LINIER
LITERATUR INFORMATIKA & KOMPUTER
e-ISSN: 3063-2218
DOI : 10.33096/linier.v3i2.3633
Literatur Informatika & Komputer
Vol 3, No 2 (2026), pp 162-173

Klasifikasi Tingkat Stres Mahasiswa Universitas Muslim Indonesia Menggunakan Decision Tree dan Random Forest

Muhammad Shabran[a], Herdianti Darwis[b], Siska Anraeni[c]
Universitas Muslim Indonesia, Makassar, Indonesia
Email: [a]realshabran@gmail.com; [b]herdianti.darwis@umi.ac.id; [c]siska.anraeni@umi.ac.id

Received: 16-06-2026 | Revised: 22-06-2026 | Accepted: 26-06-2026 | Published: 29-06-2026

Abstrak

Kesehatan mental mahasiswa menjadi isu penting di lingkungan perguruan tinggi karena berpengaruh terhadap performa akademik, interaksi sosial, dan kesejahteraan psikologis. Mahasiswa sering menghadapi tekanan akademik dan tuntutan sosial yang dapat memicu stres dengan tingkat keparahan berbeda. Penelitian ini bertujuan untuk mengklasifikasikan tingkat stres mahasiswa Universitas Muslim Indonesia berdasarkan instrumen *Depression Anxiety Stress Scale-21* (DASS-21) menggunakan algoritma *Decision Tree* berbasis *Classification and Regression Tree* (CART) dan *Random Forest*. Penelitian ini melibatkan lebih dari 3000 responden mahasiswa yang datanya diperoleh melalui kuesioner DASS-21 serta variabel pendukung seperti jenis kelamin, usia, fakultas, indeks massa tubuh, durasi tidur, olahraga, dan penggunaan media sosial. Tahapan analisis meliputi preprocessing data, *feature engineering*, *encoding variabel*, penanganan ketidakseimbangan kelas menggunakan *Synthetic Minority Oversampling Technique for Nominal and Continuous data* (SMOTENC), serta pembagian data 80% untuk pelatihan dan 20% untuk pengujian. Evaluasi model dilakukan menggunakan confusion matrix, *accuracy*, *balanced accuracy*, *precision*, *recall*, dan *Macro F1-score*. Hasil penelitian menunjukkan bahwa *Random Forest* dengan *tuning* tanpa SMOTENC memberikan performa terbaik dengan *accuracy* sebesar 0,9065 dan *Macro F1-score* sebesar 0,8427, serta lebih stabil dibandingkan *Decision Tree* dalam klasifikasi multi-kelas tingkat stres mahasiswa.

Kata kunci: DASS-21, Decision Tree, Klasifikasi, *Random Forest*, Stres Mahasiswa.

Pendahuluan

Kesehatan mental mahasiswa merupakan salah satu isu yang semakin mendapat perhatian di lingkungan perguruan tinggi karena memiliki pengaruh langsung terhadap konsentrasi belajar, performa akademik, hubungan sosial, serta kualitas hidup secara keseluruhan [1]. Mahasiswa berada pada fase kehidupan yang penuh dengan berbagai tuntutan dan perubahan, baik dalam aspek akademik maupun sosial [2]. Tekanan akademik, keterlibatan dalam kegiatan organisasi, beban tugas perkuliahan, ketidakpastian masa depan, serta proses penyesuaian diri terhadap lingkungan kampus sering kali menjadi sumber stres bagi mahasiswa [3]. Kondisi tersebut menjadikan stres sebagai salah satu permasalahan kesehatan mental yang paling umum dialami oleh mahasiswa [4]. Apabila tidak diidentifikasi dan ditangani sejak dini, stres berpotensi berkembang menjadi gangguan psikologis yang lebih serius serta dapat mengganggu keberlangsungan proses studi mahasiswa [5].

Untuk memahami kondisi psikologis mahasiswa secara lebih sistematis, berbagai penelitian memanfaatkan instrumen *Depression Anxiety Stress Scale-21* (DASS-21) sebagai alat ukur yang mampu mengidentifikasi tingkat depresi, kecemasan, dan stres secara terstruktur [6]. Instrumen ini dinilai praktis serta memiliki struktur pengukuran yang jelas sehingga banyak digunakan dalam penelitian kesehatan mental, khususnya pada populasi mahasiswa [7]. Sejumlah studi terbaru juga menunjukkan bahwa DASS-21 memiliki validitas dan reliabilitas yang baik dalam mengukur kondisi psikologis individu pada berbagai konteks penelitian, termasuk pada populasi mahasiswa di berbagai negara [8]. Oleh karena itu, penggunaan DASS-21 dalam penelitian ini memiliki dasar metodologis yang kuat untuk mengidentifikasi tingkat stres mahasiswa [9].

Seiring dengan perkembangan teknologi informasi, pendekatan analisis terhadap permasalahan kesehatan mental tidak lagi terbatas pada metode deskriptif konvensional [10]. Saat ini, teknik *machine learning* mulai banyak dimanfaatkan untuk membantu proses analisis data yang lebih kompleks, termasuk dalam mengidentifikasi dan mengklasifikasikan kondisi psikologis mahasiswa [11]. Pendekatan ini memungkinkan sistem mempelajari pola dari data yang tersedia sehingga dapat menghasilkan model prediksi yang lebih objektif dan terukur[12]. Dalam konteks kesehatan mental mahasiswa, metode *machine learning* telah digunakan untuk mengolah data survei maupun data perilaku guna mengklasifikasikan kondisi psikologis

berdasarkan karakteristik responden [13]. Dengan kemampuan tersebut, pendekatan ini menjadi solusi yang potensial dalam membantu proses identifikasi tingkat stres mahasiswa secara lebih sistematis.

Dalam penelitian klasifikasi berbasis *machine learning*, beberapa algoritma yang sering digunakan antara lain *Decision Tree* dan *Random Forest*. *Algoritma Decision Tree* dikenal sebagai metode klasifikasi yang mudah dipahami karena proses pengambilan keputusan dapat ditelusuri melalui struktur pohon yang menggambarkan hubungan antara atribut data dan kelas target [14]. Sementara itu, *Random Forest* merupakan pengembangan dari *Decision Tree* yang menggunakan pendekatan *ensemble learning*, yaitu dengan menggabungkan sejumlah pohon keputusan untuk meningkatkan stabilitas dan akurasi model [15]. Pendekatan ini memungkinkan model menghasilkan prediksi yang lebih robust terhadap variasi data [16]. Beberapa penelitian sebelumnya menunjukkan bahwa *Random Forest* sering memberikan performa klasifikasi yang lebih baik dalam berbagai permasalahan prediksi, termasuk pada analisis kesehatan mental mahasiswa yang memiliki banyak variabel dan pola data yang kompleks [17]. Meskipun demikian, *Decision Tree* tetap memiliki keunggulan dalam hal interpretasi model sehingga dapat membantu menjelaskan faktor-faktor yang memengaruhi hasil klasifikasi [18].

Penelitian mengenai klasifikasi kesehatan mental telah banyak dilakukan dengan memanfaatkan berbagai algoritma *machine learning*. Salah satu penelitian berjudul "Klasifikasi Kesehatan Mental pada Usia Remaja Menggunakan Metode SVM" menerapkan algoritma *Support Vector Machine* (SVM) untuk mengidentifikasi kondisi psikologis remaja berdasarkan data kuesioner. Metode tersebut digunakan untuk mengklasifikasikan tingkat depresi, anxiety, dan stres, dengan hasil akurasi sebesar 79% yang menunjukkan bahwa algoritma SVM cukup efektif dalam mengenali pola kesehatan mental remaja [19]. Namun, penelitian tersebut tidak secara khusus membahas klasifikasi tingkat stres pada mahasiswa di lingkungan perguruan tinggi. Oleh karena itu, penelitian ini menggunakan algoritma *Decision Tree* dan *Random Forest* untuk mengklasifikasikan tingkat stres mahasiswa Universitas Muslim Indonesia sehingga dapat memberikan analisis yang lebih kontekstual serta representatif terhadap kondisi stres mahasiswa di lingkungan perguruan tinggi.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengklasifikasikan tingkat stres mahasiswa Universitas Muslim Indonesia menggunakan algoritma *Decision Tree* dan *Random Forest*. Selain itu, penelitian ini juga membandingkan performa kedua algoritma tersebut menggunakan beberapa metrik evaluasi, yaitu *accuracy*, *balanced accuracy*, dan *macro F1-score*, sehingga dapat diperoleh model klasifikasi yang tidak hanya memiliki tingkat akurasi yang tinggi tetapi juga mampu mengenali setiap kelas secara lebih seimbang. Fokus penelitian diarahkan secara khusus pada aspek stres agar ruang lingkup analisis menjadi lebih spesifik dan hasil penelitian dapat dibahas secara lebih mendalam.

Penelitian ini diharapkan dapat memberikan kontribusi baik secara akademik maupun praktis. Secara akademik, penelitian ini memberikan analisis komparatif mengenai performa algoritma *Decision Tree* dan *Random Forest* dalam klasifikasi tingkat stres mahasiswa berbasis data survei. Secara praktis, hasil penelitian ini diharapkan dapat memberikan gambaran mengenai metode yang paling efektif dalam mengidentifikasi tingkat stres mahasiswa sehingga dapat menjadi dasar dalam pengembangan sistem deteksi dini maupun kebijakan pendampingan kesehatan mental di lingkungan perguruan tinggi. Dengan demikian, penelitian ini tidak hanya berkontribusi dalam pengembangan metode klasifikasi berbasis *machine learning*, tetapi juga mendukung upaya pemantauan kesehatan mental mahasiswa secara lebih objektif dan terstruktur.

Metode

2.1 Kerangka Dasar Penelitian

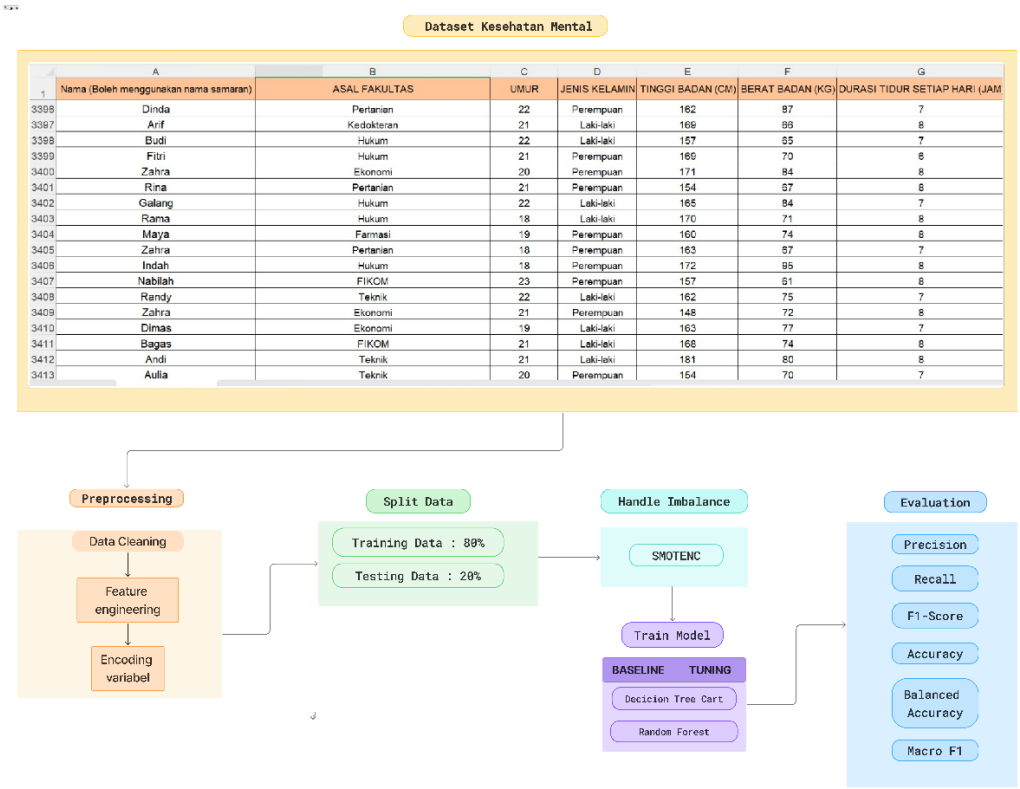
Penelitian ini merupakan penelitian kuantitatif dengan pendekatan analisis klasifikasi berbasis *machine learning*. Penelitian dilaksanakan di Universitas Muslim Indonesia dengan jumlah responden sebanyak lebih dari 3000 mahasiswa yang diperoleh melalui penyebaran kuesioner *Depression Anxiety Stress Scale-21 (DASS-21)*. Data yang digunakan meliputi skor DASS-21 pada aspek stres serta beberapa variabel pendukung, seperti jenis kelamin, usia, indeks massa tubuh, durasi tidur, durasi olahraga, dan durasi penggunaan media sosial. Variabel dependen dalam penelitian ini adalah tingkat stres mahasiswa yang dikategorikan ke dalam lima kelas, yaitu normal, ringan, sedang, berat, dan sangat berat berdasarkan konversi skor DASS-21. Sementara itu, variabel independen terdiri dari skor stres dan variabel demografis serta perilaku mahasiswa.

Penelitian ini memiliki hipotesis bahwa terdapat perbedaan performa klasifikasi antara algoritma *Decision Tree* berbasis *Classification and Regression Tree* (CART) dan *Random Forest* dalam mengklasifikasikan tingkat stres mahasiswa. Selain itu, penerapan teknik penanganan ketidakseimbangan kelas menggunakan *Synthetic Minority Oversampling Technique for Nominal and Continuous Data (SMOTENC)* diduga dapat meningkatkan performa model klasifikasi.

Proses analisis penelitian dilakukan melalui beberapa tahapan, yaitu data preprocessing yang meliputi *data cleaning*, *feature engineering*, serta proses *encoding* pada variabel kategorikal. Pada tahap *feature engineering* dilakukan pembentukan variabel indeks massa tubuh yang dihitung dari tinggi badan dan berat badan responden. Selanjutnya, dataset dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20%. Model klasifikasi dibangun menggunakan algoritma *Decision Tree* berbasis CART dan *Random Forest* dengan dua skenario pengujian, yaitu *baseline* dan *hyperparameter tuning*. Evaluasi performa model dilakukan menggunakan *confusion matrix*, *accuracy*, *balanced accuracy*, *precision*, *recall*, dan *macro F1-score* untuk memperoleh gambaran kinerja model secara lebih komprehensif.

2.2 Tahapan Penelitian

Untuk memperoleh model klasifikasi yang optimal, penelitian ini dilakukan melalui beberapa tahapan yang terstruktur mulai dari pengumpulan dataset hingga proses evaluasi model. Setiap tahapan dirancang untuk memastikan bahwa data yang digunakan telah melalui proses pengolahan yang tepat sehingga model *machine learning* yang dibangun dapat menghasilkan performa yang akurat dan stabil. Alur tahapan penelitian yang digunakan dalam penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, penelitian ini dilakukan melalui beberapa tahapan yang dimulai dari pengumpulan dataset kesehatan mental mahasiswa menggunakan kuesioner DASS-21. Data kemudian melalui proses *preprocessing*, pembagian data pelatihan dan pengujian, serta penanganan ketidakseimbangan kelas menggunakan metode *SMOTENC*. Selanjutnya dilakukan pembangunan model klasifikasi menggunakan algoritma *Decision Tree* dan *Random Forest*, kemudian dievaluasi untuk mengetahui performa model dalam mengklasifikasikan tingkat stres mahasiswa.

2.3 Pengumpulan Data

Tahap pengumpulan data dilakukan dengan menyebarkan kuesioner DASS-21 kepada mahasiswa Universitas Muslim Indonesia. Instrumen ini terdiri dari tiga aspek kesehatan mental, yaitu stres, depresi, dan *Anxiety*. Selain data utama dari DASS-21, peneliti juga mengumpulkan data pendukung seperti usia, jenis kelamin, fakultas, dan informasi lain yang relevan untuk membantu proses analisis. Pengumpulan data dilakukan secara terstruktur untuk memastikan data yang diperoleh valid, reliabel, dan sesuai dengan kebutuhan penelitian. Keseluruhan data ini menjadi dasar dalam proses klasifikasi tingkat kesehatan mental mahasiswa.

2.4 Preprocessing

Tahap *preprocessing* dilakukan untuk mempersiapkan dataset sebelum digunakan dalam proses pemodelan *machine learning*. Tahap ini bertujuan untuk meningkatkan kualitas data sehingga model dapat mempelajari pola data dengan lebih baik. Proses *preprocessing* dalam penelitian ini meliputi *data cleaning*, *feature engineering*, dan *encoding data*.

a. *Data Cleaning*
   Pada tahap ini dilakukan proses pembersihan data untuk mengidentifikasi dan mengatasi permasalahan pada dataset, seperti *missing value*, data duplikat, serta data yang tidak konsisten. Data yang teridentifikasi bermasalah kemudian diperbaiki atau dihapus agar dataset yang digunakan dalam proses pemodelan menjadi lebih akurat dan representatif [20].
b. *Feature Engineering*
   Tahap *feature engineering* dilakukan untuk membentuk variabel baru yang dapat meningkatkan kualitas informasi dalam dataset. Pada penelitian ini dilakukan pembentukan variabel indeks massa tubuh (*Body Mass Index/*BMI) yang dihitung dari tinggi badan dan berat badan responden. Variabel ini digunakan sebagai salah satu fitur tambahan dalam proses klasifikasi tingkat stres mahasiswa.
c. *Encoding Data*
   Encoding data dilakukan untuk mengubah variabel kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma *machine learning*. Teknik yang digunakan adalah *label encoding*, yaitu pemberian kode angka pada setiap kategori dalam suatu variabel sehingga data dapat diolah oleh model klasifikasi [21].

2.5 Split Data

Setelah melalui tahap *preprocessing*, dataset kemudian dibagi menjadi dua bagian, yaitu data pelatihan (*training data*) dan data pengujian (*testing data*) [22]. Pembagian data dilakukan dengan perbandingan 80% untuk data pelatihan dan 20% untuk data pengujian. Data pelatihan digunakan untuk membangun model klasifikasi menggunakan algoritma *Decision Tree* dan *Random Forest*, sedangkan data pengujian digunakan untuk mengevaluasi performa model terhadap data yang belum pernah digunakan pada proses pelatihan. Pembagian data ini bertujuan untuk memastikan bahwa model yang dihasilkan mampu melakukan generalisasi dengan baik terhadap data baru [23].

2.6 Handle Imbalance SMOTENC

Dataset yang digunakan dalam penelitian ini memiliki distribusi kelas yang tidak seimbang, di mana jumlah data pada beberapa kategori tingkat stres lebih sedikit dibandingkan kategori lainnya. Kondisi ini dapat menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas dan kurang mampu mengenali kelas minoritas. Untuk mengatasi permasalahan tersebut, digunakan metode (SMOTENC). Metode ini bekerja dengan menghasilkan data sintetis pada kelas minoritas dengan mempertimbangkan variabel numerik dan kategorikal secara bersamaan sehingga distribusi data antar kelas menjadi lebih seimbang. Dengan penerapan SMOTENC, model diharapkan dapat mempelajari pola data secara lebih merata pada setiap kategori tingkat stres.

2.7 Train Model

Pada tahap ini dilakukan proses pelatihan (training) model menggunakan dua algoritma *machine learning*, yaitu *Decision Tree* dan *Random Forest*. Proses pelatihan dilakukan menggunakan data pelatihan (*training data*) yang telah melalui tahap preprocessing dan penanganan ketidakseimbangan data.

a. Decision Tree
   Model *Decision Tree* yang digunakan adalah metode CART. Algoritma ini membentuk struktur pohon keputusan melalui proses pemisahan data berdasarkan atribut yang paling optimal hingga menghasilkan

kelas prediksi [24]. Proses pemisahan data pada metode CART menggunakan kriteria Gini Index untuk mengukur tingkat ketidakmurnian suatu node. Semakin kecil nilai Gini Index, maka semakin baik kualitas pemisahan data. Rumus Gini Index ditunjukkan pada Persamaan.

$$Gini(S) = 1 - \sum_{i=1}^{c} pi^2 \quad (1)$$

Keterangan:

S : Himpunan data atau kasus pada suatu node
c : Jumlah kelas dalam himpunan data
p_i : Proporsi jumlah kasus pada kelas ke-i terhadap total kasus dalam S

b. Random Forest

Random Forest merupakan pengembangan dari *Decision Tree* yang menggunakan pendekatan *ensemble learning*. Metode ini membangun banyak pohon keputusan dari subset data yang berbeda, kemudian hasil prediksi dari setiap pohon digabungkan menggunakan metode *majority voting* untuk menghasilkan prediksi akhir yang lebih akurat dan stabil [9].

2.8 Evaluasi Model

Tahap terakhir dalam penelitian ini adalah evaluasi model untuk mengukur kinerja algoritma *Decision Tree* dan *Random Forest* dalam mengklasifikasikan tingkat stres mahasiswa. Evaluasi dilakukan menggunakan beberapa metrik yang dihitung berdasarkan *confusion matrix*, yaitu *precision*, *recall*, *F1-score*, *accuracy*, *balanced accuracy*, dan *macro F1-score*.

Precision mengukur ketepatan model dalam memprediksi kelas positif [23].

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall mengukur kemampuan model dalam menemukan seluruh data yang termasuk dalam kelas positif [23].

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-score merupakan rata-rata harmonis dari *precision* dan *recall* yang digunakan untuk menilai keseimbangan performa model [23].

$$F1\text{-}Score = 2 \times \frac{Precision \times Recal}{Precision + Recall} \quad (4)$$

Accuracy mengukur tingkat ketepatan prediksi model secara keseluruhan [23].

$$Accuracy = \frac{Jumlah\ Prediksi\ Benar}{Total\ Data} \quad (5)$$

Balanced Accuracy menghitung rata-rata nilai *recall* dari setiap kelas sehingga dapat memberikan evaluasi yang lebih adil pada data yang tidak seimbang [25].

$$Balanced\ Accuracy = \frac{Recall_{kelas1}+Recall_{kelas2}+Recall_{kelas3}}{Total\ Recal\ Kelas} \quad (6)$$

Macro F1-score merupakan rata-rata nilai *F1-score* dari seluruh kelas tanpa mempertimbangkan jumlah data pada masing-masing kelas [25].

$$Macro\ F1\text{-}score = \frac{F1-Score_{kelas1}+F1-Score_{kelas2}+F1-Score_{kelas3}}{Total\ F1-Score\ Kelas} \quad (7)$$

Hasil Dan Pembahasan

3.1 Hasil

a. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan melalui penyebaran kuesioner kepada mahasiswa Universitas Muslim Indonesia sebagai responden. Proses pengumpulan dibantu oleh mahasiswa Fakultas Ilmu Komputer angkatan 2024 yang mengumpulkan jawaban kuesioner dari responden, kemudian data yang terkumpul dikirimkan kepada peneliti melalui Gmail dalam bentuk file spreadsheet dan digabungkan menjadi satu dataset menggunakan Microsoft Excel. Dataset akhir terdiri dari 3419 data dengan 36 kolom yang mencakup variabel untuk memprediksi risiko stres, depresi, dan kecemasan, meliputi nama, asal

fakultas, umur, jenis kelamin, tinggi badan, berat badan, durasi tidur harian, durasi olahraga per pekan, durasi penggunaan media sosial, serta item DASS-21 yang terdiri dari tujuh pertanyaan stres, tujuh pertanyaan depresi, dan tujuh pertanyaan kecemasan, dengan kategori tingkat stres, depresi, dan kecemasan sebagai variabel target dalam proses klasifikasi.

b. Data Cleaning

Pada tahap *data cleaning*, dilakukan pemeriksaan terhadap nilai yang hilang (*missing value*) pada setiap variabel dalam dataset

Gambar 2. Missing Value

Berdasarkan Gambar 2, ditemukan bahwa beberapa atribut seperti *Asal Fakultas* memiliki nilai kosong yang ditunjukkan dengan jumlah data hilang sebanyak 937 entri. Untuk memastikan kualitas data tetap terjaga, penanganan *missing value* dilakukan dengan menghapus data yang tidak lengkap atau tidak valid sehingga tidak memengaruhi proses pelatihan model.

Selanjutnya dilakukan pengecekan terhadap data duplikat untuk memastikan tidak terdapat entri yang berulang dalam dataset.

index	Nama (Boleh menggunakan nama samaran)	ASAL FAKULTAS	UMUR	JENIS KELAMIN	TINGGI BADAN (CM)	BERAT BADAN (KG)	DURASI TIDUR SETIAP HARI (JAM)
1838	Sapri	HUKUM	21	Perempuan	160.0	48	6
1839	Sapri	HUKUM	21	Perempuan	160.0	48	6
1840	Sapri	HUKUM	21	Perempuan	160.0	48	6
2139	Karobi	NaN	25	Laki-laki	200.0	80	1
2141	reksa	NaN	20	Laki-laki	167.0	60	8
2142	besse	NaN	18	Perempuan	155.0	43	8
2146	dite	NaN	19	Laki-laki	174.0	68	7
2148	nara	NaN	19	Perempuan	155.0	49	6
2152	preatty	NaN	18	Perempuan	157.0	45	6
2153	fida	NaN	20	Perempuan	154.0	48	5

Gambar 3. Data Duplikat

Berdasarkan Gambar 3, menunjukkan bahwa terdapat 81 baris data duplikat yang memiliki kesamaan pada beberapa atribut, seperti nama, fakultas, umur, jenis kelamin, dan variabel lainnya. Data duplikat tersebut kemudian dihapus agar setiap baris merepresentasikan responden yang unik dan tidak menimbulkan bias dalam proses pelatihan model.

c. *Feature Engineering*

Pada tahap *feature engineering*, dilakukan pembentukan variabel baru yang lebih representatif dari informasi tinggi badan dan berat badan, yaitu *Berat Badan Ideal (BB_IDEAL)*. Variabel ini dihitung menggunakan rumus berat badan ideal yang disesuaikan berdasarkan jenis kelamin, dimana untuk responden laki-laki digunakan rumus $(Tinggi\ Badan - 100) \times 0{,}90$, sedangkan untuk responden perempuan menggunakan rumus $(Tinggi\ Badan - 100) \times 0{,}85$. Perbedaan koefisien tersebut bertujuan untuk menyesuaikan proporsi tubuh ideal sesuai standar kesehatan. Selanjutnya, dilakukan perhitungan selisih antara berat badan aktual dan berat badan ideal *(SELISIH_BB)*, yang kemudian dikategorikan menjadi tiga kelompok, yaitu *Kurus* apabila selisih kurang dari −5, *Ideal* apabila berada pada rentang −5 hingga 5, dan *Gemuk* apabila lebih dari 5. Variabel hasil transformasi ini digunakan untuk memperkaya informasi dataset sehingga dapat meningkatkan kemampuan model dalam mengidentifikasi pola yang berkaitan dengan kondisi kesehatan mental mahasiswa.

d. *Encoding Data*

Pada tahap ini dilakukan proses *encoding* pada seluruh data kategorikal untuk mengubahnya menjadi bentuk numerik agar dapat digunakan dalam pemodelan *machine learning*. Metode yang digunakan adalah *label encoding*, yaitu pemberian kode angka pada setiap kategori. Pada variabel yang bersifat ordinal,

pemberian kode dilakukan dengan tetap mempertahankan urutan tingkatannya agar representasi numerik mencerminkan kondisi yang sebenarnya**.**

e. Split Data

Pada tahap *split data*, dilakukan pembagian dataset menjadi data pelatihan (*training set*) dan data pengujian (*testing set*) menggunakan rasio 80:20

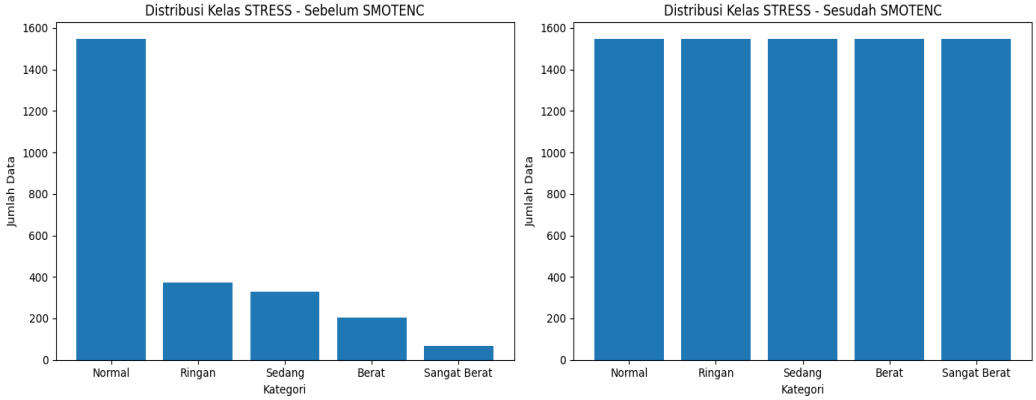
```
=== STRESS ===
Train: (2523, 27) (2523,)
Test : (631, 27) (631,)
```

Gambar 4. Split Data

Dari Gambar 4 total 3.154 data yang telah melalui proses *data cleaning* dan seleksi fitur, diperoleh sebanyak 2.523 data sebagai data latih dan 631 data sebagai data uji. Proses pembagian ini diterapkan secara konsisten pada masing-masing target klasifikasi, yaitu *Stress*, *Depresi*, dan *Anxiety*, sehingga model dapat dilatih menggunakan sebagian besar data dan diuji pada data yang belum pernah dilihat sebelumnya untuk mengevaluasi kinerja secara objektif.

f. Handle Imbalance

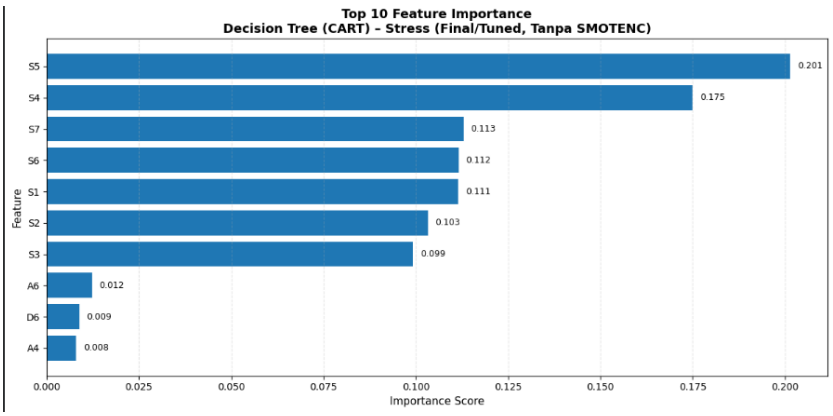
Pada tahap *handle imbalance*, dilakukan penanganan ketidakseimbangan distribusi kelas pada data target *Stress* menggunakan metode (SMOTENC).



Gambar 5. SMOTENC

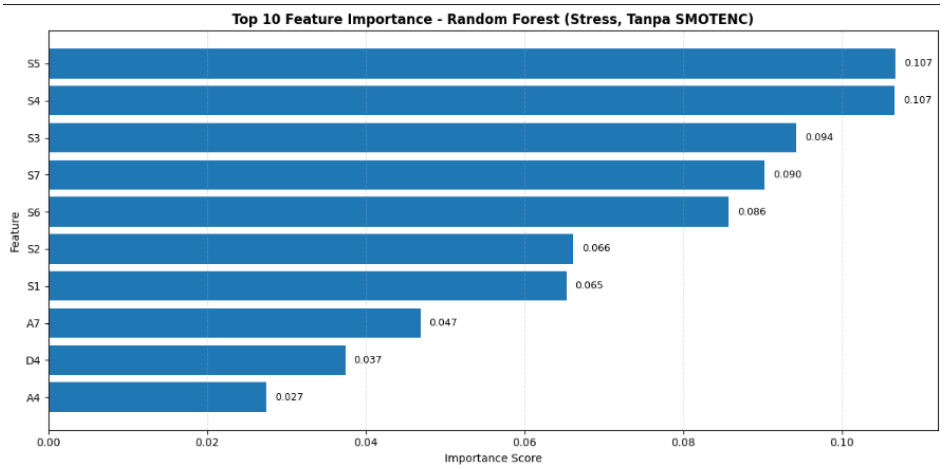
Berdasarkan Gambar 5, sebelum penerapan SMOTENC terlihat bahwa distribusi kelas tidak merata, dimana kategori *Normal* memiliki jumlah data jauh lebih besar dibandingkan kategori *Ringan*, *Sedang*, *Berat*, dan *Sangat Berat*. Kondisi ini berpotensi menyebabkan model bias terhadap kelas mayoritas. Setelah penerapan SMOTENC, distribusi data menjadi lebih seimbang karena metode ini menambahkan data sintetis pada kelas minoritas berdasarkan kemiripan karakteristik data. Dengan demikian, model dapat mempelajari pola dari setiap kelas secara lebih optimal dan menghasilkan performa klasifikasi yang lebih adil serta akurat.

3.2 Pembahasan



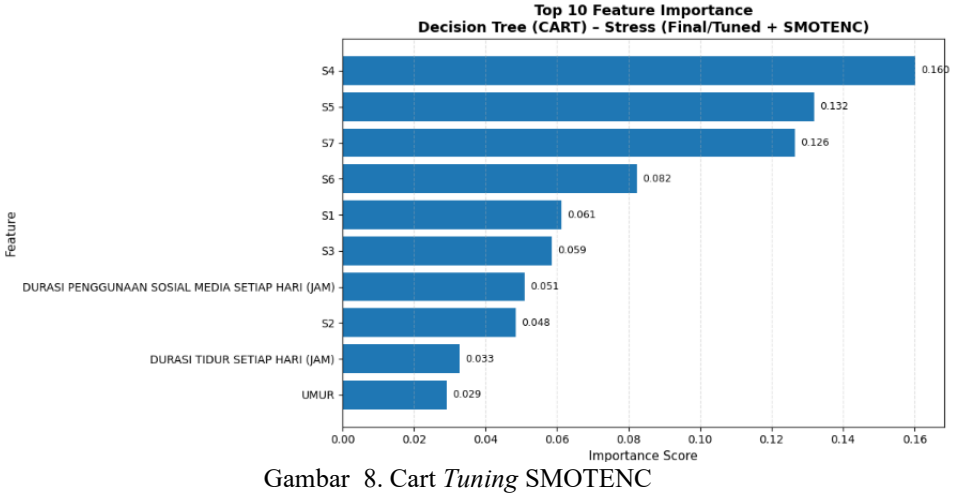
Gambar 6. Cart *Tuning* Tanpa SMOTENC

Berdasarkan Gambar 6, Visualisasi *feature importance* pada model *Decision Tree* CART untuk klasifikasi *Stress Tuning* tanpa SMOTENC menunjukkan bahwa variabel pertanyaan DASS-21 memiliki pengaruh paling dominan dalam proses klasifikasi. Berdasarkan grafik, fitur S5 memiliki nilai kepentingan tertinggi sebesar 0,201, diikuti oleh S4 sebesar 0,175, serta S7, S6, dan S1 dengan nilai di atas 0,11. Hal ini menunjukkan bahwa beberapa item pertanyaan stres memiliki kontribusi paling besar dalam menentukan tingkat klasifikasi stres mahasiswa. Sementara itu, variabel pendukung seperti A6, D6, dan A4 memiliki nilai kepentingan yang relatif kecil, sehingga pengaruhnya terhadap prediksi lebih rendah dibandingkan item utama stres. Secara keseluruhan, hasil ini menunjukkan bahwa model lebih banyak bergantung pada skor pertanyaan DASS-21 aspek stres dalam mengidentifikasi tingkat stres mahasiswa.



Gambar 7. *Random Forest Tuning* Tanpa SMOTENC

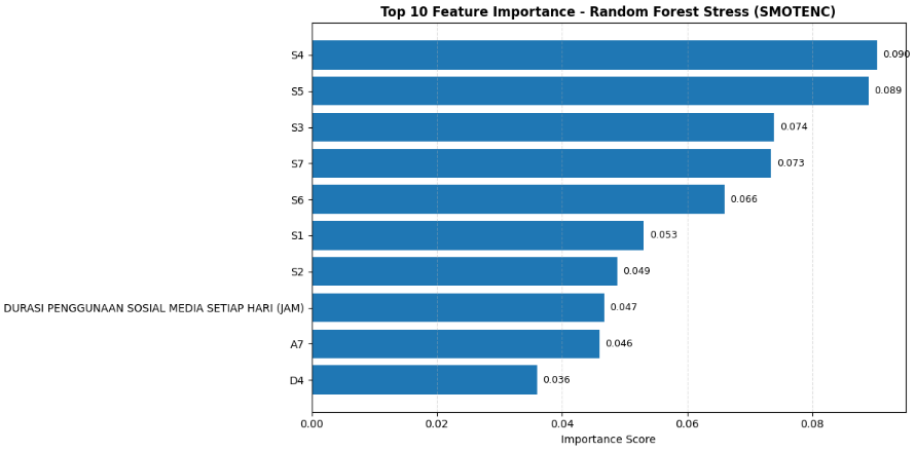
Berdasarkan Gambar 7, Visualisasi *feature importance* pada model *Random Forest* untuk klasifikasi *Stress* tanpa penerapan SMOTENC menunjukkan bahwa fitur S5 dan S4 memiliki nilai kepentingan tertinggi, masing-masing sebesar 0,107. Fitur lain yang juga memberikan kontribusi besar adalah S3, S7, dan S6, yang seluruhnya merupakan item pertanyaan DASS-21 aspek stres. Selain itu, beberapa variabel lain seperti A7, D4, dan A4 juga masuk dalam sepuluh fitur teratas, meskipun dengan nilai kontribusi yang lebih kecil. Hasil ini menunjukkan bahwa model *Random Forest* lebih banyak bergantung pada item utama stres dalam proses klasifikasi, dengan distribusi pengaruh fitur yang relatif merata dibandingkan *Decision Tree*. Secara keseluruhan, temuan ini menguatkan bahwa skor pertanyaan DASS-21 aspek stres menjadi faktor dominan dalam menentukan tingkat stres mahasiswa.



Gambar 8. Cart *Tuning* SMOTENC

Berdasarkan Gambar 8, Visualisasi *feature importance* pada model *Decision Tree* (CART) untuk klasifikasi *Stress* dengan skenario *final/tuning* dan penerapan SMOTENC menunjukkan bahwa fitur S4 memiliki nilai kepentingan tertinggi sebesar 0,168, diikuti oleh S5 sebesar 0,132 dan S7 sebesar 0,126. Hal ini menunjukkan bahwa beberapa item pertanyaan DASS-21 aspek stres tetap menjadi faktor dominan dalam proses klasifikasi meskipun data telah diseimbangkan. Selain itu, terlihat bahwa variabel pendukung

seperti Durasi Penggunaan Media Sosial per Hari, Durasi Tidur per Hari, dan Umur mulai muncul dalam sepuluh fitur teratas, meskipun dengan nilai kepentingan yang lebih rendah. Temuan ini menunjukkan bahwa setelah penerapan SMOTENC, model tidak hanya bergantung pada item utama stres, tetapi juga mulai mempertimbangkan faktor perilaku dan demografis dalam menentukan tingkat stres mahasiswa.



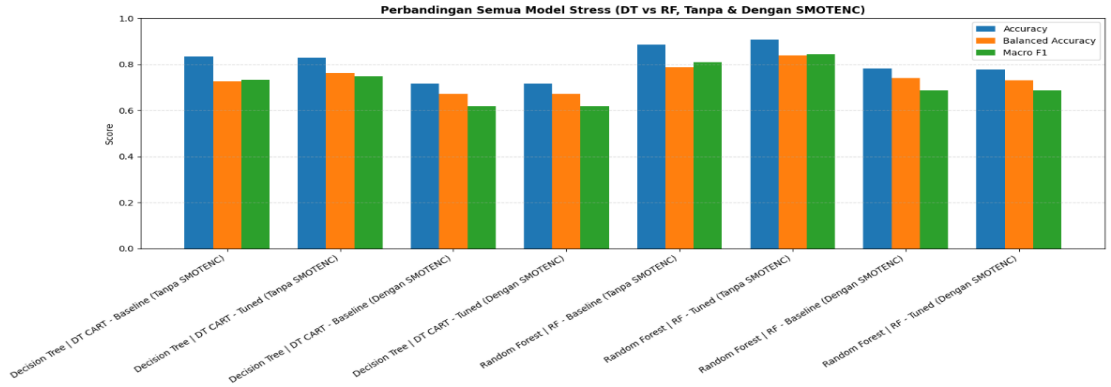
Gambar 9. *Random Forest Tuning* SMOTENC

Berdasarkan Gambar 9. Visualisasi *feature importance* pada model *Random Forest* untuk klasifikasi *Stress* dengan penerapan SMOTENC menunjukkan bahwa fitur S4 dan S5 memiliki nilai kepentingan tertinggi, masing-masing sebesar 0,090 dan 0,089. Fitur lain yang juga berkontribusi signifikan adalah S3, S7, dan S6, yang semuanya merupakan item pertanyaan DASS-21 aspek stres. Selain itu, variabel pendukung seperti Durasi Penggunaan Media Sosial per Hari, A7, dan D4 juga termasuk dalam sepuluh fitur teratas, meskipun dengan nilai kepentingan yang lebih rendah. Hal ini menunjukkan bahwa model *Random Forest* tidak hanya bergantung pada item utama stres, tetapi juga mempertimbangkan variabel perilaku dan aspek emosional lainnya dalam proses klasifikasi. Secara keseluruhan, distribusi nilai kepentingan fitur pada *Random Forest* terlihat lebih merata dibandingkan *Decision Tree*, yang menunjukkan kemampuan model dalam menangkap pola data secara lebih komprehensif.

Untuk memperoleh gambaran performa model secara menyeluruh, dilakukan perbandingan antara seluruh skenario pengujian yang meliputi algoritma *Decision Tree* dan *Random Forest*, baik pada kondisi *baseline* maupun *tuning*, serta dengan dan tanpa penerapan metode SMOTENC. Perbandingan ini dilakukan menggunakan tiga metrik evaluasi utama, yaitu *accuracy*, *balanced accuracy*, dan *macro F1-score*. Ringkasan hasil evaluasi performa seluruh model ditunjukkan pada Tabel 1, sedangkan visualisasi perbandingan performa model ditampilkan pada Gambar 10.

Tabel 1. Hasil Keseluruhan Model

Model	**Algoritma**	***accuracy***	***Balanced Accuracy***	***Macro F1***
Baseline Tanpa SMOTENC	*Decision Tree*	0.8336	0.7261	0.7345
Tuning Tanpa SMOTENC	*Decision Tree*	0.8288	0.7637	0.7475
Baseline SMOTENC	*Decision Tree*	0.7163	0.6715	0.6175
Tuning SMOTENC	*Decision Tree*	0.7163	0.6715	0.6175
Baseline Tanpa SMOTENC	*Random Forest*	0.8843	0.7871	0.8085
Tuning Tanpa SMOTENC	*Random Forest*	0.9065	0.8400	0.8427
Baseline SMOTENC	*Random Forest*	0.7813	0.7410	0.6875
Tuning SMOTENC	*Random Forest*	0.7765	0.7311	0.6862



Gambar 10. Visualisasi Keseluruhan Model

Berdasarkan Tabel 1, terlihat bahwa algoritma *Random Forest* secara umum menghasilkan performa yang lebih baik dibandingkan *Decision Tree* pada hampir seluruh skenario pengujian. Model *Random Forest tuning* tanpa SMOTENC menunjukkan performa terbaik dengan nilai *accuracy* sebesar 0,9065, *balanced accuracy* sebesar 0,8400, dan *macro F1-score* sebesar 0,8427. Hasil ini menunjukkan bahwa pendekatan *ensemble learning* pada *Random Forest* mampu menghasilkan model klasifikasi yang lebih stabil dan akurat dalam menangkap pola data tingkat stres mahasiswa.

Selain itu, model *Random Forest baseline* tanpa SMOTENC juga menunjukkan performa yang cukup tinggi dengan nilai *accuracy* sebesar 0,8843, *balanced accuracy* sebesar 0,7871, dan *macro F1-score* sebesar 0,8085. Hal ini menunjukkan bahwa bahkan tanpa proses optimasi parameter, algoritma *Random Forest* telah mampu memberikan kinerja klasifikasi yang baik.

Sebaliknya, model *Decision Tree* cenderung menghasilkan performa yang lebih rendah dibandingkan *Random Forest*. Nilai *accuracy* pada model *Decision Tree* berada pada rentang 0,7163 hingga 0,8336, dengan *macro F1-score* berkisar antara 0,6175 hingga 0,7475. Penurunan performa terutama terlihat pada skenario dengan penerapan SMOTENC, di mana nilai evaluasi model menjadi lebih rendah dibandingkan skenario tanpa penyeimbangan data.

Secara keseluruhan, hasil perbandingan ini menunjukkan bahwa algoritma *Random Forest* lebih efektif dan stabil dalam mengklasifikasikan tingkat stres mahasiswa dibandingkan *Decision Tree* pada berbagai skenario pengujian. Visualisasi perbandingan performa model pada Gambar 18 juga memperlihatkan bahwa nilai metrik evaluasi pada model *Random Forest* secara konsisten lebih tinggi dibandingkan model *Decision Tree*.

Kesimpulan

Penelitian ini bertujuan untuk mengklasifikasikan tingkat stres mahasiswa Universitas Muslim Indonesia berdasarkan instrumen DASS-21 serta beberapa variabel pendukung lainnya. Hasil analisis menunjukkan bahwa distribusi kelas tingkat stres pada dataset tidak seimbang, di mana kategori Normal memiliki jumlah data yang paling dominan dibandingkan kategori lainnya. Berdasarkan hasil pengujian model, algoritma *Random Forest* dengan skenario *tuning* tanpa penerapan SMOTENC menghasilkan performa terbaik dengan nilai *accuracy* sebesar 0,9065, *balanced accuracy* sebesar 0,8400, dan *macro F1-score* sebesar 0,8427. Hasil tersebut menunjukkan bahwa model ini paling optimal dalam mengklasifikasikan tingkat stres mahasiswa secara multi-kelas. Penerapan metode SMOTENC mampu membantu menyeimbangkan distribusi data, namun tidak selalu memberikan peningkatan performa yang signifikan pada seluruh skenario pengujian. Selain itu, hasil analisis *feature importance* menunjukkan bahwa item pertanyaan DASS-21 pada aspek stres, khususnya S4 dan S5, menjadi faktor yang paling dominan dalam proses klasifikasi. Meskipun demikian, penelitian ini masih memiliki keterbatasan pada cakupan data yang hanya berasal dari satu institusi serta jumlah variabel perilaku yang masih terbatas. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas cakupan responden, menambahkan variabel psikososial lainnya, serta menguji algoritma klasifikasi yang lebih kompleks atau pendekatan *ensemble learning* yang lebih lanjut guna meningkatkan kemampuan generalisasi dan akurasi model dalam mendeteksi kondisi kesehatan mental mahasiswa secara lebih komprehensif.

Daftar Pustaka

[1] Y. Afrillia, R. Putra Fhonna, and M. Rahma, "Yesy Afrillia, et.al Klasifikasi Kesehatan Mental Remaja Tingkat SMA di Kota Lhokseumawe Menggunakan Algoritma Random Forest," vol. 6, no. 3, pp. 2025–4019, doi: 10.55338/jpkmn.v6i3.6377.

[2] S. Widyawati, D. Mayasaroh, S. L. Aqila, K. N. Iriantina, M. Y. Al Islam, and J. T. Nugraha, "Faktor-faktor yang Berkaitan dengan Kesehatan Mental Mahasiswa," *J. ISO J. Ilmu Sos. Polit. dan Hum.*, vol. 5, no. 1, p. 11, Jun. 2025, doi: 10.53697/iso.v5i1.2534.

[3] J. Jayadi, V. Hafizh, C. Putra, A. R. Raharja, and M. Al-husaini, "Deteksi Dini Kesehatan Mental Mahasiswa dengan Machine Learning : Perbandingan Algoritma Decision Tree dan Random Forest Pendahuluan Tinjauan Pustaka," vol. 16, no. 1, pp. 134–141, 2026, doi: 10.31602/tji.v17i1.21251.

[4] 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," no. https://ejournal.upnvj.ac.id/informatik/issue/view/400, Apr. 2024, doi: https://doi.org/10.52958/iftk.v20i1.9158.

[5] R. Wardhani and N. Nafi, "Comparison of Machine Learning for Mental Health Identification (The DASS-21 Questionnaire)," vol. 18, no. 1, 2025, doi: 10.24036/jtip.v18i1.860.

[6] C. Nisa, "PREDIKSI KESEHATAN MENTAL MAHASISWA UNIVERSITAS YUDHARTA PASURUAN DENGAN KLASIFIKASI DECISION TREE," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 3S1, Oct. 2025, doi: 10.23960/jitet.v13i3S1.7925.

[7] S. G. Dzakiyyah *et al.*, "Prediction of Student Academic Stress Levels Using the Decision Tree Algorithm and Particle Swarm Optimization," vol. 8, no. 3, pp. 579–591, 2025, doi: 10.24014/ijaidm.v8i3.38081.

[8] M. A. Hakim and N. V. Aristawati, "Mengukur depresi, kecemasan, dan stres pada kelompok dewasa awal di Indonesia: Uji validitas dan reliabilitas konstruk DASS-21," *J. Psikol. Ulayat*, vol. 10, no. 2, pp. 232–250, Oct. 2023, doi: 10.24854/jpu553.

[9] M. F. Rahmatullah, P. L. Lokapitasari Belluano, and H. Darwis, "Analisis Sentimen Review Aplikasi di Google Play Store Menggunakan Random Forest," *LINIER Lit. Inform. dan Komput.*, vol. 2, no. 3, pp. 380–389, Oct. 2025, doi: 10.33096/linier.v2i3.3149.

[10] Y. Salim and L. Budi, "Analisis Sentimen terhadap Komentar Negatif di Media Sosial Facebook dengan Metode Klasifikasi Naïve Bayes," vol. 1, no. 4, pp. 259–265, 2020, doi: 10.33096/busiti.v1i4.666.

[11] A. Anjani and Y. Yamasari, "Klasifikasi Tingkat Stres Mahasiswa Menggunakan Metode Berbasis Tree," *J. Informatics Comput. Sci.*, vol. 05, Jul. 2023, doi: https://doi.org/10.26740/jinacs.v5n01.p83-89.

[12] R. Chandra *et al.*, "Decision Tree-Based Approach for Predicting Mental Health Symptoms in University Students : A Case Study at IT Del," vol. 1, no. 1, pp. 20–29, 2025, doi: https://doi.org/10.54074/jicsa.v1i01.7.

[13] S. I. Ovi, J. Hossain, R. A. Rahi, and F. Akter, "Protecting Student Mental Health with a Context-Aware Machine Learning Framework for Stress Monitoring," 2025, [Online]. Available: http://arxiv.org/abs/2508.01105

[14] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, pp. 86716–86727, Jun. 2024, doi: 10.1109/ACCESS.2024.3416838.

[15] A. Amelia, L. Nur, and H. Darwis, "Analisis Sentimen Masyarakat Terhadap Sistem Pembayaran Mypertamina dengan Metode Random Forest , SVM , dan Naïve Bayes," vol. 1, no. 1, pp. 28–44, 2024, [Online]. Available: https://doi.org/10.33096/linier.v1i1.2269

[16] M. Erkamim, "Sentiment Analysis of Shopee App Reviews Using Random Forest and Support Vector Machine," vol. 15, no. 3, pp. 427–435, 2023, doi: 10.33096/ilkom.v15i3.1610.427-435.

[17] S. Penguatan, E. Kelautan, D. Wakatobi, M. Fuad, and I. Marsela, "Pendekatan Analitik Berbasis Big Data dan Model Prediktif Sebagai," *LINIER Lit. Inform. dan Komput.*, vol. 2, no. 4, pp. 569–578, 2025, doi: 10.33096/linier.v2i4.3342.

[18] E. Priyanto, E. Itje, L. Alexander, and N. Islam, "Decision Tree C4 . 5 Performance Improvement using Synthetic Minority Oversampling Technique (SMOTE) and K-Nearest Neighbor for Debtor Eligibility Evaluation," *Ilk. J. Ilm.*, vol. 15, no. 2, pp. 373–381, 2023, doi: 10.33096/ilkom.v15i2.1676.373-381.

[19] E. E. Pratiwi, A. R. Aisy, R. Rahmaddeni, and N. Ananta, "Klasifikasi Kesehatan Mental Pada Usia Remaja Menggunakan Metode Svm," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6232.

[20] R. Puspitasari, T. Amaliah, and H. Darwis, "Comparative Machine Learning Models for Dementia Prediction Using SMOTE," *Int. J. Artif. Intell. Med. Issues*, vol. 3, no. 2, pp. 79–90, Nov. 2025, doi:

10.56705/ijaimi.v3i2.351.
[21] M. I. E. Saputra Troy, S. R. Jabir, and S. Anraeni, "Evaluation of Multi-Class Classification Performance Lung Cancer Through K-NN and SVM Approach," *Ilk. J. Ilm.*, vol. 17, no. 1, pp. 27–33, Apr. 2025, doi: 10.33096/ilkom.v17i1.2464.27-33.
[22] N. Anizah, Y. Salim, and L. Budi, "Analisis Sentimen Terhadap Event Big Sale 11 . 11 Shopee di Media Sosial Instagram Mengggunakan Metode Naïve Bayes," *Bul. Sist. Inf. dan Teknol. Islam*, vol. 4, no. 1, pp. 25–34, 2023, doi: 10.33096/busiti.v4i1.1309.
[23] N. Yanti Sangaji and I. As, "Analisis Pengguna Facebook Terhadap Objek Wisata di Maluku Tengah Menggunakan Metode K-Nearest Neighbor INFORMASI ARTIKEL ABSTRAK," *Bul. Sist. Inf. dan Teknol. Islam*, vol. 3, no. 3, pp. 196–202, 2022.
[24] G. D. Ramadhan and H. Nugroho, "Analisis Perubahan Tutupan Lahan Menggunakan Algoritma CART untuk Evaluasi Kesesuaian Lahan Terhadap RTRW Kabupaten Tangerang," *J. Rekayasa Hijau*, vol. 9, no. 1, pp. 44–57, Mar. 2025, doi: 10.26760/jrh.v9i1.44-57.
[25] K. Leung, "Mikro, Makro & Rata-rata Tertimbang dari Skor F1, Dijelaskan Secara Jelas," Towards Data Science. [Online]. Available: https://towardsdatascience.com/micro-macro-weighted-averages-of-f1-score-clearly-explained-b603420b292f/