

Explainable AI untuk Deteksi Tingkat Depresi Mahasiswa Berbasis Faktor Akademik dan Sosial: Pendekatan SHAP

Faisal¹, Megawati Busrah²

Teknik Informatika, Fakultas Teknik, Universitas Pejuang Republik Indonesia¹

Komputerisasi akuntansi, Amik Luwuk Banggai²

Faisal.ces@gmail.com¹, egha.palopo@gmail.com²

Kata Kunci:

Explainable AI, SHAP, deteksi depresi, mahasiswa, XGBoost, learning analytics kesehatan mental

Abstrak

Depresi merupakan salah satu gangguan kesehatan mental yang paling banyak dialami mahasiswa dan berdampak signifikan terhadap performa akademik serta kesejahteraan psikologis. Model machine learning telah banyak digunakan untuk mendeteksi risiko depresi secara dini, namun sifatnya yang cenderung black-box menyulitkan pemanfaatannya sebagai dasar intervensi oleh konselor maupun pengambil kebijakan pendidikan. Penelitian ini bertujuan mengembangkan model deteksi tingkat depresi mahasiswa berbasis faktor akademik dan sosial yang interpretatif menggunakan pendekatan Explainable AI (XAI), khususnya SHapley Additive exPlanations (SHAP). Dataset yang digunakan terdiri atas 502 responden mahasiswa dengan 10 fitur meliputi tekanan akademik, kepuasan belajar, durasi tidur, pola makan, riwayat pikiran untuk bunuh diri, jam belajar, tekanan finansial, riwayat keluarga dengan gangguan mental, usia, dan gender. Tiga model klasifikasi dibandingkan, yaitu Logistic Regression, Random Forest, dan XGBoost. Logistic Regression mencatat akurasi tertinggi sebesar 98,02%, diikuti XGBoost sebesar 94,06% dan Random Forest sebesar 93,07%. Interpretasi SHAP pada model XGBoost menunjukkan bahwa riwayat pikiran untuk bunuh diri, tekanan akademik, usia, kepuasan belajar, dan tekanan finansial merupakan lima faktor paling berpengaruh terhadap prediksi tingkat depresi. Temuan ini memberikan dasar empiris bagi institusi pendidikan tinggi untuk merancang sistem deteksi dini dan kebijakan dukungan kesehatan mental mahasiswa yang lebih tepat sasaran dan dapat dipertanggungjawabkan secara transparan.

Abstract

Depression is one of the most prevalent mental health disorders among university students and significantly affects academic performance and psychological well-being. Machine learning models have been widely applied for early depression risk detection, yet their black-box nature limits their use as a basis for intervention by counselors and education policymakers. This study develops an interpretable model for detecting student depression levels based on academic and social factors using an Explainable AI (XAI) approach, specifically SHapley Additive exPlanations (SHAP). The dataset consists of 502 student respondents with 10 features, including academic pressure, study satisfaction, sleep duration, dietary habits, suicidal thoughts, study hours, financial stress, family history of mental illness, age, and gender. Three classification models were compared: Logistic Regression, Random Forest, and XGBoost. Logistic Regression achieved the highest accuracy at 98.02%, followed by XGBoost at 94.06% and Random Forest at 93.07%. SHAP interpretation on the XGBoost model revealed that suicidal thoughts, academic pressure, age, study satisfaction, and financial stress were the five most influential predictors of depression level. These

Key Word:

Explainable AI, SHAP, depression detection, university students, XGBoost, mental health learning analytics

findings provide an empirical basis for higher education institutions to design early-detection systems and targeted, transparently accountable student mental health support policies.

Copyright © xxxx

This work is licensed under an Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

PENDAHULUAN

Kesehatan mental mahasiswa merupakan isu global yang terus meningkat urgensinya. Kajian sistematis terbaru melaporkan bahwa prevalensi depresi di kalangan mahasiswa di berbagai negara tetap tinggi dan bervariasi menurut wilayah, metode pengukuran, serta konteks sosial-budaya (Jacob & Maguvhe, 2026). Tekanan akademik, beban finansial, dan kurangnya dukungan sosial merupakan faktor yang berulang kali diidentifikasi sebagai kontributor utama gangguan depresi pada populasi mahasiswa di berbagai belahan dunia, termasuk pada mahasiswa di negara berkembang (Sen et al., 2025).

Perkembangan teknologi machine learning membuka peluang untuk mendeteksi risiko depresi secara dini dan pada skala besar. Zhai et al. (2025) mengembangkan model prediktif berbasis XGBoost, Random Forest, Decision Tree, dan Logistic Regression untuk mengidentifikasi mahasiswa berisiko tinggi mengalami gangguan kecemasan dan depresi menggunakan data lebih dari enam puluh ribu mahasiswa di Amerika Serikat, dengan tujuan mendukung alokasi intervensi konseling secara lebih efisien. Studi serupa pada populasi mahasiswa di Tiongkok juga menunjukkan bahwa model machine learning mampu mengidentifikasi faktor sosiodemografis, psikologis, dan sosial yang berkontribusi terhadap gejala depresi dengan akurasi yang baik (Yu et al., 2025).

Meskipun demikian, sebagian besar model machine learning bersifat black-box, yaitu mampu menghasilkan prediksi akurat namun tidak memberikan penjelasan yang dapat dipahami mengenai faktor apa yang mendasari prediksi tersebut. Keterbatasan ini menjadi hambatan serius ketika model hendak digunakan dalam konteks kesehatan mental, di mana transparansi dan akuntabilitas keputusan sangat penting bagi penerimaan klinis maupun kebijakan (Ahmed et al., 2025). Explainable AI (XAI), khususnya pendekatan SHapley Additive exPlanations (SHAP), hadir untuk menjembatani kesenjangan antara akurasi model dan interpretabilitas dengan mengukur kontribusi setiap fitur terhadap prediksi individual maupun global (Lundberg & Lee, 2017).

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk: (1) membangun model klasifikasi tingkat depresi mahasiswa berdasarkan faktor akademik dan sosial; (2) membandingkan performa beberapa algoritma machine learning; dan (3) menginterpretasikan kontribusi masing-masing faktor terhadap prediksi model menggunakan SHAP, sehingga hasilnya dapat menjadi dasar rekomendasi kebijakan bagi bagian kemahasiswaan atau unit bimbingan konseling di perguruan tinggi.

METODE

Dataset

Penelitian ini menggunakan dataset publik berisi data 502 mahasiswa dengan 11 atribut, terdiri atas 10 fitur prediktor dan 1 variabel target (Depression). Fitur prediktor meliputi Gender, Age, Academic Pressure (skala 1-5), Study Satisfaction (skala 1-5), Sleep Duration (kategorikal), Dietary Habits (kategorikal), Have You Ever Had Suicidal Thoughts? (biner), Study Hours, Financial Stress (skala 1-5), dan Family History of Mental Illness (biner). Variabel target berupa klasifikasi biner Depression (Yes/No) dengan distribusi kelas yang relatif seimbang, yaitu 252 responden (50,2%) berlabel Yes dan 250 responden (49,8%) berlabel No. Dataset tidak memiliki nilai yang hilang (missing value) pada seluruh atributnya.

Preprocessing Data

Encoding variabel kategorikal (Gender, Sleep Duration, Dietary Habits, Suicidal Thoughts, Family History of Mental Illness) menggunakan Label Encoding, Pembagian data latih dan data uji dengan rasio 80:20 menggunakan stratified split untuk menjaga proporsi kelas target, Tidak dilakukan penanganan class imbalance karena distribusi target sudah relatif seimbang.

Model Klasifikasi dan Skema Evaluasi

Tiga algoritma klasifikasi dibandingkan dalam penelitian ini, yaitu Logistic Regression sebagai baseline model linear, Random Forest sebagai model ensemble berbasis bagging, dan XGBoost sebagai model ensemble berbasis gradient boosting. Ketiga model dilatih menggunakan data latih (401 sampel) dan dievaluasi pada data uji (101 sampel) yang tidak dilibatkan dalam proses pelatihan. Evaluasi performa menggunakan lima metrik, yaitu Accuracy, Precision, Recall, F1-Score, dan Area Under the ROC Curve (AUC-ROC).

Pendekatan Explainable AI (SHAP)

Interpretasi model dilakukan menggunakan SHAP TreeExplainer yang diterapkan pada model XGBoost. Pemilihan XGBoost sebagai model dasar interpretasi didasarkan pada kompatibilitasnya dengan TreeExplainer yang efisien secara komputasi untuk model berbasis pohon, serta performanya yang kompetitif dibandingkan Random Forest. Analisis SHAP menghasilkan nilai kontribusi (SHAP value) setiap fitur terhadap prediksi individual, yang kemudian diagregasi menjadi mean absolute SHAP value untuk memperoleh peringkat kepentingan fitur secara global (Lundberg & Lee, 2017).

HASIL DAN PEMBAHASAN

Perbandingan Performa Model

Tabel 1 menyajikan hasil evaluasi ketiga model pada data uji. Logistic Regression mencatat performa tertinggi pada seluruh metrik dengan akurasi 98,02% dan AUC-ROC 99,80%, diikuti oleh XGBoost dengan akurasi 94,06% dan AUC-ROC 99,41%, serta Random Forest dengan akurasi 93,07% dan AUC-ROC 98,45%. Tingginya performa Logistic Regression mengindikasikan bahwa hubungan antara fitur prediktor dan target pada dataset ini cenderung bersifat linear atau dapat dipisahkan secara linear, sehingga model yang lebih sederhana mampu menangkap pola tersebut secara optimal.

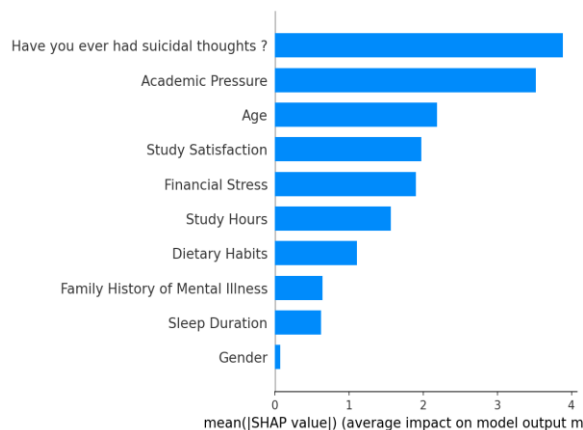
Tabel 1. Perbandingan Performa Model Klasifikasi

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	98.02%	98.04%	98.04%	98.04%	99.80%
Random Forest	93.07%	89.29%	98.04%	93.46%	98.45%
XGBoost	94.06%	90.91%	98.04%	94.34%	99.41%

Meskipun Logistic Regression unggul dalam metrik performa, model ini kurang ideal untuk menangkap interaksi non-linear antar fitur yang seringkali relevan dalam fenomena psikologis. Oleh karena itu, tahap interpretasi SHAP pada penelitian ini diterapkan pada model XGBoost, yang menawarkan keseimbangan antara akurasi kompetitif (94,06%) dan kemampuan menangkap hubungan non-linear serta interaksi antar fitur, sekaligus kompatibel secara native dengan SHAP TreeExplainer.

Interpretasi SHAP

Gambar 1 menyajikan peringkat kepentingan fitur secara global berdasarkan nilai rata-rata absolut SHAP (mean |SHAP value|) pada model XGBoost. Lima fitur dengan kontribusi tertinggi terhadap prediksi tingkat depresi adalah riwayat pikiran untuk bunuh diri (mean |SHAP| = 3,88), tekanan akademik (3,52), usia (2,19), kepuasan belajar (1,98), dan tekanan finansial (1,90). Sebaliknya, gender (0,07), durasi tidur (0,62), dan riwayat keluarga dengan gangguan mental (0,64) menunjukkan kontribusi yang relatif kecil terhadap prediksi model.



Gambar 1. Peringkat Kepentingan Fitur Global Berdasarkan Mean |SHAP Value| pada Model XGBoost

Dominannya riwayat pikiran untuk bunuh diri sebagai prediktor utama sejalan dengan temuan Ahmed et al. (2025) yang juga mengidentifikasi indikator risiko akut sebagai faktor pembeda kuat antara kelompok depresi dan non-depresi. Kontribusi tinggi tekanan akademik memperkuat temuan Sen et al. (2025) dan Yu et al. (2025) bahwa beban akademik merupakan salah satu determinan konsisten depresi pada populasi mahasiswa lintas negara. Sementara itu, kontribusi tekanan finansial yang cukup besar menegaskan peran faktor sosial-ekonomi sebagaimana ditekankan oleh Zhai et al. (2025) dalam konteks mahasiswa di Amerika Serikat, menunjukkan bahwa pola ini relatif konsisten lintas konteks geografis dan budaya.

KESIMPULAN

Penelitian ini mengembangkan model klasifikasi tingkat depresi mahasiswa berbasis faktor akademik dan sosial dengan pendekatan Explainable AI. Logistic Regression mencatat performa tertinggi (akurasi 98,02%), sementara XGBoost dipilih untuk tahap interpretasi karena kompatibilitasnya dengan SHAP TreeExplainer dan performa yang kompetitif (akurasi 94,06%). Interpretasi SHAP mengidentifikasi riwayat pikiran untuk bunuh diri, tekanan akademik, usia, kepuasan belajar, dan tekanan finansial sebagai lima faktor paling berpengaruh terhadap prediksi tingkat depresi mahasiswa. Temuan ini memberikan dasar empiris yang transparan bagi institusi pendidikan tinggi dalam merancang sistem deteksi dini dan kebijakan dukungan kesehatan mental mahasiswa.

Penelitian selanjutnya disarankan untuk memperluas ukuran sampel dan cakupan populasi, melibatkan instrumen psikometrik yang tervalidasi secara klinis, serta mengeksplorasi desain longitudinal untuk mengungkap hubungan kausal antar faktor risiko. Integrasi model ke dalam dashboard institusi juga dapat dieksplorasi sebagai tindak lanjut praktis dari penelitian ini.

DAFTAR REFERENSI

- Ahmed, I., Brahmacharimayum, A., Ali, R. H., Khan, T. A., & Ahmad, M. O. (2025). Explainable AI for depression detection and severity classification from activity data: Development and evaluation study of an interpretable framework. *JMIR Mental Health*, 12, e72038. <https://doi.org/10.2196/72038>
- Jacob, U. S., & Maguvhe, M. O. (2026). Global prevalence of depression among university students: A systematic literature review. *Frontiers in Education*, 11, Article 1809441. <https://doi.org/10.3389/educ.2026.1809441>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774.
- Nie, H., & Wu, X. Y. (2024). Research on influencing factors of depression combined with gradient boosting tree algorithm and explainable machine learning model SHAP. *Data Analysis and Knowledge Discovery*, 8(3), 41-52.
- Sen, S. K., Apurba, M. S. A., Mrityika, A. P., Anwar, M. T., Al Islam, A. B. M. A., & Noor, J. (2025). Unveiling shadows: A data-driven insight on depression among Bangladeshi university students. *Heliyon*, 11(1), e41110. <https://doi.org/10.1016/j.heliyon.2024.e41110>
- Yu, C., Kong, X., Yu, W., Ni, X., Chen, J., & Liao, X. (2025). Machine learning models for predicting the risk of depressive symptoms in Chinese college students. *Frontiers in Psychiatry*, 16, Article 1648585. <https://doi.org/10.3389/fpsy.2025.1648585>
- Zhai, Y., Zhang, Y., Chu, Z., Geng, B., Almaawali, M., Fulmer, R., Lin, Y.-W. D., Xu, Z., Daniels, A. D., Liu, Y., Chen, Q., & Du, X. (2025). Machine learning predictive models to guide prevention and intervention



**Journal Teknik Dan
Teknologi(JTT)**
Fakultas Teknik UPRI
<https://jurnalfatekupri.web.id/>

e-ISSN: 3110-6005

allocation for anxiety and depressive disorders among college students. Journal of Counseling & Development, 103(1), 110-125. <https://doi.org/10.1002/jcad.12543>**Sumber Internet**
U.S. Census Bureau. (2000). State and Country QuickFacts. Washington, D.C.: United States Bureau of the Census. Retrieved November 7, 2008, from <http://quickfacts.census.gov/qfd/>.