

Pendekatan *Random Forest* untuk Klasifikasi Tingkat *Burnout* Mahasiswa Berdasarkan Pola Aktivitas Harian

Random Forest Approach for Classifying Student Burnout Levels Based on Daily Activity Patterns

Tanaya Salsabila Arliansyah¹, Asep Budiman Kusdinar², Asriyanik³

^{1, 2, 3}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sukabumi, Sukabumi, Indonesia

E-mail : tanayasalsabilaa@ummi.ac.id¹, abkus016@ummi.ac.id², asriyanik263@ummi.ac.id³

*Corresponding author

Received 30 June 2026; Revised 1 August 2026; Accepted 14 August 2026

Abstrak – *Burnout* akademik merupakan kondisi kelelahan fisik, emosional, dan mental akibat tingginya tuntutan kuliah yang dapat menurunkan motivasi serta kualitas hidup mahasiswa. Oleh karena itu, diperlukan pendekatan yang mampu memberikan indikasi awal tingkat *burnout* secara lebih cepat dan objektif. Penelitian ini bertujuan membangun sistem klasifikasi tingkat *burnout* mahasiswa berdasarkan pola aktivitas harian menggunakan algoritma *Random Forest* dengan kerangka CRISP-DM. Data penelitian diperoleh melalui kuesioner daring dari 1.026 mahasiswa Universitas Muhammadiyah Sukabumi. Variabel prediktor meliputi jam tidur, waktu istirahat, intensitas *screen time*, keteraturan pola tidur, dan beban akademik, dengan target klasifikasi tingkat *burnout* rendah, sedang, dan tinggi. Tahap prapengolahan data meliputi transformasi data, pelabelan, penanganan *class imbalance* menggunakan SMOTE, serta *hyperparameter tuning* menggunakan *GridSearchCV*. Model divalidasi menggunakan *Stratified K-Fold Cross-Validation* dan dievaluasi dengan pembagian data 80:20. Hasil penelitian menunjukkan bahwa model *Random Forest* memperoleh akurasi 88,83%, dengan nilai *weighted average precision*, *recall*, dan *F1-score* masing-masing 0,89. Model berhasil diimplementasikan ke dalam aplikasi berbasis web menggunakan Hugging Face untuk memberikan indikasi awal tingkat *burnout* mahasiswa. Hasil tersebut menunjukkan bahwa algoritma *Random Forest* memiliki performa yang baik dalam mengklasifikasikan tingkat *burnout*.

Kata Kunci – *Burnout Akademik, Klasifikasi, Machine Learning, Pola Aktivitas Harian, Random Forest*

Abstract – *Academic burnout is a condition of physical, emotional, and mental exhaustion caused by high academic demands, which can reduce students' motivation and quality of life. Therefore, an approach is needed to provide an initial indication of burnout more quickly and objectively. This study aims to develop a student burnout classification system based on daily activity patterns using the Random Forest algorithm with the CRISP-DM framework. Research data were collected through an online questionnaire from 1,026 students at Muhammadiyah University of Sukabumi. The predictor variables included sleep duration, rest time, screen time intensity, sleep regularity, and academic workload, while burnout level was classified into low, moderate, and high categories. Data preparation included data transformation, labeling, class imbalance handling using SMOTE, and hyperparameter tuning using GridSearchCV. The model was validated using Stratified K-Fold Cross-Validation and evaluated using an 80:20 training-testing split. The Random Forest model achieved an accuracy of 88.83%, with weighted average precision, recall, and F1-score of 0.89. The model was implemented as a web-based application using Hugging Face to provide an initial indication of student burnout. These results indicate that the Random Forest algorithm performed well in classifying burnout levels.*

Keywords – *Academic Burnout, Classification, Daily Activity Patterns, Machine Learning, Random Forest*

1. PENDAHULUAN

Akselerasi teknologi digital pada ekosistem akademik berkorelasi linier terhadap eskalasi beban kognitif yang berpotensi memicu manifestasi *burnout* di kalangan mahasiswa. Berdasarkan taksonomi *World Health Organization* (WHO) dalam *International Classification of Diseases* edisi ke-11 (ICD-11), *burnout* dikonseptualisasikan sebagai sindrom psikologis yang mengonfirmasi kegagalan mitigasi stres kronis di lingkungan aktivitas, yang diidentifikasi melalui tiga dimensi utama, yaitu deplesi energi atau kelelahan emosional, peningkatan distansi mental secara sinis terhadap rutinitas, dan reduksi efikasi diri [1]. Secara klinis dan psikologis, tekanan akademis yang ekstrem terbukti mendegradasi regulasi emosi serta memicu deteriorasi kualitas tidur secara signifikan, khususnya pada fase krusial mahasiswa tingkat akhir [2]. Disregulasi waktu istirahat akibat tekanan tugas perkuliahan tersebut secara empiris menginduksi kelelahan fisik yang berimplikasi pada multiplikasi derajat stres akademis [3]. Di Indonesia, prevalensi fenomena ini menunjukkan status kedaruratan yang tinggi, merujuk pada analisis Rosmawati & Merida, tercatat sebanyak 66,5% mahasiswa tergolong dalam kategori *burnout* tingkat tinggi [4], selaras dengan temuan Arista et al. yang menempatkan mayoritas subjek studi pada rentang stres moderat hingga berat [5]. Kondisi empiris ini terkonfirmasi di Universitas Muhammadiyah Sukabumi melalui survei berbasis skala Likert terhadap 60 responden mahasiswa aktif, yang mengekstraksi data bahwa 65% di antaranya mengalami stres intensif akibat beban kurikulum. Realitas objektif tersebut menegaskan urgensi pergeseran paradigma diagnostik dari pendekatan subjektif menuju sistem klasifikasi prediktif yang objektif memanfaatkan algoritma *machine learning* berbasis pola variabilitas aktivitas harian.

Dalam ranah data mining dan komputasi prediktif, pemetaan tingkat *burnout* mahasiswa sejauh ini masih didominasi oleh metode deskriptif-retrospektif yang belum terintegrasi secara otomatis. Studi global oleh Yeskuatov et al. Membuktikan bahwa pemodelan prediktif yang hanya mengandalkan data administratif kampus menghasilkan akurasi yang rendah, yaitu berkisar antara 53,6% hingga 57,2% [6]. Rendahnya akurasi tersebut disebabkan oleh tidak dilibatkannya fitur perilaku harian (*daily behavior features*). Mengabaikan variabel krusial seperti kualitas tidur, waktu istirahat, dan durasi aktivitas harian terbukti menurunkan performa generalisasi model. Jika masalah ini dibiarkan, keterlambatan dalam memprediksi gejala dini berisiko menurunkan prestasi akademik, mengganggu kesehatan mental, hingga meningkatkan angka putus studi (*dropout*). Oleh karena itu, integrasi multivariabel pola aktivitas harian yang mencakup durasi tidur, waktu istirahat, intensitas *screen time*, konsisten pola tidur, dan beban akademik menjadi prediktor esensial karena mampu merepresentasikan kondisi fisik dan psikis mahasiswa. Guna mengekstraksi pengetahuan dari rangkaian data perilaku tersebut, metodologi Cross-Industry Standard Process for Data Mining (CRISP-DM) diimplementasikan sebagai kerangka kerja standar untuk menjalin siklus riset yang berjalan secara sistematis, adaptif, dan akurat [7].

Optimasi arsitektur kecerdasan buatan berbasis *machine learning* telah banyak digunakan untuk mengklasifikasikan variabel psikologis mahasiswa. Riset oleh Ananda Hapsari et al. Menerapkan variasi pengklasifikasi tunggal seperti Decision Tree, Logistic Regression, dan Random Forest untuk memetakan status kesehatan mental berdasarkan instrumen survei TMAS [8]. Selaras dengan itu, Nurdiansyah et al. mengombinasikan algoritma K-Nearest Neighbor (KNN) dan Random Forest untuk mendeteksi gangguan kecemasan serta memetakan kondisi kesehatan mental mahasiswa [9]. Dalam komparasi efisiensi algoritma untuk prediksi depresi, Setiawan et al. membuktikan keunggulan Random Forest yang mencatatkan akurasi tertinggi pada rentang 83,44%-83,90% dibandingkan dengan Decision Tree (82,52%-82,76%), Naïve Bayes (81,84%-83,28%), dan KNN (81,61%-83,15%) [10]. Karakteristik superior Random Forest dalam memproses data berdimensi besar bersumber dari kemampuannya mengolah fitur multivariabel nonlinier. Model ini juga menyajikan stabilitas performa yang lebih baik dibandingkan dengan model tunggal konvensional [11], [12]. Efektivitas algoritma yang mengadopsi prinsip *ensemble learning* ini juga terbukti konsisten dalam meminimalisasi risiko *overfitting* pada data yang memiliki variasi dan fluktuasi tinggi [13]. Di sisi lain, telaah konseptual mengenai manajemen

stres akademis di era kecerdasan buatan dipaparkan oleh Yusufi et al., namun pembahasannya masih terbatas pada teori tanpa adanya implementasi sistem klasifikasi otomatis [14]. Hibridisasi model juga pernah diuji oleh Oktaviani et al. memanfaatkan integrasi Random Forest dan SMOTE untuk mendeteksi tingkat stres mahasiswa tingkat akhir, tetapi fiturnya masih terikat pada parameter kuesioner psikologi umum [15]. Berdasarkan tinjauan literatur tersebut, kelemahan mendasar model klasifikasi tunggal terletak pada tingginya kerentanan terhadap bias data mahasiswa yang fluktuatif. Selain itu, belum ada penelitian terdahulu yang menggunakan variabel aktivitas harian spesifik secara komprehensif sebagai fitur masukan utama. Celah metodologi inilah yang melandasi nilai kebaruan (*novelty*) dan orisinalitas dalam penelitian ini.

Sebagai solusi ilmiah, penelitian ini mengusulkan implementasi algoritma Random Forest untuk mengklasifikasikan tingkat *burnout* mahasiswa berdasarkan rekam aktivitas harian. Pemilihan algoritma ini didasarkan pada kapabilitas arsitekturnya yang andal dalam memproses data multivariabel dan nonlinier secara bersamaan. Melalui mekanisme penggabungan pohon keputusan (*ensemble learning via bagging*), algoritma ini mampu mengatasi efek noise pada data kuesioner mandiri, sehingga menghasilkan estimasi klasifikasi yang stabil meskipun pola hidup mahasiswa tidak teratur. Penelitian ini bertujuan untuk mengekstraksi dataset aktivitas harian dari minimal 1.000 mahasiswa aktif Universitas Muhammadiyah Sukabumi melalui kuesioner digital. Model prediktif yang dihasilkan kemudian akan diterapkan (*deployment*) ke dalam aplikasi berbasis web menggunakan platform Hugging Face. Kemudian, reliabilitas model akan dievaluasi secara matematis melalui metrik pengujian *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil klasifikasi yang terbagi menjadi tiga kategori (rendah, sedang, dan tinggi) diproyeksikan dapat berfungsi sebagai sistem pendukung atau peringatan dini (*early warning system*) yang presisi untuk mendukung program pembinaan mahasiswa serta merumuskan rekomendasi pencegahan yang relevan.

2. METODE PENELITIAN

Penelitian ini menerapkan pendekatan *machine learning* untuk mengklasifikasikan *burnout* mahasiswa berdasarkan variabilitas data aktivitas harian. Pengembangan model dan pengolahan data dikendalikan secara kronologis menggunakan kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) [7]. Alur pelaksanaan eksperimen ini dilakukan dalam beberapa tahapan sistematis pada Gambar 1.



Gambar 1. Alur Metodologi CRISP-DM

2.1. Business Understanding

Tahap awal ini difokuskan untuk mengidentifikasi permasalahan utama penelitian, yaitu tingginya risiko kelelahan mental pada mahasiswa akibat pola manajemen aktivitas harian yang tidak seimbang. Solusi yang diterapkan adalah membangun sistem klasifikasi otomatis yang

mampu memetakan derajat *burnout* mahasiswa secara objektif ke dalam tiga kategori (rendah, sedang, dan tinggi) menggunakan algoritma Random Forest.

2.2. Data Understanding

Tahap selanjutnya dilakukan untuk mengumpulkan sekaligus memahami karakteristik data yang diperoleh melalui kuesioner daring. Data penelitian mencakup variabel *prediktor* (X) dan variabel target (Y). Skala pengukuran beserta rentang nilai untuk setiap variabel dijabarkan pada Tabel 1.

Tabel 1. Skala Pengukuran Variabel

Variabel	Skala	Rentang Nilai	Keterangan
Jam tidur (X)	Ordinal	1–4	Durasi tidur harian mahasiswa
Waktu istirahat (X)	Ordinal	1–4	Lama waktu istirahat dari aktivitas akademik
Intensitas <i>screen time</i> (X)	Ordinal	1–4	Tingkat penggunaan perangkat digital
Keteraturan pola tidur (X)	Ordinal (Likert)	1–4	Konsistensi waktu tidur
Beban akademik (X)	Ordinal (Likert)	1–4	Tingkat persepsi responden terhadap volume tugas dan tuntutan tanggung jawab akademik harian.
Tingkat <i>burnout</i> (Y)	Ordinal (Likert)	1–4	Frekuensi mahasiswa merasa stres atau <i>burnout</i>

Seluruh variabel diukur menggunakan skala ordinal dengan rentang nilai 1-4. Variabel keteraturan pola tidur, beban akademik, dan tingkat *burnout* diukur menggunakan skala Likert untuk menggambarkan persepsi responden, sedangkan variabel jam tidur, waktu istirahat, dan intensitas *screen time* diukur menggunakan kategori ordinal yang merepresentasikan kondisi aktual aktivitas harian mahasiswa.

Instrumen penelitian disusun dalam bentuk kuesioner yang memuat enam butir pertanyaan sesuai dengan variabel penelitian. Setiap pertanyaan digunakan untuk memperoleh informasi mengenai pola aktivitas harian mahasiswa dan tingkat stres atau *burnout*. Berikut instrumen penelitian pada Tabel 2.

Tabel 2. Instrumen Kuesioner

No	Variabel	Pertanyaan
1	Jam tidur (X)	Jam tidur rata-rata per hari.
2	Waktu istirahat (X)	Waktu istirahat rata-rata per hari.
3	Intensitas <i>screen time</i> (X)	<i>Screen time</i> rata-rata per hari (HP, laptop, TV)
4	Keteraturan pola tidur (X)	Keteraturan pola tidur.
5	Beban akademik (X)	Beban akademik yang tinggi (seperti banyaknya tugas kuliah, praktikum, atau proyek kelompok) menyita hampir seluruh rutinitas harian saya dari pagi hingga malam.
6	Tingkat <i>burnout</i> (Y)	Seberapa sering merasa stres atau <i>burnout</i> ?

Berdasarkan target (Y) yang diperoleh dari data pertanyaan mengenai frekuensi responden yang merasa *burnout* akibat aktivitas akademik. Pengukuran tersebut menggunakan pendekatan self-assessment berdasarkan persepsi responden dan tidak merujuk pada instrumen pengukuran klinis, seperti *Maslach Burnout Inventory* (MBI). Oleh karena itu, hasil klasifikasi pada penelitian

ini hanya digunakan sebagai indikasi awal Tingkat *burnout* dan tidak dimaksudkan sebagai diagnosis klinis

2.3. Data Preparation

Tahap *data preparation* dilakukan untuk menyiapkan data hasil kuesioner sebelum proses pemodelan menggunakan algoritma Random Forest. Data yang diperoleh dari Google Form diekspor ke Microsoft Excel, kemudian dilakukan *data cleaning* dengan menghapus atribut yang tidak digunakan serta memeriksa adanya *missing value*.

Selanjutnya, dilakukan transformasi skala pada variabel jam tidur, waktu istirahat, dan intensitas *screen time* dengan mengubah data kategorikal menjadi skala ordinal 1-4. Variabel keteraturan pola tidur, beban akademik, dan tingkat *burnout* tidak ditransformasikan karena telah menggunakan skala Likert 1-4. Kategori pada ketiga variabel tersebut disusun berdasarkan penelitian terdahulu yang relevan serta disesuaikan dengan kebutuhan penelitian sehingga setiap rentang kategori tidak saling tumpang tindih. Aturan transformasi skalanya ditunjukkan pada Tabel 3.

Tabel 3. Aturan Transformasi Skala Data

No	Jam Tidur	Waktu Istirahat	Intensitas <i>Screen Time</i>
1	7-9 jam	> 9 jam	< 3 jam
2	> 9 jam	7-9 jam	3-6 jam
3	4-6 jam	4-6 jam	7-10 jam
4	< 4 jam	< 4 jam	> 10 jam

Kemudian, dilakukan pelabelan pada variabel target dengan mengelompokkan skor *burnout* menjadi tiga kategori, yaitu rendah (1-2), sedang (3), dan tinggi (4) berdasarkan jawaban pada skala Likert 1-4. Pengelompokan tersebut merupakan operasionalisasi peneliti berdasarkan frekuensi *burnout* yang ditunjukkan pada setiap pilihan jawaban, sehingga skor 1 (Tidak Pernah) dan 2 (Jarang) dikategorikan sebagai *burnout* rendah, skor 3 (Sangat Sering) sebagai *burnout* sedang, dan skor 4 (Sangat Sering) sebagai *burnout* tinggi. Kriteria pelabelan pada Tabel 4.

Tabel 4. Kriteria Pelabelan Tingkat *Burnout*

Rentang Skor	Kategori	Keterangan
1-2	Rendah	Mengindikasikan bahwa responden jarang atau tidak pernah merasa <i>burnout</i> akibat aktivitas akademik.
3	Sedang	Mengindikasikan bahwa responden mulai sering merasa <i>burnout</i> akibat aktivitas akademik.
4	Tinggi	Mengindikasikan bahwa responden sangat sering merasa <i>burnout</i> akibat aktivitas akademik.

Selanjutnya, pembagian dataset menggunakan rasio 80:20, yaitu 80% sebagai data latih (*training data*) dan 20% sebagai data uji (*testing data*) menggunakan metode *stratified sampling* untuk mempertahankan proporsi setiap kelas pada kedua dataset. Setelah itu, dilakukan penanganan ketidakseimbangan kelas menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE) pada data latih. Metode ini menghasilkan sampel sintesis pada kelas minoritas sehingga distribusi data antarkelas menjadi lebih seimbang. Proses oversampling hanya diterapkan pada data latih, sedangkan data uji tetap menggunakan distribusi data asli agar evaluasi model dilakukan secara objektif.

2.4. Modeling

Tahap modelling dilakukan dengan membangun model klasifikasi menggunakan algoritma Random Forest berdasarkan data latih. Sebelum model dibangun, dilakukan proses *hyperparameter tuning* menggunakan *GridSearchCV* untuk memperoleh kombinasi parameter

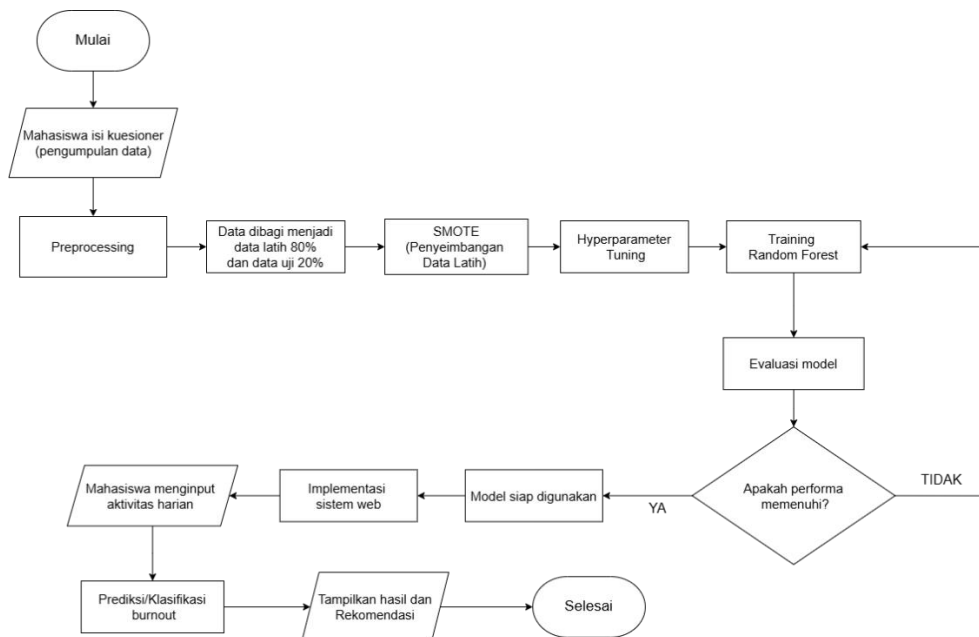
yang optimal. Parameter yang diuji meliputi $n_estimator$, max_depth , dan $max_features$. Proses pencarian parameter terbaik menggunakan *Stratified K-Fold Cross Validation* dengan lima *fold*. Kombinasi parameter yang diperoleh dari proses *hyperparameter tuning* kemudian digunakan untuk membangun model Random Forest. Selanjutnya, dilakukan *Stratified K-Fold Cross Validation* sebanyak lima *fold* pada data latih untuk mengevaluasi kemampuan generalisasi model sebelum dilakukan pengujian menggunakan data uji.

2.5. Evaluation

Tahap evaluasi bertujuan untuk mengukur performa model Random Forest dalam mengklasifikasikan tingkat *burnout* mahasiswa menggunakan data uji. Hasil prediksi divisualisasikan melalui *Confusion Matrix* untuk membandingkan kelas aktual dengan kelas prediksi. Selanjutnya, kinerja model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* guna menilai performa klasifikasi secara menyeluruh.

2.6. Deployment

Tahap *deployment* dilakukan dengan mengimplementasikan model Random Forest ke dalam sistem berbasis web menggunakan platform Hugging Face. Sistem menerima masukan berupa data pola aktivitas harian mahasiswa dalam skala ordinal dan Likert 1-4, kemudian memproses data menggunakan model untuk menghasilkan klasifikasi tingkat *burnout* ke dalam kategori rendah, sedang, atau tinggi. Selain hasil klasifikasi, sistem juga menampilkan informasi umum mengenai *burnout* beserta rekomendasi pencegahan. Alur implementasi sistem ditunjukkan pada Gambar 2.



Gambar 2. Alur Sistem Klasifikasi Tingkat *Burnout* Mahasiswa

Berdasarkan Gambar 2, proses sistem diawali dengan pengumpulan data melalui kuesioner, dilanjutkan dengan tahap *data preparation* yang meliputi *data cleaning*, transformasi skala, dan pelabelan data. Dataset kemudian dibagi menjadi data latih dan data uji untuk proses pelatihan serta evaluasi model Random Forest. Model yang telah memenuhi kriteria evaluasi selanjutnya diimplementasikan ke dalam sistem berbasis web sehingga mahasiswa dapat melakukan klasifikasi tingkat *burnout* berdasarkan aktivitas harian dan memperoleh hasil klasifikasi beserta informasi umum mengenai *burnout*.

3. HASIL DAN PEMBAHASAN

3.1. Business Understanding

Tahap *Business Understanding* diawali dengan mengidentifikasi permasalahan tingginya risiko *burnout* pada mahasiswa yang dipengaruhi oleh pola aktivitas harian, meliputi jam tidur, waktu istirahat, intensitas *screen time*, keteraturan pola tidur, dan beban akademik. Berdasarkan permasalahan tersebut, penelitian ini bertujuan membangun model klasifikasi menggunakan algoritma Random Forest untuk mengelompokkan tingkat *burnout* mahasiswa ke dalam kategori rendah, sedang, dan tinggi. Model yang dikembangkan selanjutnya diimplementasikan ke dalam sistem berbasis web untuk membantu proses identifikasi tingkat *burnout* berdasarkan data aktivitas harian mahasiswa.

3.2. Data Understanding

Data penelitian diperoleh melalui penyebaran kuesioner secara daring menggunakan Google Form kepada mahasiswa Universitas Muhammadiyah Sukabumi. Kuesioner disusun untuk memperoleh informasi mengenai pola aktivitas harian mahasiswa yang berkaitan dengan tingkat *burnout*. Berdasarkan hasil pengumpulan data, diperoleh sebanyak 1.026 responden yang digunakan sebagai dataset penelitian. Gambar 3 menunjukkan kuesioner yang digunakan untuk memperoleh data pola aktivitas harian mahasiswa.

Gambar 3. Kuesioner Penelitian Menggunakan Google Form

Dataset penelitian terdiri atas lima variabel *prediktor*, yaitu jam tidur, waktu istirahat, intensitas *screen time*, keteraturan pola tidur, dan beban akademik, serta satu variabel target, yaitu tingkat *burnout*. Variabel *prediktor* digunakan sebagai fitur dalam proses klasifikasi, sedangkan variabel target digunakan sebagai kelas yang diprediksi oleh algoritma Random Forest. Variabel riset penelitian disajikan pada Tabel 5.

Tabel 5. Variabel Penelitian

No	Variabel	Jenis Variabel	Keterangan
1	Jam Tidur	Prediktor (X)	Durasi tidur harian mahasiswa
2	Waktu Istirahat	Prediktor (X)	Lama waktu istirahat dari aktivitas
3	Intensitas <i>Screen Time</i>	Prediktor (X)	Tingkat penggunaan perangkat digital per hari
4	Keteraturan Pola Tidur	Prediktor (X)	Konsistensi jadwal tidur mahasiswa
5	Beban Akademik	Prediktor (X)	Tingkat beban akademik yang dirasakan mahasiswa
6	Tingkat <i>Burnout</i>	Target (Y)	Tingkat <i>burnout</i> yang menjadi kelas klasifikasi

Hasil Tabel 5 menunjukkan bahwa dataset menggunakan lima variabel yang merepresentasikan pola aktivitas harian mahasiswa sebagai fitur prediksi, sedangkan tingkat *burnout* digunakan sebagai variabel target. Dataset kemudian diekspor ke dalam format Microsoft

Excel untuk memudahkan proses prapengolahan data. Hasil pemeriksaan awal menunjukkan bahwa struktur data telah sesuai dengan variabel penelitian sehingga dataset dapat digunakan pada tahap *data preparation*. Tampilan dataset hasil ekspor pada Gambar 4.

Timestamp	Jam tidur rata-rata per hari	Waktu istirahat total per hari	Screen time rata-rata per hari (HP, laptop, TV)	Keteraturan pola tidur	Beban akademik yang tinggi (seperti banyak tugas kuliah, praktikum, atau proyek kelompok) menyita hampir seluruh rutinitas harian saya dari pagi hingga malam.	Seberapa sering merasa stres atau burnout?
1/15/2026 19:53:33	7-9 jam	7-9 jam	3-6 jam	2	1	2
1/15/2026 19:54:51	4-6 jam	4-6 jam	3-6 jam	2	3	4
1/15/2026 19:55:27	4-6 jam	> 10 jam	> 10 jam	2	3	4
1/15/2026 19:58:06	7-9 jam	> 10 jam	> 10 jam	2	2	2
1/15/2026 20:03:04	4-6 jam	> 10 jam	> 10 jam	1	4	4
1/15/2026 20:30:41	4-6 jam	4-6 jam	3-6 jam	2	2	3
1/15/2026 20:45:00	< 4 jam	3-6 jam	3-6 jam	2	3	3
1/15/2026 20:59:33	4-6 jam	3-6 jam	3-6 jam	2	3	3
1/15/2026 21:02:57	4-6 jam	3-6 jam	3-6 jam	3	2	4
1/15/2026 21:07:15	< 4 jam	3-6 jam	3-6 jam	2	3	2
1/15/2026 21:47:49	< 4 jam	7-10 jam	7-10 jam	3	3	4
1/16/2026 5:48:23	< 4 jam	7-10 jam	7-10 jam	2	4	4
1/16/2026 6:59:44	4-6 jam	3-6 jam	3-6 jam	3	3	3

Gambar 4. Dataset Hasil Ekspor Kuesioner

3.3. Data Preparation

Tahap *data preparation* dilakukan untuk mempersiapkan dataset agar sesuai dengan kebutuhan algoritma Random Forest. Proses ini meliputi pembersihan data, transformasi skala data, pelabelan variabel target, dan pembagian dataset.

a. Data Cleaning

Tahap *data cleaning* diawali dengan menghapus atribut *Timestamp* karena hanya berisi informasi waktu pengisian kuesioner dan tidak digunakan sebagai variabel penelitian. Dataset setelah penghapusan atribut tersebut ditunjukkan pada Gambar 5.

Jam tidur rata-rata per hari	Waktu istirahat total per hari	Screen time rata-rata per hari (HP, laptop, TV)	Keteraturan pola tidur	Beban akademik yang tinggi (seperti banyaknya tugas kuliah, praktikum, atau proyek kelompok) menyita hampir seluruh rutinitas harian saya dari pagi hingga malam.	Seberapa sering merasa stres atau burnout?
4-6 jam	7-9 jam	7-10 jam	2	1	1
7-9 jam	7-9 jam	7-10 jam	2	3	3
> 9 jam	> 9 jam	< 3 jam	3	3	1
4-6 jam	4-6 jam	> 10 jam	3	4	3
< 4 jam	< 4 jam	> 10 jam	1	4	3
4-6 jam	4-6 jam	3-6 jam	2	4	3
7-9 jam	> 9 jam	3-6 jam	3	3	3
4-6 jam	4-6 jam	3-6 jam	2	3	3
7-9 jam	7-9 jam	3-6 jam	2	2	2

Gambar 5. Dataset Sebelum Data Cleaning

Selanjutnya, dilakukan pemeriksaan missing value menggunakan Python untuk memastikan dataset lengkap. Hasil pemeriksaan pada Gambar 6 menunjukkan bahwa tidak terdapat data kosong sehingga seluruh 1.026 data responden dapat digunakan pada tahap berikutnya.

```

--- [A] PROSES DATA CLEANING ---
• Jumlah data kosong terdeteksi      : 0 baris (Dihapus)
• Validasi Jumlah Data Bersih        : Dari 1026 menjadi 1026 responden.

```

Gambar 6. Hasil Data Cleaning

b. Transformasi Skala Data

Tahap transformasi dilakukan dengan mengubah data kategorikal menjadi nilai numerik menggunakan skala 1-4 serta menyesuaikan nama variabel agar lebih ringkas dan konsisten. Berdasarkan hasil transformasi pada Gambar 7, seluruh variabel telah berada dalam format numerik sehingga memenuhi kebutuhan masukan algoritma Random Forest.

```
>>> 5 Data Pertama Setelah Transformasi:
```

	jam_tidur	waktu_istirahat	screen_time	keteraturan_tidur	beban_akademik	skala_burnout
0	3	2	3	2	1	1
1	1	2	3	2	3	3
2	2	1	1	3	3	1
3	3	3	4	3	4	3
4	4	4	4	1	4	3

Gambar 7. Dataset Setelah Transformasi Skala Data

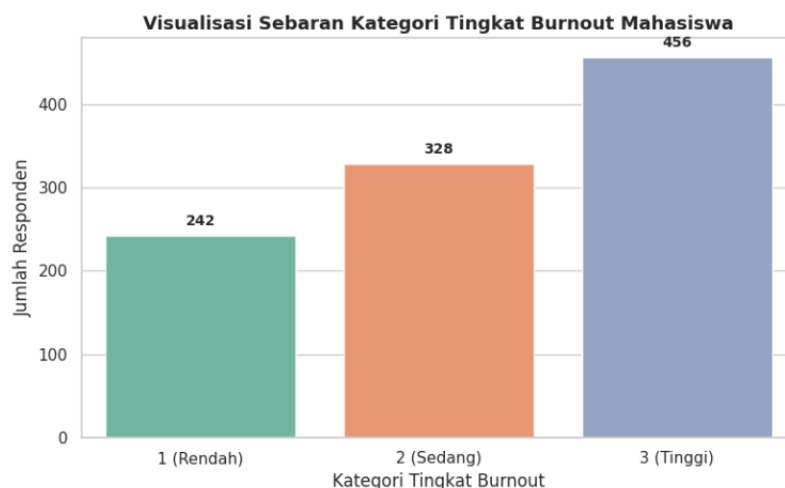
c. Labeling Data

Tahap labeling dilakukan untuk membentuk variabel target yang akan digunakan dalam proses klasifikasi. Pelabelan disesuaikan dengan jawaban responden pada variabel tingkat *burnout* yang diukur menggunakan skala Likert 1-4. Skor 1-2 dikelompokkan sebagai kategori *burnout* rendah, skor 3 sebagai *burnout* sedang, dan skor 4 sebagai *burnout* tinggi. Kriteria pelabelan yang digunakan ditunjukkan pada Tabel 6.

Tabel 6. Kriteria Pelabelan Tingkat *Burnout*

Kategori Target	Tingkat <i>Burnout</i>	Rentang Skor Rata-rata	Jumlah responden	Persentase
1	Rendah	1-2 (Tidak Pernah - Jarang)	242	23,59%
2	Sedang	3 (Sering)	328	31,97%
3	Tinggi	4 (Sangat Sering)	456	44,44%
Total			1.026	100%

Berdasarkan proses pelabelan, diperoleh tiga kategori tingkat *burnout*, yaitu. 242 responden pada kategori rendah, 328 responden pada kategori sedang, dan 456 responden pada kategori tinggi, sebagaimana ditunjukkan pada Gambar 8. Distribusi tersebut menunjukkan bahwa jumlah data pada setiap kategori belum seimbang (*class imbalance*), sehingga diperlukan penanganan pada tahap berikutnya dari proses pemodelan.



Gambar 8. Distribusi Kategori Tingkat *Burnout*

d. Pembagian Data

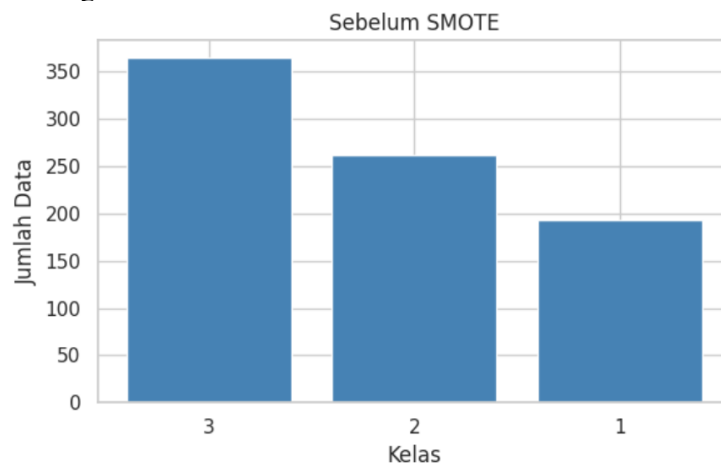
Dataset selanjutnya dibagi menggunakan rasio 80:20 menjadi 820 data latih dan 206 data uji, seperti disajikan pada Tabel 7. Sebagian besar data digunakan untuk melatih model, sedangkan data uji dimanfaatkan untuk mengevaluasi kemampuan model dalam mengklasifikasikan data yang belum pernah diproses pada tahap pelatihan.

Tabel 7. Pembagian Dataset

Jenis Data	Jumlah
Data Latih	820
Data Uji	206
Total Data	1.026

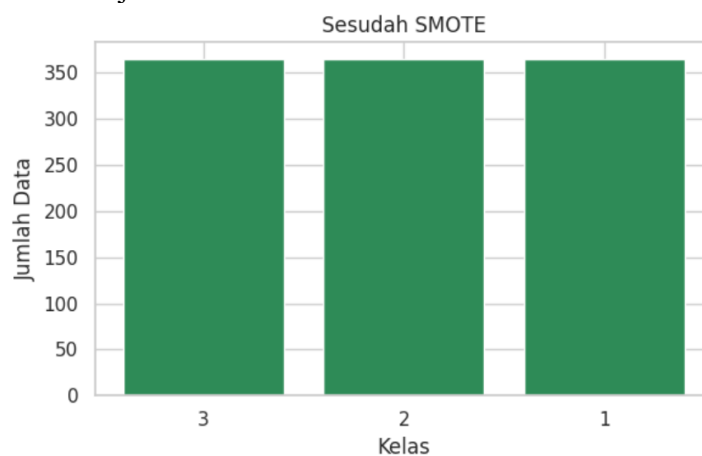
e. Penanganan *Class Imbalance* (SMOTE)

Setelah pembagian dataset, dilakukan pemeriksaan distribusi kelas pada data latih. Hasil pemeriksaan menunjukkan bahwa jumlah data pada setiap kategori burnout belum seimbang, sehingga diterapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) pada data latih untuk menghasilkan distribusi kelas yang lebih seimbang.



Gambar 9 Jumlah Data Sebelum SMOTE

Setelah proses SMOTE dilakukan, jumlah data pada setiap kategori menjadi seimbang, yaitu masing-masing sebanyak 365 data, sehingga total data latih meningkat dari 820 menjadi 1.095 data.



Gambar 10 Jumlah Data Sesudah SMOTE

Penerapan SMOTE bertujuan mengurangi bias model terhadap kelas mayoritas sehingga model Random Forest dapat mempelajari pola pada setiap kategori *burnout* secara lebih seimbang sebelum memasuki tahap modeling.

3.4. Modeling

Tahap *modeling* dilakukan menggunakan algoritma Random Forest berdasarkan data latih yang telah melalui proses *data preparation*. Sebelum model dibangun, dilakukan *hyperparameter tuning* menggunakan *GridSearchCV* untuk memperoleh kombinasi parameter yang menghasilkan performa terbaik. Parameter yang diuji meliputi $n_estimator = 100$, $max_depth = 5$, $max_features = 'sqrt'$, dan $random_state = 42$. Konfigurasi parameter tersebut digunakan untuk menghasilkan model yang stabil, mengurangi risiko *overfitting*, serta memastikan hasil pelatihan konsisten. Parameter yang digunakan ditunjukkan pada Tabel 8.

Tabel 8. Parameter Pengujian Model Random Forest

No	Parameter	Nilai Parameter	Keterangan
1	$n_estimators$	100	Jumlah pohon keputusan (<i>decision tree</i>) yang dibangun.
2	max_depth	5	Kedalaman maksimum pohon untuk mencegah <i>overfitting</i> .
3	$max_features$	'sqrt'	Jumlah fitur acak yang dipertimbangkan di setiap <i>split mode</i> .
4	$random_state$	42	Pengunci keacakan agar hasil selalu konsisten

Selanjutnya, model divalidasi menggunakan metode *Stratified K-Fold Cross Validation* sebanyak 5 *fold* pada data latih untuk mengevaluasi kemampuan generalisasi model sebelum dilakukan pengujian menggunakan data uji. Hasil validasi silang ditunjukkan pada Tabel 9.

Tabel 9 Hasil Pengujian *K-Fold Cross Validation* Performa Model Klasifikasi

Metrik Evaluasi	<i>Fold 1</i>	<i>Fold 2</i>	<i>Fold 3</i>	<i>Fold 4</i>	<i>Fold 5</i>	Rata-rata
<i>Accuracy</i>	0.8311	0.8721	0.8082	0.8493	0.8311	0.8384
<i>Precision</i>	0.8454	0.8786	0.8235	0.8581	0.8408	0.8493
<i>Recall</i>	0.8311	0.8721	0.8082	0.8493	0.8311	0.8384
<i>F1-score</i>	0.8324	0.8722	0.8094	0.8484	0.8332	0.8391
<i>Accuracy</i>	0.8311	0.8721	0.8082	0.8493	0.8311	0.8384
Hasil <i>Final Training</i> Model	0.8457					

Hasil *Stratified K-Fold Cross Validation* pada Tabel 9, model memperoleh rata-rata *accuracy* sebesar 83,84%, *precision* sebesar 84,93%, *recall* sebesar 83,84%, dan *F1-score* sebesar 83,91%. Selanjutnya, model dilatih menggunakan seluruh data latih dengan kombinasi parameter terbaik sehingga diperoleh training *accuracy* sebesar 84,57%. Model hasil pelatihan kemudian digunakan pada tahap evaluasi untuk mengukur performa klasifikasi menggunakan data uji.

3.5. Evaluation

Tahap evaluasi dilakukan menggunakan 206 data uji (20%) untuk mengukur performa model Random Forest. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta *Confusion Matrix* untuk membandingkan hasil prediksi dengan kelas sebenarnya. Hasil evaluasi yang dilakukan adalah sebagai berikut.

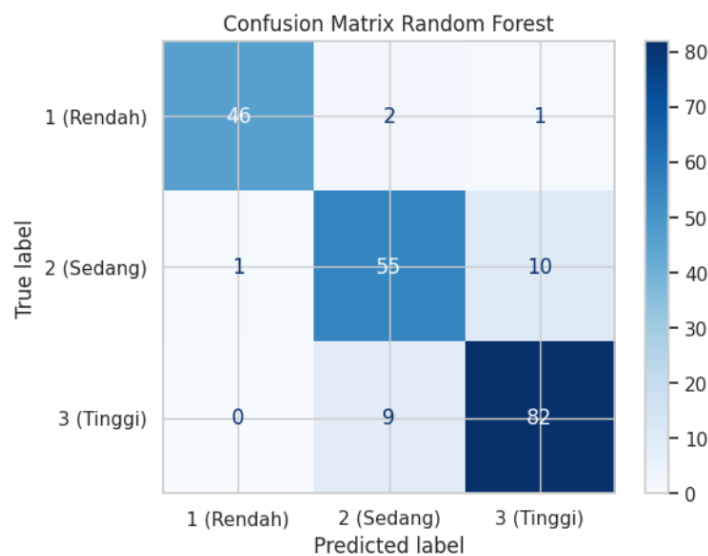
Tabel 10. Hasil Evaluasi Metrik Klasifikasi Model Random Forest

Kelas Target/Label	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
1 (Rendah)	0,98	0,94	0,96	49
2 (Sedang)	0,83	0,83	0,83	66

3 (Tinggi)	0,88	0,90	0,89	91
Macro Avg	0,90	0,89	0,89	206
Weighted Avg	0,89	0,89	0,89	206
Akurasi Akhir Model				88,83%

Model pada Tabel 10 memperoleh *accuracy* sebesar 88,83% dengan nilai *weighted average precision*, *recall*, dan *F1-score* masing-masing sebesar 0,89, menunjukkan bahwa model memiliki performa klasifikasi yang baik pada data uji dengan mempertimbangkan proporsi jumlah data pada setiap kategori *burnout*.

Hasil *Stratified K-Fold Cross Validation* pada tahap modeling memperoleh rata-rata *accuracy* sebesar 83,84%. Meskipun terdapat variasi nilai *accuracy* pada setiap *fold*, perbedaan tersebut merupakan hal yang wajar karena setiap *fold* menggunakan pembagian data yang berbeda. Secara keseluruhan, hasil tersebut menunjukkan bahwa model memiliki kemampuan generalisasi yang baik sebelum dilakukan pengujian menggunakan data uji.

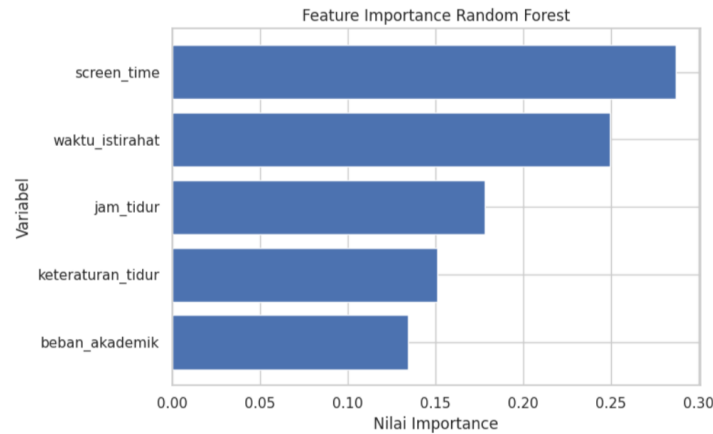


Gambar 11. Confusion Matrix Hasil Klasifikasi Random Forest

Hasil *Confusion Matrix* pada Gambar 12 menunjukkan bahwa model mengklasifikasikan 46 data pada kategori *burnout* rendah, 55 data pada kategori *burnout* sedang, dan 82 data pada kategori *burnout* tinggi secara benar. Kesalahan klasifikasi masih terjadi pada beberapa data, terutama antara kategori *burnout* sedang dan tinggi. Hal tersebut menunjukkan bahwa kedua kategori memiliki karakteristik pola aktivitas harian yang relatif berdekatan sehingga lebih sulit dibedakan oleh model dibandingkan dengan kategori *burnout* rendah.

Hasil penelitian ini memperoleh akurasi yang lebih tinggi dibandingkan dengan penelitian Setiawan et al. (2025) yang melaporkan akurasi 83,44%-83,80% menggunakan algoritma Random Forest [10]. Perbedaan tersebut kemungkinan dipengaruhi oleh karakteristik dataset, variabel yang digunakan, serta tahapan prapengolahan data, termasuk proses transformasi dan pelabelan variabel target dan penanganan ketidakseimbangan kelas menggunakan SMOTE. Dengan demikian, performa model tidak hanya dipengaruhi oleh algoritma yang digunakan, tetapi juga oleh kualitas dan karakteristik data yang digunakan dalam proses pelatihan.

Selain mengevaluasi performa klasifikasi, penelitian ini juga menganalisis *feature importance* untuk mengetahui tingkat kontribusi masing-masing variabel dalam proses pembentukan model Random Forest. Hasil analisis *feature importance* ditunjukkan sebagai berikut.



Gambar 12 Tingkat Kepentingan Variabel (*Feature importance*)

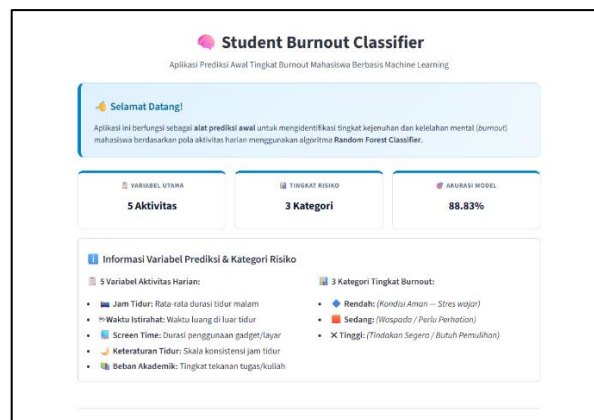
Berdasarkan Gambar 12, variabel *screen time* memiliki nilai *feature importance* tertinggi sebesar 0,286650, diikuti oleh waktu istirahat sebesar 0,249248, jam tidur sebesar 0,178421, keteraturan pola tidur sebesar 0,151030, dan beban akademik sebesar 0,134651. Hasil tersebut menunjukkan bahwa *screen time* merupakan variabel yang paling berkontribusi dalam proses klasifikasi tingkat *burnout* mahasiswa. Sementara itu, variabel lainnya juga memberikan kontribusi terhadap pembentukan model meskipun dengan tingkat pengaruh yang berbeda. Temuan ini menunjukkan bahwa klasifikasi tingkat *burnout* pada penelitian ini dipengaruhi oleh kombinasi berbagai aspek pola aktivitas harian, sehingga mendukung kebaruan penelitian dalam mengidentifikasi variabel yang paling berpengaruh terhadap proses klasifikasi menggunakan algoritma Random Forest.

3.6. Deployment

Tahap *deployment* dilakukan dengan mengimplementasikan model Random Forest yang telah melalui proses pelatihan dan evaluasi ke dalam aplikasi berbasis web menggunakan platform Hugging Face. Implementasi ini memungkinkan mahasiswa melakukan klasifikasi tingkat *burnout* berdasarkan data aktivitas harian mahasiswa secara daring, sehingga hasil prediksi dapat diperoleh secara langsung.

a. Halaman Utama dan Informasi Prediksi

Halaman utama memberikan informasi mengenai tujuan aplikasi, variabel yang digunakan dalam proses klasifikasi, serta kategori tingkat *burnout* yang dihasilkan oleh sistem. Informasi tersebut disajikan sebagai pengantar agar pengguna memahami cara kerja aplikasi sebelum melakukan prediksi. Tampilan halaman utama ditunjukkan pada Gambar 13.



Gambar 13 Halaman Utama Aplikasi

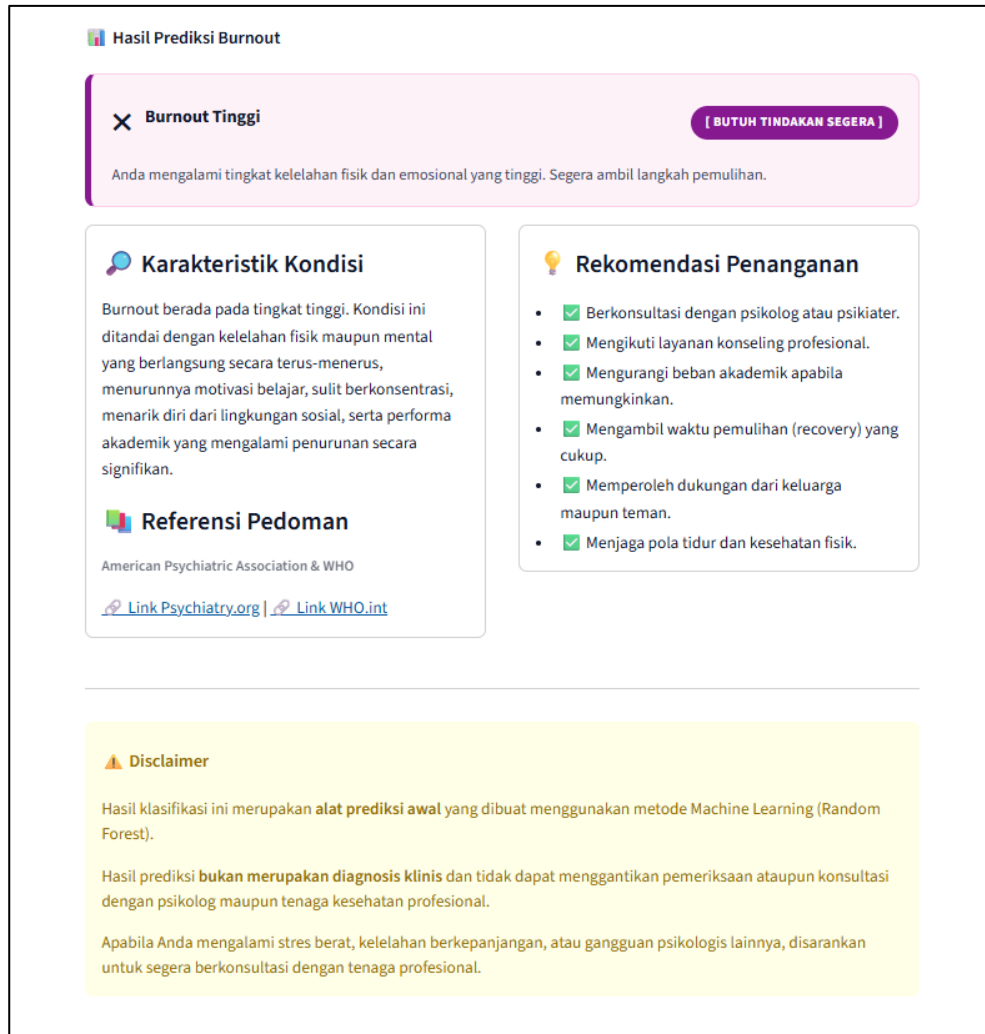
b. Halaman Form Input Data Aktivitas Harian

Pada halaman ini, pengguna diminta mengisi lima variabel aktivitas harian, yaitu jam tidur, waktu istirahat, intensitas *screen time*, keteraturan pola tidur, dan beban akademik. Ketiga variabel berupa durasi diinput menggunakan pilihan kategori ordinal, sedangkan keteraturan pola tidur dan beban akademik menggunakan skala Likert 1-4. Setelah seluruh data diisi, pengguna dapat menekan tombol “Jalankan Klasifikasi” untuk memproses data menggunakan model Random Forest. Tampilan form input ditunjukkan pada Gambar 14.

Gambar 14 Form Input Data Aktivitas Harian

c. Halaman Hasil Prediksi Klasifikasi *Burnout*

Setelah proses klasifikasi selesai, sistem menampilkan kategori tingkat *burnout* yang diprediksi, yaitu rendah, sedang, atau tinggi. Selain hasil prediksi, aplikasi juga menyajikan penjelasan singkat mengenai kondisi yang diprediksi, rekomendasi umum yang dapat dilakukan mahasiswa, serta disclaimer bahwa hasil yang diberikan hanya merupakan indikasi awal berdasarkan data aktivitas harian dan bukan diagnosis klinis. Tampilan hasil prediksi ditunjukkan pada Gambar 15.



Gambar 15 Hasil Prediksi Tingkat *Burnout*

4. KESIMPULAN

Penelitian ini berhasil memanfaatkan pola aktivitas harian mahasiswa, yang meliputi jam tidur, waktu istirahat, intensitas *screen time*, keteraturan pola tidur, dan beban akademik, sebagai data untuk mengklasifikasikan tingkat *burnout* menggunakan algoritma Random Forest. Hasil penelitian menunjukkan bahwa model memperoleh akurasi sebesar 88,83%, dengan nilai *weighted average precision*, *recall*, dan *F1-score* masing-masing sebesar 0,89, sehingga menunjukkan performa klasifikasi yang baik. Analisis *feature importance* menunjukkan bahwa intensitas screen time merupakan variabel yang paling berkontribusi dalam proses klasifikasi. Model juga berhasil diimplementasikan ke dalam aplikasi berbasis web menggunakan Hugging Face untuk memberikan indikasi awal tingkat *burnout self-assessment* responden dan tidak dimaksudkan sebagai diagnosis klinis. Penelitian selanjutnya dapat mengembangkan sistem agar dapat diterapkan pada skala yang lebih luas serta dilengkapi dengan rekomendasi yang lebih spesifik berdasarkan hasil klasifikasi, sehingga manfaat sistem bagi mahasiswa maupun pihak perguruan tinggi dapat semakin optimal.

REFERENSI

- [1] World Health Organization, “Burn-out an ‘occupational phenomenon’: International Classification of Diseases,” World Health Organization (WHO). Accessed: Jan. 17, 2026. [Online]. Available: <https://www.who.int/news/item/28-05-2019-burn-out-an-occupational-phenomenon-international-classification-of-diseases>
- [2] S. Nurhidayah *et al.*, “HUBUNGAN TINGKAT STRES DENGAN KUALITAS TIDUR PADA MAHASISWA TINGKAT AKHIR,” *JURNAL KESEHATAN TAMBUSAI*, vol. 6, no. 4, pp. 15803–15812, Dec. 2025, doi: <https://doi.org/10.31004/jkt.v6i4.51973>.
- [3] A. Dewi, P. A. Wiratmo, R. F. Mussyarifah, and I. Gilang Pamungkas, “ACADEMIC STRESS OF FINAL-YEAR STUDENTS OF THE NURSING STUDY PROGRAM: A CROSS-SECTIONAL STUDY WITH THEIR SLEEP QUALITY,” *KLABAT JOURNAL OF NURSING*, vol. 6, no. 2, pp. 2685–7154, Oct. 2024, doi: <https://doi.org/10.37771/kjn.v6i2.1172>.
- [4] R. Rosmawati and S. C. Merida, “Academic *Burnout* Pada Mahasiswa Fakultas Ekonomi dan Bisnis Universitas Esa Unggul,” 2025. doi: <https://doi.org/10.31004/jptam.v9i2.28507>.
- [5] S. N. Arista, A. Burhan, and R. N. Habdayani, “Studi Deskriptif *Burnout* Pada Mahasiswa Keperawatan Anestesiologi di Universitas Hatapan Bangsa,” *Jurnal Kesehatan Ilmiah Indonesia*, vol. 10, no. 1, Jun. 2025, doi: <https://doi.org/10.51933/health.v10i1.1698>.
- [6] E. Yeskuatov, L. K. Foo, and S.-L. Chua, “Detecting *Burnout* Among Undergraduate Computing Students with Supervised *Machine Learning*,” *Healthcare*, vol. 13, no. 23, p. 3182, Dec. 2025, doi: [10.3390/healthcare13233182](https://doi.org/10.3390/healthcare13233182).
- [7] D. B. Saputra, V. Atina, F. E. Nastiti, and F. I. Komputer, “PENERAPAN MODEL CRISP-DM PADA PREDIKSI NASABAH KREDIT MENGGUNAKAN ALGORITMA RANDOM FOREST,” *Idealis: Indonesia Journal Information System*, vol. 7, no. 2, pp. 240–247, Jul. 2024, doi: <https://doi.org/10.36080/idealis.v7i2.3244>.
- [8] A. Ananda Hapsari, A. Syafei Nursuwanda, H. Zuhriyah, and D. Junesco Vresdian, “Klasifikasi Kesehatan Mental Mahasiswa Model TMAS dengan Algoritma Decision Tree, Logistic Regression, dan Random Forest,” *Jurnal Informatika dan Teknologi Informasi (INTEK)*, vol. 7, no. 2, Nov. 2024, doi: <https://doi.org/10.37729/intek.v7i2.5690>.
- [9] N. Nurdiansyah, F. S. Febriyan, Z. G. D. Amanta, D. A. Saputra, and W. M. Baihaqi, “Analisis Kesehatan Mental untuk Mencegah Gangguan Mental pada Mahasiswa Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan Random Forest,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 1–9, Nov. 2024, doi: [10.57152/malcom.v5i1.1537](https://doi.org/10.57152/malcom.v5i1.1537).
- [10] I. Setiawan, I. Fatah Yasin, and Y. Tri Desianti, “Komparasi Kinerja Algoritma Random Forest, Decision Tree, Naïve Bayes, dan KNN dalam Prediksi Tingkat Depresi Mahasiswa Menggunakan Student Depression Dataset,” Feb. 2025. doi: doi.org/10.35960/ikomti.v6i1.1756.
- [11] S. A. Utiahman, A. Mulawati, and M. Pratama, “Penerapan Support Vector Machine dan Random Forest Classifier Untuk Klasifikasi Tingkat Obesitas,” *URNAL FASILKOM*, vol. 14, no. 3, pp. 754–760, Dec. 2024, doi: <https://doi.org/10.37859/jf.v14i3.8104>.
- [12] B. Maulana, D. Febrian, I. R. Fachrezi, and M. F. Zeen, “Comparison of Support Vector Machine, Random Forest, and C4.5 Algorithms for Customer Loss Prediction,” *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, vol. 2, no. 1, pp. 1–6, Feb. 2025, doi: [10.57152/ijatis.v2i1.1102](https://doi.org/10.57152/ijatis.v2i1.1102).
- [13] L. Maretva Cendani and A. Wibowo, “Perbandingan Metode *Ensemble Learning* pada Klasifikasi Penyakit Diabetes,” *Jurnal Masyarakat Informatika*, vol. 13, no. 1, pp. 2777–0648, May 2022, doi: <https://doi.org/10.14710/jmasif.13.1.42912>.

- [14] A. Yusufi, N. Prasetyo, G. Tyas Prayoga, M. Ghani Ramadhana, and K. Kunci, “Kesejahteraan Akademik Mahasiswa Islam di Era Kecerdasan Buatan: Teori dan Implementasi dalam Pengelolaan Stres dan *Burnout* Artikel Review,” *Jurnal Kolaboratif Sains*, vol. 8, no. 5, pp. 2577–2585, May 2025, doi: 10.56338/jks.v8i5.7618.
- [15] V. Oktaviani, N. Rosmawarni, M. Panji Muslim, F. Ilmu Komputer, U. Pembangunan Nasional Veteran Jakarta, and J. R. Fatmawati, “Perbandingan Kinerja Random Forest Dan Smote Random Forest Dalam Mendeteksi Dan Mengukur Tingkat Stres Pada Mahasiswa Tingkat Akhir,” *JURNAL INFORMATIK*, vol. 20, no. 1, pp. 43–49, Apr. 2024, doi: 10.52958/iftk.v20i1.9158.