Perancangan dan Pengembangan Learning Analytics Dashboard Berbasis Artificial Intelligence dan Explainable Artificial Intelligence: Studi Kasus Deteksi Dini Risiko Akademik

Ananda Widitomo Puntoadi[1], Fitri Setiawan[2], Eri Zuliarso[3]
Program Studi Magister Teknologi Informasi, Fakultas Teknologi Informasi
Universitas Stikubank (UNISBANK) Semarang, Jawa Tengah, Indonesia
e-mail: anandawiditomo@gmail.com[1], fitrisetiawan0009@mhs.unisbank.ac.id[2], eri@unisbank.ac.id[3]

Abstrak

Learning Analytics Dashboard (LAD) yang ditenagai kecerdasan buatan kerap gagal diadopsi bukan karena modelnya tidak akurat, melainkan karena keluarannya berupa skor tanpa alasan sehingga pendidik tidak memiliki dasar untuk bertindak maupun untuk menolak saran sistem. Penelitian ini merancang dan mengembangkan EduLens, sebuah LAD untuk deteksi dini mahasiswa berisiko akademik yang memadukan tiga pilar: prinsip Interaksi Manusia dan Komputer (IMK), Artificial Intelligence (AI), dan Explainable Artificial Intelligence (XAI). Sistem dibangun di atas dataset jejak digital LMS berisi 1.200 mahasiswa dengan 10 fitur perilaku belajar. Tiga algoritma diperbandingkan Logistic Regression, Random Forest, dan Gradient Boosting dan Logistic Regression terpilih dengan akurasi 0,867, F1-score 0,845, serta ROC-AUC 0,930 (validasi silang 5-lipat: 0,926 ± 0,017). Lapisan XAI menggunakan SHAP untuk menghasilkan penjelasan global maupun lokal, diperkuat permutation importance, analisis what-if, dan rekomendasi counterfactual. Prototipe memuat empat dashboard: monitoring, analitik dan insight, prediksi, serta explainability. Evaluasi antarmuka dilakukan melalui HCI Analyzer dengan tiga metrik komputasional: Balance (BM = 0,967), Density (kepadatan konten 0,666 → DM = 0,667), dan Figure-Ground Contrast menurut WCAG 2.1 (88,9% pasangan warna lolos AA; median rasio 6,11:1). Temuan substantif penelitian ini: frekuensi login metrik yang paling lazim dipakai pendidik justru merupakan prediktor terlemah (mean |SHAP| 0,103), sedangkan ketuntasan tugas adalah yang terkuat (0,646). Analisis HCI juga mengungkap dua pasangan warna keterangan yang belum memenuhi ambang aksesibilitas dan telah direkomendasikan perbaikannya.

Kata kunci: *learning analytics dashboard; explainable AI; SHAP; interaksi manusia dan komputer; deteksi dini risiko akademik; WCAG*

Abstract

AI-powered Learning Analytics Dashboards (LADs) often fail to gain adoption not because their models are inaccurate, but because they emit scores without reasons, leaving educators with no basis to act upon or to overrule the system's advice. This study designs and develops EduLens, a LAD for the early detection of at-risk students that integrates three pillars: Human-Computer Interaction (HCI) principles, Artificial Intelligence (AI), and Explainable AI (XAI). The system is built on an LMS digital-trace dataset of 1,200 students across 10 behavioural features. Three algorithms were compared Logistic Regression, Random Forest, and Gradient Boosting with Logistic Regression selected at 0.867 accuracy, 0.845 F1-score, and 0.930 ROC-AUC (5-fold CV: 0.926 ± 0.017). The XAI layer employs SHAP for both global and local explanations, reinforced by permutation importance, what-if analysis, and counterfactual recommendations. The prototype comprises four dashboards: monitoring, analytics and insight, prediction, and explainability. Interface evaluation was conducted through an HCI Analyzer using three computational metrics: Balance (BM = 0.967), Density (content density 0.666 → DM = 0.667), and Figure-Ground Contrast per WCAG 2.1 (88.9% of colour pairs pass AA; median ratio 6.11:1). A substantive finding: login frequency the metric educators most commonly rely on is the weakest predictor (mean |SHAP| 0.103), whereas assignment completion is the strongest (0.646).

Keywords: *learning analytics dashboard; explainable AI; SHAP; human-computer interaction; early warning system; WCAG*

I. PENDAHULUAN

Perpindahan sebagian besar aktivitas perkuliahan ke Learning Management System (LMS) menghasilkan jejak digital yang sangat kaya: setiap unggahan tugas, pemutaran video, percobaan kuis, dan sesi login terekam dengan stempel waktu. Learning analytics lahir dari premis bahwa jejak tersebut dapat dibaca kembali untuk memahami dan memperbaiki proses belajar [1]. Learning Analytics Dashboard (LAD) adalah manifestasi antarmukanya lapisan visual yang menerjemahkan tumpukan log menjadi sesuatu yang dapat ditindaklanjuti oleh dosen [2].

Persoalannya, sebagian besar LAD berhenti sebagai papan pemantau deskriptif: berapa banyak yang login, berapa persen materi dibuka, berapa rerata nilai. Ketika kecerdasan buatan mulai disematkan misalnya untuk memprediksi mahasiswa yang berpotensi gagal muncul persoalan baru yang lebih pelik. Model prediktif berkinerja tinggi umumnya bersifat kotak hitam. Sistem hanya menyodorkan angka "risiko 0,87" tanpa menjelaskan mengapa. Dosen dihadapkan pada dua pilihan yang sama-sama buruk: mempercayai angka itu secara buta, atau mengabaikannya sama sekali. Keduanya menggagalkan tujuan sistem.

Di sinilah Explainable Artificial Intelligence (XAI) menjadi kebutuhan struktural, bukan sekadar pemanis. Penjelasan bukan hanya soal transparansi teknis; ia adalah prasyarat agar manusia dapat menjalankan perannya sebagai pengambil keputusan akhir [3], [4]. Namun XAI saja belum cukup: nilai SHAP dalam satuan log-odds tidak bermakna apa pun bagi seorang dosen wali. Penjelasan harus melewati satu lapisan lagi perancangan antarmuka yang berbasis prinsip Interaksi Manusia dan Komputer (IMK) sebelum ia benar-benar dapat digunakan.

Ketiga lapisan itulah yang dipadukan dalam penelitian ini. Kontribusi yang diklaim ada tiga:

- Rancang bangun EduLens, sebuah LAD empat-dashboard (monitoring, analitik dan insight, prediksi, explainability) yang menjadikan penjelasan sebagai warga kelas satu, bukan panel tambahan.
- Integrasi XAI yang bergerak dari penjelasan pasif (SHAP global dan lokal) ke penjelasan aktif dan aktionabel (simulasi what-if dan rekomendasi counterfactual) yang berjalan langsung di antarmuka.
- Evaluasi antarmuka secara komputasional dan terukur melalui HCI Analyzer dengan tiga metrik dari dimensi berbeda komposisi visual, beban kognitif, dan aksesibilitas bukan sekadar klaim kualitatif bahwa antarmuka "mudah digunakan".

Pertanyaan penelitian yang dijawab: (RQ1) Sejauh mana model prediksi risiko akademik berbasis jejak LMS dapat mencapai kinerja yang layak dipakai? (RQ2) Fitur perilaku mana yang benar-benar menentukan risiko, dan apakah ia sejalan dengan intuisi pendidik? (RQ3) Sejauh mana rancangan antarmuka EduLens memenuhi kriteria kualitas HCI yang terukur?

II. TINJAUAN PUSTAKA

A. Learning Analytics Dashboard

Siemens dan Gašević mendefinisikan learning analytics sebagai pengukuran, pengumpulan, analisis, dan pelaporan data tentang pembelajar dan konteksnya untuk memahami serta mengoptimalkan pembelajaran [1]. Verbert dkk. memetakan LAD sebagai antarmuka yang mengagregasi jejak tersebut menjadi visualisasi yang mendukung refleksi dan pengambilan keputusan [2]. Tinjauan Bodily dan Verbert [5] atas LAD yang berorientasi mahasiswa menemukan kesenjangan yang konsisten: mayoritas dashboard berhenti pada deskripsi, sangat sedikit yang prediktif, dan hampir tidak ada yang menjelaskan dasar prediksinya.

Course Signals di Purdue University adalah salah satu penerapan awal yang paling banyak dikutip: sistem menandai mahasiswa dengan lampu lalu lintas hijau, kuning, dan merah [6]. Sistem itu efektif menaikkan angka retensi, tetapi juga memperlihatkan batas pendekatan tanpa penjelasan dosen mengetahui bahwa

seorang mahasiswa berwarna merah, namun tidak mengetahui apa yang membuatnya merah, sehingga intervensi cenderung generik.

B. Explainable Artificial Intelligence

Adadi dan Berrada [7] mengklasifikasikan pendekatan XAI ke dalam metode intrinsik (model yang memang transparan, seperti regresi logistik dan pohon keputusan) dan post-hoc (penjelasan yang dilekatkan pada model kotak hitam yang sudah terlatih). SHAP (SHapley Additive exPlanations) yang dikembangkan Lundberg dan Lee [8] menyatukan beberapa metode post-hoc di bawah kerangka teori permainan koperatif. Untuk sebuah prediksi tunggal, kontribusi setiap fitur dihitung sebagai nilai Shapley, sehingga bersifat aditif:

$$f(x) = \varphi_0 + \Sigma\square\ \varphi\square$$

dengan φ_0 sebagai nilai dasar (rerata prediksi pada kohort) dan $\varphi\square$ sebagai kontribusi fitur ke-j. Sifat aditif inilah yang membuat SHAP dapat divisualisasikan sebagai gaya dorong dan tarik terhadap nilai dasar properti yang dieksploitasi langsung pada rancangan antarmuka penelitian ini. LIME [9] menawarkan alternatif berbasis aproksimasi lokal, namun tidak menjamin konsistensi aditif. Wachter dkk. [10] melengkapi lanskap ini dengan penjelasan counterfactual: alih-alih menjawab "mengapa", counterfactual menjawab "apa yang harus berubah agar hasilnya berbeda" bentuk penjelasan yang secara langsung dapat diterjemahkan menjadi tindakan.

C. Prinsip IMK dan Metrik Antarmuka Komputasional

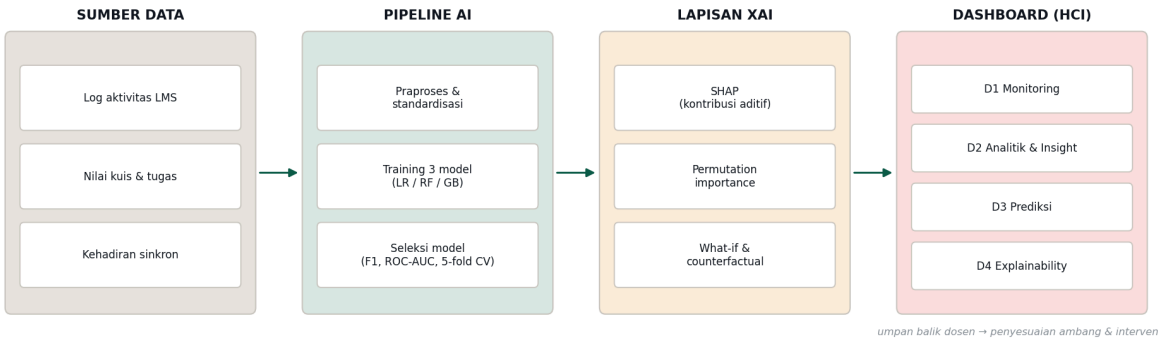
Shneiderman merumuskan visual information-seeking mantra yang menjadi tulang punggung perancangan dashboard: overview first, zoom and filter, then details on demand [11]. Sepuluh heuristik Nielsen [12] menyediakan kerangka inspeksi kegunaan yang telah teruji lama. Dalam konteks sistem berbasis AI, Shneiderman [13] menekankan bahwa otomasi tinggi dan kendali manusia yang tinggi bukanlah dua kutub yang saling meniadakan sistem yang baik memaksimalkan keduanya sekaligus.

Selain inspeksi kualitatif, kualitas antarmuka dapat diukur secara komputasional. Ngo, Teo, dan Byrne [14] memformalkan sejumlah ukuran estetika tata letak di antaranya Balance, Equilibrium, Symmetry, Density, Simplicity, dan Economy yang dihitung langsung dari kotak pembatas komponen. Oulasvirta dkk. [15] memperluas pendekatan ini melalui Aalto Interface Metrics, yang menambahkan metrik persepsi visual dan aksesibilitas. Dari khazanah tujuh belas metrik inilah tiga metrik dipilih pada penelitian ini, dengan pertimbangan yang diuraikan pada Bagian III-E.

III. METODE PENELITIAN

A. Rancangan Umum

Penelitian ini menggunakan pendekatan design science research: sebuah artefak dirancang, dibangun, lalu dievaluasi secara sistematis. Alur sistem terdiri atas empat lapisan yang berurutan sumber data, pipeline AI, lapisan XAI, dan dashboard dengan umpan balik dari dosen yang kembali menyesuaikan ambang keputusan serta bentuk intervensi. Arsitekturnya disajikan pada Gambar 1.



Gambar 1. Arsitektur sistem EduLens: dari jejak LMS hingga dashboard, dengan lapisan XAI sebagai perantara wajib antara model dan pengguna.

B. Dataset dan Fitur

Dataset terdiri atas 1.200 mahasiswa dengan sepuluh fitur jejak digital LMS. Data dibangkitkan secara sintetis melalui proses generatif terkendali (seed 42): sebuah variabel laten keterlibatan (engagement) dibangkitkan lebih dahulu, lalu seluruh fitur diturunkan darinya dengan derau independen, sehingga struktur korelasi antar-fitur menyerupai kondisi nyata. Label risiko dibangkitkan melalui proses logistik dengan bobot yang mencerminkan temuan literatur ketuntasan tugas dan performa formatif diberi bobot dominan, aktivitas login diberi bobot lemah. Penggunaan data sintetis dipilih atas pertimbangan etika dan perlindungan data pribadi; konsekuensinya dibahas secara terbuka pada Bagian V-B. Distribusi kelas relatif seimbang: 673 mahasiswa aman (56,1%) dan 527 berisiko (43,9%). Data dibagi secara stratified 70:30 menjadi 840 data latih dan 360 data uji.

Tabel 1. Fitur jejak digital LMS yang digunakan

No	Fitur	Deskripsi	Rentang
1	Frekuensi login/minggu	Jumlah sesi login ke LMS	0 – 25
2	Durasi belajar (jam/mgg)	Total waktu aktif di LMS	0,2 – 22
3	Materi diakses (%)	Proporsi modul yang dibuka	3 – 100
4	Video ditonton (%)	Proporsi durasi video yang ditonton	0 – 100
5	Tugas terkumpul (%)	Proporsi tugas yang dikumpulkan	0 – 100
6	Rerata keterlambatan (hari)	Rerata hari keterlambatan pengumpulan	0 – 14
7	Rerata nilai kuis	Skor kuis formatif	15 – 100
8	Kontribusi forum	Jumlah unggahan di forum diskusi	0 – 30
9	Kehadiran sinkron (%)	Proporsi kehadiran kelas daring	0 – 100
10	Streak hari aktif	Rentetan hari aktif terpanjang	0 – 30

C. Pemodelan dan Evaluasi

Tiga algoritma klasifikasi diperbandingkan: Logistic Regression, Random Forest (300 pohon, kedalaman maksimum 8), dan Gradient Boosting (250 estimator, laju belajar 0,06). Seluruh implementasi menggunakan scikit-learn [16] dengan seed 42 agar hasil dapat direplikasi. Fitur distandardisasi menggunakan z-score untuk model linier. Evaluasi dilakukan pada data uji yang tidak pernah dilihat model, menggunakan akurasi, presisi, recall, F1-score, ROC-AUC, dan PR-AUC, serta divalidasi dengan stratified 5-fold cross-validation untuk memastikan hasil tidak bergantung pada satu pembagian data tertentu.

Dalam konteks deteksi dini, recall memiliki bobot etis yang lebih berat daripada presisi: seorang mahasiswa berisiko yang terlewat (false negative) berpotensi gagal tanpa pernah tersentuh bantuan, sedangkan sebuah alarm palsu (false positive) hanya berbiaya satu percakapan tambahan. Karena itu antarmuka menyediakan kendali ambang keputusan yang dapat digeser dosen, bukan mengunci ambang di 0,5.

D. Lapisan XAI

Empat mekanisme penjelasan diterapkan secara berlapis. Pertama, SHAP menghasilkan penjelasan global (rerata nilai absolut SHAP di seluruh kohort uji) dan lokal (kontribusi tiap fitur pada satu mahasiswa tertentu). Kedua, permutation importance dihitung dengan 20 repetisi sebagai validasi silang metodologis bila kedua metode yang berbeda asumsi menghasilkan peringkat fitur yang konvergen, keyakinan terhadap temuan meningkat. Ketiga, simulasi what-if memungkinkan dosen menggeser nilai fitur dan menyaksikan prediksi dihitung ulang secara langsung oleh model yang sama. Keempat, rekomendasi counterfactual mencari perubahan tunggal terkecil yang cukup untuk menurunkan prediksi ke bawah ambang, dengan pembatas rentang agar saran yang dihasilkan tetap masuk akal secara praktis.

E. HCI Analyzer: Pemilihan dan Definisi Tiga Metrik

Dari tujuh belas metrik antarmuka komputasional yang tersedia, dipilih tiga metrik yang secara sengaja mewakili tiga dimensi HCI yang berbeda, agar analisis bersifat komprehensif dan tidak redundan. Memilih tiga metrik dari keluarga yang sama (misalnya Balance, Symmetry, dan Equilibrium yang ketiganya mengukur komposisi) hanya akan mengukur hal yang sama tiga kali.

Tabel 2. Tiga metrik terpilih dan justifikasinya

Metrik	Dimensi HCI	Rujukan	Alasan pemilihan
M1 Balance (BM)	Komposisi visual	Ngo, Teo & Byrne (2003)	Dashboard multipanel rawan berat sebelah; ketidakseimbangan memaksa mata bekerja lebih keras.
M2 Density (DM)	Beban kognitif	Ngo, Teo & Byrne (2003)	LAD menyajikan banyak informasi sekaligus; kepadatan berlebih adalah penyebab utama kegagalan dashboard.
M3 Figure-Ground Contrast	Aksesibilitas	WCAG 2.1 (W3C); AIM	Warna adalah pembawa makna utama (risiko rendah/sedang/tinggi); bila kontras gagal, makna itu hilang.

Balance dihitung dari bobot visual setiap komponen (luas dikalikan jarak pusat komponen ke pusat kanvas), dijumlahkan pada sisi kiri, kanan, atas, dan bawah:

$$\mathbf{BM = 1 - (|BM_{vertikal}| + |BM_{horizontal}|) / 2}$$

Density dihitung sebagai proporsi kanvas yang terisi objek, dinormalisasi terhadap kepadatan optimal 0,5 menurut Ngo dkk.: DM = 1 − |2d − 1|. Metrik ini diukur pada dua tingkat granularitas tingkat kontainer (seluruh kartu dihitung sebagai objek padat) dan tingkat konten (hanya elemen konten atomik yang dihitung, chrome kartu dikecualikan) karena, sebagaimana dibahas pada Bagian IV-D, granularitas pengukuran ternyata mengubah kesimpulan secara drastis.

Figure-Ground Contrast dihitung menurut WCAG 2.1 [17] sebagai rasio luminansi relatif antara warna latar depan dan latar belakang, $(L_1 + 0{,}05) / (L_2 + 0{,}05)$, pada delapan belas pasangan warna yang benar-benar digunakan dalam kode prototipe. Ambang kelulusan mengikuti tingkat AA: 4,5:1 untuk teks normal serta 3,0:1 untuk teks besar dan komponen antarmuka.

IV. HASIL DAN PEMBAHASAN

A. Prototipe EduLens

EduLens diimplementasikan sebagai aplikasi web berbasis React dengan empat dashboard yang tersusun mengikuti mantra Shneiderman: overview terlebih dahulu, lalu penyaringan, baru detail sesuai permintaan. Rancangan setiap dashboard dipetakan secara eksplisit terhadap prinsip IMK yang mendasarinya, sebagaimana dirangkum pada Tabel 3.

Tabel 3. Pemetaan dashboard terhadap prinsip IMK

Dashboard	Isi	Prinsip IMK yang diterapkan
D1 Monitoring	Kartu KPI, tren keterlibatan mingguan, distribusi risiko, peta sebaran mahasiswa	Overview first; visibility of system status; pengenalan pola secara pra-atentif melalui warna dan posisi
D2 Analitik & Insight	Kartu insight terkurasi, profil perilaku per kohort, permutation importance, rekomendasi intervensi	Recognition rather than recall; insight disajikan sebagai kalimat, bukan sebagai grafik yang harus ditafsirkan sendiri
D3 Prediksi	Kendali ambang keputusan, confusion matrix langsung, tabel prediksi, perbandingan model	User control and freedom; kendali atas trade-off diserahkan kepada dosen, bukan dikunci oleh sistem
D4 Explainability	SHAP global, force ribbon dan waterfall lokal, simulasi what-if, counterfactual, catatan batas penggunaan	Details on demand; help users understand; human-in-the-loop; kejujuran mengenai batas sistem

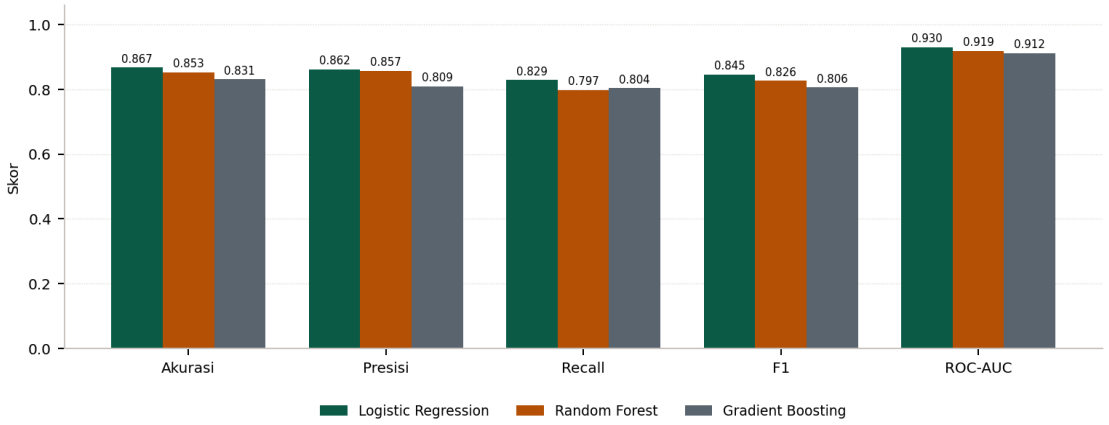
Satu keputusan rancangan patut disorot. Pada D3, ambang keputusan tidak dikunci pada 0,5 melainkan disediakan sebagai penggeser, dan confusion matrix dihitung ulang secara langsung setiap kali penggeser digerakkan. Konsekuensi dari sebuah pilihan menjadi kasatmata seketika: menurunkan ambang memperkecil jumlah mahasiswa yang terlewat, tetapi memperbesar beban intervensi. Trade-off itu bersifat pedagogis dan etis, bukan teknis karena itu ia dikembalikan kepada manusia.

B. Kinerja Model (RQ1)

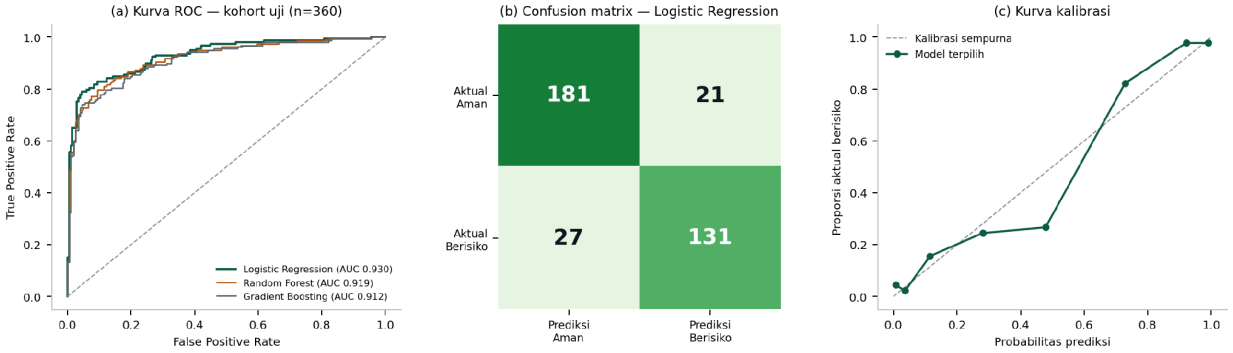
Ketiga model mencapai kinerja yang layak, dengan Logistic Regression unggul pada seluruh metrik utama (Tabel 4, Gambar 2).

Tabel 4. Perbandingan kinerja model pada data uji (n = 360)

Model	Akurasi	Presisi	Recall	F1	ROC-AUC	CV ROC-AUC
Logistic Regression	**0,867**	**0,862**	**0,829**	**0,845**	**0,930**	**0,926 ± 0,017**
Random Forest	0,853	0,857	0,797	0,826	0,919	0,923 ± 0,018
Gradient Boosting	0,831	0,809	0,804	0,806	0,912	0,908 ± 0,018



Gambar 2. Perbandingan kinerja tiga algoritma pada lima metrik evaluasi.

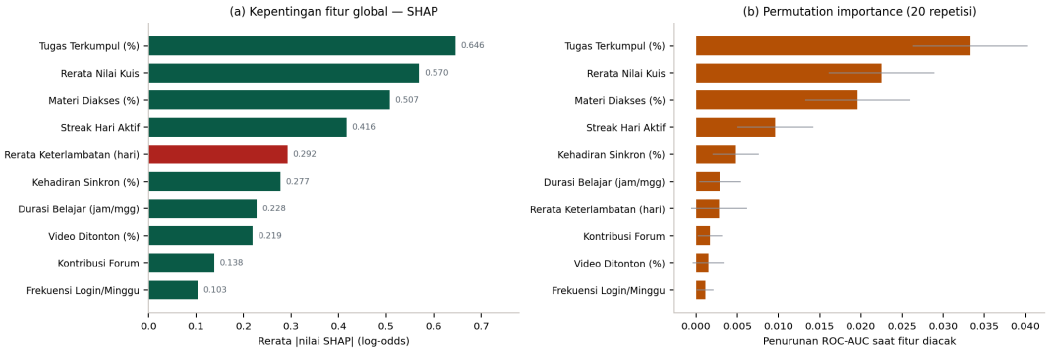


Gambar 3. (a) Kurva ROC ketiga model; (b) confusion matrix model terpilih; (c) kurva kalibrasi.

Confusion matrix model terpilih menunjukkan 181 true negative, 131 true positive, 21 false positive, dan 27 false negative. Dua puluh tujuh mahasiswa berisiko yang terlewat pada ambang 0,5 adalah angka yang tidak boleh dianggap remeh dan justru itulah alasan penggeser ambang disediakan pada antarmuka. Kurva kalibrasi (Gambar 3c) menunjukkan bahwa probabilitas keluaran model dapat dibaca apa adanya: ketika model mengatakan 0,73, proporsi aktual mahasiswa berisiko pada kelompok itu adalah 0,82. Ini penting secara HCI sebuah angka probabilitas yang tidak terkalibrasi akan menyesatkan pengguna yang mempercayainya secara harfiah.

Menariknya, model paling sederhana justru menang. Regresi logistik mengungguli dua model ensembel yang jauh lebih kompleks, dan keunggulannya bukan hanya pada angka: karena model bersifat linier, nilai SHAP yang dihasilkan bersifat eksak, bukan aproksimasi. Akurasi dan interpretabilitas dalam kasus ini tidak berkonflik sebuah pengingat bahwa asumsi "harus mengorbankan penjelasan demi akurasi" tidak selalu benar.

C. Temuan XAI (RQ2)

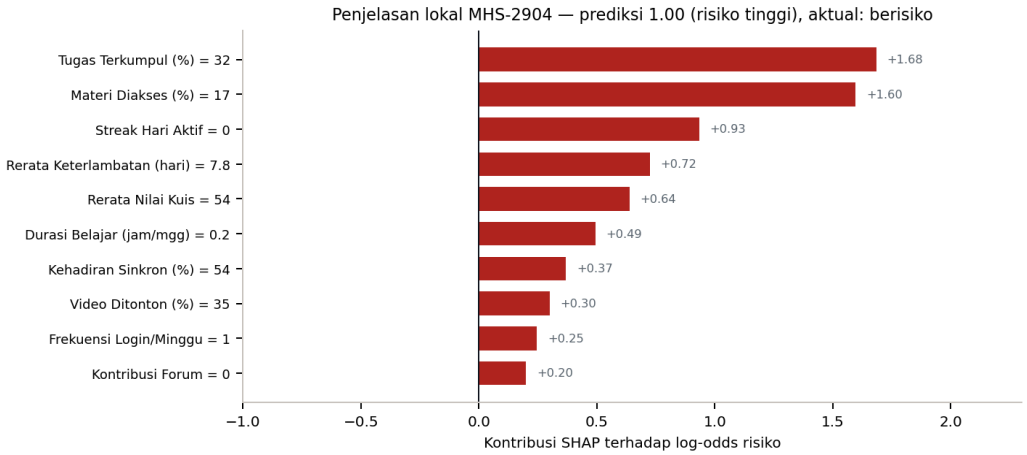


Gambar 4. (a) Kepentingan fitur global menurut SHAP; (b) permutation importance sebagai validasi silang.

Analisis SHAP global (Gambar 4a) menempatkan ketuntasan tugas sebagai prediktor terkuat (mean |SHAP| 0,646), disusul rerata nilai kuis (0,570), materi diakses (0,508), dan streak hari aktif (0,416). Permutation importance (Gambar 4b) menghasilkan peringkat empat besar yang identik konvergensi dua metode dengan asumsi berbeda ini memperkuat keyakinan terhadap temuan.

Temuan yang paling layak dibahas justru berada di dasar peringkat. Frekuensi login menempati posisi terlemah pada kedua metode (mean |SHAP| 0,103; penurunan ROC-AUC hanya 0,0011 ± 0,0010 ketika fitur diacak). Padahal, dalam praktik sehari-hari, frekuensi login adalah metrik yang paling sering dipandangi dosen ia paling mudah dilihat di halaman muka LMS. Jika sebuah LAD hanya menampilkan aktivitas login, ia menampilkan justru sinyal yang paling tidak informatif. Rajin membuka LMS bukanlah belajar. Temuan ini secara langsung menjawab RQ2 dan sekaligus menjadi argumen paling kuat mengapa lapisan XAI diperlukan: tanpa penjelasan, asumsi keliru semacam ini akan terus bertahan tanpa pernah terkoreksi.

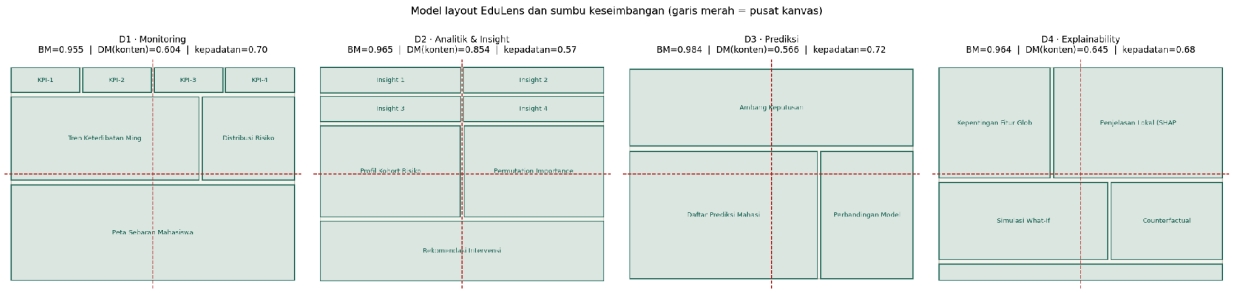
Streak hari aktif (0,416) juga mengungguli durasi belajar total (0,228) secara meyakinkan. Konsistensi lebih protektif daripada intensitas sebuah temuan yang selaras dengan literatur pembelajaran terdistribusi dan berimplikasi praktis: intervensi sebaiknya menyasar pemulihan ritme belajar, bukan menambah jam belajar.



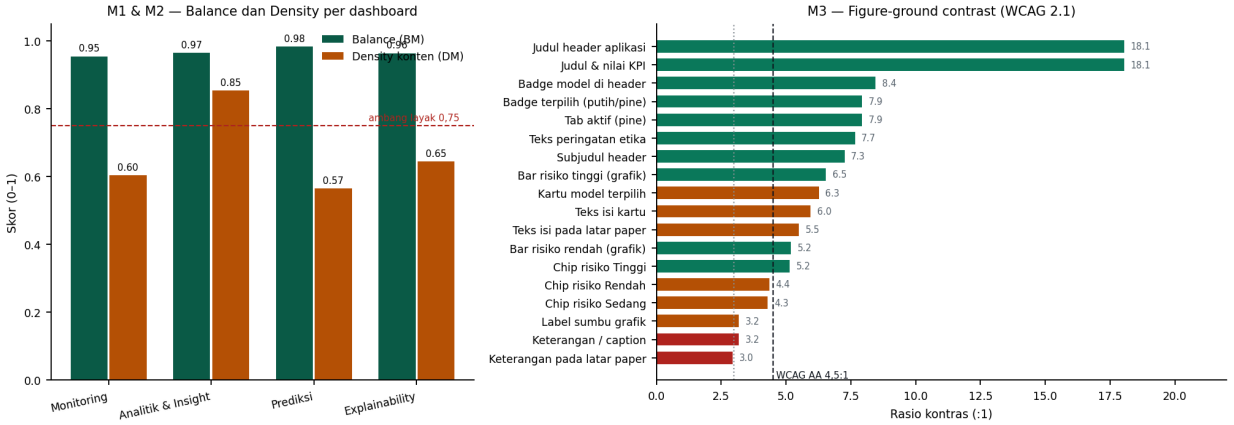
Gambar 5. Penjelasan lokal SHAP untuk MHS-2904 (prediksi 0,999; aktual berisiko).

Pada tingkat individu (Gambar 5), penjelasan menjadi konkret dan langsung dapat ditindaklanjuti. Mahasiswa MHS-2904 diprediksi berisiko dengan probabilitas 0,999. Kontribusi terbesar berasal dari ketuntasan tugas yang hanya 32% (+1,68 log-odds) dan materi diakses 17% (+1,60), diikuti streak hari aktif bernilai nol (+0,93). Dosen wali yang melihat panel ini tidak sekadar mengetahui bahwa mahasiswa tersebut merah; ia mengetahui apa yang merah, sehingga percakapan dapat dimulai dari titik yang tepat. Panel counterfactual menyempurnakannya dengan menghitung perubahan tunggal terkecil yang cukup untuk menurunkan prediksi ke bawah ambang mengubah penjelasan menjadi rencana.

D. Analisis HCI Analyzer (RQ3)



Gambar 6. Model layout keempat dashboard beserta sumbu keseimbangan.



Gambar 7. Hasil pengukuran tiga metrik HCI.

Tabel 5. Hasil HCI Analyzer pada keempat dashboard

Dashboard	M1 Balance	M2 Kepadatan konten	M2 Density (DM)	Keterangan
D1 Monitoring	0,955	0,698	0,604	Seimbang; padat
D2 Analitik & Insight	0,965	0,573	0,854	Terbaik pada kedua metrik
D3 Prediksi	0,984	0,717	0,566	Paling seimbang; paling padat
D4 Explainability	0,964	0,677	0,645	Seimbang; padat sedang
Rerata	**0,967**	**0,666**	**0,667**	—

M1 Balance. Rerata BM sebesar 0,967 (rentang 0,955–0,984) berada jauh di atas ambang kelayakan 0,75 yang lazim dipakai. Ini bukan kebetulan melainkan konsekuensi langsung dari pemakaian sistem grid dua belas kolom yang konsisten. Dashboard paling seimbang adalah D3 (0,984) yang hanya memuat tiga blok besar; sedikit ketidakseimbangan horizontal pada D1 (−0,076) berasal dari kartu peta sebaran berukuran besar di bagian bawah kanvas, dan secara sengaja dipertahankan bobot visual yang lebih berat di bawah justru memberi kesan berpijak (grounded) dan tidak mengganggu pemindaian.

M2 Density. Metrik ini menghasilkan temuan metodologis yang tidak diduga. Diukur pada tingkat kontainer, kepadatan mencapai 0,950 sehingga DM anjlok ke 0,100 bila dilaporkan begitu saja, akan menyimpulkan bahwa antarmuka gagal total. Namun kesimpulan itu keliru: pada dashboard, kartu adalah wadah yang bagian dalamnya justru didominasi ruang kosong. Setelah chrome kartu (border, padding, header) dikeluarkan dan hanya elemen konten atomik yang dihitung, kepadatan sebenarnya adalah 0,666

dengan DM = 0,667 dan ruang putih efektif 33,4%. Nilai optimal 0,5 yang diusulkan Ngo dkk. dikalibrasi pada layar entri data pada era 1990-an, bukan pada dashboard analitik yang secara hakikat memang padat informasi. Pelajarannya bersifat umum: metrik komputasional tidak boleh dipakai sebagai vonis tanpa memeriksa granularitas pengukurannya.

Meskipun demikian, temuan ini tetap menghasilkan tindakan nyata. D3 Prediksi adalah yang terpadat (0,717) karena memuat tabel panjang; rekomendasi perbaikannya adalah menambah tinggi baris tabel dan memberi jeda antar-blok. Sebaliknya D2 (0,573) mendekati optimum dan dijadikan acuan internal untuk penyesuaian dashboard lain.

M3 Figure-Ground Contrast. Dari 18 pasangan warna yang benar-benar dipakai dalam kode, 16 pasangan (88,9%) lolos WCAG AA dan 10 pasangan (55,6%) bahkan lolos AAA, dengan median rasio 6,11:1. Namun dua pasangan gagal, dan keduanya melibatkan warna keterangan #8A9199: pada latar putih rasionya 3,19:1 dan pada latar #F7F6F3 hanya 2,95:1 keduanya di bawah ambang 4,5:1 untuk teks normal.

Temuan ini penting karena warna tersebut dipakai pada teks caption yang menjelaskan makna setiap panel justru teks yang paling dibutuhkan oleh pengguna yang belum terbiasa. Rekomendasi perbaikan: mengganti #8A9199 menjadi #6B7280 (rasio 4,83:1 pada latar putih) untuk seluruh teks berukuran normal, sambil mempertahankan #8A9199 hanya untuk label sumbu grafik, yang secara kategori WCAG tergolong komponen antarmuka dengan ambang 3,0:1 sehingga tetap memenuhi syarat. Sisi positifnya, seluruh chip risiko pembawa makna terpenting dalam sistem ini lolos dengan aman (Rendah 4,36:1; Sedang 4,30:1; Tinggi 5,15:1), dan seluruhnya juga dibedakan oleh label teks, bukan hanya oleh warna, sehingga tetap terbaca oleh pengguna dengan buta warna.

Tabel 6. Ringkasan hasil ketiga metrik HCI

Metrik	Hasil	Ambang	Status
M1 Balance (BM)	0,967	≥ 0,75	Sangat baik
M2 Density (DM, konten)	0,667	≥ 0,60	Baik, perlu penyesuaian D3
M3 Kontras lolos AA	88,9% (16/18)	100%	Perlu perbaikan 2 pasangan
M3 Kontras median	6,11:1	≥ 4,5:1	Sangat baik

E. Pembahasan

Hasil ketiga metrik menjawab RQ3 secara terukur: antarmuka EduLens sangat kuat pada komposisi visual, memadai pada kepadatan informasi, dan hampir memenuhi standar aksesibilitas dengan dua cacat spesifik yang telah teridentifikasi berikut perbaikannya. Yang perlu digarisbawahi, nilai dari analisis ini bukanlah skornya yang tinggi, melainkan bahwa ia menemukan cacat yang tidak terlihat oleh mata telanjang. Dua pasangan warna yang gagal tersebut lolos dari inspeksi visual manual teksnya tampak "cukup terbaca" bagi perancang yang melihatnya di layar terang dengan penglihatan normal. Hanya pengukuran yang mampu menangkapnya. Inilah argumen bagi evaluasi antarmuka komputasional sebagai pelengkap, bukan pengganti, evaluasi berbasis pengguna.

Pada tataran yang lebih luas, penelitian ini memperlihatkan bahwa AI, XAI, dan IMK bukanlah tiga komponen yang ditumpuk melainkan tiga lapisan yang saling menopang. Model yang akurat tanpa penjelasan menghasilkan skor yang tak dapat ditindaklanjuti. Penjelasan tanpa perancangan antarmuka menghasilkan nilai log-odds yang tak terbaca. Dan antarmuka yang indah tanpa model yang benar hanya memperindah kekeliruan. Ketiganya hanya bermakna secara bersamaan.

V. KESIMPULAN

A. Kesimpulan

Penelitian ini berhasil merancang dan mengembangkan EduLens, sebuah Learning Analytics Dashboard yang memadukan prinsip IMK, AI, dan XAI untuk deteksi dini risiko akademik. Model Logistic Regression yang terpilih mencapai akurasi 0,867, F1-score 0,845, dan ROC-AUC 0,930 dengan validasi silang 0,926 ± 0,017 (RQ1). Analisis XAI mengungkap bahwa ketuntasan tugas adalah prediktor terkuat (mean |SHAP| 0,646) sementara frekuensi login metrik yang paling lazim dipandangi pendidik justru yang terlemah (0,103), sebuah temuan yang dikonfirmasi silang oleh permutation importance (RQ2). Evaluasi HCI Analyzer dengan tiga metrik menghasilkan Balance 0,967, Density konten 0,667, dan kelulusan kontras WCAG AA sebesar 88,9%, disertai identifikasi dua cacat aksesibilitas spesifik beserta perbaikannya (RQ3).

B. Keterbatasan

- Data bersifat sintetis. Meskipun proses generatifnya terkendali dan berbasis literatur, model belum divalidasi pada jejak LMS nyata. Validasi eksternal pada dataset publik seperti OULAD [18] atau pada data LMS institusional merupakan langkah lanjutan yang wajib sebelum sistem digunakan dalam keputusan sesungguhnya.
- Evaluasi antarmuka masih bersifat komputasional. Metrik Balance, Density, dan Contrast mengukur properti objektif antarmuka, bukan pengalaman pengguna. Pengujian dengan pengguna nyata menggunakan System Usability Scale [19] dan NASA-TLX diperlukan untuk memverifikasi bahwa skor yang baik benar-benar berbanding lurus dengan kegunaan yang dirasakan.
- Fitur terbatas pada jejak digital. Kondisi kesehatan, ekonomi, dan keluarga yang seringkali menjadi penyebab sesungguhnya dari kemerosotan akademik tidak terekam oleh LMS. Karena itu sistem ini secara eksplisit diposisikan sebagai pemicu percakapan, bukan sebagai penentu keputusan.

C. Saran Pengembangan

Arah pengembangan berikutnya mencakup validasi model pada data LMS nyata dengan pembagian temporal, penambahan prediksi deret waktu agar risiko dapat dipantau perkembangannya dari minggu ke minggu, pengujian kegunaan dengan dosen sebagai partisipan, serta audit keadilan (fairness) untuk memastikan model tidak memberi kerugian sistematis kepada subkelompok mahasiswa tertentu.

DAFTAR PUSTAKA

[1] G. Siemens and D. Gašević, "Guest editorial Learning and knowledge analytics," Educational Technology & Society, vol. 15, no. 3, pp. 1–2, 2012.

[2] K. Verbert, E. Duval, J. Klerkx, S. Govaerts, and J. L. Santos, "Learning analytics dashboard applications," American Behavioral Scientist, vol. 57, no. 10, pp. 1500–1509, 2013.

[3] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.

[4] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd ed., 2022.

[5] R. Bodily and K. Verbert, "Review of research on student-facing learning analytics dashboards and educational recommender systems," IEEE Transactions on Learning Technologies, vol. 10, no. 4, pp. 405–418, 2017.

[6] K. E. Arnold and M. D. Pistilli, "Course Signals at Purdue: Using learning analytics to increase student success," in Proc. 2nd Int. Conf. on Learning Analytics and Knowledge (LAK '12), 2012, pp. 267–270.

[7] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI)," IEEE Access, vol. 6, pp. 52138–52160, 2018.

[8] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems 30 (NeurIPS), 2017, pp. 4765–4774.

[9] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.

[10] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual explanations without opening the black box: Automated decisions and the GDPR," Harvard Journal of Law & Technology, vol. 31, no. 2, pp. 841–887, 2018.

[11] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in Proc. IEEE Symposium on Visual Languages, 1996, pp. 336–343.

[12] J. Nielsen, "Heuristic evaluation," in Usability Inspection Methods, J. Nielsen and R. L. Mack, Eds. New York: John Wiley & Sons, 1994, pp. 25–62.

[13] B. Shneiderman, "Human-centered artificial intelligence: Reliable, safe & trustworthy," International Journal of Human–Computer Interaction, vol. 36, no. 6, pp. 495–504, 2020.

[14] D. C. L. Ngo, L. S. Teo, and J. G. Byrne, "Modelling interface aesthetics," Information Sciences, vol. 152, pp. 25–46, 2003.

[15] A. Oulasvirta et al., "Aalto Interface Metrics (AIM): A service and codebase for computational GUI evaluation," in Proc. 31st ACM Symposium on User Interface Software and Technology (UIST '18 Adjunct), 2018, pp. 16–19.

[16] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[17] World Wide Web Consortium (W3C), "Web Content Accessibility Guidelines (WCAG) 2.1," W3C Recommendation, 2018.

[18] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open University Learning Analytics dataset," Scientific Data, vol. 4, art. 170171, 2017.

[19] J. Brooke, "SUS: A 'quick and dirty' usability scale," in Usability Evaluation in Industry, P. W. Jordan et al., Eds. London: Taylor & Francis, 1996, pp. 189–194.

[20] S. Few, Information Dashboard Design: The Effective Visual Communication of Data. Sebastopol, CA: O'Reilly Media, 2006.

LAMPIRAN A. REPRODUSIBILITAS

Seluruh angka pada naskah ini dihasilkan oleh skrip yang dapat dijalankan ulang dan menghasilkan keluaran identik (seed 42):

Tabel A1. Artefak penelitian

Berkas	Fungsi	Keluaran
train_model.py	Sintesis data, training 3 model, evaluasi, SHAP	hasil_eksperimen.json, dataset_lms.csv
hci_analyzer.py	Perhitungan Balance, Density, dan kontras WCAG	hasil_hci.json, Gambar 6 dan 7
figures.py	Pembuatan seluruh gambar naskah	Gambar 1–5
EduLens.jsx	Prototipe dashboard (React); model berjalan di sisi klien	Aplikasi web interaktif

Koefisien model terpilih (regresi logistik, pada fitur terstandardisasi) adalah: tugas terkumpul −0,840; rerata nilai kuis −0,740; materi diakses −0,636; streak hari aktif −0,514; rerata keterlambatan +0,357; kehadiran sinkron −0,333; durasi belajar −0,282; video ditonton −0,271; kontribusi forum −0,178; frekuensi login −0,128; dengan intersep −0,554. Tanda negatif berarti kenaikan nilai fitur menurunkan risiko.