Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

Komparasi Decision Tree, Random Forest, dan XGBoost untuk Deteksi Dini Risiko Akademik, Kehadiran, dan Karakter Siswa

M. Hafidhatul Fathoni*, Junadhi, Susanti, Hadi Asnal

Fakultas Teknik dan Informatika, Prodi Teknik Informatika, Universitas Sains dan Teknologi Indonesia, Pekanbaru, Indonesia
Email: [1,*]m4914@admin.smk.belajar.id, [2]junadhi@usti.ac.id, [3]susanti@usti.ac.id, [4]hadiasnal@usti.ac.id,
Email Penulis Korespondensi: m4914@admin.smk.belajar.id
Submitted: **25/02/2026**; Accepted: **30/06/2026**; Published**: 30/06/2026**

Abstrak−Sekolah memiliki data akademik, kehadiran, dan karakter siswa yang tersedia secara rutin, tetapi pemanfaatannya masih banyak berhenti pada pencatatan administratif. Akibatnya, perubahan kondisi siswa belum selalu diterjemahkan menjadi sinyal risiko yang dapat dipantau secara dini, baik berupa penurunan akademik, ketidakhadiran berulang, maupun akumulasi pelanggaran karakter. Permasalahan penelitian ini adalah bagaimana mengevaluasi model deteksi dini yang mampu membaca tiga dimensi risiko tersebut secara bersamaan pada data longitudinal bulanan yang tidak seimbang dan berurutan waktu. Penelitian ini membandingkan Decision Tree, Random Forest, dan XGBoost untuk mendeteksi Risiko Akademik, Risiko Kehadiran, dan Risiko Karakter pada bulan berikutnya. Data yang digunakan merupakan data longitudinal bulanan siswa SMKS YUM Pesantren Teknologi Riau tahun 2025 dengan 206 siswa selama 12 bulan. Dataset mentah berjumlah 2.472 baris dan menjadi 2.266 baris valid setelah feature engineering dan time-shifting target. Evaluasi dilakukan menggunakan rolling-origin temporal evaluation, baseline mayoritas, baseline persistensi temporal, serta strategi tanpa balancing, class_weight, dan SMOTE parsial pada data training. Macro F1 digunakan sebagai metrik utama karena distribusi kelas tidak seimbang. Hasil menunjukkan bahwa Random Forest dengan class_weight menjadi model terbaik untuk Risiko Akademik dengan Macro F1 0,5005 dan recall Risiko Tinggi 0,6185. Decision Tree dengan class_weight menjadi model terbaik untuk Risiko Kehadiran dengan Macro F1 0,4586 dan recall Risiko Tinggi 0,5339. Random Forest dengan class_weight menjadi model terbaik untuk Risiko Karakter dengan Macro F1 0,4761 dan recall Risiko Tinggi 0,5698. Temuan ini menunjukkan bahwa model machine learning dapat memberi sinyal awal Risiko Tinggi, tetapi prediksi tetap perlu diverifikasi oleh guru atau wali kelas.

Kata Kunci: Deteksi Dini; Evaluasi Temporal; Machine Learning; Risiko Siswa; Time-Shifting Target

Abstract−Schools routinely collect academic, attendance, and character data, but their use remains limited to administrative recording. As a result, changes in student conditions are not always transformed into early risk signals, such as academic decline, repeated absenteeism, or accumulated character-related violations. The problem addressed in this study is how to evaluate an early detection model that can capture these three risk dimensions simultaneously using imbalanced and temporally ordered monthly longitudinal data. This study compares Decision Tree, Random Forest, and XGBoost for detecting Academic Risk, Attendance Risk, and Character Risk in the following month. The dataset consists of longitudinal student records from SMKS YUM Pesantren Teknologi Riau in 2025, covering 206 students across 12 months. The raw dataset contained 2,472 rows and was reduced to 2,266 valid rows after feature engineering and time-shifting target construction. Model evaluation used rolling-origin temporal evaluation, majority baseline, temporal persistence baseline, and imbalance handling strategies: no balancing, class_weight, and partial SMOTE on training data. Macro F1 was used as the primary metric because the class distribution was imbalanced. The best models were Random Forest with class_weight for Academic Risk, with Macro F1 of 0.5005 and High-Risk recall of 0.6185; Decision Tree with class_weight for Attendance Risk, with Macro F1 of 0.4586 and High-Risk recall of 0.5339; and Random Forest with class_weight for Character Risk, with Macro F1 of 0.4761 and High-Risk recall of 0.5698. These findings indicate that machine learning models can provide early High-Risk signals, but predictions should be verified by teachers or homeroom teachers.

Keywords: Early Detection; Machine Learning; Student Risk; Temporal Evaluation; Time-Shifting Target

1. PENDAHULUAN

Data akademik, kehadiran, dan karakter siswa tersedia secara rutin dalam administrasi sekolah, tetapi pemanfaatannya sering berhenti pada pencatatan nilai, rekap absensi, laporan pelanggaran, dan dokumentasi wali kelas. Kondisi tersebut membuat sekolah memiliki data, tetapi belum selalu memiliki mekanisme analitik untuk membaca perubahan risiko siswa sebelum masalah berkembang. Kebutuhan ini relevan dengan konteks pendidikan yang masih menghadapi tantangan capaian belajar. Laporan World Bank dan UNESCO menunjukkan bahwa *learning poverty* tetap menjadi masalah penting, sedangkan PISA 2022 menunjukkan capaian belajar Indonesia masih perlu diperkuat pada literasi, matematika, dan sains [1], [2]. UNESCO juga menekankan bahwa teknologi pendidikan perlu diarahkan pada kebutuhan pembelajaran yang nyata, bukan hanya adopsi perangkat digital [3]. Dengan demikian, masalah utama penelitian ini bukan sekadar ketersediaan data, melainkan bagaimana data sekolah dapat diolah menjadi sinyal risiko yang berguna untuk pemantauan siswa.

Pemanfaatan data pendidikan untuk prediksi performa siswa telah berkembang dalam bidang *educational data mining*. Batool et al. menunjukkan bahwa prediksi performa akademik merupakan salah satu fokus utama *educational data mining* karena data pendidikan dapat digunakan untuk mengenali pola keberhasilan maupun kegagalan belajar [4]. Duong et al. mengembangkan *academic performance warning system* berbasis data untuk mendukung identifikasi mahasiswa yang berpotensi mengalami masalah akademik [5]. Skittou et al. juga menunjukkan bahwa *early warning system* dapat mendukung proses perencanaan pendidikan melalui identifikasi siswa berisiko [6]. Zaied et al. mengembangkan model penasihat (advisory model) berbasis data mining untuk mendukung identifikasi capaian akademik siswa [7]. Temuan tersebut memperlihatkan bahwa sistem deteksi dini dapat membantu institusi pendidikan

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

mengubah data historis menjadi informasi pemantauan. Namun, sistem seperti itu perlu dirancang dengan hati-hati karena prediksi risiko harus mendukung keputusan manusia, bukan menggantikan verifikasi guru, wali kelas, atau pihak bimbingan konseling.

Prediksi dini siswa berisiko tidak cukup hanya mengandalkan satu indikator. Risiko Akademik dapat muncul melalui penurunan nilai, keterlambatan tugas, atau perubahan keterlibatan belajar. Risiko Kehadiran dapat terlihat dari penurunan persentase hadir, peningkatan alpa, atau pola ketidakhadiran berulang. Risiko Karakter dapat terbaca dari perubahan sikap, catatan pembinaan, dan akumulasi poin pelanggaran. Jang et al. menekankan pentingnya prediksi dini yang praktis dan dapat dijelaskan, sehingga hasil model lebih mudah dipahami oleh pengguna pendidikan [8]. Kord et al. menunjukkan bahwa integrasi data multi-modal dapat memperkuat prediksi performa siswa [9]. Gupta et al. juga menunjukkan bahwa pola sekuensial dapat digunakan untuk prediksi dini siswa berisiko [10]. Dasar tersebut mendukung penggunaan data longitudinal bulanan dan target risiko multidimensi dalam penelitian ini.

Algoritma berbasis pohon relevan untuk data pendidikan berbentuk tabular karena dapat memodelkan hubungan nonlinier dan tetap relatif mudah ditafsirkan dibandingkan beberapa model kompleks lain. Kumar et al. menggunakan Random Forest dan XGBoost untuk mengantisipasi performa akademik siswa [11]. Aurelio dan Setiawan membandingkan Decision Tree, Random Forest, dan XGBoost dalam analisis *attrition* mahasiswa [12]. Niyogisubizo et al. menunjukkan penggunaan *ensemble learning* untuk prediksi *dropout*, sedangkan Carballo Mendívil et al. menerapkan XGBoost dalam *early warning system* berbasis data awal siswa [13], [14]. Almaqbali et al. juga membandingkan *tree-based classifiers* yang interpretatif untuk memprediksi *disengagement* pembelajar [15]. Penelitian-penelitian tersebut memberi dasar metodologis untuk membandingkan Decision Tree, Random Forest, dan XGBoost pada konteks deteksi dini risiko siswa. Akan tetapi, pemilihan algoritma tidak boleh diasumsikan sama untuk setiap target karena Risiko Akademik, Risiko Kehadiran, dan Risiko Karakter memiliki distribusi dan pola fitur yang berbeda.

Evaluasi model pada data risiko pendidikan perlu memperhatikan ketidakseimbangan kelas. Pada kasus deteksi dini, kelas Risiko Rendah biasanya lebih dominan daripada Risiko Sedang dan Risiko Tinggi. Andri et al. menggunakan pendekatan SMOTE-ENN pada data dropout siswa, sedangkan Alhammouri et al. membahas optimasi klasifikasi multi-kelas melalui ensemble learning dan teknik balancing [16], [17]. Kondisi ini membuat *accuracy* tidak memadai sebagai metrik utama. Model yang selalu memprediksi kelas mayoritas dapat menghasilkan *accuracy* yang tampak baik, tetapi gagal menemukan siswa pada Risiko Tinggi. Karena itu, evaluasi dalam penelitian ini menempatkan *Macro F1* sebagai metrik utama, disertai *recall* Risiko Tinggi, F1-score Risiko Tinggi, F1-score Risiko Sedang, *support* per kelas, dan *generalization gap*.

Berdasarkan kajian tersebut, terdapat empat *gap* yang menjadi dasar penelitian ini. Pertama, sebagian studi prediksi pendidikan masih berfokus pada performa akademik, *dropout*, atau *attrition*, sehingga belum secara eksplisit membandingkan tiga target risiko sekolah, yaitu Risiko Akademik, Risiko Kehadiran, dan Risiko Karakter. Kedua, data longitudinal bulanan pada tingkat sekolah menengah belum banyak dikaitkan dengan skema *time-shifting target* yang memetakan fitur bulan berjalan untuk memprediksi risiko bulan berikutnya. Ketiga, evaluasi model pada data berurutan waktu masih sering kurang menekankan validitas temporal. Cerqueira et al. menunjukkan bahwa evaluasi temporal penting karena metode estimasi performa dapat memengaruhi kesimpulan model pada data berbasis waktu [18]. Keempat, dalam konteks Indonesia, studi prediksi *dropout* telah berkembang, tetapi belum secara khusus menekankan baseline persistensi temporal sebagai pembanding pada deteksi dini risiko bulanan siswa [19]. *Gap* ini menunjukkan perlunya evaluasi komparatif yang tidak hanya membandingkan algoritma, tetapi juga menguji nilai tambah model terhadap baseline sederhana.

Penelitian ini bertujuan membandingkan Decision Tree, Random Forest, dan XGBoost untuk mendeteksi Risiko Akademik, Risiko Kehadiran, dan Risiko Karakter siswa pada bulan berikutnya. Penelitian menerapkan *rule-based labeling* sebagai *proxy label*, *time-shifting target*, *rolling-origin temporal evaluation*, serta baseline mayoritas dan baseline persistensi temporal sebagai pembanding. Kontribusi penelitian terletak pada rancangan evaluasi komparatif berbasis data longitudinal sekolah, bukan pada pengusulan algoritma baru. Penelitian juga menggunakan beberapa strategi penanganan ketidakseimbangan kelas, yaitu tanpa *balancing*, *class_weight*, dan SMOTE parsial pada data *training*. Model terbaik dipilih berdasarkan *selection score* yang mempertimbangkan *Macro F1*, F1-score Risiko Tinggi, *recall* Risiko Tinggi, F1-score Risiko Sedang, *accuracy*, dan *generalization gap*.

2. METODOLOGI PENELITIAN

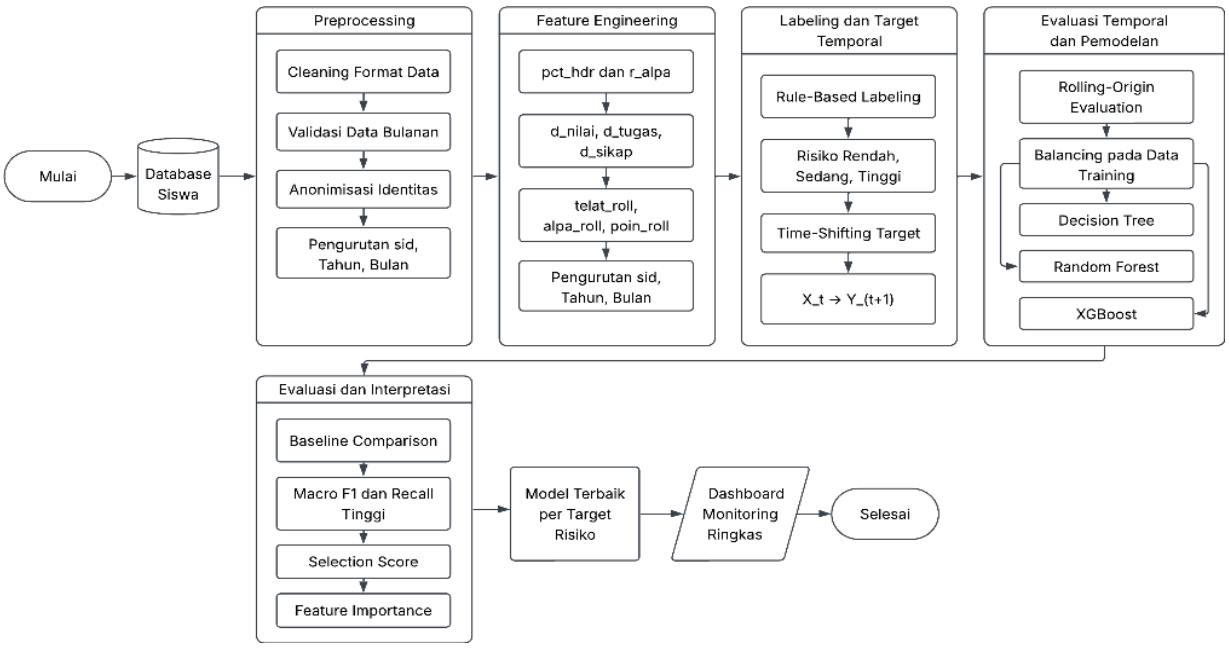
2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis *machine learning* untuk membandingkan kemampuan Decision Tree, Random Forest, dan XGBoost dalam mendeteksi risiko siswa pada bulan berikutnya. Alur penelitian dimulai dari penyusunan data siswa bulanan yang mencakup indikator akademik, kehadiran, dan karakter. Data kemudian diproses melalui tahap *preprocessing* untuk menyeragamkan format kolom, membersihkan karakter tersembunyi, memastikan tipe data numerik, dan mengurutkan data berdasarkan siswa serta bulan pengamatan. Pengurutan kronologis diperlukan karena model menggunakan data masa lalu untuk memprediksi periode setelahnya.

Tahap berikutnya adalah anonimisasi identitas siswa, *feature engineering* longitudinal, pembentukan label risiko berbasis aturan, penerapan *time-shifting target*, evaluasi temporal, penanganan ketidakseimbangan kelas pada

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

data *training*, pelatihan model, perbandingan dengan baseline, pemilihan model terbaik, dan interpretasi model. Evaluasi dirancang berbasis urutan waktu karena data siswa memiliki struktur longitudinal. Pendekatan tersebut lebih sesuai dibanding pembagian acak ketika tujuan model adalah memprediksi risiko pada periode mendatang [18]. Seluruh tahapan ini dirancang agar evaluasi model sesuai dengan konteks penggunaan nyata di sekolah, yaitu memprediksi risiko periode mendatang dari data historis yang tersedia. Alur tahapan penelitian divisualisasikan pada Gambar 1, yang menunjukkan bahwa proses penelitian dimulai dari penyusunan data siswa bulanan, *preprocessing, feature engineering, rule-based labeling, time-shifting target,* evaluasi temporal, pemodelan, hingga interpretasi model.



Gambar 1. Alur Penelitian Deteksi Dini Risiko Siswa

2.2 Dataset dan Variabel Penelitian

Dataset yang digunakan merupakan data riil administrasi akademik dan kesiswaan SMKS YUM Pesantren Teknologi Riau tahun 2025 yang digunakan atas izin institusi untuk kepentingan penelitian. Unit analisis penelitian adalah satu siswa pada satu bulan. Setiap baris data merepresentasikan kondisi akademik, kehadiran, dan karakter seorang siswa dalam satu periode pengamatan. Dataset mencakup 206 siswa selama 12 bulan, sehingga data mentah berjumlah 2.472 baris. Setelah proses *feature engineering*, dataset memiliki 2.472 baris dan 30 kolom. Setelah penerapan *time-shifting target*, data valid untuk pemodelan berjumlah 2.266 baris.

Data yang dianalisis telah melalui proses anonimisasi. Identitas langsung siswa, seperti nama siswa, tidak digunakan sebagai fitur prediktif dan tidak ditampilkan dalam hasil penelitian. Kolom sid hanya digunakan sebagai penanda anonim untuk menjaga konsistensi pengurutan data longitudinal antarbulan. Kolom kelas juga tidak digunakan sebagai fitur prediktif untuk mengurangi potensi bias administratif. Seluruh hasil analisis disajikan dalam bentuk agregat sehingga tidak mengarah pada identifikasi individu siswa tertentu. Karakteristik dataset yang digunakan dalam penelitian ini diringkas pada Tabel 1.

Tabel 1. Karakteristik Dataset dan Target Risiko

Aspek	Nilai	Keterangan
Sumber data	SMKS YUM Pesantren Teknologi Riau	Data administrasi akademik dan kesiswaan
Periode	2025	Data bulanan selama 12 bulan
Jumlah siswa	206	Identitas dianonimkan
Unit analisis	Siswa-bulan	Satu siswa pada satu bulan
Data mentah	2.472 baris	206 siswa × 12 bulan
Setelah *feature engineering*	2.472 baris × 30 kolom	Data referensi bulanan
Data valid setelah *time-shifting target*	2.266 baris	Data siap pemodelan
Target risiko	3 target	Akademik, kehadiran, karakter
Kelas risiko	3 kelas	Risiko Rendah, Risiko Sedang, Risiko Tinggi

Berdasarkan Tabel 1, data penelitian memiliki struktur longitudinal siswa-bulan sehingga sesuai untuk evaluasi temporal.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

2.3 Feature Engineering Longitudinal

Feature engineering longitudinal dilakukan untuk membentuk fitur yang mampu merepresentasikan perubahan kondisi siswa antarbulan. Tahap ini diperlukan karena risiko siswa tidak selalu terlihat dari nilai absolut pada satu periode. Risiko dapat muncul melalui tren penurunan nilai, penurunan kehadiran, peningkatan alpa, keterlambatan tugas berulang, akumulasi poin pelanggaran, atau perubahan sikap. Penggunaan fitur berbasis urutan waktu juga sejalan dengan pendekatan prediksi dini yang memanfaatkan pola sekuensial pada data pendidikan [10].

Fitur longitudinal utama yang dibentuk meliputi pct_hdr, r_alpa, d_nilai, d_tugas, d_sikap, telat_roll, poin_roll, alpa_roll, dan hdr_roll. pct_hdr menunjukkan persentase kehadiran siswa dalam satu bulan. r_alpa menunjukkan rasio alpa terhadap total hari tercatat. d_nilai, d_tugas, dan d_sikap menunjukkan perubahan nilai, tugas, dan sikap terhadap bulan sebelumnya. Fitur telat_roll, poin_roll, alpa_roll, dan hdr_roll merepresentasikan agregasi riwayat jangka pendek pada tiga bulan sebelumnya. telat_roll menghitung riwayat keterlambatan tugas, poin_roll menghitung riwayat poin pelanggaran, alpa_roll menghitung riwayat alpa, sedangkan hdr_roll menghitung riwayat persentase kehadiran. Fitur ini digunakan agar model membaca pola berulang, bukan hanya kondisi bulan berjalan. Dalam data mining, transformasi fitur seperti ini digunakan untuk memperkuat representasi data sebelum proses pemodelan [20]. Persentase kehadiran, rasio alpa, dan perubahan nilai terhadap bulan sebelumnya dihitung menggunakan Rumus:

$$pct\_hdr_t = \frac{hadir\_t}{hari\_t} \times 100 \quad (1)$$

$$r\_alpa\_t = \frac{alpa\_t}{hari\_t} \times 100 \quad (2)$$

$$d\_nilai_t = nilai\_t - nilai_{t-1} \quad (3)$$

Pada Rumus (1) dan Rumus (2), (hari_t) merupakan total hari tercatat pada bulan ke-(t), yaitu penjumlahan hadir, sakit, izin, dan alpa. Pada Rumus (3), (nilai_t) menunjukkan nilai bulan berjalan, sedangkan (nilai_{t-1}) menunjukkan nilai bulan sebelumnya. Makna fitur turunan longitudinal yang digunakan dalam penelitian ini diringkas pada Tabel 2.

Tabel 2. Definisi Fitur Turunan Longitudinal

Fitur	Makna
hari	Total hari tercatat, yaitu hadir + sakit + izin + alpa.
pct_hdr	Persentase kehadiran siswa pada bulan berjalan.
r_alpa	Rasio alpa terhadap total hari tercatat (alpa / hari) × 100
d_nilai	Perubahan nilai terhadap bulan sebelumnya.
d_tugas	Perubahan jumlah tugas atau setoran terhadap bulan sebelumnya.
d_sikap	Perubahan skor sikap terhadap bulan sebelumnya.
telat_roll	Riwayat keterlambatan tugas dalam tiga bulan terakhir.
poin_roll	Riwayat poin pelanggaran dalam tiga bulan terakhir.
alpa_roll	Riwayat alpa dalam tiga bulan terakhir.
hdr_roll	Riwayat persentase kehadiran dalam tiga bulan terakhir.

Berdasarkan Tabel 2, fitur turunan longitudinal digunakan untuk menangkap kondisi siswa dari dua sisi, yaitu kondisi bulan berjalan dan perubahan historis antarbulan. Fitur seperti pct_hdr dan r_alpa menggambarkan kondisi kehadiran pada bulan berjalan, sedangkan d_nilai, d_tugas, dan d_sikap menunjukkan perubahan kondisi siswa dibanding bulan sebelumnya. Sementara itu, telat_roll, poin_roll, alpa_roll, dan hdr_roll digunakan untuk membaca pola riwayat jangka pendek agar model tidak hanya bergantung pada data satu bulan. Fitur-fitur tersebut kemudian dipilih secara spesifik untuk masing-masing target risiko agar model tetap relevan secara domain dan lebih mudah diinterpretasikan [20]. Rincian pemetaan fitur prediktif yang digunakan pada masing-masing target risiko diringkas pada Tabel 3.

Tabel 3. Feature Set per Target Risiko

Target Risiko	Fitur Prediktif	Rasional Domain
Risiko Akademik	nilai, tugas, telat, d_nilai, d_tugas, telat_roll, pct_hdr, r_alpa	Membaca capaian belajar, keterlibatan tugas, keterlambatan, dan dukungan kehadiran terhadap proses akademik
Risiko Kehadiran	hadir, sakit, izin, alpa, pct_hdr, r_alpa, alpa_roll, hdr_roll	Membaca partisipasi fisik, ketidakhadiran, dan pola kehadiran historis
Risiko Karakter	sikap, d_sikap, poin, cat_bk, poin_roll, telat, telat_roll	Membaca sikap, perubahan perilaku, catatan pembinaan, riwayat pelanggaran, dan kedisiplinan tugas

Berdasarkan Tabel 3, setiap target risiko menggunakan kombinasi fitur yang dipilih berdasarkan relevansi domain, bukan seleksi statistik otomatis. Yang perlu dicatat adalah fitur telat dan telat_roll muncul pada dua target yang berbeda sebagai indikator keterlibatan akademik pada Risiko Akademik, dan sebagai sinyal kedisiplinan perilaku

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

pada Risiko Karakter. Fitur yang sama dapat memiliki makna prediktif yang berbeda bergantung pada target yang dimodelkan dan apakah model benar-benar memanfaatkan perbedaan peran ini akan terlihat pada hasil analisis *feature importance* di Subbab 3.6.

2.4 Rule-Based Labeling dan Time-Shifting Target

Dataset sekolah tidak memiliki label risiko formal yang langsung dapat digunakan sebagai target pemodelan. Oleh karena itu, penelitian ini membentuk label risiko menggunakan *rule-based labeling*. Label yang dihasilkan diposisikan sebagai *proxy label*, yaitu label operasional berbasis aturan, bukan diagnosis final terhadap kondisi siswa. Pendekatan ini digunakan agar status risiko dapat diturunkan secara eksplisit dari indikator akademik, kehadiran, dan karakter yang tersedia dalam data sekolah.

Tiga target yang digunakan adalah Risiko Akademik, Risiko Kehadiran, dan Risiko Karakter. Setiap target diklasifikasikan ke dalam tiga kelas, yaitu Risiko Rendah, Risiko Sedang, dan Risiko Tinggi. Risiko Rendah menunjukkan kondisi yang relatif stabil. Risiko Sedang menunjukkan gejala awal yang perlu dipantau. Risiko Tinggi menunjukkan kondisi yang memerlukan perhatian lebih cepat dari guru, wali kelas, atau pihak sekolah. Skema tiga kelas dipertahankan karena Risiko Sedang merepresentasikan fase transisi yang penting dalam sistem deteksi dini.

Label risiko dalam penelitian ini disusun agar dapat ditelusuri secara eksplisit dari indikator yang tersedia pada data sekolah. Risiko Akademik dibentuk dari kombinasi nilai, perubahan nilai, keterlambatan tugas, riwayat keterlambatan tugas, dan jumlah tugas. Risiko Kehadiran dibentuk dari persentase hadir, jumlah alpa, rasio alpa, sakit, dan izin. Risiko Karakter dibentuk dari skor sikap, perubahan sikap, poin pelanggaran, riwayat poin pelanggaran, dan catatan BK atau wali kelas. Setiap indikator terlebih dahulu dikonversi menjadi skor risiko, kemudian digabungkan menjadi skor komposit untuk menentukan kelas Risiko Rendah, Risiko Sedang, atau Risiko Tinggi. Dengan demikian, label dalam penelitian ini bersifat transparan dan dapat ditelusuri, meskipun tetap diposisikan sebagai proxy label, bukan diagnosis final terhadap siswa.

Skor komposit Risiko Akademik dihitung dari kombinasi nilai sebesar 0,35, d_nilai sebesar 0,25, telat sebesar 0,20, telat_roll sebesar 0,10, dan tugas sebesar 0,10. Skor komposit Risiko Kehadiran dihitung dari pct_hdr sebesar 0,45, alpa sebesar 0,30, r_alpa sebesar 0,15, sakit sebesar 0,05, dan izin sebesar 0,05. Skor komposit Risiko Karakter dihitung dari sikap sebesar 0,30, d_sikap sebesar 0,20, poin sebesar 0,25, poin_roll sebesar 0,15, dan cat_bk sebesar 0,10. Bobot tersebut digunakan untuk membedakan kontribusi indikator utama dan indikator pendukung pada setiap dimensi risiko.

Setiap indikator dikonversi terlebih dahulu ke dalam skor risiko ordinal 0-2, dengan skor 0 menunjukkan kondisi rendah risiko, skor 1 menunjukkan kondisi sedang, dan skor 2 menunjukkan kondisi tinggi. Ambang konversi indikator disusun berdasarkan karakteristik data sekolah, kemudian dikonsultasikan dengan pihak sekolah, yaitu guru, wali kelas, dan BK, untuk memastikan kesesuaiannya dengan kebutuhan operasional pemantauan siswa. Proses ini menjadi salah satu bentuk pertimbangan domain dalam penyusunan ambang label, meskipun belum menggantikan validasi eksternal yang lebih formal, seperti pembandingan dengan outcome riil siswa atau penilaian oleh panel ahli independen. Tabel 4 menyajikan aturan label untuk ketiga target risiko, mencakup dasar indikator yang digunakan, bentuk perhitungan skor komposit berbobot, serta ambang batas kelas yang membedakan kondisi Risiko Rendah, Risiko Sedang, dan Risiko Tinggi pada masing-masing dimensi.

Tabel 4. Aturan Label

Target	Dasar Indikator	Bentuk Skor	Ambang Kelas
Risiko Akademik	nilai, d_nilai, telat, telat_roll, tugas	skor komposit berbobot	Rendah: <0,70; Sedang: 0,70-1,25; Tinggi: >1,25
Risiko Kehadiran	pct_hdr, alpa, r_alpa, sakit, izin	skor komposit berbobot	Rendah: <0,65; Sedang: 0,65-1,20; Tinggi: >1,20
Risiko Karakter	sikap, d_sikap, poin, poin_roll, cat_bk	skor komposit berbobot	Rendah: <0,60; Sedang: 0,60-1,15; Tinggi: >1,15

Tabel 4 menunjukkan bahwa setiap target risiko menggunakan indikator dan ambang yang berbeda sesuai karakteristik domainnya. Ambang Risiko Tinggi ditetapkan di atas 1,25 untuk Risiko Akademik, di atas 1,20 untuk Risiko Kehadiran, dan di atas 1,15 untuk Risiko Karakter. Ambang Risiko Karakter yang paling rendah mencerminkan sensitivitas lebih tinggi terhadap akumulasi poin pelanggaran dan perubahan sikap, yaitu kondisi yang dinilai perlu direspons lebih cepat secara operasional. Karena ambang ini bersifat tetap dan berbasis aturan, distribusi kelas yang dihasilkan serta tingkat kesulitan model dalam mempelajari setiap kelas bergantung langsung pada sebaran nilai aktual data sekolah, bukan pada skema yang dapat dikalibrasi ulang secara otomatis. Distribusi label aktual yang terbentuk dari skema ini disajikan pada Tabel 7. Setelah label risiko bulan berjalan terbentuk, penelitian menerapkan *time-shifting target*. Skema ini digunakan agar model tidak hanya mengklasifikasikan kondisi pada bulan yang sama, tetapi memprediksi risiko pada bulan berikutnya. Prinsip tersebut dirumuskan pada Rumus (4).

$$X_t \rightarrow Y_{t+1} \tag{4}$$

Dengan skema ini, fitur Januari digunakan untuk memprediksi risiko Februari, fitur Februari digunakan untuk memprediksi risiko Maret, dan seterusnya. Pendekatan ini mendukung fungsi sistem sebagai deteksi dini karena

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

memberi jarak waktu antara sinyal data dan periode risiko [5], [6], [10]. Namun, *time-shifting target* tidak diposisikan sebagai solusi penuh terhadap seluruh risiko sirkularitas. Data siswa bulanan tetap dapat memiliki autokorelasi antarperiode. Karena itu, baseline persistensi temporal tetap digunakan pada tahap hasil sebagai pembanding penting.

2.5 Rolling-Origin Temporal Evaluation

Evaluasi model dilakukan menggunakan *rolling-origin temporal evaluation*. Skema ini dipilih karena data penelitian memiliki urutan waktu. Model dilatih menggunakan data masa lalu dan diuji pada periode setelahnya. Dengan demikian, evaluasi lebih mendekati kondisi penggunaan nyata, yaitu ketika data historis digunakan untuk memprediksi risiko siswa pada periode mendatang. Skema ini tidak sama dengan *cross-validation* acak karena tidak melakukan pengacakan urutan data.

Evaluasi dilakukan dalam enam fold. Pada setiap fold, data *training* bertambah secara kronologis, sedangkan data *testing* berada pada periode setelah data *training*. Setiap fold menguji 206 baris data, sehingga total data *testing pooled* adalah 1.236 baris. Evaluasi pooled digunakan untuk membaca performa keseluruhan model pada seluruh periode pengujian, sedangkan hasil per fold digunakan untuk melihat kestabilan performa temporal [18]. Rincian pembagian periode *training*, *testing*, target prediksi, serta jumlah data pada setiap fold disajikan pada Tabel 5.

Tabel 5. Skema Rolling-Origin Temporal Evaluation

Fold	Data Training	Data Testing	Target Prediksi	Train Rows	Test Rows
1	Bulan 1-5	Bulan 6	Bulan 7	1.030	206
2	Bulan 1-6	Bulan 7	Bulan 8	1.236	206
3	Bulan 1-7	Bulan 8	Bulan 9	1.442	206
4	Bulan 1-8	Bulan 9	Bulan 10	1.648	206
5	Bulan 1-9	Bulan 10	Bulan 11	1.854	206
6	Bulan 1-10	Bulan 11	Bulan 12	2.060	206

Berdasarkan Tabel 5, skema *rolling-origin* dilakukan dengan menambah cakupan data *training* secara bertahap dari bulan 1–5 hingga bulan 1–10, sedangkan data *testing* selalu berada setelah periode *training*. Setiap *fold* menguji 206 baris, sehingga evaluasi *pooled* dari enam *fold* menghasilkan 1.236 baris data *testing*. Pola ini memastikan model dilatih menggunakan data historis dan diuji pada periode berikutnya, bukan pada data yang tercampur secara acak. Namun, ukuran *training set* meningkat dari 1.030 baris pada *fold* 1 menjadi 2.060 baris pada *fold* 6, sehingga stabilitas performa antarfold dapat dipengaruhi oleh jumlah data historis, terutama pada kelas minoritas.

2.6 Algoritma dan Strategi Balancing

Penelitian ini membandingkan tiga algoritma *machine learning*, yaitu Decision Tree, Random Forest, dan XGBoost. Decision Tree digunakan sebagai model pohon tunggal yang mudah ditafsirkan. Random Forest digunakan sebagai metode *ensemble bagging* yang membangun banyak pohon keputusan untuk meningkatkan stabilitas prediksi. XGBoost digunakan sebagai model berbasis *gradient boosting* yang relevan untuk data tabular pendidikan, sebagaimana dijelaskan dalam literatur standar *machine learning* [20], [21]. Penggunaan XGBoost dalam konteks prediksi pendidikan juga didukung oleh studi terapan terbaru yang menerapkan Random Forest dan XGBoost pada prediksi performa akademik siswa [22].

Setiap algoritma diuji dengan tiga strategi penanganan ketidakseimbangan kelas, yaitu none, *class_weight*, dan *smote_partial*. Strategi none mempertahankan distribusi kelas asli pada data *training*. Strategi *class_weight* memberi bobot lebih besar pada kelas minoritas. Pada XGBoost, prinsip pembobotan diterapkan melalui *sample_weight*. Strategi *smote_partial* digunakan untuk menambah representasi kelas minoritas secara terbatas pada data *training*. SMOTE hanya diterapkan pada data *training* di setiap fold, sedangkan data *testing* tidak diberi SMOTE agar evaluasi tetap mencerminkan distribusi data asli. *StandardScaler* hanya digunakan pada skenario *smote_partial* sebelum proses SMOTE karena SMOTE bekerja berdasarkan jarak antar-sampel. Standardisasi tersebut bukan kebutuhan model pohon, sebab Decision Tree, Random Forest, dan XGBoost tidak memerlukan standardisasi fitur untuk proses pemisahan berbasis pohon. Konfigurasi utama setiap algoritma dan strategi penanganan ketidakseimbangan kelas yang digunakan dalam eksperimen diringkas pada Tabel 6.

Tabel 6. Konfigurasi Model dan Strategi Balancing

Model	Parameter Utama	Strategi Balancing
Decision Tree	criterion=gini, max_depth=4, min_samples_split=40, min_samples_leaf=20, random_state=42	none, class_weight, smote_partial
Random Forest	n_estimators=300, max_depth=7, min_samples_split=30, min_samples_leaf=10, max_features=sqrt, bootstrap=True, random_state=42, n_jobs=-1	none, class_weight, smote_partial
XGBoost	objective=multi:softprob, num_class=3, n_estimators=180, max_depth=3, learning_rate=0,05, subsample=0,85, colsample_bytree=0,85, min_child_weight=5, reg_alpha=0,50, reg_lambda=2,00, eval_metric=mlogloss	none, sample_weight, smote_partial

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

Konfigurasi parameter pada Tabel 6 ditetapkan melalui proses *tuning* pada ruang parameter (*parameter space*) yang dibatasi untuk masing-masing algoritma. Pembatasan ini mempertimbangkan karakteristik skema *rolling-origin*, di mana ukuran data *training* bertambah secara bertahap antarfold (Tabel 5), sehingga model pada fold-fold awal dilatih dengan jumlah data yang lebih sedikit dibanding total dataset dan berisiko lebih besar mengalami *overfitting* apabila kompleksitas model seperti max_depth, min_samples_leaf, min_samples_split, dan jumlah estimator tidak dibatasi. Oleh karena itu, ruang parameter dijaga tetap konservatif agar performa model stabil di seluruh fold temporal, bukan hanya unggul pada satu periode pengujian tertentu. Setiap kombinasi parameter dievaluasi menggunakan skema *rolling-origin temporal evaluation* yang sama dengan evaluasi akhir (Tabel 5), dan kombinasi terbaik dipilih berdasarkan *selection score* yang mempertimbangkan *Macro F1*, *recall* Risiko Tinggi, F1-score Risiko Tinggi, F1-score Risiko Sedang, *accuracy*, dan *generalization gap*, sebagaimana dijelaskan pada Subbab 2.7.

2.7 Metrik Evaluasi dan Selection score

Evaluasi model menggunakan *accuracy*, *precision*, *recall*, F1-score, *Macro F1*, *support*, dan *generalization gap*. *Accuracy* tetap dilaporkan sebagai ukuran proporsi prediksi benar secara keseluruhan, tetapi tidak dijadikan metrik utama karena distribusi kelas risiko tidak seimbang. Pada data seperti ini, model dapat memperoleh *accuracy* cukup tinggi dengan hanya mengikuti kelas mayoritas, tetapi gagal mendeteksi Risiko Sedang dan Risiko Tinggi [17].

Macro F1 digunakan sebagai metrik utama karena menghitung rata-rata F1-score setiap kelas tanpa memberi bobot lebih besar pada kelas mayoritas. Rumus *Macro F1* ditunjukkan pada Rumus (5).

$$MacroF1 = \frac{F1_{Rendah} + F1_{Sedang} + F1_{Tinggi}}{3} \tag{5}$$

Selain *Macro F1*, *recall* Risiko Tinggi digunakan untuk menilai kemampuan model menemukan siswa yang benar-benar berada pada Risiko Tinggi. F1-score Risiko Tinggi digunakan untuk menilai keseimbangan antara ketepatan dan sensitivitas pada kelas paling serius. F1-score Risiko Sedang digunakan karena kelas ini menjadi fase transisi sebelum Risiko Tinggi. *Support* dilaporkan agar interpretasi performa per kelas tidak dilepaskan dari jumlah data aktual. *Generalization gap* digunakan untuk membaca perbedaan *Macro F1* antara data *training* dan data *testing*. Model terbaik dipilih menggunakan *selection score*. Skor ini merupakan skor komposit internal penelitian, bukan metrik baku universal. Rumusnya ditunjukkan pada Rumus berikut.

$$G = |MacroF1_{train} - MacroF1_{test}| \tag{6}$$

$$S = 0{,}35M + 0{,}20F_T + 0{,}15R_T + 0{,}15F_S + 0{,}10A - 0{,}20G \tag{7}$$

Pada Rumus (7), (S) adalah *selection score*, (M) adalah *Macro F1 testing*, (F_T) adalah F1-score kelas Risiko Tinggi, (R_T) adalah *recall* kelas Risiko Tinggi, (F_S) adalah F1-score kelas Risiko Sedang, (A) adalah *accuracy testing*, dan (G) adalah *generalization gap*. Bobot tersebut dipakai agar pemilihan model tidak hanya mengejar *accuracy*, tetapi juga mempertimbangkan performa pada kelas Risiko Tinggi, kelas transisi Risiko Sedang, dan stabilitas generalisasi [18], [21].

3. HASIL DAN PEMBAHASAN

3.1 Distribusi Label dan Konsekuensi Evaluasi

Hasil pembentukan dataset melalui *feature engineering* longitudinal dan *time-shifting target* menghasilkan 2.266 baris data valid. Data ini digunakan untuk mengevaluasi tiga target risiko, yaitu Risiko Akademik, Risiko Kehadiran, dan Risiko Karakter. Setiap target memiliki tiga kelas, yaitu Risiko Rendah, Risiko Sedang, dan Risiko Tinggi. Setelah proses *feature engineering longitudinal* dan *time-shifting target*, distribusi label pada data valid dihitung untuk melihat komposisi setiap kelas pada setiap target. Ringkasan distribusi tersebut disajikan pada Tabel 7.

Tabel 7. Distribusi Label Time-Shift Valid

Target Risiko	Risiko Rendah	Risiko Sedang	Risiko Tinggi
Risiko Akademik	1.252 / 55,25%	672 / 29,66%	342 / 15,09%
Risiko Kehadiran	1.426 / 62,93%	561 / 24,76%	279 / 12,31%
Risiko Karakter	1.482 / 65,40%	536 / 23,65%	248 / 10,94%

Tabel 7 menunjukkan bahwa Risiko Rendah menjadi kelas mayoritas pada seluruh target, dengan proporsi 55,25% pada Risiko Akademik, 62,93% pada Risiko Kehadiran, dan 65,40% pada Risiko Karakter. Risiko Tinggi menjadi kelas paling kecil, yaitu 15,09% pada Risiko Akademik, 12,31% pada Risiko Kehadiran, dan 10,94% pada Risiko Karakter. Secara gabungan, Risiko Sedang dan Risiko Tinggi mencakup 44,75% data Risiko Akademik, 37,07% Risiko Kehadiran, dan 34,60% Risiko Karakter cukup untuk dievaluasi secara *pooled*, tetapi tetap berada jauh di bawah proporsi Risiko Rendah. Ketidakseimbangan ini memiliki konsekuensi langsung terhadap *precision* model pada kelas minoritas: strategi *class_weight* yang meningkatkan *recall* Risiko Tinggi akan secara bersamaan menghasilkan lebih banyak *false positive*, sehingga *precision* kelas ini cenderung terbatas meskipun *recall*-nya meningkat. Pola ini akan terlihat pada *classification report* yang disajikan di Tabel 11.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

Distribusi ini memastikan bahwa Risiko Rendah mendominasi seluruh data evaluasi, dengan proporsi 55–65% bergantung pada target. Oleh karena itu, pembacaan hasil pada bagian berikutnya menggunakan *Macro F1* sebagai metrik utama, disertai F1-score Risiko Sedang, *recall* Risiko Tinggi, F1-score Risiko Tinggi, *support* per kelas, dan *generalization gap*, sebagaimana dirancang pada Subbab 2.7.

3.2 Komparasi Algoritma pada Rolling-Origin Evaluation

Evaluasi dilakukan menggunakan *rolling-origin temporal evaluation* sebanyak enam fold. Setiap model dilatih pada data masa lalu dan diuji pada periode setelahnya. Ringkasan pada Tabel 8 menampilkan kombinasi terbaik untuk setiap algoritma pada masing-masing target, disertai baseline mayoritas dan baseline persistensi temporal sebagai pembanding awal. Pembahasan detail terhadap baseline disajikan pada bagian 3.4, sehingga bagian ini hanya menggunakan baseline sebagai konteks komparasi umum.

Tabel 8. Ringkasan Komparasi Algoritma dan Strategi Terbaik

Target	Metode	Strat.	Acc.	M-F1	R-T	F1-T	F1-S	G
Akademik	DT	CW	0,4968	0,4485	0,7034	0,4057	0,2839	0,0405
Akademik	RF	CW	0,5477	0,5005	0,6185	0,4447	0,3535	0,1351
Akademik	XGB	CW	0,5332	0,4878	0,5918	0,4131	0,3610	0,1559
Akademik	Mayoritas	Mayoritas	0,5380	0,2329	0,0000	0,0000	0,0000	0,0128
Akademik	Persistensi	Persistensi	0,5494	0,4805	0,3556	0,3541	0,3839	0,0332
Kehadiran	DT	CW	0,5178	0,4586	0,5339	0,3726	0,3439	0,0506
Kehadiran	RF	CW	0,5340	0,4705	0,5169	0,3888	0,3441	0,1151
Kehadiran	XGB	CW	0,5194	0,4649	0,5197	0,3733	0,3663	0,1240
Kehadiran	Mayoritas	Mayoritas	0,6311	0,2579	0,0000	0,0000	0,0000	0,0050
Kehadiran	Persistensi	Persistensi	0,6011	0,4833	0,3599	0,3577	0,3402	0,0388
Karakter	DT	SMOTE-P	0,5510	0,4312	0,6251	0,3486	0,2236	0,0403
Karakter	RF	CW	0,5388	0,4761	0,5698	0,3886	0,3523	0,1586
Karakter	XGB	SMOTE-P	0,6173	0,4830	0,4554	0,3639	0,3044	0,1106
Karakter	Mayoritas	Mayoritas	0,6254	0,2561	0,0000	0,0000	0,0000	0,0153
Karakter	Persistensi	Persistensi	0,5850	0,4483	0,2701	0,2687	0,3314	0,0202

Keterangan singkatan pada Tabel 8 adalah sebagai berikut:

a. DT, RF, dan XGB masing-masing merujuk pada Decision Tree, Random Forest, dan XGBoost.
b. CW merujuk pada *class_weight*, sedangkan SMOTE-P merujuk pada *smote_partial*.
c. Acc., M-F1, R-T, F1-T, F1-S, dan G masing-masing merujuk pada *accuracy*, *Macro F1*, *recall* Risiko Tinggi, F1-score Risiko Tinggi, F1-score Risiko Sedang, dan *generalization gap*.
d. Mayoritas dan Persistensi masing-masing merujuk pada baseline mayoritas dan baseline persistensi temporal.

Hasil komparasi menunjukkan bahwa tidak ada satu algoritma yang unggul pada seluruh target dan seluruh metrik. Pada Risiko Akademik, Random Forest dengan *class_weight* menghasilkan *Macro F1* tertinggi di antara model *machine learning*, yaitu 0,5005. Pada Risiko Kehadiran, Random Forest menghasilkan *Macro F1* lebih tinggi daripada Decision Tree, tetapi Decision Tree dengan *class_weight* dipilih sebagai model terbaik karena memperoleh *selection score* tertinggi. Pemilihan ini dipengaruhi oleh generalization gap yang lebih kecil dan keseimbangan metrik kelas Risiko Tinggi serta Risiko Sedang. Pada Risiko Karakter, XGBoost dengan smote_partial memiliki *Macro F1* sedikit lebih tinggi daripada Random Forest, tetapi Random Forest dengan *class_weight* menjadi model terbaik berdasarkan kriteria pemilihan komposit.

Temuan tersebut menunjukkan bahwa performa model harus dibaca melalui kombinasi metrik, bukan dari satu angka. Baseline mayoritas memperoleh *accuracy* yang relatif tinggi pada Risiko Kehadiran dan Risiko Karakter, tetapi gagal mendeteksi Risiko Sedang dan Risiko Tinggi karena *recall* dan F1-score Risiko Tinggi bernilai 0,0000. Pola ini memperlihatkan bahwa *Macro F1* Baseline Mayoritas yang sangat rendah (0,2329-0,2579) bukan anomali, melainkan konsekuensi langsung dari kegagalan mendeteksi kelas minoritas yang justru menjadi target utama sistem deteksi dini.

3.3 Model Machine learning Terbaik per Target Risiko

Berdasarkan *selection score*, model *machine learning* terbaik berbeda antar target risiko. Tahap pemilihan model terbaik dilakukan untuk memastikan bahwa model yang dipilih tidak hanya unggul pada metrik agregat, tetapi juga relevan dengan tujuan deteksi dini. Oleh karena itu, *selection score* digunakan dengan mempertimbangkan *Macro F1*, kemampuan mengenali Risiko Tinggi, performa pada kelas transisi Risiko Sedang, serta stabilitas generalisasi model. Model dengan nilai *selection score* terbaik pada setiap target risiko diringkas pada Tabel 9.

Tabel 9. Model Machine learning Terbaik per Target Risiko

Target Risiko	Model Terbaik	Strategi	Accuracy	*Macro F1*	*Recall* Tinggi	F1 Tinggi	F1 Sedang	Gap
Risiko Akademik	Random Forest	class_weight	0,5477	0,5005	0,6185	0,4447	0,3535	0,1351
Risiko Kehadiran	Decision Tree	class_weight	0,5178	0,4586	0,5339	0,3726	0,3439	0,0506

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

Target Risiko	Model Terbaik	Strategi	Accuracy	*Macro F1*	*Recall* Tinggi	F1 Tinggi	F1 Sedang	Gap
Risiko Karakter	Random Forest	class_weight	0,5388	0,4761	0,5698	0,3886	0,3523	0,1586

Pemilihan model terbaik pada Tabel 9 didasarkan pada *selection score*, bukan hanya *Macro F1* atau *accuracy*. Skor ini mempertimbangkan *Macro F1*, F1-score Risiko Tinggi, *recall* Risiko Tinggi, F1-score Risiko Sedang, *accuracy*, dan *generalization gap*. Karena itu, model dengan *Macro F1* tertinggi tidak selalu dipilih apabila memiliki *generalization gap* lebih besar atau keseimbangan performa kelas yang lebih lemah. Tabel 9 menunjukkan bahwa Random Forest dengan *class_weight* menjadi model terbaik pada Risiko Akademik dan Risiko Karakter, sedangkan Decision Tree dengan *class_weight* lebih sesuai untuk Risiko Kehadiran. Pada ketiga target, F1-score Risiko Tinggi (0,4447; 0,3726; 0,3886) secara konsisten lebih rendah dari *recall* Risiko Tinggi (0,6185; 0,5339; 0,5698). Karena F1-score merupakan rata-rata harmonik *precision* dan *recall*, nilai F1 yang lebih rendah dari *recall* secara matematis mengindikasikan bahwa *precision* Risiko Tinggi berada lebih rendah lagi artinya model mendeteksi banyak kasus Risiko Tinggi, tetapi sebagian di antaranya merupakan *false positive*. Detail *precision* per kelas akan disajikan pada Tabel 11, dan perlu dibaca bersama distribusi kelas pada Tabel 7 untuk memahami asal-usul struktural dari pola ini.

Pada Risiko Kehadiran, model terbaik adalah Decision Tree dengan *class_weight*. Hasil ini penting karena menunjukkan bahwa model yang lebih sederhana dapat menjadi pilihan terbaik pada target tertentu. Risiko Kehadiran memiliki pola indikator yang relatif langsung, seperti kehadiran, alpa, rasio alpa, dan riwayat kehadiran. Karena itu, struktur pohon tunggal yang diberi pembobotan kelas dapat menghasilkan keseimbangan performa yang kompetitif. Gap model ini juga paling kecil dibanding model terbaik pada target lain, yaitu 0,0506. Namun, gap kecil tidak berarti model tersebut paling unggul secara keseluruhan, karena *Macro F1* Risiko Kehadiran tetap berada pada 0,4586.

Strategi *class_weight* menjadi strategi paling konsisten karena muncul pada seluruh model terbaik. Strategi ini membantu model memberi perhatian lebih besar pada kelas minoritas tanpa mengubah distribusi data *testing*. Nilai *accuracy* model terbaik berada pada kisaran 0,51-0,55. Rentang ini harus dibaca sebagai hasil evaluasi temporal yang realistis pada klasifikasi tiga kelas yang tidak seimbang. Nilai tersebut tidak menunjukkan kegagalan model secara langsung, tetapi menunjukkan bahwa tugas deteksi dini bulanan lebih sulit daripada evaluasi acak atau evaluasi yang hanya menekankan kelas mayoritas. Oleh karena itu, hasil utama pada bagian ini adalah variasi model terbaik antar target dan pentingnya membaca performa berdasarkan *Macro F1*, Risiko Sedang, Risiko Tinggi, serta *generalization gap*.

3.4 Perbandingan Model ML dengan Baseline

Perbandingan dengan baseline diperlukan untuk menilai apakah model *machine learning* benar-benar memberikan nilai tambah terhadap strategi prediksi sederhana. Baseline mayoritas merepresentasikan skenario ketika model hanya mengikuti kelas dominan pada data *training*, sedangkan baseline persistensi temporal merepresentasikan skenario ketika risiko bulan berikutnya diasumsikan sama dengan risiko bulan berjalan. Dalam penelitian ini, baseline persistensi digunakan sebagai pembanding utama karena lebih relevan dengan data longitudinal siswa. Selisih performa model *machine learning* terbaik terhadap baseline persistensi temporal pada setiap target risiko disajikan pada Tabel 10.

Tabel 10. Baseline vs Model ML Terbaik

Target Risiko	Δ Accuracy	Δ *Macro F1*	Δ F1 Sedang	Δ *Recall* Tinggi	Δ F1 Tinggi	Interpretasi
Risiko Akademik	-0,0016	+0,0201	-0,0305	+0,2630	+0,0905	ML meningkatkan *Macro F1* dan deteksi Risiko Tinggi, tetapi menurunkan F1 Risiko Sedang
Risiko Kehadiran	-0,0833	-0,0247	+0,0037	+0,1740	+0,0149	*Baseline* persistensi lebih kuat pada *accuracy* dan *Macro F1*, tetapi ML lebih sensitif terhadap Risiko Tinggi
Risiko Karakter	-0,0461	+0,0278	+0,0209	+0,2997	+0,1199	ML meningkatkan *Macro F1*, F1 Sedang, dan deteksi Risiko Tinggi, tetapi *accuracy* lebih rendah

Jika dibaca lintas target, Tabel 10 menunjukkan adanya pola *trade-off* antara *accuracy* dan sensitivitas terhadap Risiko Tinggi. Model *machine learning* terbaik memiliki *accuracy* lebih rendah daripada *baseline* persistensi pada seluruh target, yaitu -0,0016 pada Risiko Akademik, -0,0833 pada Risiko Kehadiran, dan -0,0461 pada Risiko Karakter. Namun, pada saat yang sama, model meningkatkan *recall* Risiko Tinggi pada seluruh target, yaitu +0,2630 pada Risiko Akademik, +0,1740 pada Risiko Kehadiran, dan +0,2997 pada Risiko Karakter. Pola ini menunjukkan bahwa model tidak terutama memberi nilai tambah pada prediksi keseluruhan, tetapi pada kemampuan menangkap lebih banyak kasus Risiko Tinggi. Dalam konteks deteksi dini, peningkatan sensitivitas ini lebih relevan daripada kenaikan *accuracy* kecil karena tujuan sistem adalah membantu sekolah menemukan siswa yang perlu diverifikasi lebih cepat.

Pada Risiko Akademik, *Random Forest* dengan *class_weight* menunjukkan nilai tambah yang relatif seimbang dibanding *baseline* persistensi. Meskipun *accuracy* turun sangat kecil sebesar -0,0016, model meningkatkan *Macro*

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

F1 sebesar +0,0201, *recall* Risiko Tinggi sebesar +0,2630, dan *F1-score* Risiko Tinggi sebesar +0,0905. Peningkatan ini menunjukkan bahwa model akademik mampu membaca sinyal Risiko Tinggi lebih baik daripada sekadar mempertahankan status risiko bulan berjalan. Namun, *F1-score* Risiko Sedang turun sebesar -0,0305, sehingga kemampuan model dalam membaca fase transisi akademik masih perlu ditafsirkan secara hati-hati.

Pada Risiko Kehadiran, *baseline* persistensi masih lebih kuat pada metrik agregat karena model *Decision Tree* dengan *class_weight* memiliki selisih *accuracy* -0,0833 dan *Macro F1* -0,0247. Hasil ini menunjukkan bahwa pola kehadiran siswa memiliki kesinambungan temporal yang kuat, sehingga asumsi risiko bulan berikutnya sama dengan bulan berjalan masih menjadi pembanding yang sulit dikalahkan. Meskipun demikian, model tetap memberi nilai tambah pada tujuan deteksi dini karena *recall* Risiko Tinggi meningkat sebesar +0,1740 dan *F1-score* Risiko Tinggi meningkat sebesar +0,0149. Dengan demikian, pada target kehadiran, model *machine learning* lebih tepat diposisikan sebagai pelengkap untuk menandai potensi Risiko Tinggi, bukan sebagai model yang unggul pada seluruh metrik.

Pada Risiko Karakter, *Random Forest* dengan *class_weight* menunjukkan pola yang paling mendukung kebutuhan deteksi dini. Walaupun *accuracy* turun sebesar -0,0461, model meningkatkan *Macro F1* sebesar +0,0278, *F1-score* Risiko Sedang sebesar +0,0209, *recall* Risiko Tinggi sebesar +0,2997, dan *F1-score* Risiko Tinggi sebesar +0,1199. Peningkatan pada Risiko Sedang dan Risiko Tinggi menunjukkan bahwa fitur karakter dan riwayat perilaku memberikan sinyal tambahan yang membantu model membaca perubahan risiko secara lebih luas daripada *baseline* persistensi. Secara keseluruhan, hasil pada Tabel 10 memperlihatkan bahwa kontribusi utama model *machine learning* bukan kemenangan pada seluruh metrik, melainkan peningkatan kemampuan mendeteksi Risiko Tinggi dengan konsekuensi penurunan *accuracy*. Temuan ini menjadi dasar untuk membaca hasil per kelas pada subbagian berikutnya.

3.5 Analisis Per Kelas dan Kelas Transisi Risiko Sedang

Analisis per kelas diperlukan untuk melihat bagian model yang sudah kuat dan bagian yang masih lemah pada setiap target risiko. Evaluasi agregat seperti *accuracy* dan *Macro F1* belum cukup untuk menjelaskan perilaku model, terutama pada data yang tidak seimbang. Oleh karena itu, pembahasan pada subbagian ini difokuskan pada *precision*, *recall*, *F1-score*, dan *support* untuk kelas Risiko Rendah, Risiko Sedang, dan Risiko Tinggi pada masing-masing target. Rincian *classification report pooled* dari model terbaik pada setiap target risiko disajikan pada Tabel 11.

Tabel 11. Classification Report Pooled Model Terbaik

Target Risiko	Kelas	Precision	*Recall*	F1-score	Support
Risiko Akademik	Rendah	0,7724	0,6481	0,7048	665
Risiko Akademik	Sedang	0,3846	0,3279	0,3540	366
Risiko Akademik	Tinggi	0,3443	0,6146	0,4413	205
Risiko Kehadiran	Rendah	0,8177	0,5577	0,6631	780
Risiko Kehadiran	Sedang	0,2968	0,4067	0,3432	300
Risiko Kehadiran	Tinggi	0,2833	0,5321	0,3697	156
Risiko Karakter	Rendah	0,8303	0,5886	0,6889	773
Risiko Karakter	Sedang	0,3157	0,4019	0,3536	311
Risiko Karakter	Tinggi	0,2945	0,5658	0,3874	152

Berdasarkan Tabel 11, komposisi *support* menunjukkan bahwa Risiko Rendah menjadi kelas paling dominan pada seluruh target, sedangkan Risiko Tinggi memiliki jumlah data paling kecil. Kondisi ini membuat nilai rata-rata tidak cukup untuk menggambarkan kualitas model secara utuh. Oleh karena itu, pembacaan hasil perlu diarahkan pada perbedaan karakter setiap kelas: Risiko Rendah sebagai kondisi stabil, Risiko Sedang sebagai fase transisi, dan Risiko Tinggi sebagai kelas prioritas dalam sistem deteksi dini.

Kelas Risiko Rendah memiliki *F1-score* tertinggi pada seluruh target, yaitu 0,7048 pada Risiko Akademik, 0,6631 pada Risiko Kehadiran, dan 0,6889 pada Risiko Karakter. Pola ini menunjukkan bahwa model relatif lebih mudah mengenali siswa yang berada pada kondisi stabil. Namun, performa yang baik pada Risiko Rendah tidak dapat dijadikan satu-satunya dasar keberhasilan sistem, karena tujuan utama deteksi dini bukan hanya mengenali siswa yang stabil, tetapi juga membaca siswa yang mulai menunjukkan gejala risiko.

Kelas Risiko Sedang menjadi bagian paling menantang dalam pemodelan. *F1-score* Risiko Sedang hanya mencapai 0,3540 pada Risiko Akademik, 0,3432 pada Risiko Kehadiran, dan 0,3536 pada Risiko Karakter. Nilai ini menunjukkan bahwa model masih kesulitan membedakan fase transisi dari kondisi Risiko Rendah dan Risiko Tinggi. Secara domain, kondisi ini dapat dipahami karena Risiko Sedang tidak selalu memiliki pola ekstrem. Siswa pada kelas ini dapat menunjukkan sebagian gejala risiko, tetapi belum cukup kuat untuk masuk ke Risiko Tinggi. Meskipun sulit diprediksi, kelas Risiko Sedang tetap penting dipertahankan karena merepresentasikan ruang pemantauan awal sebelum kondisi siswa memburuk.

Pada kelas Risiko Tinggi, model menunjukkan kecenderungan lebih sensitif dibandingkan tepat. Risiko Akademik memperoleh *recall* 0,6146 dengan *precision* 0,3443, Risiko Kehadiran memperoleh *recall* 0,5321 dengan *precision* 0,2833, dan Risiko Karakter memperoleh *recall* 0,5658 dengan *precision* 0,2945. Pola ini menunjukkan bahwa model mampu menangkap sebagian besar siswa berisiko tinggi, tetapi masih menghasilkan prediksi positif

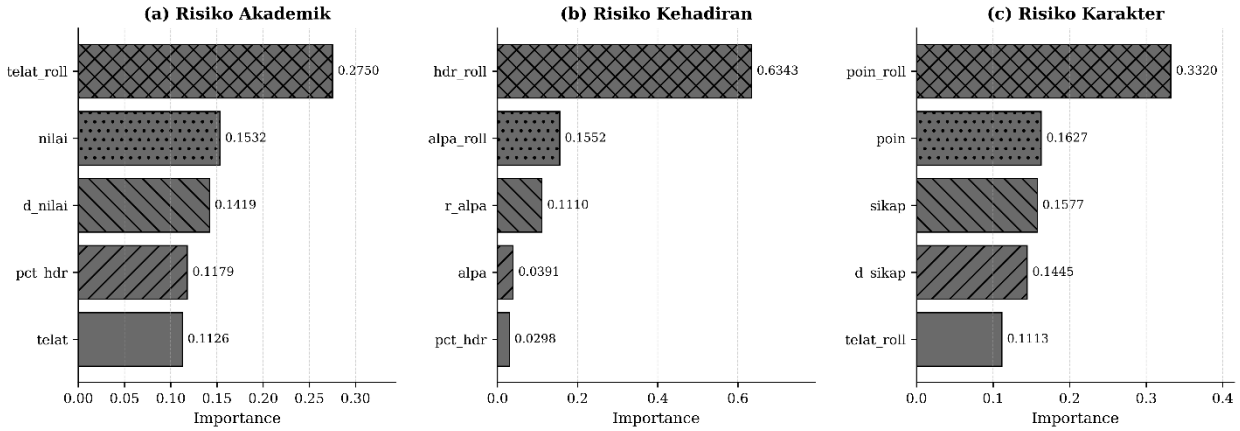
Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

yang tidak seluruhnya tepat. Dalam konteks *early warning system*, karakter seperti ini masih dapat diterima selama *output* model digunakan sebagai sinyal awal untuk verifikasi, bukan sebagai keputusan otomatis.

Dengan demikian, hasil per kelas memperjelas pola kekuatan dan kelemahan model secara konkret: model relatif andal mengenali Risiko Rendah, cukup sensitif terhadap Risiko Tinggi meskipun *precision*-nya masih terbatas, dan masih lemah pada Risiko Sedang. Rendahnya performa pada Risiko Sedang menunjukkan perlunya pengayaan fitur, perluasan periode data, atau validasi tambahan pada penelitian berikutnya agar fase transisi risiko dapat dibaca lebih stabil.

3.6 Analisis Feature Importance

Analisis *feature importance* digunakan untuk membaca fitur yang paling berkontribusi dalam proses pemisahan kelas pada model terbaik. Nilai ini hanya menunjukkan kontribusi prediktif di dalam model, bukan hubungan sebab-akibat. Oleh karena itu, fitur dengan nilai *importance* tinggi tidak boleh ditafsirkan sebagai penyebab langsung munculnya risiko siswa. Interpretasi dibatasi pada peran fitur dalam membantu model membedakan Risiko Rendah, Risiko Sedang, dan Risiko Tinggi. Lima fitur teratas pada setiap target risiko disajikan pada Gambar 2.



Gambar 2. Lima Fitur Teratas Model Terbaik per Target Risiko

Pada Risiko Akademik, fitur paling dominan adalah telat_roll dengan nilai *importance* 0,2750, diikuti nilai, d_nilai, pct_hdr, dan telat. Pola ini menunjukkan bahwa model akademik banyak memanfaatkan riwayat keterlambatan tugas, capaian nilai bulan berjalan, dan perubahan nilai. Masuknya pct_hdr juga menunjukkan bahwa kehadiran tetap berkontribusi dalam pemisahan kelas Risiko Akademik. Namun, kontribusi tersebut tetap dibaca sebagai hubungan prediktif, bukan bukti bahwa keterlambatan tugas atau kehadiran menjadi penyebab langsung risiko akademik.

Pada Risiko Kehadiran, fitur paling dominan adalah hdr_roll dengan nilai *importance* 0,6343. Nilai ini jauh lebih tinggi dibanding fitur lain, seperti alpa_roll, r_alpa, alpa, dan pct_hdr. Temuan ini menunjukkan bahwa riwayat persentase kehadiran menjadi sinyal paling kuat dalam model kehadiran. Dominasi fitur riwayat juga konsisten dengan karakter data kehadiran yang cenderung memiliki pola berulang antarbulan. Dengan demikian, model Risiko Kehadiran lebih banyak membaca pola temporal daripada kondisi satu bulan saja.

Pada Risiko Karakter, fitur paling dominan adalah poin_roll dengan nilai *importance* 0,3320, diikuti poin, sikap, d_sikap, dan telat_roll. Pola ini menunjukkan bahwa model karakter memanfaatkan kombinasi antara riwayat poin pelanggaran, poin bulan berjalan, skor sikap, perubahan sikap, dan riwayat keterlambatan tugas. Masuknya telat_roll menunjukkan bahwa kedisiplinan tugas ikut membantu pemisahan kelas Risiko Karakter. Interpretasi ini tetap bersifat prediktif karena *feature importance* tidak menjelaskan hubungan kausal.

3.7 Implementasi Ringkas Dashboard Monitoring

Sebagai implementasi ringkas, keluaran model dirancang untuk ditampilkan dalam *dashboard monitoring* yang menyajikan tiga target risiko secara bersamaan, yaitu Risiko Akademik, Risiko Kehadiran, dan Risiko Karakter. *Dashboard* bukan kontribusi utama penelitian, tetapi berfungsi sebagai media penyajian hasil prediksi agar lebih mudah dibaca oleh pengguna sekolah. Pada konteks operasional, kelas Risiko Rendah, Risiko Sedang, dan Risiko Tinggi dapat dipetakan menjadi istilah Aman, Waspada, dan Kritis. Pemetaan ini hanya digunakan pada *dashboard* untuk mempermudah komunikasi risiko. Di dalam analisis ilmiah, istilah yang digunakan tetap Risiko Rendah, Risiko Sedang, dan Risiko Tinggi.

Dashboard dapat menampilkan ringkasan status risiko siswa, target risiko yang perlu diperhatikan, dan informasi fitur yang berkontribusi terhadap prediksi. Informasi ini membantu guru, wali kelas, atau BK menentukan prioritas pemeriksaan awal. *Dashboard* dirancang agar informasi risiko dapat dibaca oleh pengguna tanpa latar belakang teknis. Setiap tampilan status siswa disertai indikator fitur pendukung sehingga guru atau wali kelas dapat

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

menilai konteks prediksi secara langsung, tidak hanya melihat label risiko akhir. Dengan pendekatan ini, *dashboard* mendukung alur kerja pemantauan yang sudah berjalan di sekolah, bukan menggantikannya.

3.8 Pembahasan

Hasil penelitian ini menunjukkan bahwa algoritma berbasis pohon tetap relevan untuk data pendidikan berbentuk tabular, tetapi pemilihan model terbaik tidak dapat dilepaskan dari karakter target risiko dan tujuan evaluasi. Studi Kumar et al. dan Zoralioğlu et al. menunjukkan bahwa Random Forest dan XGBoost dapat digunakan untuk memprediksi performa akademik siswa [11], [22]. Penelitian ini sejalan dengan arah tersebut, tetapi memperluas fokusnya dari prediksi performa akademik menjadi deteksi tiga dimensi risiko siswa, yaitu Risiko Akademik, Risiko Kehadiran, dan Risiko Karakter. Dibandingkan penelitian *dropout* dan *attrition* yang umumnya menekankan satu keluaran utama [12], [13], [14], pembagian tiga target dalam penelitian ini lebih sesuai dengan kebutuhan sekolah karena perubahan kondisi siswa dapat muncul melalui jalur akademik, kehadiran, atau karakter secara berbeda. Temuan bahwa model terbaik tidak sama pada setiap target juga menunjukkan bahwa deteksi dini risiko siswa tidak cukup diselesaikan dengan satu asumsi algoritmik yang berlaku umum.

Dari sisi evaluasi, temuan baseline mayoritas pada penelitian ini yang memperoleh *accuracy* 0,5380–0,6311 tetapi F1-score Risiko Tinggi 0,0000 pada seluruh target memberikan ilustrasi empiris atas kekhawatiran yang diangkat Andri et al. dan Alhammouri et al. tentang ketidakcukupan metrik tunggal pada data pendidikan tidak seimbang [16], [17]. Penggunaan *Macro F1*, *recall* Risiko Tinggi, F1-score Risiko Sedang, dan *generalization gap* secara bersamaan memungkinkan evaluasi yang tidak hanya membaca akurasi keseluruhan, tetapi juga kualitas model pada kelas yang paling relevan untuk pemantauan dini. Dengan demikian, kontribusi model tidak dibaca dari kemampuan mengikuti kelas mayoritas, tetapi dari kemampuannya memberi sinyal pada kelas yang lebih penting untuk pemantauan. Penggunaan *selection score* juga membedakan penelitian ini dari evaluasi yang hanya memilih model berdasarkan satu metrik, karena model terbaik dipilih dengan mempertimbangkan keseimbangan performa kelas dan stabilitas generalisasi.

Dari sisi temporal, penggunaan *rolling-origin temporal evaluation* dan *baseline* persistensi temporal membuat hasil penelitian ini lebih hati-hati dibanding evaluasi acak biasa. Cerqueira et al. menegaskan bahwa evaluasi pada data berbasis waktu perlu mempertahankan urutan kronologis agar estimasi performa tidak terlalu optimistis [18]. Dalam penelitian ini, *baseline* persistensi berfungsi sebagai pembanding yang ketat karena menguji apakah model *machine learning* benar-benar memberi nilai tambah dibanding asumsi sederhana berbasis risiko bulan berjalan. Dengan demikian, kontribusi utama penelitian ini bukan pada pengusulan algoritma baru, melainkan pada rancangan evaluasi komparatif yang lebih sesuai untuk konteks deteksi dini risiko siswa: menggunakan data longitudinal bulanan, *time-shifting target*, evaluasi temporal, pembanding *baseline*, metrik per kelas, serta interpretasi *feature importance* sebagai dasar verifikasi oleh guru atau wali kelas.

4. KESIMPULAN

Penelitian ini menghasilkan evaluasi komparatif terhadap Decision Tree, Random Forest, dan XGBoost untuk deteksi dini Risiko Akademik, Risiko Kehadiran, dan Risiko Karakter siswa menggunakan data longitudinal bulanan sekolah. Melalui penerapan *time-shifting target* dan *rolling-origin temporal evaluation*, proses pemodelan diarahkan untuk meniru kebutuhan nyata sekolah, yaitu menggunakan data historis untuk membaca potensi risiko pada periode berikutnya. Hasil penelitian menunjukkan bahwa model terbaik berbeda pada setiap target risiko, dengan Random Forest berbasis *class_weight* lebih sesuai untuk Risiko Akademik dan Risiko Karakter, sedangkan Decision Tree berbasis *class_weight* lebih sesuai untuk Risiko Kehadiran. Temuan ini menegaskan bahwa deteksi dini risiko siswa tidak dapat diselesaikan dengan satu algoritma yang dianggap unggul untuk seluruh dimensi, karena setiap target memiliki karakteristik fitur, distribusi kelas, dan pola temporal yang berbeda. Perbandingan dengan *baseline* juga menunjukkan bahwa model *machine learning* tidak selalu unggul pada *accuracy*, tetapi memberi nilai tambah pada kemampuan mengenali Risiko Tinggi yang lebih relevan untuk kebutuhan pemantauan dini. Dengan demikian, keluaran model lebih tepat diposisikan sebagai sinyal awal untuk membantu guru, wali kelas, atau BK menentukan prioritas verifikasi, bukan sebagai keputusan otomatis terhadap siswa. Keterbatasan penelitian ini terletak pada penggunaan data dari satu sekolah dan satu tahun pengamatan, pembentukan label berbasis *proxy rule-based labeling*, serta *precision* Risiko Tinggi yang masih terbatas. Penelitian lanjutan perlu menguji data lintas tahun dan lintas sekolah, memperbaiki kalibrasi probabilitas, serta mengevaluasi penerimaan pengguna agar sistem lebih siap digunakan secara operasional.

REFERENCES

[1] World Bank, UNESCO, “Indonesia - Learning Poverty Brief 2024,” World Bank, Washington, DC, 2024. [Online]. Available: https://documents.worldbank.org/curated/en/099082924151529593
[2] OECD, *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. Paris: OECD Publishing, 2023. doi: 10.1787/53f23881-en.
[3] UNESCO, *Global Education Monitoring Report 2023: Technology in Education: A Tool on Whose Terms?*, First. Paris: UNESCO, 2023. doi: 10.54676/UZQV8501.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 497−509
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9452

[4] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Educ. Inf. Technol.*, vol. 28, no. 1, pp. 905–971, 2023, doi: 10.1007/s10639-022-11152-y.
[5] H. T.-H. Duong, L. T.-M. Tran, H. Q. To, and K. Van Nguyen, "Academic performance warning system based on data driven for higher education," *Neural Comput. Appl.*, vol. 35, no. 8, pp. 5819–5837, 2023, doi: 10.1007/s00521-022-07997-6.
[6] M. Skittou, S. Raghay, and A. Zellou, "Development of an Early Warning System to Support Educational Planning Process by Identifying At-Risk Students," *IEEE Access*, vol. 12, p. 2260, 2024, doi: 10.1109/ACCESS.2023.3348091.
[7] A. N. H. Zaied, E. M. Hamza, and R. W. Ismael, "An Advisory Student Achievement Model Based on Data Mining Techniques," in *2022 International Conference on Computing and Information*, 2022. doi: 10.1109/ICCI54321.2022.9756121.
[8] Y. Jang, S. Choi, H. Jung, and H. Kim, "Practical early prediction of students' performance using *machine learning* and eXplainable AI," *Educ. Inf. Technol.*, vol. 27, no. 9, pp. 12855–12889, 2022, doi: 10.1007/s10639-022-11120-6.
[9] A. M. Kord, A. Aboelfetouh, and S. M. Shohieb, "Academic Course Planning Recommendation and Students' Performance Prediction Multi-Modal Based on Educational Data Mining Techniques," *J. Comput. High. Educ.*, 2025, doi: 10.1007/s12528-024-09426-0.
[10] A. Gupta, D. Garg, and P. Kumar, "Mining Sequential Learning Trajectories with Hidden Markov Models for Early Prediction of At-Risk Students in E-Learning Environments," *IEEE Trans. Learn. Technol.*, vol. 15, no. 6, pp. 783–799, 2022, doi: 10.1109/TLT.2022.3202277.
[11] M. Kumar, N. Singh, J. Wadhwa, P. Singh, G. Kumar, and A. Qtaishat, "Utilizing Random Forest and XGBoost DataMining Algorithms for Anticipating Students' Academic Performance," *Int. J. Mod. Educ. Comput. Sci.*, vol. 16, no. 2, 2024, doi: 10.5815/ijmecs.2024.02.03.
[12] C. Aurelio and J. Setiawan, "In-Depth Examination of Student Attrition in Higher Education Through Decision Tree, Random Forest, and XGBoost Algorithms," 2024. doi: 10.1109/EECSI63442.2024.10776324.
[13] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, and P. C. Nshimyumukiza, "Predicting Student's Dropout in University Classes Using Two-Layer Ensemble *Machine learning* Approach: A Novel Stacked Generalization," *Comput. Educ. Artif. Intell.*, vol. 3, p. 100066, 2022, doi: 10.1016/j.caeai.2022.100066.
[14] B. Carballo Mendívil, A. Arellano González, N. J. Ríos-Vázquez, and M. del P. Lizardi-Duarte, "Predicting Student Dropout from Day One: XGBoost-Based Early Warning System Using Pre-Enrollment Data," *Appl. Sci.*, vol. 15, no. 16, p. 9202, 2025, doi: 10.3390/app15169202.
[15] S. Almaqbali, M.-C. Leow, B. Shannaq, O. Farouk, A. H. Marhoubi, and L.-Y. Ong, "Predicting Learner Disengagement in E-Learning Platforms Using Interpretable *Machine learning*: A Comparative Study of Tree-Based Classifiers," 2025. doi: 10.1109/ISCI65687.2025.11167693.
[16] Andri, R. Yunis, and Tanti, "Optimizing Random Forest Classification Using Chi-Square and SMOTE-ENN on Student Drop-Out Data," 2023. doi: 10.1109/ICIC60109.2023.10382055.
[17] M. Alhammouri, Z. A. A. Hammouri, I. T. Almalkawi, and A. Lafee, "Optimizing Multi-Class Classification in Educational Data with Ensemble Learning and Data Balancing Techniques," 2024. doi: 10.1109/IDSTA62194.2024.10746987.
[18] V. Cerqueira, L. Torgo, and I. Mozetič, "Evaluating Time Series Forecasting Models: An Empirical Study on Performance Estimation Methods," *Mach. Learn.*, vol. 109, pp. 1997–2028, 2020, doi: 10.1007/s10994-020-05910-7.
[19] T. A. Marzuqi, E. Kristiani, and M. Marcel, "Prediksi Mahasiswa Drop-Out di Universitas XYZ," *J. Teknol. Inf. dan Ilmu Komput.*, 2024, doi: 10.25126/jtiik.2024118689.
[20] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to Data Mining*, 2nd ed. Boston, MA: Pearson, 2019.
[21] A. Géron, *Hands-On Machine learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. Sebastopol, CA: O'Reilly Media, 2022.
[22] Y. Zoralioğlu, M. Gul, F. Azizoğlu, G. Azizoğlu, and A. N. Toprak, "Predicting Academic Performance of Students Using *Machine learning* Techniques," in *2023 Innovations in Intelligent Systems and Applications Conference*, 2023. doi: 10.1109/ASYU58738.2023.10296648.