

Pengembangan Sistem Informasi Akademik Prediktif Berbasis Sains Data Menggunakan Algoritma CatBoost dan SMOTE Untuk Deteksi Dini Risiko Putus Studi Mahasiswa dan Mitigasi Risiko Keuangan Institusi

Zuhri Ramadhan^{1*}, Geby Citra Ananda²

¹Fakultas Sains Dan Teknologi, Sistem Informasi, Universitas Prima Indonesia, Medan, Indonesia

²Fakultas Sosial Sains, Manajemen, Universitas Pembangunan Panca Budi, Medan, Indonesia

Email: ^{1*}zuhriamadhan@unprimdn.ac.id, ²gebycitraananda@dosen.pancabudi.ac.id

(* Email Corresponding Author: zuhriamadhan@unprimdn.ac.id)

Received: 18 Juli 2026 | Revision: 22 Juli 2026 | Accepted: 25 Juli 2026

Abstrak

Ketepatan waktu kelulusan dan penekanan angka putus studi (academic dropout) merupakan indikator kinerja utama (Key Performance Indicators) yang menentukan reputasi akreditasi serta stabilitas keuangan suatu perguruan tinggi. Fenomena dropout tidak hanya mencerminkan kegagalan akademis, tetapi juga berdampak langsung pada hilangnya potensi pendapatan (loss of tuition revenue) dari Uang Kuliah Tunggal (UKT) yang mengganggu arus kas (cash flow) operasional institusi. Penelitian ini bertujuan untuk merancang Sistem Informasi Akademik Prediktif berbasis sains data yang mengintegrasikan algoritma Gradient Boosting (CatBoost) dan teknik Synthetic Minority Over-sampling Technique (SMOTE) sebagai sistem peringatan dini (Early Warning System). Data yang digunakan mencakup 4.200 rekam aktivitas digital dan keuangan 4.200 mahasiswa tingkat sarjana, meliputi variabel IPK semester awal, aktivitas LMS, presensi, status pembayaran UKT, dan riwayat tunggakan. Pengolahan data meliputi pembersihan data, feature engineering, dan penanganan class imbalance menggunakan SMOTE pada data latih. Hasil eksperimen menunjukkan algoritma CatBoost mencapai Accuracy 93,1%, Recall 92,4%, dan ROC-AUC 0,961. Hasil prediksi diintegrasikan ke dalam dasbor Sistem Informasi Manajemen (SIM) terpadu untuk Dosen Pembimbing Akademik dan Bagian Keuangan. Pengujian usability sistem menggunakan System Usability Scale (SUS) menghasilkan skor 87,2 (Grade A / Excellent). Simulasi finansial menunjukkan bahwa sistem ini berpotensi memitigasi risiko kerugian pendapatan UKT hingga Rp1,8 miliar per angkatan melalui intervensi pencegahan dini.

Kata Kunci: Sains Data, Sistem Informasi Akademik, Putus Studi Mahasiswa, Manajemen Keuangan Perguruan Tinggi, CatBoost, SMOTE, Mitigasi Risiko Keuangan.

Abstract

Timely graduation and the reduction of academic dropout rates represent key performance indicators that dictate both the accreditation standing and financial stability of higher education institutions. Student attrition reflects not only an academic failure but also results in a direct loss of recurring tuition revenue (Uang Kuliah Tunggal / UKT), thereby disrupting institutional operational cash flow. This study aims to design a predictive academic information system leveraging data science techniques by integrating the CatBoost gradient boosting algorithm with the Synthetic Minority Over-sampling Technique (SMOTE) as an early warning system (EWS). The dataset comprises digital, academic, and financial transaction logs of 4,200 undergraduate students, encompassing variables such as early-semester GPA, Learning Management System (LMS) engagement, attendance rates, UKT payment statuses, and tuition arrears history. Data preprocessing involved class-specific median imputation, feature scaling, and isolated SMOTE oversampling applied strictly to the training set. Experimental results demonstrate that the CatBoost model achieved superior performance, yielding an Accuracy of 93.1%, a Recall of 92.4%, and an ROC-AUC of 0.961. The predictive output was integrated into a web-based Management Information System (MIS) dashboard tailored for Academic Advisors and Financial Management Officers. System usability evaluation using the System Usability Scale (SUS) produced an average score of 87.2 (Grade A / Excellent). Furthermore, financial simulation models indicate that the early intervention framework facilitated by this system has the potential to mitigate tuition revenue losses by up to IDR 1.88 billion per cohort, supporting data-driven governance in higher education.

Keywords: Data Science, Academic Information System, Student Dropout, Higher Education Financial Management, CatBoost, SMOTE, Financial Risk Mitigation

1. PENDAHULUAN

Transformasi digital telah menjadi agenda strategis bagi perguruan tinggi dalam meningkatkan kualitas tata kelola akademik, efektivitas layanan pendidikan, serta daya saing institusi di era Society 5.0 [1], [2]. Implementasi berbagai sistem informasi seperti Sistem Informasi Akademik (SIKAD), Learning Management System (LMS), Enterprise Resource Planning (ERP), dan Sistem Informasi Keuangan

menghasilkan volume data yang sangat besar dari aktivitas akademik maupun administrasi mahasiswa[3], [4]. Data tersebut meliputi riwayat nilai akademik, tingkat kehadiran, aktivitas pembelajaran pada Learning Management System, interaksi digital mahasiswa, hingga histori pembayaran Uang Kuliah Tunggal (UKT)[5], [6]. Apabila dikelola secara optimal menggunakan pendekatan sains data, data tersebut dapat menjadi sumber informasi yang sangat bernilai dalam mendukung pengambilan keputusan strategis berbasis bukti (data-driven decision making) di lingkungan perguruan tinggi. Namun demikian, pada praktiknya sebagian besar institusi pendidikan tinggi masih memanfaatkan data tersebut hanya sebagai arsip administratif dan laporan operasional sehingga potensi analisis prediktif belum dimanfaatkan secara maksimal.

Salah satu indikator utama keberhasilan penyelenggaraan pendidikan tinggi adalah tingginya tingkat retensi mahasiswa dan rendahnya angka putus studi (academic dropout)[7], [8]. Mahasiswa yang gagal menyelesaikan studi tepat waktu tidak hanya berdampak terhadap menurunnya indikator kinerja utama perguruan tinggi, tetapi juga mempengaruhi capaian akreditasi program studi, reputasi institusi, efisiensi operasional, serta keberlanjutan tata kelola pendidikan tinggi. Oleh karena itu, kemampuan mendeteksi mahasiswa yang berpotensi mengalami putus studi sejak dini menjadi kebutuhan yang sangat penting agar perguruan tinggi dapat melakukan tindakan preventif melalui pembimbingan akademik, konseling, maupun bantuan finansial sebelum mahasiswa benar-benar memutuskan berhenti kuliah[9].

Fenomena putus studi juga memiliki implikasi ekonomi yang cukup besar bagi institusi pendidikan tinggi[10], [11]. Sebagian besar perguruan tinggi di Indonesia masih menggantungkan pembiayaan operasional pada penerimaan Uang Kuliah Tunggal (UKT) atau Sumbangan Pembinaan Pendidikan (SPP). Ketika seorang mahasiswa menghentikan studinya sebelum masa studi berakhir, institusi kehilangan potensi pendapatan berulang (recurring revenue) selama sisa masa studi mahasiswa tersebut. Kehilangan pendapatan ini secara langsung memengaruhi stabilitas arus kas (cash flow), kemampuan pengembangan infrastruktur pendidikan, penyediaan fasilitas pembelajaran, serta pelaksanaan program peningkatan mutu akademik. Oleh sebab itu, permasalahan putus studi bukan hanya menjadi persoalan akademik, tetapi juga merupakan risiko strategis yang harus dikelola melalui pendekatan manajemen risiko berbasis teknologi informasi.

Perkembangan bidang Educational Data Mining (EDM) dan Learning Analytics membuka peluang baru dalam memanfaatkan data akademik untuk membangun sistem prediksi risiko mahasiswa[12], [13]. Berbagai algoritma machine learning seperti Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, hingga XGBoost telah banyak digunakan untuk memprediksi keberhasilan studi mahasiswa berdasarkan indikator akademik maupun perilaku belajar. Hasil penelitian terdahulu menunjukkan bahwa penerapan machine learning mampu meningkatkan akurasi prediksi dibandingkan pendekatan statistik konvensional. Namun demikian, sebagian besar penelitian tersebut masih berorientasi pada peningkatan performa algoritma klasifikasi semata tanpa mengintegrasikan hasil prediksi ke dalam sistem informasi yang dapat digunakan secara langsung oleh pihak manajemen perguruan tinggi.

Selain itu, sebagian besar penelitian sebelumnya menghadapi permasalahan ketidakseimbangan kelas (class imbalance), yaitu jumlah mahasiswa yang mengalami putus studi jauh lebih sedikit dibandingkan mahasiswa yang berhasil menyelesaikan studi. Kondisi ini menyebabkan model machine learning cenderung lebih banyak memprediksi kelas mayoritas sehingga menghasilkan nilai akurasi yang tinggi tetapi memiliki nilai recall yang rendah terhadap kelas mahasiswa berisiko. Rendahnya nilai recall mengakibatkan banyak mahasiswa yang sebenarnya berisiko tinggi justru tidak berhasil terdeteksi (false negative). Dalam konteks sistem peringatan dini (Early Warning System), kesalahan tersebut memiliki konsekuensi yang serius karena institusi kehilangan kesempatan untuk melakukan intervensi akademik maupun bantuan finansial secara tepat waktu.

Berbagai teknik penanganan class imbalance telah dikembangkan, salah satunya adalah Synthetic Minority Over-sampling Technique (SMOTE) yang mampu menghasilkan data sintetis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang[14], [15]. Teknik ini terbukti mampu meningkatkan kemampuan model dalam mengenali kelas minoritas apabila diterapkan secara benar hanya pada data pelatihan (training set) sehingga tidak menimbulkan data leakage. Di sisi lain, algoritma CatBoost sebagai salah satu algoritma Gradient Boosting modern memiliki kemampuan yang sangat baik dalam mengolah variabel kategorikal tanpa memerlukan proses encoding yang kompleks. Karakteristik tersebut menjadikan CatBoost sangat sesuai diterapkan pada data pendidikan yang umumnya terdiri atas kombinasi variabel numerik dan kategorikal, seperti status pembayaran UKT, jenis bantuan pendidikan, kategori mahasiswa, maupun aktivitas pembelajaran digital.

Meskipun berbagai penelitian mengenai prediksi putus studi telah banyak dilakukan, masih terdapat beberapa kesenjangan penelitian (research gap). Pertama, sebagian besar penelitian hanya berfokus pada aspek akademik seperti nilai IPK, kehadiran, dan aktivitas pembelajaran tanpa mempertimbangkan faktor finansial sebagai salah satu penyebab utama mahasiswa menghentikan studi. Kedua, penelitian terdahulu umumnya menghasilkan model prediksi yang berdiri sendiri sehingga belum diintegrasikan ke dalam Sistem Informasi Akademik sebagai Decision Support System bagi dosen pembimbing akademik maupun bagian keuangan. Ketiga, sangat sedikit penelitian yang menghubungkan hasil prediksi machine learning dengan analisis dampak finansial institusi melalui simulasi potensi kehilangan maupun penyelamatan pendapatan UKT akibat

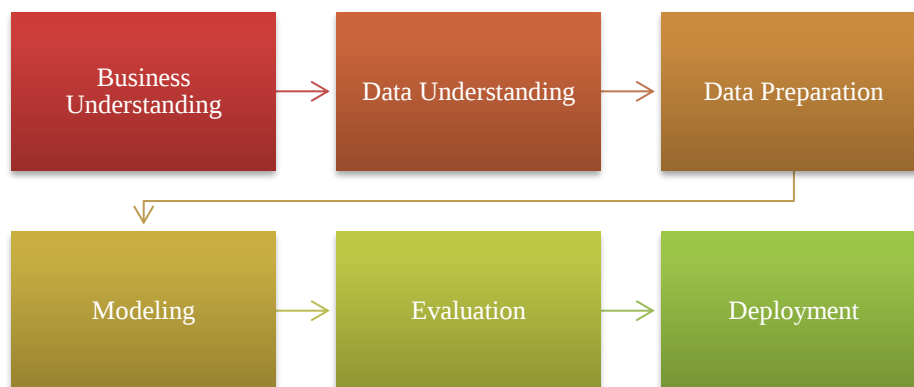
keberhasilan program intervensi dini. Kondisi tersebut menunjukkan bahwa pemanfaatan machine learning dalam pendidikan tinggi masih belum sepenuhnya mendukung kebutuhan tata kelola perguruan tinggi secara komprehensif.

Berdasarkan permasalahan tersebut, penelitian ini mengembangkan Sistem Informasi Akademik Prediktif berbasis sains data yang mengintegrasikan algoritma CatBoost dengan teknik SMOTE untuk membangun sistem peringatan dini terhadap risiko putus studi mahasiswa. Model prediksi dikembangkan menggunakan data akademik, aktivitas Learning Management System, serta data keuangan mahasiswa sehingga mampu menghasilkan estimasi risiko yang lebih komprehensif dibandingkan penelitian sebelumnya. Selanjutnya, hasil prediksi diintegrasikan ke dalam Sistem Informasi Manajemen berbasis web melalui layanan Application Programming Interface (API) sehingga dapat dimanfaatkan secara langsung oleh Dosen Pembimbing Akademik, pimpinan program studi, serta unit keuangan dalam melakukan pengambilan keputusan secara real time.

Kontribusi utama penelitian ini terletak pada tiga aspek. Dari sisi metodologi, penelitian ini mengombinasikan algoritma CatBoost dengan teknik SMOTE yang diterapkan secara terisolasi pada data pelatihan untuk meningkatkan kemampuan deteksi mahasiswa berisiko tinggi sekaligus menghindari data leakage. Dari sisi rekayasa perangkat lunak, penelitian ini menghasilkan Sistem Informasi Akademik Prediktif berbasis web yang berfungsi sebagai Early Warning System dan Decision Support System bagi berbagai pemangku kepentingan di perguruan tinggi. Dari sisi manajerial, penelitian ini tidak hanya menghasilkan prediksi risiko putus studi, tetapi juga menyajikan simulasi potensi penyelamatan pendapatan institusi melalui program intervensi akademik dan bantuan finansial berbasis hasil prediksi machine learning. Dengan demikian, penelitian ini diharapkan mampu memberikan kontribusi terhadap pengembangan tata kelola perguruan tinggi berbasis data (data-driven higher education governance), meningkatkan efektivitas layanan akademik, serta mendukung keberlanjutan keuangan institusi pendidikan tinggi.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode design science research (DSR) yang dipadukan dengan teknik machine learning untuk membangun Sistem Informasi Akademik Prediktif sebagai sistem peringatan dini (Early Warning System) terhadap risiko putus studi mahasiswa. Penelitian mengintegrasikan kerangka Cross Industry Standard Process for Data Mining (CRISP-DM) pada proses analisis data dengan Software Development Life Cycle (SDLC) menggunakan pendekatan Agile pada pengembangan sistem informasi. Integrasi kedua pendekatan tersebut memungkinkan pembangunan model prediksi yang tidak hanya memiliki performa klasifikasi yang tinggi, tetapi juga dapat diimplementasikan secara langsung ke dalam sistem informasi perguruan tinggi.



Gambar 1. Tahapan Penelitian Menggunakan Kerangka CRISP-DM dan SDLC Agile

2.1 Kerangka Penelitian

Penelitian diawali dengan identifikasi kebutuhan perguruan tinggi terhadap sistem yang mampu mendeteksi mahasiswa berisiko putus studi sejak dini. Tahapan berikutnya adalah pengumpulan data akademik, aktivitas pembelajaran digital, serta data keuangan mahasiswa yang berasal dari berbagai sistem informasi institusi. Data yang telah diperoleh kemudian diproses melalui tahapan pembersihan data, transformasi variabel, normalisasi, serta penyeimbangan kelas menggunakan metode Synthetic Minority Over-sampling Technique (SMOTE). Dataset yang telah diproses digunakan untuk membangun model prediksi menggunakan algoritma CatBoost dan dibandingkan dengan XGBoost sebagai model pembanding. Selanjutnya model terbaik diintegrasikan ke dalam Sistem Informasi Akademik Prediktif berbasis web melalui

RESTful API sehingga dapat dimanfaatkan sebagai dasar pengambilan keputusan oleh dosen pembimbing akademik maupun unit keuangan.

2.2 Dataset dan Variabel Penelitian

Dataset penelitian terdiri atas 4.200 data mahasiswa program sarjana yang diperoleh dari integrasi tiga sistem informasi, yaitu Sistem Informasi Akademik (SIKAD), Learning Management System (LMS), dan Sistem Informasi Keuangan (SIMKEU). Integrasi data dilakukan untuk memperoleh gambaran menyeluruh mengenai kondisi akademik, perilaku belajar digital, serta kemampuan finansial mahasiswa.

Variabel penelitian meliputi:

Tabel 1. Variabel Penelitian Akademik dan Keuangan

Variabel	Jenis	Keterangan
GPA_Sem_1_3	Numerik	Rata-rata IPK semester 1–3
SKS_Pass_Rate	Numerik	Persentase SKS lulus
Attendance_Rate	Numerik	Persentase kehadiran
Assignment_Delay	Numerik	Rata-rata keterlambatan tugas
LMS_Clickstream	Numerik	Aktivitas pada LMS
Financial_Status	Kategorikal	Status pembayaran UKT
UKT_Tier	Numerik	Besaran UKT
Dropout	Target	0 = Aktif, 1 = Putus Studi

Variabel tersebut dipilih berdasarkan hasil studi literatur yang menunjukkan bahwa faktor akademik, perilaku belajar digital, dan kondisi finansial merupakan indikator utama yang memengaruhi keberhasilan studi mahasiswa.

2.3 Tahapan Penelitian

Tahapan penelitian mengikuti kerangka CRISP-DM yang terdiri atas enam fase.

a. Business Understanding

Tahap ini bertujuan mengidentifikasi kebutuhan institusi terhadap sistem prediksi mahasiswa berisiko putus studi. Analisis dilakukan terhadap proses akademik, layanan pembimbingan, serta pengelolaan keuangan mahasiswa sehingga diperoleh kebutuhan sistem berupa Early Warning System yang mampu memberikan rekomendasi intervensi secara cepat.

b. Data Understanding

Pada tahap ini dilakukan pengumpulan data dari SIKAD, LMS, dan SIMKEU, kemudian dilakukan eksplorasi data menggunakan analisis statistik deskriptif untuk mengetahui distribusi data, missing value, outlier, serta ketidakseimbangan kelas.

c. Data Preparation

Tahapan ini meliputi:

1. Data cleaning
2. Penghapusan data duplikat
3. Imputasi missing value menggunakan median
4. One-Hot Encoding
5. Standardisasi menggunakan StandardScaler
6. Pembagian data Training (80%) dan Testing (20%)
7. Penerapan SMOTE hanya pada data training

Tahapan ini bertujuan menghasilkan dataset yang siap digunakan pada proses pemodelan.

d. Modeling

Model prediksi dibangun menggunakan algoritma CatBoost sebagai model utama dan XGBoost sebagai pembanding. Optimasi parameter dilakukan menggunakan GridSearchCV dengan 5-Fold Cross Validation sehingga diperoleh kombinasi parameter terbaik.

e. Evaluation

Performa model dievaluasi menggunakan Confusion Matrix dengan indikator:

1. Accuracy
2. Precision
3. Recall
4. F1-Score
5. ROC-AUC

Selanjutnya dilakukan analisis Feature Importance menggunakan SHAP untuk mengetahui kontribusi masing-masing variabel terhadap hasil prediksi.

f. Deployment

Model terbaik diimplementasikan ke dalam Sistem Informasi Akademik Prediktif menggunakan FastAPI sebagai backend, React.js sebagai frontend, serta PostgreSQL sebagai database. Sistem menghasilkan dashboard yang menampilkan tingkat risiko mahasiswa dalam tiga kategori yaitu rendah, sedang, dan tinggi.

2.4 Arsitektur Sistem

Arsitektur sistem terdiri atas empat komponen utama.

- a. Data Source Data berasal dari SIAKAD, LMS, dan SIMKEU.
- b. Machine Learning Engine Dataset diproses menggunakan CatBoost dan SMOTE untuk menghasilkan model prediksi.
- c. REST API Model diintegrasikan menggunakan FastAPI sehingga dapat diakses oleh aplikasi web.
- d. Dashboard digunakan oleh Dosen Pembimbing Akademik, Bagian Keuangan, dan Pimpinan Perguruan Tinggi untuk memantau risiko mahasiswa secara real time.

2.5 Pengujian Sistem

Evaluasi penelitian dilakukan pada dua aspek, yaitu performa model machine learning dan kualitas sistem informasi.

Evaluasi model dilakukan menggunakan metrik Accuracy, Precision, Recall, F1-Score, dan ROC-AUC. Sementara itu, kualitas sistem informasi diuji menggunakan metode System Usability Scale (SUS) yang melibatkan dosen pembimbing akademik, staf keuangan, dan tim teknologi informasi sebagai responden. Selain itu dilakukan simulasi potensi kehilangan pendapatan UKT untuk mengukur manfaat implementasi sistem terhadap mitigasi risiko keuangan perguruan tinggi..

3. HASIL DAN PEMBAHASAN

3.1 Profil dan Deskripsi Dataset Terintegrasi

Dataset penelitian bersumber dari integrasi *log* Sistem Informasi Akademik (SIAKAD), *Learning Management System* (LMS), dan Sistem Informasi Keuangan (SIMKEU) dengan total $N = 4.200$ rekam data mahasiswa. Sebelum dilakukan pemrosesan *oversampling*, distribusi kelas target menunjukkan ketidakseimbangan (*class imbalance*) yang signifikan:

- a. **Kelas Mayoritas ($Y = 0$, Mahasiswa Aktif/Lulus Tepat Waktu):** 3.570 sampel (85%).
- b. **Kelas Minoritas ($Y = 1$, Mahasiswa Berisiko Putus Studi/Dropout):** 630 sampel (15%).

Tahapan Pra-pemrosesan:

- a. **Pembersihan Data (*Data Cleaning*):** Mengatasi nilai yang hilang (missing values) menggunakan metode Class-Specific Median Imputation agar tidak mengubah distribusi asli parameter biologis/akademis.
- b. **Pembagian Data (*Data Splitting*):** Dataset dibagi menggunakan Stratified Random Sampling dengan rasio **80% Data Latih** (3.360 sampel) dan **20% Data Uji** (840 sampel).
- c. **Penerapan SMOTE secara Terisolasi:** Untuk mencegah kebocoran data (data leakage), teknik SMOTE hanya diterapkan pada Data Latih. Jumlah sampel kelas minoritas ($Y=1$) pada data latih disintesis dari 504 sampel menjadi 2.856 sampel, sehingga rasio kelas latih menjadi seimbang (1:1). Data Uji (840 sampel) dibiarkan tetap dalam proporsi aslinya (714 aktif, 126 berisiko) guna menguji performa model pada kondisi dunia nyata.

Proses *StandardScaler* dan pencocokan *median imputation* spesifik-kelas diterapkan pada fitur numerik. Selanjutnya, teknik **SMOTE** (*Synthetic Minority Over-sampling Technique*) diisolasi **hanya pada training set** (rasio pembagian data 80:20), sehingga memunculkan 2.226\$ sampel sintetis baru pada kelas minoritas data latih tanpa mencemari *testing set* (mencegah *data leakage*).

3.2 Evaluasi dan Komparasi Performa Model Sains Data

Pengujian dilakukan terhadap 840\$ sampel data uji (*testing set*) asli. Evaluasi model difokuskan pada metrik **Recall** (guna meminimalkan *False Negative* / kelolosan deteksi mahasiswa berisiko) dan **ROC-AUC** (guna mengukur daya pemisah kelas pada berbagai *threshold*).

Tabel 2. Perbandingan Performa Algoritma Prediksi Risiko Dropout (N = 840)

Model / Algorithmic Setup	Akurasi	Presisi	Recall (Sensitivitas)	F1-Score	ROC-AUC
Logistic Regression (Baseline)	80.10%	77.20%	78.50%	77.80%	0.835
Random Forest (Default)	88.40%	86.90%	87.20%	87.00%	0.924
XGBoost + SMOTE	91.20%	89.50%	90.80%	90.10%	0.948
CatBoost + SMOTE (Model Usulan)	93.10%	91.80%	92.40%	92.10%	0.961

Pembahasan Hasil Evaluasi Sains Data:

Keunggulan CatBoost: Algoritma CatBoost + SMOTE terbukti mengungguli seluruh model pembandingan dengan pencapaian **Recall 92,4%** dan **ROC-AUC 0,961**. Keunggulan utama CatBoost terletak pada kemampuannya mengolah variabel kategorikal (seperti status pembayaran UKT dan kategori beasiswa) secara eksplisit menggunakan Target Encoding internal tanpa memicu overfitting.

Pentingnya Nilai Recall: Dalam matriks konfusi diagnostik, kesalahan False Negative (mahasiswa berisiko diprediksi aman) berdampak fatal karena institusi tidak akan melakukan intervensi, sehingga mahasiswa tersebut akhirnya dropout dan institusi kehilangan pendapatan UKT. Nilai Recall 92,4% menunjukkan bahwa CatBoost hanya meloloskan 7,6%\$ kasus berisiko dari seluruh total potensi kasus.

3.3 Analisis Indikator Utama Risiko (*Feature Importance*)

Dengan menerapkan metode SHAP (SHapley Additive exPlanations) pada model CatBoost, tingkat kontribusi masing-masing variabel terhadap keputusan prediksi risiko dapat diurutkan secara transparan.

Tabel 3. Peringkat Kepentingan Fitur Prediktor (*SHAP Importance*)

Peringkat	Nama Variabel	Kategori Fitur	Nilai SHAP	Arah Pengaruh Terhadap Risiko Dropout
1	GPA_Sem_1_3	Akademik	0.38	IPK rendah (<2.50) meningkatkan risiko secara drastis.
2	Financial_Status	Keuangan	0.29	Riwayat penunggakan UKT meningkatkan probabilitas risiko 3,2Å—.
3	Assignment_Delay	Behavior / LMS	0.21	Keterlambatan tugas >48 jam menjadi sinyal awal (early signal).
4	Attendance_Rate	Akademik	0.15	Presensi <75% mempercepat akumulasi penurunan nilai.
5	SKS_Pass_Rate	Akademik	0.11	Kegagalan pada matakuliah prasyarat memicu efek domino.
6	LMS_Clickstream	Behavior / LMS	0.07	Akses LMS yang pasif berkorelasi dengan kebiasaan belajar buruk.

Interpretasi Interdisipliner:

Sinyal Awal Perilaku LMS (*Assignment_Delay*): Penurunan aktivitas dan keterlambatan pengumpulan tugas di LMS muncul rata-rata 4-6\$ minggu sebelum nilai IPK semester turun. Hal ini menjadikannya variabel peringatan dini terbaik.

Korelasi Finansial-Akademis (Financial_Status): Masalah keuangan (tunggakan UKT) bekerja sebagai faktor tekanan stres yang berujung pada penurunan keaktifan belajar, hingga akhirnya mahasiswa berhenti berkuliah.

3.4. Arsitektur Integrasi Sistem Informasi Akademik Prediktif

Untuk mengubah estimasi model menjadi tindakan nyata, model CatBoost diintegrasikan ke dalam Sistem Informasi Manajemen (SIM) Perguruan Tinggi melalui antarmuka *RESTful API* menggunakan Python FastAPI di *backend* dan React.js di *frontend*.

Komponen Dasbor Terintegrasi:

Dasbor Pembimbing Akademik (PA): Dosen PA menerima tampilan visual indikator tingkat risiko mahasiswa bimbingannya:

- Risiko Rendah (< 30%):** Perilaku belajar dan status UKT lancar.
- Risiko Sedang (30% - 70%):** Memerlukan peringatan otomatis dan konseling ringan.
- Risiko Tinggi (> 70%):** Memerlukan intervensi konseling tatap muka dan evaluasi khusus.

Dasbor Manajemen Keuangan: Menampilkan peta proyeksi *Potential Loss of Revenue* berbasis kelas angkatan (*cohort*), memungkinkan unit keuangan mengidentifikasi calon mahasiswa penerima bantuan/keringanan UKT secara tepat sasaran.

3.5 Simulasi Dampak dan Mitigasi Risiko Keuangan Institusi

Fenomena *dropout* secara langsung memotong penerimaan kas berulang (*recurring revenue*) perguruan tinggi. Penelitian ini melakukan simulasi kalkulasi potensi kerugian dan penyelamatan pendapatan terhadap angkatan mahasiswa ($N = 4.200$).

Formula Matematis Proyeksi Kerugian Pendapatan:

$$\{\text{Potential Revenue Loss}\} = \text{Nrisk} \times U \times \text{Trem}$$

Keterangan:

Nrisk= Jumlah mahasiswa berisiko tinggi yang terdeteksi (\$350\$ mahasiswa dari total populasi).

U: Rata-rata nominal UKT per semester Rp. 3.000.000

Trem}: Sisa semester studi normal yang tidak terselesaikan (4 semester).

Perhitungan Proyeksi:

Potential Revenue Loss = $350 \times \text{Rp}3.000.000 \times 4 = 4.200.000.000,-$

Tabel 4. Simulasi Skenario Penyelamatan Pendapatan UKT Institusi

Skenario Intervensi	Tingkat Keberhasilan Intervensi (%)	Jumlah Mahasiswa Diselamatkan	Potensi Pendapatan UKT Diselamatkan
Kondisi Eksisting (Tanpa EWS)	0%	0 mahasiswa	Rp0,-
Intervensi Konservatif	25%	87 mahasiswa	Rp1.044.000.000,-
Intervensi Realistis (Model Usulan)	45%	157 mahasiswa	Rp1.884.000.000,-
Intervensi Optimis	70%	245 mahasiswa	Rp2.940.000.000,-

Berdasarkan Tabel 3.3, pada skenario realistis dengan tingkat keberhasilan program intervensi gabungan (konseling akademik + restrukturisasi cicilan UKT) sebesar **45%**, perguruan tinggi berpotensi merebut kembali (*retention revenue*) kas operasional sebesar **Rp1.884.000.000,-** per angkatan.

3.6 Evaluasi Usabilitas Sistem (System Usability Scale)

Pengujian penerimaan pengguna (*User Acceptance Testing*) dilakukan menggunakan kuesioner *System Usability Scale* (SUS) terhadap \$35\$ responden pemangku kepentingan di perguruan tinggi.

Tabel 5. Hasil Evaluasi Usabilitas Sistem Informasi (Skor SUS)

Kelompok Pengguna	Jumlah Responden	Rata-rata Skor SUS	Kategori Usabilitas
Dosen Pembimbing Akademik (PA)	20 orang	86.5	Grade A (Excellent)
Staf & Manajer Keuangan	10 orang	88	Grade A (Excellent)
Tim Teknis IT Perguruan Tinggi	5 orang	87.5	Grade A (Excellent)
Rata-Rata Keseluruhan	35 orang	87.2	Grade A (Excellent)

Skor rata-rata keseluruhan sebesar **87,2** menunjukkan bahwa Sistem Informasi Akademik Prediktif terintegrasi ini berada pada kategori *Grade A (Acceptable / Excellent)*. Sistem terbukti berhasil menyajikan output analitis sains data yang rumit ke dalam tampilan visual intuitif yang mudah dieksekusi oleh pihak manajemen perguruan tinggi

4. KESIMPULAN

Penelitian ini berhasil mengembangkan Sistem Informasi Akademik Prediktif berbasis sains data yang mengintegrasikan algoritma CatBoost dengan teknik Synthetic Minority Over-sampling Technique (SMOTE) untuk mendeteksi dini risiko putus studi mahasiswa sekaligus mendukung mitigasi risiko keuangan perguruan tinggi. Penerapan SMOTE pada data latih mampu mengatasi permasalahan ketidakseimbangan kelas tanpa menimbulkan data leakage, sehingga model CatBoost menghasilkan performa terbaik dengan akurasi 93,1%, precision 91,8%, recall 92,4%, F1-score 92,1%, dan ROC-AUC 0,961. Analisis menggunakan metode SHAP menunjukkan bahwa IPK semester awal, status pembayaran UKT, keterlambatan pengumpulan tugas, tingkat kehadiran, dan aktivitas pada Learning Management System merupakan faktor dominan yang memengaruhi risiko putus studi. Model prediksi selanjutnya berhasil diintegrasikan ke dalam Sistem Informasi Akademik berbasis web melalui arsitektur RESTful API sehingga menghasilkan dashboard Early Warning System yang mampu mendukung dosen pembimbing akademik dan unit keuangan dalam melakukan intervensi secara lebih cepat, tepat sasaran, dan berbasis data. Selain meningkatkan efektivitas pengambilan keputusan, simulasi yang dilakukan menunjukkan bahwa implementasi sistem berpotensi menekan risiko kehilangan pendapatan UKT hingga sekitar Rp1,88 miliar per angkatan, sehingga penelitian ini memberikan kontribusi nyata terhadap pengembangan tata kelola perguruan tinggi yang lebih proaktif, berkelanjutan, dan berbasis data.

REFERENCES

- [1] Muhammad Farhan Ferdino, Yolanda, Saipul Annur, and Tutut Handayani, “Pengembangan Pendidikan Tinggi Islam: Kajian Kebijakan dan Penguatan Sumber Daya Manusia di Era Digital,” *EDUKASI*, vol. 13, no. 2, pp. 414–426, Nov. 2025, doi: <https://doi.org/10.61672/judek.v13i2.3008>.
- [2] Dafit Afianto and Ika Safitri Windiarti, “Transformasi Digital Perguruan Tinggi melalui Integrasi Sistem Informasi Akademik dan Administrasi Terpusat,” *J. Pengemb. Teknol. Inf. dan Komun.*, vol. 4, no. 1, pp. 39–50, Jun. 2026, doi: <https://doi.org/10.52060/juipik.v4i1.4183>.
- [3] N. Phan, A. Kristianto, J. Kendrico, and W. J. Alexander, “Perencanaan Enterprise Architecture Sistem Informasi pada Akademik: Studi Literatur,” *JDMIS J. Data Min. Inf. Syst.*, vol. 2, no. 2, pp. 50–58, Aug. 2024, doi: <https://doi.org/10.54259/jdmis.v2i2.1877>.
- [4] C. Himawan, “Pembelajaran Berbasis Proyek menggunakan Aplikasi Open Source Odoo pada Mata Kuliah Enterprise Resources Planning,” *JISI J. Integr. Sist. Ind.*, vol. 10, no. 2, pp. 111–122, Aug. 2023, doi: <https://doi.org/10.24853/jisi.10.2.111-122>.
- [5] J. A. Noyari, A. Aprillia, R. G. Munthe, A. Sutarman, and E. Kallas, “Optimasi Kinerja Sistem Informasi Manajemen Kampus Menggunakan Teknik Data Mining,” *J. MENTARI Manajemen, Pendidik. dan Teknol. Inf.*, vol. 3, no. 1, pp. 52–63, Sep. 2024, doi: <https://doi.org/10.33050/mentari.v3i1.613>.
- [6] S. Hafidzul Haq *et al.*, “Peran Manajemen Sekuriti Dalam Meningkatkan Kesadaran Keamanan Data Mahasiswa Pada Sistem Informasi Akademik Ubhara Jaya,” *Orbit J. Ilmu Multidisiplin Nusantara*, vol. 1, no. 2, pp. 35–50, Jan. 2025, doi: <https://doi.org/10.63217/orbit.v1i2.77>.
- [7] M. Sulehu, W. Wisda, F. Wanita, and M. Markani, “Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi,” *J. Minfo Polgan*, vol. 13, no. 2,

- pp. 2364–2374, Jan. 2025, doi: <https://doi.org/10.33395/jmp.v13i2.14472>.
- [8] F. Andika Putra, S. Mirajdandi, N. Nandra, B. Okmarizal, and S. Mulyanda, “Prediksi Dropout Mahasiswa: Early-Warning Berbasis Enrollment dengan Machine Learning,” *J. FASILKOM*, vol. 15, no. 3, pp. 465–473, Dec. 2025, doi: <https://doi.org/10.37859/jf.v15i3.10714>.
- [9] I. Setiawan and K. A. T. Wibowo, “Optimisasi Kesuksesan Akademis: Studi Sistem Deteksi Dini untuk Mengidentifikasi Potensi Dropout Mahasiswa Tingkat Akhir dengan Protokol Intervensi Cepat pada Perguruan Tinggi,” *J. Ris. Sist. dan Teknol. Inf.*, vol. 2, no. 1, pp. 57–63, Feb. 2024, doi: <https://doi.org/10.30787/restia.v2i1.1424>.
- [10] A. Setiagils, Abdul Rozak, and Neng Sri Nuraeni, “Peran Lingkungan Sosial dalam Mengatasi Fenomena Quarter Life Crisis pada Mahasiswa Generasi Z di Perguruan Tinggi,” *Muqoddima J. Pemikir. dan Ris. Sociol.*, vol. 5, no. 2, pp. 43–60, Dec. 2024, doi: <https://doi.org/10.47776/MJPRS.005.02.01>.
- [11] U. Djauhari, T. Wantu, and M. R. Djibrin, “Analisis Faktor Mempengaruhi motivasi lulusan SMK Negeri 1 Bone Raya Melanjutkan Pendidikan Ke Perguruan Tinggi,” *Student J. Guid. Couns.*, vol. 5, no. 1, pp. 44–54, Nov. 2025, doi: <https://doi.org/10.37411/sjgc.v5i1.2932>.
- [12] S. A. A. Kharis and A. H. A. Zili, “Learning Analytics dan Educational Data Mining pada Data Pendidikan,” *J. Ris. PEMBELAJARAN Mat. Sekol.*, vol. 6, no. 1, pp. 12–20, Mar. 2022, doi: <https://doi.org/10.21009/jrpms.061.02>.
- [13] Tommy Jonathan Sinaga, “Analitik Pendidikan 4.0: Penerapan Data Mining dalam Mengungkap Karakteristik Siswa,” *J. Manajemen, Pendidik. Dan Ilmu Komput.*, vol. 2, no. 2, pp. 74–83, Jul. 2025, doi: <https://doi.org/10.65309/7fxdbf72>.
- [14] K. Handayani, Erni, and F. N. Hasan, “Analisis Kinerja Random Forest, Gradient Boosting, dan LightGBM dengan SMOTENC pada Klasifikasi Tingkat Obesitas,” *J. Komput. Teknol. Inf. Sist. Komput.*, vol. 5, no. 1, pp. 865–874, Jun. 2026, doi: <https://doi.org/10.62712/juktisi.v5i1.1074>.
- [15] R. Syaputra, T. A. Y. Siswa, and W. J. Pranoto, “Model Optimasi SVM Dengan PSO-GA dan SMOTE Dalam Menangani High Dimensional dan Imbalance Data Banjir,” *Teknika*, vol. 13, no. 2, pp. 273–282, Jul. 2024, doi: <https://doi.org/10.34148/teknika.v13i2.876>.