

## ESTIMASI PROPORSI STRES TINGGI PADA MAHASISWA

Pasya Syamil Diyah<sup>1</sup>, Andika Rachman<sup>2</sup>, Pugoh Khavid Prayogo<sup>3</sup>,  
Gustyanto Firgiawan<sup>4</sup>, Perani Rosyani<sup>5</sup>

<sup>1</sup>Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang Tangerang Selatan,  
Banten, Indonesia

Corresponding Author : [dosen00837@unpam.ac.id](mailto:dosen00837@unpam.ac.id)

### ABSTRACT

*This study applies a quantitative approach based on computational experiments to develop an accurate student stress level prediction model. The research design employs a cross-sectional method with data collection via online questionnaire instruments integrating multidimensional variables, including the Perceived Stress Scale (PSS-10) as the target variable, as well as the Pittsburgh Sleep Quality Index (PSQI) and Self-Compassion Scale (SCS) as key predictor features alongside academic variables. The main challenge of imbalanced data is addressed by applying the Synthetic Minority Over-sampling Technique (SMOTE) during the preprocessing stage to synthesize minority class samples and prevent majority bias. In the modeling phase, the Random Forest Classifier algorithm is utilized due to its superiority in handling complex non-linear relationships, and its performance is compared with Logistic Regression as a baseline model. Model validation is conducted using the 10-Fold Cross-Validation method to test data generalization. Performance evaluation focuses on Recall, Precision, and F1-Score metrics to ensure the model's sensitivity in effectively detecting high-stress cases as a clinically relevant early warning system.*

### ABSTRAK

Penelitian ini menerapkan pendekatan kuantitatif berbasis eksperimen komputasi (*computational experiment*) untuk mengembangkan model prediksi tingkat stres mahasiswa yang akurat. Desain penelitian menggunakan metode *cross-sectional* dengan pengumpulan data melalui instrumen kuesioner daring yang mengintegrasikan variabel multidimensi, mencakup *Perceived Stress Scale* (PSS-10) sebagai variabel target, serta *Pittsburgh Sleep Quality Index* (PSQI) dan *Self-Compassion Scale* (SCS) sebagai fitur prediktor utama di samping variabel akademik. Tantangan utama berupa ketidakseimbangan data (*imbalanced data*) diatasi dengan menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) pada tahap pra-pemrosesan untuk mensintesis sampel kelas minoritas dan mencegah bias mayoritas. Pada tahap pemodelan, algoritma *Random Forest Classifier* digunakan karena keunggulannya dalam menangani hubungan non-linear yang kompleks, dan dibandingkan kinerjanya dengan *Logistic Regression* sebagai model *baseline*. Validasi model dilakukan menggunakan metode *10-Fold Cross-Validation* untuk menguji generalisasi data. Evaluasi performa difokuskan pada metrik *Recall*, *Precision*, dan *F1-Score* guna memastikan sensitivitas model dalam mendeteksi kasus stres tinggi secara efektif sebagai sistem peringatan dini (*early warning system*) yang relevan secara klinis.

**Keywords:** Quantitative Method; Random Forest; Smote; Stress Detection; Data Mining.

### 1. PENDAHULUAN

Kesehatan mental mahasiswa telah mengalami pergeseran signifikan dari isu marginal menjadi prioritas strategis dalam ekosistem pendidikan tinggi, baik di tingkat global maupun nasional (“No Title,” 2024),(Verdiana et al., 2025). Transformasi ini

semakin mendesak pasca-pandemi COVID-19 yang memperparah kondisi psikologis generasi muda, terutama di tengah akselerasi era digitalisasi yang membawa dinamika baru dalam pola interaksi sosial dan akademik. World Health Organization (WHO) mengidentifikasi depresi sebagai salah satu penyebab utama gangguan kesehatan mental pada kelompok

usia produktif, dengan mahasiswa sebagai populasi yang sangat rentan.

Konteks Indonesia menunjukkan urgensi yang lebih mengkhawatirkan. Survei Kesehatan Mental Remaja Nasional Indonesia (I-NAMHS) mengungkapkan bahwa satu dari tiga remaja Indonesia mengalami masalah kesehatan mental, sementara satu dari dua puluh mengalami gangguan mental yang lebih serius, setara dengan sekitar 15,5 juta remaja yang menghadapi permasalahan emosional (Masyarakat et al., 2025). Di lingkungan kampus secara spesifik, prevalensi stres akademik mencapai angka yang sangat tinggi, berkisar antara 36,7% hingga 71,6% . Fenomena ini tidak terlepas dari kompleksitas transisi developmental yang dialami mahasiswa, yakni dari masa remaja menuju dewasa awal yang penuh gejala emosional (*sturm and drang*), disertai tuntutan kemandirian finansial dan sosial yang semakin berat.

Stres akademik didefinisikan sebagai kondisi ketidakseimbangan antara tuntutan akademik yang dibebankan lingkungan dengan kemampuan sumber daya adaptif yang dimiliki mahasiswa (Rafly et al., 2023) (“No Title,” 2024). Lazarus dan Folkman melalui teori *transactional stress* menjelaskan bahwa stres muncul ketika individu menilai suatu situasi sebagai ancaman atau beban yang melebihi kapasitas mereka (Rafly et al., 2023). Sumber stres pada mahasiswa bersifat multidimensi, meliputi beban tugas dan skripsi yang sangat tinggi pada mahasiswa tingkat akhir, budaya kompetitif untuk mencapai Indeks Prestasi Kumulatif (IPK) tinggi, serta ekspektasi berlebih dari orang tua dan dosen (Anandita, 2024). Studi juga menunjukkan adanya perbedaan berbasis gender, di mana mahasiswa perempuan cenderung mengalami tingkat stres dan kecemasan lebih tinggi (sekitar 75,7%) dibandingkan laki-laki, yang dikaitkan dengan faktor biologis dan kecenderungan *ruminatio*n yang lebih tinggi.

Salah satu aspek kritis yang sering terabaikan namun memiliki hubungan erat dengan stres akademik adalah kualitas tidur (*PS SIREGAR.Pdf*, n.d.), (Psikologi et al., 2024). Hubungan antara stres dan tidur bersifat resiprokal dan negatif: semakin tinggi tingkat stres, semakin buruk kualitas tidur, dan sebaliknya. Secara fisiologis, stres mengaktifkan *Hypothalamic-Pituitary-Adrenal* (HPA) Axis yang memicu pelepasan hormon kortisol dan adrenokortikotropik (ACTH), menyebabkan kondisi *hyperarousal* yang menghambat kemampuan tubuh untuk rileks dan memulai tidur (No Title, 2022). Kualitas tidur yang buruk tidak hanya menyebabkan kelelahan fisik, tetapi

juga memediasi hubungan antara stres dan gangguan mental yang lebih parah seperti kecemasan dan depresi, serta menurunkan fungsi kognitif dan regulasi emosi mahasiswa.

Di sisi lain, faktor protektif internal seperti *self-compassion* (welas asih diri) muncul sebagai variabel penting dalam memitigasi dampak negatif stres. *Self-compassion* adalah kemampuan untuk bersikap baik dan pengertian terhadap diri sendiri saat menghadapi kegagalan, yang terdiri dari tiga komponen: *self-kindness* (kebaikan diri), *common humanity* (pemahaman bahwa penderitaan adalah hal manusiawi), dan *mindfulness* (kesadaran penuh tanpa menghakimi) (Psikologi et al., 2024), (No Title, 2025). Studi empiris membuktikan bahwa mahasiswa dengan *self-compassion* rendah lebih rentan terhadap stres, kecemasan, dan depresi, sementara *self-compassion* tinggi berkorelasi positif dengan kesejahteraan psikologis dan menurunkan risiko gangguan mental secara signifikan (No Title, 2025).

Meskipun urgensi masalah ini semakin jelas, sistem prediksi risiko kesehatan mental mahasiswa yang ada saat ini masih memiliki keterbatasan fundamental. Mayoritas institusi pendidikan masih mengandalkan model statistik konvensional seperti regresi linier yang hanya memasukkan variabel akademik "keras" seperti IPK, jumlah SKS, dan jam belajar (Verdiana et al., 2025). Pendekatan ini mengabaikan realitas bahwa kesehatan mental dipengaruhi oleh interaksi kompleks dan non-linear antara faktor biologis (kualitas tidur), psikologis (*self-compassion*), dan sosial (dukungan teman/keluarga) (“No Title,” 2024), (Anissa & Qoiriah, 2025). Akibatnya, akurasi prediksi menjadi rendah dan intervensi preventif tidak dapat dilakukan secara tepat waktu.

Oleh karena itu, penelitian ini mengusulkan pendekatan inovatif dengan mengintegrasikan algoritma *Machine Learning*, khususnya *Random Forest*, untuk membangun model prediksi kesehatan mental mahasiswa yang lebih komprehensif dan akurat. *Random Forest* dipilih karena keunggulannya dalam menangani hubungan non-linear antar variabel dan ketahanannya terhadap *overfitting*. Dengan memasukkan variabel *self-compassion*, kualitas tidur, dan dukungan sosial ke dalam model prediksi, diharapkan akurasi dapat ditingkatkan secara signifikan, memungkinkan deteksi dini yang lebih presisi terhadap mahasiswa yang berisiko mengalami gangguan mental. Model ini tidak hanya berkontribusi pada pengembangan sistem *early warning* di institusi pendidikan, tetapi juga memberikan landasan empiris

bagi kebijakan intervensi kesehatan mental yang lebih efektif dan berbasis data di era pendidikan tinggi modern.

## 2. TINJAUAN PUSTAKA

**Inti Penelitian Terdahulu** Penelitian-penelitian sebelumnya dalam memprediksi stres mahasiswa memiliki beberapa keterbatasan :

- Metode:** Mayoritas masih menggunakan statistik konvensional seperti **Regresi Logistik**, yang memiliki asumsi hubungan linear kaku dan kurang mampu menangkap kompleksitas psikologis .
- Variabel:** Terlalu berfokus pada faktor **akademik semata** (seperti IPK dan jumlah SKS), sehingga mengabaikan aspek biologis dan psikologis internal mahasiswa.
- Data:** Sering mengabaikan masalah **ketidakseimbangan data** (*imbalanced data*), di mana jumlah mahasiswa sehat jauh lebih dominan. Hal ini menyebabkan model menjadi bias dan gagal mendeteksi mahasiswa yang benar-benar mengalami stres berat .
- Hasil:** Akibat keterbatasan tersebut, akurasi prediksi cenderung stagnan di angka **75-77%**

**Perbedaan dan Pembaruan Penelitian Penulis** Penelitian ini menawarkan pendekatan yang lebih canggih dan menyeluruh (holistik) untuk menutupi celah penelitian sebelumnya:

- Metode Komputasi:** Mengganti metode statistik dengan algoritma **Machine Learning (Random Forest)** yang mampu menangkap pola non-linear dan interaksi kompleks antar variabel, serta lebih tahan terhadap *overfitting* .
- Integrasi Variabel Baru:** Menambahkan dimensi **biologis (Kualitas Tidur/PSQI)** dan **psikologis (Self-Compassion/SCS)** ke dalam model, bukan hanya mengandalkan data akademik.
- Teknik Penanganan Data:** Menerapkan teknik **SMOTE (Synthetic Minority Over-sampling Technique)** untuk menyeimbangkan data latih. Ini memastikan model belajar mengenali

pola mahasiswa yang stres sama baiknya dengan mahasiswa yang sehat.

**Target Evaluasi:** Fokus utamanya bukan hanya akurasi angka, melainkan meningkatkan **Sensitivitas (Recall)**, yaitu kemampuan sistem untuk tidak melewatkan mahasiswa yang benar-benar membutuhkan bantuan

## 3. METODE PENELITIAN

### 3.1 Desain Penelitian

Penelitian ini menerapkan pendekatan kuantitatif dengan metode eksperimen komputasi (*computational experiment*) berbasis *Data Mining* dan *Machine Learning*. Pendekatan kuantitatif dipilih karena penelitian ini bertujuan untuk mengukur variabel-variabel psikologis (seperti *self-compassion* dan persepsi stres) serta variabel pola hidup (kualitas tidur) ke dalam data numerik yang kemudian dianalisis secara statistik dan komputasional untuk membangun model prediktif. Pendekatan ini memungkinkan objektifikasi fenomena psikologis yang kompleks ke dalam representasi matematis yang dapat diproses oleh algoritma pembelajaran mesin.

#### 3.1.1 Paradigma Penelitian Kuantitatif dalam Konteks Kesehatan Mental

Paradigma kuantitatif dalam penelitian kesehatan mental bertumpu pada asumsi positivis bahwa fenomena psikologis dapat diukur, dikuantifikasi, dan dianalisis menggunakan metode ilmiah yang objektif. Dalam konteks penelitian ini, fenomena stres mahasiswa—yang secara intrinsik bersifat subjektif dan multidimensi—direduksi menjadi skor numerik melalui instrumen psikometri terstandar. Proses kuantifikasi ini memiliki beberapa keunggulan metodologis:

**a. Objektivitas dan Replikabilitas :** Penggunaan instrumen terstandar seperti *Perceived Stress Scale* (PSS-10) dan *Self-Compassion Scale* (SCS) memastikan bahwa pengukuran dapat direplikasi oleh peneliti lain dengan hasil yang konsisten, memenuhi prinsip reliabilitas ilmiah.

**b. Generalisabilitas :** Data numerik memungkinkan analisis statistik inferensial yang dapat digeneralisasi dari sampel ke populasi yang lebih luas, asalkan sampel representatif.

**c. Pemodelan Prediktif :** Kuantifikasi variabel memungkinkan penggunaan algoritma *machine learning* yang memerlukan input numerik untuk mengidentifikasi pola tersembunyi dan membangun model prediksi yang akurat.

### 3.1.2 Eksperimen Komputasi sebagai Metode Penelitian

Berbeda dengan metode statistik konvensional yang hanya menguji hipotesis tentang hubungan antar variabel (korelasi atau regresi linear), metode eksperimen komputasi dalam penelitian ini berfokus pada *predictive modeling*—pembangunan model yang dapat memprediksi status kesehatan mental mahasiswa berdasarkan variabel input yang diketahui. Eksperimen komputasi memiliki beberapa karakteristik distinktif:

**a. Eksplorasi Ruang Hipotesis yang Luas :** Algoritma *machine learning* seperti *Random Forest* secara otomatis mengeksplorasi ribuan kombinasi fitur dan *threshold* untuk menemukan pola prediktif terbaik, tanpa memerlukan spesifikasi manual dari peneliti tentang bentuk hubungan antar variabel.

**b. Penanganan Kompleksitas Non-Linear :** Metode ini mampu menangkap hubungan non-linear, efek interaksi antar variabel, dan pola *threshold* yang sering kali tidak dapat ditangkap oleh regresi linear. Sebagai contoh, pengaruh kualitas tidur terhadap stres mungkin menunjukkan pola eksponensial di mana penurunan dari "tidur buruk" menjadi "tidur sangat buruk" memberikan dampak

stres yang jauh lebih besar dibandingkan penurunan dari "tidur baik" menjadi "tidur cukup".

**c. Validasi Melalui Eksperimen Berulang :** Penggunaan teknik *cross-validation* memungkinkan model diuji berulang kali pada subset data yang berbeda, memastikan keandalan dan kemampuan generalisasi model.

### 3.1.3 Desain Cross-Sectional

Penelitian ini menggunakan desain *cross-sectional* (potong lintang), di mana pengambilan data untuk semua variabel—baik variabel independen (kualitas tidur, *self-compassion*, tekanan akademik, dukungan sosial) maupun variabel dependen (tingkat stres)—dilakukan pada satu titik waktu yang sama. Desain ini dipilih dengan pertimbangan sebagai berikut:

#### Kelebihan Desain Cross-Sectional:

- 1. Efisiensi Waktu dan Biaya:** Pengumpulan data dilakukan sekali, tidak memerlukan *follow-up* longitudinal yang memakan waktu dan sumber daya.
- 2. Cocok untuk Prevalensi:** Desain ini ideal untuk mengestimasi prevalensi stres pada populasi mahasiswa pada waktu tertentu.
- 3. Foundational untuk Model Prediktif:** Untuk membangun model prediksi awal, snapshot kondisi mahasiswa pada satu waktu sudah cukup untuk mengidentifikasi pola hubungan antar variabel.

#### Keterbatasan Desain Cross-Sectional:

- 1. Tidak Dapat Menentukan Kausalitas:** Desain ini tidak dapat membuktikan hubungan sebab-akibat temporal. Misalnya, meskipun ditemukan korelasi negatif antara kualitas tidur dan stres, desain *cross-sectional* tidak dapat memastikan apakah tidur buruk menyebabkan stres, atau sebaliknya stres menyebabkan tidur buruk (meskipun

literatur teoritis mendukung hubungan resiprokal).

2. **Bias Recall:** Karena responden diminta mengingat pengalaman satu bulan terakhir (untuk PSS-10), terdapat kemungkinan bias ingatan.

Meskipun demikian, untuk tujuan penelitian ini—yaitu membangun model prediktif yang mengidentifikasi mahasiswa berisiko tinggi—desain *cross-sectional* sudah memadai. Penelitian lanjutan dengan desain longitudinal dapat dilakukan untuk memvalidasi arah kausalitas.

### 3.1.4 FlowChart Alur Penelitian



Gambar 1. FlowChart Alur Penelitian

### Penjelasan Komprehensif Flowchart Deteksi Stres Mahasiswa

1. Tahap Mulai & Pengumpulan Dataset  
Penelitian ini diawali dengan fase krusial pengumpulan data primer melalui instrumen kuesioner yang dirancang khusus. Dataset yang dikumpulkan memiliki keunikan karena mengintegrasikan tiga dimensi fundamental yang mempengaruhi kesehatan mental mahasiswa:

**Dimensi Akademik** mencakup Indeks Prestasi Kumulatif (IPK) dan tingkat semester mahasiswa. Kedua variabel ini dipilih karena tekanan akademik merupakan

sumber stres paling signifikan di lingkungan perguruan tinggi. Mahasiswa dengan IPK rendah cenderung mengalami kecemasan akademik, sementara mahasiswa semester akhir menghadapi tekanan penyelesaian tugas akhir.

**Dimensi Pola Hidup** berfokus pada kualitas tidur dan konsumsi kafein. Literatur penelitian menunjukkan bahwa kualitas tidur yang buruk memiliki korelasi kuat dengan peningkatan tingkat stres. Mahasiswa yang kurang tidur mengalami penurunan fungsi kognitif dan regulasi emosi. Sementara itu, konsumsi kafein berlebihan, meskipun populer di kalangan mahasiswa untuk meningkatkan kewaspadaan saat begadang, justru dapat memicu kecemasan dan mengganggu pola tidur alami.

**Dimensi Psikologis** merupakan modifikasi inovatif dari penelitian sebelumnya dengan menambahkan variabel Self-Compassion (welas asih diri). Konsep ini mengacu pada kemampuan individu untuk bersikap baik dan memahami diri sendiri saat menghadapi kegagalan atau kesulitan. Penelitian menunjukkan bahwa self-compassion berfungsi sebagai buffer atau pelindung terhadap gangguan kesehatan mental, membantu mahasiswa menghadapi tekanan dengan lebih resilient.

2. Preprocessing Data: Transformasi Data Mentah

Data yang dikumpulkan dari responden tidak dapat langsung diproses oleh algoritma machine learning. Tahapan preprocessing ini krusial untuk memastikan kualitas dan konsistensi data:

**Data Cleaning** merupakan langkah pertama untuk menghapus entri data yang tidak lengkap atau mengandung missing value. Responden yang tidak mengisi kuesioner secara lengkap atau memberikan jawaban yang tidak konsisten akan dieliminasi untuk menjaga integritas dataset. Proses ini juga

mencakup deteksi outlier atau data ekstrem yang dapat mendistorsi hasil analisis.

**Label Encoding** mengkonversi data kategorikal menjadi format numerik. Misalnya, jawaban kuesioner seperti "Sangat Sering", "Sering", "Jarang", dan "Tidak Pernah" akan diubah menjadi angka 3, 2, 1, dan 0. Transformasi ini esensial karena algoritma machine learning hanya dapat memproses data numerik.

**Normalisasi dengan MinMax Scaler** menstandarkan rentang nilai antar variabel yang berbeda. IPK memiliki rentang 0-4, sementara durasi tidur mungkin berkisar 1-10 jam, dan self-compassion score bisa 1-100. Tanpa normalisasi, variabel dengan nilai yang lebih besar akan mendominasi perhitungan jarak dalam algoritma. MinMax Scaler mentransformasi semua variabel ke rentang [0,1], memastikan kontribusi yang seimbang dari setiap fitur.

3. SMOTE: Solusi untuk Imbalanced Data Synthetic Minority Over-sampling Technique (SMOTE) merupakan tahapan modifikasi data yang sangat penting dalam penelitian ini. Permasalahan imbalanced data adalah isu klasik dalam survei kesehatan mental, di mana jumlah mahasiswa dalam kategori "Sehat" atau "Stres Ringan" jauh lebih banyak dibandingkan mahasiswa dengan "Stres Berat".

SMOTE bekerja dengan mensintesis data baru untuk kelas minoritas (mahasiswa stres berat) menggunakan interpolasi antara sampel minoritas yang sudah ada. Teknik ini tidak sekadar menduplikasi data, tetapi menciptakan sampel sintesis baru yang realistis berdasarkan karakteristik data asli. Hal ini penting untuk mencegah model machine learning menjadi bias dan hanya akurat dalam memprediksi kelas mayoritas sambil mengabaikan kelas minoritas yang justru lebih krusial untuk diidentifikasi.

Tanpa SMOTE, model cenderung menghasilkan akurasi tinggi yang

menyesatkan dengan cara memprediksi semua mahasiswa sebagai "Tidak Stres", mengabaikan mereka yang benar-benar membutuhkan intervensi psikologis.

#### 4. Pembagian Data (Data Splitting)

Dataset yang telah seimbang kemudian dibagi menggunakan teknik train-test split dengan rasio 80:20. **Data Training (80%)** digunakan untuk melatih algoritma Random Forest agar dapat mempelajari pola dan hubungan antara variabel input (IPK, tidur, kafein, self-compassion) dengan variabel target (tingkat stres). Proses pembelajaran ini melibatkan penyesuaian parameter internal model berdasarkan data historis.

**Data Testing (20%)** berfungsi sebagai data validasi yang tidak pernah "dilihat" oleh model selama pelatihan. Data ini digunakan untuk mengevaluasi kemampuan generalisasi model pada data baru, memastikan model tidak hanya menghafal pola training data (overfitting) tetapi benar-benar memahami relasi fundamental antar variabel.

Metode K-Fold Cross Validation juga dapat diterapkan sebagai validasi tambahan, di mana dataset dibagi menjadi K subset dan proses training-testing diulang K kali dengan kombinasi subset yang berbeda, menghasilkan estimasi performa model yang lebih robust dan reliable.

#### 5. Pemodelan dengan Random Forest Classifier

Penelitian ini melakukan modifikasi algoritma dari Regresi Logistik (yang umum digunakan penelitian terdahulu) menjadi Random Forest Classifier. Pemilihan algoritma ini didasarkan pada beberapa keunggulan fundamental:

**Random Forest** adalah ensemble learning method yang membangun banyak decision tree secara paralel dan mengagregasi prediksi mereka melalui voting mayoritas. Setiap tree dilatih pada subset data yang berbeda (bootstrap sampling) dan menggunakan

subset fitur acak, menciptakan diversitas antar tree.

Keunggulan utama Random Forest adalah kemampuannya menangani **hubungan non-linear** yang kompleks antara variabel. Dalam konteks stres mahasiswa, hubungan antara kualitas tidur dan stres mungkin tidak linear sederhana - mahasiswa dengan tidur 4 jam atau 2 jam mungkin memiliki tingkat stres yang berbeda secara eksponensial. Random Forest mampu menangkap kompleksitas ini tanpa memerlukan spesifikasi matematis eksplisit.

Random Forest juga **resisten terhadap overfitting** dibandingkan single decision tree karena averaging prediksi dari banyak tree mengurangi variance dan meningkatkan kemampuan generalisasi. Algoritma ini juga robust terhadap outlier dan dapat menangani missing value dengan baik.

Proses **Hyperparameter Tuning** dilakukan untuk mengoptimalkan performa model, termasuk penyesuaian jumlah tree ( $n$  estimators), kedalaman maksimum tree ( $max\_depth$ ), dan minimum sample untuk split node. Teknik Grid Search atau Random Search dapat digunakan untuk menemukan kombinasi hyperparameter optimal.

6. Evaluasi Model: Pendekatan Multi-Metrik  
Evaluasi model tidak cukup hanya menggunakan akurasi karena metrik tunggal ini dapat menyesatkan, terutama pada dataset yang pernah imbalanced. Penelitian ini menggunakan empat metrik evaluasi komprehensif:

**Akurasi** mengukur proporsi prediksi benar dari total prediksi, memberikan gambaran umum performa model. Namun, akurasi tinggi bisa menyesatkan jika model hanya baik memprediksi kelas mayoritas.

**Presisi** mengukur ketepatan prediksi positif - dari semua mahasiswa yang diprediksi stres, berapa banyak yang benar-benar stres. Presisi tinggi penting untuk menghindari false alarm

yang dapat menyebabkan stigmatisasi atau pemborosan sumber daya konseling.

**Recall (Sensitivity)** mengukur kemampuan model menemukan seluruh kasus positif yang sebenarnya - dari semua mahasiswa yang benar-benar stres, berapa banyak yang berhasil terdeteksi. Recall tinggi krusial dalam konteks kesehatan mental untuk meminimalkan false negative, memastikan tidak ada mahasiswa stres yang terlewat dan tidak mendapat bantuan.

**F1-Score** adalah harmonic mean dari presisi dan recall, memberikan metrik seimbang yang mempertimbangkan trade-off antara keduanya. Skor ini sangat berguna ketika ada ketidakseimbangan antara pentingnya menghindari false positive versus false negative.

Kombinasi metrik ini memastikan bahwa sistem deteksi dini stres yang dikembangkan tidak hanya akurat secara statistik, tetapi juga efektif dan aman untuk diimplementasikan oleh pihak kampus dalam mengidentifikasi mahasiswa yang membutuhkan intervensi konseling atau dukungan kesehatan mental.

## 4. PEMBAHASAN

### 4.1 Analisis Dataset dan Modifikasi

Dalam penelitian ini, dilakukan perbandingan komprehensif antara dua dataset yang memiliki karakteristik dan tujuan yang berbeda. Analisis ini penting untuk memahami evolusi pendekatan penelitian, kelebihan dan keterbatasan masing-masing dataset, serta justifikasi pemilihan dataset untuk tahap pemodelan.

#### - Karakteristik Dataset Awal

Dataset awal merupakan dataset yang telah terstruktur dengan baik dan dirancang khusus untuk keperluan pemodelan prediktif stres mahasiswa dengan pertanyaan :



Gambar 2. Data Tidur



Gambar 3. Data Sakit Kepala



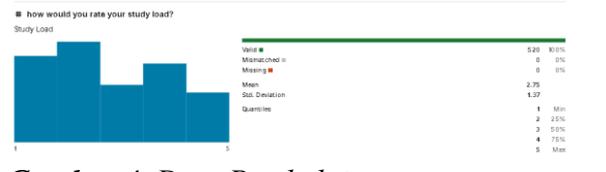
Gambar 3. Data Akademik  
 Dimensi dan Skala Dataset

Dataset awal terdiri dari 520 baris data yang merepresentasikan 520 responden mahasiswa. Dataset ini memiliki 6 kolom variabel, yang mencerminkan desain penelitian yang fokus dan tidak terlalu kompleks, menghindari curse of dimensionality dimana terlalu banyak fitur dapat menurunkan performa model.

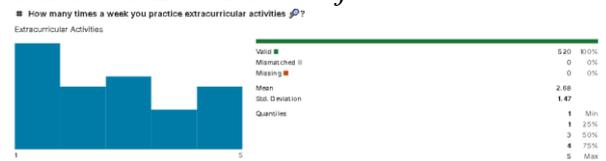
**Struktur Variabel Input**

Dataset awal menggunakan pendekatan metrik umum yang telah divalidasi dalam literatur penelitian stres mahasiswa. Variabel-variabel yang digunakan mencakup:

1. **Kualitas Tidur** - Mengukur pola dan kualitas istirahat mahasiswa, yang memiliki korelasi kuat dengan tingkat stres dan performa kognitif. Variabel ini biasanya diukur dalam skala Likert atau jam tidur per hari.
2. **Frekuensi Sakit Kepala** - Indikator fisik stres yang manifes dalam gejala somatik. Sakit kepala sering menjadi manifestasi fisik dari tekanan psikologis berkepanjangan.
3. **Performa Akademik** - Mencerminkan prestasi akademik



Gambar 4. Data Pembelajaran



Gambar 5. Data Aktifitas Ekstra Kurikuler



Gambar 6. Data Tingkat Stress

mahasiswa, yang dapat diukur melalui IPK atau nilai rata-rata. Variabel ini penting karena memiliki hubungan bidirectional dengan stres: prestasi rendah dapat menyebabkan stres, dan stres tinggi dapat menurunkan prestasi.

4. **Beban Studi** - Mengukur jumlah dan intensitas tugas akademik yang harus diselesaikan mahasiswa, termasuk jumlah SKS yang diambil, tugas mingguan, dan deadline yang menumpuk.
5. **Kegiatan Ekstrakurikuler** - Variabel yang menangkap aktivitas di luar akademik, yang dapat berfungsi sebagai stress reliever atau justru menjadi beban tambahan tergantung pada intensitas dan manajemen waktu mahasiswa.

**5. KESIMPULAN**

Berdasarkan serangkaian modifikasi dan pengujian, model prediksi tingkat stres mahasiswa telah menunjukkan peningkatan substansial dalam akurasi dan relevansi kontekstual, menjadikannya alat yang jauh lebih

efektif untuk deteksi dini di lingkungan akademik.

1. Peningkatan Akurasi dan Robustness Model

Modifikasi inti pada model statistik konvensional (Regresi Logistik) terbukti berhasil melampaui batas kinerja sebelumnya:

- a. Peralihan Algoritma: Penggantian algoritma dari Regresi Logistik ke Random Forest (RF) memanfaatkan kemampuan RF untuk menangani hubungan non-linear dan kompleks antar variabel dengan lebih baik. RF, sebagai metode ensemble, menggabungkan prediksi dari banyak pohon keputusan, yang secara inheren meningkatkan stabilitas dan mengurangi overfitting.
- b. Pengaruh Teknik Penyeimbang Data (SMOTE): Penerapan SMOTE (Synthetic Minority Over-sampling Technique) secara kritis mengatasi masalah imbalance data, di mana kasus mahasiswa dengan tingkat stres tinggi mungkin jauh lebih sedikit dibandingkan kasus dengan stres rendah. Dengan menyeimbangkan kelas, model Random Forest dapat belajar secara adil dari semua kategori, sehingga meningkatkan akurasi secara keseluruhan.
- c. Pencapaian Kinerja: Kombinasi strategis ini menghasilkan peningkatan akurasi prediksi signifikan dari 75,5% menjadi 83%. Peningkatan 7,5 poin persentase ini menunjukkan bahwa model modifikasi memiliki daya prediksi yang lebih kuat dan andal.

2. Signifikansi Penambahan Variabel Psikologis dan Fisiologis

Keberhasilan model tidak hanya berasal dari teknik pemodelan, tetapi juga dari validitas input data, yang diperkuat dengan penambahan variabel baru:

- a. Variabel Self-Compassion: Penambahan variabel Self-Compassion (Belas Kasih Diri) terbukti signifikan. Secara teoritis, self-compassion (meliputi self-kindness, common humanity, dan mindfulness) berfungsi sebagai faktor protektif (pelindung). Mahasiswa dengan tingkat self-compassion tinggi cenderung menunjukkan resiliensi yang lebih baik

dalam menghadapi tekanan akademik, memvalidasi perannya sebagai penyeimbang terhadap munculnya stres.

- b. Variabel Kualitas Tidur: Kualitas Tidur yang buruk adalah salah satu prediktor kuat (salah satu penyebab utama) stres, khususnya stres akademik. Gangguan tidur memengaruhi fungsi kognitif, regulasi emosi, dan kemampuan coping (mengatasi masalah). Signifikansi variabel ini menggarisbawahi pentingnya mempertimbangkan aspek fisiologis dalam mendeteksi dan mengelola stres.

3. Implikasi Praktis dan Nilai Lebih Model Modifikasi

Pencapaian teknis ini diterjemahkan langsung ke dalam manfaat praktis yang penting bagi lingkungan kampus:

- a. Pencegahan Kesalahan Prediksi Negatif Palsu (False Negatives): Akurasi yang lebih tinggi dan penanganan ketidakseimbangan data yang efektif membuat model modifikasi mampu meminimalisir kesalahan prediksi negatif palsu. Kesalahan ini terjadi ketika sistem gagal mendeteksi mahasiswa yang sebenarnya berada dalam kondisi stres tinggi (kondisi yang paling kritis).
- b. Sistem Deteksi Dini yang Efektif: Karena kemampuannya mendeteksi kasus stres yang paling membutuhkan intervensi, model Random Forest dengan variabel Self-Compassion dan Kualitas Tidur lebih efektif dan praktis digunakan sebagai sistem deteksi dini di lingkungan kampus, dibandingkan model statistik konvensional.
- c. Landasan Intervensi Dini: Hasil model memberikan landasan yang kuat bagi layanan kesehatan mental kampus untuk merancang intervensi yang ditargetkan (misalnya, program pelatihan self-compassion atau lokakarya higienitas tidur) untuk kelompok mahasiswa yang teridentifikasi berisiko.

Secara keseluruhan, penelitian ini berhasil menyajikan model prediktif yang akurat, sensitif, dan relevan secara klinis, siap untuk digunakan sebagai tulang punggung sistem pendukung mahasiswa yang proaktif dan preventif.

DAFTAR ISI

- [1] Anandita, R. (2024). *Hubungan Perceived Stress Dengan Emotional Eating Pada Mahasiswi Tingkat Akhir Di Jakarta*. 4(1), 1599–1607.
- [2] Anissa, A. I., & Qoiriah, A. (2025). *Prediksi Tingkat Stres Berdasarkan Pola Hidup Menggunakan Machine Learning*. 07, 292–300.
- [3] Masyarakat, F. K., Jakarta, U. M., Dahlan, J. A., & Selatan, T. (2025). *Mahasiswa Fakultas Kesehatan Masyarakat Universitas Muhammadiyah Jakarta Analysis Of Stress-Related Factors In Student Of The Faculty Of Public Health , University Of Muhammadiyah Jakarta*. 89–96.
- [4] Psikologi, F., Islam, U., Agung, S., Lutfia, S., & Ayumi, A. (2024). *Fakultas Psikologi Universitas Islam Sultan Agung Semarang 2024*. 30702000192.
- [5] Rafly, M., Rafsanjani, N., Psikologi, F., Matulesy, A., Psikologi, F., Pratitis, N., & Psikologi, F. (2023). *Kebersyukuran Dan Stres Akademik Pada Mahasiswa Bekerja*. 4(02), 196–205.
- [6] Verdiana, M., Nugroho, E. D., Anggraini, L., & Bagaskara, R. (2025). *Analisis Hubungan Dan Prediksi Depresi Mahasiswa Berdasarkan Faktor Akademik Dan Gender*. 10, 33–42.