



Peringatan Dini Kegagalan Mata Kuliah Berbasis Kehadiran Awal Semester pada Data SIAKAD

Sani Nurmalasari^{1*}, Ircham Ali², Verry³

¹ Program Studi Informatika, Fakultas Matematika dan Ilmu Komputer, Universitas Nahdlatul Ulama Al Ghazali, Indonesia

² Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nahdlatul Ulama Indonesia, Indonesia

³ Program Studi Sistem Informasi, Fakultas Matematika dan Ilmu Komputer, Universitas Nahdlatul Ulama Al Ghazali, Indonesia

Info Artikel

Riwayat Artikel:

Diterima : **03 Juni 2026**

Direvisi : **15 Juni 2026**

Diterbitkan : **29 Juni 2026**

Kata Kunci:

peringatan dini,
kegagalan mata kuliah,
presensi,
kebocoran data,
regresi logistik Firth

ABSTRAK

Kegagalan mata kuliah adalah tanda awal yang mendahului putus studi. Namun, sebagian besar sistem peringatan dini (early warning system, EWS) di perguruan tinggi Indonesia dibangun berdasarkan nilai akhir semester dan diuji dengan pembagian data acak, sehingga peringatannya datang terlambat dan kinerjanya tampak lebih baik daripada kenyataan. Kami menguji seberapa dini presensi per pertemuan di SIAKAD mampu menandai mahasiswa yang akan gagal. Data berupa 193 pasangan mahasiswa-kelas dari 10 kelas pada tahun akademik 2025/2026 (20 kegagalan, 2.948 catatan presensi). Prediksi dibuat pada pertemuan ke-2, 3, 4, 6, dan 8 hanya dengan informasi yang sudah ada saat itu, memakai aturan ambang sederhana dan regresi logistik Firth, lalu diuji lintas kelas (leave-one-class-out, termasuk varian ketat yang juga memisahkan mahasiswa) dengan interval kepercayaan bootstrap. Absen sekali saja pada dua pertemuan pertama sudah menandai 21,2% mahasiswa dan menangkap 80% kasus gagal; dari mahasiswa yang hadir penuh, hanya 2,6% yang gagal. Model Firth dengan status mengulang mencapai AUC lintas kelas 0,850 [0,721–0,955] pada pertemuan keenam. Sebaliknya, fitur yang baru tersedia setelah cutoff (UAS, total kehadiran) memberi AUC 0,93–0,96, dan kami tunjukkan bahwa angka setinggi itu adalah kebocoran data, bukan kemampuan prediksi.

Keywords:

early warning system,
course failure,
attendance,
data leakage,
Firth logistic regression

ABSTRACT

Course failure is an early sign that precedes dropping out. Most early warning systems (EWS) in Indonesian higher education, however, are built from end-of-semester grades and tested with random data splits, so their warnings come late and their reported performance looks better than it really is. We ask how early per-session attendance stored in the academic information system (SIAKAD) can flag students who will fail. The data consist of 193 student-class pairs from 10 classes in the 2025/2026 academic year (20 failures, 2,948 attendance records). Predictions are made after sessions 2, 3, 4, 6, and 8 using only information available at that point, with a simple threshold rule and Firth logistic regression, and are tested across classes (leave-one-class-out, including a strict variant that also separates students) with bootstrap confidence intervals. A single absence in the first two sessions flags 21.2% of students and catches 80% of failures; among students with full attendance, only 2.6% fail. The Firth model with repeat status reaches a cross-class AUC of 0.850 [0.721–0.955] by the sixth session. By contrast, features that only become available after the cutoff (final exam, total attendance) give AUC 0.93–0.96, and we show that such figures reflect data leakage rather than predictive power.

Penulis Korespondensi:

Sani Nurmalasari,
Program Studi Informatika,
Universitas Nahdlatul Ulama Al Ghazali, Indonesia.
Email: saninurmala89@gmail.com

This is an open access article under the [CC BY-SA](#) license



1. PENDAHULUAN

Mahasiswa jarang berhenti kuliah secara tiba-tiba. Biasanya yang terjadi adalah rangkaian tanda kecil yang bisa dilihat lebih dulu: jarang masuk, tugas tidak dikumpulkan, satu dua mata kuliah tidak lulus, lalu semester berikutnya tidak lagi mendaftar. Sejak studi awal sistem peringatan dini (*early warning system*, EWS) pada jenjang sekolah menengah, kehadiran sudah termasuk indikator paling awal yang digunakan untuk mengenali siswa yang berisiko putus sekolah [1]. Meta-analisis Credé dkk. [2] menemukan bahwa kehadiran kuliah lebih kuat memprediksi nilai daripada skor tes masuk atau kebiasaan belajar. Bahkan pada jenjang sekolah dasar-menengah, estimasi selisih pertama (*first-differences*) pada data administratif memperlihatkan bahwa setiap hari absen menggerus capaian sedikit demi sedikit, dengan besaran yang berubah menurut konteks [3]. Generasi EWS berbasis data di perguruan tinggi, mulai dari penambahan log LMS [4], Course Signals [5], hingga inisiatif sumber terbuka lintas institusi [6], juga selalu bertumpu pada indikator keterlibatan yang sudah terlihat sejak dini. Di Indonesia sendiri, EWS berbasis *machine learning* umumnya diterapkan pada tingkat program studi, fakultas, atau institusi untuk memprediksi status putus studi (*dropout*) atau kelulusan tepat waktu dari data akademik seperti IPS tiap semester [7], [8], jumlah SKS dan jalur masuk [8], status pembayaran [9], atau status pekerjaan [7]. Sebagian lagi menggunakan *dataset* publik dari Portugal [10] sehingga tidak pernah menyentuh data kampusnya sendiri [11]. Bahkan EWS yang sudah diterapkan dan diuji kegunaannya pada dosen wali masih menggunakan IPS semester dan rekap presensi akhir semester sebagai sinyal [12]. Pendekatan seperti ini bermanfaat bagi perencanaan institusi, tetapi peringatannya baru keluar setelah nilai semester terbit, ketika sebagian mahasiswa sudah lewat titik kritis; pengecualiannya hanya skenario *enrollment-only* [11], yang bekerja sebelum kuliah dimulai.

Terdapat dua masalah metodologis yang mengikuti pola tersebut. Masalah pertama menyangkut waktu. Tugas prediksi jarang didefinisikan secara temporal, sehingga tidak jelas kapan prediksi dibuat, informasi apa yang sudah tersedia saat itu, dan siapa saja yang masih layak disebut "berisiko". Fauszt [13] merumuskan syarat-syarat ini dalam kerangka Time-Consistent Dropout Prediction (TCDP). Eksperimennya memperlihatkan bahwa pembagian data acak pada data panel mahasiswa mendongkrak AUC Random Forest dari 0,57 menjadi 1,00, karena baris-baris dari mahasiswa yang sama tersebar di data latih dan uji. Kebocoran data (*data leakage*) dalam pengertian Kaufman dkk. [14], yaitu masuknya informasi yang belum ada pada saat prediksi, juga terjadi ketika komponen nilai akhir seperti UAS dipakai untuk "memprediksi" kelulusan; nilai akhir memang dihitung langsung dari komponen itu.

Sebuah studi pada data SIAKAD Indonesia melaporkan akurasi 99% dengan hanya tiga kasus putus studi di data uji [8]. Polanya persis yang diuraikan Fauszt [13]: fitur yang sudah memuat hasil akhir, misalnya jumlah SKS semester berikutnya yang bernilai nol pada mahasiswa putus studi, ikut masuk model. Masalah kedua menyangkut kegunaan. Kinerja model jarang diterjemahkan ke dalam keputusan yang bisa dijalankan dosen. Shaikhanova dkk. [15] mengingatkan bahwa EWS yang benar-benar dipakai pengajar dibatasi oleh tiga hal sekaligus: kapan prediksi keluar (hari ke-14), berapa banyak mahasiswa yang sanggup ditindaklanjuti (15%), dan apakah alasannya bisa dijelaskan.

Dengan batasan tersebut, hasil yang terlihat pada data mereka, yaitu presisi 84% pada *recall* 35% sudah layak pakai, dan regresi logistik hampir setara dengan *gradient boosting*. Pada tingkat mata kuliah, Marbouti dkk. [16] menunjukkan bahwa data kinerja pekan-pekan awal, termasuk kehadiran dan penyelesaian tugas, sudah memberi daya pembeda yang berarti untuk mengenali mahasiswa berisiko, sementara Chen dkk. [17] mengevaluasi prediksi pekan demi pekan pada kelas daring sebuah kampus Indonesia dan menemukan pengumpulan tugas sebagai fitur terpenting. Keduanya mengarah pada pertanyaan yang sama: bukan seberapa akurat, melainkan seberapa dini. Putra dkk. [11] mengusulkan simulasi beban intervensi dan menabelkan *recall* pada berbagai proporsi mahasiswa yang ditandai, tetapi pada data publik, bukan data SIAKAD Indonesia.

Oleh karena itu, penelitian ini mencoba mengisi celah tersebut dengan analisis data yang paling dekat dengan jangkauan dosen, yaitu kegagalan mata kuliah, dengan menggunakan sumber data yang paling awal tersedia di SIAKAD: presensi pada tiap pertemuan. Tiga pertanyaan yang kami ajukan: (1) pada pertemuan ke berapa (k) kehadiran sudah bisa memprediksi kegagalan mata kuliah; (2) berapa bagian mahasiswa yang perlu ditindaklanjuti agar sebagian besar kasus gagal tertangkap; dan (3) seberapa jauh kinerja yang dilaporkan menggelembung jika fitur pasca-cutoff ikut dimasukkan. Kebaruannya ada pada tiga hal. Pertama, presensi per pertemuan dipakai sebagai sinyal dini, sedangkan studi Indonesia yang kami temukan [7]–[9], [11], [12],

kalaupun memakai kehadiran, hanya memakai rekap akhir semester [12]. Kedua, tugas prediksi didefinisikan secara konsisten menurut waktu mengikuti TCDP [13] pada data SIAKAD Indonesia dan diuji lintas kelas dengan regresi logistik terpenalisasi Firth [18], [19], yang memang dirancang untuk jumlah kejadian kecil. Ketiga, kinerja disajikan sebagai kurva *recall* terhadap proporsi mahasiswa yang ditandai, mengikuti [11], [15], bukan sebagai satu angka akurasi.

2. METODE PENELITIAN

2.1. Data dan Etika

Sumber data adalah laporan nilai perkuliahan dan laporan presensi SIAKAD Universitas Nahdlatul Ulama Al Ghazali Cilacap untuk semua kelas yang diampu peneliti pada tahun akademik 2025/2026, yaitu semester Gasal (September–Desember 2025) dan Genap (Februari–Juni 2026); SIAKAD menamai keduanya 2025 Gasal dan 2025 Genap, dan penamaan itu dipakai pada Tabel 1. Laporan nilai final memuat 198 baris pada 11 kelas; lima mahasiswa pada satu kelas (semester 1 Sistem Informasi, semuanya lulus) tidak memiliki laporan presensi dan tidak disertakan dalam analisis. Presensi dipasangkan ke laporan nilai melalui NIM, sehingga diperoleh 193 pasangan mahasiswa–kelas (137 mahasiswa berbeda) di 10 kelas dari 9 mata kuliah semester 1–7 program studi Informatika dan Sistem Informasi, dengan 2.948 catatan presensi. Setiap kelas mencatat 13–16 pertemuan. Pada tiga kelas Gasal ada tujuh sel presensi yang kosong, masing-masing satu pertemuan di tengah semester, sehingga jumlah catatan sedikit di bawah hasil kali jumlah mahasiswa dan pertemuan. Sel kosong diperlakukan sebagai absen; analisis sensitivitas yang menghitungnya sebagai hadir dilaporkan pada 3.2. Sebelum analisis, identitas mahasiswa dan kelas diganti dengan kode acak; kunci pemetaannya disimpan terpisah oleh peneliti. Ekstraksi dari berkas PDF diperiksa dengan membandingkan nilai akhir tercetak dengan $\Sigma(\text{komponen} \times \text{bobot})$ pada setiap baris, dan tidak ditemukan selisih.

Penggunaan data untuk penelitian ini telah mendapat izin tertulis dari Rektor Universitas Nahdlatul Ulama Al Ghazali. Satu hal perlu diakui secara terbuka: seluruh kelas diampu dan dinilai oleh peneliti sendiri, sehingga variabel keluaran ditentukan oleh orang yang sama yang melakukan analisis. Karena presensi dicatat sebelum dan terlepas dari penilaian, fitur ini tidak terpengaruh; namun, rangkap peran tersebut membatasi klaim independensi dan menjadi alasan tambahan mengapa replikasi oleh dosen lain diperlukan.

Luaran yang diprediksi adalah kegagalan mata kuliah dengan status "tidak lulus" pada laporan nilai (nilai huruf D+ ke bawah; nilai akhir tertinggi pada kelompok ini 55,37 dan terendah pada kelompok lulus 60,55). Ada 20 kegagalan (10,4%), seluruhnya pada lima kelas semester genap; lima kelas lainnya — keempat kelas gasal dan satu kelas karyawan genap (K02), 81 pasangan atau 42% data — tidak memiliki satu pun kegagalan. Dua hal perlu dicatat karena menyangkut kemandirian fitur dari label: kehadiran juga merupakan komponen nilai akhir (bobot 10–15%), dan ketentuan institusi mensyaratkan kehadiran minimal 75% untuk mengikuti UAS; implikasi keduanya diperiksa pada 3.1. Ringkasan kelas diberikan pada Tabel 1.

Tabel 1. Kelas yang dianalisis (identitas kelas dianonimkan)

Kelas	Periode	Semester MK	Jenis kelas	<i>n</i>	Pertemuan	Tidak lulus	Mahasiswa mengulang
K01	2025 Gasal	3	Karyawan	5	15	0	0
K02	2025 Genap	6	Karyawan	9	16	0	1
K03	2025 Genap	4	Reguler	24	16	6	4
K04	2025 Genap	4	Karyawan	5	16	1	0
K05	2025 Genap	6	Reguler	41	16	4	2
K06	2025 Genap	6	Karyawan	10	15	1	1
K07	2025 Gasal	7	Karyawan	10	15	0	1
K08	2025 Genap	6	Reguler	32	16	8	5
K09	2025 Gasal	5	Reguler	21	16	0	2
K10	2025 Gasal	1	Reguler	36	13	0	0
Total				193		20	16

2.2. Definisi Tugas Prediksi

Mengikuti kerangka TCDP [13], tugas prediksi kami tetapkan lewat empat komponen. *Cutoff*: prediksi dibuat setelah pertemuan ke- k , dengan $k \in \{2, 3, 4, 6, 8\}$; $k = 2-4$ kira-kira mewakili bulan pertama perkuliahan, sedangkan $k = 6-8$ mewakili masa menjelang UTS. Himpunan informasi: yang boleh dipakai hanya presensi pada pertemuan 1 sampai k dan atribut yang sudah diketahui saat registrasi, yaitu status mengulang, yang didefinisikan secara operasional sebagai angkatan: mahasiswa yang berasal dari angkatan lebih awal daripada angkatan yang semestinya menempuh semester itu. Laporan SIAKAD tidak membedakan pengulangan mata kuliah yang sama dari keterlambatan studi, sehingga "mengulang" di sini adalah proksi. Komponen nilai (tugas, UTS, UAS, praktikum) dan skor kehadiran akhir semester sengaja tidak digunakan.

Risk set: semua mahasiswa yang terdaftar di kelas tersebut, karena tidak ada yang mengundurkan diri sebelum *cutoff*. Validasi: pemisahan antarkelas (*leave-one-class-out*, LOCO). Model dilatih pada sembilan kelas dan diuji pada kelas kesepuluh, bergiliran untuk setiap kelas, sehingga tidak ada pasangan mahasiswa-kelas yang sama di data latih dan uji. Skema ini paling mendekati situasi nyata ketika model digunakan pada kelas baru, tetapi skema ini memiliki dua kelemahan yang harus disampaikan. Karena lima kelas tidak memiliki kejadian, hanya lima *fold* yang benar-benar menguji daya pembeda; lima *fold* lainnya hanya menyumbang kasus negatif pada AUC gabungan. Karena itu, AUC juga dilaporkan per *fold* untuk kelas yang memiliki kejadian. Selain itu, LOCO memisahkan kelas, bukan mahasiswa: 39 mahasiswa muncul pada dua atau tiga kelas (95 pasangan), dan 4 dari 16 mahasiswa yang gagal mengalaminya di lebih dari satu kelas. Dalam istilah TCDP, LOCO per kelas belum sepenuhnya memenuhi syarat pemisahan entitas, yaitu tidak ada mahasiswa yang sama di data latih dan uji. LOCO ketat memenuhinya dengan membuang baris latih yang mahasiswanya juga ada di kelas uji. Keduanya dilaporkan berdampingan pada Tabel 2.

Data presensi SIAKAD yang kami gunakan memiliki empat status, yaitu: hadir (H), izin (I), sakit (S), dan alpa (A). Ketika diperiksa, status I ternyata mendominasi catatan tidak hadir (181 dari 229) dan sebarannya hampir sama antara kelompok lulus (96) dan gagal (85); A (18 berbanding 18) dan S (5 berbanding 7) juga tidak membedakan keduanya. Karena itu, fitur kehadiran kami definisikan sebagai jumlah status H pada k pertemuan pertama (*hadir_k*), dan semua status selain H diperlakukan sebagai absen.

2.3. Model

Pada setiap *cutoff*, kami menguji dua pendekatan. Pendekatan pertama berupa aturan ambang sederhana: mahasiswa ditandai berisiko bila *hadir_k* < k , yakni tercatat absen sekurang-kurangnya sekali pada k pertemuan pertama. Aturan ini tidak memerlukan pelatihan model dan dapat diterapkan oleh dosen secara manual langsung dari rekap presensi, tanpa perangkat lunak apa pun. Pendekatan kedua berupa regresi logistik terpenalisasi Firth [18], yang memaksimumkan *likelihood* dengan penalti berbasis prior Jeffreys sehingga estimasi koefisien tetap terhingga dan bias orde pertamanya terkoreksi, bahkan ketika kejadian sedikit atau terjadi separasi — kondisi ketika suatu nilai prediktor memisahkan kelompok gagal dan lulus secara sempurna [19]. Pendekatan kedua diestimasi dalam dua spesifikasi: Model A dengan *hadir_k* sebagai satu-satunya prediktor, dan Model B yang menambahkan status mengulang.

Pilihan ini mengikuti kaidah jumlah kejadian per variabel: dengan 20 kejadian, regresi *maximum likelihood* dengan lebih dari dua prediktor sudah jatuh di bawah ambang sepuluh kejadian per variabel yang disarankan Peduzzi dkk. [20]. Alasan kedua adalah risiko separasi: dengan 20 kejadian dan prediktor diskret, separasi kuasi-sempurna pada sebagian *fold* latih merupakan risiko nyata, dan dalam kondisi itu estimasi *maximum likelihood* gagal konvergen; Firth dipakai sebagai pengaman, dan 3.2 melaporkan apakah risiko itu terwujud. Rahman dan Sultana [21] menunjukkan bahwa pada data kecil atau jarang, penalisasi Firth memberi kalibrasi dan diskriminasi setara dengan metode *log-F* dan menyelesaikan separasi, meski cenderung menaikkan rerata probabilitas prediksi; hal terakhir ini kami periksa pada 3.4. Regresi logistik *maximum likelihood* biasa tetap dilaporkan sebagai pembanding untuk memeriksa separasi.

2.4. Evaluasi

Daya pembeda dilaporkan sebagai AUC beserta interval kepercayaan 95% hasil *bootstrap* berstrata menurut luaran (2.000 ulangan, *seed* 42). Karena 39 mahasiswa muncul di lebih dari satu kelas, baris tidak sepenuhnya independen; *bootstrap* klaster per mahasiswa dilaporkan sebagai pembanding pada 3.2. Untuk model Firth, AUC utama dihitung dari gabungan prediksi LOCO seluruh kelas, sedangkan AUC per *fold*

dilaporkan sebagai pelengkap. Ukuran kegunaan praktis tidak dapat ditetapkan lewat persentil karena seluruh prediktor bersifat diskret sehingga skor model hanya mengambil sedikit nilai berbeda; pada $k = 2$, misalnya, model B menghasilkan empat kombinasi skor yang teramati — hadir 1 atau 2 pertemuan, disilangkan dengan status mengulang. Menerapkan aturan "tandai 10% teratas" pada skor sekasar itu menuntut pemecahan mahasiswa berskor identik secara sewenang-wenang. Karena itu, presisi dan recall dilaporkan pada batas antarsel skor: tiap kombinasi nilai prediktor diperlakukan sebagai satu sel, sel diurutkan menurut rerata risiko prediksi LOCO, lalu presisi dan recall kumulatif dihitung pada setiap batas; hitungan ini memakai luaran teramati seluruh data, sedangkan LOCO hanya menentukan urutan sel. Pada prevalensi kegagalan sekitar 10%, pasangan ukuran ini lebih informatif daripada AUC tunggal [22].

Penilaian kalibrasi kami tempuh melalui dua jalur. Skor Brier dihitung berdasarkan prediksi LOCO dan disandingkan dengan Brier model nol yang selalu mengeluarkan prevalensi dasar, sehingga terbaca seberapa jauh model melampaui tebakan konstan. Selanjutnya, pada tiap sel skor, rerata risiko prediksi LOCO diperbandingkan dengan proporsi kegagalan yang benar-benar teramati di sel itu; prosedur ini mengikuti kerangka Steyerberg dkk. [23]. Kami juga menghitung AUC bagi empat fitur yang baru tersedia setelah cutoff atau setelah semester usai — proporsi kehadiran satu semester penuh, skor kehadiran SIAKAD, nilai praktikum, dan nilai UAS. Keempatnya tidak masuk ke model mana pun; angkanya dilaporkan semata-mata sebagai tolok banding untuk memperlihatkan seberapa besar performa menggelembung ketika batas temporal diabaikan. Analisis dikerjakan pada Python 3.11 (pandas, scikit-learn). Estimator Firth ditulis sendiri oleh penulis dengan iterasi Newton–Raphson atas fungsi skor terpenalisasi [18]; iterasi dihentikan bila perubahan langkah di bawah 10^{-8} (maksimum 200 iterasi). Implementasi diverifikasi dengan dua cara: pada model jenuh berprediktor biner (model A, $k = 2$) hasilnya identik hingga 10^{-12} dengan koreksi $+0,5$ per sel tabel 2×2 [18], [19], dan maksimisasi langsung likelihood terpenalisasi menghasilkan koefisien yang sama hingga 10^{-7} . Regresi logistik maximum likelihood biasa dijalankan pada skema LOCO yang sama sebagai pembanding. Kode dan dataset teranonimisasi tersedia atas permintaan kepada penulis.

3. HASIL DAN ANALISIS

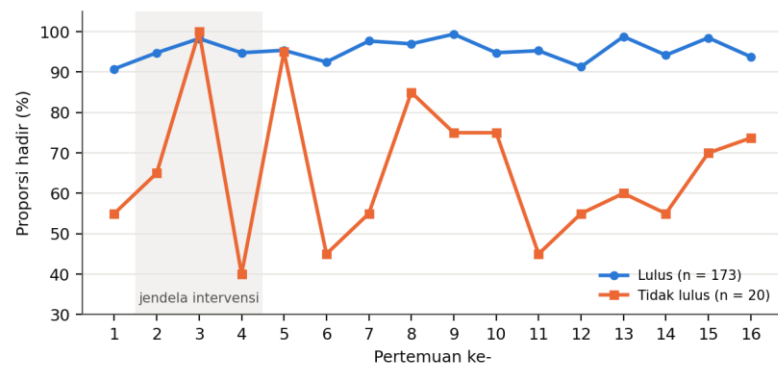
3.1. Profil Kegagalan

Kegagalan pada data ini hampir selalu berwujud komponen nilai yang kosong, bukan komponen yang bernilai rendah. Dari 20 kasus gagal — 20 pasangan mahasiswa–kelas yang berasal dari 16 mahasiswa berbeda — sebanyak 15 tidak mengumpulkan tugas individu, 12 tidak mengikuti UTS, 15 tidak mengikuti UAS, dan 13 tercatat hadir kurang dari 75%, ambang institusi untuk mengikuti UAS. Ambang itu tidak menjadikan kegagalan sekadar konsekuensi administratif: tujuh kasus gagal berkehadiran $\geq 75\%$, empat dari 15 yang tidak mengikuti UAS berada di atas ambang, dan penerapan ambang melibatkan pertimbangan dosen, sehingga tidak bersifat otomatis. Sumbangan kehadiran ke nilai akhir pun kecil — paling banyak sekitar 2 poin dari dua pertemuan pertama dan 6 poin dari enam pertemuan — sedangkan nilai akhir tertinggi kelompok gagal terpaut 5,2 poin dari terendah kelompok lulus dan 18 dari 20 kasus gagal bernilai di bawah 50, sehingga bagian *hadir_k* di dalam label tidak mungkin menjelaskan daya pembeda yang dilaporkan pada 3.2. Dengan kata lain, kegagalan di sini berwatak penarikan diri dari perkuliahan, bukan ketidakmampuan menyelesaikan penilaian. Faktor tunggal yang paling menonjol adalah status mengulang. Dari 16 pasangan mahasiswa–kelas dengan status mengulang, 10 berakhir gagal (62,5%), perbandingannya adalah 10 dari 177 pasangan pada mahasiswa seangkatan (5,6%).

Gambar 1 menyajikan lintasan kehadiran dari pertemuan ke pertemuan. Kelompok lulus bertahan pada 91–99% sepanjang semester. Kelompok gagal hanya mencapai 55% pada pertemuan pertama dan 65% pada pertemuan kedua, lalu berayun tajam antara 40% dan 100%, dengan sebagian besar pertemuan berada pada kisaran 45–75%.

Dua lonjakan menuju titik 95–100% pada pertemuan ke-3 dan ke-5 kemungkinan besar merupakan artefak pencatatan — seluruh peserta ditandai hadir — bukan pertanda pulihnya keterlibatan. Lepasnya

keterlibatan karenanya sudah terbaca sejak dua pertemuan pertama, sehingga jendela intervensi yang realistis terletak pada pertemuan 2–4, jauh sebelum UTS.



Gambar 1. Proporsi hadir per pertemuan, kelompok lulus ($n = 173$) dan tidak lulus ($n = 20$)

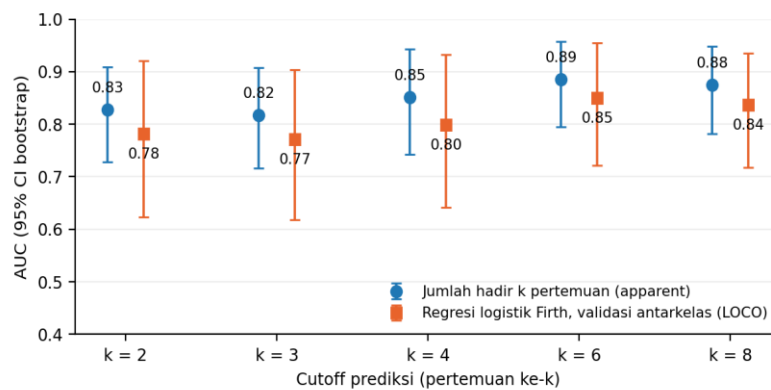
3.2. Kinerja pada Setiap Cutoff

Tabel 2. Kinerja aturan ambang dan model per *cutoff* k ($n = 193$; 20 kegagalan)

k	Ditandai (aturan)	Presisi	Recall	Gagal jika hadir penuh	AUC hadir_ k (apparent) [95% CI]	AUC Firth A, LOCO [95% CI]	AUC Firth B, LOCO [95% CI]	AUC Firth B, LOCO ketat [95% CI]
2	41 (21%)	39%	80%	2,6%	0,828 [0,728–0,909]	0,728 [0,592–0,853]	0,781 [0,623–0,921]	0,796 [0,648–0,921]
3	44 (23%)	36%	80%	2,7%	0,817 [0,716–0,908]	0,714 [0,575–0,834]	0,771 [0,618–0,903]	0,786 [0,647–0,907]
4	50 (26%)	32%	80%	2,8%	0,851 [0,742–0,943]	0,786 [0,630–0,911]	0,799 [0,642–0,933]	0,795 [0,640–0,923]
6	64 (33%)	28%	90%	1,6%	0,885 [0,795–0,957]	0,842 [0,715–0,936]	0,850 [0,721–0,955]	0,849 [0,732–0,946]
8	70 (36%)	26%	90%	1,6%	0,875 [0,782–0,949]	0,829 [0,702–0,927]	0,836 [0,718–0,935]	0,839 [0,717–0,934]

Tabel 2 merangkum kinerja aturan ambang dan model Firth pada kelima cutoff. Pada $k = 2$, aturan "absen sekurang-kurangnya satu kali dari dua pertemuan pertama" menandai 41 mahasiswa (21,2%) dengan presisi 39,0% dan recall 80%; sebaliknya, di antara mahasiswa yang hadir penuh pada dua pertemuan pertama hanya 2,6% yang berakhir gagal. Memperpanjang jendela ke $k = 6$ menaikkan recall ke 90%, tetapi proporsi yang ditandai turut membengkak menjadi 33,2% — sepertiga kelas. AUC jumlah hadir bergerak dari 0,828 [0,728–0,909] pada $k = 2$ ke 0,885 [0,795–0,957] pada $k = 6$, lalu tidak lagi membaik pada $k = 8$. Penurunan kecil pada $k = 3$ (0,817) konsisten dengan artefak pertemuan ketiga yang dibahas pada 3.1: karena pada pertemuan itu delapan dari sepuluh kelas tercatat hadir 100% dan seluruh kelompok gagal tercatat hadir, hadir_3 nyaris tidak menambah informasi apa pun di atas hadir_2.

Pola ini selaras dengan temuan Shaikhanova dkk. [15] dan Marbouti dkk. [16], yang melaporkan bahwa data yang terkumpul setelah pekan-pekan awal hanya sedikit menambah daya pembeda; pada studi Shaikhanova dkk. [15], AUC hanya naik 0,03 antara hari ke-28 dan hari ke-112. Ketujuh sel presensi kosong (lihat 2.1) tidak mengubah gambaran ini. Bila ketujuhnya diperlakukan sebagai hadir, jumlah mahasiswa yang ditandai menjadi 43, 49, 63, dan 67 berturut-turut pada $k = 3, 4, 6$, dan 8; recall tidak berubah, dan pergeseran AUC paling besar hanya 0,01.



Gambar 2. AUC per cutoff: jumlah hadir k pertemuan (*apparent*) dan model Firth B pada validasi antarkelas

Pada validasi antarkelas, AUC model Firth A (hanya hadir_k) turun menjadi 0,728 [0,592–0,853] pada $k = 2$ dan 0,842 [0,715–0,936] pada $k = 6$, yakni 0,04–0,10 di bawah nilai dalam-sampel (*apparent*). Selisih sebesar ini merupakan optimisme yang lazim menyertai performa dalam-sampel, diperbesar oleh tingkat kegagalan yang memang beragam antarkelas (0–25%); selisih tersebut bukan indikasi kebocoran, sebab himpunan informasi telah dikunci pada cutoff. Penambahan status mengulang (model B) memperbaiki AUC LOCO pada cutoff paling dini — 0,781 [0,623–0,921] pada $k = 2$ — sedangkan pada $k = 6$ kedua model praktis setara (0,850 berbanding 0,842). Selisih pada tiap cutoff tidak konklusif karena interval kepercayaannya tumpang tindih; yang lebih meyakinkan adalah polanya: model B mengungguli model A pada kelima cutoff, dengan selisih yang mengecil dari 0,05 pada $k = 2$ menjadi 0,01 pada $k = 6$ –8.

Koefisien Firth model B pada $k = 2$ adalah $-2,50$ [$-3,79; -1,36$] per pertemuan hadir (rasio odds 0,08) dan $+2,29$ [0,99; 3,67] untuk status mengulang (rasio odds 9,9 [2,7–39,2]); pada $k = 6$ keduanya $-1,29$ [$-1,92; -0,75$] dan $+1,83$ [0,32; 3,31] (interval kepercayaan 95% *profile penalized likelihood*). Karena hadir₂ hanya bernilai 1 atau 2, kemiringan pada $k = 2$ ditaksir dari satu kontras dan tidak boleh diekstrapolasi ke nilai lain. Pada model A, yang hanya memuat hadir_k, koefisiennya lebih curam, $-3,06$ [$-4,30; -2,01$] (rasio odds 0,047), karena sebagian pengaruh status mengulang terserap ke dalam koefisien kehadiran.

Estimator alternatif tidak mengubah gambaran. Regresi logistik *maximum likelihood* tanpa penalti menaksir koefisien model A pada $-3,17$, sedikit lebih ekstrem dibanding $-3,06$ milik Firth — arah yang memang diharapkan, karena penalti Jeffreys menyusutkan taksiran ke arah nol — dan tidak ada fold yang mengalami separasi sempurna. Pada skema LOCO, AUC keduanya identik untuk model B (0,781 pada $k = 2$; 0,850 pada $k = 6$) — hasil yang memang diharapkan, karena AUC hanya bergantung pada urutan sel skor dan penyusutan Firth tidak mengubahnya; skor Brier LOCO, yang peka terhadap besaran probabilitas, pun berdekatan (0,0623 berbanding 0,0623 pada $k = 2$; 0,0614 berbanding 0,0619 pada $k = 6$). Penalti Firth di sini berfungsi sebagai pengaman terhadap fold berkejadian sedikit, bukan sebagai penentu hasil.

Pemisahan entitas yang lebih ketat pun tidak mengubah kesimpulan. LOCO ketat membuang dari data latih seluruh baris milik mahasiswa yang muncul di kelas uji (Tabel 2, kolom terakhir), namun AUC model B tidak turun: 0,796 [0,648–0,921] pada $k = 2$ dan 0,849 [0,732–0,946] pada $k = 6$. Tumpang tindih mahasiswa antarkelas, dengan kata lain, tidak menggelembungkan hasil — sejalan dengan penjelasan Fauszt [13] bahwa fitur berkardinalitas rendah hanya meninggalkan jejak identitas yang lemah sehingga sukar dimanfaatkan model linear untuk mengenali kembali mahasiswa yang sama.

Bootstrap klaster per mahasiswa memberi interval yang sedikit lebih lebar untuk model B — [0,59–0,95] pada $k = 2$ dan [0,69–0,97] pada $k = 6$ — dengan batas bawah turun 0,03–0,04 tanpa mengubah kesimpulan. Sebaran per fold memperlihatkan keragaman yang besar. Pada kelima kelas yang memiliki kejadian, AUC model B pada $k = 2$ dan $k = 6$ berturut-turut adalah 0,97 dan 0,97 untuk K03 ($n = 24$; 6 kejadian), 0,88 dan 1,00 untuk K04 (5; 1), 0,67 dan 0,77 untuk K05 (41; 4), 0,89 dan 0,94 untuk K06 (10; 1), serta 0,87 dan 0,93 untuk K08 (32; 8). Nilai 1,00 pada K04 hanya berarti satu mahasiswa yang gagal berperingkat di atas empat rekannya, dan nilai terendah jatuh pada K05, kelas tempat dua di antara mahasiswa yang akhirnya gagal justru hadir penuh

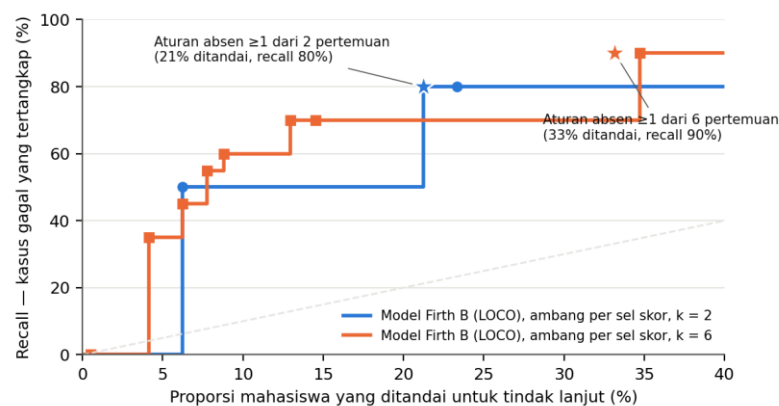
sampai pertemuan ke-4. Rentang selebar itu, berikut interval kepercayaan yang menyertainya, merupakan konsekuensi wajar dari 20 kejadian dan dilaporkan apa adanya.

3.3. Beban Tindak Lanjut

Pertanyaan yang dihadapi dosen di lapangan bukan seberapa tinggi AUC sebuah model, melainkan: jika hanya sebagian mahasiswa yang sanggup ditindaklanjuti, berapa kasus gagal yang tertangkap? Gambar 3 dan Tabel 3 menjawabnya dengan menyusun skor risiko LOCO model B menurut kapasitas tindak lanjut. Karena seluruh prediktor diskret, titik operasinya adalah batas antarsel skor, bukan persentil.

Empat sel tersedia pada $k = 2$. Yang paling berisiko memuat mahasiswa mengulang yang absen satu kali: 12 orang, 10 di antaranya berakhir gagal (83%). Menindaklanjuti sel ini saja berarti memberi perhatian pada 6% mahasiswa dan sudah menangkap separuh kasus gagal. Selanjutnya, mahasiswa seangkatan yang absen satu kali, membawa cakupan ke 21% dengan recall 80% dan presisi 39% — dan pada titik ini, himpunan yang ditandai identik dengan himpunan aturan "absen ≥ 1 dari 2 pertemuan pertama". Kesamaan itu muncul karena model menempatkan kedua sel yang hadir penuh di urutan paling bawah, sementara salah satunya — empat mahasiswa yang mengulang hadir penuh — tidak menyumbang kegagalan sama sekali. Urutan itu rapuh: rerata risiko LOCO sel "seangkatan, hadir 1" (0,261) hanya sedikit di atas sel "mengulang, hadir 2" (0,222), selisih yang jauh lebih kecil daripada ketidakpastian koefisien, meski data teramati searah dengannya (6 dari 29 berbanding 0 dari 4).

Sel yang lebih banyak pada $k = 6$ menghasilkan kurva yang lebih halus. Cakupan 9% (hadir ≤ 5 sekaligus mengulang, atau hadir ≤ 3) menangkap 60% kasus gagal dengan presisi 71%, sedangkan cakupan 15% (seluruh pengulang, atau hadir ≤ 4) menangkap 70% dengan presisi 50%. Pembacaan praktisnya: keunggulan model atas aturan biner hanya muncul pada kapasitas tindak lanjut di bawah 15%, sementara pada kapasitas 20% ke atas aturan absen sekali di $k = 2$ sudah menyamainya.



Gambar 3. Recall terhadap proporsi mahasiswa yang ditandai (model Firth B, prediksi LOCO) dan titik aturan ambang

Tabel 3. Presisi dan *recall* kumulatif model Firth B (LOCO) pada batas sel skor; sel diurutkan menurut rerata risiko prediksi LOCO; hitungan memakai luaran teramati seluruh data

<i>k</i>	Sel skor yang ditandai (kumulatif)	<i>n</i> sel (gagal)	Ditandai kumulatif	Presisi	Recall
2	mengulang, hadir 1	12 (10)	12 (6%)	83%	50%
2	+ seangkatan, hadir 1	29 (6)	41 (21%)	39%	80%
2	+ mengulang, hadir 2	4 (0)	45 (23%)	36%	80%
2	+ seangkatan, hadir 2 (semua)	148 (4)	193 (100%)	10%	100%
6	hadir 2, seangkatan	1 (0)	1 (1%)	0%	0%
6	+ hadir 3, mengulang	7 (7)	8 (4%)	88%	35%
6	+ hadir 4, mengulang	4 (2)	12 (6%)	75%	45%
6	+ hadir 3, seangkatan	3 (2)	15 (8%)	73%	55%
6	+ hadir 5, mengulang	2 (1)	17 (9%)	71%	60%
6	+ hadir 4, seangkatan	8 (2)	25 (13%)	56%	70%
6	+ hadir 6, mengulang	3 (0)	28 (15%)	50%	70%
6	+ hadir 5, seangkatan	39 (4)	67 (35%)	27%	90%
6	+ hadir 6, seangkatan	126 (2)	193 (100%)	10%	100%

3.4. Kalibrasi

Skor Brier prediksi LOCO model B berada pada kisaran 0,059–0,067, di bawah Brier model nol yang selalu mengeluarkan prevalensi dasar (0,093); *Brier skill score* 0,27–0,37. Rerata prediksi LOCO model B (0,105–0,109) juga mendekati prevalensi teramati (0,104), sehingga bias rerata probabilitas yang dilaporkan menyertai penalisasi Firth [21] tidak tampak pada data ini.

Pada tingkat sel skor di $k = 2$, rerata prediksi LOCO dan proporsi kegagalan teramati berturut-turut adalah 0,691 berbanding 0,833 (mengulang, hadir 1; $n = 12$), 0,261 berbanding 0,207 (seangkatan, hadir 1; $n = 29$), 0,222 berbanding 0,000 (mengulang, hadir 2; $n = 4$), dan 0,024 berbanding 0,027 (seangkatan, hadir 2; $n = 148$). Model menaksir risiko terlalu rendah pada sel tertinggi dan terlalu tinggi pada sel pengulang berkehadiran penuh — sel terakhir hanya memuat empat orang, sehingga proporsi teramati 0,000 pada dirinya sendiri tidak informatif. Pada $k = 6$, sel dengan $n \geq 8$ juga tergolong dekat (seangkatan hadir 5: 0,084 berbanding 0,103; seangkatan hadir 6: 0,023 berbanding 0,016).

Dengan 20 kejadian yang tersebar di sejumlah sel berukuran kecil, gambaran kalibrasi ini sebaiknya dibaca sebagai indikasi kasar, bukan sebagai penilaian kalibrasi yang definitif.

3.5. Demonstrasi Kebocoran Data

Tabel 4 memperlihatkan apa yang terjadi bila fitur pasca-cutoff turut dipakai. Jumlah baris untuk UAS dan praktikum lebih kecil (157 dan 152) karena satu kelas semester 1 tidak memiliki komponen UAS maupun praktikum, dan satu kelas karyawan tidak memiliki praktikum. Keduanya adalah kelas tanpa kegagalan, sehingga seluruh 20 kejadian tetap tercakup pada setiap baris Tabel 4; nilai nol pada mahasiswa yang gagal merupakan nol yang sebenarnya — tidak mengikuti komponen tersebut — bukan data hilang.

Proporsi kehadiran satu semester penuh menghasilkan AUC 0,946, skor kehadiran SIAKAD 0,925, nilai praktikum 0,930, dan nilai UAS 0,959. Seluruhnya tampak lebih unggul daripada model dini mana pun pada Tabel 2, namun tidak satu pun dapat dipakai untuk peringatan dini. Kedua fitur kehadiran baru lengkap setelah semester berakhir, sedangkan nilai UAS dan praktikum bukan sekadar terlambat: keduanya adalah komponen yang secara aljabar membentuk nilai akhir yang mendefinisikan kelulusan, sehingga hubungannya bersifat sirkular alih-alih prediktif.

Pola semacam ini — fitur yang sudah memuat hasil akhir menghasilkan akurasi mendekati sempurna — merupakan jenis kebocoran yang diperingatkan Kaufman dkk. [14] dan dirinci Fauszt [13], dan patut diperiksa setiap kali akurasi 95–100% dilaporkan untuk prediksi kelulusan. Bentuk paling ekstremnya terjadi ketika label didefinisikan langsung dari fitur masukan: status lulus ditetapkan berdasarkan rata-rata nilai,

sementara rata-rata itu sendiri berperan sebagai prediktor, sehingga akurasi 100% yang dilaporkan pada data sekolah kejuruan [24] mengikuti definisi labelnya.

Tabel 4. AUC fitur pasca-cutoff (kebocoran) dibandingkan fitur pra-cutoff

Fitur	Tersedia pada	<i>n</i>	AUC
Jumlah hadir 2 pertemuan pertama	Pertemuan ke-2	193	0,828
Jumlah hadir 6 pertemuan pertama	Pertemuan ke-6	193	0,885
Proporsi hadir seluruh semester	Akhir semester	193	0,946
Skor kehadiran SIAKAD (komponen nilai)	Akhir semester	193	0,925
Nilai praktikum (komponen nilai)	Akhir semester	152	0,930
Nilai UAS (komponen nilai)	Setelah UAS	157	0,959

3.6. Pembahasan dan Keterbatasan

Sebelum membahas implikasi, perlu disampaikan batas alami pendekatan ini yang tampak dari data. Recall aturan ambang bertahan di 80% dari $k = 2$ sampai $k = 4$; 16 mahasiswa yang sama tertangkap pada setiap k , dan empat kejadian yang lolos (tiga mahasiswa; satu di antaranya gagal di dua kelas sekaligus) ternyata hadir penuh sampai pertemuan keempat. Ketiganya bukan pengulang, kehadiran totalnya 10–15 dari 16 pertemuan, dan mereka gagal karena tugas individu, praktikum, UTS, atau UAS bernilai nol. Ini profil "hadir tetapi tidak mengerjakan" yang tidak dapat ditangkap oleh EWS berbasis presensi mana pun, dan menunjukkan bahwa sinyal berikutnya yang layak ditambahkan adalah pengumpulan tugas pertama, bukan kehadiran yang lebih panjang.

Ada tiga implikasi praktis dari hasil ini. Pertama, dosen tidak perlu menunggu model: daftar mahasiswa yang absen pada dua pertemuan pertama sudah menangkap empat dari lima kasus gagal, dan daftar itu bisa ditarik langsung dari SIAKAD pada pekan ketiga. Kedua, bila kapasitas tindak lanjut terbatas, model Firth dengan status mengulang memberi urutan prioritas yang lebih tajam daripada aturan biner; mahasiswa mengulang yang sudah absen pada dua pertemuan pertama adalah prioritas paling atas. Ketiga, sistem informasi akademik perlu membenahi makna status presensi, karena selama "izin" tampaknya terisi sebagai nilai bawaan, kategori itu tidak bisa dipakai untuk apa pun.

Ditemukan bahwa *disengagement* sudah tampak sebelum UTS dan bahwa kegagalan berupa komponen kosong sejalan dengan literatur EWS: indikator keterlibatan (kehadiran, pengumpulan tugas, aktivitas LMS) adalah prediktor terkuat, di atas faktor demografis [1], [15], [16]. Studi pada data LMS daring [17], [25] juga menempatkan pengumpulan tugas dan intensitas akses sebagai fitur teratas. Arah temuan tentang status mengulang sama dengan temuan Shaikhanova dkk. [15], yang menemukan risiko naik 1,26 kali per percobaan ulang; besaran pada data kami (rasio odds 9,9 [2,7–39,2]) jauh lebih besar dan intervalnya sangat lebar karena hanya ada 16 pengulang; status itu pun proksi berbasis angkatan (2.2), bukan catatan pengulangan mata kuliah. Konsekuensinya, intervensi yang masuk akal adalah pengingat dan kontak personal seperti yang dijalankan Course Signals [5], bukan remedial akademik. Meta-analisis intervensi berbasis *learning analytics* [26] mengingatkan bahwa efek intervensi semacam itu di perguruan tinggi kecil hingga moderat (*effect size* 0,32 [0,21–0,42]) dan sangat heterogen; peringatan dini hanya bernilai sejauh benar-benar ditindaklanjuti [15].

Keterbatasan penelitian ini perlu disebut secara terbuka. Jumlah kejadian (20) kecil, sehingga interval kepercayaan lebar dan kalibrasi hanya bersifat indikatif. Semua kelas diampu oleh satu dosen di satu institusi, sehingga generalisasi ke dosen dan program studi lain masih menunggu replikasi. Luarannya adalah kegagalan mata kuliah, bukan putus studi, meski keduanya berada di lintasan yang sama. Kualitas pencatatan presensi membatasi fitur itu sendiri: selain "izin" yang tampaknya menjadi nilai bawaan, pada pertemuan ketiga hampir semua kelas tercatat hadir 100%, sehingga *hadir_k* untuk $k \geq 3$ sudah tercampur derau pencatatan; ini justru memperkuat alasan memakai dua pertemuan pertama, tetapi juga berarti pola kehadiran setelahnya harus dibaca hati-hati. Hampir separuh data (81 pasangan pada empat kelas gasal dan satu kelas karyawan genap) tidak memiliki kejadian, sehingga hasil validasi bertumpu pada lima kelas genap lainnya; mengapa kelas-kelas itu, termasuk kelas semester 1 dengan 36 mahasiswa, tidak memiliki satu pun kegagalan, belum dapat dijelaskan dari data ini. Validasi antarkelas mendekati, tetapi belum sama dengan, validasi antarangkatan yang disyaratkan oleh TCDP [13]. Dua celahnya: LOCO per kelas belum sepenuhnya memisahkan entitas karena 39 mahasiswa muncul di lebih dari satu kelas, sehingga LOCO ketat dilaporkan berdampingan; dan

penggabungan kelas Gasal dan Genap dalam satu skema LOCO menyentuh syarat urutan kohort, meski tanpa dampak pada AUC karena kelas Gasal tidak memiliki kejadian. Validasi antarangkatan sendiri belum bisa dilakukan dengan data dua semester. Pengalaman Jayaprakash dkk. [6] memperlihatkan bahwa model peringatan dini dapat dipindahkan antarinstansi dengan penurunan kinerja yang moderat, sehingga replikasi lintas program studi adalah langkah lanjutan yang wajar.

4. KESIMPULAN

Data presensi yang sudah rutin terekam di SIAKAD ternyata cukup untuk menyusun peringatan dini kegagalan mata kuliah, dan temuannya sudah terbaca pada pertemuan kedua. Satu kali ketidakhadiran pada dua pertemuan pertama menandai 21,2% mahasiswa dan menjaring 80% kasus gagal; sebaliknya, 97,4% mahasiswa yang hadir penuh pada kedua pertemuan tersebut lulus, sehingga sebagian besar kelas dapat dikesampingkan dari perhatian sejak dini. Regresi logistik Firth dengan status mengulang mencapai AUC antarkelas 0,850 pada pertemuan keenam dan berguna terutama untuk menyusun prioritas ketika kapasitas tindak lanjut sangat sempit; pada kapasitas 20% ke atas, aturan ambang yang dapat dihitung oleh dosen secara manual sudah menyamainya. Fitur pasca-cutoff memang mencapai AUC 0,93–0,96, tetapi angka tersebut merupakan konsekuensi kebocoran informasi dan tidak dapat dipakai untuk peringatan dini.

Temuan ini bersandar pada 20 kegagalan di sepuluh kelas yang diampu oleh satu dosen, sehingga interval kepercayaannya lebar dan cakupan generalisasinya terbatas. Arah lanjutan yang diperlukan mencakup pengulangan protokol pada presensi seluruh program studi yang melibatkan banyak dosen dan beberapa angkatan, penautan kegagalan mata kuliah dengan status putus studi, serta pengujian kuasi-eksperimental atas pertanyaan apakah kontak dosen pada pekan ketiga sungguh menurunkan angka kegagalan.

REFERENSI

- [1] R. Balfanz, L. Herzog, and D. J. Mac Iver, "Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions," *Educ. Psychol.*, vol. 42, no. 4, pp. 223–235, 2007, doi: 10.1080/00461520701621079.
- [2] M. Credé, S. G. Roch, and U. M. Kieszczynka, "Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics," *Rev. Educ. Res.*, vol. 80, no. 2, pp. 272–295, 2010, doi: 10.3102/0034654310362998.
- [3] T. Swiderski, S. C. Fuller, and K. C. Bastian, "The relationship between student attendance and achievement, pre- and post-COVID," *AERA Open*, vol. 11, no. 1, pp. 1–19, 2025, doi: 10.1177/23328584251371041.
- [4] L. P. Macfadyen and S. Dawson, "Mining LMS data to develop an 'early warning system' for educators: A proof of concept," *Comput. Educ.*, vol. 54, no. 2, pp. 588–599, 2010, doi: 10.1016/j.compedu.2009.09.008.
- [5] K. E. Arnold and M. D. Pistilli, "Course Signals at Purdue: Using learning analytics to increase student success," in *Proc. 2nd Int. Conf. Learning Analytics and Knowledge (LAK '12)*, Vancouver, Canada, 2012, pp. 267–270, doi: 10.1145/2330601.2330666.
- [6] S. M. Jayaprakash, E. W. Moody, E. J. M. Lauría, J. R. Regan, and J. D. Baron, "Early alert of academically at-risk students: An open source analytics initiative," *J. Learn. Anal.*, vol. 1, no. 1, pp. 6–47, 2014, doi: 10.18608/jla.2014.11.3.
- [7] A. M. Romadhoni, W. P. P. Witra, and I. G. Anugrah, "Penerapan algoritma random forest sebagai early warning system prediksi kelulusan mahasiswa tepat waktu," *J. Inf. Technol. Comput. Sci. (INTECOMS)*, vol. 9, no. 3, pp. 459–470, 2026.
- [8] P. Prima, A. R. Adriansyah, and A. N. Usyaid, "Prediksi retensi mahasiswa menggunakan algoritma random forest dengan optimasi algoritma genetika," *MIND J.*, vol. 11, no. 1, pp. 115–126, 2026, doi: 10.26760/mindjournal.v11i1.115-126.
- [9] T. A. Marzuqi, E. Kristiani, and Marcel, "Prediksi mahasiswa drop-out di Universitas XYZ," *J. Teknol. Inf. Ilmu Komput. (JTIK)*, vol. 11, no. 6, pp. 1345–1350, 2024, doi: 10.25126/jtiik.2024118689.
- [10] V. Realinho, J. Machado, L. Baptista, and M. V. Martins, "Predicting student dropout and academic success," *Data*, vol. 7, no. 11, Art. no. 146, 2022, doi: 10.3390/data7110146.
- [11] F. A. Putra, S. Mirajdandi, 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, 2025, doi: 10.37859/jf.v15i3.10714.

- [12] L. Handayani and Priyadi, "A machine learning-based early warning system for student performance prediction: System development and empirical evaluation in higher education," *J. Artif. Intell. Inf. Technol. (JAI-IT)*, vol. 1, no. 1, pp. 171–190, 2026. [Online]. Available: <https://journal.perkivi.or.id/index.php/jai-it/article/view/78>
- [13] T. Fauszt, "Time-consistent prediction in higher education: A framework for preventing data leakage in longitudinal models," *Information*, vol. 17, no. 6, Art. no. 581, 2026, doi: 10.3390/info17060581.
- [14] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," *ACM Trans. Knowl. Discov. Data*, vol. 6, no. 4, pp. 1–21, 2012, doi: 10.1145/2382577.2382579.
- [15] A. Shaikhanova, O. Kuznetsov, K. Iklassova, A. Tokkuliyeva, and L. Sugurova, "Interpretable predictive modeling for educational equity: A workload-aware decision support system for early identification of at-risk students," *Big Data Cogn. Comput.*, vol. 9, no. 11, Art. no. 297, 2025, doi: 10.3390/bdcc9110297.
- [16] F. Marbouti, H. A. Diefes-Dux, and K. Madhavan, "Models for early prediction of at-risk students in a course using standards-based grading," *Comput. Educ.*, vol. 103, pp. 1–15, 2016, doi: 10.1016/j.compedu.2016.09.005.
- [17] H.-C. Chen, E. Prasetyo, S.-S. Tseng, K. T. Putra, Prayitno, S. S. Kusumawardani, and C.-E. Weng, "Week-wise student performance early prediction in virtual learning environment using a deep explainable artificial intelligence," *Appl. Sci.*, vol. 12, no. 4, Art. no. 1885, 2022, doi: 10.3390/app12041885.
- [18] D. Firth, "Bias reduction of maximum likelihood estimates," *Biometrika*, vol. 80, no. 1, pp. 27–38, 1993, doi: 10.1093/biomet/80.1.27.
- [19] G. Heinze and M. Schemper, "A solution to the problem of separation in logistic regression," *Stat. Med.*, vol. 21, no. 16, pp. 2409–2419, 2002, doi: 10.1002/sim.1047.
- [20] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein, "A simulation study of the number of events per variable in logistic regression analysis," *J. Clin. Epidemiol.*, vol. 49, no. 12, pp. 1373–1379, 1996, doi: 10.1016/S0895-4356(96)00236-3.
- [21] M. S. Rahman and M. Sultana, "Performance of Firth- and log F-type penalized methods in risk prediction for small or sparse binary data," *BMC Med. Res. Methodol.*, vol. 17, Art. no. 33, 2017, doi: 10.1186/s12874-017-0313-9.
- [22] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, p. e0118432, 2015, doi: 10.1371/journal.pone.0118432.
- [23] E. W. Steyerberg, A. J. Vickers, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M. J. Pencina, and M. W. Kattan, "Assessing the performance of prediction models: A framework for traditional and novel measures," *Epidemiology*, vol. 21, no. 1, pp. 128–138, 2010, doi: 10.1097/EDE.0b013e3181c30fb2.
- [24] A. Fatunnisa and H. Marcos, "Prediksi kelulusan tepat waktu siswa SMK teknik komputer menggunakan algoritma random forest," *J. Manaj. Inform. (JAMIKA)*, vol. 14, no. 1, pp. 101–111, 2024, doi: 10.34010/jamika.v14i1.12114.
- [25] B. Ujkani, D. Minkovska, and N. Hinov, "Course success prediction and early identification of at-risk students using explainable artificial intelligence," *Electronics*, vol. 13, no. 21, Art. no. 4157, 2024, doi: 10.3390/electronics13214157.
- [26] Y. Liu, W. Wang, and E. Xu, "The effectiveness of learning analytics-based interventions in enhancing students' learning effect: A meta-analysis of empirical studies," *SAGE Open*, vol. 15, no. 2, pp. 1–22, 2025, doi: 10.1177/21582440251336707.