Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

IMPLEMENTASI *MACHINE LEARNING* MELALUI PENDEKATAN ALGORITMA *RANDOM FOREST* DALAM PREDIKSI TINGKAT STRES BERDASARKAN POLA GAYA HIDUP

Anita Khansa Ramadanti[1], April Lia Hananto[2], Bayu Priyatna[3], Agustia Hananto[4]
1,2,3,4) Sistem Informasi, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang

Article Info

Article history:

Received: 14 Maret 2026
Revised: 29 Maret 2026
Accepted: 29 Maret 2026

ABSTRACT

Abstrak

Stres yang tidak terkelola berisiko berkembang menjadi gangguan serius seperti depresi hingga risiko bunuh diri. *Machine learning* dapat dioptimalkan sebagai solusi deteksi dini berdasarkan kombinasi faktor gaya hidup. Penelitian ini bertujuan untuk mengembangkan model prediksi tingkat stres melalui pendekatan algoritma *Random Forest* dengan dataset yang diperoleh *platform* Kaggle. Tahapan penelitian meliputi data *preprocessing*, penanganan ketidakseimbangan kelas menggunakan SMOTE, hingga evaluasi dan integrasi model. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 0.80, dengan nilai *precision, recall,* dan *F1-Score* secara keseluruhan berada pada angka 0.80. Performa terbaik diperoleh pada klasifikasi tingkat stres kategori *High* dengan F1-Score sebesar 0.86. Model yang telah tervalidasi kemudian diintegrasikan ke dalam antarmuka melalui Streamlit, sehingga mampu memberikan hasil prediksi secara *real-time* berdasarkan *input* data pengguna. Penelitian ini membuktikan bahwa algoritma *Random Forest* efektif dalam mengidentifikasi tingkat stres, dan implementasinya dalam bentuk aplikasi *web* berpotensi menjadi alat bantu deteksi dini yang fungsional dan sederhana.

Kata Kunci: Prediksi Stres, *Machine Learning, Random Forest*, SMOTE, Gaya Hidup.

Abstract

Unmanaged stress can develop into serious disorders such as depression and even suicide, therefore machine learning can be optimized as an early detection solution based on a combination of lifestyle factors. This research aims to develop a stress level prediction method using the Random Forest algorithm approach with a dataset obtained from the Kaggle platform. The research stages include data preprocessing, class imbalance handling using SMOTE, to model evaluation and integration model. The evaluation results show that the model achieves an accuracy of 0.80, with precision, recall, and F1-Score values overall at 0.80. The best performance was obtained in the classification of the High stress level category with an F1-Score of 0.86. The validated model was then integrated into the interface through Streamlit, enabling it to provide real-time prediction results based on user data input. This study proves that the Random Forest algorithm is effective in identifying stress levels, and its implementation in the form of a web application has the potential to become a functional and straightforward early detection tool.

Keywords: Stress Prediction, Machine Learning, Random Forest, SMOTE, Lifestyle.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Corresponding Author:
E-mail : si22.anitaramadanti@mhs.ubpkarawang.ac.id

1. PENDAHULUAN

Permasalahan terkait kondisi mental menjadi isu internasional yang terus meningkat dan menjadi perhatian penuh sejak beberapa tahun belakangan [1],[2]. Salah satu permasalahan utama yang banyak dialami individu adalah stres, yang dapat dipahami sebagai reaksi fisik maupun mental saat menghadapi beban yang melebihi kapasitas adaptasi individu [3],[4]. Stres yang kurang terkontrol dapat berpotensi memicu berkembangan lebih serius, seperti depresi dan ancaman terhadap keselamatan individu [5]. Oleh sebab itu, diperlukan pendekatan deteksi dini yang efektif guna meminimalkan dampak jangka panjang terhadap kesehatan mental.

Perkembangan teknologi, khususnya dalam bidang *machine learning*, memfasilitasi transformasi dalam pengembangan sistem prediksi kesehatan mental. *Machine learning* memberikan kemampuan pada sistem untuk memahami kecenderungan dari data sebelumnya guna menghasilkan prediksi secara otomatis [6],[7]. Dalam kategori model klasifikasi, *Random Forest* dikenal andal dalam memproses fitur data yang kompleks. Melalui pendekatan *ensemble*, algoritma ini mampu menjaga stabilitas model dan menghindari *overfitting* secara optimal [8],[9].

Sejumlah penelitian sebelumnya telah mengeksplorasi penggunaan *Random Forest* dalam kesehatan mental. Penelitian dalam [10] membuktikan efektivitas faktor gaya hidup dalam mengklasifikasikan kondisi depresi dengan tingkat akurasi yang mencapai 91%. Senada dengan hal tersebut, penelitian pada [11], menunjukan keandalan algoritma klasifikasi dalam mendeteksi tingkat stres melalui studi komparatif. Selain itu, penelitian sebelumnya [12] memperkuat argumen bahwa variabel pola hidup merupakan prediktor signifikan dalam memetakan tingkat tekanan psikologis.

Berbagai penelitian telah membuktikan bahwa algoritma *Random Forest* memiliki performa yang baik dalam mengidentifikasi kondisi kesehatan mental. Namun, selain evaluasi performa model, analisis terhadap kontribusi variabel yang berkaitan dengan pola gaya hidup dan kondisi kesehatan individu juga penting guna menganalisis secara luas penyebab yang melatarbelakangi tingkat stres. Hal ini didasarkan pada karakteristik stres yang melibatkan respons psikologis sekaligus fisiologis, sehingga penggunaan berbagai jenis data pendukung diharapkan dapat memperkuat keandalan hasil klasifikasi model.

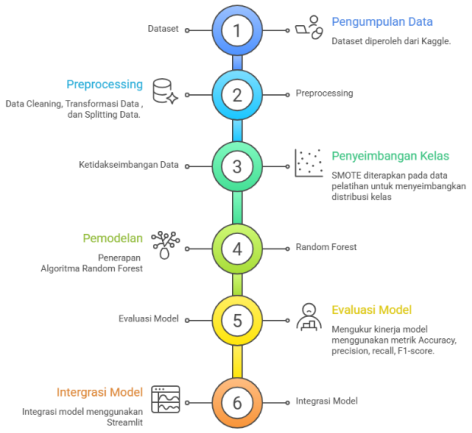
Berdasarkan pada uraian tersebut, penelitian ini bertujuan mengimplementasikan machine learning sebagai model prediksi tingkat stres melalui pendekatan algoritma *Random Forest* pada dataset *Stress Detection Data* yang diperoleh dari *platform* Kaggle.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Penelitian ini tidak hanya mengevaluasi performa model secara kuantitatif melalui metrik akurasi, tetapi juga menerapkan *feature importance* guna menentukan variabel dalam dataset yang memberikan kontribusi paling besar terhadap hasil klasifikasi model. Hasil penelitian ini kemudian diimplementasikan melalui antarmuka interaktif yang dibangun dengan Streamlit guna memfasilitasi pengujian fungsionalitas model.

2. METODE PENELITIAN

Penelitian ini dilaksanakan melalui serangkaian tahapan, diawali dengan tahap pengambilan dataset hingga evaluasi dan diakhiri oleh intergrasi model. Rangkaian tahapan tersebut disajikan pada Gambar berikut.



Gambar 1. Tahapan Penelitian

Tahapan-tahapan tersebut menjadi kerangka utama dalam pelaksanaan penelitian ini dan akan dijelaskan pada subbab berikut.

2.1 Dataset

Penelitian ini memanfaatan sekumpulan data publik yang diakses melalui *platform* Kaggle, yakni *Stress Prediction Data* sebagai sumber data utama untuk melatih model prediksi yang akan dikembangkan. Dataset tersebut mencakup 773 sampel data yang terdiri dari 22 fitur berbeda. Variabel yang digunakan diklasifikasikan menjadi dua komponen utama, yaitu fitur (*feature*) dan target. Variabel yang ditetapkan sebagai target adalah *Stress Detection*, yang terbagi ke dalam tiga kategori, yaitu *Low, Medium*, dan *High*.

Sementara itu, fitur penelitian mencakup berbagai faktor yang merepresentasikan pola gaya hidup individu, seperti aktivitas fisik, pola tidur, dan interaksi sosial, serta indikator fisiologis yang meliputi konsentrasi gula darah, tekanan darah, serta kadar kolesterol. Pemilihan fitur tersebut dinilai relevan karena pola gaya hidup berkontribusi terhadap kondisi fisiologis dan psikologis individu, yang berdampak pada tingkat stres.

2.2 Pra-pemrosesan

Tahap Pra-pemrosesan guna menjamin kebersihan serta reliabilitas data sebelum pemodelan dimulai[13]. Prosedur ini difokuskan pada identifikasi *missing values* dan data duplikat guna memastikan dataset terbebas dari redundansi atau ketidakkonsistenan nilai. Pembersihan data menjadi bagian penting dalam pengolahan data, sebab kualitas input yang digunakan berpengaruh langsung terhadap model yang dihasilkan[14], [15].

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Selain pembersihan data, tahap transformasi juga dilakukan untuk menyesuaikan format data guna mengoptimalkan kinerja pada *machine learning*. Transformasi yang diterapkan meliputi konversi variabel waktu ke dalam bentuk desimal serta penerapan metode *One-Hot Encoding* terhadap fitur bertipe kategorikal. Proses ini dilakukan karena algoritma pembelajaran mesin memerlukan representasi numerik dalam proses perhitungannya, sehingga fitur kategorikal perlu dikonversi ke bentuk numerik sebelum digunakan dalam proses pemodelan [16]. Kemudian, data dibagi menjadi data pelatihan dan data pengujian dengan perbandingan 80:20.

2.3 SMOTE

Ketidakseimbangan jumlah sampel antar kelas berpotensi mendominasi hasil klasifikasi pada kelompok mayoritas dibandingkan kelas minoritas [17], [18]. Untuk meminimalisir potensi permasalahan tersebut, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan pada data pelatihan. Teknik *oversampling* ini menggunakan karakteristik data yang sudah ada sehingga menghasilkan sampel buatan pada kelas minoritas guna meningkatkan keseimbangan distribusi kelas [19], [20]. Oleh karena itu, teknik *oversampling* SMOTE digunakan dalam penelitian ini untuk membantu meningkatkan keseimbangan data serta mengoptimalkan performa model klasifikasi yang dihasilkan [21], [22].

Proses penyeimbang kelas hanya diimplementasikan pada data pelatihan (*training data*) setelah proses pembagian data dilakukan. Pendekatan ini bertujuan guna menghindari kebocoran data dan mempertahankan evaluasi yang tidak bias pada data pengujian[23].

2.4 Penerapan Random Forest

Pendekatan *supervised learning* pada *Random Forest* diimplementasikan dengan membangun sekumpulan pohon keputusan guna menghasilkan keputusan yang terbaik [24]. Dalam penelitian ini, penerapan algoritma tersebut dilakukan menggunakan pendekatan *pipeline* guna menjamin seluruh tahapan pemodelan berjalan secara sistematis dan terstruktur.

Data yang telah melewati tahap praproses kemudian diintegrasikan ke dalam *pipeline* tersebut untuk diimplementasikan pada pengembangan model klasifikasi. Dalam pembangunan model, proses optimasi diterapkan pada sejumlah parameter, khususnya untuk menentukan jumlah pohon, kedalaman maksimal, dan jumlah minimum sampel daun guna mencapai performa terbaik. Berikut adalah representasi dari struktur pohon dengan batasan sampel tersebut:

Tabel 1. *Hyperparameter Tunning*

Hyperparameter	Nilai Uji
n_estimators	100, 200, 300
max_depth	None, 10, 15
min_samples_leaf	1, 2, 3

Berdasarkan Gambar 2, proses pemilihan parameter dilakukan dengan menguji beberapa kombinasi nilai menggunakan teknik *Grid Search* yang dipadukan bersama dengan validasi silang melalui 5-*Fold Cross-Validation*. Optimasi ini bertujuan untuk

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

memperoleh parameter terbaik serta meningkatkan kapabilitas model dalam mengklasifikasikan data baru.

2.5 Evaluasi Model

Evaluasi model diimplementasikan guna menguji kapabilitas algoritma *Random Forest* dalam melakukan klasifikasi tingkat stres. Proses evaluasi menggunakan metrik *accuracy, precision, recall*, dan *F1-score* untuk menilai performa model [25]. *Confusion matrix* juga dimanfaatkan dalam menganalisis prediksi model pada setiap kelas.

Selain itu, diterapkan juga pengecekan bobot fitur untuk melihat seberapa besar peran tiap fitur saat proses pengelompokan dilakukan oleh algoritma *Random Forest*. Langkah ini bertujuan untuk mencari tahu fitur-fitur yang memberikan pengaruh terbesar dalam menentukan hasil tingkat stres.

2.6 Integrasi Model

Sebagai tahap akhir, hasil pengembangan model selanjutnya diuji melalui tampilan situs web yang dibangun dengan *framework* Streamlit. Tampilan tersebut menggunakan model yang sudah diekspor dalam bentuk pickle, sehingga nilai fitur yang dimasukkan dapat diproses langsung oleh model untuk menghasilkan prediksi tingkat stres.

Tahap ini bertujuan untuk memastikan kemampuan model dalam memberikan prediksi secara konsisten ketika menerima data baru, tanpa mengubah konfigurasi maupun parameter yang telah ditentukan pada tahap sebelumnya.

3. HASIL DAN PEMBAHASAN

3.1 Pra-pemrosesan

Tahapan awal Pra-pemrosesan data diawali dengan pemeriksaan terhadap struktur serta kualitas dataset yang digunakan. Langkah ini dilakukan guna memastikan data tersebut telah memenuhi standar kelayakan sebelum memasuki tahap pengolahan selanjutnya.

1. *Data Cleaning*

   Pada tahap ini dilakukan pemeriksaan terhadap keberadaan nilai yang hilang (*missing values*) serta data duplikat pada dataset yang terdiri dari 773 baris dan 22 fitur. Hasil pengecekan menunjukkan bahwa seluruh kolom tidak memiliki nilai yang hilang (0 *missing value*) dan tidak ditemukan data yang terduplikasi (0 *duplicate*). Kondisi ini menunjukkan bahwa dataset yang digunakan memiliki kualitas yang baik sehingga tidak memerlukan proses penghapusan maupun imputasi data sebelum memasuki tahapan transformasi selanjutnya.

2. Transformasi Waktu

   Pada tahap preprocessing, fitur *Wake_Up_Time* dan *Bed_Time* yang semula bertipe *object* dengan format waktu 12 jam (*AM/PM*) ditransformasikan ke dalam bentuk numerik berupa jam desimal. Proses konversi ini menghasilkan representasi waktu dalam format 24 jam sehingga dapat digunakan dalam proses pelatihan model. Hasil transformasi fitur waktu sebagaimana ditampilkan pada Tabel 2.

Tabel 2. Hasil Transformasi Waktu

Wake_Up_Hour	*Bed_Hour*

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

7.0	22.0
6.0	23.0
7.0	22.0
7.0	22.0
6.5	22.5

Transformasi ini mengubah data waktu menjadi representasi numerik sehingga memudahkan model dalam memproses informasi waktu istirahat sebagai salah satu fitur dalam proses klasifikasi tingkat stres.

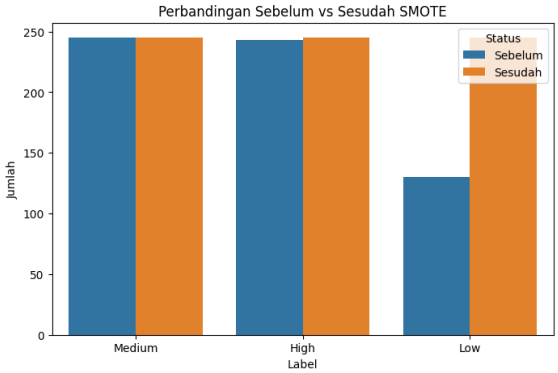
3. Encoding

Proses *encoding* dilakukan terhadap fitur kategorikal selain target untuk mengubah data berbentuk teks menjadi representasi numerik. Setelah proses *one hot encoding*, jumlah fitur mengalami peningkatan dari 21 fitur menjadi sejumlah fitur baru hasil transformasi kategori ke dalam bentuk biner.

Peningkatan jumlah fitur ini terjadi karena setiap kategori direpresentasikan sebagai kolom terpisah. Meskipun jumlah fitur bertambah, pendekatan ini dipilih untuk mempertahankan informasi dari setiap kategori sehingga dapat dimanfaatkan secara optimal oleh model *Random Forest* dalam proses klasifikasi.

3.2 SMOTE

Distribusi kelas pada fitur *Stress_Detection* menunjukkan adanya ketidakseimbangan. Kelas *Low* memiliki proporsi data yang lebih rendah dibandingkan kelas *Medium* dan *High*. Secara rinci, jumlah data pada kelas *Low* sebanyak 130 data, sedangkan kelas *Medium* dan *High* masing-masing berjumlah 245 dan 243 data. Kondisi ini mengindikasikan adanya potensi ketidakseimbang data.



Gambar 2. Hasil Perbandingan SMOTE

Hasil penerapan SMOTE menghasilkan komposisi kelas yang seragam, dengan total 245 data untuk setiap kategori yang ada. Terjadi peningkatan pada kelas *Low* sebesar 88,46% dari 130 menjadi 245 data, sementara kelas lainnya tetap pada jumlah yang sama. Penyeimbangan ini guna meminimalkan kecenderungan pada kategori dominan sekaligus memperkuat ketajaman model dalam mengidentifikasi pola pada kelas yang minoritas.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

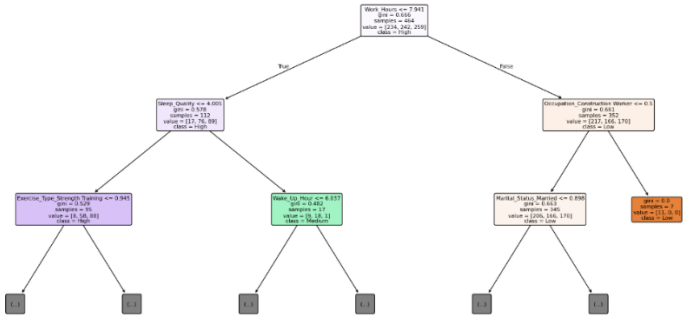
3.3 Penerapan Random Forest

Optimasi model dilakukan melalui *tuning hyperparameter* dengan metode *Grid Search* untuk menentukan konfigurasi parameter yang memberikan performa terbaik pada algoritma *Random Forest*. Hasil Pencarian parameter terbaik ini dipaparkan secara rinci pada Tabel 3.

Tabel 3. Hasil *Hyperparameter* Terbaik

Hyperparameter	Nilai Uji	Hasil Terbaik
n_estimators	100, 200, 300	300
max_depth	None, 10, 15	None
min_samples_leaf	1, 2, 3	1

Temuan ini menunjukkan bahwa algoritma Random Forest memanfaatkan sejumlah besar pohon keputusan sehingga proses pengambilan keputusan melalui kombinasi berbagai pohon yang terbentuk. Pendekatan ini memungkinkan model untuk mengenali pola dalam data secara lebih efektif dan menyeluruh. Di samping itu, tidak adanya pembatasan pada kedalaman pohon memberikan ruang yang lebih luas bagi model dalam menyusun aturan keputusan berdasarkan data yang digunakan dalam penelitian ini. Berlandaskan perpaduan parameter paling optimal yang diperoleh, visualisasi pada gambar 3 mendemonstrasikan mekanisme pasca optimasi parameter. Struktur tersebut memberikan gambaran bagaimana fitur-fitur diproses melalui serangkaian percabangan logika untuk menentukan kategori tingkat stres.



Gambar 3. Struktur Pohon Keputusan

Gambar 3 sebagai salah satu struktur pohon keputusan yang terbentuk dalam model *Random Forest*. Pada struktur ini, *Work Hours* menjadi pemisah utama, yang menunjukkan bahwa durasi jam kerja berperan penting dalam proses awal pengambilan keputusan model.

Pada cabang berikutnya, model mempertimbangkan fitur lain seperti *Sleep Quality, Exercise Type,* dan *Wake Up Hour*. Hal ini menunjukan setelah faktor jam kerja terpenuhi, kualitas istirahat menjadi penentu berikutnya. Sementara itu, pada cabang lainnya model juga mempertimbangkan faktor pekerjaan serta status pernikahan sebagai fitur tambahan dalam proses klasifikasi.

Secara keseluruhan, struktur pohon ini menunjukkan bahwa model *Random Forest* memanfaatkan perpaduan faktor pekerjaan, kualitas tidur, gaya hidup, dan kondisi sosial untuk mengidentifikasi pola yang berkaitan dengan tingkat stres individu.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

3.4 Evaluasi Model

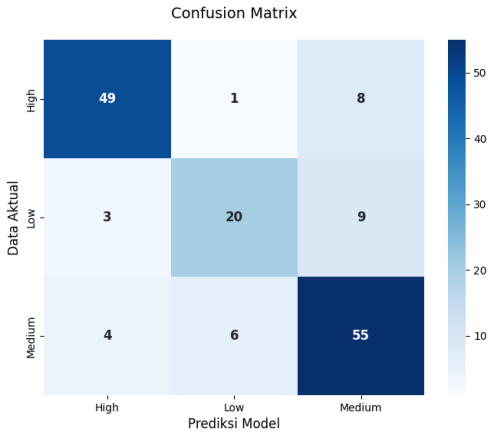
Model hasil optimasi *Random Forest* selanjutnya dievaluasi menggunakan data pengujian guna menilai kemampuan generalisasinya. Hasil pengujian membuktikan model memperoleh skor akurasi 0.80, hal ini mengindikasikan kemampuan klasifikasi yang cukup baik dalam memprediksi tingkat stres berdasarkan indikator gaya hidup dan kesehatan.

Tabel 4. Hasil Evaluasi Model

Stres_Detection	*Precision*	*Recall*	*F1-Score*
High	0.88	0.84	0.86
Low	0.74	0.62	0.68
Medium	0.76	0.85	0.80

Berdasarkan tabel 4, model mencapai performa optimal pada kelas *High (F1-Score* 0.86), unggul 18% dibandingkan kelas *Low* (0.68) yang mencatatkan performa terendah. Rendahnya capaian pada kelas *Low* dipicu oleh misklasifikasi terhadap kelas *Medium*, yang memiliki selisih performa 12% lebih tinggi.

Meskipun terdapat perbedaan hasil antar-kelas, rata-rata *precision, recall, dan F1-Score* di angka 0.80 membuktikan stabilitas model secara seimbang pada seluruh kategori. Gambar 4 menunjukkan hasil analisis yang dilakukan menggunakan *confusion matrix* untuk memahami pola kesalahan klasifikasi.

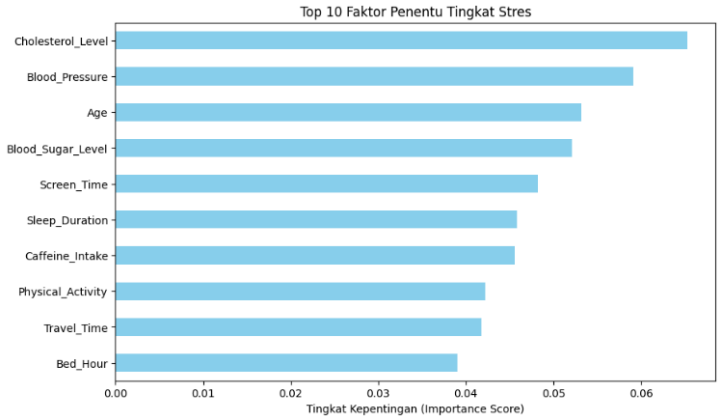


Gambar 4. Hasil Confusion Matrix

Berdasarkan Gambar 4, kesalahan klasifikasi yang terjadi umumnya berada pada kategori tingkat stres yang berdekatan, khususnya antara kelas *Low* dan *Medium*. Sementara itu, kesalahan klasifikasi antara kategori yang memiliki perbedaan tingkat stres yang jauh relatif jarang terjadi.

Pola ini menunjukkan bahwa model mampu mempertahankan batas pemisah antar kategori dengan cukup baik, sehingga kesalahan prediksi yang terjadi masih berada pada tingkat stres yang relatif berdekatan. Selanjutnya, fitur yang paling berpengaruh dalam prediksi tingkat stres oleh model *Random Forest* diidentifikasi melalui analisis fitur penting. Visualisasi tingkat kepentingan masing-masing fitur ditampilkan pada Gambar 5.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

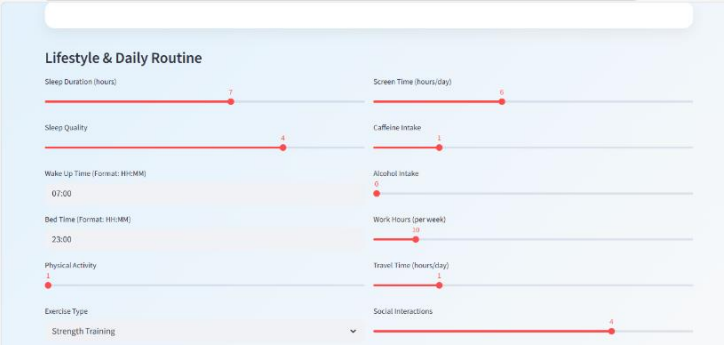


Gambar 5. Hasil Feature Importance

Berdasarkan hasil tersebut, Dominasi variabel kesehatan fisik pada peringkat atas menunjukkan kuatnya pola keterkaitan antara parameter fisiologis dengan tingkat stres dalam dataset ini. Meskipun demikian, faktor gaya hidup tetap memberikan kontribusi pada pemodelan ini, seperti fitur *Screen_Time, Sleep_Duration,* dan *Caffeine_Intake* menempati tingkat kepentingan menengah, diikuti oleh *Physical_Activity, Travel_Time,* serta *Bed_Hour*. Hal ini membuktikan kemampuan machine learning dalam mengidentifikasi pola kompleks dari berbagai dimensi data yang digunakan.

3.5 Integrasi Model

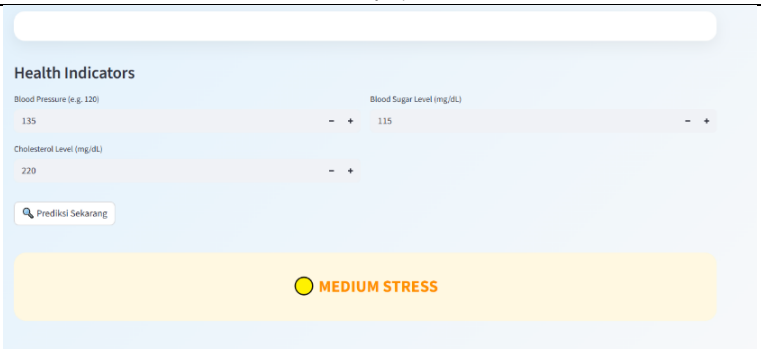
Model klasifikasi yang telah dibangun diintegrasikan ke dalam aplikasi web dengan Streamlit untuk mendukung pemanfaatan model secara interaktif. Pengujian prediksi dapat dilakukan secara langsung oleh pengguna dengan memasukkan indikator gaya hidup dan kesehatan melalui form input, seperti kategori usia, durasi tidur, aktivitas fisik, dan parameter kesehatan. Antarmuka aplikasi yang dirancang ditampilkan pada Gambar 6.



Gambar 6. Antarmuka Form Input Berbasis Streamlit

Setelah pengguna memasukkan nilai pada fitur yang tersedia, kemudian data tersebut dianalisis dengan model Random Forest yang telah dilatih sebelumnya. Proses ini menghasilkan prediksi kategori tingkat stres berdasarkan kombinasi indikator yang diberikan oleh pengguna. Hasil prediksi kategori tingkat stres yang dikeluarkan oleh model berdasarkan data input tersebut ditampilkan pada Gambar 7.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno



Gambar 7. Hasil Pengujian Menggunakan Streamlit

Integrasi ke dalam antarmuka Streamlit ini berfungsi sebagai sarana pengujian interaktif untuk memvalidasi respons model terhadap data input baru secara langsung. Hal tersebut membuktikan bahwa model *Random Forest* yang dibangun telah siap dioperasikan dalam lingkungan aplikasi dan menghasilkan output prediksi yang konsisten sesuai berdasarkan data input yang diberikan.

4. SIMPULAN

Berdasarkan hasil akhir penelitian, model *Random Forest* berhasil dikembangkan untuk memprediksi tingkat stres dengan tingkat akurasi sebesar 80%. Evaluasi model menunjukkan bahwa performa terbaik diperoleh pada klasifikasi tingkat stres kategori *High* dengan nilai *F1-Score* tertinggi sebesar 0.86. Meskipun terdapat perbedaan performa antar kategori, nilai rata-rata *precision, recall*, dan *F1-Score* yang konsisten di angka 0.80 mengindikasikan bahwa model mampu mengenali berbagai pola data secara seimbang dan merata.

Selain itu, hasil *feature importance* membuktikan bahwa indikator fisiologis dan kebiasaan gaya hidup, khususnya kolesterol dan tekanan darah, merupakan fitur yang paling berpengaruh dalam menentukan tingkat stres seseorang. Sejalan dengan hal tersebut, model yang telah dibangun berhasil diimplementasikan ke dalam antarmuka berbasis web menggunakan Streamlit sebagai sarana pengujian interaktif. Sistem terbukti mampu memproses input pengguna secara real-time dan menghasilkan prediksi yang konsisten, sehingga model ini dapat menjembatani kompleksitas algoritma *machine learning* ke dalam bentuk aplikasi yang fungsional dan sederhana.

REFERENCES

[1] L. Anggraeni, “Perspektif Global Kesehatan Mental Kaum Pemuda (Remaja, Adolesen & Dewasa Awal) Di Amerika Serikat, Eropa, Negara Persemakmuran & Asia Tenggara Tahun 2024: Sebuah Tinjauan Pustaka Sistematis,” *J. Ilmu Kesehat. Karya Bunda Husada*, vol. 10, no. 2, hal. 34–61, Nov 2024, doi: 10.56861/jikkbh.v10i2.140.

[2] F. Firmansyah dan A. Yulianto, “Pemodelan Pembelajaran Mesin untuk Prediksi Kesehatan Mental di Tempat Kerja,” *J. Minfo Polgan*, vol. 13, no. 1, hal. 397–407, 2024, doi: 10.33395/jmp.v13i1.13674.

[3] C. Nursadrina, K. Rahmayanti, dan M. Sofia, “PENGARUH STRES TERHADAP KESEHATAN MENTAL DAN FISIK : TINJAUAN PSIKOLOGI KLINIS The Impact Of Stress On Mental And Physical Health : A Clinical Psychology Review,” vol. 11, no. 1, hal. 701–705, 2025, doi: https://doi.org/10.33143/jhtm.v11i1.5288.

[4] R. Jannah dan H. Santoso, “Tingkat Stres Mahasiswa Mengikuti Pembelajaran Daring pada Masa

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Pandemi Covid-19," *J. Ris. dan Pengabdi. Masy.*, vol. 1, no. 1, hal. 130–146, Jan 2021, doi: 10.22373/jrpm.v1i1.638.

[5] H. Cai *dkk.*, "Prevalence of Suicidality in Major Depressive Disorder: A Systematic Review and Meta-Analysis of Comparative Studies," *Front. Psychiatry*, vol. 12, no. September, 2021, doi: 10.3389/fpsyt.2021.690130.

[6] R. Wajhillah, S. Bahri, dan A. Wibowo, "Komparasi Metode Machine Learning pada Diagnosa Gangguan Kejiwaan Depresi," *Syntax J. Inform.*, vol. 9, no. 1, hal. 26–31, 2020, doi: 10.35706/syji.v9i1.2050.

[7] L. Alzubaidi *dkk.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.

[8] A. Agustina, T. Tukino, B. Huda, dan E. Novalia, "Prediksi Volume Penjualan Gadget Berdasarkan Promo dan Channel Penjualan Menggunakan Random Forest," *JUSIFOR J. Sist. Inf. dan Inform.*, vol. 4, no. 1, hal. 85–91, 2025, doi: 10.70609/jusifor.v4i1.6962.

[9] M. Hussein Umar, R. Daud Antony Pangaribuan, R. Primadana, I. Budiawan, dan R. Pakpahan, "Penerapan Algoritma Random Forest Untuk Prediksi Tingkat Stres Dari Aktivitas Media Sosial," *J. Media Inform. [JUMIN]*, vol. 6, no. 6, hal. 3113–3112, 2025.

[10] M. A. Hamid, N. Anisa, F. Sains, dan U. S. Mulia, "Penerapan Algoritma Random Forest untuk Klasifikasi Depresi Berdasarkan Faktor Tekanan Kerja dan Kebiasaan Hidup," vol. 5, no. 1, hal. 46–52, 2025, doi: https://doi.org/10.46880/tamika.Vol5No1.pp46-52.

[11] S. S. A. Larasati, E. N. K. Dewi, B. H. Farhansyah, F. A. Bachtiar, dan F. Pradana, "Penerapan Decision Tree dan Random Forest dalam Deteksi Tingkat Stres Manusia Berdasarkan Kondisi Tidur," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 5, hal. 1043–1050, 2024, doi: 10.25126/jtiik.2024117993.

[12] A. I. Anissa dan A. Qoiriah, "Prediksi Tingkat Stres Berdasarkan Pola Hidup Menggunakan Machine Learning," *J. Informatics Comput. Sci.*, vol. 07, hal. 292–300, 2025.

[13] Juniardi, Tukino, Bayu Priyatna, dan Shofa Shofia Hilabi, "Klasifikasi Sentimen Analisis Ulasan Aplikasi Alfagift Menggunakan Algoritma Long Short Term Memory," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 4, no. 2, hal. 48–55, 2025, doi: 10.55123/storage.v4i2.5138.

[14] T. Gori, A. Sunyoto, dan H. Al Fatta, "Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, hal. 215–224, 2024, doi: 10.25126/jtiik.20241118074.

[15] M. A. Setiawan, M. Hasan Efendi, M. Farizal Akbar, dan W. Septian Pratama, "Pengaruh Teknik Preprocessing terhadap Kinerja Model Explainable Boosting Machine (EBM) untuk Prediksi Serangan Jantung," *J. Nusant. Eng.*, vol. 8, no. 2, hal. 317–326, 2025, [Daring]. Tersedia pada: https://ojs.unpkediri.ac.id/index.php/noe

[16] F. Aprilia, R. A. Anggraini, dan Y. D. Putri, "Prediksi Kelulusan Siswa dengan Algoritma Pembelajaran Mesin: Aplikasi Regresi Linear dan Logistik pada Faktor-Faktor Pendidikan," *ROUTERS J. Sist. dan Teknol. Inf.*, vol. 3, no. 1, hal. 55–64, 2025, doi: 10.25181/rt.v3i1.3897.

[17] F. Dwi Astuti dan F. Nova Lenti, "Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN," *J. JUPITER*, vol. 13, no. 1, hal. 89–98, 2021.

[18] A. Kurniawan, I., Hananto, A., Hilabi, S., Hananto, A., Priyatna, B., & Rahman, "Perbandingan Algoritma Naive Bayes Dan Svm Dalam Analisis," vol. 4, no. 1, hal. 392–400, 2025, doi: https://doi.org/https://doi.org/10.35957/jatisi.v10i1.3582.

[19] S. A. Putri dan R. Rachmatika, "Penerapan Algoritma Random Forest dan SMOTE untuk Prediksi Risiko Putus Sekolah Siswa Sekolah Menengah Kejuruan," *Decod. J. Pendidik. Teknol. Inf.*, vol. 5, no. 3, hal. 903–910, 2025, doi: 10.51454/decode.v5i3.1360.

[20] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, dan F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, hal. 677–690, 2022, doi: 10.30812/matrik.v21i3.1726.

[21] A. Syukron, S. Sardiarinto, E. Saputro, dan P. Widodo, "Penerapan Metode Smote Untuk Mengatasi

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 1, April 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung," *J. Teknol. Inf. dan Terap.*, vol. 10, no. 1, hal. 47–50, 2023, doi: 10.25047/jtit.v10i1.313.

[22] C. Cahyaningtyas, Y. Nataliani, dan I. R. Widiasari, "Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE," *Aiti*, vol. 18, no. 2, hal. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.

[23] W. N. Rohman dan I. M. A. Agastya, "Evaluation of SMOTE Technique in the Comparison of XGBoost and Random Forest Algorithms for Liver Disease Prediction," *J. Appl. Informatics Comput.*, vol. 9, no. 6, hal. 3157–3167, 2025, doi: 10.30871/jaic.v9i6.10239.

[24] L. Amaliana *dkk.*, "Dengan Voting Classifier Dan Random Forest Prediction of Death Risk in Chronic Kidney Failure Patients," vol. 12, no. 4, hal. 859–866, 2025.

[25] Z. Zuriati, D. K. Widyawati, O. Arifin, K. Saputra, S. Sriyanto, dan A. Ahmad, "Hybrid Machine Learning Approach for Nutrient Deficiency Detection in Lettuce," *TIERS Inf. Technol. J.*, vol. 6, no. 2, hal. 187–204, 2025, doi: 10.38043/tiers.v6i2.7143.