

Implementasi Random Forest untuk Klasifikasi Tingkat stres Mahasiswa Berdasarkan Data Kuesioner DASS-21

Oleh:

Syariful Rifky Bachtiar,

Mochamad Alfian Rosid

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Juli, 2026



Pendahuluan

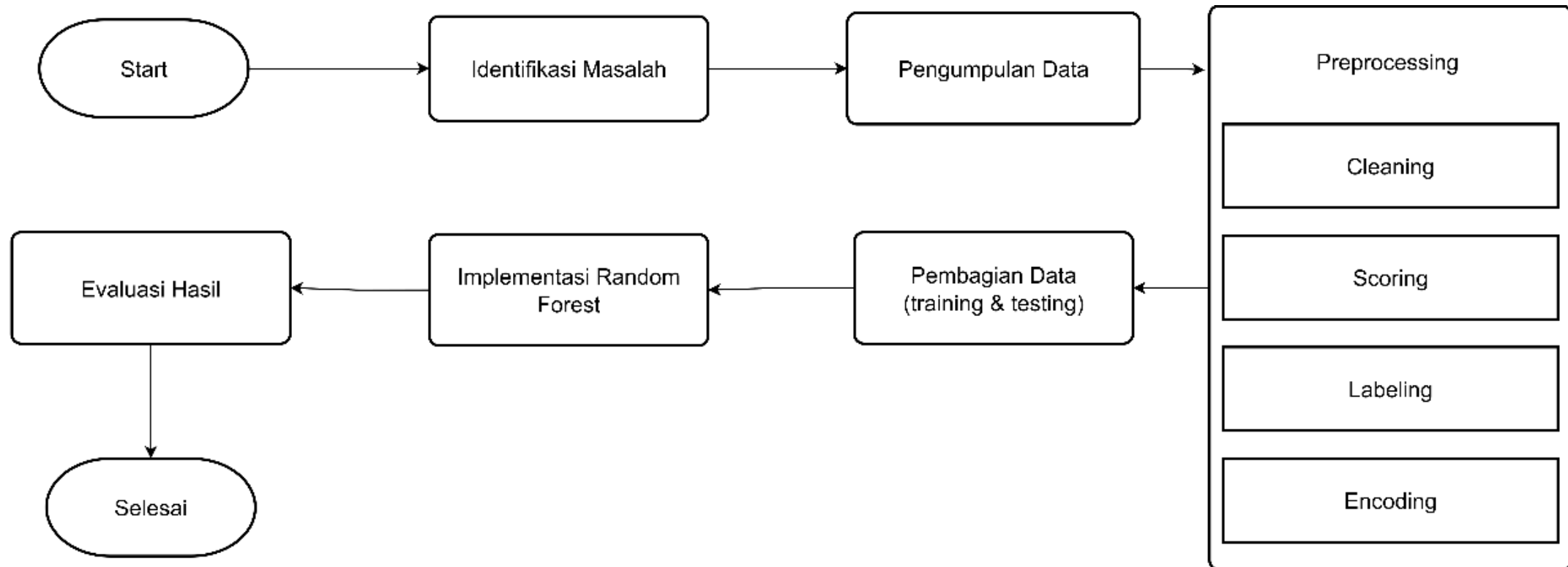
Stres merupakan masalah psikologis yang sering dialami mahasiswa. Penyebab meliputi tekanan akademik, tuntutan prestasi, dan faktor pribadi maupun sosial. Penilaian tingkat stres secara manual membutuhkan waktu dan berpotensi terjadi kesalahan. Machine Learning dapat dimanfaatkan untuk melakukan klasifikasi tingkat stres secara otomatis. Random Forest dipilih karena memiliki performa yang baik pada berbagai permasalahan klasifikasi.

Rumusan Masalah

Rumusan masalah yang diangkat dalam penelitian ini adalah:
“Bagaimana mengimplementasikan algoritma Random Forest untuk mengklasifikasikan tingkat stres mahasiswa berdasarkan data DASS-21?”

Metode

Tahapan Penelitian:



Metode

Pengumpulan Data:

- Menggunakan kuesioner DASS-21 (fokus: 7 item stres).
- Disebarkan via Google Form.
- Responden: 1000 mahasiswa aktif UMSIDA (Informed Consent).
- Pengisian dilakukan mandiri untuk kondisi 1 minggu terakhir.

No	Pernyataan
1	Saya merasa sulit beristirahat
2	Saya cenderung bereaksi berlebihan terhadap suatu situasi
3	Saya merasa menggunakan banyak energi karena gugup
4	Saya merasa mudah gelisah
5	Saya merasa sulit bersantai
6	Saya menemukan diri saya menjadi tidak sabar ketika mengalami penundaan
7	Saya merasa bahwa saya mudah tersinggung

Skala Respons DASS-21

No	Keterangan
0	Tidak Pernah
1	Kadang-Kadang
2	Cukup Sering
3	Sangat Sering

Metode

Pra-pemrosesan Data:

1. Cleaning : Tidak ditemukan *missing value* pada seluruh atribut (S1–S7). Seluruh data valid sehingga tidak dilakukan penghapusan data.
2. Scoring : Menjumlahkan skor dari 7 item subskala stres. Hasil penjumlahan dikalikan 2 sesuai pedoman DASS-21.
3. Labeling : Mengelompokkan skor menjadi 5 kategori:

Kategori	Rentang Skor
Normal	0-14
Ringan	15-18
Sedang	19-25
Berat	26-33
Sangat Berat	≥ 34

Metode

Pra-pemrosesan Data:

4. Encoding : Mengubah label menjadi bentuk numerik

Kategori	Label
Normal	0
Ringan	1
Sedang	2
Berat	3
Sangat Berat	4

Hasil preprocessing menghasilkan distribusi data pada lima kategori tingkat stres sebagai berikut.

Kategori	Jumlah Data
Normal	261
Ringan	231
Sedang	219
Berat	191
Sangat Berat	98

Pembagian Dataset
Training : **800 data (80%)**
Testing : **200 data (20%)**

Metode

Pembangunan Model:

Model dibangun menggunakan algoritma Random Forest. Hyperparameter dioptimasi menggunakan GridSearchCV.

Parameter	Nilai yang Diuji
N estimators	100, 200, 300
Max depth	None, 10, 20, 30
Min samples split	2, 5, 10
Min samples leaf	1, 2, 4

Metode

Pembangunan Model:

Hasil Hyperparameter Tuning menggunakan GridSearchCV

No	Rank	Mean Test Score	Parameter			
			n estimators	max depth	min samples split	min samples leaf
1	1	0,78000	200	10	2	1
2	2	0,77750	200	10	2	2
3	3	0,77625	200	10	5	1
4	4	0,77500	200	None	2	2
5	4	0,77500	200	30	2	2
...						
108	105	0,74375	300	30	10	4

Metode

Pembangunan Model:

Hasil Validasi Menggunakan 5-Fold Cross Validation

Fold	Accuracy (%)
1	76,88
2	75,00
3	80,00
4	73,75
5	75,62
Rata-rata ± Standar Deviasi	76,25 ± 2,13

Hasil

Performa model Random Forest dievaluasi menggunakan data uji dengan metrik accuracy, precision, recall, dan F1-score Hasilnya menunjukkan accuracy sebesar 73,50% dengan precision, recall, dan F1-score yang relatif seimbang.

Metrik	Nilai (%)
Accuracy	73,50
Precision	74,17
Recall	73,50
F1-Score	73,61

Hasil Analisis

Hasil Klasifikasi

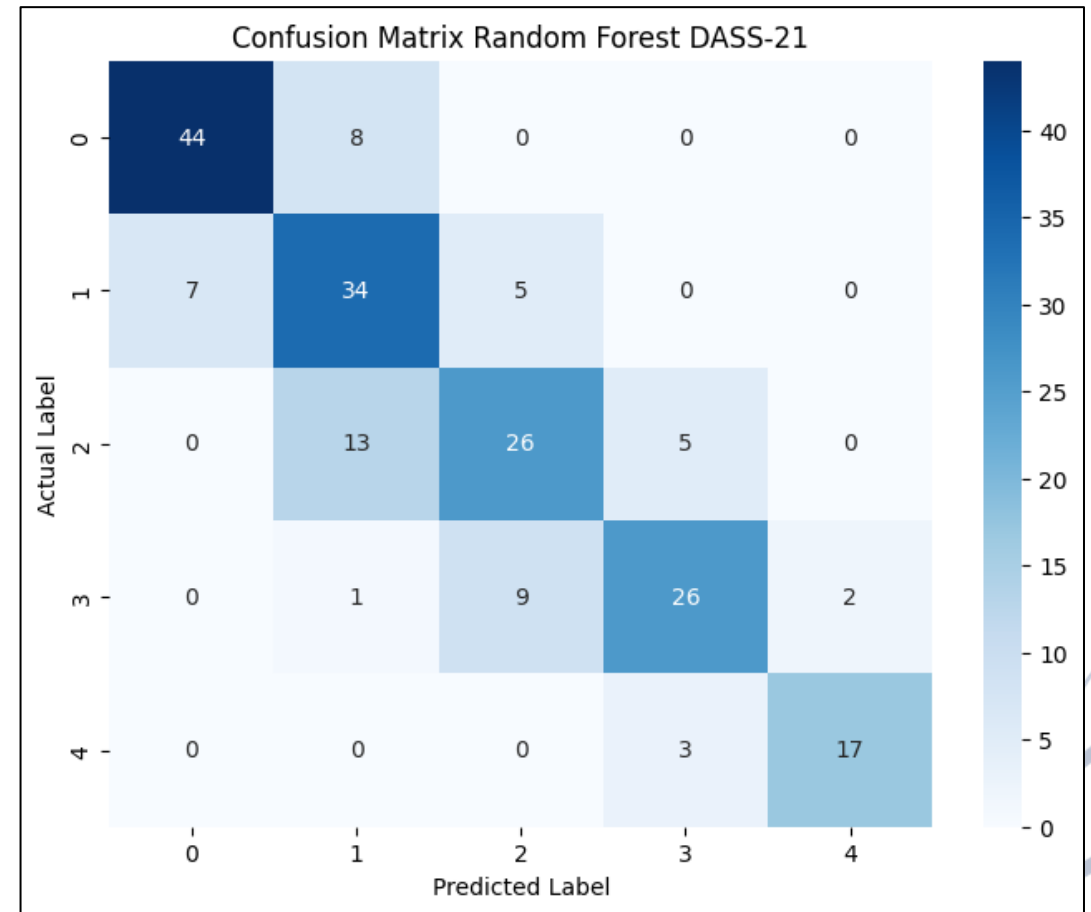
- Kategori Normal dan Sangat Berat memiliki performa terbaik.
- Kategori Sedang memiliki nilai F1-Score paling rendah.

	precision	recall	f1-score	support
Normal	0.86	0.85	0.85	52
Ringan	0.61	0.74	0.67	46
Sedang	0.65	0.59	0.62	44
Berat	0.76	0.68	0.72	38
Sangat Berat	0.89	0.85	0.87	20
accuracy			0.73	200
macro avg	0.76	0.74	0.75	200
weighted avg	0.74	0.73	0.74	200

Hasil Analisis

Confusion Matrix :

Kesalahan klasifikasi paling banyak terjadi pada kategori Ringan, Sedang, dan Berat. Kategori Normal dan Sangat Berat memiliki tingkat klasifikasi yang lebih baik.



Hasil Analisis

Feature Importance :

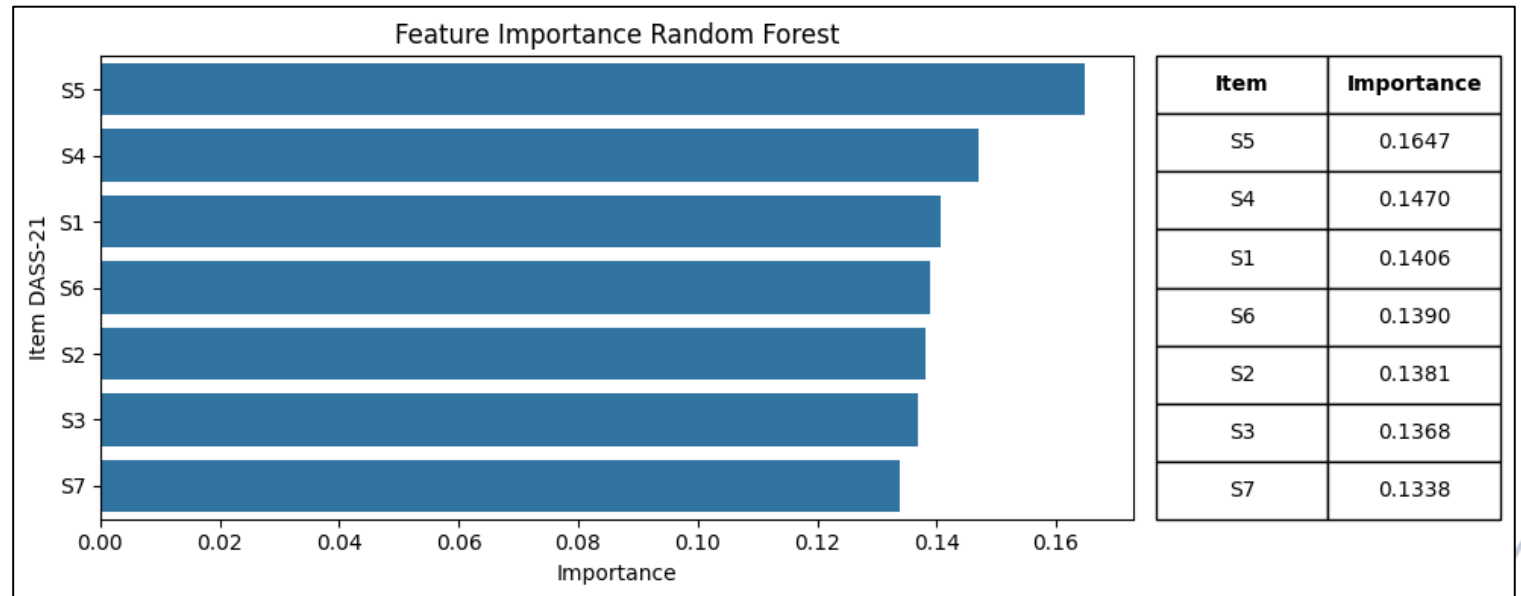
Hasil menunjukkan seluruh fitur DASS-21 memberikan kontribusi terhadap proses klasifikasi.

Tiga fitur paling berpengaruh

S5 – Saya merasa sulit bersantai

S4 – Saya merasa mudah gelisah

S1 – Saya merasa sulit beristirahat



Pembahasan

- Random Forest mampu mengklasifikasikan tingkat stres mahasiswa dengan accuracy 73,50% serta nilai precision, recall, dan F1-score yang relatif seimbang.
- Kategori Normal dan Sangat Berat memiliki performa klasifikasi yang lebih baik dibandingkan kategori lainnya.
- Kesalahan klasifikasi lebih banyak terjadi pada kategori Ringan, Sedang, dan Berat karena karakteristik ketiga kategori tersebut saling berdekatan.
- Analisis feature importance menunjukkan seluruh item DASS-21 berkontribusi dalam proses klasifikasi tanpa adanya satu fitur yang mendominasi secara signifikan.

Temuan Penting Penelitian

- Algoritma Random Forest berhasil mengklasifikasikan tingkat stres mahasiswa ke dalam lima kategori berdasarkan data kuesioner DASS-21 dengan accuracy sebesar 73,50%.
- Optimasi menggunakan GridSearchCV menghasilkan kombinasi hyperparameter terbaik, yaitu $n_estimators = 200$, $max_depth = 10$, $min_samples_split = 2$, dan $min_samples_leaf = 1$, sehingga menghasilkan performa validasi tertinggi.
- Kategori Normal dan Sangat Berat memiliki performa klasifikasi yang lebih baik, sedangkan kategori Sedang menjadi kategori yang paling sulit diklasifikasikan karena karakteristiknya berdekatan dengan kategori Ringan dan Berat.
- Analisis feature importance menunjukkan bahwa S5 (Sulit bersantai), S4 (Mudah gelisah), dan S1 (Sulit beristirahat) merupakan fitur yang memberikan kontribusi terbesar dalam proses klasifikasi.
- Seluruh item pada subskala stres DASS-21 memiliki nilai feature importance yang relatif berdekatan, sehingga menunjukkan bahwa semua fitur berkontribusi terhadap proses klasifikasi tanpa ada satu fitur yang mendominasi secara signifikan.

Manfaat Penelitian

Penelitian ini memberikan gambaran mengenai penerapan algoritma Random Forest dalam mengklasifikasikan tingkat stres mahasiswa berdasarkan data kuesioner DASS-21. Hasil penelitian diharapkan dapat menjadi referensi dalam pengembangan sistem klasifikasi tingkat stres serta mendukung penelitian selanjutnya melalui penggunaan dataset yang lebih besar, penanganan ketidakseimbangan data, maupun perbandingan dengan algoritma klasifikasi lainnya.

Referensi

- Thasya Nur Oktaviona, Herlina, and T. H. Sari, "Faktor Yang Mempengaruhi Tingkat Stres Pada Mahasiswa Akhir," *J. Keperawatan Trop. Papua*, vol. 6, no. 1, pp. 26–32, 2023, doi: 10.47539/jktp.v6i1.344.
- W. Musikhah and D. Nastiti, "Academic Stress of University of Muhammadiyah Sidoarjo Students Who Study While Working Class of 2021," *J. Islam. Muhammadiyah Stud.*, vol. 2, pp. 1–6, 2022, doi: <https://doi.org/10.21070/jims.v2i0.1543>.
- Q. Nada, I. Herdiana, and F. Andriani, "Testing the validity and reliability of the Depression Anxiety Stress Scale (DASS)-21 instrument for individuals with Psychodermatology," *Psikohumaniora*, vol. 7, no. 2, pp. 153–168, 2022, doi: 10.21580/pjpp.v7i2.11802.
- A. S. N. Isha et al., "Validation of 'Depression, Anxiety, and Stress Scales' and 'Changes in Psychological Distress during COVID-19' among University Students in Malaysia," *Sustain.*, vol. 15, no. 5, 2023, doi: 10.3390/su15054492.
- R. Rois, M. Ray, A. Rahman, and S. K. Roy, "Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms," *J. Heal. Popul. Nutr.*, vol. 40, no. 1, pp. 1–12, 2021, doi: 10.1186/s41043-021-00276-5.
- S. Arya, Anju, and N. A. Ramli, "Predicting the stress level of students using Supervised Machine Learning and Artificial Neural Network (ANN)," *Indian J. Eng.*, vol. 21, no. 56, 2024, doi: 10.54905/disssi.v21i55.e9ije1684.
- K. Rahayu, V. Fitria, D. Septhya, and ..., "Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning: Text Classification for Detecting Depression and ...," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 108–114, 2023, doi: <https://doi.org/10.57152/malcom.v3i2.780>.
- I. Adriansyah, M. D. Mahendra, E. Rasywir, and Y. Pratama, "Perbandingan Metode Random Forest Classifier dan SVM Pada Klasifikasi Kemampuan Level Beradaptasi Pembelajaran Jarak Jauh Siswa," *Bull. Informatics Data Sci.*, vol. 1, no. 2, pp. 98–103, 2022, doi: <http://dx.doi.org/10.61944/bids.v1i2.49>.
- D. Haganta Depari, Y. Widiastiwi, and M. Mega Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *J. Inform.*, vol. 18, no. 3, pp. 239–248, Dec. 2022, doi: <https://doi.org/10.52958/iftk.v18i3.4694>.
- A. Yusuf Permana, H. Noer Fazri, M. Fakhri Nur Athoilah, M. Robi, and R. Firmansyah, "Penerapan Data Mining Dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *J. Ilm. Tek. Inform. dan Komun.*, vol. 3, no. 2, pp. 27–41, Jun. 2023, doi: <https://doi.org/10.35970/infotekmesin.v14i1.1751>.
- D. Fuji Astari, Y. Herry Chrisnanto, and M. Melina, "KLASIFIKASI TINGKAT STRES SAAT TIDUR MENGGUNAKAN ALGORITMA RANDOM FOREST," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 5, pp. 3676–3684, 2024, doi: 10.36040/jati.v7i5.7750.
- M. A. Rosid, F. Muharram, and G. R. Affandi, "Perbandingan Kinerja Algoritma Machine Learning Untuk Mendeteksi Kalimat Sarkasme Dalam Bahasa Indonesia," *Procedia Soc. Sci. Humanit.*, vol. 3, no. c, pp. 1192–1195, 2022, doi: <https://doi.org/10.21070/pssh.v3i.253>.

Referensi

- M. L. Wallace *et al.*, "Use and misuse of random forest variable importance metrics in medicine: demonstrations through incident stroke prediction," *BMC Med. Res. Methodol.*, vol. 23, no. 1, pp. 1–12, 2023, doi: 10.1186/s12874-023-01965-x.
- L. Langsetmo *et al.*, "Advantages and Disadvantages of Random Forest Models for Prediction of Hip Fracture Risk Versus Mortality Risk in the Oldest Old," *JBMR Plus*, vol. 7, no. 8, pp. 1–11, 2023, doi: 10.1002/jbm4.10757.
- Y. Rothacher and C. Strobl, *Identifying Informative Predictor Variables With Random Forests*, vol. 49, no. 4. 2024. doi: 10.3102/10769986231193327.
- R. Ririn, A. Rizki Rinaldi, and F. M Basysyar, "Prediksi Kesehatan Mental Menggunakan Random Forest Berdasarkan Faktor Demografi Dan Lingkungan Kerja," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4515–4522, 2025, doi: 10.36040/jati.v9i3.13720.
- Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, and Susanti, "Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest," *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 122–131, Dec. 2022, doi: 10.33372/stn.v8i2.885.
- A. Hapsari, A. S. Nursuwanda, H. Zuhriyah, and D. J. Vresdian, "Klasifikasi Kesehatan Mental Mahasiswa Model TMAS dengan Algoritma Decision Tree, Logistic Regression, dan Random Forest," *INTEK J. Inform. dan Teknol. Inf.*, vol. 7, no. 2, pp. 55–64, 2024, doi: 10.37729/intek.v7i2.5690.
- L. Pefrianti, I. R. Munthe, I. Irmayanti, and M. Masrizal, "Klasifikasi Tingkat Stres Mahasiswa Dalam Penyelesaian Tugas Akhir Menggunakan Naïve Bayes Dan K-Nearest Neighbor," *J. Comput. Sci. Inf. Syst.*, vol. 7, no. 1, pp. 129–137, 2026, doi: <https://doi.org/10.36987/jcoins.v7i1.9060>.
- M. Chanafy and H. Sulistiani, "Comparative Analysis Of Machine Learning Algorithm Performance For Student Mental Health Classification," *SMATIKA STIKI Inform. J.*, vol. 16, no. 1, pp. 19–29, 2026, doi: <https://doi.org/10.32664/smatika.v16i01.2080>.
- I. M. M. Matin, "Hyperparameter Tuning Menggunakan GridsearchCV pada Random Forest untuk Deteksi Malware," *Multinetics*, vol. 9, no. 1, pp. 43–50, 2023, doi: 10.32722/multinetics.v9i1.5578.
- J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, 2022. [Online]. Available: <https://books.google.co.id/books?id=Rkh4EAAAQBAJ>
- M. U. Nadeem, S. J. Kulich, and I. H. Bokhari, "The assessment and validation of the depression, anxiety, and stress scale (DASS-21) among frontline doctors in Pakistan during fifth wave of COVID-19," *Front. Public Heal.*, vol. 11, 2023, doi: 10.3389/fpubh.2023.1192733.
- Y. A. Prasetyo, E. Utami, and A. Yaqin, "Pengaruh Komposisi Split Data Terhadap Performa Akurasi Analisis Sentimen Algoritma Naïve Bayes dan SVM," *J. Electr. Eng. Comput.*, vol. 6, no. 2, pp. 382–390, 2024, doi: 10.33650/jeecom.v6i2.9188.

