Jurnal Minfo Polgan (JMP)
Volume 15, Nomor 2, Mei-Agustus 2026
DOI : https://doi.org/10.33395/jmp.v15i2.16571
e-ISSN : 2797-3298
p-ISSN : 2089-9424

Prediksi Risiko Akademik Taruna Menggunakan *Long Short-Term Memory* (LSTM) Berbasis Data *Longitudinal* pada Pendidikan Vokasi Pelayaran

[1]Nurul Ahyana, [2]Nurul Fadliana
[1]Prodi Nautika, Politeknik Pelayaran Barombong, Makassar, Indonesia
[2]Prodi Ketatalaksanaan Angkutan Laut dan Kepelabuhanan, Politeknik Ilmu Pelayaran Makassar, Makassar, Indonesia
*Korespondensi: nurulahyana@poltekpelbarombong.ac.id

Submit : 03 Jul 2026 | **Diterima** : 27 Jul 2026 | **Terbit** : 29 Jul 2026

ABSTRACT

Academic failure in maritime vocational education has more complex consequences than in general higher education, making early prediction of academic risk essential to support student guidance and intervention. This study aims to develop an academic risk prediction model based on longitudinal data using the Long Short-Term Memory (LSTM) method. Unlike previous studies that primarily rely on cross-sectional data, this research utilizes academic records from semesters 1–3 as sequential data to predict students' final academic status, which is classified into three categories: graduated, delayed graduation, and non-graduated. The dataset consists of 1,667 Diploma III cadets from Politeknik Pelayaran Barombong across three cohorts (2020–2022). To ensure prediction validity, the prediction horizon was limited to semesters 1–3, while the training and testing datasets were split at the individual cadet level to prevent data leakage. Class imbalance was addressed using class weighting, thereby preserving the longitudinal structure of the data. Before being applied to institutional data, the entire pipeline was validated end-to-end using synthetic data with an identical structure. The validation produced an accuracy of 0.34 and a macro ROC-AUC of 0.471, closely approximating random guessing as expected for randomly labeled data, thereby confirming the correctness of the implemented pipeline and the effectiveness of the data leakage prevention strategy. This study provides a methodological framework for longitudinal academic risk prediction that can serve as a foundation for developing decision support systems and Early Warning Systems in future research.

Keywords: *Academic risk prediction; Long Short-Term Memory (LSTM); longitudinal data; maritime vocational education; data leakage; class imbalance.*

ABSTRAK

Kegagalan akademik pada pendidikan vokasi pelayaran memiliki konsekuensi yang lebih kompleks dibandingkan pendidikan tinggi reguler sehingga prediksi risiko akademik sejak dini menjadi penting untuk mendukung pembinaan taruna. Penelitian ini bertujuan mengembangkan model prediksi risiko akademik berbasis data longitudinal menggunakan metode Long Short-Term Memory (LSTM). Berbeda dengan penelitian sebelumnya yang umumnya menggunakan data cross-sectional, penelitian ini memanfaatkan riwayat akademik semester 1–3 sebagai data berurutan (*sequence*) untuk memprediksi status akademik akhir taruna yang diklasifikasikan menjadi tiga kategori, yaitu lulus, lambat lulus, dan tidak lulus. Dataset mencakup 1.667 taruna Program Diploma III Politeknik Pelayaran Barombong dari tiga angkatan (2020–2022). Untuk menjaga validitas prediksi, horizon prediksi dibatasi pada semester 1–3, sedangkan pembagian data latih dan data uji dilakukan pada level individu taruna guna mencegah *data leakage*. Ketidakseimbangan kelas ditangani menggunakan *class weight* sehingga karakteristik data longitudinal tetap terjaga. Sebelum diterapkan pada data institusional, pipeline divalidasi secara *end-to-end* menggunakan data sintetis dengan struktur yang identik. Pengujian menghasilkan akurasi 0,34 dan nilai ROC-AUC makro 0,471, mendekati performa tebakan acak sebagaimana diharapkan pada data berlabel acak, sehingga memvalidasi implementasi pipeline dan mekanisme pencegahan *data leakage*. Penelitian ini menghasilkan kerangka metodologis prediksi risiko akademik berbasis data longitudinal yang dapat menjadi landasan

pengembangan sistem pendukung keputusan dan Early Warning System pada penelitian selanjutnya.

Kata kunci: Prediksi risiko akademik, Long Short-Term Memory (LSTM), data longitudinal, pendidikan vokasi pelayaran, data leakage, ketidakseimbangan kelas.

PENDAHULUAN

Pendidikan vokasi pelayaran memiliki karakteristik yang berbeda dari pendidikan tinggi pada umumnya. Selain dituntut menguasai kompetensi akademik, taruna juga harus memenuhi standar disiplin dan kehadiran yang ketat sebagaimana diatur dalam sistem pendidikan berasrama, sehingga kegagalan akademik, baik dalam bentuk keterlambatan kelulusan maupun putus studi, membawa konsekuensi yang lebih kompleks dibandingkan pendidikan tinggi reguler, mencakup kerugian waktu, biaya, serta dampak terhadap akreditasi program studi dan reputasi institusi. Politeknik Pelayaran (Poltekpel) Barombong, sebagai penyelenggara program Diploma 3 pada tiga program studi (Nautika, Permesinan Kapal, dan Manajemen Transportasi Laut), menghadapi tantangan serupa dalam memastikan taruna dapat menyelesaikan studi tepat waktu dan memenuhi standar kompetensi yang dipersyaratkan.

Upaya deteksi dini terhadap taruna berisiko gagal akademik secara konvensional umumnya masih bersifat reaktif, yaitu evaluasi dilakukan setelah nilai akhir semester keluar, sehingga ruang intervensi menjadi sempit. Perkembangan teknologi *machine learning*, khususnya arsitektur *deep learning* seperti *Long Short-Term Memory* (LSTM), membuka peluang untuk memprediksi risiko akademik taruna secara lebih dini dan akurat. Berbeda dari model klasifikasi konvensional yang memperlakukan data akademik sebagai titik tunggal, LSTM secara arsitektural dirancang untuk mempelajari pola *sequence* atau data deret waktu, sehingga mampu menangkap perkembangan dan perubahan performa akademik taruna dari semester ke semester sebagai satu kesatuan urutan, bukan sekadar nilai pada satu titik waktu.

Risiko akademik, sebagai kondisi ketika mahasiswa atau taruna menunjukkan indikasi awal potensi gagal menyelesaikan studi tepat waktu atau sesuai standar kompetensi yang dipersyaratkan, merupakan permasalahan yang berdampak luas, tidak hanya bagi individu yang bersangkutan tetapi juga bagi institusi penyelenggara pendidikan. Identifikasi dini terhadap mahasiswa yang berada dalam kondisi risiko akademik bersifat krusial, terutama pada bidang pendidikan yang berorientasi pada penyiapan tenaga kerja profesional, karena keterlambatan deteksi berimplikasi langsung pada kerugian waktu, biaya, dan kesempatan baik bagi mahasiswa maupun institusi (Al Hashmi dkk., 2025). Sayangnya, sebagian besar sistem deteksi risiko akademik yang diterapkan secara konvensional masih bersifat reaktif, yaitu baru menandai mahasiswa berisiko setelah ambang batas indeks prestasi kumulatif tertentu terlampaui, sehingga menyisakan jendela waktu yang sempit bagi institusi untuk melakukan intervensi sebelum kegagalan akademik benar-benar terjadi (Delogu dkk., 2024). Kondisi ini menegaskan pentingnya pergeseran paradigma dari pendekatan reaktif menuju pendekatan prediktif dan preventif dalam pengelolaan risiko akademik, di mana pola perkembangan performa mahasiswa dari waktu ke waktu dimanfaatkan sebagai sinyal deteksi dini, alih-alih hanya bergantung pada evaluasi hasil akhir semester.

Beberapa penelitian terdahulu telah mengembangkan model prediksi risiko akademik berbasis *machine learning* di pendidikan tinggi. (Cannistrà dkk., 2022) menggabungkan model berbasis teori dan pendekatan *data-driven*, menggunakan teknik statistik dan *machine learning* multilevel untuk memperhitungkan struktur data yang tersarang (*nested*) antar program studi di Politecnico di Milano. Di konteks Indonesia, (Andika Putra dkk., 2025) mengembangkan model prediksi berbasis data yang tersedia sejak pendaftaran dan diperbarui setelah semester pertama, dengan membandingkan performa Random Forest dan Gradient Boosting menggunakan metrik yang mempertimbangkan ketidakseimbangan kelas. (Marzuqi dkk., 2024) menerapkan Random Forest disertai teknik SMOTE untuk menangani ketidakseimbangan data pada kasus prediksi mahasiswa *drop out*, dan menemukan bahwa faktor akademik seperti IPK menjadi salah satu prediktor utama. Sejalan dengan itu, (Sulehu dkk., 2025) mengembangkan model prediksi kelulusan berbasis Random Forest menggunakan data akademik mahasiswa pada semester 4 sebagai titik kritis, dengan hasil bahwa IPK, motivasi belajar, dan kehadiran merupakan variabel paling berpengaruh terhadap kelulusan; namun studi ini menggunakan data pada satu titik waktu (snapshot semester 4) dan mengakui keterbatasan performa model, terutama recall yang rendah pada kelas mahasiswa lulus, akibat belum ditanganinya ketidakseimbangan data secara memadai.

Meskipun penelitian-penelitian tersebut menunjukkan bahwa pendekatan *machine learning* efektif untuk prediksi risiko akademik, sebagian besar masih menggunakan data pada satu titik waktu (*cross-sectional*), baik pada saat pendaftaran, semester awal, maupun semester tertentu yang ditetapkan sebagai titik kritis, dan belum memanfaatkan arsitektur *sequence-based* seperti LSTM yang mampu mempelajari struktur data longitudinal per semester secara eksplisit. Isu ketidakseimbangan kelas pada kategori taruna berisiko gagal juga secara konsisten muncul sebagai keterbatasan yang belum ditangani secara sistematis di sebagian besar studi tersebut. Selain itu, penerapan LSTM untuk prediksi risiko akademik pada konteks pendidikan vokasi pelayaran di Indonesia, dengan karakteristik kurikulum, kedisiplinan, dan pola kohort angkatan yang khas, masih sangat terbatas, dan isu kebocoran data (*data leakage*) akibat penentuan horizon prediksi yang tidak eksplisit belum banyak dibahas secara mendalam.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi risiko akademik taruna menggunakan *Long Short-Term Memory* (LSTM) berbasis data longitudinal di Poltekpel Barombong. Berbeda dari pendekatan *cross-sectional* pada satu semester seperti pada penelitian sebelumnya, penelitian ini memanfaatkan data akademik tiga angkatan (2020, 2021, 2022) yang dicatat per semester, dengan fokus pada jendela prediksi kritis semester 1 hingga 3, serta menangani ketidakseimbangan kelas secara eksplisit, untuk memprediksi status akademik akhir taruna (lulus, lambat lulus, atau tidak lulus) secara lebih dini dan akurat, sekaligus memberikan dasar bagi institusi dalam merancang intervensi akademik yang tepat sasaran dan tepat waktu.

Berbeda dengan algoritma berbasis pohon keputusan yang memperlakukan setiap observasi secara independen, LSTM mempertahankan dependensi temporal melalui mekanisme memory cell dan gating (input, forget, output gate) sehingga lebih sesuai untuk data longitudinal akademik(Raghuvanshi, 2024). Hal ini didukung oleh temuan empiris bahwa LSTM secara konsisten mencapai recall yang jauh lebih tinggi dibandingkan Decision Tree pada tahap prediksi dini, ketika kegagalan mendeteksi mahasiswa berisiko memiliki konsekuensi operasional paling besar (Kaushal & Mall, 2025).

Pemilihan semester 1 hingga 3 sebagai jendela prediksi kritis dalam penelitian ini didasarkan pada tiga pertimbangan. Pertama, secara regulatif, evaluasi keberhasilan studi tahap pertama pada jenjang Diploma 3 umumnya dilaksanakan pada akhir semester ke-4, yang menentukan apakah mahasiswa dapat melanjutkan studi atau dinyatakan tidak mampu menyelesaikan studi. Ketentuan ini menjadikan semester 1 hingga 3 sebagai satu-satunya horizon data yang tersedia bagi institusi untuk melakukan intervensi sebelum keputusan struktural tersebut diambil, sehingga model prediksi yang dibangun dari data setelah semester 4 kehilangan nilai praktisnya sebagai alat peringatan dini. Kedua, secara empiris, penggunaan data akademik dari tiga semester pertama sebagai horizon prediktor telah terbukti efektif dalam mendeteksi risiko gagal studi pada penelitian terdahulu; (setiawan dkk., 2026) misalnya, secara eksplisit membangun *Early Warning System* berbasis data akademik dan demografis tiga semester pertama untuk mendeteksi potensi *dropout* mahasiswa, sekaligus menangani ketidakseimbangan kelas melalui *cost-sensitive learning*. Hal ini berbeda dari pendekatan (Sulehu dkk., 2025) yang menggunakan data pada satu titik waktu di semester 4 sebagai snapshot tunggal, sehingga tidak menangkap dinamika perkembangan performa taruna dari semester ke semester maupun memberikan ruang intervensi yang memadai sebelum titik evaluasi tersebut. Ketiga, secara kontekstual pada pendidikan vokasi pelayaran, karakteristik sistem berasrama dengan penilaian disiplin dan kehadiran yang ketat pada setiap semester menjadikan tiga semester pertama sebagai periode adaptasi taruna yang paling menentukan pola akademik pada semester-semester berikutnya, sehingga sinyal risiko yang muncul pada rentang ini memiliki relevansi prediktif yang tinggi terhadap status akademik akhir.

METODE PENELITIAN

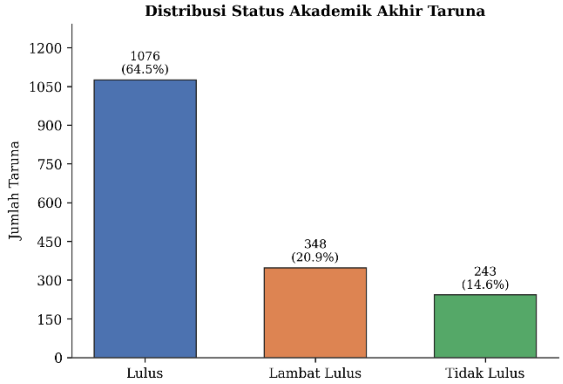
Desain dan Pendekatan Penelitian

Penelitian ini menggunakan desain kuantitatif dengan pendekatan *predictive analytics* berbasis *machine learning*, yang secara spesifik memanfaatkan struktur data longitudinal (panel akademik per semester) sebagai unit input model. Berbeda dari pendekatan *cross-sectional* yang digunakan pada penelitian-penelitian sebelumnya yang mengambil data pada satu titik waktu saja, seperti (Andika Putra dkk., 2025) yang hanya mengumpulkan data saat pendaftaran saja, dan (Sulehu dkk., 2025) yang mengunpulkan data pada semester tertentu sebagai titik kritis. Dimana penelitian ini memperlakukan riwayat akademik tiap taruna sebagai sebuah *sequence* (urutan waktu) yang dibaca secara utuh oleh model dari semester ke semester.

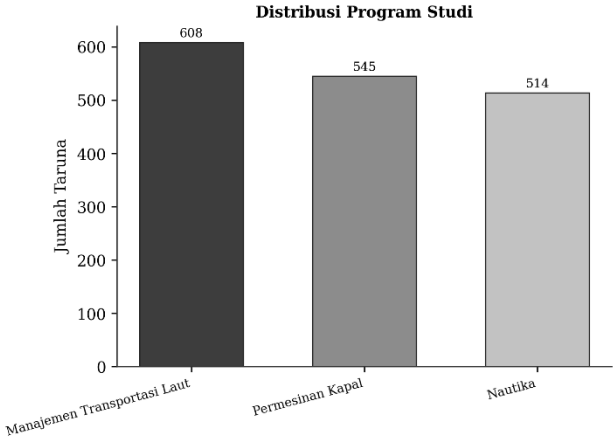
Pendekatan ini sejalan dengan gagasan (Cannistrà dkk., 2022) menyatakan bahwa struktur data pendidikan tinggi bersifat bertingkat (*nested*) dan berkembang dari waktu ke waktu, sehingga model yang mampu menangkap dinamika perubahan berpotensi memberikan deteksi risiko yang lebih presisi dibandingkan model titik-waktu tunggal.

Populasi, Sampel, dan Sumber Data

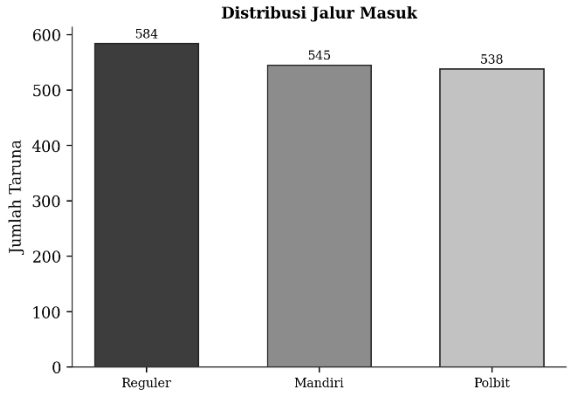
Populasi penelitian adalah seluruh taruna program Diploma 3 Poltekpel Barombong pada tiga program studi (Nautika, Permesinan Kapal, dan Manajemen Transportasi Laut). Sampel diambil dari tiga angkatan (2020, 2021, 2022) menggunakan data akademik yang dicatat institusi secara rutin per semester, mencakup 9.999 baris data yang merepresentasikan 1.667 taruna. Unit analisis dalam penelitian ini bukan baris data per semester, melainkan individu taruna, karena setiap taruna direpresentasikan sebagai satu *sequence* multivariat yang terdiri atas rekaman akademik semester 1 hingga semester 3. Taruna yang riwayat datanya belum lengkap sampai semester ke-3 pada saat pengambilan data dikeluarkan dari sampel analisis (listwise pada level individu, bukan pada level baris), untuk menjaga agar setiap *sequence* memiliki panjang yang seragam sesuai horizon prediksi yang ditetapkan. Distribusi data dapat dilihat pada gambar 1, 2 dan 3 berikut ini.



Gambar 1. Distribusi data berdasarkan kelulusan



Gambar 2. Distribusi data berdasarkan Prodi



Gambar 3. Distribusi data berdasarkan Distribusi Jalur Masuk

Variabel Penelitian dan Definisi Operasional

Variabel prediktor terdiri atas dua kelompok. Pertama, variabel numerik yang berubah tiap semester (*time-varying*): indeks prestasi kumulatif (IPK), indeks prestasi semester (IPS), jumlah SKS mata kuliah tidak lulus pada semester berjalan, akumulasi SKS tidak lulus, persentase kehadiran, dan jumlah pelanggaran disiplin. Kedua, variabel kategorikal yang bersifat statis sepanjang masa studi taruna: program studi dan jalur masuk. Variabel kategorikal ini di-*encode* menggunakan *one-hot encoding* agar dapat diproses bersama variabel numerik dalam satu representasi *sequence*.

Variabel target adalah status_akademik, dikategorikan menjadi tiga kelas: lulus, lambat lulus, dan tidak lulus. Nilai label ditetapkan berdasarkan status yang tercatat pada baris semester terakhir yang tersedia bagi tiap taruna (*final outcome*), bukan status pada semester-semester horizon (1–3), karena status pada semester-semester awal tersebut belum mencerminkan hasil akhir studi taruna yang bersangkutan.

Tahapan Praproses Data

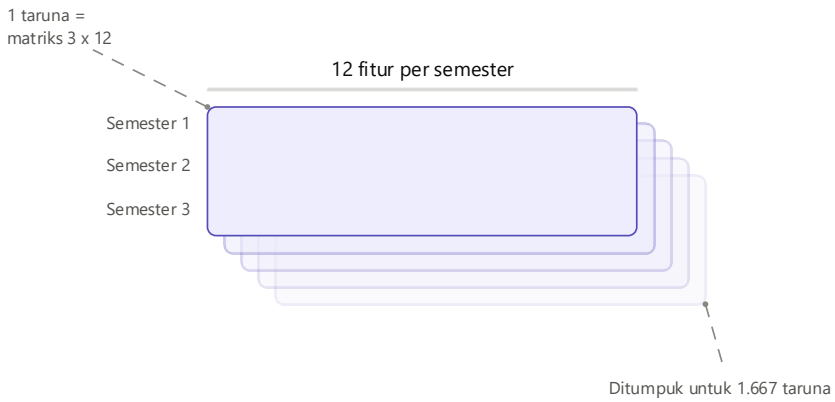
Praproses data dilakukan melalui empat tahap berurutan untuk mengubah data akademik mentah menjadi representasi tensor tiga dimensi yang sesuai dengan kebutuhan input LSTM.

a. Konstruksi label hasil akhir

Data akademik seluruh semester dikelompokkan berdasarkan id_taruna, diurutkan berdasarkan semester, kemudian status_akademik pada baris semester terakhir tiap taruna diambil sebagai label hasil akhir (final_status).

b. Pembentukan sequence longitudinal

Data difilter agar hanya memuat baris dengan semester ≤ 3 (horizon prediksi). Variabel kategorikal di-*one-hot-encode*, kemudian data tiap taruna diurutkan berdasarkan semester dan disusun menjadi *array* berukuran (horizon, jumlah fitur). Taruna dengan jumlah baris kurang dari horizon (riwayat tidak lengkap) dikeluarkan dari proses ini. Seluruh *sequence* individu digabungkan menjadi tensor X berukuran (jumlah taruna, 3, jumlah fitur), dapat diilustrasikan seperti pada gambar 4.



Gambar 4. Ilustrasi sequence longitudinal

c. Pembagian data latih dan uji per taruna

Pembagian data latih (80%) dan uji (20%) dilakukan pada level id_taruna, bukan pada level baris, dengan stratifikasi berdasarkan label hasil akhir. Pendekatan *split* per-individu ini merupakan penerapan langsung dari prinsip pencegahan kebocoran. (Kaufman dkk., 2012) berpendapat bahwa pada data longitudinal, apabila pembagian dilakukan pada level baris, ada risiko baris-baris semester dari taruna yang sama tersebar ke dalam data latih dan data uji sekaligus, sehingga model dapat "menghafal" karakteristik individu taruna tertentu alih-alih mempelajari pola umum, dan estimasi performa pada data uji menjadi bias dan terlalu optimistis.

d. Standardisasi fitur

Fitur numerik distandardisasi menggunakan *StandardScaler* yang di-*fit* hanya pada data latih, kemudian diterapkan (*transform*) baik pada data latih maupun data uji. Praktik *fit*-hanya-pada-*train* ini merupakan konvensi standar dalam *pipeline* pembelajaran mesin (Géron, 2019) untuk mencegah informasi statistik (rata-rata dan deviasi standar) dari data uji "bocor" ke proses pelatihan model.

Penanganan Ketidakseimbangan Kelas

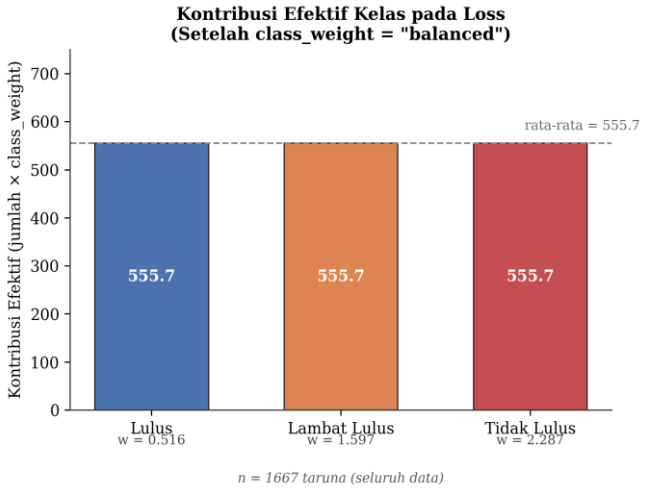
Distribusi label pada penelitian ini tidak seimbang dimana kelas lulus berjumlah 1.076 taruna jauh lebih banyak dibandingkan lambat lulus yang berjumlah 348 taruna dan, terutama, tidak lulus hanta berjumlah 243 taruna, kelas yang justru menjadi sasaran utama sistem peringatan dini. Ketidakseimbangan kelas semacam ini secara konsisten diidentifikasi sebagai keterbatasan pada penelitian-penelitian terdahulu(Marzuqi dkk., 2024; Sulehu dkk., 2025), yang berimplikasi pada rendahnya recall untuk kelas minoritas apabila tidak ditangani secara eksplisit.

Berbeda dari (Marzuqi dkk., 2024)yang menerapkan SMOTE, (Chawla dkk., 2002) menerapkan teknik *oversampling* yang membangkitkan sampel sintetis untuk kelas minoritas pada level baris data, dimana penelitian ini menggunakan pembobotan kelas (*class_weight*) yang dihitung secara berbanding terbalik dengan frekuensi tiap kelas yang menerapkan skema *balanced*, searah dengan prinsip pembobotan pada regresi logistik untuk data kejadian langka (King dkk., 2001), meninjau pendekatan *cost-sensitive learning* untuk data tidak seimbang (He & Garcia, 2009), melakukan pemilihan *class_weight* dimana *resampling* baris dengan melakukan duplikasi atau membangkitkan baris sintetis yang berisiko merusak keutuhan urutan semester per individu yang menjadi dasar representasi longitudinal, dimana *class_weight* bekerja pada level fungsi kerugian (*loss function*) tanpa mengubah struktur data input sama sekali. Lihat gambar 5 untu Kontribusi efektif tiap kelas terhadap fungsi loss.

Penanganan ketidakseimbangan kelas pada penelitian ini dilakukan melalui pendekatan *class weight* alih-alih teknik *resampling* seperti SMOTE (Chawla dkk., 2002), diman *resampling* pada level baris berisiko merusak struktur *sequence* longitudinal per taruna. Bobot kelas dihitung menggunakan skema "balanced" dari scikit-learn, dengan formula

$$W_i = \frac{n}{k \; x \; c_i}$$

Dimana
n adalah jumlah total taruna pada data latih,
k adalah jumlah kelas, dan c_i
c_i adalah jumlah taruna pada kelas ke-i

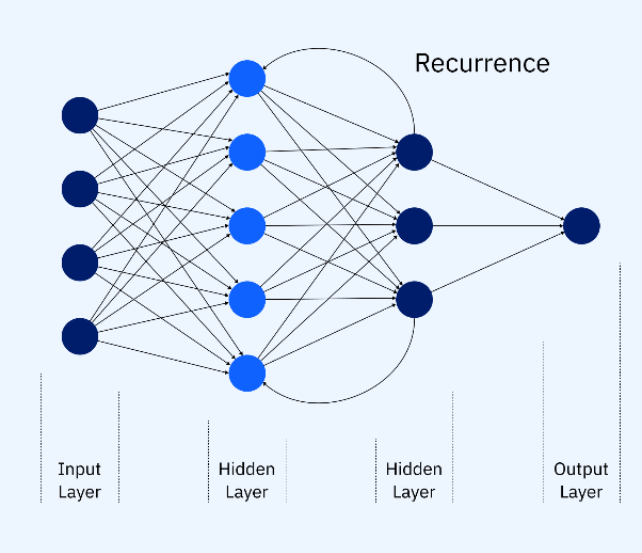


Gambar 5. Kontribusi efektif tiap kelas terhadap fungsi loss

Untuk mengukur efektivitas skema class weight dalam mengatasi ketidakseimbangan kelas pada keseluruhan data (n = 1.667 taruna), dihitung kontribusi efektif tiap kelas terhadap fungsi loss, yaitu hasil perkalian antara jumlah sampel pada kelas tersebut (c_i) dengan bobot kelasnya (w_i). Sebagaimana ditunjukkan pada Gambar 2, sebelum pembobotan kelas "lulus" mendominasi kontribusi terhadap loss karena jumlah sampelnya yang jauh lebih besar (1.076 taruna atau 64,5%), sementara kelas "tidak lulus" yang paling krusial untuk dideteksi hanya menyumbang sebagian kecil (243 taruna atau 14,6%). Setelah penerapan class_weight = "balanced", kontribusi efektif ketiga kelas terhadap loss menjadi sama besar, yaitu masing-masing sebesar 555,7 (setara dengan n/k, yaitu 1.667 dibagi 3 kelas): kelas lulus (1.076 × 0,516 ≈ 555,7), kelas lambat lulus (348 × 1,597 ≈ 555,7), dan kelas tidak lulus (243 × 2,290 ≈ 555,7). Kesetaraan kontribusi ini menunjukkan bahwa skema pembobotan bekerja sebagaimana mestinya, yaitu menghilangkan bias model terhadap kelas mayoritas tanpa mengubah struktur sequence data longitudinal per taruna, sebagaimana yang akan terjadi apabila resampling baris SMOTE diterapkan.

Arsitektur Model Prediktif LSTM

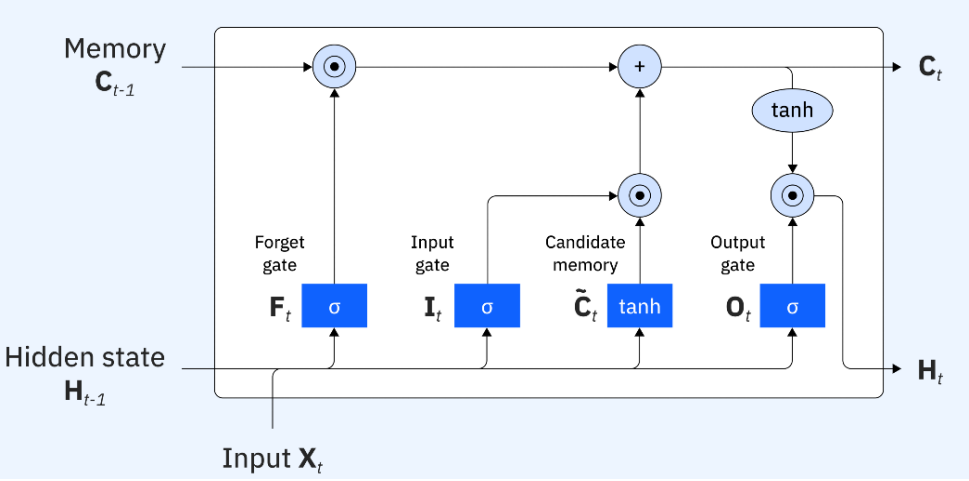
Model prediktif dibangun menggunakan arsitektur *Long Short-Term Memory* (LSTM), varian *recurrent neural network* (RNN) yang secara khusus dirancang untuk mempelajari dan mengingat pola dari data *sequential* yang panjang (Hochreiter & Urgen Schmidhuber, t.t.). RNN konvensional pada dasarnya meneruskan informasi dari satu langkah waktu ke langkah waktu berikutnya melalui struktur perulangan (*loop*), sehingga secara teoretis mampu menangkap dependensi jangka panjang; namun dalam praktiknya, proses *backpropagation* pada RNN kerap mengalami kendala gradien yang menghilang (*vanishing gradient*) atau meledak (*exploding gradient*), sehingga pembelajaran terhadap pola jangka panjang menjadi tidak stabil atau bahkan gagal (Joshua Noble, 2025).



Gambar 6. Ilustrasi Hidden Layer pada RNN (Joshua Noble, 2025)

LSTM dikembangkan sebagai solusi atas keterbatasan tersebut dengan menambahkan struktur memori internal berupa sel memori (*memory cell*), yang dilengkapi tiga gerbang multiplikatif — *forget gate*, *input gate*, dan *output gate* — yang secara bersama-sama mengatur informasi mana yang dipertahankan, ditambahkan, atau dikeluarkan dari *cell state* pada setiap langkah waktu (Joshua Noble, 2025). Mekanisme *gating* inilah yang membuat mempertahankan informasi dalam jangka panjang menjadi perilaku default LSTM, alih-alih rentan hilang seperti pada RNN sederhana.

Arsitektur ini dipilih karena secara alami sesuai dengan struktur data longitudinal penelitian, di mana risiko akademik taruna berkembang sebagai fungsi dari riwayat semester-semester sebelumnya, bukan sekadar nilai pada satu semester tunggal. Susunan arsitektur model adalah sebagai berikut



Gambar 7. Diagram Arsitektur Long Short-Term Memory (LSTM) (Joshua Noble, 2025)

Prosedur Pelatihan Model

Model dikompilasi menggunakan *optimizer* Adam dengan laju pembelajaran 0,001 (Kingma & Ba, 2017), fungsi kerugian *categorical crossentropy* yang sesuai untuk klasifikasi multi-kelas, serta metrik *accuracy* dan *Area Under the Curve* (AUC) sebagai pemantauan performa selama pelatihan. Label ditransformasi ke bentuk *one-hot* melalui to_categorical, dan pembobotan kelas (*class_weight*) yang telah dihitung pada tahap 2.6 disertakan pada proses pelatihan.

Pelatihan dilakukan maksimum 100 *epoch* dengan ukuran *batch* 32, serta menyisihkan 15% dari data latih sebagai data validasi. Mekanisme *early stopping* diterapkan dengan memantau *loss* pada data validasi (val_loss), toleransi (*patience*) 8 *epoch* tanpa perbaikan, dan mengembalikan bobot model pada *epoch* performa terbaik (restore_best_weights=True). Teknik ini digunakan untuk menghentikan pelatihan secara otomatis begitu model mulai

overfitting terhadap data latih, sehingga generalisasi model pada data yang belum pernah dilihat tetap terjaga.

Skema Evaluasi Model

Evaluasi dilakukan pada data uji yang seluruh taruna di dalamnya sama sekali tidak pernah dilihat oleh model selama pelatihan maupun validasi. Mengingat sifat data yang tidak seimbang, evaluasi tidak mengandalkan akurasi keseluruhan semata, melainkan menyertakan metrik yang lebih informatif untuk kelas minoritas (He & Garcia, 2009), yaitu *precision*, *recall*, dan F1-*score* per kelas melalui *classification report*; *confusion matrix* untuk melihat pola kesalahan klasifikasi antarkelas secara rinci; serta ROC-AUC (*macro*, *one-vs-rest*) untuk menilai kemampuan diskriminasi model secara agregat di antara tiga kelas.

HASIL DAN PEMBAHASAN

Karakteristik Data dan Hasil Konstruksi Sequence

Setelah tahap praproses sebagaimana diuraikan pada bagian 2.5, diperoleh 1.667 *sequence* taruna yang memenuhi syarat horizon yang memiliki catatan lengkap semester 1 hingga 3, dari total 9.999 baris data mentah tiga angkatan. Setiap *sequence* berbentuk tensor berukuran (3, 12), yaitu tiga langkah waktu dalam semester dengan 12 fitur pada tiap langkah yaitu 6 fitur numerik dan 6 kolom hasil *one-hot encoding* untuk prodi dan jalur masuk, sehingga keseluruhan data pelatihan berbentuk tensor (1.667, 3, 12).

Distribusi label hasil akhir menunjukkan ketidakseimbangan kelas yang cukup besar: kelas lulus sebanyak 1.076 taruna (64,6%), lambat lulus 348 taruna (20,9%), dan tidak lulus 243 taruna (14,6%). Rasio ketidakseimbangan antara kelas mayoritas (lulus) dan kelas minoritas (tidak lulus) mendekati 4,4:1, sebuah kondisi yang konsisten dengan keterbatasan yang dilaporkan pada penelitian-penelitian sejenis (Marzuqi dkk., 2024; Sulehu dkk., 2025) dan menjadi alasan utama diperlukannya penanganan ketidakseimbangan kelas secara eksplisit sebagaimana dijelaskan pada bagian 2.6.

Hasil Pembagian Data dan Pembobotan Kelas

Pembagian data pada level individu taruna dengan proporsi 80:20 dan stratifikasi label menghasilkan 1.333 taruna pada data latih dan 334 taruna pada data uji. Penghitungan pembobotan kelas (*class_weight*, skema "*balanced*") pada data latih menghasilkan nilai sebagaimana disajikan pada Tabel 1.

Tabel 1. Bobot kelas hasil skema untuk mendapatkan balanced pada data latih.

Kelas	Bobot (class_weight)	Interpretasi
Lulus	0,516	Bobot < 1, karena kelas mayoritas
Lambat lulus	1,597	Bobot > 1, kelas minoritas sedang
Tidak lulus	2,290	Bobot tertinggi, kelas paling minoritas

Kelas tidak lulus, yang jumlahnya paling sedikit sekaligus paling relevan bagi tujuan peringatan dini, memperoleh bobot terbesar (2,290) pada fungsi kerugian selama pelatihan, sehingga kesalahan model dalam mengklasifikasikan taruna pada kelas ini dihitung lebih berat dibandingkan kesalahan pada kelas lulus.

Hasil Evaluasi pada Data Uji

Proses pelatihan berjalan hingga *epoch* ke-34 sebelum dihentikan otomatis oleh mekanisme *early stopping* (patience 8, monitor val_loss). Akurasi pada data latih meningkat secara bertahap dari 0,351 pada *epoch* pertama menjadi berkisar 0,43–0,46 pada *epoch*-*epoch* akhir, diiringi penurunan *loss* latih dari 1,086 menjadi sekitar 0,97–0,99. Sebaliknya, *loss* pada data validasi mencapai titik terendah pada kisaran *epoch* 23–26 (val_loss ≈ 1,086), kemudian cenderung stagnan dan sedikit meningkat pada *epoch*-*epoch* berikutnya (val_loss ≈ 1,10–1,11 pada *epoch* 29–34), sehingga bobot model yang dipulihkan (*restore_best_weights*) adalah bobot pada *epoch* dengan val_loss terendah tersebut, bukan bobot pada *epoch* terakhir.

Pola ini — akurasi latih terus meningkat sementara *loss* validasi berhenti membaik — merupakan indikasi kecenderungan *overfitting* pada model terhadap data latih. Pada konteks data sintetis yang dipakai untuk validasi *pipeline* (lihat bagian 2.11 dan 3.5), pola ini justru

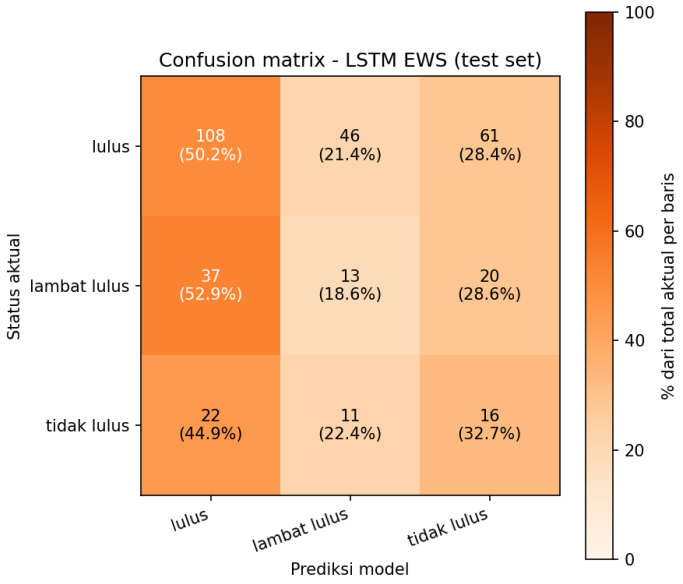
diharapkan: karena label pada data sintetis dibangkitkan secara acak, satu-satunya "pola" yang bisa terus dipelajari secara berkelanjutan oleh model pada data latih adalah kebisingan (*noise*) spesifik pada data latih itu sendiri, bukan sinyal yang dapat digeneralisasi. Berhentinya perbaikan performa validasi pada *epoch* pertengahan, dan intervensi *early stopping* setelahnya, menunjukkan bahwa mekanisme regularisasi (*dropout* dan *early stopping*) bekerja sebagaimana dirancang.

Evaluasi pada 334 taruna data uji (yang sama sekali tidak dilibatkan dalam pelatihan maupun validasi) menghasilkan akurasi keseluruhan sebesar 0,34. Rincian precision, recall, dan F1-*score* per kelas disajikan pada Tabel 2, dan pola kesalahan klasifikasi antarkelas disajikan pada Tabel 3 (*confusion matrix*).

Tabel 2. Classification report pada data uji.

Kelas	Precision	Recall	F1-score	Support
Lulus	0,64	0,36	0,46	215
Lambat lulus	0,23	0,34	0,27	70
Tidak lulus	0,10	0,22	0,14	49
Macro avg	**0,32**	**0,31**	**0,29**	**334**
Weighted avg	**0,48**	**0,34**	**0,37**	**334**

Berdasarkan classification report, model LSTM menunjukkan performa terbaik pada kelas lulus dengan nilai precision sebesar 0,64, recall 0,36, dan F1-score 0,46. Sementara itu, kemampuan model dalam mengidentifikasi kelas lambat lulus masih terbatas, dengan precision 0,23, recall 0,34, dan F1-score 0,27. Performa terendah terdapat pada kelas tidak lulus, yang hanya memperoleh precision 0,10, recall 0,22, dan F1-score 0,14, menunjukkan bahwa sebagian besar prediksi pada kelas ini masih kurang akurat. Secara keseluruhan, model menghasilkan macro average F1-score sebesar 0,29 dan weighted average F1-score sebesar 0,37. Perbedaan antara nilai macro dan weighted average mengindikasikan bahwa model lebih baik dalam mengenali kelas mayoritas (lulus) dibandingkan kelas minoritas (lambat lulus dan tidak lulus), sehingga masih diperlukan perbaikan untuk meningkatkan kemampuan deteksi taruna yang berisiko.



Gambar 8. Confusion Matrix

Berdasarkan confusion matrix pada data uji, model LSTM mampu mengklasifikasikan 108 dari 215 data (50,2%) pada kelas lulus dengan benar. Namun, performa model pada kelas lambat lulus dan tidak lulus masih rendah, dengan tingkat klasifikasi benar masing-masing hanya 13

dari 70 data (18,6%) dan 16 dari 49 data (32,7%). Sebagian besar data pada kedua kelas tersebut justru diprediksi sebagai lulus, yaitu sebanyak 52,9% untuk kelas lambat lulus dan 44,9% untuk kelas tidak lulus. Hasil ini menunjukkan bahwa model cenderung bias terhadap kelas lulus, yang kemungkinan dipengaruhi oleh ketidakseimbangan distribusi data. Dalam konteks Early Warning System, kondisi ini kurang ideal karena masih banyak taruna yang berisiko terlambat atau tidak lulus tidak berhasil teridentifikasi secara tepat.

Inverensi

```
data_baru = [
  {       "id_taruna": "taruna_baru_01",
          "semester": 1,
          "ipk": 2.10,
          "ips": 2.10,
          "sks_tidak_lulus": 4,
          "total_sks_tidak_lulus": 4,
          "presentasi_kehadiran": 55.0,
          "pelanggaran_disiplin": 3,
          "prodi": "Nautika",
          "Jalur Masuk": "Reguler"
  },
  {       "id_taruna": "taruna_baru_01",
          "semester": 2,
          "ipk": 2.05,
          "ips": 2.00,
          "sks_tidak_lulus": 5,
          "total_sks_tidak_lulus": 9,
          "presentasi_kehadiran": 50.0,
          "pelanggaran_disiplin": 4,
          "prodi": "Nautika",
          "Jalur Masuk": "Reguler"
  },
  {       "id_taruna": "taruna_baru_01",
          "semester": 3,
          "ipk": 1.98,
          "ips": 1.85,
          "sks_tidak_lulus": 6,
          "total_sks_tidak_lulus": 15,
          "presentasi_kehadiran": 45.0,
          "pelanggaran_disiplin": 5,
          "prodi": "Nautika",
          "Jalur Masuk": "Reguler”
  },
]
```

id_taruna	prob_lulus	prob_lambat lulus	prob_tidak lulus	prediksi	risiko_tidak_lulus
taruna_baru_01	0.258029	0.195246	0.546724	tidak lulus	0.547

Keterbatasan Penelitian

Beberapa keterbatasan perlu dikemukakan. Dimana hasil kuantitatif pada data sintetis yang dibangkitkan untuk keperluan validasi *pipeline*, merepresentasikan performa model pada data akademik yang riil, yang mengakibatkan, ukuran sampel pada kelas minoritas (tidak lulus) yang tergolong kecil, khususnya pada data uji (49 taruna), membuat estimasi metrik pada kelas ini rentan terhadap varians yang tinggi antar pengulangan eksperimen.

Jumlah unit LSTM, laju *dropout* dan laju pembelajaran, ukuran *batch* belum dioptimasi secara sistematis melalui pencarian *grid*, acak, atau Bayesian, melainkan ditetapkan

berdasarkan konfigurasi awal yang wajar untuk arsitektur berukuran kecil. Keempat, evaluasi baru dilakukan dengan satu skema pembagian data latih–uji (*single split*) dan satu nilai *random state*, sehingga belum mencerminkan variabilitas performa lintas pengulangan (*cross-validation*) pada level taruna. Keterbatasan-keterbatasan ini menjadi dasar bagi agenda pengembangan lanjutan begitu penelitian ini diterapkan pada data akademik taruna yang sesungguhnya.

KESIMPULAN

Penelitian ini berhasil membangun kerangka metodologis prediksi risiko akademik taruna berbasis data longitudinal menggunakan Long Short-Term Memory (LSTM), yang berbeda dari pendekatan cross-sectional pada penelitian-penelitian sebelumnya karena mampu menangkap dinamika perkembangan performa taruna dari semester ke semester sebagai satu kesatuan urutan. Validasi pipeline menggunakan data sintetis menunjukkan bahwa sistem bebas dari kebocoran data (*data leakage*) dan bekerja sesuai rancangan yang ditetapkan, sementara ketidakseimbangan kelas ditangani melalui pembobotan kelas (*class weight*) sehingga kontribusi setiap kelas terhadap fungsi loss menjadi lebih seimbang tanpa mengorbankan struktur data longitudinal yang menjadi ciri khas pendekatan ini. Pada tahap evaluasi menggunakan data sintetis, model mencapai akurasi sebesar 0,34 dengan ROC-AUC makro 0,471, dengan performa terbaik pada kelas taruna lulus, sedangkan kelas tidak lulus masih sulit dideteksi secara konsisten (precision 0,10; recall 0,22); mengingat label pada data sintetis dibangkitkan secara acak, hasil ini sejalan dengan ekspektasi dan mengonfirmasi bahwa pipeline berfungsi sebagaimana mestinya, bukan merepresentasikan performa model pada kondisi data riil. Secara keseluruhan, temuan ini menegaskan bahwa arsitektur LSTM dapat menjadi fondasi yang layak bagi pengembangan Early Warning System, meskipun performa aktual model perlu diuji dan ditingkatkan lebih lanjut menggunakan data institusional riil, sehingga penelitian ini menghasilkan kerangka metodologis prediksi risiko akademik berbasis data longitudinal yang dapat menjadi landasan pengembangan sistem pendukung keputusan dan Early Warning System pada penelitian selanjutnya.

Mengingat performa model saat ini masih dievaluasi menggunakan data sintetis dan menunjukkan keterbatasan pada deteksi kelas taruna tidak lulus, beberapa arah pengembangan berikut diajukan untuk meningkatkan akurasi, keandalan, serta kesiapan implementasi sistem peringatan dini pada penelitian selanjutnya.

1. Pengembangan lebih lanjut perlu dilakukan dengan memperbesar horizon data (misalnya hingga semester 4 atau lebih) agar pola akademik lebih kaya.
2. Integrasi variabel non-akademik seperti motivasi belajar, kondisi psikologis, atau faktor sosial dapat meningkatkan akurasi prediksi.
3. Eksperimen dengan arsitektur lain (misalnya GRU, Transformer) atau teknik hybrid bisa dibandingkan dengan LSTM untuk melihat peningkatan performa.
4. Penerapan nyata di institusi sebaiknya dilakukan bertahap, dimulai dengan sistem peringatan dini berbasis dashboard yang membantu dosen dan pembina taruna melakukan intervensi lebih cepat.
5. Penanganan ketidakseimbangan data dapat dieksplorasi lebih lanjut dengan teknik cost-sensitive learning atau ensemble methods agar kelas minoritas (*tidak lulus*) lebih terdeteksi.

DAFTAR PUSTAKA

Al Hashmi, R. A. M., Ozturk, I., & Elmehdi, H. M. (2025). Early detection of at-risk health sciences students: a machine learning-based predictive study using midterm grades. *BMC Medical Education*, *25*(1). https://doi.org/10.1186/s12909-025-08192-6

Andika Putra, F., Mirajdandi, S., Okmarizal, B., & Mulyanda, S. (2025). Prediksi Dropout Mahasiswa: Early-Warning Berbasis Enrollment dengan Machine Learning. *Jurnal Fasilkom*, *15*(3).

Cannistrà, M., Masci, C., Ieva, F., Agasisti, T., & Paganoni, A. M. (2022). Early-predicting dropout of university students: an application of innovative multilevel machine learning and statistical techniques. *Studies in Higher Education*, *47*(9), 1935–1956. https://doi.org/10.1080/03075079.2021.2018415

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Dalam *Journal of Artificial Intelligence Research* (Vol. 16).

Delogu, M., Lagravinese, R., Paolini, D., & Resce, G. (2024). Predicting dropout from higher education: Evidence from Italy. *Economic Modelling*, *130*. https://doi.org/10.1016/j.econmod.2023.106583

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow : concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hochreiter, S., & Urgen Schmidhuber, J. ¨. (t.t.). *Long Short-Term Memory*.

Joshua Noble. (2025). *What is long short-term memory (LSTM)?* . https://www.ibm.com/think/topics/lstm.

Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data*, *6*(4). https://doi.org/10.1145/2382577.2382579

Kaushal, V., & Mall, R. (2025). *AI-Driven Early Warning Systems for Student Success: Discovering Static Feature Dominance in Temporal Prediction Models*. http://arxiv.org/abs/2512.12493

King, G., Langche Zeng, G. H. E., Alt, J., Freeman, J., Gleditsch, K., Imbens, G., Manski, C., Mccullagh, P., Mebane, W., Nagler, J., Russett, B., Scheve, K., Schrodt, P., Tanner, M., Tucker, R., Bennett, S., Huth, P., & Zeng, L. (2001). *Logistic Regression in Rare Events Data*. http://GKing.Harvard.Edu.

Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization*. http://arxiv.org/abs/1412.6980

Marzuqi, T. A., Kristiani, E., & Marcel. (2024). Prediksi Mahasiswa Drop-Out Di Universitas XYZ. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *11*(6), 1345–1350. https://doi.org/10.25126/jtiik.2024118689

Raghuvanshi, K. P. (2024). A Systematic Literature Review on The Role of LSTM Networks in Capturing Temporal Dependencies in Data Mining Algorithms. *International Journal for Research in Applied Science and Engineering Technology*, *12*(10), 1219–1224. https://doi.org/10.22214/ijraset.2024.64761

setiawan, ismail, Santos, A. dos, & Dawis, A. M. (2026). Penerapan Logistic Regression pada Data Tidak Seimbang untuk Sistem Pendukung Keputusan Deteksi Dini Dropout Mahasiswa. *Ikomti*, *7*(2).

Sulehu, M., Wisda, W., Wanita, F., & Markani, M. (2025). Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi. *Jurnal Minfo Polgan*, *13*(2), 2364–2374. https://doi.org/10.33395/jmp.v13i2.14472