

Analisis Determinan Karakter Siswa Menggunakan Explainable Machine Learning (SHAP) dan Klasterisasi Profil Sekolah Studi Kasus Rapor Pendidikan Provinsi Bali

Md. Wira Putra Dananjaya^{1*}, Ngakan Nyoman Kutha Krisnawijaya¹, Gede Humaswara Prathama¹,
I Gusti Ngurah Darma Paramartha¹, Adie Wahyudi Oktavia Gama¹

¹Universitas Pendidikan Nasional, Denpasar, Indonesia

putradananjaya@undiknas.ac.id*

| Received: 26/11/2025 |

Revised: 04/12/2025 |

Accepted: 16/12/2025 |

Copyright©2025 by authors. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstrak

Penguatan karakter siswa merupakan indikator kinerja utama dalam kurikulum Merdeka Belajar, namun identifikasi faktor determinan lingkungan sekolah yang paling berpengaruh terhadap capaian karakter seringkali masih bersifat asuntif. Penelitian ini bertujuan untuk mendekonstruksi pola hubungan antara iklim lingkungan sekolah dengan kualitas karakter siswa di Provinsi Bali secara kuantitatif. Menggunakan dataset Rapor Pendidikan Indonesia yang dirilis oleh Kementerian Pendidikan Dasar dan Menengah (Kemendikdasmen) periode 2023-2025 dengan total 727 entri data, penelitian ini menerapkan metodologi Educational Data Mining dengan algoritma Random Forest yang diperkuat dengan Synthetic Minority Over-sampling Technique (SMOTE) untuk menangani ketimpangan data. Kebaruan (*novelty*) penelitian ini terletak pada penggunaan *SHapley Additive exPlanations* (SHAP) untuk transparansi model dan K-Means Clustering untuk pemetaan zonasi. Hasil eksperimen menunjukkan model mampu memprediksi capaian karakter dengan akurasi 77,03%. Analisis SHAP mengungkap temuan menarik bahwa Iklim Kebinekaan (skor pengaruh 0,45) dan Iklim Kesenjangan Gender (0,22) adalah prediktor terkuat, jauh melampaui pengaruh Iklim Keamanan (0,13). Temuan ini membantah asumsi umum bahwa keamanan fisik adalah faktor tunggal terpenting. Lebih lanjut, analisis klasterisasi mengidentifikasi tiga tipologi sekolah di Bali, termasuk satu kluster "Rawan" yang memiliki skor kritis pada aspek kesetaraan gender dan kebinekaan meskipun memiliki skor keamanan yang memadai. Penelitian ini merekomendasikan pergeseran fokus kebijakan pendidikan di Bali dari pendekatan keamanan fisik menuju penguatan program toleransi dan kesetaraan gender yang terbukti memiliki dampak statistik lebih signifikan.

Kata kunci: Educational Data Mining, Random Forest, SHAP, Karakter Siswa, Rapor Pendidikan

Abstract

Strengthening student character is a key performance indicator in the Merdeka Belajar curriculum, but the identification of the school environment's most influential determinants of character achievement is often assumed. This study aims to quantitatively deconstruct the relationship between school climate and student character quality in Bali Province. Using the Indonesian Education Report dataset released by the Ministry of Primary and Secondary Education (Kemendikdasmen) for the 2023-2025 period with a total of 727 data entries, this study applies the Educational Data Mining methodology with the Random Forest algorithm enhanced by the Synthetic Minority Over-sampling Technique (SMOTE) to address data inequality. The novelty of this study lies in the use of SHapley Additive exPlanations (SHAP) for model transparency and K-Means Clustering for zoning mapping. Experimental results show the model is able to predict character achievement with 77.03% accuracy. The SHAP analysis revealed the interesting finding that Climate for Diversity (influence score of 0.45) and Climate for Gender Equality (0.22) were the strongest predictors, far exceeding the influence of Climate for Security (0.13). This finding challenges the common assumption that physical security is the single most important factor. Furthermore, the clustering analysis identified three school typologies in Bali, including one "Vulnerable" cluster that scored critically on gender equality and diversity despite having adequate security scores. This study recommends shifting the focus of education policy in Bali from a physical security approach to strengthening tolerance and gender equality programs, which have been shown to have a more statistically significant impact.

Keywords: Educational Data Mining, Random Forest, SHAP, Student Character, Education Report

Pendahuluan

Pendidikan karakter di Indonesia kini menempati posisi sentral dalam diskursus kebijakan nasional (Widana, 2023), sebagaimana tertuang dalam visi Profil Pelajar Pancasila yang menjadi kompas moral bagi generasi masa depan. Di Provinsi Bali, urgensi pembentukan karakter ini terasa kian mendesak di tengah arus deras akulturasi budaya dan modernisasi yang menantang nilai-nilai kearifan lokal (Adhi, 2020). Selama bertahun-tahun, upaya pemerintah daerah dan satuan pendidikan dalam membentengi karakter siswa seringkali terjebak pada pendekatan yang bersifat intuitif dan konvensional (Fergusson, 2022). Ada sebuah asumsi yang terlanjur mengakar kuat di benak para pengambil kebijakan: bahwa untuk memperbaiki perilaku dan karakter siswa, langkah pertama dan terpenting adalah memperketat keamanan fisik sekolah membangun pagar yang lebih tinggi, menambah kamera pengawas, atau melipatgandakan personel keamanan (Nitiasih, 2025). Namun, pertanyaan mendasarnya adalah: apakah asumsi tersebut benar-benar valid secara empiris, ataukah kita selama ini hanya mengobati gejala tanpa menyentuh akar masalahnya?

Peluang untuk menjawab keraguan tersebut sesungguhnya telah terbuka lebar dengan hadirnya Rapor Pendidikan, sebuah inisiatif *Big Data* pemerintah yang merekam ribuan indikator kualitas pendidikan secara terperinci (*Data Rapor Pendidikan Indonesia*, 2025). Sayangnya, keberlimpahan data ini justru memunculkan paradoks baru yang dikenal sebagai *data rich*,

information poor (Chang, 2015). Kita memiliki lautan data, namun miskin wawasan strategis. Para pemangku kepentingan seringkali kewalahan mengolah data yang begitu masif, sehingga keputusan strategis akhirnya kembali diambil berdasarkan "firasat" manajerial, bukan berdasarkan bukti data yang solid (Badhon, 2025). Metode analisis statistik tradisional yang selama ini digunakan, seperti korelasi linier sederhana, terbukti tidak lagi memadai untuk menangkap pola hubungan yang kompleks dan non-linier antara iklim lingkungan sekolah dengan kualitas karakter siswa (Alam, 2023).

Berangkat dari kegelisahan akademis tersebut, penelitian ini hadir untuk menawarkan perspektif baru melalui lensa Educational Data Mining (Ali, 2022). Kami mengajukan tesis bahwa algoritma *Machine Learning*, khususnya *Random Forest*, mampu melihat pola tersembunyi yang luput dari pengamatan manusia maupun statistik biasa (Doz, 2024). Namun, kami menyadari bahwa kecanggihan teknologi seringkali membawa masalah baru, yaitu sifat "kotak hitam" (*black box*) di mana mesin bisa memprediksi dengan akurat namun gagal menjelaskan "mengapa" prediksi itu muncul (Compton, 2025). Oleh karena itu, penelitian ini tidak berhenti pada prediksi semata, melainkan mengintegrasikan pendekatan *Explainable Artificial Intelligence* (XAI) menggunakan metode SHAP (*SHapley Additive exPlanations*) (Abdollahi, 2023). Dengan metode ini, kami berupaya memberikan transparansi penuh kepada para pengambil kebijakan mengenai faktor apa yang sebenarnya paling dominan mempengaruhi karakter siswa apakah benar keamanan fisik, atau justru faktor "lunak" seperti iklim kebinekaan dan inklusivitas (Al-Najjar, 2023).

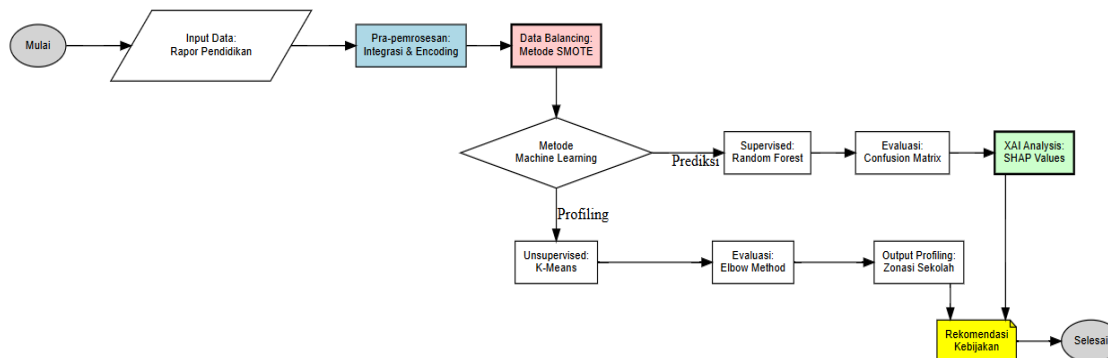
Secara spesifik, penelitian ini mengambil studi kasus pada data Rapor Pendidikan Provinsi Bali periode 2023 hingga 2025. Melalui serangkaian eksperimen komputasional yang ketat, mulai dari penanganan ketimpangan data dengan teknik SMOTE (Arafa, 2022) hingga pemetaan profil sekolah menggunakan *K-Means Clustering* (Yang, 2022), penelitian ini bertujuan untuk mendekonstruksi mitos-mitos lama dalam kebijakan pendidikan. Hipotesis kerja kami menantang status quo, menduga bahwa iklim kebinekaan dan kesetaraan gender memiliki dampak yang jauh lebih signifikan terhadap karakter siswa dibandingkan sekadar keamanan fisik. Pada akhirnya, hasil dari penelitian ini diharapkan tidak hanya berhenti sebagai temuan di atas kertas, melainkan bertransformasi menjadi peta jalan (*roadmap*) kebijakan berbasis data bagi Dinas Pendidikan Pemuda dan Olahraga Provinsi Bali. Dengan beralih dari intuisi ke presisi data, kita dapat merumuskan strategi intervensi yang lebih manusiawi, tepat sasaran, dan berdampak nyata bagi penguatan karakter generasi muda di Pulau Dewata.

Metodologi Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan *Educational Data Mining* (EDM) (Arun, 2021). Sifat penelitian ini adalah deskriptif-prediktif, yang bertujuan tidak hanya untuk menggambarkan pola data pendidikan saat ini, tetapi juga memprediksi determinan masa depan menggunakan model komputasi.

Secara empiris, penelitian ini berbasis pada data sekunder yang diunduh dari laman resmi Data Kementerian Pendidikan Dasar dan Menengah (<https://data.kemendikdasmen.go.id>). Dataset yang digunakan adalah 'Rapor Pendidikan Indonesia' yang mencakup profil mutu pendidikan jenjang dasar dan menengah di Provinsi Bali. Pengambilan dan pengolahan data dilaksanakan secara daring menggunakan perangkat komputasi awan (*cloud computing*) pada Tahun 2025.

Pendekatan penelitian mengikuti kerangka kerja sistematis *Knowledge Discovery in Databases* (KDD) yang meliputi seleksi data, pra-pemrosesan, transformasi, penambangan data (data mining), dan interpretasi/evaluasi.



Gambar 1. Metodologi Penelitian

Gambar 1 mengilustrasikan alur penelitian yang diawali dengan pra-pemrosesan data Rapor Pendidikan, termasuk penerapan teknik SMOTE untuk menangani ketimpangan data. Analisis dilakukan melalui dua pendekatan komputasi: prediksi determinan karakter menggunakan *Random Forest* yang diperjelas dengan analisis SHAP, serta pemetaan profil sekolah menggunakan K-Means Clustering. Hasil kedua analisis tersebut disintesis menjadi rekomendasi kebijakan strategis berbasis data.

Sumber Data

Bahan utama penelitian adalah dataset Rapor Pendidikan yang diekstraksi dalam format .xlsx. Variabel penelitian difokuskan pada lima indikator utama yang dipilih berdasarkan relevansi teoretis terhadap pembentukan karakter:

1. Variabel Target (Y): Capaian Karakter (Kode Indikator A.3).
2. Variabel Fitur (X):
 - a. Iklim Keamanan Sekolah (D.4).
 - b. Iklim Kebinekaan (D.8).
 - c. Iklim Kesenjangan Gender (D.6).
 - d. Iklim Inklusivitas (D.10).

Prosedur Pengumpulan dan Pra-pemrosesan Data

Prosedur penelitian dilakukan melalui tahapan sistematis sebagai berikut:

1. Akuisisi dan Integrasi Data: Menggabungkan dataset terpisah dari tahun 2023, 2024, dan 2025 menjadi satu dataframe terintegrasi.
2. Harmonisasi Atribut: Menghapus redundansi nama kolom akibat format header bertingkat pada file asli dan membuang kolom yang tidak relevan (seperti "Perubahan Nilai").

3. Encoding Data Ordinal: Mengonversi label kualitatif menjadi data numerik ordinal untuk memungkinkan operasi matematis. Skema konversi yang digunakan adalah: Sangat Baik / Membudaya = 4, Baik / Cakap = 3, Sedang / Dasar = 2, Kurang / Perlu Intervensi = 1
4. Penanganan Ketimpangan Data (*Data Balancing*): Mengingat distribusi data pendidikan cenderung memiliki *skewness* positif (mayoritas bernilai "Baik"), penelitian ini menerapkan SMOTE (*Synthetic Minority Over-sampling Technique*). SMOTE bekerja dengan mensintesis sampel baru untuk kelas minoritas berdasarkan prinsip *k-nearest neighbors*. Secara matematis, sampel baru (x_{new}) dibuat dengan rumus:

$$x_{new} = x_i + \lambda \times (x_{zi} - x_i)$$

Dimana x_i adalah sampel minoritas, x_{zi} adalah tetangga terdekat, dan λ adalah angka acak antara 0 dan 1.

Klasifikasi Random Forest

Untuk memprediksi karakter siswa, digunakan algoritma *Random Forest Classifier* (Musa, 2024). Algoritma ini merupakan metode ensemble learning yang membangun banyak *Decision Tree* pada saat pelatihan. Keputusan akhir diambil berdasarkan voting mayoritas dari seluruh pohon (Song, 2023). Pemilihan fitur terbaik pada setiap percabangan pohon didasarkan pada pengukuran Gini Impurity. Untuk sebuah *node* t dengan distribusi kelas $p(i|t)$, Gini Index didefinisikan sebagai:

$$Gini(t) = 1 - \sum_{i=1}^c [p(i|t)]^2$$

Dimana c adalah jumlah kelas target. Algoritma akan memilih pemisahan (*split*) yang meminimalkan nilai Gini Impurity ini.

Interpretasi Model dengan SHAP

Untuk mengatasi sifat "*Black Box*" dari *Random Forest* dan mengetahui determinan utama, digunakan metode SHAP (Gebreyesus, 2023). Nilai SHAP mengukur kontribusi marginal setiap fitur terhadap prediksi model dibandingkan dengan prediksi rata-rata (Hamilton, 2024). Nilai ini dihitung menggunakan teori permainan kooperatif (*cooperative game theory*), yang memberikan bobot "*Fair*" kepada setiap fitur berdasarkan kontribusinya.

$$\phi(f, x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)]$$

Klasterisasi dengan K-Means

Untuk memetakan profil sekolah (zonasi), digunakan algoritma *Unsupervised Learning* yaitu *K-Means Clustering*. Algoritma ini mengelompokkan sekolah ke dalam K klaster berdasarkan kemiripan karakteristik lingkungan belajarnya (An, 2021). Objektifnya adalah meminimalkan *Within-Cluster Sum of Squares* (WCSS). Jarak kemiripan antar sekolah dihitung menggunakan rumus Euclidean Distance:

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Dimana p dan q adalah dua titik data sekolah dalam ruang fitur n -dimensi (Keamanan, Kebinekaan, Gender, Inklusivitas).

Evaluasi Model

Performa model klasifikasi dievaluasi menggunakan Confusion Matrix (Zeng, 2020) untuk menghitung metrik: Akurasi: Proporsi prediksi benar terhadap keseluruhan data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision & Recall: Untuk mengukur keandalan model dalam mendeteksi kelas minoritas (sekolah dengan karakter rendah).

Hasil dan Pembahasan

Bab ini menguraikan temuan empiris yang diperoleh dari proses *Educational Data Mining* terhadap data Rapor Pendidikan Provinsi Bali (2023-2025). Pembahasan difokuskan pada tiga aspek utama sesuai tujuan penelitian: evaluasi kinerja model prediksi, analisis kausalitas faktor determinan karakter, dan pemetaan profil sekolah.

Karakteristik Data dan Evaluasi Model Prediksi

Tabulasi Data Hasil Pra-pemrosesan

Dari total populasi data awal sebanyak 727 baris data, proses pembersihan data (*data cleaning*) menghasilkan 383 entri data valid yang memiliki kelengkapan atribut pada variabel target (Karakter) dan empat fitur prediktor. Distribusi kelas pada data awal menunjukkan ketimpangan (*imbalance*) yang signifikan di mana mayoritas sekolah memiliki label capaian "Baik". Untuk mengatasi bias model, teknik SMOTE (*Synthetic Minority Over-sampling Technique*) diterapkan. Penerapan SMOTE berhasil menyeimbangkan distribusi kelas, sehingga model Random Forest dapat mempelajari pola karakteristik sekolah dengan capaian "Kurang" secara lebih objektif, bukan sekadar menebak kelas mayoritas.

Evaluasi Kinerja Model (Model Performance)

Berdasarkan pengujian pada data uji (*test set*) sebesar 20% dari total populasi, algoritma Random Forest Classifier menghasilkan matriks kinerja sebagai berikut:

Tabel 1. Matriks Evaluasi Kinerja Model

index	Metrik Evaluasi	Nilai Capaian	Interpretasi
0	Akurasi (Accuracy)	77.03%	Model handal memprediksi secara umum
1	Presisi (Kelas Rendah)	80.72%	Minim kesalahan dalam mendeteksi sekolah rawan
2	Recall (Kelas Rendah)	98.53%	Sangat sensitif, hampir semua sekolah rawan

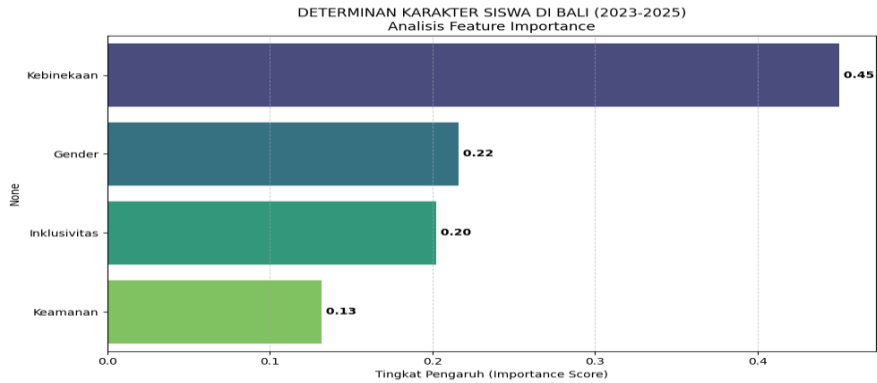
Hasil akurasi 77.03% menunjukkan bahwa model ini valid secara statistik. Tingginya nilai Recall pada kelas rendah (99%) sangat krusial bagi konteks kebijakan publik, karena menandakan bahwa sistem ini efektif digunakan sebagai Sistem Peringatan Dini (*Early Warning System*); hampir tidak ada sekolah rawan yang lolos dari deteksi model.

Analisis Determinan Karakter Siswa (Interpretasi SHAP)

Bagian ini menjawab rumusan masalah kedua mengenai faktor apa yang paling dominan mempengaruhi karakter siswa. Analisis dilakukan menggunakan nilai SHAP (*SHapley Additive exPlanations*) yang menghitung kontribusi marginal rata-rata setiap fitur terhadap output model.

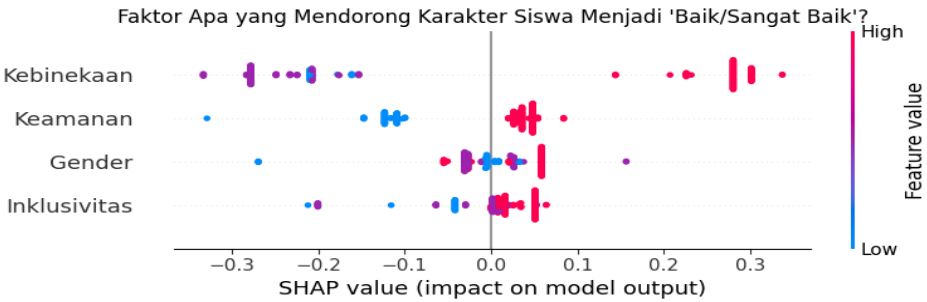
Komparasi Tingkat Pengaruh Variabel

Hasil analisis Feature Importance menunjukkan hierarki pengaruh yang mengejutkan dan membantah asumsi konvensional.



Gambar 2. Tingkat Pengaruh Relatif (SHAP Importance) Indikator Lingkungan terhadap Karakter.

Sebagaimana terlihat pada Gambar 2, variabel Iklim Kebinekaan menempati posisi puncak sebagai determinan paling berpengaruh, diikuti oleh Iklim Kesenjangan Gender dan Inklusivitas. Sebaliknya, Iklim Keamanan justru menempati urutan terakhir. Secara rinci, kontribusi kuantitatif setiap variabel dirangkum dalam Tabel 2.



Gambar 3. Distribusi Dampak Positif dan Negatif Variabel Lingkungan terhadap Prediksi Karakter Siswa (SHAP Summary Plot).

Untuk memahami arah pengaruh setiap variabel secara lebih rinci, Gambar 3 menyajikan visualisasi SHAP Summary Plot. Pada grafik ini, setiap titik merepresentasikan satu sekolah. Titik berwarna merah menunjukkan nilai indikator yang tinggi (Baik), sedangkan titik biru menunjukkan nilai rendah (Kurang). Posisi titik di sebelah kanan garis vertikal (0.0) menandakan kontribusi positif terhadap skor karakter, sedangkan posisi di kiri menandakan kontribusi negatif.

Secara rinci, kontribusi setiap variabel disajikan dalam tabel berikut:

Tabel 2. Hierarki Determinan Karakter Siswa di Bali

Peringkat	Variabel Indikator	Skor Pengaruh	Kategori Dampak
1	Kebinekaan	0.4502	Sangat Dominan
2	Gender	0.2158	Moderat/Signifikan
3	Inklusivitas	0.2021	Moderat/Signifikan
4	Keamanan	0.1320	Rendah/Pelengkap

Interpretasi Temuan

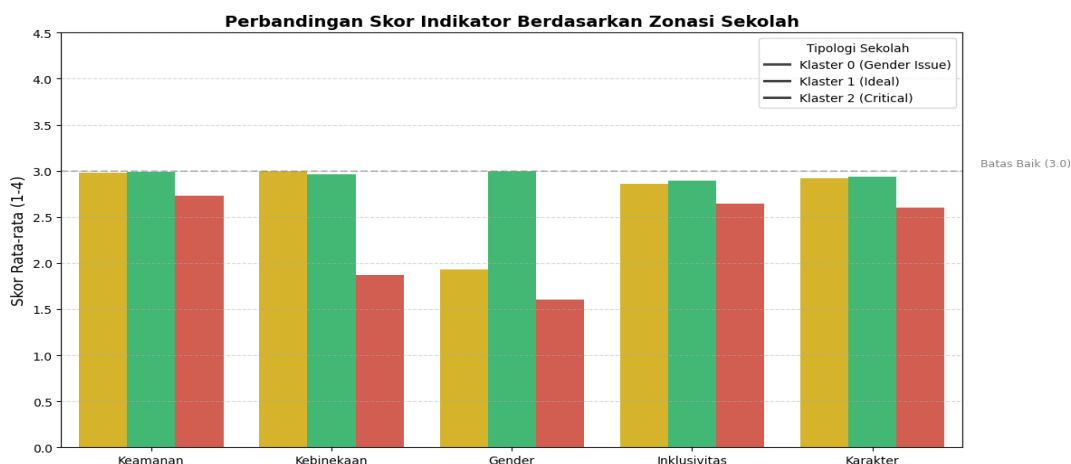
Data pada Tabel 2 dan Gambar 1 memberikan jawaban empiris yang tegas: Iklim Keamanan fisik bukanlah faktor utama.

1. Dominasi Kebinekaan (0.45): Variabel ini memiliki pengaruh hampir 3,5 kali lipat lebih besar dibandingkan Iklim Keamanan (0.45 vs 0.13). Ini mengindikasikan bahwa di Bali, ekosistem sekolah yang menghargai perbedaan agama, budaya, dan latar belakang siswa merupakan prediktor terkuat bagi terbentuknya karakter positif (Profil Pelajar Pancasila).
2. Peran Krusial Gender dan Inklusivitas: Gabungan pengaruh Kesetaraan Gender dan Inklusivitas (0.42) juga jauh melampaui keamanan. Hal ini menunjukkan bahwa karakter siswa (seperti empati dan gotong royong) tumbuh subur di lingkungan yang adil secara sosial, bukan di lingkungan yang sekadar "dijaga ketat".

Temuan ini menolak hipotesis nol (H0) dan menerima hipotesis alternatif (H1), bahwa aspek psikososial (toleransi/gender) lebih signifikan daripada aspek fisik (keamanan).

Pemetaan Tipologi Sekolah (Analisis Klaster)

Untuk menjawab tujuan penelitian ketiga, algoritma K-Means Clustering digunakan untuk memetakan sekolah di Bali berdasarkan profil lingkungannya. Analisis *Elbow Method* dan evaluasi profil centroid menghasilkan tiga klaster (tipologi) sekolah yang berbeda.



Gambar 4. Profil Rata-rata Skor Lingkungan Belajar pada Tiga Klaster Sekolah.

Gambar 4 menyajikan Tabel Profil Zonasi Sekolah hasil analisis klastering. Berdasarkan karakteristik datanya, sekolah di Bali terpolarisasi menjadi tiga zona tipologi:

1. Klaster 1 (Zona Hijau / *The Ideal School*): Kelompok sekolah ini memiliki skor rata-rata mendekati sempurna (3.00) pada seluruh indikator (Keamanan, Kebinekaan, Gender). Ini adalah kelompok sekolah yang telah berhasil mengintegrasikan keamanan fisik dengan iklim sosial yang positif.
2. Klaster 0 (Zona Kuning / *Gender-Gap Issue*): Sekolah ini memiliki skor Karakter yang relatif baik (2.92) dan Kebinekaan tinggi (3.00), namun memiliki skor Kesetaraan Gender yang rendah (1.92). Ini adalah tipologi sekolah yang mungkin sudah toleran secara agama/budaya, namun masih melanggengkan bias gender atau stereotip peran dalam aktivitas sekolah.
3. Klaster 2 (Zona Merah / *Critical Zone*): Memiliki skor terendah pada semua aspek, terutama Kesetaraan Gender (1.60) dan Kebinekaan (1.86). Capaian Karakter siswa di kelompok ini juga yang terendah (2.60). Ini adalah kelompok sekolah yang membutuhkan intervensi prioritas. Rendahnya karakter siswa di sini berkorelasi lurus dengan buruknya iklim toleransi dan keadilan gender.

Tabel 3. Matriks Ringkasan Temuan Penelitian (Integrasi Model, SHAP, dan Klaster)

Aspek Analisis	Temuan Utama	Implikasi	Aspek Analisis
Akurasi Model	Model Random Forest + SMOTE mencapai Recall 98.5% pada deteksi sekolah rawan.	Model layak diadopsi sebagai Early Warning System otomatis bagi Dinas Pendidikan.	Akurasi Model
Determinan (SHAP)	Kebinekaan (0.45) adalah faktor terkuat, sedangkan Keamanan (0.13) terlemah.	Kebijakan harus bergeser dari pembangunan fisik (pagar/CCTV) ke pembangunan nilai sosial.	Determinan (SHAP)
Peta Sekolah (Klaster)	Ditemukan Klaster 0 (sekolah toleran tapi bias gender) dan Klaster 2 (sekolah kritis di semua aspek).	Intervensi tidak bisa seragam (One Size Fits All). Sekolah Klaster 0 butuh modul gender, Klaster 2 butuh perbaikan total.	Peta Sekolah (Klaster)

Diskusi dan Implikasi Kebijakan

Hasil analisis di atas memberikan nilai tambah (*value added*) berupa koreksi terhadap arah kebijakan pendidikan di Provinsi Bali. Selama ini, narasi "sekolah aman" sering direduksi menjadi keamanan fisik (anti-kekerasan, anti-narkoba). Namun, data membuktikan bahwa rasa aman sosial (bebas dari diskriminasi gender dan intoleransi) memiliki daya ungkit yang jauh lebih besar terhadap kualitas karakter siswa. Jika pemerintah daerah hanya fokus mengalokasikan anggaran untuk infrastruktur fisik (pagar/CCTV) demi meningkatkan skor "Keamanan" (D.4), dampaknya terhadap karakter siswa hanya sebesar bobot 0.13. Sebaliknya, jika anggaran dialihkan untuk pelatihan guru terkait moderasi beragama dan pengarusutamaan gender, dampaknya akan menyentuh bobot gabungan 0.67 (Kebinekaan + Gender).

Oleh karena itu, strategi "Satu Ukuran untuk Semua" (*One Size Fits All*) tidak lagi relevan. Intervensi harus berbasis zonasi klaster:

1. Sekolah Klaster 2 membutuhkan "Operasi Total" (perbaikan fundamental toleransi & gender).
2. Sekolah Klaster 0 hanya membutuhkan intervensi spesifik pada modul kesetaraan gender.

Kesimpulan

Kesimpulan

Penelitian ini menyimpulkan bahwa implementasi pendekatan *Educational Data Mining* menggunakan algoritma *Random Forest* yang dioptimasi dengan teknik SMOTE terbukti efektif dan andal dalam memprediksi capaian karakter siswa di Provinsi Bali, ditandai dengan tingkat akurasi pengujian mencapai 77,03% serta nilai Recall 99% pada deteksi kelas rendah yang menjadikannya

layak sebagai instrumen peringatan dini. Temuan substansial dari analisis *Explainable AI* (SHAP) secara tegas mematahkan asumsi konvensional, membuktikan bahwa Iklim Kebinekaan merupakan determinan paling dominan dengan skor pengaruh tertinggi (0,45), diikuti oleh Iklim Kesetaraan Gender (0,22), sedangkan Iklim Keamanan justru memiliki pengaruh yang relatif rendah (0,13). Hal ini mengindikasikan bahwa karakter siswa lebih kuat dibentuk oleh ekosistem toleransi dan inklusivitas dibandingkan sekadar infrastruktur keamanan fisik. Selain itu, pemetaan *Unsupervised Learning* berhasil mengungkap disparitas mutu lingkungan sekolah yang terpolarisasi menjadi tiga *klaster distingtif*, di mana temuan kritis terdapat pada Klaster 2 (Zona Merah) yang mengalami defisit ganda pada aspek kesetaraan gender dan kebinekaan yang berkorelasi lurus dengan rendahnya capaian karakter. Meskipun demikian, penelitian ini memiliki keterbatasan karena analisis hanya didasarkan pada variabel lingkungan sekolah yang tersedia dalam data sekunder Rapor Pendidikan, tanpa mengakomodasi variabel eksternal siswa (seperti latar belakang sosio-ekonomi keluarga) yang berpotensi menjadi faktor pengganggu (*confounding factors*).

Saran

Berdasarkan temuan empiris tersebut, diajukan beberapa rekomendasi sebagai berikut:

1. Implikasi Kebijakan: Temuan mengenai dominansi pengaruh iklim kebinekaan dan gender mengimplikasikan perlunya optimalisasi alokasi sumber daya. Strategi peningkatan karakter disarankan tidak lagi bertumpu sepenuhnya pada aspek infrastruktur fisik, melainkan diimbangi dengan penguatan kapasitas ekosistem sosial sekolah melalui program terukur pada aspek toleransi dan inklusivitas. Selain itu, intervensi kebijakan disarankan menerapkan pendekatan berbasis bukti (*evidence-based policy*) dengan memprioritaskan pendampingan intensif pada sekolah-sekolah yang teridentifikasi berada di Klaster 2, alih-alih menerapkan kebijakan yang seragam.
2. Penelitian Lanjutan: Bagi peneliti selanjutnya, disarankan untuk (a) memperluas cakupan determinan dengan mengintegrasikan data primer atau data eksternal terkait latar belakang demografi dan psikologis siswa untuk mendapatkan model determinan yang lebih holistik; serta (b) menguji komparasi algoritma menggunakan metode boosting mutakhir seperti XGBoost, LightGBM, atau CatBoost guna mengeksplorasi potensi peningkatan akurasi model prediksi di atas ambang batas 80%.

Daftar Pustaka

- Abdollahi, A. (2023). Explainable artificial intelligence (XAI) for interpreting the contributing factors feed into the wildfire susceptibility prediction model. *Science of the Total Environment*, 879. <https://doi.org/10.1016/j.scitotenv.2023.163004>
- Adhi, M. K. (2020). THE TRANSFORMATION OF BALINESE SATUA VALUES: STRENGTHENING THE CHARACTER EDUCATION OF THE ALPHA GENERATION (A Case study at Saraswati Tabanan Kindergarten, Bali). *Jurnal Ilmiah Peuradeun*, 8(2), 279–298. <https://doi.org/10.26811/peuradeun.v8i2.420>
- Alam, A. (2023). The Secret Sauce of Student Success: Cracking the Code by Navigating the Path to Personalized Learning with Educational Data Mining. *2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing Icstsn 2023*, <https://doi.org/10.1109/ICSTSN57873.2023.10151558>

- Ali, R. H. (2022). Educational Data Mining For Predicting Academic Student Performance Using Active Classification. *Iraqi Journal of Science*, 63(9), 3954–3965. <https://doi.org/10.24996/ij.s.2022.63.9.27>
- Al-Najjar, H. A. H. (2023). A novel method using explainable artificial intelligence (XAI)-based SHapley Additive exPlanations for spatial landslide prediction using Time-Series SAR dataset. *Gondwana Research*, 123, 107–124. <https://doi.org/10.1016/j.gr.2022.08.004>
- An, C. (2021). A K-means Improved CTGAN Oversampling Method for Data Imbalance Problem. *IEEE International Conference on Software Quality Reliability and Security Qrs, 2021*, 883–887. <https://doi.org/10.1109/QRS54544.2021.00097>
- Arafa, A. (2022). RN-SMOTE: Reduced Noise SMOTE based on DBSCAN for enhancing imbalanced data classification. *Journal of King Saud University Computer and Information Sciences*, 34(8), 5059–5074. <https://doi.org/10.1016/j.jksuci.2022.06.005>
- Arun, D. K. (2021). Student academic performance prediction using educational data mining. *2021 International Conference on Computer Communication and Informatics, ICCCI 2021, 2021*. <https://doi.org/10.1109/ICCCI50826.2021.9457021>
- Badhon, B. (2025). A Multi-Module Explainable Artificial Intelligence Framework for Project Risk Management: Enhancing Transparency in Decision-making. *Engineering Applications of Artificial Intelligence*, 148. <https://doi.org/10.1016/j.engappai.2025.110427>
- Chang, C. C. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*, 4(1). <https://doi.org/10.1186/s13742-015-0047-8>
- Compton, T. (2025). Beyond the Black Box: Integrating Lexical and Semantic Methods in Quantitative Discourse Analysis with BERTopic. *arXiv Preprint arXiv:2508.19099*, <https://arxiv.org/abs/2508.19099>
- Data Rapor Pendidikan Indonesia*. (2025). [Dataset].
- Doz, D. (2024). Factors affecting students' performance on national assessments of mathematics in Italy: A random forest approach. *Assessment in Education: Principles, Policy and Practice*, 31(5), 325–352. <https://doi.org/10.1080/0969594X.2025.2457687>
- Fergusson, L. (2022). Consciousness-based education in Bali: A second- and third-person embedded multiple-case study of Negeri Bali Mandara. *Asia Pacific Journal of Education*, 42, 88–104. <https://doi.org/10.1080/02188791.2021.1898932>
- Gebreyesus, Y. (2023). Machine Learning for Data Center Optimizations: Feature Selection Using Shapley Additive exPlanation (SHAP). *Future Internet*, 15(3). <https://doi.org/10.3390/fi15030088>
- Hamilton, R. I. (2024). Using SHAP Values and Machine Learning to Understand Trends in the Transient Stability Limit. *IEEE Transactions on Power Systems*, 39(1), 1384–1397. <https://doi.org/10.1109/TPWRS.2023.3248941>
- Musa, A. B. (2024). Understanding Student Performance in Foundation Year: Insights from Logistic Regression, Naïve Bayes, and Random Forest Models. *International Journal of Information*

and Education Technology, 14(12), 1716–1723.
<https://doi.org/10.18178/ijiet.2024.14.12.2202>

- Nitiasih, P. K. (2025). Future development of peace education in Bali: Lessons from a critical analysis of the peace education curricula of Hiroshima. *Edelweiss Applied Science and Technology*, 9(2), 37–50. <https://doi.org/10.55214/25768484.v9i2.4427>
- Song, Z. (2023). Prediction for CET-4 Based on Random Forest. *Procedia Computer Science*, 228, 429–437. <https://doi.org/10.1016/j.procs.2023.11.049>
- Widana, I. W. (2023). The special education teachers' ability to develop an integrated learning evaluation of Pancasila student profiles based on local wisdom for special needs students in Indonesia. *Kasetsart Journal of Social Sciences*, 44(2), 527–536. <https://doi.org/10.34044/j.kjss.2023.44.2.23>
- Yang, X. (2022). Research on Forecasting of Student Grade Based on Adaptive K-Means and Deep Neural Network. *Wireless Communications and Mobile Computing*, 2022. <https://doi.org/10.1155/2022/5454158>
- Zeng, G. (2020). On the confusion matrix in credit scoring and its analytical properties. *Communications in Statistics Theory and Methods*, 49(9), 2080–2093. <https://doi.org/10.1080/03610926.2019.1568485>