Jurnal Riset Sistem dan Teknologi Informasi (RESTIA)
Vol. 4 No. 1 | Februari 2026
Hal 12-21

Prediksi Depresi Mahasiswa Menggunakan Algoritma *Random Forest* Berbasis Data Psikososial

Depression Prediction Among University Students Using a Random Forest Algorithm Based on Psychosocial Data

Abiyya Alfahrizi Putra Arifiansyah[1*], Muhammad Afandi[2], Dodi Dwi Riskianto[3], Sudriyanto[3]
[1,2,3] *Informatika, Teknik, Universitas Nurul Jadid, Probolinggo, Indonesia*
E-mail: abyalfahri18@gmail.com; muhammadafandi0112@gmail.com; dwidr07@gmail.com; sudriyanto@unuja.ac.id

Article History
Submitted : July 18, 2025
Revised : Nov 25, 2025
Accepted : Jan 29, 2026
Available Online : Feb 2, 2026
Published Regularly : Feb 2, 2026

Kata Kunci: *Depresi Mahasiswa; Pembelajaran Mesin; Prediksi; Random Forest*

Keywords: *Student Depression; Machine Learning; Prediction; Random Forest*

Contact
Author
abyalfahri18@gmail.com

ABSTRAK
Kesehatan mental mahasiswa merupakan isu penting yang semakin mendapat perhatian, khususnya terkait depresi yang berdampak signifikan pada kualitas hidup dan capaian akademik. Penelitian ini bertujuan mengembangkan model prediksi depresi pada mahasiswa berbasis data psikososial menggunakan algoritma Random Forest. Data yang digunakan merupakan dataset sekunder publik dari Kaggle berjumlah 1000 sampel, mencakup variabel demografis, gaya hidup, dan indikator psikologis. Proses analisis mencakup pra-pemrosesan data, penyeimbangan kelas, pelatihan model, dan evaluasi menggunakan metrik accuracy, precision, recall, F1-score, serta confusion matrix. Hasil pengujian menunjukkan bahwa model Random Forest mampu memprediksi kondisi depresi dengan akurasi 87,0%, precision 86,1%, recall 87,4%, dan F1-score 86,7%, menunjukkan performa yang baik dan stabil. Visualisasi word cloud mengidentifikasi tekanan akademik, stres, dan kecemasan sebagai faktor dominan. Dibandingkan penelitian sebelumnya dengan algoritma SVM, Random Forest menunjukkan peningkatan performa, terutama dalam menangani data kompleks dan tidak seimbang. Penelitian ini menegaskan bahwa pendekatan machine learning berbasis Random Forest efektif untuk mendukung deteksi dini depresi mahasiswa, sekaligus menyediakan dasar bagi pengembangan sistem monitoring kesehatan mental di lingkungan pendidikan tinggi

ABSTRACT
College students' mental health is a critical issue that is gaining increasing attention, particularly regarding depression, which significantly impacts quality of life and academic achievement. This study aims to develop a predictive model for depression in college students based on psychosocial data using the Random Forest algorithm. The data used is a public secondary dataset from Kaggle with 1,000 samples, covering demographic variables, lifestyle, and psychological indicators. The analysis process included data preprocessing, class balancing, model training, and evaluation using accuracy, precision, recall, F1-score, and confusion matrix metrics. Test results showed that the Random Forest model was able to predict depression with 87.0% accuracy, 86.1% precision, 87.4% recall, and 86.7% F1-score, demonstrating good and stable performance. Word cloud visualization identified academic pressure, stress, and anxiety as dominant factors. Compared to previous research using the SVM algorithm, Random Forest demonstrated improved performance, particularly in handling complex and imbalanced data. This study confirms the effectiveness of the Random Forest-based machine learning approach in supporting the early detection of college students' depression and provides a foundation for the development of mental health monitoring systems in higher education settings.

1. Pendahuluan

Kesehatan mental telah menjadi isu global yang mendapat perhatian luas dalam beberapa dekade terakhir. Depresi merupakan salah satu gangguan mental yang paling umum dan berdampak signifikan terhadap kualitas hidup individu [1]. Menurut *World Health Organization* (WHO), lebih dari 322 juta orang di seluruh dunia mengalami depresi, dengan prevalensi yang meningkat secara signifikan di kalangan remaja dan dewasa muda [2]. Gangguan ini tidak hanya menimbulkan penderitaan emosional, tetapi juga berdampak pada penurunan produktivitas, peningkatan risiko bunuh diri, serta membebani sistem layanan kesehatan secara global [3].

Di Indonesia, perhatian terhadap kesehatan mental, khususnya di kalangan mahasiswa, semakin meningkat seiring dengan tingginya tekanan akademik, ketidakpastian masa depan, permasalahan ekonomi, serta tantangan sosial yang dihadapi. Mahasiswa berada pada fase transisi kehidupan yang kompleks dan rentan mengalami tekanan psikologis [4]. Beberapa studi melaporkan bahwa prevalensi depresi pada mahasiswa Indonesia berada pada kisaran 20–30%, tergantung pada karakteristik responden dan instrumen pengukuran yang digunakan [5]. Kondisi ini menegaskan pentingnya upaya deteksi dini guna mencegah dampak jangka panjang terhadap kesehatan mental dan capaian akademik mahasiswa.

Seiring dengan perkembangan teknologi, pendekatan berbasis kecerdasan buatan, khususnya *machine learning*, telah banyak dimanfaatkan dalam bidang prediksi dan klasifikasi gangguan mental [6]. Berbagai algoritma seperti *Support Vector Machine* (SVM), *Decision Tree*, dan *K-Nearest Neighbors* (KNN) telah digunakan dalam penelitian sebelumnya [7]. Meskipun SVM unggul dalam menangani data berdimensi tinggi, metode ini sensitif terhadap pemilihan kernel dan kurang optimal pada data tidak seimbang. *Decision Tree* menawarkan interpretabilitas yang baik, namun rentan terhadap *overfitting*, sedangkan KNN memiliki keterbatasan dari sisi efisiensi komputasi dan sensitivitas terhadap *noise* serta skala fitur.

Dalam penelitian ini, algoritma *Random Forest* dipilih karena mampu mengatasi keterbatasan metode-metode tersebut melalui mekanisme *ensemble learning* yang mengombinasikan banyak pohon keputusan untuk meningkatkan stabilitas dan akurasi prediksi. *Random Forest* lebih tahan terhadap *overfitting*, efektif dalam menangani data berdimensi tinggi dan tidak seimbang, serta mampu mengidentifikasi tingkat kepentingan fitur (*feature importance*) [8], yang relevan dalam analisis faktor-faktor psikososial penyebab depresi [9]. Oleh karena itu, penelitian ini bertujuan untuk membangun model prediksi depresi pada mahasiswa berbasis data psikososial menggunakan algoritma *Random Forest* sebagai upaya mendukung deteksi dini kondisi kesehatan mental di lingkungan pendidikan tinggi.

2. Metode Penelitian

Penelitian ini dilakukan untuk mengembangkan dan mengevaluasi model prediksi tingkat depresi pada mahasiswa berdasarkan data psikososial dengan pendekatan *machine learning.* Metode penelitian yang digunakan bersifat kuantitatif dengan pendekatan deskriptif-prediktif, yang bertujuan untuk mengidentifikasi pola dan hubungan antara variabel psikososial terhadap tingkat depresi mahasiswa [10].

2.1 Desain Penelitian

Penelitian ini merupakan studi kuantitatif dengan pendekatan prediktif yang bertujuan untuk membangun model klasifikasi guna memprediksi tingkat depresi mahasiswa berdasarkan data psikososial [11]. Pendekatan ini dipilih karena mampu mengukur pengaruh variabel input terhadap hasil klasifikasi secara objektif menggunakan algoritma *machine learning*, khususnya *Random Forest*. Penelitian ini bersifat eksperimental, dengan prosedur validasi dan evaluasi model secara terukur.

2.2 Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari dataset publik yang tersedia di *platform Kaggle* 1000 data yaitu *Student Depression Dataset* yang dikembangkan oleh HopesB. Dataset tersebut berisi data kuantitatif dan kualitatif yang dikumpulkan melalui kuesioner terhadap mahasiswa, mencakup informasi demografis, gaya hidup, dan indikator psikologis yang berkaitan dengan depresi. Data diunduh dan dianalisis secara sekunder [12].

2.3 Variabel dan Fitur

Variabel dalam penelitian ini terdiri dari sejumlah fitur psikososial seperti usia, jenis kelamin, status akademik, kebiasaan tidur, aktivitas fisik, hubungan sosial, tingkat stres, serta hasil penilaian depresi. Variabel target adalah label *depressed* atau *not depressed* yang telah ditentukan berdasarkan skoring standar yang melekat pada dataset [12]. Pemilihan fitur dilakukan berdasarkan relevansi psikologis terhadap indikator depresi serta hasil analisis korelasi.

2.4 Pra-pemrosesan Data

Langkah pra-pemrosesan mencakup penghapusan data duplikat dan nilai kosong (*missing values*), konversi tipe data yang tidak sesuai, dan transformasi data kategorikal menjadi numerik menggunakan teknik label *encoding* [13]. Selain itu, data diseimbangkan dengan teknik undersampling untuk mengatasi ketidakseimbangan kelas (*class imbalance*) [14]. Proses ini bertujuan untuk meningkatkan performa dan stabilitas model *machine learning*.

2.5 Pelatihan Model

Algoritma yang digunakan dalam penelitian ini adalah *Random Forest Classifier*, sebuah metode ensembel berbasis pohon keputusan yang dikenal tangguh dalam menghadapi *noise* dan data multivariable [15]. Model dikembangkan menggunakan pustaka Scikit-learn, dengan

parameter seperti *n_estimators = 100* dan *max_depth = None*. Model dilatih pada data latih, dan selanjutnya diujikan pada data uji untuk menilai performanya [16].

2.6 Evaluasi dan Validasi

Evaluasi model dilakukan menggunakan metrik evaluasi klasifikasi yaitu: *accuracy*, *precision, recall*, dan F1-*score*. Selain itu, confusion matrix digunakan untuk memberikan gambaran performa model dalam membedakan antara kelas depresi dan tidak depresi. Visualisasi tambahan berupa *word cloud* dan bar *chart* juga digunakan untuk mendukung interpretasi hasil model secara visual [17].

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

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

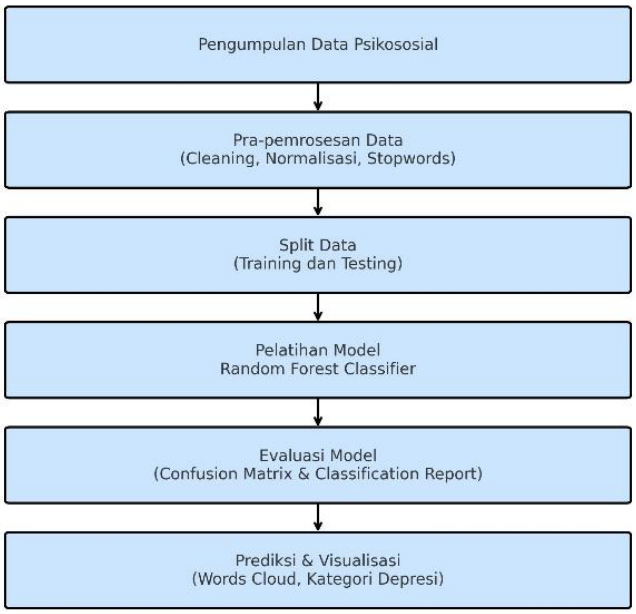
$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

Keterangan:

TP : *True Positive*, jumlah data yang benar-benar positif dan berhasil diprediksi positif oleh model.
TN : *True Negative*, jumlah data yang benar-benar negatif dan berhasil diprediksi negatif oleh model.
FP : *False Positive*, jumlah data yang sebenarnya negatif tetapi diprediksi positif oleh model.
FN : *False Negative*, jumlah data yang sebenarnya positif tetapi diprediksi negatif oleh model.

2.7 *Flowchart* Alur Penelitian

Langkah-langkah dalam penelitian ini mencakup pengumpulan data, prapemrosesan data, pembagian dataset, pembangunan model menggunakan algoritma *Random Forest*, serta evaluasi performa model menggunakan metrik klasifikasi. Seluruh proses dilaksanakan secara sistematis dan terstruktur sebagaimana dijelaskan pada diagram alir penelitian di tunjukan pada Gambar 1.



Gambar 1. *Flowchart* Alur Kerja Model

Berikut ini adalah tahapan alur kerja model:

1. Pengumpulan Data Psikososial

   Pada tahap ini data psikososial mahasiswa dikumpulkan dari dataset publik berbasis kuesioner yang mencakup variabel demografis, akademik, gaya hidup, dan kondisi psikologis sebagai dasar pembentukan model prediksi depresi.

2. Pra-pemrosesan Data (*Cleaning*, Normalisasi, *Stopwords*)

   Pada tahap ini data dipersiapkan melalui pembersihan data, penanganan nilai kosong, normalisasi fitur numerik, serta penghapusan *stopwords* pada data teks untuk meningkatkan kualitas input model.

3. Split Data (Training dan *Testing*)

   Dataset dibagi menjadi data latih dan data uji untuk melatih model serta menguji kemampuan generalisasi dan menghindari *overfitting*.

4. Pelatihan Model *Random Forest Classifier*

   Pada tahap ini model Random Forest dilatih dengan mengombinasikan beberapa pohon keputusan untuk menangkap hubungan non-linear antar fitur dan meningkatkan akurasi prediksi.

5. Evaluasi Model (*Confusion* Matrix & *Classification Report*)

Performa model dievaluasi menggunakan *confusion matrix* dan *classification report* dengan metrik *akurasi*, *precision*, *recall*, dan f1-*score*.

6. Prediksi & Visualisasi (*Word Cloud*, Kategori Depresi)

Tahap akhir model digunakan untuk memprediksi kategori depresi mahasiswa, dan hasilnya divisualisasikan menggunakan *word cloud* serta klasifikasi tingkat depresi untuk memudahkan interpretasi.

3. Hasil dan Pembahasan

3.1 Hasil Penerapan *Model Random Forest*

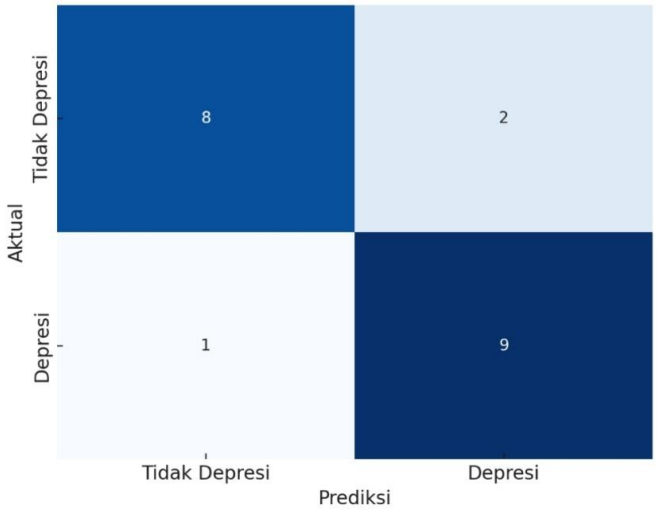
Setelah melalui tahap pra-pemrosesan dan pembagian data, algoritma *Random Forest* diterapkan untuk melakukan klasifikasi kondisi depresi mahasiswa berdasarkan data psikososial. Model menghasilkan keluaran berupa kategori depresi dan tidak depresi pada data uji. Hasil penerapan model ditunjukkan melalui *confusion matrix*, yang merepresentasikan jumlah prediksi benar dan salah pada masing-masing kelas. Berdasarkan hasil pengujian, diperoleh nilai *True Positive* (TP) sebanyak 426 data**,** *True Negative* (TN) sebanyak 444 data**,** *False Positive* (FP) sebanyak 69 data, dan *False Negative* (FN) sebanyak 61 data. Nilai-nilai ini menjadi dasar dalam perhitungan metrik evaluasi model.

3.1 Hasil Evaluasi Model

Berdasarkan nilai TP, TN, FP, dan FN yang diperoleh, dilakukan perhitungan metrik evaluasi model. Akurasi dihitung menggunakan rumus persamaan 1, dengan memasukkan nilai hasil pengujian ke dalam rumus tersebut, diperoleh nilai akurasi sebesar 87,0% yang menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar oleh model. Selain akurasi, evaluasi juga dilakukan menggunakan metrik precision, recall, dan f1-score untuk memberikan gambaran performa model secara lebih komprehensif [18]. Hasil evaluasi terhadap model ditampilkan pada Tabel 1.

Tabel 1. Hasil Evaluasi Kinerja *Model Random Forest*

Metrik Evaluasi	**Akurasi Validasi (%)**
Akurasi	87,0
Precision	86,1
Recall	87,4
F1-*Score*	86,7



Gambar 2. *Confusion Matrix* Model *Random Forest*

Gambar 2. menampilkan *confusion matrix* hasil klasifikasi model Random Forest. Tingginya jumlah *True Positive* dan *True Negative* menunjukkan bahwa model mampu mengklasifikasikan data secara tepat pada kedua kelas. Sebaliknya, jumlah *False Positive* dan *False Negative* relatif rendah, yang menandakan tingkat kesalahan prediksi yang kecil. Hasil ini mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik terhadap data uji, serta mampu menangkap hubungan antara fitur-fitur psikososial seperti tekanan akademik, kondisi emosional, dan hubungan social dengan status depresi mahasiswa. Tolok ukur keberhasilan suatu model klasifikasi ditentukan oleh tingginya nilai metrik evaluasi serta keseimbangan antara *precision* dan *recall*. Dalam penelitian ini, seluruh metrik berada di atas 85%**,** yang secara umum dikategorikan sebagai kinerja baik hingga sangat baik dalam studi klasifikasi berbasis *machine learning.*

3.2 Pembahasan Temuan

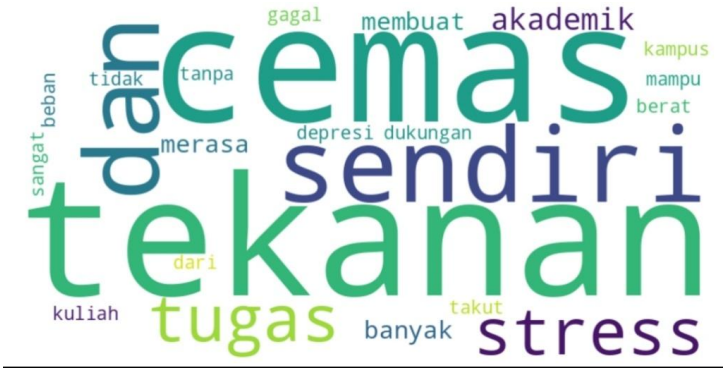
Hasil evaluasi model *Random Forest* yang ditunjukkan pada Tabel. 1 dan Gambar. 2 memberikan indikasi bahwa pendekatan ini mampu mengidentifikasi kondisi depresi mahasiswa secara efektif berdasarkan data psikososial. Tingginya nilai akurasi, *precision, recall*, dan f1-*score* menguatkan bahwa model tidak hanya andal secara statistik, tetapi juga relevan secara praktis dalam konteks deteksi awal masalah psikologis di lingkungan akademik.

Nilai akurasi sebesar 87,0% menunjukkan bahwa model memiliki kemampuan klasifikasi yang andal dalam memprediksi kondisi depresi mahasiswa. Secara umum, nilai akurasi di atas 80% telah dianggap baik, sedangkan nilai di atas 85% menunjukkan performa model yang kuat dan stabil. Nilai recall sebesar 87,4% menunjukkan bahwa model mampu mengidentifikasi lebih dari 87% mahasiswa yang benar-benar mengalami depresi**.** Dalam konteks kesehatan mental, nilai recall yang tinggi sangat penting karena mengurangi risiko *false negative*, yaitu kondisi ketika individu yang mengalami depresi tidak terdeteksi oleh sistem.

Sementara itu, nilai precision sebesar 86,1% mengindikasikan bahwa lebih dari 86% prediksi depresi yang dihasilkan model adalah benar. *Precision* di atas 85% dapat dikategorikan tinggi dan menunjukkan bahwa tingkat kesalahan berupa *false positive* masih berada dalam batas yang dapat diterima, sehingga meminimalkan risiko *over-diagnosis*. Nilai f1-*score* sebesar 86,7% mencerminkan keseimbangan yang baik antara precision dan recall. Hal ini menunjukkan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga konsisten dalam mendeteksi kelas depresi secara proporsional, meskipun terdapat potensi ketidakseimbangan kelas pada data.

Secara keseluruhan, performa model yang ditunjukkan pada penelitian ini dapat dijadikan acuan untuk pengembangan sistem monitoring kesehatan mental berbasis teknologi di lingkungan kampus. Temuan ini sekaligus menegaskan bahwa pendekatan machine learning, khususnya *Random Forest*, dapat diadaptasi secara efektif untuk menangani isu-isu sosial yang kompleks seperti kesehatan mental mahasiswa.

3.3 Analisis Visualisasi *Word Cloud*



Gambar 3. *Word Cloud* Model

Gambar 3 menampilkan visualisasi *word cloud* yang menggambarkan kata-kata dominan dalam data teks mahasiswa. Kata-kata seperti *"stres"*, *"tekanan"*, *"cemas"*, dan *"tugas"* muncul dengan frekuensi tinggi, yang menunjukkan bahwa tekanan akademik merupakan faktor dominan dalam kondisi psikologis mahasiswa. Visualisasi ini berfungsi sebagai analisis pendukung yang memperkuat hasil klasifikasi model, sekaligus memberikan gambaran kontekstual mengenai faktor psikososial yang berkontribusi terhadap depresi mahasiswa.

3.4 Perbandingan dengan Penelitian Sebelumnya

Dibandingkan dengan penelitian sebelumnya yang menggunakan algoritma SVM, penelitian ini menunjukkan peningkatan performa meskipun jumlah data berbeda. Studi [12] menggunakan sekitar 924 responden dan menghasilkan akurasi sebesar 85,2% dengan F1-*score* 80,7%. Sementara itu, penelitian ini menggunakan dataset berjumlah 1000 sampel dan menerapkan algoritma *Random Forest*, yang terbukti lebih stabil dalam menangani data kompleks dan tidak seimbang. Keunggulan *Random Forest* terletak pada kemampuannya membangun banyak pohon keputusan (*decision tree*) secara independen dan menggabungkan hasil voting untuk prediksi akhir, sehingga mengurangi risiko overfitting dan lebih tahan terhadap outlier dibanding SVM. Selain itu, *Random Forest* mampu menangkap interaksi non-*linear* antar fitur dengan lebih baik dan memberikan estimasi pentingnya setiap fitur, sehingga meningkatkan akurasi dan F1-*score* pada dataset dengan variabilitas tinggi atau distribusi kelas yang tidak seimbang. Kombinasi kemampuan ini memungkinkan model *Random Forest* menghasilkan performa yang lebih tinggi dibandingkan SVM pada kasus yang serupa.

3.5 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Pertama, data yang digunakan merupakan data sekunder dari platform publik *Kaggle*, sehingga peneliti tidak memiliki kontrol penuh terhadap proses pengumpulan data, kualitas aslinya, maupun distribusi sampel. Kedua, meskipun dataset relatif besar, representativitas data terhadap populasi mahasiswa secara keseluruhan masih terbatas karena asal sampel dan distribusi karakteristik responden tidak sepenuhnya diketahui. Ketiga, data bersifat *cross-sectional* dan bergantung pada informasi yang sudah dikumpulkan

sebelumnya, sehingga potensi bias atau keterbatasan konteks tetap ada. Selain itu, model yang digunakan belum melalui optimasi *hyperparameter* secara mendalam, sehingga performa model masih berpotensi ditingkatkan.

3.6 Arah Penelitian Selanjutnya

Penelitian mendatang disarankan untuk menggunakan data longitudinal guna mengamati perubahan kondisi psikologis mahasiswa secara dinamis. Selain itu, integrasi fitur tambahan seperti aktivitas media sosial atau pola tidur dapat meningkatkan kedalaman analisis. Penggunaan teknik ensemble learning lanjutan atau optimasi *hyperparameter* juga diharapkan dapat memperbaiki performa model.

4. Kesimpulan

Penelitian ini berhasil mengembangkan model prediksi depresi mahasiswa berbasis data psikososial menggunakan algoritma *Random Forest*. Model menunjukkan performa yang baik dengan akurasi 87,0%, *precision* 86,1%, *recall* 87,4%, dan F1-*score* 86,7%, yang menandakan kemampuan klasifikasi yang stabil dan andal. *Random Forest* terbukti unggul dalam menangani data kompleks dan tidak seimbang, serta mampu mengidentifikasi hubungan non-linear antar fitur psikososial yang berkontribusi terhadap depresi. Visualisasi *word cloud* mendukung temuan bahwa tekanan akademik, stres, dan kecemasan merupakan faktor dominan. Dibandingkan penelitian sebelumnya menggunakan SVM, *Random Forest* menunjukkan peningkatan performa. Keterbatasan penelitian mencakup penggunaan data sekunder publik, keterbatasan representativitas populasi, sifat *cross-sectional* data, serta belum dilakukannya optimasi *hyperparameter* secara mendalam. Penelitian selanjutnya disarankan menggunakan data longitudinal, menambahkan fitur relevan tambahan, dan mengeksplorasi optimasi model untuk meningkatkan performa prediksi lebih lanjut. Temuan ini dapat menjadi dasar pengembangan sistem monitoring kesehatan mental mahasiswa berbasis teknologi.

Daftar Pustaka

[1] G. Limenih, A. MacDougall, M. Wedlake, and E. Nouvet, “Depression and global mental health in the global south: a critical analysis of policy and discourse,” *Int. J. Soc. Determ. Heal. Heal. Serv.*, vol. 54, no. 2, pp. 95–107, 2024.

[2] K. S. Chaudhari, M. P. Dhapkas, A. Kumar, and R. G. Ingle, “Mental disorders–a serious global concern that needs to address,” *Int J Pharm Qual Assur*, vol. 15, no. 02, pp. 973–978, 2024.

[3] G. I. Al Jowf *et al.*, “A public health perspective of post-traumatic stress disorder,” *Int. J. Environ. Res. Public Health*, vol. 19, no. 11, p. 6474, 2022.

[4] N. R. Rohmah and M. Mahrus, “Mengidentifikasi Faktor-faktor Penyebab Stres Akademik pada Mahasiswa dan Strategi Pengelolaannya,” *JIEM J. Islam. Educ. Manag.*, vol. 5, no. 1, pp. 36–43, 2024.

[5] V. Blanco, M. Salmerón, P. Otero, and F. L. Vázquez, “Symptoms of Depression, Anxiety, and Stress and Prevalence of Major Depression and Its Predictors in Female University Students.,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 11, May 2021, doi: 10.3390/ijerph18115845.

[6] S. Verma, C. Sharma, G. Aggarwal, and P. Upadhya, “Artificial intelligence-based approach for classification and prediction of mental health,” in *2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, IEEE, 2024, pp. 708–713.

[7] B. Acharya, “Comparative analysis of machine learning algorithms: KNN, SVM, decision

tree and logistic regression for efficiency and performance," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 12, no. 11, pp. 614–619, 2024.
[8] L. F. Voges, L. C. Jarren, and S. Seifert, "Exploitation of surrogate variables in random forests for unbiased analysis of mutual impact and importance of features," *Bioinformatics*, vol. 39, no. 8, p. btad471, 2023.
[9] J. Buesa *et al.*, "Predictors of postpartum depression in threatened preterm labour: importance of psychosocial factors," *Spanish J. Psychiatry Ment. Heal.*, vol. 17, no. 1, pp. 51–54, 2024.
[10] H. S. BALTACI, D. Kucuker, I. Ozkilic, U. Y. Karatas, and H. A. Ozdemir, "Investigation of Variables Predicting Depression in College Students.," *Eurasian J. Educ. Res.*, no. 93, 2021.
[11] W. Narkbunnum and K. Wisaeng, "Prediction of Depression for Undergraduate Students Based on Imbalanced Data by Using Data Mining Techniques," *Appl. Syst. Innov.*, vol. 5, no. 6, p. 120, 2022.
[12] G. S. Dhillon and S. Kaur, "Depression Among College Students: Prevalence And Associated Risk Factors," *Indian J. Ment. Heal.*, vol. 9, no. 2, 2022.
[13] N. Kosaraju, S. R. Sankepally, and K. Mallikharjuna Rao, "Categorical data: Need, encoding, selection of encoding method and its emergence in machine learning models—a practical review study on heart disease prediction dataset using pearson correlation," in *Proceedings of International Conference on Data Science and Applications: ICDSA 2022, Volume 1*, Springer, 2023, pp. 369–382.
[14] A. Bansal, A. Verma, S. Singh, and Y. Jain, "Combination of oversampling and undersampling techniques on imbalanced datasets," in *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022, Volume 3*, Springer, 2022, pp. 647–656.
[15] M. Maindola *et al.*, "Utilizing random forests for high-accuracy classification in medical diagnostics," in *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, IEEE, 2024, pp. 1679–1685.
[16] K. Vita, P. Yana, B. Liliia, and V. Dmytro, "AUTOMATED DETECTION OF POTENTIALLY DANGEROUS URL ADDRESSES USING THE SCIKIT-LEARN LIBRARY," *Міжнародна науково-технічна конференція Інформаційні технології в металургії та машинобудуванні*, pp. 353–357, 2024.
[17] F. Aziz, S. Abasa, and A. Andyka, "Pengembangan dan Validasi Model Hybrid Machine Learning untuk Diagnosis Awal Depresi," *J. Pharm. Appl. Comput. Sci.*, vol. 3, no. 1, pp. 8–15, 2025.
[18] O. Iparraguirre-Villanueva, C. Paulino-Moreno, A. Epifanía-Huerta, and C. Torres-Ceclén, "Machine Learning Models to Classify and Predict Depression in College Students.," *Int. J. Interact. Mob. Technol.*, vol. 18, no. 14, 2024.