

Prediksi Kesehatan Mental Remaja Berdasarkan Faktor Lingkungan Sekolah Menggunakan Machine Learning

Mutiara Rahma, Muhammad Fikry, Yesy Afrillia

Department of Informatic, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia

Info Artikel

Riwayat Artikel:

Received 2025-03-09

Revised 2025-04-18

Accepted 2025-05-06

Abstract – Adolescent mental health is a crucial aspect that significantly influences academic performance, social relationships, and overall well-being. The school environment is one of the primary factors affecting adolescents' mental conditions. This study developed a Random Forest-based prediction model for adolescent mental health using data from 229 students in Lhokseumawe City. The data were categorized into four mental health levels (no indication, level 1–3) based on HSCL-25 scores and school environmental factors, with the application of SMOTE to address class imbalance. The research methodology included data preprocessing, model training, and performance evaluation using standard accuracy metrics. The model achieved an accuracy of 80.43%, with F1-scores of 0.90 (no indication), 0.73 (level 1), 0.57 (level 2), and 0.00 (level 3), respectively. Feature importance analysis identified loneliness, feelings of worthlessness, academic pressure, and household stress as key influencing factors in prediction. Although the model classified most data accurately, misclassification persisted at certain mental disorder levels. This study contributes to the field of intelligent systems through the application of machine learning algorithms to holistically understand and map mental health conditions, particularly in developing intelligent decision support systems and e-health monitoring applications for adolescent mental well-being. Therefore, the Random Forest model can serve as an effective predictive tool for identifying adolescents at risk of mental health disorders. The findings of this study may serve as a foundation for educational institutions to design more targeted and adaptive mental health intervention strategies.

Keywords: Adolescent; Machine Learning; Mental Health; Prediction; Random Forest.

Corresponding Author:

Mutiara Rahma

Email:

mutiara.210170178@mhs.unimal.ac.id



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstrak – Kesehatan mental remaja adalah aspek krusial yang berpengaruh terhadap prestasi akademik, hubungan sosial, dan kesejahteraan secara keseluruhan. Lingkungan sekolah menjadi salah satu faktor utama yang dapat memengaruhi kondisi mental remaja. Penelitian ini mengembangkan model prediksi kesehatan mental remaja berbasis Random Forest menggunakan data dari 229 remaja di Kota Lhokseumawe. Data dikategorikan ke dalam empat kelas tingkat gangguan mental (tidak ada indikasi, level 1-3) berdasarkan skor HSCL-25 dan faktor lingkungan sekolah, dengan penerapan SMOTE untuk menangani ketidakseimbangan kelas. Metode penelitian mencakup preprocessing data, pelatihan model, serta evaluasi kinerja menggunakan metrik akurasi evaluasi. Hasil menunjukkan bahwa model mencapai akurasi sebesar 80.43%, dengan f1-score masing-masing sebesar 0.90 (tidak ada indikasi), 0.73 (level 1), 0.57 (level 2), dan 0.00 (level 3). Analisis feature importance mengidentifikasi bahwa perasaan kesepian, perasaan tidak berharga, tekanan akademik, dan stres di rumah merupakan faktor utama yang memengaruhi prediksi. Meskipun model mampu mengklasifikasikan sebagian besar data dengan baik, masih terdapat kesalahan klasifikasi pada beberapa tingkat gangguan mental. Penelitian ini memberikan kontribusi pada bidang sistem cerdas melalui pemanfaatan algoritma pembelajaran mesin untuk memahami dan memetakan kondisi mental secara holistik, khususnya dalam pengembangan sistem pendukung keputusan cerdas dan aplikasi monitoring e-health untuk kesehatan mental remaja. Dengan demikian, model Random Forest dapat digunakan sebagai alat bantu prediksi yang efektif dalam mendeteksi potensi gangguan mental remaja. Hasil penelitian ini dapat menjadi dasar bagi institusi pendidikan dalam merancang strategi intervensi yang lebih tepat sasaran.

Kata Kunci: Kesehatan Mental; Machine Learning; Prediksi; Random Forest; Remaja.

I. PENDAHULUAN

Kesehatan mental menjadi isu penting dalam perkembangan seorang remaja yang berdampak pada prestasi akademik, hubungan sosial, kualitas hidup, dan kesejahteraan secara keseluruhan. Remaja merupakan salah satu kelompok yang paling rentan mengalami gangguan kesehatan mental, terutama dalam hal mengendalikan keseimbangan emosi [1]. Selama masa remaja, seseorang menghadapi berbagai perubahan fisik, emosional, dan sosial yang dapat menyebabkan stres dan gangguan psikologis. Perubahan ini seringkali diperburuk oleh tekanan dari lingkungan termasuk tekanan akademis, dan interaksi sosial.

World Health Organization (WHO) menegaskan bahwa kesehatan mental menjadi bagian penting dari kesehatan dan kesejahteraan manusia [2]. Namun, data global menunjukkan tingginya angka gangguan kesehatan mental, dengan perkiraan 322 juta jiwa menderita depresi dan 264 juta mengalami gangguan kecemasan [3]. WHO juga memperkirakan 14% remaja mengalami gangguan kesehatan mental [4].

Di Indonesia masalah ini juga menjadi perhatian serius, data dari *National Adolescent Mental Health Survey* (I-NAMHS) yang dilakukan pada tahun 2022 menunjukkan bahwa 15,5 juta remaja Indonesia, memiliki masalah kesehatan mental dalam setahun terakhir. Ironisnya, hanya 2,6 persen remaja yang pernah mendapatkan dukungan emosional atau konseling [5]. Oleh karena itu, deteksi dini dan intervensi yang cepat sangat penting untuk mencegah dampak negatif yang lebih besar.

Lingkungan sekolah secara signifikan memengaruhi kesehatan mental dan kesejahteraan siswa [6]. Sejumlah faktor di lingkungan sekolah, termasuk dukungan kesehatan mental, hubungan teman sebaya, dan lingkungan sekolah, secara signifikan memengaruhi hasil kesehatan mental siswa [7]. Sebagai contoh, siswa yang merasa sekolah mereka ramah melaporkan hasil kesehatan mental yang lebih baik, tetapi siswa yang mengalami tekanan akademik yang besar biasanya melaporkan hasil yang lebih buruk. Selain itu, lingkungan sekolah, termasuk keamanan dan kualitas udara, memengaruhi kesehatan mental siswa [8]. Oleh karena itu, elemen-elemen yang ada di lingkungan sekolah dapat berfungsi sebagai pengukur yang signifikan terhadap kesehatan mental remaja.

Dengan kemajuan teknologi saat ini, metode *Machine Learning* dapat digunakan untuk menganalisis dan memprediksi [9] kondisi mental remaja berdasarkan berbagai faktor yang mempengaruhinya, serta untuk menemukan pola dan korelasi dalam data siswa tentang kesehatan mental [10]. Penelitian sebelumnya menunjukkan keunggulan *Random Forest*, salah satu algoritma *Machine Learning* dalam memprediksi kesehatan mental, seperti dalam penelitian [11] dalam klasifikasi data kesehatan mental di industri teknologi. Selain itu, penelitian [12] menunjukkan bahwa *Random Forest* mencapai akurasi tinggi dalam mendeteksi stres berdasarkan pola tidur dan fisiologis. Meskipun *Naive Bayes* dan *Decision Tree* untuk klasifikasi kesehatan mental remaja [13] hasilnya menunjukkan bahwa *Naive Bayes* lebih akurat sebesar 89,03% daripada *Decision Tree* sebesar 85,55%, algoritma ini memiliki keterbatasan karena asumsi independensi antar fitur.

Dalam penelitian ini, *Hopkins Symptom Checklist-25* (HSCL-25), sebuah alat laporan diri yang terdiri dari 25 item (10 item kecemasan dan 15 item depresi) dengan skala *Likert* 4 poin, digunakan untuk mengukur gejala depresi dan kecemasan pada remaja. HSCL-25 telah terbukti bermanfaat dalam menentukan individu yang membutuhkan tes psikologis tambahan dan telah divalidasi di berbagai demografi [14]. Studi terbaru menunjukkan bahwa HSCL-25 dapat digunakan untuk prediksi dalam menentukan remaja yang berisiko mengalami depresi [15]. Namun faktor lingkungan juga perlu untuk diidentifikasi agar mendapatkan pemahaman yang lebih mendalam.

Dalam konteks pendidikan, kesehatan mental yang buruk memengaruhi motivasi belajar remaja, yang berdampak signifikan pada keberhasilan akademik. Gangguan mental seperti stres pada siswa memengaruhi fungsi kognitif dan sosial, termasuk kemampuan untuk mendengarkan dan berbicara dengan orang lain [16]. Oleh karena itu, penelitian ini juga mengkaji hubungan antara motivasi belajar dan kesehatan mental.

Berdasarkan penelitian sebelumnya, metode *Random Forest* biasa digunakan dalam menangani masalah prediksi. *Random Forest* memiliki beberapa kelebihan dibandingkan dengan model yang hanya menggunakan satu pohon keputusan, seperti memiliki tingkat akurasi yang tinggi, mampu menangani *overfitting*, mampu menangani data yang berdimensi tinggi [17]. Namun, penelitian-penelitian tersebut masih berfokus pada populasi umum atau kelompok profesional tertentu, seperti karyawan teknologi. Penelitian yang secara khusus membahas remaja khususnya di wilayah Indonesia seperti Kota Lhokseumawe.

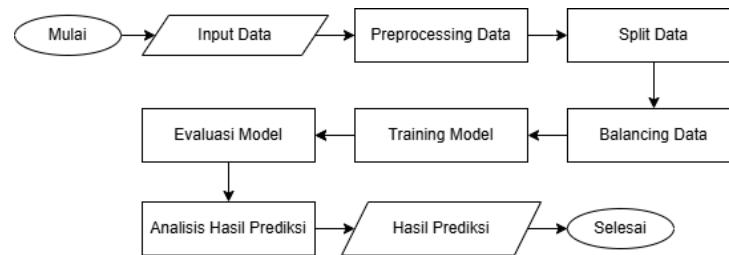
Penelitian ini bertujuan untuk mengembangkan sistem yang dapat memprediksi kondisi mental remaja menggunakan metode *Random Forest* dengan HSCL-25 dan motivasi belajar sebagai indikator penilaian. Selain itu, penelitian ini juga mengevaluasi performa model dengan metrik akurasi, presisi, *recall*, dan *F1-score*. Evaluasi tersebut dilakukan untuk memastikan keandalan dan konsistensi akurasi prediksi.

Kebaruan penelitian ini terletak pada integrasi indikator psikologis dan faktor lingkungan sekolah dalam sistem prediksi berbasis *Random Forest* yang dirancang khusus untuk memetakan kondisi kesehatan mental remaja. Dengan mempertimbangkan pengaruh lingkungan sekolah terhadap kesehatan mental serta korelasinya dengan motivasi belajar, penelitian ini juga memberikan kontribusi pada bidang sistem cerdas melalui pemanfaatan algoritma pembelajaran mesin untuk memahami dan memetakan kondisi mental secara holistik. Sistem ini diharapkan menjadi bagian dari solusi teknologi berbasis AI dalam upaya peningkatan layanan kesehatan mental yang adaptif, kontekstual, dan berbasis data, serta dapat diintegrasikan ke dalam layanan kesehatan masyarakat maupun program kesehatan mental di institusi pendidikan.

II. METODE

Metodologi kuantitatif digunakan bersama dengan pendekatan eksploratif dan prediktif dalam penelitian ini. Tujuan utamanya adalah untuk membuat model prediksi dengan algoritma *Random Forest* untuk kesehatan mental remaja dengan menggunakan data dari *Checklist Symptom Hopkins-25* (HSCL-25) dan faktor lingkungan sekolah. Penelitian ini memiliki beberapa tahapan sistematis seperti, pengumpulan data (*data input*), tahapan

preprocessing data, split data, training model dengan *Random Forest*, evaluasi model, serta analisis hasil dengan *feature importance*, yang ditampilkan pada Gambar 1.



Gambar 1. Alur penelitian

Pada Penelitian ini, Google Colab digunakan sebagai *platform* pengembangan berbasis *cloud* dengan bahasa pemrograman *Python* versi 3.11, dengan beberapa *library* seperti *scikit-learn* untuk implementasi model *Random Forest* dan juga penerapan teknik SMOTE. Selain itu pustaka *pandas* juga digunakan untuk manipulasi data, *seaborn* dan *matplotlib* untuk memvisualisasikan data, serta *numpy* untuk kalkulasi numerik.

A. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan secara langsung di lapangan dengan melibatkan siswa Sekolah Menengah Atas (SMA) di Kota Lhokseumawe. Metode purposive sampling digunakan untuk menentukan sampel, dengan mempertimbangkan keterwakilan siswa yang sesuai dengan kriteria penelitian. Sampel yang terlibat dalam penelitian ini berjumlah 229 siswa yang berasal dari SMA Negeri 5, SMA Negeri 6, dan SMA Negeri 7 di Kota Lhokseumawe. Responden dalam penelitian ini terdiri dari 64% perempuan dan 36% laki-laki, dengan rentang usia 15 hingga 18 tahun. Data dikumpulkan secara langsung di lingkungan sekolah dengan koordinasi pihak sekolah dan izin dari pihak terkait. Proses pengambilan data dilakukan dalam suasana yang kondusif agar responden dapat memberikan jawaban secara jujur dan objektif.

Selain mengumpulkan data dari siswa, penelitian ini juga mempertimbangkan faktor lingkungan sekolah yang dapat memengaruhi kondisi mental remaja. Informasi tambahan mengenai tekanan akademik, dukungan sosial, serta kondisi lingkungan sekolah dikumpulkan untuk memahami lebih dalam faktor-faktor yang berkontribusi terhadap kesehatan mental siswa.

Melalui instrumen *Checklist Symptom Hopkins-25* (HSCL-25) pengukuran gejala kecemasan dan depresi dapat dilakukan [18]. HSCL-25 dihitung dengan membagi skor total semua item dengan jumlah item dari siswa [19]. Secara sistematis dapat dirumuskan seperti pada Rumus (1):

$$S = \frac{\sum_{i=1}^N xi}{N} \quad (1)$$

Dimana S adalah skor rata-rata gejala kecemasan dan depresi yang diperoleh dari hasil HSCL-25, xi adalah skor per item, dan N adalah jumlah item yang dihasilkan oleh siswa.

Selain itu, faktor lingkungan sekolah merupakan variabel seperti hubungan dengan teman sebaya, suhu ruangan, suasana dan aktivitas dalam kelas, serta aksesibilitas sumber daya kesehatan mental juga dikumpulkan untuk tujuan meningkatkan analisis. Proses pengumpulan data pada penelitian ini termasuk mengajar enumerator untuk memahami cara mengisi formulir, membuat kuesioner dalam bentuk cetak dan digital, dan memvalidasi data untuk memastikan bahwa data akurat dan lengkap.

B. Data Preprocessing

Tahap *preprocessing* merupakan langkah penting dalam pengolahan data untuk memastikan kualitas data sebelum digunakan dalam pemodelan. Langkah pertama yang dilakukan adalah pembersihan data, yaitu mengidentifikasi dan menghapus data yang mengandung *outlier*, menangani nilai yang hilang, serta melakukan verifikasi terhadap kesesuaian data. Proses ini bertujuan untuk menghindari bias atau kesalahan yang dapat memengaruhi hasil prediksi model.

Selain itu, teknik *label encoding* digunakan untuk mengubah data kategorikal menjadi format numerik. Teknik tersebut dilakukan agar data dapat diolah secara efektif oleh algoritma machine learning. Selanjutnya, proses penerapan normalisasi data guna menyamakan skala pada setiap fitur agar model dapat bekerja secara optimal tanpa bias akibat rentang nilai antar variabel

C. Split Data

Tahapan ini dilakukan untuk validasi data menggunakan *train-test split*. Data dibagi menjadi dua kelompok, yaitu kelompok pelatihan (*training set*) dan kelompok uji (*testing set*), dengan perbandingan 80:20

untuk digunakan dalam pembangunan dan evaluasi model. *Split data* bertujuan untuk mencegah kesalahan data dan menghasilkan performa yang objektif terhadap kinerja model.

D. Balancing Data

Setelah tahapan *preprocessing* dan *split data* terdapat ketidakseimbangan data kelas, maka tahapan selanjutnya adalah penyeimbangan data yang dilakukan untuk mengatasi ketidakseimbangan dalam distribusi kelas target, dengan menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE). Teknik ini bekerja dengan mensintesis sampel baru dari kelas minoritas sehingga distribusi kelas menjadi lebih seimbang [20]. Dengan penerapan SMOTE, performa model diharapkan meningkat, terutama dalam mengklasifikasikan kategori dengan jumlah sampel yang lebih sedikit, sehingga akurasi prediksi dapat lebih optimal. Secara sistematis dapat dirumuskan dengan persamaan berikut:

$$X_{new} = X_i + \lambda (X_j - X_i) \quad (2)$$

Dimana X_i dan X_j adalah sampel minoritas yang dipilih secara acak dari tetangga terdekatnya, dan λ merupakan nilai acak antara 0 dan 1 yang menentukan interpolasi dua sampel minoritas. Namun, keterbatasan dataset bisa menjadi faktor yang memengaruhi kemampuan generalisasi model.

E. Training Model

Pada tahap ini, model prediksi dilatih menggunakan algoritma *Random Forest*, yang merupakan salah satu metode *machine learning* berbasis pohon keputusan. Algoritma ini dipilih karena kemampuannya untuk mengenali pola kompleks dalam data. Selain itu, *Random Forest* dapat mengatasi masalah *overfitting* yang sering terjadi pada model berbasis pohon tunggal dengan menggabungkan hasil dari beberapa pohon keputusan.

Proses pembangunan model diawali dengan pelatihan menggunakan data yang telah melalui tahap *preprocessing*. Data pelatihan digunakan untuk melatih model dalam mengenali pola serta hubungan antara fitur dengan variabel target. Dalam tahap ini, model akan membangun sejumlah pohon keputusan secara acak dan menggabungkan hasil prediksi dari masing-masing pohon untuk menghasilkan prediksi yang lebih akurat dan stabil. Keunggulan utama dari metode *ensemble Random Forest* adalah kemampuannya dalam meningkatkan akurasi serta ketahanan model terhadap variasi dalam data. Dengan menggabungkan prediksi dari banyak pohon keputusan, model menjadi lebih *robust* terhadap ketidakpastian [21], [22], [23] serta lebih adaptif dalam mengelola data yang memiliki banyak fitur. Dengan demikian, metode ini diharapkan mampu memberikan prediksi kesehatan mental remaja yang lebih akurat dan dapat diandalkan. Secara sistematis dapat dirumuskan sebagai berikut :

$$\hat{y} = \text{mode} \{T_1(x), T_2(x), \dots, T_n(x)\} \quad (3)$$

Dimana $\hat{y}$ merupakan hasil prediksi akhir dari model *Random Forest*, $T_1(x), T_2(x), \dots, T_n(x)$ mewakili hasil prediksi dari masing-masing pohon keputusan yang ada dalam *ensemble* model, serta fungsi *mode* digunakan untuk menentukan prediksi akhir berdasarkan hasil mayoritas dari semua pohon keputusan. Dalam kasus klasifikasi, *mode* akan memilih kelas yang paling sering diprediksi oleh pohon-pohon keputusan.

Algoritma *Random Forest* membangun banyak pohon keputusan, $(T_1(x), T_2(x), \dots, T_n(x))$ berdasarkan subset data yang dipilih secara acak (*bootstrap sampling*). Setiap pohon menghasilkan prediksi sendiri terhadap suatu data input x . Untuk data klasifikasi, hasil akhir diambil berdasarkan kelas yang paling sering dipilih oleh pohon-pohon keputusan (*mode*), kelas yang memiliki frekuensi tertinggi dari hasil prediksi pohon-pohon akan menjadi prediksi akhir ($\hat{y}$).

Untuk meningkatkan kinerja model, *GridSearchCV* digunakan untuk menemukan parameter terbaik. Parameter pencarian grid termasuk penggunaan *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *max_features*. Untuk menghindari *overfitting*, *GridSearchCV* menggunakan *cross-validation* (*cv*=5) untuk memilih parameter terbaik. Parameter yang dioptimalkan dalam tuning yaitu, *n_estimators* [50, 100, 200], *max_depth* [5, 10, 20, None], *min_samples_split* [2, 10, 15], *min_samples_leaf*: [1, 4, 5], *max_features* ['sqrt', 'log2'].

F. Evaluasi Model

Confusion Matrix digunakan untuk menilai akurasi model, dengan membandingkan nilai prediksi dan nilai aktual [24]. Evaluasi model dilakukan untuk memastikan performa prediksi yang dihasilkan. Evaluasi model dilakukan dengan menghitung metrik-metrik performa seperti akurasi, *precision*, *recall*, dan *F1-score*. Secara sistematis dapat dirumuskan pada persamaan berikut :

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{F-1 Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Dimana TP merupakan True Positive untuk jumlah prediksi positif yang benar, TN adalah True Negative, jumlah prediksi negatif yang benar, FP adalah False Positive untuk jumlah prediksi positif yang salah, dan FN merupakan False Negative untuk Jumlah prediksi negatif yang salah. Akurasi mengukur seberapa sering model memberikan prediksi yang benar dari seluruh data yang diuji. Semakin tinggi nilai akurasi, semakin baik performa model dalam mengklasifikasikan data. Namun, akurasi saja tidak cukup jika terdapat ketidakseimbangan kelas dalam data.

Presisi digunakan untuk mengukur seberapa andal model dalam memprediksi kelas positif, nilai presisi yang tinggi menunjukkan bahwa dari semua prediksi positif yang dibuat oleh model, sebagian besar benar, ini berguna dalam kasus di mana kesalahan positif harus diminimalkan. *Recall* mengukur seberapa baik model dalam mendeteksi semua kasus positif sebenarnya, nilai *recall* yang tinggi menunjukkan bahwa model mampu menangkap hampir semua sampel positif yang ada, *recall* penting dalam kasus di mana kesalahan negatif harus diminimalkan, seperti dalam deteksi dini penyakit atau ancaman keamanan. F1-Score merupakan rata-rata harmonik dari Precision dan Recall, F1-Score digunakan ketika terdapat ketidakseimbangan antara kelas positif dan negatif, karena memberikan keseimbangan antara kedua metrik tersebut, nilai F1-Score tinggi menunjukkan bahwa model memiliki presisi dan recall yang seimbang dalam membuat prediksi.

Metode evaluasi ini digunakan untuk menilai performa model *Random Forest* dalam memprediksi kesehatan mental remaja. Nilai yang lebih tinggi dalam metrik ini menunjukkan bahwa model dapat memberikan prediksi yang lebih akurat dan dapat diandalkan.

G. Analisis Hasil Prediksi

Pada langkah terakhir, analisis hasil prediksi dilakukan, di mana kondisi kesehatan mental remaja dikategorikan ke dalam empat kategori yaitu tidak ada indikasi (kelas 3), level 1 (kelas 0), level 2 (kelas 1), dan level 3 (kelas 2) seperti ditunjukkan pada Tabel 1. Selain itu, faktor-faktor lingkungan sekolah dianalisis menggunakan *feature importance*, untuk menentukan faktor yang memiliki dampak paling signifikan terhadap kesehatan mental remaja.

TABEL 1
KATEGORI PREDIKSI

Kelas	Interpretasi (kategori)
0	Level 1
1	Level 2
2	Level 3
3	Tidak Ada Indikasi

III. HASIL DAN PEMBAHASAN

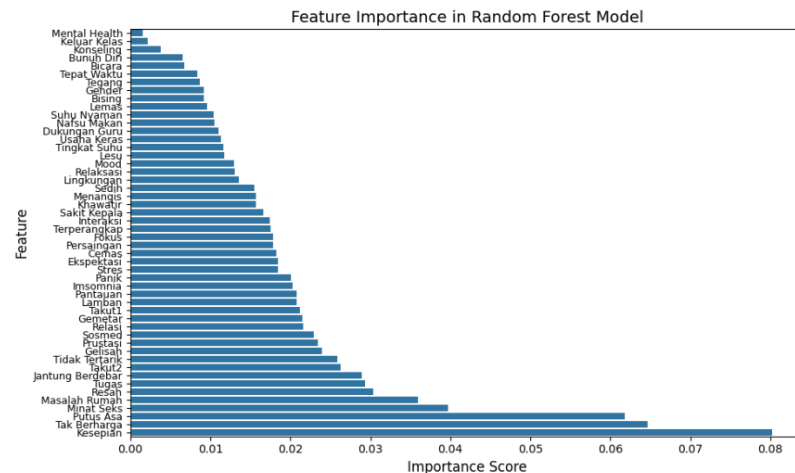
Pada penelitian ini, Model *Random Forest* yang digunakan untuk memprediksi kesehatan mental remaja berdasarkan faktor lingkungan sekolah, mencapai akurasi yang baik sebesar 80.43%. *Hyperparameter* terbaik yang ditemukan melalui proses tuning adalah *max_depth* = 10, *n_estimator* = 200, dan *random_state* = 42. *Hyperparameter* ini dipilih karena memiliki kemampuan untuk menangkap pola data yang kompleks sekaligus mengurangi kemungkinan *overfitting*.

Evaluasi menggunakan *classification report* menunjukkan bahwa setiap kelas memiliki nilai *F1-score*, *precision*, dan *recall* yang berbeda untuk tiap kelas. Dengan presisi 0.88, recall 0.92, dan F1-score 0.90, model ini menunjukkan performa yang luar biasa di kelas 3 (tidak ada indikasi). Hal ini menunjukkan bahwa kemampuan model untuk mengidentifikasi remaja tanpa masalah kesehatan mental hampir selalu akurat. Namun, presisi model sebesar 0.67 dan *recall* 0.80 pada kelas 0 (level 1) menunjukkan bahwa meskipun model ini cukup efektif dalam mengidentifikasi remaja dengan risiko rendah, masih terdapat beberapa kesalahan klasifikasi, terutama dalam membedakan kategori remaja ini dari kategori lainnya.

Kesalahan prediksi yang lebih besar terjadi pada kelas 1 (level 2), di mana presisi mencapai 1.00 tetapi recall hanya 0.40. Hal ini menunjukkan bahwa, meskipun model sangat akurat dalam memprediksi kasus kelas 1, hanya 40% dari kasus yang sebenarnya termasuk dalam kategori ini dapat diidentifikasi dengan benar. Hal ini menunjukkan bahwa model menghadapi masalah dalam mengidentifikasi remaja dengan risiko sedang. Pada

kelas 2 (level 3), meskipun presisi mencapai 1.00, tetapi *recall* 0.00 yang menunjukkan bahwa model tidak mampu memprediksi kategori ini dengan baik.

Berdasarkan hasil tersebut, penerapan teknik *balancing data* menggunakan SMOTE secara efektif dapat meningkatkan sensitivitas model terhadap kelas minoritas. Namun, ukuran dataset yang tidak seimbang menunjukkan bahwa SMOTE juga dapat meningkatkan risiko *overfitting*, seperti dibuktikan dengan ketidakseimbangan *recall* yang signifikan antar kelas, terutama pada kelas 2 (Level 3) yang tidak terdeteksi sama sekali. Hal ini dikarenakan pola yang dibuat dari data sintesis sangat mirip dengan data asli sehingga tidak dapat menunjukkan variasi data yang sebenarnya.



Gambar 2. Feature importance

Penjelasan pada gambar 2, bisa dilihat pada tabel 2 berikut:

TABEL 2
FITUR KESEHATAN MENTAL REMAJA

Label	Keterangan
Gender	Jenis Kelamin
Mental Health	Apakah Anda memiliki riwayat masalah kesehatan mental?
Tugas	Tugas sekolah menumpuk dan membuat saya stres
Ekspektasi	Saya merasa terbebani dengan ekspektasi orang tua
Relasi	Saya membangun berkomunikasi dengan teman sekelas
Dukungan Guru	Saya merasa didukung oleh guru-guru di sekolah
Lingkungan	Saya merasa lingkungan sekolah saya nyaman
Masalah Rumah	Saya merasa tertekan dengan masalah di rumah
Relaksasi	Saya memiliki waktu yang cukup untuk bersantai dan berolahraga
Sosmed	Saya merasa tertekan karena media sosial
Cemas	Saya merasa cemas mengenai masa depan saya
Persaingan	Saya merasa ada persaingan yang kuat di antara teman sekelas
Suhu Nyaman	Apakah Anda merasa suhu di kelas nyaman?
Tingkat Suhu	Seberapa sering Anda merasa kelas terlalu panas atau terlalu dingin
Tepat Waktu	Seberapa sering Anda datang tepat waktu ke kelas?
Keluar Kelas	Apakah Anda sering keluar-masuk kelas selama jam pelajaran?
Fokus	Seberapa sering Anda merasa sulit untuk tetap terjaga dan fokus selama pelajaran?
Interaksi	Seberapa sering Anda berbicara atau berinteraksi dengan teman sekelas selama jam pelajaran?
Bicara	Apakah Anda merasa nyaman berbicara di depan kelas?
Bising	Apakah Anda merasa suara di kelas terlalu bising sehingga mengganggu konsentrasi?
Lesu	Seberapa sering Anda merasa lelah atau lesu di kelas?
Stres	Apakah Anda merasa cemas atau stres selama jam pelajaran?
Mood	Seberapa sering Anda merasa suasana hati Anda buruk selama di kelas?
Konseling	Apakah Anda pernah melakukan konsultasi dengan guru konseling?
Pantauan	Seberapa sering guru konseling melakukan pemantauan kesehatan mental Anda?
Takut1	Tiba-Tiba Merasa Ketakutan Tanpa Sebab Yang Jelas
Takut2	Merasa Ketakutan.
Lemas	Limbung, Pening, Atau Lemas
Gelisah	Kegelisahan Atau Gemetar Di Dalam Diri.
Jantung Berdebar	Jantung Berdebar Kuat Atau Amat Cepat
Gemetar	Gemetaran
Tegang	Merasa Tegang Dan Terhimpit.

Sakit Kepala	Sakit Kepala.
Panik	Saat Merasa Amat Ketakutan Atau Panik.
Resah	Merasa Resah, Tidak Dapat Diam Tenang.
Lamban	Merasa Kurang Bertenaga, Melamban.
Frustasi	Menyalahkan Diri Sendiri Untuk Berbagai Hal.
Menangis	Mudah Menangis.
Minat Seks	Kehilangan Minat Atau Kesenangan Seksual (Tidak Mau Berada Di Antara Perempuan Atau Laki-Laki, Lawan Jenis).
Nafsu Makan	Selera Makan Terganggu (Berkurang).
Insomnia	Sulit Tidur, Mudah Terbangun/Terjaga.
Sedih	Merasa Sedih
Putus Asa	Merasa Tidak Punya Harapan Mengenai Masa Depan
Kesepian	Merasa Kesepian.
Bunuh Diri	Berpikir Untuk Mengakhiri Hidup Anda
Terperangkap	Merasa Terperangkap Atau Terjebak, Tidak Dapat Keluar Dari Suatu Situasi.
Khawatir	Terlalu Mengkhawatirkan Banyak Hal.
Tidak Tertarik	Merasa Tidak Tertarik Atau Tidak Berminat Terhadap Segala Hal.
Usaha Keras	Merasa Bahwa Segala Sesuatu Memerlukan Usaha Keras Atau Terasa Berat.
Tak Berharga	Merasa Tidak Berharga.

Selain itu, *feature importance* yang ditampilkan pada Gambar 2 menunjukkan bahwa fitur yang memiliki pengaruh signifikan dalam prediksi model adalah fitur “Kesepian”, “Tak Berharga”, dan “Putus Asa” yang merupakan instrumen dari HSCL-25 dengan *score* tertinggi. Hal ini menunjukkan bahwa perasaan kesepian, perasaan tidak berharga, dan tidak punya harapan untuk masa depan dapat menjadi gejala dari banyak masalah kesehatan mental, seperti depresi, kecemasan, isolasi sosial, dan kesulitan membangun hubungan dengan teman. Selanjutnya faktor lingkungan remaja seperti pada fitur “Masalah Rumah” dan “Tugas” juga menjadi fitur tertinggi selanjutnya. Remaja yang mengalami tekanan karena masalah di rumah cenderung membawa beban emosional ke sekolah, yang dapat mempengaruhi konsentrasi dan kesehatan mental. Tugas sekolah yang menumpuk juga dapat meningkatkan kecemasan dan kelelahan emosional, yang dapat memperburuk kondisi kesehatan mental. Hal ini menunjukkan bahwa lingkungan rumah dan tekanan akademik memiliki peran yang signifikan dalam menentukan seberapa rentan remaja terhadap masalah kesehatan mental.

Sebagai bagian dari evaluasi kinerja model, dilakukan perbandingan menggunakan algoritma *Decision Tree* tanpa teknik penyeimbangan data (SMOTE). Hasil evaluasi menunjukkan bahwa *Decision Tree* tanpa SMOTE hanya mampu mencapai akurasi 70%. Pada *classification report*, kelas 3 (tidak ada indikasi) memperoleh skor *F1-score* tertinggi sebesar 0.81, *recall* 0.84, *precision* 0.79. Sementara pada kelas 0 (level 1) memperoleh *recall* 0.50, *precision*, dan *F1-score* dengan nilai masing-masing 0.46 dan 0.48. Namun, pada kelas 1 (level 2) memperoleh *recall*, *precision*, dan *F1-score* dengan nilai 0.00, bahkan kelas 2 (level 3) tidak terdefinisi dalam hasil klasifikasi. Ini menunjukkan bahwa model baseline tidak dapat secara efektif mengidentifikasi kelas minoritas. Hal ini diperkuat oleh *confusion matrix*, yang menunjukkan bahwa sebagian besar prediksi model berpusat pada kelas mayoritas (kelas 3), sedangkan kelas minoritas sering salah diklasifikasikan. Model *Decision Tree* memiliki kelemahan yang signifikan dalam generalisasi dibandingkan dengan Random Forest yang dikombinasikan dengan SMOTE, terutama untuk distribusi data yang tidak seimbang.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa model Random Forest yang dikembangkan mampu memprediksi kondisi kesehatan mental remaja dengan tingkat akurasi yang baik. Dengan akurasi sebesar 80.43% serta evaluasi berdasarkan metrik presisi, recall, dan *F1-score*, model ini telah membuktikan kemampuannya dalam mengklasifikasikan kondisi kesehatan mental remaja berdasarkan faktor lingkungan sekolah dan hasil HSCL-25. Selain itu, analisis *feature importance* menegaskan bahwa faktor psikologis seperti perasaan kesepian dan ketidakberhargaan, serta faktor lingkungan seperti tekanan akademik dan masalah di rumah, memiliki peran signifikan dalam memengaruhi kesehatan mental remaja. Dengan demikian, penelitian ini berhasil mencapai tujuannya dalam mengembangkan sistem prediksi kesehatan mental remaja serta memberikan wawasan mengenai hubungan antara lingkungan sekolah, kesehatan mental, dan motivasi belajar. Temuan ini dapat menjadi dasar bagi pengembangan sistem deteksi dini yang dapat diintegrasikan dalam program kesehatan mental di sekolah maupun layanan kesehatan masyarakat lokal, seperti sistem monitoring kesehatan mental berbasis aplikasi. Sistem ini juga dapat diintegrasikan dalam bentuk web ataupun mobile yang dapat memberikan peringatan dini atau rekomendasi tindak lanjut kepada guru, konselor, maupun orang tua. Namun keterbatasan jumlah data juga menjadi tantangan utama dalam penelitian ini, yang mempengaruhi kestabilan dan generalisasi model. Hal ini menunjukkan perlunya penelitian lanjutan dengan jumlah sampel yang lebih besar dan representatif agar hasil model dapat diterapkan secara lebih luas dan akurat.

IV. SIMPULAN

Penelitian ini berhasil mengembangkan model prediksi kesehatan mental remaja menggunakan metode *Random Forest* dengan indikator HSCL-25 dan faktor lingkungan sekolah. Model *Random Forest* menunjukkan performa yang baik dengan akurasi mencapai 80.43%. Model sangat baik memprediksi kesehatan mental remaja di kelas 3 (tidak ada indikasi) dengan *F1-score* 0.90 dan *recall* 0.92. Namun, model mengalami kesulitan dalam memprediksi remaja pada kategori level 2 dan level 3, ini disebabkan kemiripan pola data dan ketidakseimbangan jumlah sampel antar kategori. Analisis *feature importance* menunjukkan bahwa perasaan kesepian, tidak berharga, dan kehilangan harapan untuk masa depan merupakan faktor utama yang memengaruhi prediksi model. Hal ini menunjukkan bahwa faktor psikologis paling banyak memengaruhi kesehatan mental remaja. Faktor lingkungan seperti tekanan akibat masalah di rumah dan stres akademik akibat tugas sekolah yang menumpuk juga berkontribusi secara signifikan. Ini menunjukkan bahwa lingkungan rumah dan tekanan akademik memiliki dampak yang signifikan terhadap kesejahteraan mental remaja, dimana tekanan dari keluarga dapat menyebabkan gangguan emosional, sementara beban akademik yang berlebihan berpotensi meningkatkan kecemasan dan kelelahan psikologis. Penelitian ini menggarisbawahi betapa pentingnya menjaga kesehatan mental remaja melalui lingkungan di rumah dan sekolah yang mendukung program intervensi yang dapat meningkatkan percaya diri dan motivasi belajar. Model prediksi yang dikembangkan dapat digunakan sebagai alat skrining awal untuk memprediksi remaja yang berisiko, terutama pada remaja yang menunjukkan gejala dengan tingkat ringan atau sedang. Karena itu, ada beberapa keterbatasan dalam penelitian ini, seperti ukuran sampel yang tidak seimbang terutama dalam memprediksi kesehatan mental pada kategori level 2 dan 3. Untuk meningkatkan akurasi dan generalisasi model, disarankan penelitian lanjutan dengan sampel yang lebih luas, data objektif, serta eksplorasi algoritma dan teknik *pre-processing* yang lebih canggih. Secara keseluruhan, penelitian ini memberikan kontribusi dalam bidang kesehatan mental remaja dengan mengintegrasikan pendekatan Machine Learning untuk prediksi dini, khususnya pengembangan sistem cerdas dan sistem pendukung keputusan untuk memantau kesehatan mental. Model prediksi yang dibangun dapat digunakan sebagai bagian penting dari aplikasi *e-health* berbasis web atau mobile yang mendukung intervensi berbasis data dan deteksi dini di lingkungan sekolah. Metode ini memungkinkan lembaga pendidikan untuk memanfaatkan teknologi informasi untuk memantau kondisi psikologis siswa dengan cara yang lebih sistematis, fleksibel, dan berkelanjutan.

DAFTAR PUSTAKA

- [1] Aloysius dan Salvia, "Analisis Kesehatan Mental Mahasiswa Perguruan Tinggi Pada Awal Terjangkitnya Covid-19 di Indonesia," *Jurnal Citizenship Virtues*, vol. 2021, no. 2, pp. 83–97.
- [2] who.int, "Mental health," who.int.
- [3] O. Shisana, D. J. Stein, N. P. Zungu, and G. Wolvaardt, "The rationale for South Africa to prioritise mental health care as a critical aspect of overall health care," *Compr Psychiatry*, vol. 130, Apr. 2024, doi: 10.1016/j.comppsy.2024.152458.
- [4] who.int, "Mental health of adolescents," who.int.
- [5] kemenpppa.go.id, "Kolaborasi KemenPPPA dan UNICEF: Sinergikan Dukungan Kesehatan Mental dan Psikososial di Kementerian dan Lembaga," kemenpppa.go.id.
- [6] M. Fikry, Bustami, and E. Suzanna, "IoT-Based Environmental Monitoring and Its Implications For Student Mental Health", *RGSA*, vol. 19, no. 2, p. e011183, Feb. 2025.
- [7] E. Long, C. Zucca, and H. Sweeting, "School Climate, Peer Relationships, and Adolescent Mental Health: A Social Ecological Perspective," *Youth Soc*, vol. 53, no. 8, pp. 1400–1415, Nov. 2021, doi: 10.1177/0044118X20970232.
- [8] U. Chourasia, S. Tendulkar, K. Gogia, N. Beerakayala, and K. Sumit, "Health Impact of School Built Environment on Children," *National Journal of Community Medicine*, vol. 14, no. 9, pp. 618–623, Sep. 2023, doi: 10.55489/njcm.140920233063.
- [9] M. Fikry and S. Inoue, "Optimizing Forecasted Activity Notifications with Reinforcement Learning," *Sensors*, vol. 23, no. 14, p. 6510, 2023.
- [10] N. Moningga, M. Hafidurrohman, and W. R. Ajri Tri, "Klasifikasi Mental Mahasiswa Menggunakan Metode Machine Learning," 2023. [Online]. Available: <https://www.kaggle.com/datasets/shariful07/student-mental->
- [11] E. Rosta Br Sebayang, Y. Herry Chrisnanto, U. Jenderal Achmad Yani Cimahi, J. Terusan Jend Sudirman, C. Selatan, and J. Barat, "Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest," 2023. [Online]. Available: <http://ijespjournal.org>
- [12] S. Sza et al., "Penerapan Decision Tree Dan Random Forest Dalam Deteksi Tingkat Stres Manusia Berdasarkan Kondisi Tidur", doi: 10.25126/jtiik.2024117993.
- [13] S. Oleh, "Klasifikasi Kesehatan Mental Usia Remaja Menggunakan Algoritma Decision Tree Dan Naïve Bayes."
- [14] S. Sinikumpu, J. Jokelainen, K. Tasanen, M. Timonen, and L. Huilaja, "Association of Rosacea With Depressive and Anxiety Symptoms: A General Population Study," *Dermatology*, vol. 240, no. 2, pp. 189–194, 2023, doi: 10.1159/000535034.
- [15] M. Rodríguez-Barragán et al., "Validation and psychometric properties of the spanish version of the hopkins symptom checklist-25 scale for depression detection in primary care," *Int J Environ Res Public Health*, vol. 18, no. 15, Aug. 2021, doi: 10.3390/ijerph18157843.
- [16] H. Dwi Putra, L. Khairani, D. Hastari, P. Studi Sistem Informasi, F. Sains dan Teknologi, and C. Author, "SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Comparison of Naive Bayes Classifier and Support Vector Machine Algorithms for Classifying Student Mental Health Data Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Data Kesehatan Mental Mahasiswa." [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>

- [17] E. S. Mohamed, T. A. Naqishbandi, S. A. C. Bukhari, I. Rauf, V. Sawrikar, and A. Hussain, "A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms," *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2023.100185.
- [18] D. Deasyanti and F. Muzdalifah, "Kesehatan mental mahasiswa ditinjau dari two continua model: Pengujian multiple analysis of variance," *Persona: Jurnal Psikologi Indonesia*, vol. 10, no. 1, pp. 147–166, Jun. 2021, doi: 10.30996/persona.v10i1.4660.
- [19] S. Park, P. Bangirana, E. Mupere, R. I. Baluku, E. S. Helgeson, and S. E. Cusick, "Association of COVID-19-related perceptions and experiences with depression and anxiety in Ugandan caregivers of young children with malaria and iron deficiency: A cross-sectional study," *PLoS One*, vol. 19, no. 12, Dec. 2024, doi: 10.1371/journal.pone.0314409.
- [20] R. Zhou *et al.*, "Prediction Model for Infectious Disease Health Literacy Based on Synthetic Minority Oversampling Technique Algorithm," *Comput Math Methods Med*, vol. 2022, 2022, doi: 10.1155/2022/8498159.
- [21] U. J. Ganai, S. Sachdev, and B. Bhushan, "Predictive modelling of stress, anxiety and depression: A network analysis and machine learning study," *British Journal of Clinical Psychology*, vol. 63, no. 4, pp. 522–542, Nov. 2024, doi: <https://doi.org/10.1111/bjc.12487>.
- [22] N. Acharya, P. Kar, M. Ally, and J. Soar, "Predicting Co-Occurring Mental Health and Substance Use Disorders in Women: An Automated Machine Learning Approach," *Applied Sciences (Switzerland)*, vol. 14, no. 4, Feb. 2024, doi: 10.3390/app14041630.
- [23] A. H. Januansyah, M. Fikry, and Y. Afrillia, "Machine Learning Algorithms Comparison for Gender Identification," p. 7, 2024, doi: 10.29103/micoms.v4.2024.
- [24] Jubeile Mark Baladjay, Nisce Riva, Ladine Ashley Santos, Dan Michael Cortez, Criselle Centeno, and Ariel Antwaun Rolando Sison, "Performance evaluation of random forest algorithm for automating classification of mathematics question items," *World Journal of Advanced Research and Reviews*, vol. 18, no. 2, pp. 034–043, May 2023, doi: 10.30574/wjarr.2023.18.2.0762.