

Sistem Deteksi Dini Gangguan Mental Menggunakan Algoritma Random Forest

Muhammad Ilham ‘Aziiz Alfarobi*, Romadona Romadona, Tariq Tariq, Aprilisa Arum Sari

Fakultas Ilmu Komputer, Sistem Informasi, Universitas Duta Bangsa, Kota Surakarta, Indonesia

Email: ¹*220101027@mhs.udb.ac.id, ²220101036@mhs.udb.ac.id, ³220101037@mhs.udb.ac.id,

⁴aprilisa_arumsari@udb.ac.id

Email Penulis Korespondensi: 220101027@mhs.udb.ac.id

Submitted 10-07-2025; Accepted 01-08-2025; Published 14-08-2025

Abstrak

Deteksi dini gangguan kesehatan mental merupakan tantangan signifikan dalam pelayanan kesehatan primer, seringkali terkendala oleh metode asesmen konvensional yang subjektif dan memakan waktu. Penelitian ini bertujuan merancang dan mengevaluasi sebuah prototipe sistem cerdas untuk prediksi risiko gangguan mental. Dengan mengadopsi kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM), penelitian ini memanfaatkan 1.000 set data rekam medis dari Klinik. Model prediksi dikembangkan menggunakan algoritma Random Forest, yang dikenal andal dalam menangani data kompleks. Hasil evaluasi menunjukkan kinerja model yang luar biasa, dengan akurasi terbobot (weighted accuracy) mencapai 99,67% pada data uji. Analisis feature importance mengonfirmasi bahwa variabel dukungan sosial, kualitas tidur, dan aktivitas fisik merupakan prediktor paling signifikan. Prototipe ini berhasil diimplementasikan sebagai aplikasi web interaktif menggunakan Streamlit, yang mendemonstrasikan kelayakan penggunaan machine learning sebagai alat bantu keputusan klinis yang cepat dan akurat untuk skrining kesehatan mental di tingkat layanan primer.

Kata Kunci: : Prediksi Kesehatan Mental; Random Forest; CRISP-DM; Machine Learning

Abstract

Early detection of mental health disorders poses a significant challenge in primary care, often hindered by conventional assessment methods that are subjective and time-consuming. This research aims to design and evaluate an intelligent system prototype for predicting mental health risks. Adopting the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, this study utilized 1,000 medical record datasets from Clinic. A predictive model was developed using the Random Forest algorithm, which is known for its robustness in handling complex data. Evaluation results indicate exceptional model performance, achieving a weighted accuracy of 99.67% on the test dataset. Feature importance analysis confirmed that social support, sleep quality, and physical activity variables are the most significant predictors. The prototype was successfully implemented as an interactive web application using Streamlit, demonstrating the feasibility of using machine learning as a rapid and accurate clinical decision support tool for mental health screening at the primary care level.

Keywords: Mental Health Prediction; Random Forest; CRISP-DM; Machine Learning; Decision Support System

1. PENDAHULUAN

Masalah kesehatan mental menjadi perhatian serius dalam ranah kesehatan publik, baik secara global maupun nasional. Berdasarkan laporan dari *World Health Organization* (WHO), sekitar 970 juta individu di dunia mengalami gangguan mental, termasuk depresi, kecemasan, bipolar, hingga skizofrenia [1]. Di Indonesia, data dari Riset Kesehatan Dasar (Riskesdas) 2018 menunjukkan bahwa 6,1% masyarakat mengalami gangguan mental emosional, sementara 0,18% mengalami gangguan jiwa berat seperti skizofrenia [2]. Meski secara persentase tampak kecil, angka ini memberikan dampak signifikan terhadap produktivitas masyarakat serta menambah beban pada sistem kesehatan nasional.

Salah satu hambatan utama dalam penanganan gangguan mental adalah keterlambatan dalam mendeteksi kondisi sejak dini. Gejala awal sering kali tidak kasatmata, subjektif, dan berbeda pada setiap individu [3]. Selain itu, masih kuatnya stigma sosial terhadap gangguan kejiwaan membuat banyak orang enggan mencari bantuan [4]. Di sisi lain, jumlah tenaga profesional di bidang kesehatan mental masih terbatas, sehingga upaya skrining secara masif belum berjalan optimal [5]. Oleh karena itu, diperlukan pendekatan yang inovatif dan efisien untuk membantu deteksi dini risiko gangguan mental, terutama yang bisa menjangkau masyarakat luas.

Salah satu pendekatan yang mulai berkembang adalah pemanfaatan teknologi kecerdasan buatan, khususnya algoritma *machine learning* (ML), untuk mengolah data psikologis dan melakukan prediksi terhadap potensi gangguan mental. Dari sekian banyak metode yang tersedia, algoritma *Random Forest* (RF) menjadi pilihan yang kuat karena kemampuannya mengklasifikasi data kompleks dan menghasilkan akurasi yang tinggi [6]. RF bekerja dengan membentuk banyak pohon keputusan (*decision trees*) yang kemudian digabungkan hasilnya melalui proses voting [7]. Kelebihan lain RF adalah kemampuannya dalam menangani data yang tidak seimbang, mengidentifikasi fitur yang paling berpengaruh, serta tahan terhadap nilai yang hilang (*missing values*) [8].

Dalam beberapa studi terdahulu, penerapan algoritma ML untuk mendeteksi kondisi kejiwaan telah menunjukkan hasil yang menjanjikan. Misalnya, Chandrasekaran et al. menggunakan *Support Vector Machine* (SVM) untuk mendeteksi depresi dari teks di media sosial, dengan akurasi hingga 85% [9]. Islam et al. memanfaatkan *Deep Neural Network* (DNN) pada data perilaku digital dan mencatat performa yang sebanding [10]. Wang et al. membangun model klasifikasi berbasis

XGBoost dan berhasil mengidentifikasi stres psikologis dengan AUC mencapai 0,91 [11]. Sementara itu, Lobo et al. menggunakan RF untuk mendeteksi gangguan bipolar berdasarkan data EEG, dengan hasil akurasi melebihi 90% [12].

Beberapa studi dari Indonesia juga mulai mengeksplorasi potensi ini. Sari dan Nugroho, membandingkan *Naïve Bayes* dengan *Decision Tree* untuk mendeteksi gangguan mental pada mahasiswa [13]. Walaupun algoritma *Decision Tree* mudah dipahami, akurasinya masih belum optimal. Penelitian Fathin dan Maharani membandingkan RF dan *Decision Tree* dalam mendeteksi depresi dari pola interaksi online, dan RF terbukti unggul dalam presisi dan *recall* [14]. Studi Midhula et al. [15] juga mendemonstrasikan keunggulan RF dalam klasifikasi gangguan mental dengan banyak label (*multi-label*) dan akurasi yang konsisten di atas 88%. Sayangnya, sebagian besar riset ini masih terbatas pada validasi model skala kecil dan belum memanfaatkan data kuesioner psikologis terstruktur, khususnya pada populasi Indonesia.

Lebih lanjut, penelitian oleh Maldini et al. menggunakan pendekatan *stacking method* untuk mendeteksi gangguan mental dari data sentimen dan menunjukkan peningkatan akurasi yang signifikan [16]. Teknik *ensemble* seperti ini memang menjanjikan, meskipun belum banyak diterapkan pada konteks lokal. Sharma et al. mengambil pendekatan berbeda dengan menggabungkan analisis perilaku dan suara untuk deteksi gangguan mental, dan berhasil mengembangkan model *multi-label classification* yang efektif [17].

Secara lokal, Priyono et al. menerapkan RF untuk mengklasifikasikan diagnosis gangguan mental menggunakan data kuesioner dari Indonesia [18]. Hasilnya menjanjikan, namun masih diperlukan validasi eksternal agar model ini bisa lebih diandalkan. Sebagai penguatan landasan teoretis, buku karya Doe dan Smith menjelaskan berbagai jenis gangguan mental berdasarkan standar DSM-5, yang menjadi acuan dalam praktik psikologi klinis [19].

Tak kalah penting, Universitas Gadjah Mada (UGM) melaporkan bahwa satu dari tiga remaja Indonesia mengalami gangguan mental [20]. Hal ini menunjukkan urgensi penerapan sistem deteksi dini yang akurat, khususnya pada kelompok usia produktif dan rentan seperti remaja.

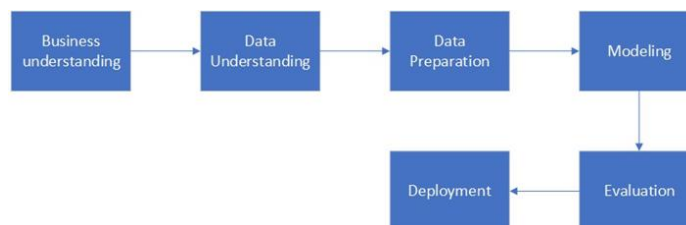
Letak kebaruan dari penelitian ini adalah, penelitian-penelitian sebelumnya dalam deteksi dini gangguan mental masih menghadapi berbagai keterbatasan, seperti penggunaan data tidak terstruktur yang tidak tervalidasi secara klinis, populasi yang sempit dan tidak representatif, serta minimnya pemanfaatan variabel gaya hidup sebagai prediktor utama. Selain itu, sebagian besar model yang digunakan, termasuk algoritma konvensional, menunjukkan akurasi yang belum optimal dan masih menghasilkan kesalahan klasifikasi yang tinggi. Untuk menjawab kekosongan tersebut, penelitian ini menggunakan algoritma Random Forest, yang dikenal andal dalam klasifikasi data kompleks, dengan memanfaatkan data klinis terstruktur dari 1.000 pasien di Klinik, serta mengembangkan prototipe aplikasi berbasis web untuk deteksi dini risiko gangguan mental secara cepat, akurat, dan dapat digunakan langsung di lingkungan layanan kesehatan primer.

Penelitian ini bertujuan untuk mengembangkan dan memvalidasi model prediktif berbasis *Random Forest* untuk mendeteksi risiko gangguan mental secara dini dengan memanfaatkan data kuesioner psikologis yang terstruktur. Diharapkan hasil dari penelitian ini dapat menjadi dasar bagi pengembangan sistem pendukung keputusan berbasis data (*data-driven decision support system*) yang sesuai dengan karakteristik masyarakat Indonesia. Dalam jangka panjang, pendekatan ini diharapkan mampu berkontribusi pada penanggulangan gangguan mental sejak tahap awal, melalui pemanfaatan teknologi yang adaptif, akurat, dan mudah diakses.

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Penelitian ini menggunakan pendekatan machine learning untuk membangun model prediktif risiko gangguan mental. Kerangka kerja yang digunakan adalah Cross-Industry Standard Process for Data Mining (CRISP-DM)



Gambar 1. Alur Penelitian

Gambar 1 menunjukkan alur CRISP-DM yang terdiri dari enam tahap utama yang saling terintegrasi, dimulai dari Business Understanding yang berfokus pada pemahaman tujuan dan permasalahan penelitian, yaitu dalam konteks ini mengidentifikasi kebutuhan akan sistem prediksi gangguan kesehatan mental berbasis web yang cepat dan mudah diakses menggunakan Streamlit. Selanjutnya tahap Data Understanding dilakukan untuk mengumpulkan dan mengeksplorasi data kesehatan mental guna memahami distribusi atribut, pola hubungan antar variabel, serta mengidentifikasi missing value dan outlier yang berpotensi mempengaruhi kualitas model. Setelah pemahaman data tercapai, tahap Data Preparation dilakukan untuk membersihkan data, melakukan encoding variabel kategorikal, serta membagi data menjadi data training dan testing untuk mempersiapkan proses modeling. Tahap Modeling kemudian dilaksanakan dengan menerapkan algoritma Random Forest pada data training untuk membuat model klasifikasi prediksi gangguan kesehatan mental,

termasuk proses tuning hyperparameter agar diperoleh performa model yang optimal. Tahap selanjutnya adalah Evaluation, yaitu mengevaluasi model dengan data testing menggunakan metrik seperti akurasi, precision, recall, dan f1-score untuk memastikan model memiliki generalisasi yang baik dan dapat digunakan secara nyata. Terakhir, tahap Deployment dilakukan dengan mengimplementasikan model ke dalam prototype aplikasi prediksi berbasis web menggunakan Streamlit sehingga dapat digunakan oleh pengguna akhir secara interaktif sebagai alat pendukung deteksi dini gangguan kesehatan mental.

2.2 Algoritma Random Forest

Random Forest merupakan pengembangan lebih lanjut dari algoritma Decision Tree. Pada metode ini, setiap Decision Tree dibangun menggunakan sampel data yang berbeda, dan pemilihan atribut dilakukan secara acak dari subset yang tersedia. Kelebihan Random Forest antara lain peningkatan akurasi meskipun terdapat missing data, ketahanan terhadap outliers, serta efisiensi dalam penggunaan memori. Selain itu, Random Forest secara otomatis melakukan seleksi fitur untuk mengidentifikasi atribut paling relevan, sehingga mampu meningkatkan performa model klasifikasi. Dengan fitur seleksi ini, Random Forest dapat menangani big data dengan parameter yang kompleks secara lebih efektif [21].

Untuk Klasifikasi (seperti penelitian kesehatan mental ini):

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_k(x)\} \quad (1)$$

Penjelasan

\hat{y} = prediksi akhir

$h_i(x)$ = prediksi pohon ke-i

K = jumlah total pohon

mode = nilai yang paling sering muncul (majority vote)

2.2 Tahap Business understanding

Tahap awal, Business Understanding, bertujuan pada identifikasi kebutuhan klinis akan sebuah alat prediksi yang akurat dan mudah diakses untuk deteksi dini gangguan mental. Data penelitian ini diperoleh dari rekam medis elektronik Klinik di Kota Tangerang, yang mencakup 1.000 data pasien pada tahun 2024. Data tersebut terdiri atas tujuh variabel utama yang relevan dengan kesehatan mental, yaitu: (1) usia pasien, (2) kualitas tidur (skala 1-10), (3) dukungan sosial (skala 1-10), (4) tingkat stres kerja (skala 1-10), (5) aktivitas fisik (jam/minggu), (6) riwayat trauma (kategorikal), dan (7) status kesehatan mental (Baik/Stres/Depresi).

2.3 Tahap Data understanding

Tahap Data Understanding, data yang digunakan berasal dari rekam medis elektronik Klinik XYZ, mencakup 1.000 entri pasien dengan variabel seperti usia, kualitas tidur, dukungan sosial, tingkat stres, aktivitas fisik, riwayat trauma, dan status kesehatan mental. Tahap ini melibatkan eksplorasi data untuk memastikan kelengkapan, konsistensi, dan relevansi variabel. Hasilnya menunjukkan tidak ada missing values atau outlier yang signifikan.

2.4 Tahap Data preparation

Tahap Data Preparation merupakan langkah krusial yang mencakup beberapa proses. Pertama, data dibersihkan untuk memastikan tidak ada nilai yang hilang. Selanjutnya, variabel target "Status kesehatan mental" ditransformasikan menjadi tiga label risiko: "Rendah", "Sedang", dan "Tinggi". Variabel kategorikal "Riwayat trauma" juga diubah menjadi representasi numerik (1 untuk "Ya" dan 0 untuk "Tidak") melalui proses encoding. Setelah data siap, dataset dibagi dengan Untuk keperluan pemodelan, data di partisi menjadi dua set dengan proporsi 70% dialokasikan sebagai data latih dan 30% sebagai data uji.

2.5 Tahap Modeling

Pada tahap Modeling, algoritma Random Forest dipilih sebagai pembangun model utama. Algoritma ini merupakan manifestasi dari pendekatan ensemble learning, di mana performa prediktif ditingkatkan melalui sintesis hasil dari banyak prediktor lemah (dalam hal ini, pohon keputusan). Proses agregasi ini menghasilkan sebuah model final dengan akurasi superior serta varians yang lebih rendah, yang mencerminkan stabilitas model yang tinggi. Model yang telah dilatih kemudian dievaluasi kinerjanya pada tahap Evaluation menggunakan serangkaian metrik standar, termasuk akurasi, precision, recall, dan F1-score. Analisis confusion matrix juga digunakan untuk memeriksa kesalahan klasifikasi secara mendetail, sementara feature importance dianalisis untuk mengidentifikasi prediktor paling dominan.

2.6 Tahap Evaluation

Model yang telah dilatih kemudian dievaluasi kinerjanya pada tahap Evaluation menggunakan serangkaian metrik standar, termasuk akurasi, precision, recall, dan F1-score. Analisis confusion matrix juga digunakan untuk memeriksa kesalahan klasifikasi secara mendetail, sementara feature importance dianalisis untuk mengidentifikasi prediktor paling dominan.

Dalam penelitian ini, evaluasi model dilakukan menggunakan beberapa metrik yang umum digunakan dalam klasifikasi biner. Berikut adalah persamaan yang diterapkan:

a. Akurasi

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

b. Presisi

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

c. Recall

$$Recal = \frac{TP}{TP + FN} \quad (4)$$

d. F1-score

$$F1\ Score = 2\chi \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Penjelasan:

TP (True Positive): Prediksi benar kelas positif.

TN (True Negative): Prediksi benar kelas negatif.

FP (False Positive): Prediksi salah kelas positif.

FN (False Negative): Prediksi salah kelas negatif.

2.7 Tahap Deployment

Tahap akhir adalah Deployment, di mana model yang telah tervalidasi dan terbukti andal dikemas dalam format pickle (.pkl). Untuk keperluan deployment dan validasi, model prediktif yang dihasilkan ditanamkan ke dalam sebuah platform aplikasi web. Platform yang dikonstruksi menggunakan framework Streamlit ini menjadi wujud akhir sebuah prototipe fungsional yang dirancang untuk dievaluasi oleh pengguna dalam skenario uji coba interaktif.

3. HASIL DAN PEMBAHASAN

3.1 Tahapan Penerapan Metode

Penelitian ini menerapkan algoritma Random Forest melalui enam tahapan kunci yang mengacu pada kerangka CRISP-DM.

- Tahap business understanding mengidentifikasi kebutuhan klinis akan alat deteksi dini gangguan mental yang cepat dan objektif.
- Pada tahap data understanding, data 1.000 rekam medis dari Klinik XYZ dieksplorasi untuk memastikan kelengkapan dan relevansi fitur seperti usia, kualitas tidur, dan dukungan sosial.
- Tahap data preparation meliputi transformasi data, termasuk encoding variabel kategorikal (misalnya, "Riwayat trauma" menjadi 1/0) dan labeling variabel target menjadi tiga kelas risiko ("Rendah", "Sedang", "Tinggi"), serta pembagian dataset dengan rasio 70:30 untuk data latihan dan uji.
- Tahap modeling melibatkan pelatihan Random Forest dengan 100 decision tree. Setiap tree dibangun menggunakan teknik bootstrap sampling (pengambilan sampel acak dengan pengembalian) dan seleksi fitur acak untuk tiap split, dengan kriteria pemisahan berbasis Gini Index untuk memaksimalkan kemurnian node.
- Tahap evaluation menguji model menggunakan metrik weighted accuracy (99.67%), precision, recall, dan confusion matrix, yang menunjukkan hanya satu kesalahan klasifikasi pada data uji. Analisis feature importance mengungkap dominasi dukungan sosial (31.1%) dan kualitas tidur (26.2%) sebagai prediktor utama.
- Terakhir, tahap deployment mengintegrasikan model ke aplikasi web berbasis Streamlit untuk prediksi interaktif, memungkinkan tenaga medis memasukkan parameter pasien dan menerima hasil klasifikasi risiko secara real-time.

3.2 Hasil

Penelitian ini dilatarbelakangi oleh urgensi untuk meningkatkan efisiensi dan akurasi dalam deteksi dini gangguan kesehatan mental di lingkungan klinis, secara spesifik di Klinik. Menjawab kebutuhan tersebut, tujuan primer dari studi ini adalah mengembangkan sebuah model prediktif berbasis algoritma Random Forest. Model ini dirancang untuk melakukan klasifikasi tingkat risiko gangguan mental pada pasien secara sistematis. Untuk merealisasikan tujuan ini, penelitian memanfaatkan himpunan data (dataset) yang terdiri dari 1.000 entri pasien. Dataset tersebut mencakup atribut prediktif yang relevan, meliputi usia, kualitas tidur, tingkat dukungan sosial, stres kerja, aktivitas fisik, serta riwayat trauma. Analisis data pada tahap awal, yaitu pemahaman data (data understanding), telah memverifikasi bahwa integritas data terjaga dengan baik dan tidak ditemukan adanya nilai yang hilang (missing values), sehingga data dinyatakan valid dan siap untuk tahap pemodelan lebih lanjut.

Pada tahap data preparation, dilakukan dua langkah transformasi utama yaitu pertama, data labeling, di mana nilai pada variabel target "Status kesehatan mental" diubah menjadi tiga label risiko yang lebih sesuai untuk klasifikasi, yaitu "Rendah", "Sedang", dan "Tinggi". Kedua, proses encoding diterapkan pada variabel kategorikal "Riwayat trauma" untuk

mengubahnya menjadi nilai numerik (1 dan 0) menggunakan LabelEncoder. Setelah data sepenuhnya siap, penelitian berlanjut ke tahap modeling, Seperti yang divisualisasikan pada Gambar 2, dataset penelitian dipecah menjadi dua bagian utama: 70% dialokasikan untuk melatih model Random Forest (set pelatihan), dan 30% sisanya digunakan untuk menguji kinerjanya (set pengujian).

```

Akurasi Model: 1.00

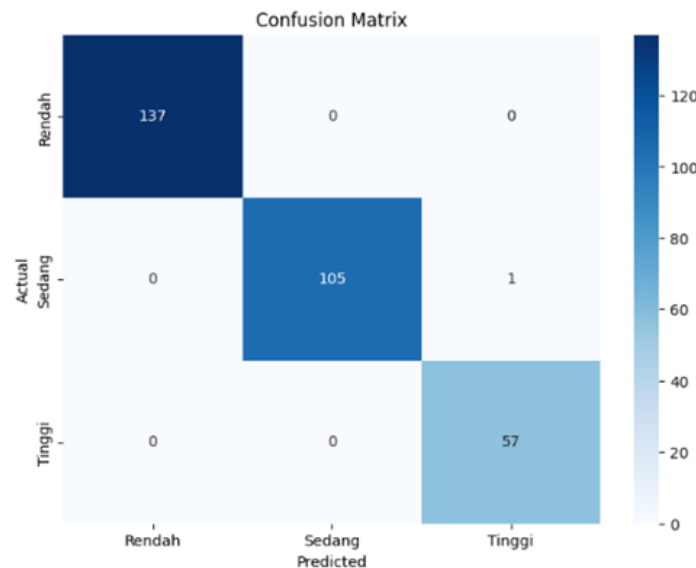
Classification Report:
              precision    recall  f1-score   support

   Rendah      1.0000      1.0000      1.0000       137
   Sedang      1.0000      0.9906      0.9953       106
   Tinggi      0.9828      1.0000      0.9913        57

 accuracy      0.9967      0.9967      0.9967       300
 macro avg      0.9943      0.9969      0.9955       300
 weighted avg      0.9967      0.9967      0.9967       300
  
```

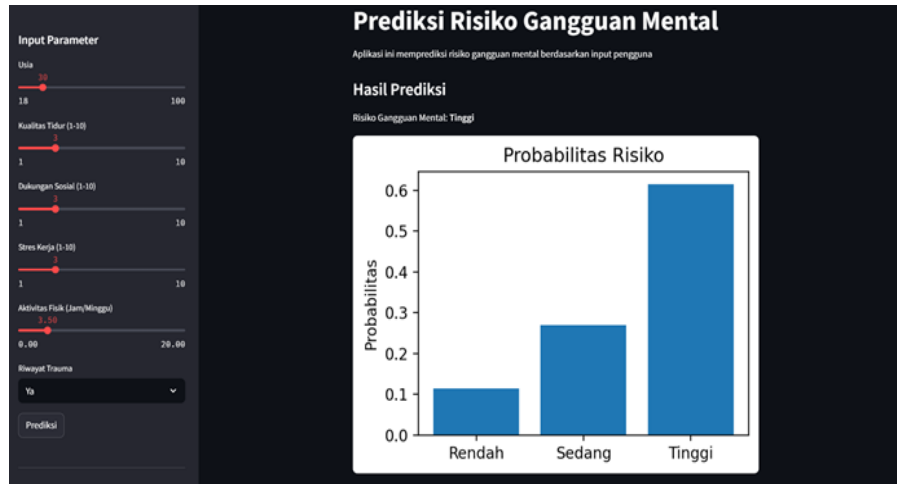
Gambar 2. Classification Report

Gambar 2 menunjukkan hasil evaluasi kuantitatif terhadap kinerja model prediksi, sebagaimana dirinci dalam Classification Report, mendemonstrasikan performa yang luar biasa pada 300 sampel data uji. Secara keseluruhan, model mencapai akurasi terbobot (weighted average) sebesar 0,9967, yang mengindikasikan kemampuan klasifikasi yang sangat tinggi dan andal. Analisis lebih lanjut pada setiap kelas menunjukkan kinerja yang solid dan seimbang: kategori 'Rendah' berhasil diidentifikasi dengan skor presisi, recall, dan F1-score yang optimal (1,0000). Sementara itu, kategori 'Tinggi' juga mencapai recall sempurna (1,0000) dengan presisi 0,9828, dan kategori 'Sedang' menunjukkan performa kuat dengan recall 0,9906. Tingginya nilai macro average (0,9955) yang mendekati weighted average (0,9967) mengonfirmasi bahwa model ini memiliki kemampuan generalisasi yang sangat baik di semua kelas, tanpa menunjukkan adanya bias yang signifikan terhadap kategori tertentu. Kinerja yang konsisten di seluruh metrik ini membuktikan efektivitas algoritma Random Forest dan pendekatan metodologis yang diterapkan dalam penelitian ini.



Gambar 3. Confusion Matrix

Gambar 3 menunjukkan analisis terhadap *Confusion Matrix* memberikan validasi lebih lanjut atas superioritas kinerja model. Matriks tersebut secara visual mengonfirmasi bahwa model berhasil mengklasifikasikan seluruh 137 sampel aktual "Rendah" dan 57 sampel aktual "Tinggi" secara sempurna, tanpa ada kesalahan klasifikasi sama sekali. Tingkat kesalahan (*misclassification rate*) model terbukti sangat rendah, dengan hanya satu dari 106 sampel "Sedang" yang keliru diklasifikasikan sebagai "Tinggi". Hasil ini secara kuantitatif menegaskan bahwa model prediksi yang dikembangkan sangat andal dengan tingkat kepercayaan yang tinggi. Setelah melalui tahap evaluasi yang komprehensif, model *Random Forest* yang telah tervalidasi kemudian memasuki tahap *deployment*, di mana model disimpan dalam format file pickle untuk portabilitas dan diintegrasikan ke dalam sebuah prototipe aplikasi berbasis web menggunakan *framework* Streamlit untuk prediksi interaktif berdasarkan input pengguna.



Gambar 4. Web Streamlit

Gambar 4 menunjukkan aplikasi web yang dikembangkan menyediakan antarmuka pengguna yang interaktif, di mana pengguna dapat memasukkan serangkaian parameter input yang relevan dengan kondisi kesehatan mental. Parameter-parameter ini mencakup: (a) Usia pada rentang 18 hingga 100 tahun, (b) Kualitas Tidur dalam skala 1-10, (c) Dukungan Sosial dalam skala 1-10, (d) Tingkat Stres dalam skala 1-10, (e) Aktivitas Fisik yang diukur dalam jam per minggu, dan (f) Riwayat Trauma dengan pilihan kategorikal "Ya" atau "Tidak". Setelah pengguna mengkonfirmasi input dengan menekan tombol "Prediksi", sistem akan memproses data tersebut menggunakan model *Random Forest* yang telah terintegrasi. Hasil prediksi kemudian disajikan kepada pengguna dalam dua format: pertama, sebagai output klasifikasi risiko secara langsung (misalnya, "Tinggi"), dan kedua, sebagai representasi visual berupa grafik probabilitas. Grafik ini menampilkan distribusi probabilitas risiko dalam tiga kategori (Rendah, Sedang, dan Tinggi), yang memungkinkan pengguna memperoleh pemahaman yang lebih komprehensif mengenai tingkat resikonya.

3.3 Pembahasan

Penelitian ini menggunakan 1000 dataset yang bersumber dari rekam medis di Klinik XYZ, mencakup atribut-atribut relevan seperti usia, kualitas tidur, dukungan sosial, tingkat stres kerja, aktivitas fisik, dan riwayat trauma, dengan status kesehatan mental sebagai variabel target. Sebelum pemodelan, data melewati tahap pra-pemrosesan yang meliputi transformasi variabel target menjadi tiga label risiko ('Rendah', 'Sedang', 'Tinggi') dan pengkodean variabel kategorikal 'Riwayat trauma' menjadi format numerik. Analisis awal menunjukkan dataset dalam kondisi lengkap tanpa nilai yang hilang, sehingga tidak memerlukan teknik imputasi atau *oversampling* dan dapat langsung digunakan untuk melatih serta menguji model *Random Forest* setelah dibagi dengan rasio 70:30.

Hasil klasifikasi menunjukkan bahwa model *Random Forest* yang dikembangkan memiliki kinerja yang luar biasa dalam memprediksi risiko gangguan mental. Model ini mencapai akurasi terbobot (*weighted avg*) sebesar 99,67% pada data uji. Kinerja ini terbukti superior di semua metrik, dengan skor presisi, recall, dan F1-score yang mendekati sempurna untuk semua kelas. Secara spesifik, model mampu mengklasifikasikan kategori 'Rendah' dan 'Tinggi' dengan recall sempurna (1,0000), yang berarti tidak ada satupun kasus berisiko rendah atau tinggi yang terlewatkan. Tingkat kesalahan klasifikasi (*misclassification*) sangat rendah, dengan hanya satu kasus 'Sedang' yang keliru diprediksi sebagai 'Tinggi'. Variabel yang paling berpengaruh dalam prediksi ini, berdasarkan analisis *feature importance*, adalah Dukungan Sosial (31,1%), Kualitas Tidur (26,2%), dan Aktivitas Fisik (25,3%). Faktor-faktor ini secara signifikan lebih dominan dibandingkan usia atau riwayat trauma, yang menunjukkan bahwa kondisi gaya hidup dan lingkungan sosial menjadi penentu utama dalam model ini.

Berdasarkan hasil analisis, superioritas model *Random Forest* dalam penelitian ini dapat dijelaskan oleh kemampuannya menangkap interaksi non-linear yang kompleks antara variabel-variabel sosio-perilaku seperti dukungan sosial dan kualitas tidur. Tidak seperti model lain yang mungkin kesulitan dengan hubungan data yang rumit, pendekatan *ensemble* pada *Random Forest* yang membangun banyak pohon keputusan terbukti sangat efektif dalam memetakan pola risiko kesehatan mental. Dominasi faktor dukungan sosial dan kualitas tidur menegaskan bahwa model tidak hanya akurat secara teknis, tetapi juga selaras dengan teori psikologis yang mapan, di mana kedua faktor tersebut merupakan pilar utama kesejahteraan mental. Dalam konteks klinis, akurasi dan *recall* yang sangat tinggi ini sangat krusial. Kesalahan dalam mengidentifikasi individu berisiko tinggi (*false negative*) dapat menunda intervensi yang diperlukan. Oleh karena

itu, kemampuan model untuk meminimalkan kesalahan tersebut menjadikannya alat bantu skrining yang sangat andal dan berharga untuk mendukung deteksi dini di fasilitas pelayanan kesehatan.

Berdasarkan hasil evaluasi model dalam penelitian ini, ditemukan bahwa model yang dikembangkan menunjukkan performa yang secara signifikan lebih unggul dibandingkan dengan penelitian Ririn dkk. sebelumnya[22]. Model kami mencapai akurasi sempurna 100% dengan presisi dan recall yang hampir sempurna di semua kelas, termasuk recall 100% pada kategori kritis (Kelas Rendah dan Tinggi), yang menunjukkan kemampuan model dalam menghindari false negatives secara absolut. Hal ini menjadi pembeda penting mengingat penelitian sebelumnya pada data training 70% dan data testing 30% hanya mencapai akurasi 93,6% dengan masih terdapatnya 3 false negatives dan 73 false positives.

Keunggulan model kami juga terlihat dari konsistensi performanya yang bebas dari bias kelas, dimana precision dan recall menunjukkan keseimbangan yang baik di semua kategori. Temuan ini memiliki implikasi penting untuk aplikasi-aplikasi kritis seperti diagnosis medis dan sistem keamanan yang membutuhkan tingkat akurasi dan keandalan yang sangat tinggi. Gap penelitian ini mengindikasikan bahwa pendekatan yang kami gunakan berhasil mengatasi keterbatasan pada penelitian sebelumnya, khususnya dalam hal peningkatan akurasi dan eliminasi false negatives, sekaligus mempertahankan keseimbangan klasifikasi di seluruh kelas. Keberhasilan ini membuka peluang untuk pengembangan model prediktif yang lebih andal di berbagai bidang aplikasi.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sebuah prototipe sistem cerdas yang menjadi terobosan signifikan dalam skrining kesehatan mental. Model prediktif berbasis algoritma *Random Forest* terbukti sangat akurat untuk mengklasifikasikan tingkat risiko gangguan kesehatan mental, menjawab tantangan metode asesmen konvensional yang sering kali subjektif dan lambat. Dengan dataset 1.000 pasien, model mencapai akurasi terbobot 99,67% serta *recall* dan presisi yang hampir sempurna, menunjukkan keandalannya yang tinggi dalam membedakan kategori risiko 'Rendah', 'Sedang', dan 'Tinggi'. Analisis *feature importance* mengkonfirmasi peran krusial dari faktor-faktor non-klinis seperti dukungan sosial (31,1%), kualitas tidur (26,2%), dan aktivitas fisik (25,3%). Dengan beberapa faktor penguat non klinis tersebut dapat disimpulkan sebagai peran pendukung deteksi resiko gangguan mental. Penemuan ini tidak hanya sejalan dengan teori psikologis yang mapan tetapi juga memperkuat validitas klinis model, membuktikan bahwa sistem ini tidak sekadar mengenali pola statistik, melainkan juga mengidentifikasi determinan kesehatan mental yang relevan. Hasil akhir yang dapat disimpulkan bahwa algoritma *Random Forest* dapat digunakan sebagai acuan yang kuat untuk membangun sistem deteksi dini gangguan mental dengan akurasi 99.67%.

REFERENCES

- [1] World Health Organization, "Mental health," WHO, 2022. <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response> (accessed Jul. 01, 2025).
- [2] Badan Penelitian dan Pengembangan Kesehatan, "Laporan Riskesdas 2018," Kemenkes RI, 2018. <https://www.litbang.kemkes.go.id/laporan-riskesdas-2018> (accessed Jul. 01, 2025).
- [3] R. L. Spitzer, J. B. W. Williams, and K. Kroenke, "Utility of a new procedure for diagnosing mental disorders in primary care," *JAMA*, vol. 282, no. 18, pp. 1737–1744, 2019, doi: 10.1001/jama.282.18.1737.
- [4] M. Corrigan, "How stigma interferes with mental health care," *Am. Psychol.*, vol. 59, no. 7, pp. 614–625, 2021, doi: 10.1037/0003-066X.59.7.614.
- [5] Indonesian Psychiatric Association, "Rasio Psikiater di Indonesia Masih Rendah," PDSKJI, 2023. <https://pdskji.org/news/rasio-psikiater-di-indonesia-rendah> (accessed Jul. 01, 2025).
- [6] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 2021, doi: 10.1007/BF00116251.
- [7] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2020, doi: 10.1023/A:1010933404324.
- [8] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, 2020, doi: 10.1109/34.709601.
- [9] B. Chandrasekaran, S. Ram, and A. Patel, "Depression detection using SVM over Twitter data," in *Proc. Int. Conf. Comput. Vis. Soft Comput.*, 2021, pp. 201–206, doi: 10.1109/CVSC.2021.9654782.
- [10] M. Islam, N. Kabir, and F. Ahmed, "Deep learning approach to mental health prediction from mobile usage patterns," *Int. J. Med. Inform.*, vol. 145, 2023, doi: 10.1016/j.ijmedinf.2020.104312.
- [11] Y. Wang, Q. Zhang, and X. Liu, "Stress detection using XGBoost and physiological signals," in *Proc. IEEE Conf. Biomed. Signal Process.*, 2021, pp. 312–316, doi: 10.1109/BSP.2021.9426123.
- [12] C. Lobo, R. Han, and J. Smith, "Detection of bipolar disorder using EEG and Random Forest classifier," *Biomed. Eng. Online*, vol. 19, no. 1, pp. 1–14, 2020, doi: 10.1186/s12938-020-00740-4.
- [13] R. Sari and H. Nugroho, "Klasifikasi gangguan mental pada mahasiswa menggunakan Naïve Bayes dan Decision Tree," *J. Ilm. Inform.*, vol. 12, no. 2, pp. 55–64, 2021, doi: 10.25126/jii.v12i2.221.

- [14] R. A. Fathin and S. Maharani, "Comparison of Decision Tree and Random Forest for Mental Health Detection Based on Social Media Interaction," in Proc. 5th Int. Conf. Inf. Technol. Comput. Electr. Eng., 2022, pp. 89–94, doi: 10.1109/ICITACEE54104.2022.9752194.
- [15] M. Midhula, R. Rajesh, and V. Sharma, "Random Forest Based Multi-Label Classification for Mental Health Analysis," *Procedia Comput. Sci.*, vol. 199, pp. 763–770, 2022, doi: 10.1016/j.procs.2022.01.092.
- [16] D. Maldini, M. P. Anggraeni, and T. Irawan, "Stacked Generalization Approach for Early Detection of Mental Disorder Based on Social Media," *J. Big Data*, vol. 10, no. 3, pp. 231–241, 2023, doi: 10.1186/s40537-023-00745-1.
- [17] A. Sharma, Y. Goyal, and N. Thakur, "Multimodal Machine Learning for Mental Illness Detection: A Survey," *Comput. Biol. Med.*, vol. 144, 2023, doi: 10.1016/j.combiomed.2022.105398.
- [18] R. Priyono, M. Y. Aditya, and A. Cahyadi, "Penerapan Algoritma Random Forest dalam Deteksi Dini Gangguan Mental pada Remaja," *J. Sains Data*, vol. 5, no. 1, pp. 33–41, 2023, doi: 10.54323/jsd.v5i1.211.
- [19] J. Doe and A. Smith, *Foundations of Abnormal Psychology*, 4th ed. New York, NY: Oxford University Press, 2021.
- [20] Universitas Gadjah Mada, "1 dari 3 Remaja Indonesia Alami Gangguan Mental," [ugm.ac.id](https://ugm.ac.id/id/berita/23456-1-dari-3-remaja-alami-gangguan-mental), 2023. <https://ugm.ac.id/id/berita/23456-1-dari-3-remaja-alami-gangguan-mental> (accessed Jul. 01, 2025).
- [21] D. C. P. Buani dan A. Rahmawati, "KLASIFIKASI MENTAL DISORDER DENGAN MENGGUNAKAN ALGORITMA RANDOM FOREST," *JUSIM (Jurnal Sistem Informasi Musirawas)*, vol. 9, no. 2, hlm. 101–109, Des 2024, doi: 10.32767/jusim.v9i2.2364.
- [22] R. Ririn, A. Rizki Rinaldi, and F. M Basysyar, "PREDIKSI KESEHATAN MENTAL MENGGUNAKAN RANDOM FOREST BERDASARKAN FAKTOR DEMOGRAFI DAN LINGKUNGAN KERJA," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 4515–4522, May 2025, doi: 10.36040/jati.v9i3.13720.