

Prediksi Tingkat Stres Mahasiswa Menggunakan Algoritma Decision Tree

Predicting Student Stress Levels Through Decision Tree Algorithm: A Machine Learning Approach

Azka Muhammad Mujaky, Berkat Telaumbanua, Diah Safitri*, Fakhri Aliy Ardiansyah, Muhammad Ibnu Hafiz

Teknologi Informasi, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

E-mail: *diaaahsaafii@gmail.com

Abstract. This study aims to apply the Decision Tree algorithm to predict stress levels among university students based on non-psychological variables. Academic stress is a common issue that affects students' performance and overall well-being, making early prediction important for timely intervention. The model uses features such as gender, age, CGPA, year of study, marital status, and help-seeking behavior to avoid data leakage and ensure that predictions rely only on indirect factors. Data processing and model development were conducted using Python and the scikit-learn library. The Decision Tree model achieved an accuracy of 55%, performing better in identifying students with low stress levels. However, the model still struggles to detect high-stress cases due to limited feature relevance. Overall, the approach provides an interpretable baseline model that can assist educational institutions in early identification and preventive support for students at risk of stress.

Keywords: Decision Tree, Student Stress, Machine Learning, Classification, Prediction

Abstrak. Penelitian ini bertujuan menerapkan algoritma *Decision Tree* untuk memprediksi tingkat stres pada mahasiswa menggunakan variabel non-psikologis. Stres akademik merupakan permasalahan umum yang dapat mempengaruhi performa dan kesejahteraan mahasiswa sehingga deteksi dini menjadi penting. Model menggunakan fitur jenis kelamin, usia, IPK, tahun studi, status pernikahan, dan perilaku pencarian bantuan guna menghindari kebocoran data dan memastikan prediksi berbasis faktor tidak langsung. Pengolahan data dilakukan menggunakan *Python* dan pustaka *scikit-learn*. Model *Decision Tree* memperoleh akurasi sebesar 55% dan menunjukkan kinerja lebih baik dalam mengidentifikasi mahasiswa dengan stres rendah. Namun, model masih kesulitan mengenali kasus stres tinggi karena keterbatasan relevansi fitur. Secara keseluruhan, pendekatan ini memberikan model dasar yang mudah dipahami dan dapat membantu institusi pendidikan dalam mengidentifikasi mahasiswa berisiko serta mendukung langkah pencegahan dan intervensi awal.

Kata kunci: Decision Tree, Stres Mahasiswa, Machine Learning, Klasifikasi, Prediksi

Submitted: 13-01-2026 | Accepted: 25-03-2026 | Published: 31-03-2026

How to Cite:

F. Firdaus, B. Telaumbanua, D. Safitri, F. A. Ardiansyah and M. I. Hafiz, "Prediksi Tingkat Stres Mahasiswa Menggunakan Algoritma Decision Tree" *Journal of Information System and Application Development (JISAD)*, vol. 4, no. 1, pp. 43-49, 2026, doi: 10.26905/jisad.v4i1.16616



PENDAHULUAN

Kesehatan mental mahasiswa menjadi perhatian penting dalam pendidikan tinggi karena tekanan akademik, tuntutan sosial, dan faktor lingkungan dapat memicu stres yang berdampak pada performa belajar serta kesejahteraan psikologis. Kondisi stres yang tidak ditangani berpotensi menurunkan konsentrasi, motivasi, hingga memicu perilaku berisiko yang merugikan mahasiswa serta institusi pendidikan. Seiring dengan perkembangan teknologi, pendekatan berbasis *machine learning* semakin banyak digunakan untuk menganalisis dan memprediksi kondisi mental mahasiswa secara lebih cepat, objektif, dan berbasis data [1], [2], [3], [4].

Selain itu, berbagai penelitian menunjukkan bahwa pandemi COVID-19 memberikan dampak signifikan terhadap kondisi stres mahasiswa akibat perubahan metode pembelajaran, isolasi sosial, dan keterbatasan akses layanan pendidikan [5], [6]. Selain faktor psikologis langsung, sejumlah studi modern juga memanfaatkan data perilaku seperti *passive sensing*, rekam jejak digital, dan pola mobilitas untuk mengidentifikasi stres mahasiswa secara *real-time* dan tingkat akurasi yang cukup tinggi [7], [8].

Berbagai penelitian juga menegaskan bahwa algoritma *Decision Tree* efektif digunakan untuk klasifikasi kesehatan mental karena kemampuannya memproses data kategorik dan numerik serta sifatnya yang mudah diinterpretasikan [1]. Teknik *oversampling* seperti *SMOTE* terbukti dapat meningkatkan performa model dengan menyeimbangkan distribusi kelas pada dataset [9]. Namun, penggunaan variabel psikologis sebagai fitur prediksi sering kali menimbulkan masalah data *leakage*, yaitu ketika fitur terlalu dekat dengan label sehingga membuat hasil prediksi tidak realistis dan sulit digeneralisasikan [10].

Penelitian lintas negara seperti di Rwanda dan Hongkong juga menunjukkan bahwa faktor akademik dan demografis memiliki kontribusi terhadap kondisi kesehatan mental mahasiswa [11], [12], [13]. Selain itu, tinjauan literatur terbaru menyatakan bahwa penerapan kecerdasan buatan dalam bidang kesehatan mental semakin berkembang dan memberikan hasil yang menjanjikan [14].

Berdasarkan berbagai temuan tersebut, penelitian ini berfokus pada pembangunan model prediksi stres mahasiswa menggunakan fitur non-psikologis untuk menghindari data *leakage*. Fitur yang digunakan meliputi *gender*, usia, IPK, tahun studi, status pernikahan, dan perilaku pencarian bantuan. Penelitian ini bertujuan untuk menerapkan algoritma *Decision Tree* dalam memprediksi tingkat stres mahasiswa, mengevaluasi performa model, serta menganalisis kontribusi setiap fitur terhadap proses klasifikasi.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi *machine learning* untuk memprediksi tingkat stres mahasiswa. Proses penelitian meliputi beberapa tahapan, yaitu pra pemrosesan data, pembentukan label, pelatihan model, evaluasi, dan interpretasi hasil. Pemilihan algoritma *Decision Tree* dilakukan karena sifatnya yang mudah diinterpretasikan dan kemampuan menangani data numerik maupun kategorikal [1].

Dataset yang digunakan berasal dari platform *Kaggle*, berisi informasi demografis, akademik, serta variabel psikologis mahasiswa. Variabel dependen pada penelitian ini adalah *Stress_Level*, sedangkan variabel independennya terdiri dari *Gender*, *Age*, *Year of Study*, *CGPA*, *Marital Status*, dan *Seek Help*. Penelitian serupa yang menggunakan survei akademik dan data demografis sebagai fitur prediksi juga dilakukan pada beberapa studi terkait kesehatan mental mahasiswa [12], [13].

Langkah pertama penelitian adalah pembersihan data untuk memastikan tidak ada nilai kosong dan format data konsisten. Label stres dibentuk berdasarkan variabel *Anxiety*, *Depression*, dan *Panic Attack*. Namun, ketiga fitur tersebut dihapus dari input model untuk mencegah data *leakage* sehingga model hanya belajar dari faktor tidak langsung [10]. Selanjutnya, data diubah ke bentuk numerik melalui teknik *one-hot encoding*, kemudian dibagi menjadi data latih dan data uji menggunakan *train-test split*. Teknik *SMOTE* diterapkan pada data latih untuk menyeimbangkan distribusi kelas [9].

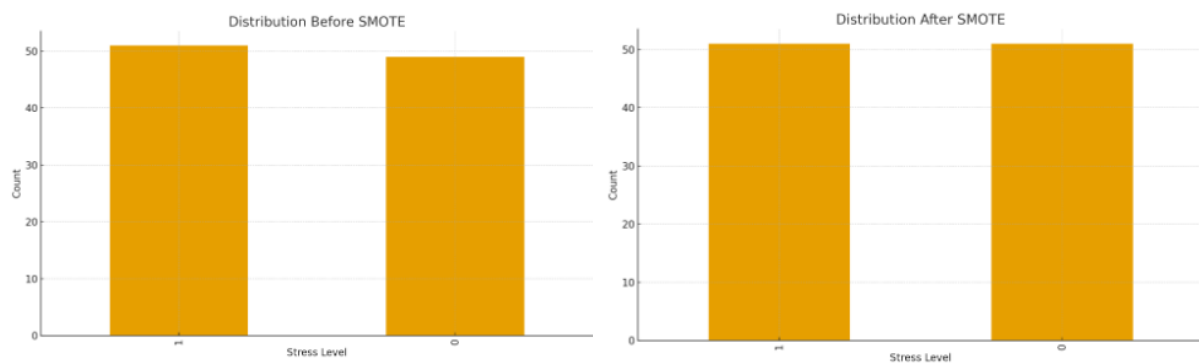
Pelatihan model menggunakan algoritma *Decision Tree Classifier* dari *scikit-learn* dengan parameter *max_depth* = 6 untuk mengurangi risiko *overfitting*. Pemilihan parameter ini didasarkan

pada praktik umum di penelitian *machine learning* terkait kesehatan mental [1], [2], [4]. Evaluasi model dilakukan menggunakan matrik *accuracy*, *precision*, *recall*, *f1-score*, serta *confusion matrix*. *Cross-validation* juga dilakukan untuk memastikan performa model tetap konsisten di berbagai skenario pembagian data.

HASIL DAN PEMBAHASAN

Pembentukan label Stress Level dibentuk berdasarkan kombinasi tiga variabel psikologis, yaitu *Anxiety*, *Depression*, dan *Panic Attack*. Kemudian dapat dikategorikan menjadi dua kelas: *Low Stress* (0) dan *High Stress* (1). Meskipun digunakan dalam pembentukan label, fitur-fitur tersebut dihapus dari *input* model guna mencegah data *leakage* sesuai rekomendasi literatur [10].

Sedangkan pendistribusi data sebelum dan sesudah *SMOTE* dilakukan dengan mencari ketidakseimbangan. Dataset awal menunjukkan ketidakseimbangan kelas antara *Low Stress* dan *High Stress*. Untuk menghindari bias model, teknik *SMOTE* diterapkan untuk menyeimbangkan distribusi kelas [1]. Penerapan *SMOTE* membuat model lebih mampu mengenali kelas minoritas, meskipun tidak selalu menjamin peningkatan akurasi keseluruhan. Gambar 1 menunjukkan perbandingan distribusi kelas sebelum dan sesudah penerapan *SMOTE*.



Gambar 1. Distribusi Kelas Sebelum dan Sesudah SMOTE

Penerapan *SMOTE* membantu meningkatkan kemampuan model dalam mengenali kelas minoritas, terutama pada prediksi *High Stress*. Namun demikian, *oversampling* tidak menjamin peningkatan performa keseluruhan apabila fitur yang digunakan memiliki korelasi lemah terhadap label. Selanjutnya mencari model *Decision Tree Classifier* yang dilatih menggunakan parameter *max_depth* = 6 untuk menghindari *overfitting*. Proses pelatihan dilakukan pada data yang telah melalui tahap pra pemrosesan dan *balancing*.

Tabel 1. Hasil Classification Report Model Decision Tree

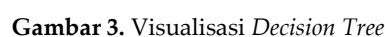
	Precision	Recall	F1-score	Support
Low Stress (0)	0.54	0.70	0.61	10
High Stress (1)	0.57	0.40	0.47	10
Accuracy			0.55	20
Macro avg	0.55	0.55	0.54	20
Weightd avg	0.55	0.55	0.54	20

Dari tabel 1 terlihat bahwa kelas *Low Stress* memiliki nilai *recall* lebih tinggi (0.70) dibanding kelas *High Stress* (0.40). Hal ini menunjukkan bahwa model lebih efektif dalam mendeteksi mahasiswa yang benar-benar berada pada kondisi stres rendah. Sebaliknya, kemampuan model untuk mengenali *High Stress* masih rendah, yang tercermin dari nilai *false negative* yang cukup tinggi.

Kontribusi performa per kelas dapat dilihat pada Gambar 2 yang menggambarkan analisis *Confusion Matrix*



Dalam menginterpretasi *Decision Tree* berupa struktur pohon keputusan pada Gambar 3 memperlihatkan bahwa variabel *CGPA* menjadi fitur paling dominan dalam proses pemisahan data.



Secara keseluruhan, hasil penelitian menunjukkan bahwa model *Decision Tree* mampu memberikan gambaran awal mengenai pola stres mahasiswa berdasarkan variabel non-psikologis. Namun, performa model yang hanya mencapai akurasi 55% mengindikasikan bahwa fitur yang

digunakan belum cukup kuat untuk membedakan dua kategori stres secara efektif. Jumlah kesalahan prediksi pada kelas *High Stress* mengonfirmasi bahwa model masih kesulitan mendeteksi mahasiswa yang sebenarnya berada pada kondisi berisiko tinggi.

Penelitian ini memberikan kontribusi penting berupa model prediksi bebas data *leakage* yang dapat dijadikan *baseline* dalam pengembangan sistem deteksi dini stres mahasiswa. Hasil penelitian menjawab permasalahan awal bahwa prediksi stres dapat dilakukan menggunakan variabel non-psikologis, namun efektivitasnya masih terbatas. Hal ini konsisten dengan penelitian lain yang menggunakan perilaku digital, *passive sensing*, dan data akademik sebagai fitur prediktif [7], [8], [11]. Selain itu, berbagai studi menegaskan bahwa metode kompleks seperti *SVM*, *Random Forest*, dan *XGBoost* sering memberikan hasil yang lebih stabil dalam domain kesehatan mental [15]. Penelitian ini memperkuat temuan bahwa model sederhana seperti *Decision Tree* dapat menjadi langkah awal sebelum penggunaan metode yang lebih canggih.

Rumus Matematika

Evaluasi performa model dilakukan menggunakan beberapa metrik umum pada *machine learning*, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Metrik ini dihitung berdasarkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Selain itu, algoritma *Decision Tree* menggunakan ukuran impurity seperti *Gini Index* untuk menentukan titik pemisahan terbaik. Rumus-rumus digunakan sebagai dasar perhitungan evaluasi, sebagaimana dituliskan pada Persamaan (1) hingga Persamaan (5).

1. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Keterangan:

TP = data positif yang berhasil diprediksi benar

TN = data negatif yang berhasil diprediksi benar

FP = data negatif yang salah diprediksi sebagai positif

FN = data positif yang salah diprediksi sebagai *negative*

2. Precision

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

Keterangan:

Precision menunjukkan seberapa akurat prediksi kelas positif dibandingkan total prediksi positif.

3. Recall

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

Keterangan:

Recall mengukur kemampuan model dalam menemukan seluruh data positif sebenarnya.

4. F1-Score

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

Keterangan:

F1-score adalah rata-rata harmonik antara precision dan recall, digunakan ketika dataset tidak seimbang.

5. Gini Impurity (Digunakan pada *Decision Tree*)

$$Gini(t) = 1 - \sum_{i=1}^c p_i^2 \quad (5)$$

Keterangan:

C = jumlah kelas

p_i = probabilitas data pada kelas ke- i di node t

SIMPULAN DAN SARAN

Penelitian ini berhasil menerapkan algoritma *Decision Tree* untuk memprediksi tingkat stres mahasiswa menggunakan variabel non-psikologis guna menghindari *data leakage*. Variabel yang digunakan meliputi jenis kelamin, usia, IPK, tahun studi, status pernikahan, serta perilaku pencarian bantuan. Pembentukan label *Stress_Level* dilakukan berdasarkan kondisi *Anxiety*, *Depression*, dan *Panic Attack*, namun variabel tersebut dihilangkan dari proses pemodelan sehingga model mempelajari pola yang lebih objektif.

Hasil evaluasi menunjukkan bahwa model menghasilkan akurasi sebesar 55% pada data uji. Model lebih mampu mengenali kelas *Low Stress* dengan nilai *recall* yang lebih tinggi, sedangkan performa pada kelas *High Stress* masih rendah karena jumlah *false negative* yang cukup besar. Kesulitan model dalam mendeteksi stres tinggi mengindikasikan bahwa variabel non-psikologis memiliki pengaruh terbatas dalam menentukan tingkat stres mahasiswa. Meski demikian, struktur pohon keputusan menunjukkan kontribusi penting dari IPK dan tahun studi sebagai prediktor utama, sedangkan variabel demografis memberikan pengaruh yang lebih kecil. Secara keseluruhan, model ini dapat berfungsi sebagai *baseline model* yang interpretatif untuk deteksi dini stres mahasiswa dan mendukung institusi pendidikan dalam pemantauan awal kesehatan mental.

Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan menambahkan fitur relevan seperti perilaku digital, jam tidur, beban akademik, dan interaksi sosial. Selain itu, algoritma lain seperti *Random Forest*, *XGBoost*, dan *SVM* perlu dieksplorasi karena sering memberikan hasil yang lebih stabil [6], [11]. Integrasi data *passive sensing* juga terbukti meningkatkan kemampuan prediksi dalam penelitian lain [7], [8] dan dapat memberikan peningkatan performa dibandingkan model *Decision Tree* tunggal.

Dari sisi evaluasi, penggunaan teknik validasi yang lebih komprehensif serta perbandingan metrik tambahan seperti ROC-AUC dapat memberikan gambaran lebih tepat mengenai kualitas model. Pada tahap implementasi, model dapat dikembangkan menjadi sistem pemantauan stres berbasis *dashboard* atau *decision support system* yang dapat digunakan oleh konselor kampus untuk mengidentifikasi mahasiswa berisiko secara lebih cepat dan tepat. Dengan pengembangan lanjutan, model ini diharapkan mampu memberikan kontribusi lebih besar bagi upaya pencegahan dan intervensi kesehatan mental mahasiswa.

UCAPAN TERIMA KASIH

Tugas akhir ini disusun sebagai bentuk penerapan ilmu yang telah diperoleh selama perkuliahan, khususnya dalam memahami konsep pembelajaran mesin (*machine learning*) dan penerapannya pada kasus nyata di bidang sosial. Penelitian ini bertujuan untuk mengimplementasikan algoritma *Decision Tree* dalam menganalisis hubungan antara faktor sosial, psikologis, dan lingkungan terhadap Tingkat stress dan perilaku berisiko di kalangan mahasiswa. Melalui pendekatan ini, diharapkan dapat diperoleh pemahaman yang lebih mendalam tentang pola perilaku dan kondisi mental serta potensi penerapan teknologi kecerdasan buatan dalam bidang sosial dan kesehatan..

Dalam penyusunan makalah ini, penulis mengucapkan terima kasih kepada Ibu **Giatika Chrisnawati, ST, M.Kom** selaku dosen pengampu yang telah memberikan bimbingan serta semua pihak yang telah membantu dalam penyusunan makalah ini. Penulis menyadari bahwa jurnal ini dibuat masih jauh dari sempurna, oleh karena itu saran dan kritik yang membangun sangat diharapkan demi penyempurnaan karya ini. Semoga makalah ini dapat memberikan manfaat dan menambah wawasan bagi pembaca. Penulis berharap apa yang dibuat ini akan memberikan manfaat bagi penulis sendiri, rekan-rekan mahasiswa dan khalayak pembaca lain yang tertarik pada penerapan Machine

Learning dalam bidang sosial, khususnya dalam upaya memahami tentang kondisi stress di kalangan mahasiswa maupun lainnya.

DAFTAR PUSTAKA

- [1] T. Jain, A. Jain, P. S. Hada, H. Kumar, V. K. Verma, and A. Patni, "Machine Learning Techniques for Prediction of Mental Health," *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021*, pp. 1606–1613, Sep. 2021, doi: 10.1109/ICIRCA51532.2021.9545061.
- [2] M. Drira, S. Ben Hassine, M. Zhang, and S. Smith, "Machine Learning Methods in Student Mental Health Research: An Ethics-Centered Systematic Literature Review," *Applied Sciences* 2024, Vol. 14, Page 11738, vol. 14, no. 24, p. 11738, Dec. 2024, doi: 10.3390/APP142411738.
- [3] U. Madububambachu, A. Ukpebor, and U. Ihezue, "Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review," *Clin Pract Epidemiol Ment Health*, vol. 20, no. 1, p. e17450179315688, Jul. 2024, doi: 10.2174/0117450179315688240607052117.
- [4] B. L. Schaab *et al.*, "How do machine learning models perform in the detection of depression, anxiety, and stress among undergraduate students? A systematic review," *Cad Saude Publica*, vol. 40, no. 11, p. e00029323, Dec. 2024, doi: 10.1590/0102-311XEN029323.
- [5] M. H. E. M. Browning *et al.*, "Psychological impacts from COVID-19 among university students: Risk factors across seven states in the United States," *PLoS One*, vol. 16, no. 1 January, Jan. 2021, doi: 10.1371/JOURNAL.PONE.0245327.
- [6] D. T. L. Shek, W. Chai, X. Li, and D. Dou, "Profiles and predictors of mental health of university students in Hong Kong under the COVID-19 pandemic," *Front Psychol*, vol. 14, p. 1211229, Jul. 2023, doi: 10.3389/FPSYG.2023.1211229/FULL.
- [7] A. Shvetcov *et al.*, "Passive sensing data predicts stress in university students: a supervised machine learning method for digital phenotyping," *Front Psychiatry*, vol. 15, p. 1422027, Aug. 2024, doi: 10.3389/FPSYT.2024.1422027/BIBTEX.
- [8] M. I. Siraji *et al.*, "Impact of mobile connectivity on students' wellbeing: Detecting learners' depression using machine learning algorithms," *PLoS One*, vol. 18, no. 11, p. e0294803, Nov. 2023, doi: 10.1371/JOURNAL.PONE.0294803.
- [9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/JAIR.953.
- [10] A. Apicella, F. Isgro, and R. Prevete, "Don't push the button! Exploring data leakage risks in machine learning and transfer learning," *Artificial Intelligence Review* 2025 58:11, vol. 58, no. 11, pp. 339–, Aug. 2025, doi: 10.1007/S10462-025-11326-3.
- [11] F. Ndikumana *et al.*, "Machine learning-based predictive modelling of mental health in Rwandan Youth," *Scientific Reports* 2025 15:1, vol. 15, no. 1, pp. 16032–, May 2025, doi: 10.1038/s41598-025-00519-z.
- [12] A. Baba and K. Bunji, "Prediction of Mental Health Problem Using Annual Student Health Survey: Machine Learning Approach," *JMIR Ment Health*, vol. 10, no. 1, p. e42420, May 2023, doi: 10.2196/42420.
- [13] P. Córdova Olivera, P. Gasser Gordillo, H. Naranjo Mejía, I. La Fuente Taborga, A. Grajeda Chacón, and A. Sanjinés Unzueta, "Academic stress as a predictor of mental health in university students," *Cogent Education*, vol. 10, no. 2, Dec. 2023, doi: 10.1080/2331186X.2023.2232686;WGROU:STRING:PUBLICATION.
- [14] A. Wajid, F. Azam, and M. W. Anwar, "Applications of artificial intelligence in mental health: a systematic literature review," *Discover Artificial Intelligence* 2025 5:1, vol. 5, no. 1, pp. 332–, Nov. 2025, doi: 10.1007/S44163-025-00569-2.
- [15] K. T. Lisnyj, R. Russell, and A. Papadopoulos, "Risk and protective factors for anxiety impacting academic performance in post-secondary students," *Canadian Journal of Higher Education*, vol. 50, no. 2, pp. 71–88, 2020, doi: 10.47678/CJHE.V50I2.188715.