan sedikit penurunan pada presisi untuk kelas tersebut. Sementara itu, model K-Nearest Neighbor (KNN) menunjukkan performa yang paling rendah di antara ketiga algoritma yang diuji, dengan akurasi dan nilai metrik lainnya yang secara signifikan lebih rendah, mengindikasikan keterbatasan dalam kemampuannya untuk memisahkan kedua kelas secara efektif. Secara keseluruhan, hasil penelitian ini secara empiris mendukung keunggulan algoritma Decision Tree dalam memprediksi outcome "Lulus" dan "DO" berdasarkan dataset yang digunakan. Kinerja yang optimal dan konsisten di seluruh metrik evaluasi menjadikannya model yang paling andal untuk tugas klasifikasi ini.

REFERENCES (10 PT)

[1] R. B. R. Masinambow and A. Nugroho, "Analisis Strategi Pemasaran Biznet Dan Indihome Di Salatiga Menggunakan Game Theory," *J. Pendidik. Teknol. Inf.*, vol. 6, no. 2, pp. 299–307, 2023, doi: 10.37792/jukanti.v6i2.1042.

[2] M. T. N. Hidayat, A. I. Abdurrahman, and A. Suseno, "Strategi Sosialisasi Dinas Pendidikan Kabupaten Tangerang Dalam Menguran gi Anak Putus Sekolah (Studi Kasus Pelajar SMP di Desa Sukamulya Kecamatan Sukamulya)," *J. Ilm. Wahana Pendidik.*, vol. 10, no. 8, pp. 665–689, 2024.

[3] G. C. Ramadhani and P. Oktamianti, "Faktor Yang Mempengaruhi Drop Out Pasien Pada Pelayanan Fisioterapi : Systematic Literature Review," *J. Kesehat. Masy.*, vol. 8, no. 2, pp. 3248–3255, 2024.

[4] N. Rohmah, F. Wahyudi, U. Mudhifatul Jannah, and Z. Nyndia Rahmawati, "Audit Sistem Informasi Manajemen Akademik (SIMA) Menggunakan Framework COBIT 5.0 Domain Align, Plan and Organise (APO) Studi Kasus : Universitas Islam Raden Rahmat Malang," *J. Sist. Inf. dan Inform.*, vol. 1, no. 2, pp. 98–103, 2022, doi: 10.33379/jusifor.v1i2.1598.

[5] A. R. Afandi and H. Kurnia, "Revolusi Teknologi: Masa Depan Kecerdasan Buatan (AI) dan Dampaknya Terhadap Masyarakat," *Acad. Soc. Sci. Glob. Citizsh. J.*, vol. 3, no. 1, pp. 9–13, 2023, doi: 10.47200/aossagcj.v3i1.1837.

[6] I. M. Faiza, G. Gunawan, and W. Andriani, "Tinjauan Pustaka Sistematis: Penerapan Metode *Machine learning* untuk Deteksi Bencana Banjir," *J. Minfo Polgan*, vol. 11, no. 2, pp. 59–63, 2022, doi: 10.33395/jmp.v11i2.11657.

[7] M. Rizki *et al.*, "Aplikasi Metode Kano Dalam Menganalisis Sistem Pelayanan Online Akademik FST UIN SUSKA Riau pada masa Pandemi Covid-19," *Ejournal.Uin-Suska.Ac.Id*, vol. 18, no. 02, pp. 180–187, 2021, [Online]. Available: http://ejournal.uin-suska.ac.id/index.php/sitekin/article/view/12710

[8] M. T. Anwar, L. Heriyanto, and F. Fanini, "Model Prediksi *Dropout* Mahasiswa Menggunakan Teknik Data Mining," *J. Inform. Upgris*, vol. 7, no. 1, pp. 56–60, 2021, doi: 10.26877/jiu.v7i1.8023.

[9] M. Arif and A. D. Hartono, "Jurnal Informatika : Jurnal pengembangan IT Menggunakan Metode *Machine learning* Untuk Memprediksi Nilai Mahasiswa Dengan Model Prediksi Multiclass," vol. 10, no. 1, pp. 190–204, 2025, doi: 10.30591/jpit.v9ix.xxx.

[10] M. A. Harriz, "Implementasi Random Forest dan SMOTE untuk Prediksi Risiko Putus Sekolah Dasar Menuju Indonesia Emas 2045," *J. Indones. Sch. forS ocial Res.*, vol. 5, no. 1, pp. 97–106, 2025.

[11] N. Sitohang, "Penerapan Data Mining Untuk Peringatan Dini Banjir Menggunakan Metode Klastering K-Means," *J. Sains Inform. Terap. (JSIT)*, vol. 2, no. 1, pp. 16–20, 2023.

[12] Armansyah and Suhardi, "MODEL REGRESI LOGISTIK DAN JARINGAN SYARAF TIRUAN UNTUK KLASIFIKASI MAHASISWA BERPOTENSI DROP OUTa," *urnal Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 13, no. 1, pp. 8–16, 2025.

[13] U. Kristen, K. Wacana, and J. Barat, "PREDIKSI MAHASISWA DROP-OUT DI UNIVERSITAS XYZ PREDICTION OF STUDENT DROP-OUT AT XYZ UNIVERSITY," vol. 11, no. 6, pp. 1345–1350, 2024, doi: 10.25126/jtiik.2024118689.

[14] H. Abijono, P. Santoso, and N. L. Anggreini, "Algoritma Supervised Learning Dan Unsupervised Learning Dalam Pengolahan Data," *J. Teknol. Terap. G-Tech*, vol. 4, no. 2, pp. 315–318, 2021, doi: 10.33379/gtech.v4i2.635.

[15] A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree," *JATISI (Jurnal Tek. Inform.*

dan Sist. Informasi), vol. 9, no. 2, pp. 1109–1119, 2022, doi: 10.35957/jatisi.v9i2.2048.

[16] Rayuwati, Husna Gemasih, and Irma Nizar, "IMPLEMENTASI AlGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.

[17] Nikmatun, I. Alvi, Waspada, and Indra, "Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.

[18] M. Iqbal, H. Destiana, A. Hamid, U. Bina, and S. Informatika, "Data Mining Berbasis *Machine learning* Untuk Analitik Prediktif Dalam Kelulusan," vol. 10, no. 2, pp. 1–10, 2024.

[19] I. S. Phillnov Yohanes Pinontoan, "Implementasi dan Analisis Deteksi Serangan Jaringan Pada Web Server NFT Menggunakan Suricata," *EduTIK J. Pendidik. Teknol. Inf. dan Komun.*, vol. 4, pp. 65–78, 2024.

[20] A. Fajri, N. H. Safaat, and M. Affandes, "Analisis Manajemen Risiko TI Menggunakan Framework COBIT 5 Domain APO12 dan EDM03," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 3, pp. 1523–1530, 2023, doi: 10.30865/klik.v4i3.1396.

[21] S. Dwi Putra, H. Herman, and A. Yudhana, "Evaluasi Tata Kelola Layanan Jaringan Menggunakan COBIT 2019 Pada Sekolah Tinggi Ilmu Kesehatan," *Resist. (Elektronika Kendali Telekomun. Tenaga List. Komputer)*, vol. 5, no. 2, p. 119, 2022, doi: 10.24853/resistor.5.2.119-126.

[22] V. H. Pranatawijaya, W. Widiatry, R. Priskila, and P. B. A. A. Putra, "Penerapan Skala Likert dan Skala Dikotomi Pada Kuesioner Online," *J. Sains dan Inform.*, vol. 5, no. 2, p. 128, 2019, doi: 10.34128/jsi.v5i2.185.