

# IMPLEMENTASI DATA MINING UNTUK PREDIKSI DAN KLASIFIKASI TINGKAT STRES MENGGUNAKAN ALGORITMA RANDOM FOREST

Fitriani Noer Jamilah<sup>1</sup>, Dadang Mulyana Iskandar<sup>2</sup>

<sup>1,2</sup> Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Jl. Radin Inten II No.8 5, RT.5/RW.14, Duren Sawit, Kec. Duren Sawit, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta 13440

## Keywords:

Data Mining, Random Forest, Prediksi Stres, Perceived Stress Scale, Streamlit

## Correspondent Email:

jamilahnoerfitri-  
ani@gmail.com

**Abstrak.** Stres merupakan permasalahan kesehatan mental yang dapat menurunkan produktivitas serta kualitas hidup individu, sehingga diperlukan upaya identifikasi dini yang akurat dan efisien. Penelitian ini bertujuan mengimplementasikan teknik data mining untuk prediksi dan klasifikasi tingkat stres menggunakan algoritma Random Forest. Data penelitian diperoleh dari hasil survei *Perceived Stress Scale* (PSS-10) yang dipadukan dengan indikator gaya hidup, seperti pola tidur dan aktivitas fisik, yang dikumpulkan dari 512 responden melalui Google Form. Sistem klasifikasi dikembangkan berbasis Streamlit untuk mendukung proses *preprocessing* data, pelatihan model, pengujian internal dan eksternal, serta visualisasi hasil prediksi. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *f1-score*. Hasil pengujian menunjukkan bahwa algoritma Random Forest mencapai akurasi sebesar 99,03% pada pengujian internal dan 99,80% pada pengujian eksternal, dengan nilai *precision*, *recall*, dan *f1-score* yang mendekati 1,00 pada seluruh kategori tingkat stres. Hasil ini menunjukkan bahwa model memiliki kinerja yang sangat baik dan konsisten, sehingga berpotensi diterapkan sebagai alat bantu skrining awal tingkat stres di lingkungan pendidikan, perusahaan, dan layanan kesehatan.



Copyright © JITET (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

**Abstract.** Stress is a mental health problem that can reduce individual productivity and quality of life, making early and accurate identification essential. This study aims to implement data mining techniques for stress level prediction and classification using the Random Forest algorithm. The dataset was obtained from a *Perceived Stress Scale* (PSS-10) survey combined with lifestyle indicators, such as sleep patterns and physical activity, and collected from 512 respondents through Google Forms. A Streamlit-based system was developed to support data preprocessing, model training, internal and external testing, and visualization of prediction results. Model performance was evaluated using accuracy, *precision*, *recall*, and *f1-score* metrics. Experimental results show that the Random Forest algorithm achieved an accuracy of 99.03% in internal testing and 99.80% in external testing, with *precision*, *recall*, and *f1-score* values approaching 1.00 across all stress level categories. These findings indicate that the proposed model demonstrates excellent and consistent performance, and has strong potential to be applied as an early stress screening tool in educational institutions, workplaces, and healthcare services.

## 1. PENDAHULUAN

Stres merupakan respons psikologis individu terhadap tuntutan atau tekanan lingkungan yang dirasakan melebihi kemampuan adaptasinya. Kondisi stres dapat muncul akibat berbagai faktor, seperti tuntutan akademik, pekerjaan, hubungan sosial, serta perubahan gaya hidup yang cepat. Stres yang tidak dikelola dengan baik dapat berdampak pada kesehatan mental dan fisik, menurunkan produktivitas, serta memengaruhi kualitas hidup individu secara keseluruhan [1]. Dalam konteks kehidupan modern, peningkatan aktivitas dan tuntutan sehari-hari menyebabkan individu semakin rentan mengalami stres, terutama pada kelompok usia produktif dan pelajar [2].

Pengukuran tingkat stres secara objektif menjadi hal penting untuk mendukung upaya pencegahan dan penanganan stres sejak dini. Salah satu instrumen psikologis yang banyak digunakan dalam penelitian adalah **Perceived Stress Scale (PSS-10)**. Instrumen ini dirancang untuk mengukur persepsi individu terhadap situasi yang dianggap penuh tekanan dalam kurun waktu satu bulan terakhir. PSS-10 telah digunakan secara luas karena memiliki validitas dan reliabilitas yang baik serta mudah diimplementasikan dalam berbagai konteks penelitian [3][4]. Dengan menggunakan PSS-10, tingkat stres individu dapat dikategorikan ke dalam beberapa tingkat, sehingga memudahkan proses analisis dan pengambilan keputusan.

Selain faktor psikologis, **gaya hidup dan pola tidur** diketahui memiliki keterkaitan yang erat dengan tingkat stres individu. Pola tidur yang buruk, durasi tidur yang tidak mencukupi, serta kebiasaan hidup yang kurang sehat dapat meningkatkan risiko terjadinya stres dan gangguan kesehatan mental. Penelitian menunjukkan bahwa individu dengan kualitas tidur yang rendah cenderung memiliki tingkat stres yang lebih tinggi dibandingkan individu dengan pola tidur yang baik [5]. Oleh karena itu, gaya hidup dan pola tidur menjadi variabel penting yang relevan untuk dianalisis dalam penelitian terkait prediksi tingkat stres.

Seiring dengan meningkatnya jumlah data yang bersifat multivariat, pengolahan data secara konvensional menjadi kurang efisien dan rentan terhadap kesalahan. Perkembangan teknologi informasi memungkinkan penerapan **data mining dan machine learning** untuk membantu proses pengolahan dan analisis data dalam jumlah besar. Metode machine learning mampu mengekstraksi pola tersembunyi dari data dan menghasilkan prediksi yang lebih akurat dibandingkan pendekatan tradisional [6]. Dalam bidang kesehatan mental, penerapan machine learning memberikan peluang untuk melakukan analisis tingkat stres secara lebih cepat dan sistematis.

Salah satu algoritma machine learning yang banyak digunakan adalah **Random Forest**, yaitu algoritma ensemble berbasis decision tree yang bekerja dengan membangun banyak pohon keputusan dan menggabungkan hasilnya. Random Forest memiliki keunggulan dalam menangani data berdimensi tinggi, hubungan non-linear, serta mampu mengurangi risiko overfitting. Beberapa penelitian menunjukkan bahwa Random Forest memiliki performa yang baik dalam tugas klasifikasi, termasuk dalam klasifikasi tingkat stres berbasis data survei psikologis dan variabel gaya hidup [7]. Keunggulan tersebut menjadikan Random Forest sebagai algoritma yang potensial untuk digunakan dalam penelitian ini.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menerapkan algoritma Random Forest dalam memprediksi dan mengklasifikasikan tingkat stres berdasarkan **gaya hidup dan pola tidur**. Model yang dibangun tidak hanya dievaluasi dari sisi akurasi, tetapi juga diimplementasikan ke dalam **aplikasi berbasis web** agar dapat digunakan secara praktis sebagai alat bantu skrining awal tingkat stres. Diharapkan penelitian ini dapat memberikan kontribusi dalam pengembangan sistem prediksi tingkat stres berbasis teknologi informasi yang akurat, efisien, dan mudah diakses.

## 2. PUSTAKA

### 2.1 Stres dan Tingkat Stres

Stres merupakan respons individu terhadap tekanan atau tuntutan yang berasal dari lingkungan internal maupun eksternal. Stres dapat muncul akibat faktor akademik, pekerjaan, sosial, maupun gaya hidup sehari-hari [1]. Tingkat stres umumnya diklasifikasikan menjadi beberapa kategori seperti rendah, sedang, dan tinggi berdasarkan intensitas serta dampaknya terhadap kondisi fisik dan psikologis individu [2]. Pengukuran tingkat stres penting dilakukan sebagai langkah awal untuk mencegah dampak lanjutan yang lebih serius [3].

### 2.2 Perceived Stress Scale (PSS-10)

Perceived Stress Scale (PSS) merupakan instrumen psikologis yang digunakan untuk mengukur tingkat stres yang dirasakan individu dalam satu bulan terakhir [4]. Versi PSS-10 terdiri dari sepuluh pernyataan dengan skala Likert yang mengukur persepsi individu terhadap situasi yang dianggap penuh tekanan. PSS-10 banyak digunakan dalam penelitian karena validitas dan reliabilitasnya yang tinggi serta kemudahan dalam pengumpulan data berbasis survei [5].

### 2.3 Gaya Hidup dan Pola Tidur

Gaya hidup dan pola tidur memiliki pengaruh signifikan terhadap tingkat stres seseorang. Faktor-faktor seperti durasi tidur, kualitas tidur, aktivitas

fisik, konsumsi kafein, serta penggunaan media sosial dapat meningkatkan atau menurunkan tingkat stres [6]. Pola tidur yang buruk dan gaya hidup tidak sehat terbukti berkorelasi dengan peningkatan stres dan gangguan kesehatan mental [7], sehingga variabel-variabel ini relevan untuk digunakan dalam proses klasifikasi tingkat stres.

## 2.4 Data Mining

Data mining merupakan proses penggalian informasi dan pola tersembunyi dari kumpulan data berukuran besar [8]. Proses data mining meliputi tahapan seleksi data, preprocessing, transformasi data, penerapan algoritma, serta evaluasi hasil. Dalam bidang kesehatan mental, data mining banyak digunakan untuk memprediksi dan mengklasifikasi kondisi psikologis individu berdasarkan data survei maupun data perilaku [9].

## 2.5 Algoritma Random Forest

Random Forest adalah algoritma machine learning berbasis ensemble yang menggabungkan banyak pohon keputusan untuk menghasilkan prediksi yang lebih akurat dan stabil [10]. Algoritma ini bekerja dengan cara melakukan pelatihan pada sejumlah subset data dan fitur secara acak, kemudian menggabungkan hasil prediksi dari setiap pohon. Random Forest dikenal memiliki kemampuan tinggi dalam menangani data berdimensi besar, mengurangi overfitting, serta menghasilkan performa klasifikasi yang baik pada data non-linear [11].

## 2.6 Klasifikasi Tingkat Stres Menggunakan Machine Learning

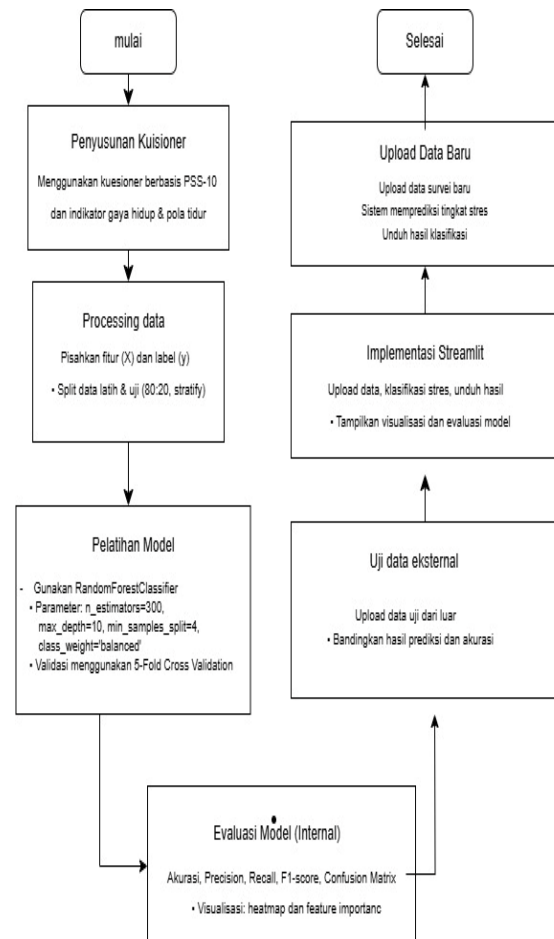
Berbagai penelitian sebelumnya menunjukkan bahwa algoritma machine learning seperti Decision Tree, Random Forest, dan Support Vector Machine mampu mengklasifikasikan tingkat stres dengan akurasi yang tinggi [12][13]. Namun, sebagian besar penelitian masih berfokus pada pengembangan model tanpa implementasi sistem yang mudah diakses oleh pengguna. Oleh karena itu, diperlukan pengembangan sistem klasifikasi tingkat stres yang tidak hanya akurat, tetapi juga aplikatif dan mudah digunakan.

## 3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan **eksperimental kuantitatif** untuk mengembangkan dan menguji model klasifikasi tingkat stres berdasarkan gaya hidup dan pola tidur. Metode yang digunakan adalah **machine learning** dengan algoritma **Random Forest**. Data penelitian diperoleh dari survei daring yang mengacu pada instrumen **Perceived Stress Scale-10 (PSS-10)** serta variabel pendukung terkait kebiasaan tidur dan gaya hidup responden.

Model yang dibangun bertujuan untuk mengklasifikasi tingkat stres ke dalam lima kategori, yaitu

**sangat rendah, rendah, sedang, tinggi, dan sangat tinggi**. Proses penelitian meliputi pengumpulan data, pelabelan tingkat stres, pra-pemrosesan data, pelatihan model Random Forest, serta evaluasi performa model menggunakan data uji internal dan eksternal. Tahapan penelitian secara umum ditunjukkan pada **Gambar 1**.



Gambar 1. Tahapan Metode Penelitian

## 3.1 Data Penelitian

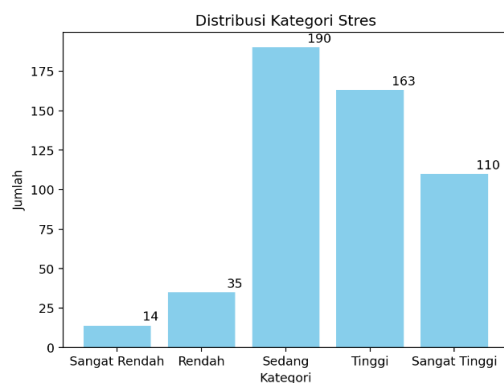
Tahap awal penelitian adalah pengumpulan data menggunakan pendekatan kuantitatif. Data diperoleh melalui survei daring yang disebarakan kepada responden menggunakan Google Form. Instrumen penelitian terdiri dari dua bagian utama, yaitu Perceived Stress Scale (PSS-10) yang digunakan untuk mengukur tingkat stres perseptual responden, serta pertanyaan terkait gaya hidup dan pola tidur. Variabel gaya hidup meliputi durasi dan kualitas tidur, aktivitas fisik, konsumsi kafein, tekanan keuangan, dan intensitas penggunaan media sosial. Data yang terkumpul berjumlah 512 responden dan seluruh data dikumpulkan secara anonim untuk menjaga kerahasiaan dan privasi responden..

Gambar 2. Atribut Dataset Penelitian

### 3.2 Pelabelan Tingkat Stres

Pelabelan tingkat stres dilakukan berdasarkan skor total PSS-10. Skor tersebut diklasifikasikan ke dalam lima kategori tingkat stres. Hasil pelabelan ini digunakan sebagai label target dalam proses pelatihan model.

Gambar 3. Tampilan Data sudah Terlabeli



Gambar 4. Distribusi Hasil Pelabelan

Grafik menunjukkan bahwa kategori stres sedang dan stres tinggi memiliki jumlah responden yang lebih dominan dibandingkan kategori lainnya. Terdapat 512 data yang sudah terimpor. Dari total 512 responden, sebanyak 14 orang (2,73%) berada pada kategori Sangat Rendah, 35 orang

(6,84%) kategori Rendah, 190 orang (37,11%) kategori Sedang, 163 orang (31,84%) kategori Tinggi, dan 110 orang (21,48%) kategori Sangat Tinggi.

### 3.3 Pra-pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk memastikan data siap digunakan dalam proses pelatihan model. Proses ini meliputi pembersihan data dari nilai kosong atau tidak valid, penyeragaman format data, serta transformasi variabel kategorikal ke dalam bentuk numerik menggunakan teknik encoding. Selanjutnya, dataset dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% menggunakan metode *stratified split* untuk mempertahankan proporsi setiap kelas tingkat stres. Metode oversampling seperti SMOTE tidak digunakan agar distribusi data tetap merepresentasikan kondisi asli responden.

### 3.4 Klasifikasi Menggunakan Random Forest

Algoritma Random Forest digunakan sebagai metode klasifikasi utama. Model dibangun menggunakan sejumlah pohon keputusan yang dilatih secara acak pada subset data dan fitur. Untuk mengatasi ketidakseimbangan kelas, model dikonfigurasi dengan parameter `class_weight = balanced`. Model kemudian dilatih menggunakan data latih untuk mempelajari pola hubungan antara variabel input dan tingkat stres.

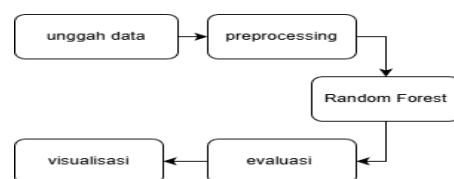
### 3.5 Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja dan kemampuan generalisasi model yang dibangun. Pengujian dilakukan menggunakan data uji internal dan data uji eksternal. Kinerja model diukur menggunakan metrik akurasi, precision, recall, f1-score, serta confusion matrix untuk menilai kesesuaian hasil prediksi dengan kondisi aktual tingkat stres.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Implementasi Sistem dan Alur Pengujian

Sistem dikembangkan berbasis **Streamlit** untuk mendukung alur end-to-end: unggah data survei, pelabelan otomatis tingkat stres, pra-pemrosesan, pelatihan model, pengujian internal dan eksternal, serta visualisasi hasil. Alur ini memastikan kemudahan penggunaan oleh pengguna non-teknis sekaligus menjaga reproduktibilitas eksperimen.

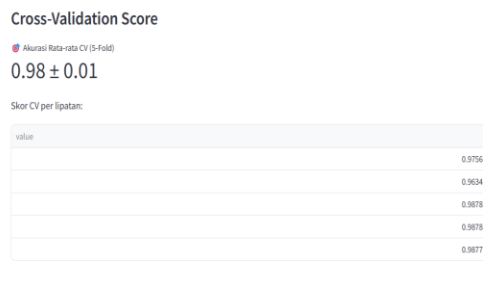


Gambar 5. Diagram Arsitektur Sistem

#### 4.2 Hasil Pelatihan Model Random Forest

Model pada penelitian ini dilatih menggunakan algoritma Random Forest dengan tujuan mengklasifikasikan tingkat stres berdasarkan data PSS-10, gaya hidup, dan pola tidur. Pemilihan algoritma Random Forest didasarkan pada kemampuannya dalam menangani data dengan banyak fitur, mengurangi risiko overfitting, serta menghasilkan performa klasifikasi yang stabil.

Parameter yang digunakan dalam proses pelatihan disesuaikan dengan implementasi pada skripsi, yaitu  $n\_estimators = 300$ ,  $max\_depth = 10$ , dan  $min\_samples\_split = 4$ . Selain itu, parameter  $class\_weight$  diatur menjadi *balanced* untuk mengatasi potensi ketidakseimbangan distribusi kelas tingkat stres. Proses pelatihan model dilakukan dengan menerapkan metode **5-Fold Cross Validation** guna menguji kestabilan performa model pada berbagai subset data. Pendekatan ini memastikan bahwa model tidak hanya bekerja optimal pada data latih, tetapi juga memiliki kemampuan generalisasi yang baik.

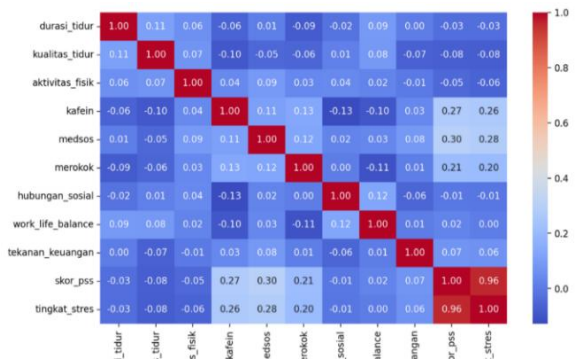


Gambar 6. Proses Pelatihan Model Random Forest

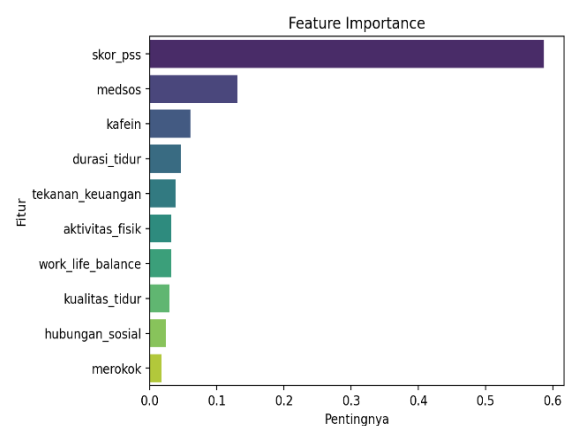
Pada Hasil pengujian pada Gambar 6 menggunakan Cross-Validation dengan metode 5-Fold menunjukkan bahwa model memiliki akurasi rata-rata sebesar 0,98 dengan deviasi  $\pm 0,01$ . Nilai ini mengindikasikan bahwa model mampu memberikan prediksi yang sangat akurat dan konsisten pada berbagai subset data. Skor akurasi pada masing-masing lipatan (fold) berturut-turut adalah 0,9756; 0,9634; 0,9878; 0,9878; dan 0,9877. Perbedaan skor antar lipatan relatif kecil, menandakan model memiliki performa yang stabil dan tidak terlalu bergantung pada data latih tertentu, sehingga potensi terjadinya *overfitting* lebih rendah.

#### 4.3 Analisis Korelasi Fitur dan Feature Importance

Analisis Korelasi Fitur dan Feature Importance Analisis korelasi menunjukkan bahwa variabel terkait kualitas dan durasi tidur serta skor PSS-10 memiliki hubungan paling kuat terhadap tingkat stres. *Feature importance* mengonfirmasi bahwa fitur-fitur tersebut berkontribusi dominan pada keputusan model.



Gambar 7. Tampilan Korelasi Antar Fitur



Gambar 8. Tampilan Feature Importance

Berdasarkan Gambar 7, matriks korelasi menunjukkan bahwa skor\_pss memiliki hubungan yang sangat kuat dengan tingkat\_stres ( $r = 0,96$ ), sedangkan variabel lain seperti penggunaan media sosial, konsumsi kafein, dan kebiasaan merokok memiliki korelasi positif yang lebih rendah. Selanjutnya, Gambar 8 yang menampilkan tingkat kepentingan fitur (*feature importance*) dari model Random Forest memperkuat hasil tersebut, di mana skor\_pss menjadi faktor paling dominan dalam prediksi tingkat stres, diikuti oleh penggunaan media sosial dan konsumsi kafein, sementara variabel gaya hidup lainnya memberikan kontribusi yang lebih kecil.

#### 4.4 Evaluasi Internal Model

Evaluasi internal dilakukan menggunakan data uji internal untuk mengukur performa model dalam mengklasifikasikan tingkat stres. Metrik evaluasi yang digunakan meliputi akurasi, precision, recall, dan F1-score, serta confusion matrix untuk melihat distribusi prediksi pada masing-masing kelas.

Hasil evaluasi menunjukkan bahwa model Random Forest mencapai akurasi sebesar 99,03%, dengan nilai precision, recall, dan F1-score yang tinggi pada seluruh kategori tingkat stress.



Aktual: Sangat Rendah	3	0	0	0	0
Aktual: Rendah	0	6	1	0	0
Aktual: Sedang	0	0	38	0	0
Aktual: Tinggi	0	0	0	33	0
Aktual: Sangat Tinggi	0	0	0	0	22
	Prediksi: Sangat Rendah	Prediksi: Rendah	Prediksi: Sedang	Prediksi: Tinggi	Prediksi: Sangat Tinggi

Gambar 9. Confusion Matrix Hasil Evaluasi Internal Model

Berdasarkan Gambar 9, confusion matrix menunjukkan bahwa model memiliki kinerja klasifikasi yang sangat baik, dengan mayoritas prediksi tepat pada kelas yang sesuai. Seluruh data pada kelas *Sangat Rendah*, *Sedang*, *Tinggi*, dan *Sangat Tinggi* diprediksi dengan benar tanpa kesalahan, sedangkan pada kelas *Rendah* terdapat satu data yang salah diprediksi sebagai *Sedang*. Hasil ini mengindikasikan tingkat akurasi yang sangat tinggi serta kemampuan model dalam membedakan kategori tingkat stres secara konsisten.

Classification Report				
	precision	recall	f1-score	support
Sangat Rendah	1.00	1.00	1.00	3.00
Rendah	1.00	0.86	0.92	7.00
Sedang	0.97	1.00	0.99	38.00
Tinggi	1.00	1.00	1.00	33.00
Sangat Tinggi	1.00	1.00	1.00	22.00
accuracy	0.99	0.99	0.99	0.99
macro avg	0.99	0.97	0.98	103.00
weighted avg	0.99	0.99	0.99	103.00

Gambar 10. Classification Report Internal Model

Gambar 10 menunjukkan hasil uji internal model Random Forest dengan akurasi 99%. Hampir semua kategori memiliki precision, recall, dan f1-score sempurna, kecuali kategori **Rendah** dengan recall 0,86. Nilai **macro avg** 0,98 dan **weighted avg** 0,99 menunjukkan performa model sangat baik dan konsisten dalam membedakan tingkat stres.

#### 4.5 Evaluasi Eksternal Model

Pengujian eksternal menggunakan data baru menunjukkan **akurasi 99,80%**, menegaskan kemampuan generalisasi model. Konsistensi performa

antara pengujian internal dan eksternal mengindikasikan minimnya overfitting.

Sangat Rendah	14	0	0	0	0
Rendah	0	34	1	0	0
Sedang	0	0	190	0	0
Tinggi	0	0	0	163	0
Sangat Tinggi	0	0	0	0	110
	Sangat Rendah	Rendah	Sedang	Tinggi	Sangat Tinggi

Gambar 11. Confusion Matrix Hasil Evaluasi Eksternal Model

Classification Report				
	precision	recall	f1-score	support
Sangat Rendah	1.00	1.00	1.00	14.00
Rendah	1.00	0.97	0.99	35.00
Sedang	0.99	1.00	1.00	190.00
Tinggi	1.00	1.00	1.00	163.00
Sangat Tinggi	1.00	1.00	1.00	110.00
accuracy	1.00	1.00	1.00	1.00
macro avg	1.00	0.99	1.00	512.00
weighted avg	1.00	1.00	1.00	512.00

Akurasi Model: 99.80%

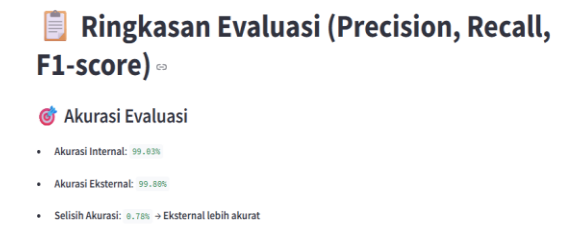
Gambar 12. Classification Report Eksternal Model

Gambar 11 dan 12 menunjukkan hasil uji eksternal model Random Forest. Confusion matrix pada Gambar 11 memperlihatkan bahwa hampir seluruh data berhasil diklasifikasikan dengan benar, dengan hanya satu kesalahan pada kelas Rendah yang diprediksi sebagai Sedang, sehingga menunjukkan tingkat ketepatan prediksi yang sangat tinggi.

Selanjutnya, classification report pada Gambar 12 menunjukkan akurasi sebesar 99,80% dengan nilai precision, recall, dan f1-score yang mendekati sempurna pada seluruh kelas. Nilai macro avg dan weighted avg sebesar 1,00 menandakan performa model yang konsisten pada semua kategori. Hasil ini membuktikan bahwa model memiliki kemampuan generalisasi yang sangat baik pada data baru.

#### 4.6 Perbandingan Akurasi Internal dan Eksternal

Perbandingan akurasi memperlihatkan selisih yang sangat kecil antara pengujian internal dan eksternal. Hal ini menguatkan bahwa pendekatan *class weight = balanced* efektif tanpa perlu oversampling sintetis.



Gambar 13. Ringkasan Evaluasi Internal dan Eksternal

Gambar 13 menampilkan ringkasan evaluasi akurasi model berdasarkan pengujian internal dan eksternal. Pada pengujian internal, akurasi model tercatat sebesar 99,03%, sedangkan pada pengujian eksternal akurasinya sedikit lebih tinggi, yaitu 99,80%. Selisih akurasi sebesar 0,78% menunjukkan bahwa performa model pada data eksternal justru lebih baik dibandingkan data internal. Hal ini menandakan bahwa model memiliki kemampuan generalisasi yang sangat baik, sehingga mampu mempertahankan bahkan meningkatkan kinerjanya ketika dihadapkan pada data baru yang belum pernah digunakan dalam proses pelatihan.

	precision	recall	f1-score	support
1	1.00	1.00	1.00	3.00
2	1.00	0.86	0.92	7.00
3	0.97	1.00	0.99	38.00
4	1.00	1.00	1.00	33.00
5	1.00	1.00	1.00	22.00

	precision	recall	f1-score	support
1	1.00	1.00	1.00	14.00
2	1.00	0.97	0.99	35.00
3	0.99	1.00	1.00	190.00
4	1.00	1.00	1.00	163.00
5	1.00	1.00	1.00	110.00

Gambar 14. Classification Report Internal dan Eksternal

Gambar 14 menampilkan perbandingan detail hasil evaluasi model Random Forest pada pengujian internal dan eksternal. Pada evaluasi internal, sebagian besar kelas memiliki precision, recall, dan f1-score sempurna (1,00), kecuali kelas ke-2 yang memiliki recall 0,86 dan f1-score 0,92, serta kelas ke-3 dengan precision 0,97. Hal ini menunjukkan masih terdapat sedikit kesalahan klasifikasi pada data internal, khususnya pada kelas dengan jumlah data kecil.

Sementara itu, pada evaluasi eksternal, nilai precision, recall, dan f1-score hampir seluruhnya sempurna (1,00), hanya kelas ke-2 yang memiliki recall 0,97 dan f1-score 0,99. Perbedaan ini menunjukkan bahwa performa model justru lebih stabil dan akurat

pada data eksternal. Jumlah data (support) pada evaluasi eksternal juga lebih besar dibandingkan internal, menandakan bahwa model mampu mempertahankan kinerja tinggi meskipun diuji pada dataset baru yang lebih beragam.

#### 4.7 Pembahasan dan Implikasi

Hasil penelitian menunjukkan bahwa algoritma Random Forest sangat efektif dalam mengklasifikasi tingkat stres berdasarkan kombinasi skor PSS-10, gaya hidup, dan pola tidur. Akurasi yang sangat tinggi pada pengujian internal (99,03%) dan eksternal (99,80%) mengindikasikan bahwa model tidak hanya mampu mempelajari pola pada data latih, tetapi juga memiliki kemampuan generalisasi yang sangat baik ketika dihadapkan pada data baru. Hal ini penting mengingat data stres bersifat subjektif dan sangat dipengaruhi oleh variasi individu.

Jika dibandingkan dengan penelitian sebelumnya, seperti Larasati et al. dan Oktaviani et al., performa model dalam penelitian ini berada pada tingkat yang sama atau lebih tinggi, namun dengan pendekatan yang lebih aplikatif. Penelitian terdahulu umumnya berhenti pada tahap evaluasi model, sementara penelitian ini mengintegrasikan model ke dalam sistem berbasis web yang dapat digunakan secara langsung oleh pengguna non-teknis. Dengan demikian, kontribusi penelitian tidak hanya terletak pada peningkatan akurasi, tetapi juga pada aspek implementasi sistem dan pengalaman pengguna.

Penggunaan parameter *class\_weight = balanced* terbukti mampu menjaga performa model pada seluruh kelas tingkat stres tanpa harus menerapkan teknik oversampling sintesis seperti SMOTE. Pendekatan ini dinilai lebih aman dalam konteks data psikologis karena tidak mengubah distribusi alami responden. Selain itu, analisis *feature importance* menunjukkan bahwa variabel terkait kualitas dan durasi tidur memiliki kontribusi besar dalam proses klasifikasi, sehingga memperkuat temuan teoretis bahwa pola tidur merupakan faktor signifikan dalam pembentukan stres. Dari sisi implementasi, sistem berbasis Streamlit yang dikembangkan memungkinkan proses klasifikasi dilakukan secara end-to-end, mulai dari unggah data hingga visualisasi hasil evaluasi model. Visualisasi seperti confusion matrix, classification report, dan grafik perbandingan akurasi memudahkan pengguna dalam memahami performa sistem. Implikasi praktis dari penelitian ini adalah tersedianya alat bantu skrining awal tingkat stres yang dapat diterapkan di lingkungan pendidikan, organisasi, maupun layanan kesehatan sebagai langkah preventif sebelum dilakukan penanganan lebih lanjut oleh tenaga profesional.

#### 5. KESIMPULAN

- Penelitian ini berhasil menerapkan algoritma **Random Forest** untuk mengklasifikasikan

tingkat stres berdasarkan skor PSS-10, gaya hidup, dan pola tidur ke dalam lima kategori, yaitu sangat rendah, rendah, sedang, tinggi, dan sangat tinggi.

- b. Model yang dibangun menunjukkan performa yang sangat baik dengan tingkat akurasi sebesar **99,03% pada pengujian internal** dan **99,80% pada pengujian eksternal**, disertai nilai *precision*, *recall*, dan *f1-score* yang tinggi pada seluruh kelas tingkat stres.
- c. Penggunaan parameter *class\_weight = balanced* efektif dalam menangani potensi ketidakseimbangan kelas tanpa menerapkan teknik oversampling sintesis, sehingga distribusi data tetap merepresentasikan kondisi asli responden
- d. Implementasi model ke dalam aplikasi berbasis web menggunakan **Streamlit** memungkinkan proses klasifikasi dilakukan secara end-to-end, mulai dari unggah data hingga visualisasi hasil, sehingga sistem dapat digunakan oleh pengguna non-teknis.
- e. Secara keseluruhan, sistem yang dikembangkan memiliki potensi untuk dimanfaatkan sebagai **alat bantu skrining awal tingkat stres** yang akurat dan efisien

#### UCAPAN TERIMA KASIH

Penulis mengucapkan puji dan syukur ke hadirat **Allah SWT** atas rahmat dan karunia-Nya sehingga penelitian dan penulisan artikel ini dapat diselesaikan dengan baik. Penulis juga mengucapkan terima kasih kepada seluruh responden yang telah berpartisipasi dalam pengisian kuesioner penelitian, serta kepada semua pihak yang telah memberikan dukungan dan bantuan selama proses pelaksanaan penelitian ini.

#### DAFTAR PUSTAKA

- [1] S. Sza *et al.*, “Penerapan Decision Tree Dan Random Forest Dalam Deteksi the Application of Decision Tree and Random Forest in Detecting Human Stress Levels Based on Sleep Conditions,” vol. 10, no. 7, pp. 1503–1510, 2023, doi: 10.25126/jtiik.2024117993.
- [2] V. Oktaviani, N. Rosmawarni, and M. P. Muslim, “Perbandingan Kinerja Random Forest Dan Smote Random Forest Dalam Mendeteksi Dan Mengukur Tingkat Stres Pada Mahasiswa Tingkat Akhir,” *Inform. J. Ilmu Komput.*, vol. 20, no. 1, pp. 43–49, 2024, doi: 10.52958/iftk.v20i1.9158.
- [3] M. Maulidah and N. Hidayati, “Prediksi Kesehatan Tidur Dan Gaya Hidup Menggunakan Machine Learning,” *CONTENT Comput. Netw. Technol.*, vol. 4, no. 1, pp. 81–86, 2024, [Online]. Available: <http://jurnal.bsi.ac.id/index.php/content>
- [4] V. Marcella, A. Simalango, and W. Franciska, “Evaluasi Algoritma Decision Tree dan Random Forest serta Efektivitas Feature Selection dalam Memprediksi Kesehatan Mental,” vol. 7, pp. 201–215, 2025.
- [5] I. A. Hidayat, “Classification of Sleep Disorders Using Random Forest on Sleep Health and Lifestyle Dataset,” *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 3, no. 2, pp. 71–76, 2023, doi: 10.20895/dinda.v3i2.1215.
- [6] A. Van Fadhila *et al.*, “Implementasi Metode Machine Learning Untuk Mendeteksi Tingkat Stres Manusia Berdasarkan Kualitas Tidur,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, vol. 4, no. 1, pp. 130–143, 2023.
- [7] G. Zahra, N. Fadhilah, R. A. Saputra, and A. H. Wibowo, “Deteksi Tingkat Gangguan Kecemasan Menggunakan Metode Random Forest,” *J. Fak. Tek. UMT*, vol. 13, no. 1, pp. 38–47, 2024, [Online]. Available: <http://jurnal.umat.ac.id/index.php/jt/index>
- [8] I. Nawawi and Z. Fatah, “Penerapan Decision Trees dalam Mendeteksi Pola Tidur Sehat Berdasarkan Kebiasaan Gaya Hidup,” vol. 2, no. 4, pp. 34–41, 2024.
- [9] W. Widyastuty and M. A. Azis, “Classification and Evaluation of Sleep Disorders Using Random Forest Algorithm in Health and Lifestyle Dataset,” *Compiler*, vol. 13, no. 1, p. 11, 2024, doi: 10.28989/compiler.v13i1.2184.
- [10] I. F. Wijayanti, E. R. Widasari, and B. H. Prasetyo, “Implementasi Wearable Device untuk Sistem Pendeteksi Stres pada Manusia berdasarkan Suhu Tubuh dan Detak Jantung,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4486–4492, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] E. Abdelfattah, S. Joshi, and S. Tiwari, “Machine and Deep Learning Models for Stress Detection Using Multimodal Physiological Data,” *IEEE Access*, vol. 13, pp. 4597–4608, 2025, doi: 10.1109/ACCESS.2024.3525459.
- [12] A. Anjani and Y. Yamasari, “Klasifikasi Tingkat Stres Mahasiswa Menggunakan Metode Berbasis Tree,” *J. Informatics Comput. Sci.*, vol. 05, pp. 83–89, 2023.
- [13] A. A. Hapsari, A. S. Nursuwanda, H. Zuhriyah, and D. J. Vresdian, “Klasifikasi Kesehatan Mental Mahasiswa Model TMAS dengan Algoritma Decision Tree, Logistic Regression, dan Random Forest,” vol. 7, no. November, 2024.
- [14] J. Sarita, I. Lu, I. Maylani, and P. Setyawati, “Pengaruh Kesehatan Mental Terhadap Stress Akademik Pada Mahasiswa Di Universitas X Dengan Menggunakan Metode Algoritma Decision Tree J48”.
- [15] R. R. Fitri, A. Asriyanik, dan W. Apriandari,



“Penggunaan Random Forest dalam Sistem  
Klasifikasi Kecemasan pada Generasi Z,”  
*Jurnal Informatika dan Teknik Elektro  
Terapan*, vol. 13, no. 3, 2025,  
doi:10.23960/jitet.v13i3.6905.