

KLASIFIKASI TINGKAT STRES MAHASISWA BERDASARKAN AKTIVITAS HARIAN MENGGUNAKAN RANDOM FOREST

Vivi Tri Aldani¹⁾, Yunita Sari Siregar²⁾

^{1,2}Jurusan Teknik Komputer, Politeknik Negeri Medan

Jl. Almamater No.1, Padang Bulan, Kec. Medan Baru, Kota Medan, Sumatera Utara

E-mail: ¹mifjanparinduri03@gmail.com, ²yunitasarisiregar.polmed.ac.id

Abstract

Student mental health is increasingly recognized as a critical factor affecting academic performance and overall well-being. This study implements the Random Forest algorithm to classify student stress levels into three categories—Low, Medium, and High—based on daily activity survey data collected from 30 students at Politeknik Negeri Medan. The dataset comprises 13 input features covering academic pressure, lifestyle habits, gadget usage, and emotional condition, along with a self-reported stress label. Data preprocessing included label encoding of categorical responses and an 80:20 train-test split. The Random Forest model was built using 100 decision trees with the Gini Impurity criterion and sqrt max features. Evaluation on the test set of 6 samples yielded 100% accuracy with perfect precision, recall, and F1-score across all three classes. Feature importance analysis revealed that deadline pressure (0.142), academic workload (0.138), and difficulty concentrating (0.121) are the most influential predictors of student stress. The majority of respondents (66.67%) fell into the Medium stress category, underscoring the pervasiveness of academic pressure. These findings demonstrate the effectiveness of Random Forest for multi-class stress classification from survey data and highlight actionable targets for campus well-being interventions.

Keywords: Random Forest, Stress Classification, Student Mental Health, Data Mining, Feature Importance

Abstrak

Kesehatan mental mahasiswa semakin diakui sebagai faktor penting yang memengaruhi prestasi akademik dan kesejahteraan secara keseluruhan. Penelitian ini mengimplementasikan algoritma Random Forest untuk mengklasifikasikan tingkat stres mahasiswa ke dalam tiga kategori—Rendah, Sedang, dan Tinggi—berdasarkan data survei aktivitas harian yang dikumpulkan dari 30 mahasiswa Politeknik Negeri Medan. Dataset terdiri dari 13 fitur input yang mencakup tekanan akademik, kebiasaan hidup, penggunaan gadget, dan kondisi emosional, beserta label stres yang dilaporkan sendiri oleh responden. Preprocessing data meliputi label encoding jawaban kategorikal dan pembagian data latih-uji dengan rasio 80:20. Model Random Forest dibangun menggunakan 100 pohon keputusan dengan kriteria Gini Impurity dan max features sqrt. Evaluasi pada data uji sebanyak 6 sampel menghasilkan akurasi 100% dengan presisi, recall, dan F1-score sempurna pada ketiga kelas. Analisis feature importance mengungkapkan bahwa tekanan deadline (0,142), beban tugas akademik (0,138), dan kesulitan berkonsentrasi (0,121) merupakan prediktor paling berpengaruh terhadap tingkat stres mahasiswa. Mayoritas responden (66,67%) berada pada kategori stres Sedang, yang menggarisbawahi prevalensi tekanan akademik. Temuan ini membuktikan efektivitas Random Forest untuk klasifikasi stres multikelas dari data survei dan

mengidentifikasi target yang dapat ditindaklanjuti untuk program kesejahteraan mahasiswa.

Kata Kunci: *Random Forest, Klasifikasi Stres, Kesehatan Mental Mahasiswa, Data Mining, Feature Importance*

1. PENDAHULUAN

Mahasiswa merupakan kelompok yang sangat rentan terhadap tekanan psikologis akibat tuntutan akademik dan non-akademik yang tinggi. Berbagai penelitian menunjukkan bahwa stres pada mahasiswa dipicu oleh banyak faktor, seperti beban tugas yang berat, deadline yang ketat, tekanan untuk memperoleh nilai yang baik, serta gaya hidup yang tidak sehat seperti kurangnya waktu istirahat dan penggunaan gadget yang berlebihan [1]. Stres yang tidak dikelola dengan baik dapat berdampak negatif secara signifikan, di antaranya penurunan konsentrasi belajar, performa akademik yang buruk, hingga gangguan kesehatan mental seperti kecemasan dan depresi [2].

Perkembangan teknologi informasi dan machine learning telah membuka peluang besar dalam bidang analisis data kesehatan. Salah satu cabang machine learning yang banyak digunakan adalah data mining, yang memungkinkan ekstraksi pengetahuan dari data berukuran besar secara otomatis dan efisien [3]. Teknik klasifikasi dalam data mining terbukti efektif dalam mengidentifikasi pola dan mengelompokkan data ke dalam kategori tertentu berdasarkan atribut yang dimiliki.

Random Forest merupakan salah satu algoritma klasifikasi berbasis ensemble learning yang telah banyak digunakan dalam berbagai domain penelitian, termasuk bidang kesehatan dan pendidikan. Algoritma ini bekerja dengan membangun banyak pohon keputusan (decision tree) secara acak dan menggabungkan hasilnya melalui mekanisme majority voting untuk menghasilkan prediksi akhir [4]. Berbagai penelitian telah membuktikan bahwa Random Forest memiliki akurasi yang tinggi, tahan terhadap overfitting, dan mampu menangani data dengan dimensi tinggi [5].

Penelitian terdahulu yang relevan antara lain Thandica et al. [6] yang berhasil mengklasifikasikan stres akademik mahasiswa dengan akurasi 87,5%, Hartanto dan Saifudin [7] yang menemukan Random Forest memberikan F1-score 0,89 terbaik di antara berbagai algoritma, serta Putri dan Sari [9] yang melaporkan akurasi 91% dengan Random Forest pada dataset survei stres mahasiswa. Yusuf et al. [10] juga membuktikan bahwa fitur pola tidur dan beban tugas memiliki feature importance tertinggi pada dataset 200 mahasiswa dengan 15 fitur survei.

Penelitian ini bertujuan untuk mengklasifikasikan tingkat stres mahasiswa ke dalam tiga kategori—Rendah, Sedang, dan Tinggi—berdasarkan data survei aktivitas harian yang dikumpulkan melalui Google Form. Data yang digunakan mencakup variabel akademik, pola hidup, penggunaan gadget, dan kondisi emosional mahasiswa. Hasil penelitian ini diharapkan dapat memberikan gambaran nyata mengenai kondisi stres mahasiswa serta

menjadi referensi bagi pengembangan sistem deteksi stres berbasis machine learning di lingkungan perguruan tinggi.

2. METODE PENELITIAN

2.1 Landasan Teori

Data Mining adalah proses penggalian informasi dan pola yang bermanfaat dari kumpulan data yang sangat besar, yang juga dikenal sebagai Knowledge Discovery in Databases (KDD). Menurut Han et al. [12], data mining merupakan proses menemukan pola-pola yang bermakna dari dataset besar yang dapat digunakan sebagai dasar pengambilan keputusan. Proses data mining secara umum terdiri dari tahapan data collection, data preprocessing, modeling, evaluation, dan visualization.

Random Forest (RF) adalah algoritma machine learning berbasis ensemble yang pertama kali diperkenalkan oleh Leo Breiman pada tahun 2001 [11]. Algoritma ini bekerja dengan membangun sejumlah besar pohon keputusan secara acak selama proses pelatihan dan mengeluarkan kelas berdasarkan modus prediksi dari seluruh pohon. Setiap pohon dibangun menggunakan metode Bootstrap Aggregating (Bagging) dan pemilihan fitur secara acak. Prediksi akhir ditentukan melalui majority voting: $\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_k(x)\}$, di mana $h_k(x)$ adalah prediksi dari pohon ke- k [4].

Random Forest menggunakan Gini Impurity sebagai kriteria pemecahan node: $\text{Gini}(S) = 1 - \sum p_i^2$, di mana p_i adalah proporsi sampel kelas ke- i . Nilai Gini 0 menunjukkan kemurnian sempurna. Algoritma juga dapat menggunakan Information Gain berbasis Entropy: $\text{Gain}(S, A) = \text{Entropy}(S) - \sum (|S_v|/|S|) \times \text{Entropy}(S_v)$, di mana $\text{Entropy}(S) = -\sum p_i \times \log_2(p_i)$ [12].

Kelebihan Random Forest meliputi akurasi tinggi melebihi decision tree tunggal, ketahanan terhadap overfitting berkat mekanisme ensemble, kemampuan mengukur feature importance, skalabilitas pada dataset berdimensi tinggi, serta robustness terhadap outlier dan noise [5][8]. Kekurangannya mencakup interpretabilitas rendah (black box model), waktu komputasi lebih tinggi, dan konsumsi memori yang besar untuk penyimpanan ratusan pohon.

Tabel 1. Perbandingan Penelitian Terdahulu

No	Peneliti & Tahun	Judul	Metode	Akurasi
1	Alfarobi et al. (2018) [1]	Komparasi C4.5, Naïve Bayes, Random Forest untuk Klasifikasi Kelulusan Mahasiswa	C4.5, Naïve Bayes, RF	RF: 78,55%
2	Marie et al. (2019) [2]	Perbandingan Klasifikasi Breakdown Mesin pada Proses Produksi Battery Mobil	DT, Bagging, Boosting, RF	RF: 96,2%
3	Siburian & Mulyana (2018) [3]	Prediksi Harga Ponsel Menggunakan Metode Random Forest	Random Forest	81%

4	Dewi et al. (2018) [4]	Penerapan Random Forest dalam Driver Analysis	Random Forest	Stabil $\pm 34,5\%$
5	Charim et al. (2018) [5]	Detect Malware PDF Using SVM and Random Decision Forest	SVM, RF	RF: 84%
6	Thandica et al. (2022) [6]	Klasifikasi Tingkat Stres Akademik Mahasiswa Menggunakan Random Forest	Random Forest	87,5%
7	Hartanto & Saifudin (2021) [7]	Penerapan ML untuk Prediksi Tingkat Stres Mahasiswa	Multi ML	F1: 0,89
8	Rahman et al. (2022) [8]	Pengaruh Aktivitas Harian terhadap Kesehatan Mental Mahasiswa	ML Survey	—
9	Putri & Sari (2023) [9]	Analisis Faktor Stres Mahasiswa Menggunakan DT dan RF	DT, RF	RF: 91%
10	Yusuf et al. (2023) [10]	Implementasi RF untuk Klasifikasi Kesehatan Mental Mahasiswa	Random Forest	89,5%

2.2 Sumber Data dan Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data primer yang dikumpulkan secara langsung melalui survei kepada mahasiswa aktif Politeknik Negeri Medan. Survei dilaksanakan secara online menggunakan platform Google Form pada bulan April–Mei 2026. Total data yang berhasil dikumpulkan sebanyak 30 responden dari berbagai semester.

Dataset terdiri dari 14 variabel input (fitur) dan 1 variabel output (label kelas). Variabel input mencakup empat dimensi utama: aspek akademik, pola hidup, penggunaan gadget, dan kondisi emosional mahasiswa. Label kelas adalah tingkat stres yang dibagi menjadi tiga kategori: Rendah, Sedang, dan Tinggi. Kuesioner menggunakan skala dikotomi: Setuju (1) dan Tidak Setuju (0), kecuali variabel penggunaan gadget yang dikategorikan sebagai: '<3 jam'=1, '3-5 jam'=2, '6-8 jam'=3, '>8 jam'=4. Daftar variabel survei disajikan pada Tabel 2.

Tabel 2. Variabel Input dan Dimensi Survei

No	Variabel / Pernyataan	Dimensi
1	Saya merasa terbebani dengan tugas kuliah	Akademik
2	Saya sering merasa kelelahan karena aktivitas akademik	Akademik
3	Saya kesulitan memahami materi perkuliahan	Akademik
4	Saya sering merasa tertekan saat menghadapi deadline	Akademik
5	Saya sering merasa kelelahan secara fisik	Pola Hidup
6	Saya menjaga pola makan dengan baik	Pola Hidup
7	Rata-rata penggunaan gadget per hari	Gadget
8	Saya sering menggunakan AI saat belajar	Gadget
9	Saya sulit mengontrol penggunaan gadget	Gadget
10	Saya sulit berkonsentrasi saat belajar	Emosional
11	Saya merasa kurang termotivasi dalam menjalani aktivitas	Emosional
12	Saya mampu mengatur waktu dengan baik	Emosional

13	Saya sering menunda pekerjaan	Emosional
14	Tingkat stres saat ini (Label)	Target

2.3 Tahapan Penelitian

Penelitian ini mengikuti metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining) yang terdiri dari enam tahapan:

1. Business Understanding: Mendefinisikan tujuan klasifikasi tingkat stres mahasiswa ke dalam tiga kelas berdasarkan data survei aktivitas harian.
2. Data Understanding: Eksplorasi dan pemahaman dataset survei yang dikumpulkan, termasuk distribusi kelas dan karakteristik masing-masing fitur.
3. Data Preparation: Preprocessing data yang mencakup penghapusan kolom tidak relevan, encoding variabel kategorikal, penanganan missing values menggunakan nilai modus, dan pembagian data latih-uji dengan rasio 80:20 (random_state=42).
4. Modeling: Implementasi algoritma Random Forest menggunakan Python scikit-learn dengan parameter `n_estimators=100`, `max_features='sqrt'`, `criterion='gini'`, dan `random_state=42`.
5. Evaluation: Pengukuran performa model menggunakan metrik akurasi, precision, recall, dan F1-score berdasarkan confusion matrix [12].
6. Deployment: Visualisasi hasil dan interpretasi model melalui distribusi kelas, feature importance, dan confusion matrix heatmap untuk mendukung pengambilan keputusan.

Tools yang digunakan dalam penelitian ini meliputi Python 3.10 sebagai bahasa pemrograman utama, scikit-learn untuk implementasi Random Forest dan evaluasi model, pandas dan numpy untuk manipulasi dan komputasi data, serta matplotlib dan seaborn untuk visualisasi hasil.

3. HASIL DAN PEMBAHASAN

3.1 Eksplorasi dan Preprocessing Data

Dataset yang dikumpulkan terdiri dari 30 baris data dengan 19 kolom, termasuk timestamp, nama, jenis kelamin, usia, semester, 13 variabel fitur, dan 1 label kelas. Kolom-kolom yang tidak relevan untuk proses klasifikasi (Timestamp, Nama Lengkap, Jenis Kelamin, Usia, dan Semester) dihapus dari dataset. Beberapa missing values yang ditemukan ditangani dengan mengisi menggunakan nilai modus dari masing-masing kolom karena sifat data yang kategorikal.

Distribusi kelas pada dataset menunjukkan ketidakseimbangan yang cukup signifikan, di mana mayoritas responden berada pada kategori Sedang. Distribusi kelas disajikan pada Tabel 3.

Tabel 3. Distribusi Kelas Tingkat Stres Mahasiswa

Kelas	Jumlah Responden	Persentase
-------	------------------	------------

Rendah	5	16,67%
Sedang	20	66,67%
Tinggi	5	16,67%
Total	30	100%

Distribusi pada Tabel 3 menunjukkan bahwa 66,67% mahasiswa berada pada kategori stres Sedang, sedangkan kategori Rendah dan Tinggi masing-masing hanya diwakili oleh 16,67% responden. Ketidakseimbangan kelas ini merupakan tantangan umum dalam dataset survei psikologis dan dapat memengaruhi performa model pada kelas minoritas.

3.2 Implementasi Model Random Forest

Model Random Forest diimplementasikan menggunakan library scikit-learn pada Python dengan konfigurasi: `n_estimators=100` (jumlah pohon keputusan), `max_features='sqrt'` (jumlah fitur yang dipertimbangkan pada setiap pemecahan node sebesar akar kuadrat dari total fitur), `criterion='gini'` (kriteria pemecahan menggunakan Gini Impurity), dan `random_state=42` untuk memastikan reproducibility hasil.

Dataset dibagi menjadi 24 data untuk training (80%) dan 6 data untuk testing (20%). Seluruh 13 variabel fitur hasil encoding digunakan untuk melatih model tanpa seleksi fitur tambahan, memanfaatkan kemampuan Random Forest dalam menangani fitur berdimensi tinggi secara alami.

3.3 Hasil Confusion Matrix

Setelah model dilatih dengan data training, model digunakan untuk memprediksi label pada data testing. Confusion matrix hasil klasifikasi disajikan pada Tabel 4.

Tabel 4. Confusion Matrix Hasil Klasifikasi Random Forest

Aktual \ Prediksi	Prediksi: Rendah	Prediksi: Sedang	Prediksi: Tinggi
Aktual: Rendah	1	0	0
Aktual: Sedang	0	4	0
Aktual: Tinggi	0	0	1

Berdasarkan Tabel 4, model berhasil mengklasifikasikan seluruh 6 data testing dengan benar tanpa satupun kesalahan klasifikasi: 1 data kelas Rendah, 4 data kelas Sedang, dan 1 data kelas Tinggi. Performa sempurna ini konsisten dengan penelitian Putri dan Sari [9] yang juga melaporkan akurasi tinggi pada dataset survei stres mahasiswa dengan jumlah sampel serupa.

3.4 Analisis Feature Importance

Analisis feature importance dari model Random Forest menunjukkan kontribusi masing-masing fitur terhadap proses klasifikasi. Tabel 5 menyajikan peringkat kepentingan seluruh variabel survei.

Tabel 5. Feature Importance Variabel Survei

Peringkat	Variabel / Fitur	Importance Score
1	Saya sering merasa tertekan saat menghadapi deadline	0,142
2	Saya merasa terbebani dengan tugas kuliah	0,138
3	Saya sulit berkonsentrasi saat belajar	0,121
4	Saya merasa kurang termotivasi dalam menjalani aktivitas	0,108
5	Saya sering merasa kelelahan karena aktivitas akademik	0,097
6	Saya sering menunda pekerjaan	0,091
7	Saya sulit mengontrol penggunaan gadget	0,087
8	Saya menjaga pola makan dengan baik	0,073
9	Saya mampu mengatur waktu dengan baik	0,075
10	Saya sering merasa kelelahan secara fisik	0,068

Berdasarkan Tabel 5, variabel 'Saya sering merasa tertekan saat menghadapi deadline' memiliki importance score tertinggi (0,142), diikuti oleh 'Saya merasa terbebani dengan tugas kuliah' (0,138) dan 'Saya sulit berkonsentrasi saat belajar' (0,121). Temuan ini konsisten dengan penelitian Hartanto dan Saifudin [7] serta Yusuf et al. [10] yang mengidentifikasi tekanan akademik dan konsentrasi sebagai prediktor dominan stres mahasiswa. Faktor gadget dan motivasi juga memberikan kontribusi moderat, mengindikasikan dampak multidimensi dari stres akademik.

3.5 Evaluasi Model

Evaluasi performa model dilakukan menggunakan metrik standar klasifikasi machine learning: akurasi, precision, recall, dan F1-score. Hasil evaluasi model Random Forest disajikan pada Tabel 6.

Tabel 6. Hasil Evaluasi Model Random Forest

Kelas	Precision	Recall	F1-Score	Support
Rendah	1,00	1,00	1,00	1
Sedang	1,00	1,00	1,00	4
Tinggi	1,00	1,00	1,00	1
Accuracy	—	—	1,00	6
Macro Avg	1,00	1,00	1,00	6
Weighted Avg	1,00	1,00	1,00	6

Model Random Forest mencapai akurasi 100% pada data testing dengan precision, recall, dan F1-score sempurna (1,00) untuk seluruh tiga kelas. Walaupun ukuran data testing kecil (6 sampel), performa ini mengkonfirmasi kemampuan Random Forest dalam mengklasifikasikan tingkat stres berdasarkan data survei multidimensi. Penelitian terdahulu menunjukkan bahwa Random Forest umumnya mencapai akurasi di atas 85% pada dataset survei kesehatan mental dengan jumlah sampel yang lebih besar [6][9][10], sehingga akurasi model pada dataset yang lebih representatif diharapkan tetap kompetitif.

3.6 Analisis dan Pembahasan

Distribusi tingkat stres mahasiswa yang diteliti menunjukkan bahwa mayoritas (66,67%) berada pada kategori Sedang, yang mengindikasikan bahwa sebagian besar mahasiswa mengalami tekanan yang cukup signifikan namun masih dalam taraf yang dapat dikelola.

Kondisi ini sejalan dengan temuan Rahman et al. [8] yang menemukan bahwa pola tidur, penggunaan gadget, dan beban akademik merupakan prediktor utama stres pada mahasiswa berdasarkan data survei. Kategori Tinggi yang mencapai 16,67% menunjukkan adanya mahasiswa dengan kondisi stres yang perlu mendapat perhatian dan intervensi segera dari pihak institusi.

Analisis feature importance mengungkapkan bahwa tekanan akademik dalam bentuk deadline dan beban tugas merupakan faktor paling dominan penyebab stres, jauh lebih berpengaruh dibandingkan faktor gadget dan pola hidup. Temuan ini memiliki implikasi praktis yang penting: program intervensi kampus sebaiknya diprioritaskan pada manajemen beban akademik, seperti penjadwalan deadline yang lebih terencana, workshop manajemen waktu, dan sesi konseling akademik, daripada hanya berfokus pada aspek digitalisasi atau gaya hidup semata.

Performa model yang sempurna pada data testing perlu disikapi dengan hati-hati mengingat ukuran dataset yang kecil. Pada dataset yang lebih besar dan beragam, akurasi model mungkin akan berfluktuasi. Namun demikian, konsistensi dengan penelitian terdahulu seperti Thandica et al. [6] (87,5%) dan Putri & Sari [9] (91%) memberikan keyakinan bahwa Random Forest merupakan pilihan algoritma yang tepat untuk klasifikasi stres mahasiswa berbasis data survei.

Dari sisi implikasi teknologi, sistem berbasis Random Forest dapat diintegrasikan ke dalam platform akademik untuk membantu mendeteksi mahasiswa yang berisiko mengalami stres tinggi secara otomatis dan memberikan rekomendasi penanganan yang personal. Implementasi survei berkala menggunakan instrumen yang telah tervalidasi dapat menjadi sistem monitoring stres mahasiswa yang berkelanjutan [6][10].

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai klasifikasi tingkat stres mahasiswa menggunakan algoritma Random Forest, dapat ditarik beberapa kesimpulan. Pertama, algoritma Random Forest berhasil mengklasifikasikan tingkat stres mahasiswa ke dalam tiga kategori (Rendah, Sedang, Tinggi) berdasarkan data aktivitas harian dengan akurasi sebesar 100% pada data testing yang terdiri dari 6 sampel, dengan precision, recall, dan F1-score sempurna pada seluruh kelas.

Kedua, faktor-faktor yang paling berpengaruh terhadap tingkat stres mahasiswa berdasarkan analisis feature importance adalah tekanan deadline tugas (0,142), beban tugas kuliah (0,138), kesulitan berkonsentrasi (0,121), kurangnya motivasi (0,108), dan kelelahan akademik (0,097). Ketiga, distribusi tingkat stres pada 30 responden menunjukkan mayoritas berada pada kategori Sedang (66,67%), mengindikasikan bahwa tekanan akademik merupakan tantangan umum yang dihadapi mahasiswa Politeknik Negeri Medan.

Keempat, algoritma Random Forest terbukti efektif untuk klasifikasi data survei multidimensi karena kemampuannya menangani berbagai jenis fitur, ketahanan terhadap noise dan overfitting, serta kemampuan mengidentifikasi fitur yang paling berpengaruh melalui mekanisme feature importance. Untuk penelitian selanjutnya, disarankan memperluas jumlah responden minimal 200–500 orang dari berbagai institusi, menggunakan instrumen pengukuran stres yang telah terstandarisasi seperti Perceived Stress Scale (PSS), membandingkan performa Random Forest dengan algoritma lain seperti XGBoost dan Deep Learning, serta mengembangkan aplikasi berbasis web atau mobile untuk deteksi stres mahasiswa secara real-time.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada Dosen Pengampu mata kuliah Praktik Grafik dan Visualisasi Komputer, Ibu Yunita Sari Siregar, S.Pd., M.Si., atas bimbingan dan arahnya selama proses penelitian. Ucapan terima kasih juga ditujukan kepada seluruh mahasiswa Politeknik Negeri Medan yang telah bersedia menjadi responden survei dan memberikan data yang valid demi kelancaran penelitian ini.

REFERENCES

- [1] I. Alfaro, T. A. Tutupoly, dan A. Suryanto, "Komparasi Algoritma C4.5, Naïve Bayes, dan Random Forest untuk Klasifikasi Data Kelulusan Mahasiswa Jakarta," *Jurnal Sistem Informasi dan Teknologi*, vol. 5, no. 2, pp. 45–53, 2018.
- [2] I. A. Marie, L. Hakim, D. Sugiarto, dan W. Septiani, "Perbandingan Performansi Teknik Klasifikasi Breakdown Mesin pada Proses Produksi Pembuatan Battery Mobil," *Jurnal Teknik Industri*, vol. 9, no. 1, pp. 12–20, 2019.
- [3] V. W. Siburian dan I. E. Mulyana, "Prediksi Harga Ponsel Menggunakan Metode Random Forest," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 3, no. 1, pp. 67–74, 2018.
- [4] N. K. Dewi, U. D. Syafitri, dan S. Y. Mulyadi, "Penerapan Metode Random Forest dalam Driver Analysis," *Jurnal Statistika dan Komputasi*, vol. 6, no. 2, pp. 88–96, 2018.
- [5] A. Charim, S. Basuki, dan D. R. Akbi, "Detect Malware in Portable Document Format Files (PDF) Using Support Vector Machine and Random Decision Forest," *Journal of Computer Science and Information Technology*, vol. 4, no. 2, pp. 102–109, 2018.
- [6] A. Thandica, B. Setiawan, dan C. Rahayu, "Klasifikasi Tingkat Stres Akademik Mahasiswa Menggunakan Algoritma Random Forest," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 9, no. 3, pp. 501–510, 2022.
- [7] S. Hartanto dan A. Saifudin, "Penerapan Machine Learning untuk Prediksi Tingkat Stres Mahasiswa," *Jurnal Informatika Universitas Pamulang*, vol. 6, no. 1, pp. 34–41, 2021.
- [8] A. Rahman, D. Kusuma, dan F. Pratiwi, "Pengaruh Aktivitas Harian Terhadap Kesehatan Mental Mahasiswa: Studi Berbasis Machine Learning," *Indonesian Journal of Health Science*, vol. 12, no. 2, pp. 45–58, 2022.
- [9] D. A. Putri dan R. N. Sari, "Analisis Faktor Penyebab Stres Mahasiswa Menggunakan Decision Tree dan Random Forest," *Jurnal Ilmu Komputer dan Informatika*, vol. 7, no. 1, pp. 25–34, 2023.
- [10] M. F. Yusuf, T. Hidayat, dan E. Nugroho, "Implementasi Random Forest untuk Klasifikasi Kesehatan Mental Mahasiswa Berbasis Data Survei," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 11, no. 4, pp. 312–321, 2023.
- [11] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.



- [12] J. Han, M. Kamber, dan J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Waltham, MA: Morgan Kaufmann, 2011.