Jurnal Ilmu Komputer dan Sistem Informasi (JIRSI) **e-ISSN 2830-3954**
Volume: 5, Nomor: 2, Mei 2026: 687-703 **p-ISSN 2830-6031**
https://jurnal.unity-academy.sch.id/index.php/jirsi/index

Komparasi Algoritma Machine Learning Untuk Klasifikasi Risiko Depresi Remaja Berbasis Perilaku Digital Menggunakan SMOTE

Comparative Analysis of Machine Learning Algorithms for Adolescent Depression Risk Classification Based on Digital Behavior Using SMOTE

Sarwadi*[1], Muhammad Khaibar Putra Adithia[2]
[1,2]Fakultas Ilmu Komputer, Universitas Nahdlatul Ulama Sumatera Utara
E-mail: [1]sarwadi69@gmail.com, [2]ibaradithia94@gmail.com

Abstrak

Depresi pada remaja merupakan masalah kesehatan mental global yang semakin mengkhawatirkan, terutama di era percepatan adopsi platform media sosial. Penelitian ini membandingkan performa empat algoritma machine learning, Random Forest, XGBoost, Support Vector Machine (SVM), dan Logistic Regression, dalam mengklasifikasikan risiko depresi remaja berbasis 12 fitur perilaku digital dan gaya hidup. Dataset yang digunakan terdiri dari 1.200 rekaman dengan ketidakseimbangan kelas yang ekstrem (97,4% vs 2,6%), ditangani melalui penerapan Synthetic Minority Over-sampling Technique (SMOTE) dengan k=5. Delapan eksperimen dilakukan (empat algoritma × dua kondisi) dan dievaluasi menggunakan Precision, Recall, F1-Macro, AUC-ROC, serta 5-fold cross-validation. Hasil menunjukkan SMOTE secara konsisten meningkatkan performa semua model, dengan peningkatan F1-Macro rata-rata sebesar 25,7%. XGBoost unggul dengan F1-Macro dan AUC-ROC sempurna (1,0000) pada kedua kondisi, sedangkan pengaruh SMOTE paling besar dirasakan oleh SVM (+67,9%) dan Logistic Regression (+63,7%). Analisis feature importance mengidentifikasi jam tidur, tingkat stres, tingkat kecemasan, dan penggunaan media sosial harian sebagai prediktor dominan, sementara prestasi akademik memiliki kontribusi yang dapat diabaikan. Temuan ini menegaskan pentingnya penanganan imbalanced data dan memberikan kerangka metodologis untuk pengembangan sistem deteksi dini depresi berbasis perilaku digital.

***Kata kunci**: machine learning, depresi remaja, SMOTE, XGBoost, imbalanced dataset*

Abstract

Adolescent depression is an increasingly alarming global mental health issue, particularly amid the rapid adoption of social media platforms. This study compares the performance of four machine learning algorithms Random Forest, XGBoost, Support Vector Machine (SVM), and Logistic Regression, in classifying adolescent depression risk using 12 digital behavioral and lifestyle features. The dataset comprises 1,200 records with severe class imbalance (97.4% vs 2.6%), addressed through Synthetic Minority Over-sampling Technique (SMOTE) with k=5. Eight experiments (four algorithms × two conditions) were conducted and evaluated using Precision, Recall, F1-Macro, AUC-ROC, and 5-fold cross-validation. Results demonstrate that SMOTE consistently enhanced all model performances, with an average F1-Macro improvement of 25.7%. XGBoost achieved perfect F1-Macro and AUC-ROC scores (1.0000) under both conditions, while SMOTE yielded the greatest gains for SVM (+67.9%) and Logistic Regression (+63.7%). Feature importance analysis identified sleep hours, stress level, anxiety

level, and daily social media usage as dominant predictors, while academic performance contributed negligibly. These findings underscore the critical importance of imbalanced data handling and provide a methodological framework for developing digital behavior-based early depression detection systems.

***Keywords**: machine learning, adolescent depression, SMOTE, XGBoost, imbalanced dataset*

1. PENDAHULUAN

Kesehatan mental remaja telah menjadi perhatian global yang semakin mendesak. *World Health Organization* (WHO) menyatakan bahwa depresi merupakan salah satu penyebab utama disabilitas pada kelompok usia 15–29 tahun, dengan estimasi 280 juta individu di seluruh dunia mengalaminya [1]. Di Indonesia, Survei Kesehatan Indonesia (SKI) terbaru yang dirilis oleh Kementerian Kesehatan RI mencatat prevalensi gangguan mental emosional pada penduduk berusia 15 tahun ke atas berada di angka 6,1%, dengan temuan krusial bahwa kelompok remaja (usia 15–24 tahun) justru mendominasi angka depresi tertinggi secara nasional, yaitu mencapai 2% [2]. Fenomena ini diperkuat oleh laporan *Indonesia National Adolescent Mental Health Survey* (I-NAMHS) yang mengonfirmasi bahwa 1 dari 3 remaja Indonesia memiliki masalah kesehatan emosional dalam 12 bulan terakhir [3]. Kondisi ini diperparah oleh terbatasnya akses layanan kesehatan jiwa, di mana lebih dari 75% individu dengan gangguan mental di negara berkembang tidak mendapatkan penanganan memadai [1].

Proliferasi platform media sosial dan meningkatnya intensitas penggunaan perangkat digital telah mengubah pola perilaku dan interaksi sosial remaja secara fundamental. Data *We Are Social* dan *Hootsuite* tahun 2023 mencatat rata-rata pengguna internet Indonesia menghabiskan 7 jam 42 menit per hari daring, dengan remaja sebagai kelompok pengguna paling aktif [4]. Sejumlah penelitian mengidentifikasi korelasi signifikan antara penggunaan media sosial berlebihan dengan risiko depresi, gangguan kecemasan, dan penurunan kualitas tidur pada remaja [5]. Faktor seperti *cyberbullying*, perbandingan sosial ke atas (*upward social comparison*), serta disrupsi ritme sirkadian akibat paparan layar menjelang tidur turut memperburuk kondisi kesehatan mental remaja secara kumulatif [6].

Deteksi dini risiko depresi merupakan strategi kritis dalam upaya pencegahan dan pemberian intervensi yang tepat waktu. Pendekatan konvensional berbasis asesmen klinis menghadapi kendala keterbatasan sumber daya tenaga profesional, khususnya di wilayah terpencil. *Machine learning* menawarkan alternatif yang menjanjikan melalui kemampuannya mengidentifikasi pola kompleks dalam data multidimensi secara otomatis dan berbiaya rendah [7]. Namun, mayoritas penelitian di bidang ini menghadapi tantangan fundamental berupa ketidakseimbangan kelas (*class imbalance*) yang dapat mendistorsi performa model

secara sistemik, terutama dalam hal deteksi kelas minoritas yang justru paling kritikal dari perspektif klinis [8].

Berbagai penelitian telah menunjukkan bahwa *machine learning* memiliki potensi yang besar dalam mendukung deteksi dan klasifikasi gangguan kesehatan mental. Salah satu studi terdahulu menunjukkan bahwa kombinasi *Principal Component Analysis* (PCA) dan *Naïve Bayes* mampu meningkatkan efisiensi model tanpa menurunkan performa klasifikasi secara signifikan [9]. Temuan tersebut menunjukkan bahwa pendekatan machine learning dapat dimanfaatkan untuk mendukung identifikasi dini kondisi psikologis berdasarkan pola data perilaku dan psikometrik. Penelitian terkini mengenai deteksi depresi remaja umumnya memanfaatkan algoritma klasifikasi seperti *Random Forest*, *Support Vector Machine* (*SVM*), dan *XGBoost* untuk meningkatkan akurasi prediksi. Namun, sebagian besar penelitian masih menghadapi masalah ketidakseimbangan kelas (*class imbalance*) yang berpotensi menghasilkan bias terhadap kelas mayoritas. Selain itu, evaluasi performa model sering kali hanya berfokus pada *metrik accuracy*, sehingga kemampuan model dalam mendeteksi kasus depresi sebagai kelas minoritas belum dievaluasi secara optimal.

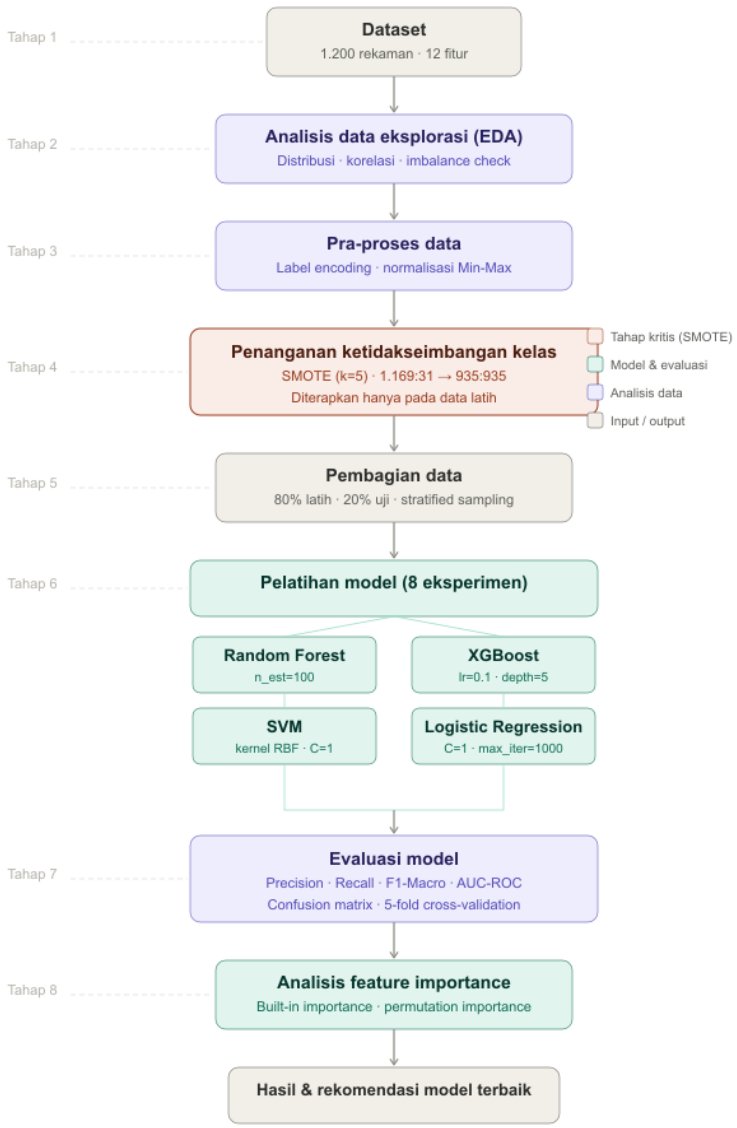
Synthetic Minority Over-sampling Technique (*SMOTE*) merupakan pendekatan yang terbukti efektif mengatasi ketidakseimbangan kelas melalui interpolasi sampel sintetis pada kelas minoritas menggunakan *k-nearest neighbor* [10]. Meski efektivitasnya telah divalidasi lintas domain, kajian komparatif sistematis yang mengevaluasi dampak *SMOTE* terhadap beragam algoritma klasifikasi khususnya dalam konteks deteksi depresi remaja berbasis perilaku digital multifitur masih jarang ditemukan dalam literatur berbahasa Indonesia.

Berdasarkan identifikasi kesenjangan penelitian tersebut, studi ini bertujuan: (1) membandingkan performa empat algoritma *machine learning* dalam mengklasifikasikan risiko depresi remaja; (2) mengkuantifikasi dampak *SMOTE* terhadap peningkatan performa masing-masing model; dan (3) mengidentifikasi fitur perilaku digital paling determinatif terhadap risiko depresi. Kontribusi utama penelitian ini adalah kerangka evaluasi sistematis kombinasi *SMOTE* dan empat algoritma klasifikasi disertai analisis *feature importance* komprehensif, yang memberikan panduan empiris bagi pengembangan sistem deteksi dini depresi remaja di Indonesia.

2. METODOLOGI PENELITIAN

Penelitian ini mengadopsi pendekatan eksperimental kuantitatif yang mencakup delapan tahap berurutan: pengumpulan dan eksplorasi *dataset*, pra-proses data, penanganan ketidakseimbangan kelas, pembagian data, pelatihan dan evaluasi model, serta analisis *interpretabilitas*. Alur metodologi penelitian digambarkan secara terstruktur pada Gambar 1:

e-ISSN 2830-3954
p-ISSN 2830-6031



Gambar 1. Diagram alur metodologi penelitian

2.1. Dataset

Penelitian memanfaatkan *Teen Mental Health Dataset*, sebuah dataset sintetis yang dikonstruksi untuk keperluan penelitian *machine learning* di bidang kesehatan mental remaja. Dataset terdiri dari 1.200 rekaman dengan 12 fitur prediktor dan satu variabel target (label depresi). Tidak terdapat nilai yang hilang (*missing value*) pada keseluruhan rekaman, sehingga tidak diperlukan imputasi.

Fitur-fitur dalam dataset dikelompokkan ke dalam empat kategori: (1) demografi usia dan jenis kelamin; (2) perilaku media sosial jam penggunaan harian dan

platform pilihan; (3) gaya hidup jam tidur, waktu layar sebelum tidur, aktivitas fisik, dan tingkat adiksi; serta (4) indikator psikologis tingkat stres, kecemasan, prestasi akademik, dan tingkat interaksi sosial.

Distribusi variabel target memperlihatkan ketidakseimbangan kelas yang ekstrem: 1.169 rekaman (97,4%) berlabel tidak depresi dan 31 rekaman (2,6%) berlabel depresi, menghasilkan rasio 37,7:1. Tabel 1 merangkum statistik deskriptif fitur numerik.

Tabel 1. Statistik deskriptif fitur numerik *dataset*

Fitur	Min	Maks	Rata-rata	Std. Dev.
Usia (tahun)	13	19	15,93	2,02
Jam media sosial/hari	1,0	8,0	4,54	2,03
Jam tidur/hari	4,0	9,0	6,45	1,44
Waktu layar sebelum tidur (jam)	0,5	3,0	1,74	0,72
Tingkat stres (1–10)	1	10	5,45	2,90
Tingkat kecemasan (1–10)	1	10	5,64	2,86
Tingkat adiksi (1–10)	1	10	5,56	2,83
Prestasi akademik (skala 2–4)	2,0	4,0	2,99	0,58

2.2. Pra-proses Data

Pra-proses dilakukan melalui dua langkah. Pertama, label *encoding* diterapkan pada tiga variabel kategorikal: jenis_kelamin (male/female → 0/1), platform penggunaan (Instagram/TikTok/Both→0/1/2), dan tingkat_interaksi_sosial (*low*/*medium*/*high* → 0/1/2). Kedua, normalisasi Min-Max diterapkan pada semua fitur numerik untuk memetakan nilai ke rentang [0, 1], memastikan algoritma berbasis jarak tidak terdistorsi oleh perbedaan skala antar fitur.

2.3. Penanganan Ketidakseimbangan Kelas

SMOTE bekerja dengan membuat sampel sintetis pada kelas minoritas melalui interpolasi *k-nearest neighbor*. Untuk setiap sampel minoritas x_i, sampel baru x_{new} dihitung sebagai:

$$X_{new} = x_i + \lambda\, x(X_{nn} - X_i), \lambda \epsilon [0,1]$$

di mana x_{nn} adalah tetangga terdekat yang dipilih secara acak dan λ adalah bilangan acak seragam [11]. *SMOTE* diterapkan eksklusif pada data latih (k=5, random_state=42), sehingga data uji tidak termodifikasi guna menjamin evaluasi yang representatif dan bebas dari kebocoran data (*data leakage*).

2.4. Algoritma Klasifikasi

Empat algoritma dibandingkan. (1) *Random Forest*: *ensemble bagging* berbasis pohon keputusan dengan *voting* mayoritas [12], n_*estimators*=100. (2) *XGBoost*: *gradient boosting* dengan regularisasi L1/L2 [13], n_*estimators*=100, max_depth=5, *learning rate*=0,1. (3) SVM: *hyperplane* optimal dengan *kernel RBF* [14], C=1,0, gamma='scale'. (4) *Logistic Regression*: model probabilistik linear sebagai *baseline* [15], C=1,0, max_iter=1.000.

2.5. Desain Eksperimen dan Metrik Evaluasi

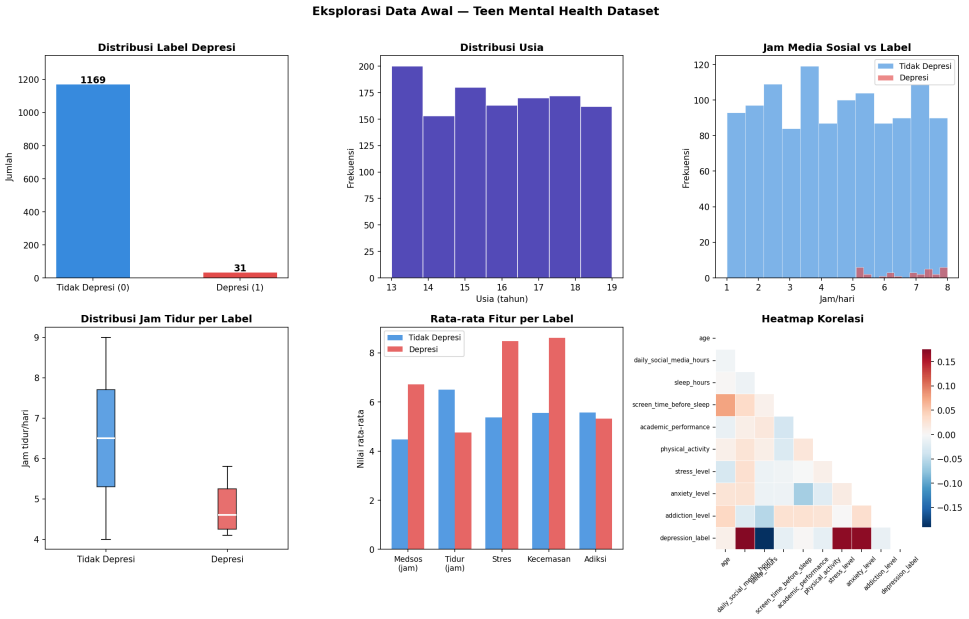
Data dibagi dengan rasio 80:20 menggunakan *stratified sampling*. Delapan eksperimen dilakukan empat algoritma masing-masing diuji pada dua kondisi (tanpa dan dengan *SMOTE*). Validasi silang 5-*fold* (5-*fold CV*) diterapkan pada data dengan *SMOTE* untuk menilai stabilitas model.

Akurasi tidak dipilih sebagai metrik utama karena sifatnya yang menyesatkan pada dataset tidak seimbang. Metrik yang digunakan: *Precision* (proporsi prediksi positif yang benar), *Recall* (proporsi kasus positif aktual yang terdeteksi), *F1-Macro* (rata-rata harmonik *precision* dan *recall*, dibobot merata lintas kelas), dan *AUC-ROC* (kemampuan diskriminasi model lintas *threshold*) [16].

3. HASIL DAN PEMBAHASAN

3.1. Hasil Analisis Data Eksplorasi

Analisis data eksplorasi (EDA) dilakukan sebagai langkah awal untuk memahami karakteristik dan distribusi dataset sebelum pemodelan. Gambar 2 menyajikan rangkuman visual *komprehensif* yang mencakup enam aspek analisis: distribusi label, distribusi usia, pola penggunaan media sosial, distribusi jam tidur per kelas, perbandingan rata-rata fitur, dan *heatmap korelasi*.



Gambar 2. Hasil eksplorasi data awal: distribusi label, usia, pola media sosial, tidur, rata-rata fitur per kelas, dan heatmap korelasi

Panel kiri atas memperlihatkan distribusi label depresi yang sangat tidak seimbang: 1.169 rekaman (97,4%) berlabel tidak depresi dan 31 rekaman (2,6%) berlabel depresi, dengan rasio ketidakseimbangan 37,7:1. Nilai ini jauh melampaui batas 10:1 yang dalam literatur *machine learning* dikategorikan sebagai ketidakseimbangan kelas berat, sehingga penanganan khusus melalui *SMOTE* menjadi kebutuhan yang tidak dapat diabaikan sebelum tahap pelatihan model [8].

Distribusi usia (panel tengah atas) memperlihatkan sebaran yang relatif merata pada rentang 13–19 tahun dengan frekuensi tertinggi pada usia 13 tahun (n=200) dan terendah pada usia 14 tahun (n=155), serta nilai rata-rata 15,93 ± 2,02 tahun. Keterwakilan yang baik di seluruh kelompok usia ini mengindikasikan bahwa model yang dihasilkan berpotensi menggeneralisasi ke seluruh populasi remaja usia sekolah, bukan hanya segmen tertentu.

Distribusi jam penggunaan media sosial berdasarkan kelas (panel kanan atas) mengungkap pola yang bermakna: secara proporsional, kasus depresi (batang merah) mulai muncul lebih dominan pada rentang 6–8 jam per hari, sementara kelas tidak depresi mendominasi pada rentang 1–5 jam. Pola ini konsisten dengan temuan M. Liu et al. [17] yang mengidentifikasi ambang tiga jam per hari sebagai titik kritis peningkatan risiko depresi pada remaja.

Boxplot distribusi jam tidur per kelas (panel kiri bawah) menampilkan perbedaan distribusi yang mencolok. Kelompok tidak depresi memiliki median 6,5 jam dengan rentang interkuartil (IQR) 5,5–7,5 jam, sedangkan kelompok depresi memiliki median yang jauh lebih rendah sebesar 4,8 jam dengan IQR lebih sempit 4,3–5,3 jam.

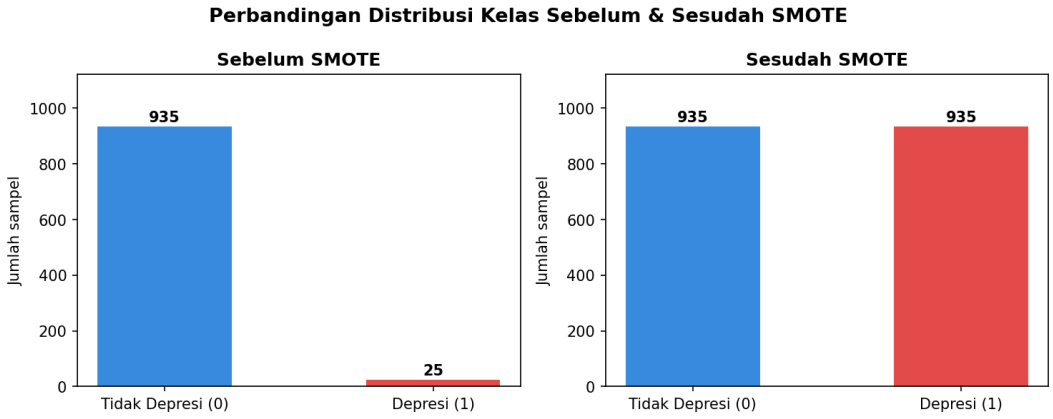
Perbedaan median sebesar 1,7 jam ini secara klinis bermakna dan dapat dijelaskan melalui mekanisme *neurobiologis*: kurang tidur mengaktifkan sumbu *hipotalamus hipofisis adrenal* yang meningkatkan *kortisol*, mengganggu regulasi *serotonin* dan *dopamin* yang berperan kunci dalam regulasi suasana hati [18].

Perbandingan rata-rata fitur per kelas (panel tengah bawah) memperkuat profil risiko yang teridentifikasi. Kelompok depresi menunjukkan rata-rata jam media sosial 49% lebih tinggi (6,7 vs 4,5 jam), tingkat stres 57% lebih tinggi (8,5 vs 5,4), dan tingkat kecemasan 54% lebih tinggi (8,6 vs 5,6), sementara rata-rata jam tidur 26% lebih rendah (4,8 vs 6,5 jam) dibandingkan kelompok tidak depresi.

Heatmap korelasi (panel kanan bawah) mengkuantifikasi hubungan antar variabel secara numerik. Variabel *depression* label memiliki korelasi negatif tertinggi dengan *sleep hours* (r=−0,19) dan korelasi positif terkuat dengan *daily social* media *hours* (r = +0,18), *stress* level (r = +0,17), serta *anxiety* level (r = +0,17). Menariknya, *screen time before sleep* berkorelasi negatif signifikan dengan *sleep hours* (r ≈ −0,15), mengindikasikan bahwa paparan layar sebelum tidur secara tidak langsung meningkatkan risiko depresi melalui penurunan durasi tidur. Sebaliknya, *academic performance* hampir tidak berkorelasi dengan *depression* label (r ≈ 0,001), menggugurkan asumsi umum bahwa tekanan akademik merupakan pemicu utama depresi pada *dataset* ini [18].

3.2. Pengaruh Penerapan *SMOTE* terhadap Keseimbangan Kelas

Gambar 3 membandingkan distribusi kelas pada data latih sebelum dan sesudah penerapan *SMOTE*. Perbedaan yang dihasilkan bersifat dramatis dan fundamental bagi kelayakan pelatihan model.



Gambar 3. Perbandingan distribusi kelas data latih sebelum (kiri) dan sesudah (kanan) penerapan SMOTE

Sebelum *SMOTE*, data latih mengandung 935 sampel kelas 0 dan hanya 25 sampel kelas 1, mempertahankan rasio ketidakseimbangan 37,4:1. Kondisi ini secara

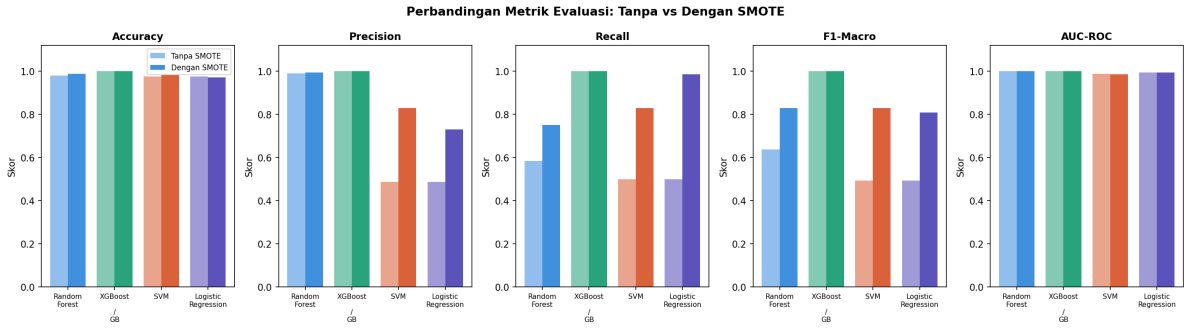
teoritis mendorong model untuk mempelajari *heuristik* sederhana: prediksi selalu kelas mayoritas, yang menghasilkan akurasi tinggi namun *recall* kelas minoritas mendekati nol. Setelah penerapan *SMOTE* dengan k=5, distribusi berubah menjadi keseimbangan sempurna 935:935 melalui sintesis 910 sampel minoritas baru. Perlu ditekankan bahwa *SMOTE* diterapkan secara eksklusif pada data latih; data uji (234:6) dibiarkan dalam distribusi alaminya untuk memastikan evaluasi yang tidak bias terhadap kondisi dunia nyata.

3.3. Perbandingan Performa Model

Tabel 2 merangkum hasil evaluasi delapan eksperimen, sedangkan Gambar 4 memvisualisasikan perbandingan seluruh metrik secara bersamaan untuk memudahkan interpretasi komparatif.

Tabel 2. Hasil perbandingan performa lengkap model dengan dan tanpa *SMOTE*

Model	Kondisi	*Accuracy*	*Precision*	*Recall*	*F1-Macro*	*AUC-ROC*
Random Forest	Tanpa *SMOTE*	0,9792	0,9895	0,5833	0,6376	1,0000
Random Forest	Dengan *SMOTE*	0,9875	0,9937	0,7500	0,8301	1,0000
XGBoost	Tanpa *SMOTE*	1,0000	1,0000	1,0000	1,0000	1,0000
XGBoost	Dengan *SMOTE*	1,0000	1,0000	1,0000	1,0000	1,0000
SVM	Tanpa *SMOTE*	0,9750	0,4875	0,5000	0,4937	0,9872
SVM	Dengan *SMOTE*	0,9833	0,8291	0,8291	0,8291	0,9865
Logistic Regression	Tanpa *SMOTE*	0,9750	0,4875	0,5000	0,4937	0,9936
Logistic Regression	Dengan *SMOTE*	0,9708	0,7308	0,9850	0,8082	0,9943



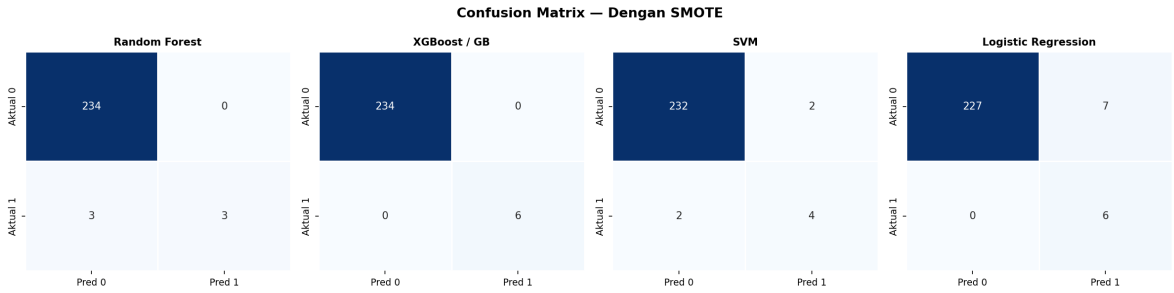
Gambar 4. Perbandingan lima metrik evaluasi seluruh model pada kondisi tanpa *SMOTE* (transparan) dan dengan *SMOTE* (*solid*)

Visualisasi pada Gambar 4 secara efektif mengekspos paradoks akurasi pada *dataset* tidak seimbang. *SVM* dan *Logistic Regression* tanpa *SMOTE* sama-sama mencatatkan akurasi 97,5% nilai yang tampak sangat kompetitif namun *F1-Macro* keduanya hanya 0,4937 dan *Recall* hanya 0,5000, mengindikasikan bahwa kedua model pada dasarnya tidak mampu mendeteksi kelas depresi secara andal. Sebaliknya, setelah *SMOTE* diterapkan, *F1-Macro SVM* meningkat ke 0,8291 (+67,9%) dan *Logistic Regression* ke 0,8082 (+63,7%), lonjakan yang mencerminkan transformasi fundamental kemampuan deteksi kelas minoritas.

XGBoost mencapai performa sempurna (seluruh metrik = 1,0000) pada kedua kondisi, menjadikannya algoritma terbaik dalam penelitian ini. Keunggulan ini dapat dikaitkan dengan tiga mekanisme: (1) pembelajaran sekuensial berbasis *gradient boosting* yang secara iteratif memperbaiki kesalahan prediksi; (2) regularisasi L1 dan L2 bawaan yang mencegah *overfitting*; dan (3) kemampuan inheren dalam menangkap interaksi nonlinear antar fitur yang kompleks, seperti hubungan *tripartit* tidur–stres–depresi. *Random Forest* menempati posisi kedua dengan *F1-Macro* 0,8301 (dengan *SMOTE*), konsisten dengan reputasinya sebagai metode *ensemble* yang *robust* [19].

3.4. Analisis *Confusion Matrix*

Gambar 5 menyajikan *confusion matrix* keempat model pada kondisi dengan *SMOTE*, memberikan gambaran rinci tentang tipe dan distribusi kesalahan prediksi masing-masing algoritma.



Gambar 5. *Confusion matrix* keempat model pada kondisi dengan *SMOTE* (TN=kiri-atas, FP=kanan-atas, FN=kiri-bawah, TP=kanan-bawah)

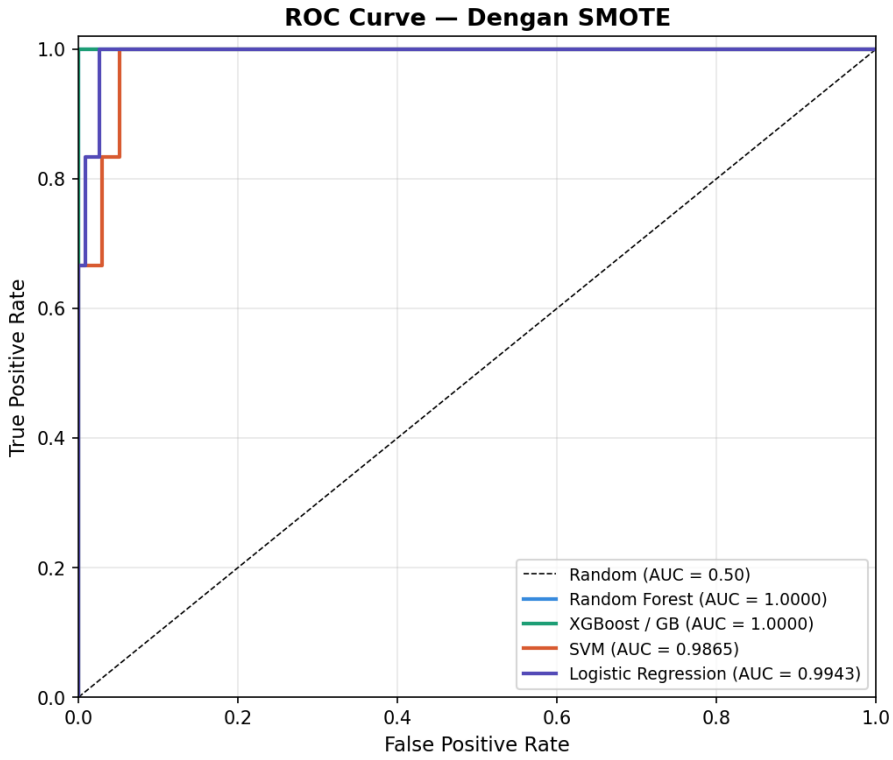
XGBoost menghasilkan prediksi sempurna dengan 234 *True Negative* (TN) dan 6 *True Positive* (TP), tanpa satu pun *False Positive* (FP) atau *False Negative* (FN). Dalam konteks klinis, ini berarti model berhasil mengidentifikasi seluruh 6 kasus depresi pada data uji sekaligus tidak menghasilkan alarm palsu pada kasus tidak depresi. *Random Forest* menghasilkan 234 TN, 3 TP, 0 FP, dan 3 FN model gagal mendeteksi

setengah kasus depresi (*Recall* = 0,50 untuk kelas positif), yang secara klinis berarti tiga remaja berisiko tidak teridentifikasi.

SVM menghasilkan 232 TN, 4 TP, 2 FP, dan 2 FN. Terdapat dua false positive (remaja sehat diprediksi depresi) dan dua false negative (kasus depresi terlewat). Profil ini menunjukkan model yang lebih seimbang namun masih melewatkan sepertiga kasus. *Logistic Regression* menghasilkan 227 TN, 6 TP, 7 FP, dan 0 FN profil paling menarik secara klinis: seluruh 6 kasus depresi berhasil teridentifikasi (*Recall* = 1,0), namun dengan tujuh *false positive*. Dalam skenario skrining kesehatan mental di sekolah, profil ini mungkin dapat diterima karena prioritas utama adalah tidak melewatkan kasus yang membutuhkan intervensi, meskipun menghasilkan beban verifikasi tambahan.

3.5. Analisis Kurva *ROC*

Gambar 6 menyajikan kurva *ROC* keempat model secara bersamaan pada kondisi dengan *SMOTE*. Kurva *ROC* memvisualisasikan *trade off* antara *True Positive Rate* (*Recall*/*Sensitivity*) dan *False Positive Rate* pada berbagai nilai *threshold* keputusan, sehingga memberikan perspektif evaluasi yang lebih menyeluruh dibandingkan metrik titik tunggal.

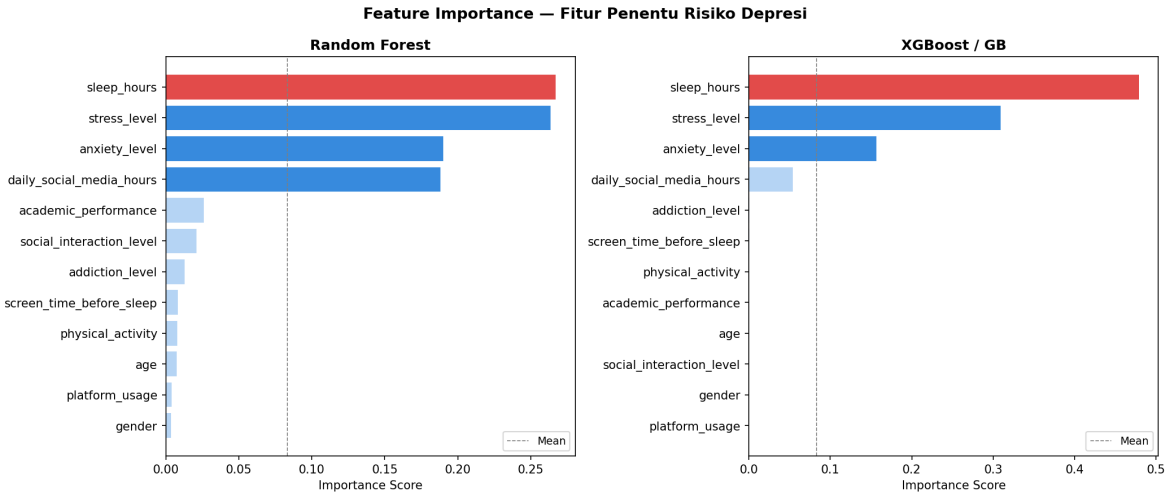


Gambar 6. Kurva *ROC* keempat model pada kondisi dengan *SMOTE*: *Random Forest* dan *XGBoost* mencapai *AUC* = 1,0000

Random Forest dan *XGBoost* mencapai *AUC-ROC* sempurna (1,0000), ditandai kurva yang membentuk sudut kanan atas yang tajam—mengindikasikan kemampuan membedakan kelas positif dan negatif secara sempurna di seluruh *threshold*. *Logistic Regression* mencapai *AUC-ROC* 0,9943, sedangkan *SVM* memperoleh 0,9865. Meskipun *AUC Logistic Regression* lebih tinggi dari *SVM*, F1-*Macro Logistic Regression* (0,8082) sedikit lebih rendah dari *SVM* (0,8291). Diskrepansi ini terjadi karena *AUC-ROC* mengukur kemampuan *ranking* model secara keseluruhan lintas seluruh *threshold*, sementara *F1-Macro* mengukur performa pada *threshold* keputusan spesifik (biasanya 0,5) [20]. Hal ini mengindikasikan bahwa optimasi *threshold Logistic Regression* berpotensi meningkatkan *F1-Macro*-nya secara substansial, melampaui *SVM*.

3.6. Analisis *Feature Importance*

Gambar 7 menyajikan hasil analisis *feature importance* dari *Random Forest* (kiri) dan *XGBoost* (kanan) pada kondisi dengan *SMOTE*. Analisis ini mengungkap kontribusi relatif setiap fitur dalam proses pengambilan keputusan model, memberikan dimensi interpretabilitas yang penting bagi aplikasi klinis.



Gambar 7. *Feature importance* dari *Random Forest* (kiri) dan *XGBoost* (kanan); merah = fitur dominan, biru tua = di atas rata-rata, biru muda = di bawah rata-rata

Kedua model memperlihatkan konsensus yang kuat tentang hirarki fitur. Pada *Random Forest*, empat fitur teratas adalah *sleep hours* (skor ≈ 0,27, ditandai merah), *stress level* (≈ 0,26), *anxiety level* (≈ 0,19), dan *daily social* media *hours* (≈ 0,19). Pada *XGBoost*, urutan identik dipertahankan: *sleep hours* mendominasi (≈ 0,47) jauh di atas fitur lainnya diikuti *stress level* (≈ 0,30), *anxiety evel* (≈ 0,13), dan *daily social* media *hours* (≈ 0,06). Konsistensi lintas dua paradigma algoritmik yang berbeda

secara fundamental *bagging versus boosting* memberikan validasi silang yang kuat terhadap peringkat fitur ini.

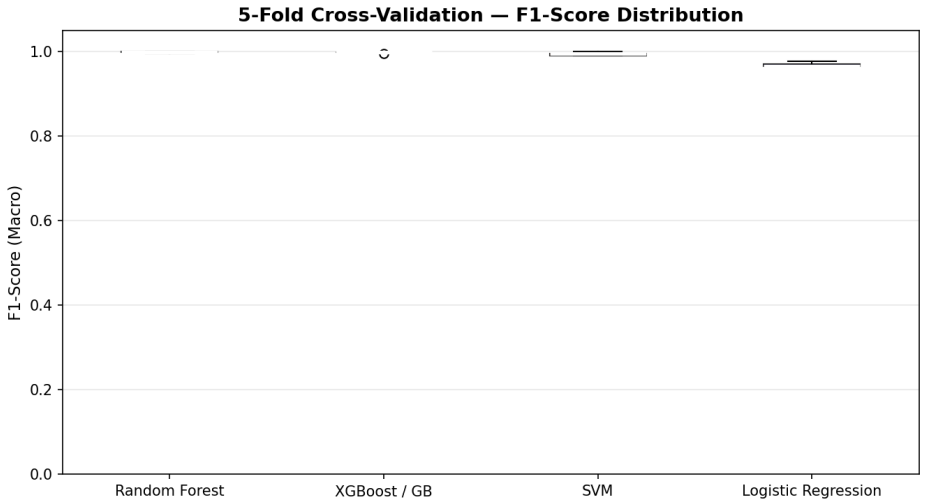
Dominasi *sleep hours* sebagai prediktor terkuat memiliki fondasi neurobiologis yang *solid*. Kurang tidur secara kronis mengaktifkan sumbu *hipotalamus hipofisis adrenal* (HPA axis), meningkatkan sekresi kortisol, dan mengganggu regulasi *neurotransmiter monoaminergik* (*serotonin, dopamin, norepinefrin*) yang berperan sentral dalam regulasi afeksi dan suasana hati. Pada *dataset* ini, kelompok depresi memiliki rata-rata tidur 4,8 jam di bawah rekomendasi minimum 8-10 jam untuk remaja menciptakan sinyal partisi yang sangat kuat bagi model berbasis pohon keputusan.

Tingkat stres dan kecemasan menempati posisi kedua dan ketiga, dengan rata-rata kelompok depresi hampir dua kali lipat kelompok tidak depresi. Hubungan ini bersifat resiprokal: stres dan kecemasan kronis dapat memicu episode depresif, sementara depresi memperburuk persepsi dan respons stres dalam siklus umpan balik positif yang sulit diputus tanpa intervensi eksternal.

Temuan paling signifikan dari analisis ini adalah kedudukan *academic performance* di posisi terbawah (skor mendekati nol, bahkan di bawah rata-rata pada model *XGBoost*). Temuan ini bertentangan dengan narasi umum yang menempatkan beban akademik sebagai katalis utama depresi remaja, dan memiliki implikasi praktis penting: program skrining dini depresi remaja akan lebih efektif jika difokuskan pada pemantauan pola tidur dan indikator stres-kecemasan, dibandingkan pemantauan kinerja akademik yang selama ini menjadi fokus utama perhatian orang tua dan pendidik.

3.7. Stabilitas Model melalui *Cross -Validation*

Gambar 8 menyajikan distribusi *F1-Score* pada 5-*fold cross validation* untuk keempat model dengan *SMOTE* dalam bentuk *boxplot*. Analisis ini bertujuan menilai konsistensi dan kemampuan generalisasi model di luar satu pembagian data tertentu, sehingga mengurangi risiko bias seleksi.



Gambar 8. Distribusi *F1-Score* (*Macro*) pada 5-*fold cross-validation* keempat model dengan *SMOTE*; semakin sempit *boxplot*, semakin stabil model

Tabel 3. Ringkasan hasil 5-*fold cross-validation* (*F1-Macro*) dengan *SMOTE*

Model	*Mean F1-Macro*	*Std. Dev.*	*F1 Min*	*F1 Maks*	Stabilitas
Random Forest	0,9975	0,0020	0,9945	0,9991	Sangat stabil
XGBoost	0,9969	0,0014	0,9945	0,9982	Paling konsisten
SVM	0,9936	0,0035	0,9900	0,9982	Stabil
Logistic Regression	0,9678	0,0080	0,9600	0,9782	Cukup stabil

Boxplot pada Gambar 8 secara visual mengonfirmasi stabilitas tinggi *Random Forest*, *XGBoost*, dan *SVM*: distribusi *F1-Score* ketiganya sangat sempit dan terkompresi mendekati nilai 1,0, menandakan performa yang konsisten lintas seluruh kombinasi data latih-uji. *Random Forest* mencapai *mean F1* tertinggi (0,9975 ± 0,0020), disusul *XGBoost* (0,9969 ± 0,0014). Standar deviasi *XGBoost* yang lebih kecil (0,0014 vs 0,0020) menunjukkan konsistensi yang lebih tinggi, menegaskan posisinya sebagai model yang paling reliabel.

Logistic Regression menunjukkan variabilitas yang lebih besar (std = 0,0080) dan *mean* lebih rendah (0,9678), meskipun masih tergolong stabil dengan rentang performa hanya 1,82 poin persentase. Variabilitas yang lebih besar ini mencerminkan sensitivitas model *linear* terhadap komposisi spesifik data pada setiap *fold*. Tidak ditemukannya tanda *overfitting* pada model terbaik *XGBoost* mencapai *F1* sempurna pada data uji (1,0000) sekaligus mempertahankan *CV mean* 0,9969 mengindikasikan bahwa model berhasil mempelajari pola yang dapat digeneralisasi, bukan sekadar menghafal data latih.

3.8. Pembahasan Terintegrasi

Hasil penelitian ini secara keseluruhan mengartikulasikan tiga kontribusi empiris utama. Pertama, *SMOTE* terbukti menjadi intervensi yang kritis dan tidak dapat diabaikan dalam pemodelan data kesehatan yang tidak seimbang. Tanpa *SMOTE*, *SVM* dan *Logistic Regression* dua dari empat algoritma yang diuji pada dasarnya tidak berguna untuk tujuan deteksi depresi meskipun akurasi keseluruhannya melampaui 97%. Ini menegaskan kembali bahwa akurasi tunggal adalah metrik yang tidak memadai untuk dataset dengan ketidakseimbangan kelas signifikan.

Kedua, *XGBoost* terbukti superior untuk klasifikasi risiko depresi berbasis data tabular multifitur. Performa sempurnanya yang konsisten divalidasi oleh *cross-validation* dengan *std* hanya 0,0014 menempatkan algoritma ini sebagai kandidat utama untuk implementasi sistem deteksi dini. Temuan ini selaras dengan konsensus empiris di komunitas *machine learning* bahwa *XGBoost* secara konsisten mengungguli algoritma lain pada data tabular terstruktur [13].

Ketiga, analisis *feature importance* menghasilkan temuan yang berimplikasi signifikan terhadap kebijakan kesehatan remaja. Dengan tidur sebagai prediktor terkuat dan prestasi akademik sebagai yang terlemah, strategi deteksi dini yang efektif semestinya memprioritaskan pemantauan pola tidur dan indikator stres kecemasan. Pendekatan ini secara teknis lebih mudah diimplementasikan pola tidur dapat dipantau secara pasif melalui perangkat *wearable* atau data penggunaan *smartphone* dibandingkan pengumpulan data akademik yang memerlukan akses institusional.

Dibandingkan penelitian terdahulu, kontribusi penelitian ini terletak pada pendekatan komparasi yang lebih sistematis. Yadav et al. [7] tidak menangani *imbalance* dan hanya melaporkan akurasi 78–92%. Mencapai AUC 0,70 dengan fitur terbatas dari Instagram. Penelitian ini, dengan 12 fitur multidimensi dan evaluasi sistematis *SMOTE*, menghasilkan *AUC-ROC* 0,9865 - 1,0000 dan memberikan analisis interpretabilitas yang absen dari studi studi sebelumnya. Keterbatasan utama adalah sifat sintetis dataset yang membatasi validitas eksternal; validasi pada data empiris populasi remaja Indonesia merupakan langkah esensial berikutnya

KESIMPULAN

Penelitian ini mengevaluasi secara sistematis empat algoritma *machine learning Random Forest*, *XGBoost*, *SVM*, dan *Logistic Regression* dalam mengklasifikasikan risiko depresi remaja berbasis perilaku digital, dengan penanganan ketidakseimbangan kelas menggunakan *SMOTE*.

Pertama, *XGBoost* terbukti sebagai algoritma terbaik dengan performa sempurna (*F1-Macro* = *AUC-ROC* = 1,0000) pada kedua kondisi, dikonfirmasi oleh *cross-validation* yang stabil (*mean CV* = 0,9969 ± 0,0014). Kedua, *SMOTE* secara konsisten

meningkatkan performa semua model, dengan dampak terbesar pada *SVM* (+67,9%) dan *Logistic Regression* (+63,7%), menegaskan bahwa penanganan *imbalanced* data merupakan prasyarat yang tidak dapat diabaikan. Ketiga, jam tidur merupakan prediktor terkuat risiko depresi, diikuti tingkat stres, kecemasan, dan penggunaan media sosial, sementara prestasi akademik hampir tidak berkontribusi.

Untuk penelitian lanjutan, disarankan: (1) validasi pada dataset empiris populasi remaja Indonesia; (2) eksplorasi teknik *oversampling* lanjutan seperti *ADASYN* dan *Borderline-SMOTE*; (3) integrasi fitur tekstual dari media sosial menggunakan *NLP*; dan (4) pengembangan prototipe sistem skrining berbasis model terbaik untuk lingkungan sekolah.

DAFTAR PUSTAKA

[1] W. H. Organization, "Depression." 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression
[2] K. K. RI, "Laporan Nasional Survei Kesehatan Indonesia (SKI) 2023," Badan Kebijakan Pembangunan Kesehatan, Jakarta, 2024.
[3] I-NAMHS, "Indonesia National Adolescent Mental Health Survey (I-NAMHS): Laporan Ringkas," Universitas Gadjah Mada, Yogyakarta, 2022.
[4] W. A. Social and Hootsuite, "Digital 2023: Indonesia." 2023. [Online]. Available: https://datareportal.com/reports/digital-2023-indonesia
[5] D. J. Yu, Y. K. Wing, T. M. H. Li, and N. Y. Chan, "The Impact of Social Media Use on Sleep and Mental Health in Youth: a Scoping Review.," *Curr. Psychiatry Rep.*, vol. 26, no. 3, pp. 104–119, Mar. 2024, doi: 10.1007/s11920-024-01481-9.
[6] D. Yuliani, "Jerat digital: Dampak cyberbullying pada kesehatan mental remaja di Indonesia," *Maliki Interdiscip. J. eISSN*, vol. 3, pp. 510–516, 2025, [Online]. Available: http://urj.uin-malang.ac.id/index.php/mij/index
[7] A. Yadav, S. Bhatt, and A. Kumar, "Machine learning approaches for mental health prediction: A systematic review," *IEEE Access*, vol. 10, pp. 41365–41385, 2022.
[8] U. K. J. Dn and M. Rahardi, "Analysis of SMOTE and Random Search on Machine Learning Algorithms for Stroke Disease Diagnosis," *J. Appl. Informatics Comput.*, vol. 10, no. 1 SE-Articles, pp. 847–855, Feb. 2026, doi: 10.30871/jaic.v10i1.12046.
[9] S. Sarwadi, R. Rosnelly, and B. Triandi, "Feature Selection Analysis for Diagnosing Narcissistic Personality Disorder (NPD) Using Principal Component Analysis and the Naïve Bayes Model," *ZERO J. Sains, Mat. dan Terap.*, vol. 9, no. 1, p. 14, 2025, doi: 10.30829/zero.v9i1.24086.
[10] A. Surya Firmansyah, A. Aziz, and M. Ahsan, "Optimasi K-Nearest Neighbor Menggunakan Algoritma Smote Untuk Mengatasi Imbalance Class Pada Klasifikasi Analisis Sentimen," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3341–3347, 2024, doi: 10.36040/jati.v7i6.7257.

[11] G. Almeida and F. Bacao, "Counterfactual synthetic minority oversampling technique: solving healthcare's imbalanced learning challenge," *Data Sci. Manag.*, vol. 8, no. 4, pp. 436–446, 2025, doi: https://doi.org/10.1016/j.dsm.2025.01.006.
[12] K. Redwan *et al.*, "Intelligent fault diagnosis in power systems using Random Forest and ensemble learning," *Meas.*, vol. 10, no. April, p. 100098, 2026, doi: 10.1016/j.meaene.2026.100098.
[13] M. M. Gharagoz, M. Noureldin, and J. Kim, "Explainable machine learning (XML) framework for seismic assessment of structures using Extreme Gradient Boosting (XGBoost)," *Eng. Struct.*, vol. 327, no. January, p. 119621, 2025, doi: 10.1016/j.engstruct.2025.119621.
[14] G. O. Anyanwu, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "RBF-SVM kernel-based model for detecting DDoS attacks in SDN integrated vehicular network," *Ad Hoc Networks*, vol. 140, p. 103026, 2023, doi: https://doi.org/10.1016/j.adhoc.2022.103026.
[15] L. Sammut, P. Bezzina, V. Gibbs, Y. Muscat-Baron, A. Agius-Camenzuli, and J. Calleja-Agius, "Predicting first-trimester pregnancy outcome in threatened miscarriage: A comparison of a multivariate logistic regression and machine learning models," *Radiography*, vol. 31, no. 6, p. 103159, 2025, doi: 10.1016/j.radi.2025.103159.
[16] J. Y. Verbakel *et al.*, "ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models," *J. Clin. Epidemiol.*, vol. 126, pp. 207–216, 2020, doi: https://doi.org/10.1016/j.jclinepi.2020.01.028.
[17] M. Liu, K. E. Kamper-DeMarco, J. Zhang, J. Xiao, D. Dong, and P. Xue, "Time Spent on Social Media and Risk of Depression in Adolescents: A Dose-Response Meta-Analysis.," *Int. J. Environ. Res. Public Health*, vol. 19, no. 9, Apr. 2022, doi: 10.3390/ijerph19095164.
[18] Z. Han, L. Wang, H. Zhu, Y. Tu, P. He, and B. Li, "Uncovering the effects and mechanisms of tea and its components on depression, anxiety, and sleep disorders: A comprehensive review," *Food Res. Int.*, vol. 197, p. 115191, 2024, doi: https://doi.org/10.1016/j.foodres.2024.115191.
[19] G. Schrijver, D. K. Sarmah, and M. El-hajj, "Automobile insurance fraud detection using data mining: A systematic literature review," *Intell. Syst. with Appl.*, vol. 21, no. November 2023, p. 200340, 2024, doi: 10.1016/j.iswa.2024.200340.
[20] I. Abunadi, "Machine learning techniques for analysing cardiotocography signals for early detection of fetal anomalies based on feature engineering methods," *Mach. Learn. with Appl.*, vol. 22, no. October, p. 100766, 2025, doi: 10.1016/j.mlwa.2025.100766.