KOMPARASI ALGORITMA MACHINE LEARNING UNTUK MEMPREDIKSI DEPRESI PADA MAHASISWA MENGGUNAKAN TEKNIK SMOTE

Kaleda Zea Arisyandi[1], Andi Taufik[2]

[1,2]*Universitas Bina Sarana Informatika*
E-mail : [1]kaledazea9964@gmail.com, [2]andi.iuf@bsi.ac.id

ABSTRAK

Depresi merupakan salah satu masalah kesehatan mental yang sering dialami oleh mahasiswa dan dapat memberikan dampak buruk terhadap prestasi akademik maupun kualitas hidup. Penelitian ini bertujuan untuk membandingkan kinerja beberapa algoritma *machine learning* yaitu, *Random Forest*, *XGBoost*, dan *Logistic Regression* dalam memprediksi depresi pada mahasiswa serta menganalisis pengaruh penerapan teknik SMOTE (Synthetic Minority Oversampling Technique) dalam mengatasi ketidakseimbangan kelas pada data. Dataset yang digunakan adalah *Student Depression Dataset* yang diperoleh dari platform *Kaggle* dengan jumlah data 27.901 data dan 13 fitur setelah dilakukan filter berdasarkan atribut *Profession* dengan nilai *Student*. Metode yang digunakan meliputi *preprocessing data*, *splitting data* dengan rasio 80:20, penerapan SMOTE pada data *training*, pemodelan menggunakan tiga algoritma *machine learning*, dan evaluasi model menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma *Logistic Regression* mencapai kinerja terbaik dengan akurasi sebesar 85,68%, *precision* sebesar 87,15%, *recall* sebesar 88,60%, dan *F1-score* sebesar 87,87%. Penerapan teknik SMOTE berhasil menyeimbangkan distribusi kelas data *training* dari 58,45% dan 41,55% menjadi 50% dan 50% dan meningkatkan nilai *precision, recall, dan F1-score* pada ketiga algoritma secara konsisten.

Kata Kunci : *Depresi Mahasiswa, Machine Learning, Random Forest, XGBoost, Logistic Regression, SMOTE*

ABSTRACT

Depression is one of the most common mental health issue experienced by university students and can negatively affect both academic performance and quality of life. This study aims to compare the performance of several machine learning algorithms, namely Random Forest, XGBoost, and Logistic Regression, in predicting depression among students, as well as to analyze the impact of applying the Shynthetic Monority Oversampling Technique (SMOTE) to address class imbalance. The dataset used in this study was the Student Depression Dataset obtained from Kaggle, consisting of 27,901 records and 13 features after filtering the Profession attribute to include only students. The research methodology involved data preprocessing, data splitting with an 80:20 ratio, the application of SMOTE to the training data, model development using the three machine learning algorithms, and model evaluation based on accuracy, precision, recall, and F1-score metrics. The results indicate that Logistic Regression achieved the best performance,with an accuracy of 85.68%, precision of 87.15%, recall of 88.60%, and an F1-score of 87.87%. The implementation of SMOTE technique successfully balanced the class distribution in the training data from 58.45% and 41.55% to 50% and 50%, and consistently improved the precision, recall, and F1-score values across all three algorithms.

Keywords : *Student Depression, Machine Learning, Random Forest, XGBoost, Logistic Regression, SMOTE.*

PENDAHULUAN

Fenomena kesehatan mental belakangan ini telah menjadi isu krusial yang memerlukan penanganan segera, terutama di kalangan mahasiswa. Stabilitas mental secara konseptual merujuk pada kemampuan seseorang dalam mengelola stres, mempertahankan hubungan sosial yang sehat, dan menjalankan kewajiban akademik dengan maksimal (Chanafy & Sulistiani, 2026). Mahasiswa merupakan kelompok yang rentan terhadap berbagai tekanan yang berpotensi memicu depresi. Sebagai kelompok usia produktif, mereka menghadapi beban yang kompleks baik dari lingkungan akademik maupun non-akademik (Fadhilla et al., 2025). Tuntunan akademik yang tinggi, persaingan akademik, dan ekspetasi akan prestasi dapat menyebabkan stres kronis. Selain itu, kendala finansial, kualitas tidur yang kurang optimal, gaya hidup yang tidak sehat, serta dinamika sosial yang bermasalah, termasuk dampak penggunaan media sosial juga dapat memperburuk kondisi kesehatan mental para mahasiswa (Alfahrizi et al., 2026).

Menurut laporan Riskesdas 2018 oleh Kementrian Kesehatan Republik Indonesia, tingkat depresi yang menyerang masyarakat dalam kelompok usia ≥15 tahun tercatat sebesar 6,1% dengan presentase gangguan emosional mental yang lebih tinggi yaitu 9,8% (RI, n.d.). Sejalan dengan hal tersebut, *World Health Organization* (WHO) menyatakan bahwa depresi merupakan masalah kesehatan mental yang paling masif terjadi di tingkat global dan menjadi faktor dominan yang memperburuk standar kualitas hidup di berbagai negara (Organization, n.d.).

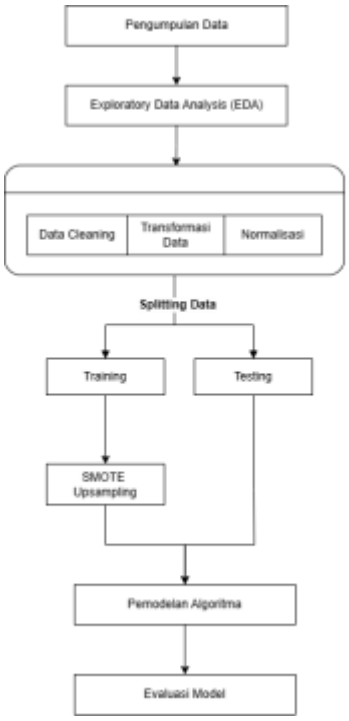
Penelitian terdahulu mengungkapkan bahwa sekitar 26,9% mahasiswa mengalami depresi pada tingkat ringan, 18,5% di tingkat sedang, dan 9,3% berada dalam kondisi depresi berat. Data ini menandakan bahwa mayoritas atau lebih dari setengah total responden memperlihatkan gejala klinis dengan derajat keparahan yang berbeda-beda, sehingga menegaskan pentingnya tindakan intervensi yang lebih serius guna menangani masalah kesehatan mental mahasiswa.

Berbagai penelitian terdahulu telah menggunakan algoritma *machine learning* untuk prediksi depresi dengan hasil yang cukup baik. Penggunaan algoritma *Random Forest* menunjukkan performa yang baik dalam klasifikasi depresi, sementara penelitian lainnya telah membandingkan kinerja algoritma *XGBoost* dan *Lightbm*. Namun, sebagian besar penelitian masih berfokus pada penggunaan satu algoritma atau perbandingan dua algoritma saja. Selain itu, penelitian-penelitian sebelumnya sebagaian besar hanya berfokus pada peningkatan kinerja model melalui *turning parameter* dan implementasi, tanpa melakukan evaluasi menyeluruh terhadap kolaborasi algoritma *machine learning* dengan Teknik *balancing data* seperti SMOTE.

Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja algoritma *machine learning* yaitu, *Random Forest, XGBoost,* dan *Logistic Regression* dengan menerapkan Teknik SMOTE guna mengidentifikasi algoritma dengan performa terbaik dalam memprediksi depresi di kalangan mahasiswa.

METODE

Metode penelitian yang digunakan dalam penelitian ini terdiri dari beberapa tahap untuk mendukung proses komparasi algoritma *machine learning* dalam prediksi depresi pada mahasiswa menggunakan *Altair AI Studio*. Tahapan-tahapan penelitian digambarkan sebagai berikut:



Gambar 1. Alur Penelitian

Pengumpulan Data

Dataset yang digunakan merupakan data sekunder yang diperoleh dari platform *Kaggle* (www.kaggle.com) yang berjudul *Student Depression Dataset*, yang terdiri dari 27.901 baris data dengan 18 fitur dan 1 label target yaitu *Depression*, dengan dua kelas yaitu, 1 (mengalami depresi), dan 0 (Tidak mengalami depresi). Penelitian ini membatasi data pada kategori *Student* melalui proses *filtering* berdasarkan atribut *Profession*. Fitur yang digunakan meliputi faktor demografi (*Gender*, *Age*, *City*), akademik (*Academic Pressure*, *Study Satisfaction, Study Hours, CGPA, Degree*), gaya hidup (*Sleep Duration, Dietary Habits, Financial Stress*), serta Riwayat kesehatan (*Family History of Mental Illness, Have you ever had suicidal thoughts?*).

***Exploratory Data Analysis* (EDA)**

Tahapan awal analisis ini berfokus pada identifikasi pola, menemukan outliner, serta verifikasi asumsi terhadap kumpulan data. Proses ini mencakup identifikasi tipe data, jumlah data, distribusi nilai, pemeriksaan *missing values,* serta analisis ketidakseimbangan kelas pada variabel depresi. Pada tahap ini dilakukan proses *filter examples* dengan memilih hanya data yang memiliki nilai *Profession = Student*, sehingga data yang digunakan dalam penelitian ini hanya mencakup responden yang berstatus mahasiswa (Haryanto et al., 2023).

Preprocessing Data

Adapun tahapan *preprocessing data* adalah sebagai berikut:

1. *Data Cleaning*

Tahapan ini dilakukan untuk mengatasi *missing values* yang terdeteksi pada dataset (M & Harianto, 2025). Setelah dilakukan pemeriksaan, terdeteksi 3 *missing values* pada fitur *Financial Stress*. *Missing values* tersebut diperbaiki dengan menggunakan metode *Replace Missing Values* dengan cara menggantinya menggunakan nilai rata-rata (*average*).

2. Transformasi Data

Tahapan ini mencakup perubahan data kategorikal menjadi data numerik. Metode *transformasi data* yang digunakan adalah *Label Encoding,* yaitu metode yang mengubah data ketagori atau data ordinal menjadi data numerik melalui pemberian label atau kode numerik untuk setiap kategori (Gori et al., 2024). Fitur-fitur yang bersifat kategorikal diubah menjadi bentuk angka dengan menggunakan metode *unique integers* melalui operator *Nominal to Numerical*.

3. Normalisasi

Tahapan ini bertujuan untuk mengubah ukuran data ke dalam jangkauan yang lebih teratur, dimana penggunaannya sangat penting saat dataset memiliki perbedaan skala antar fitur (Allorerung et al., 2024). Pada penelitian ini normalisasi dilakukan dengan menggunakan metode *Z-transformation* untuk menyamakan skala nilai antar atribut pada seluruh fitur sehingga proses pemodelan dapat berjalan lebih optimal.

Splitting Data

Pada tahapan ini proses pemisahan data dilakukan dengan menerapkan perbandingan proporsi 80% untuk data *training* dan 20% untuk data *testing.* Guna memastikan distribusi data tetap representatif, pendekatan *stratified sampling* dengan ketetapan *seed* 42 dipilih dalam mengimplementasikan langkah ini (Afqi et al., 2025). Hasil akhir dari pembagian kuantitas data tersebut mencakup 22.296 data *training* dan 5.574 data *testing*.

SMOTE Upsampling

Tahapan ini dilakukan untuk memperbaiki ketidakseimbangan kelas pada data *training*. Teknik manipulasi data ini diaplikasikan untuk menyelaraskan ketimpangan kelas dalam dataset melalui peningkatan jumlah sampel minoritas secara buatan (Mualfah et al., 2022). Sebelum diterapkan SMOTE, distribusi kelas dalam data *training* tidak seimbang, di mana kelas Depresi = 1 mencapai 58,45% dan kelas Depresi = 0 sebanyak 41,55%. Setelah penerapan SMOTE, distribusi kelas menjadi seimbang dengan masing-masing kelas memiliki 13.406 data, sehingga total data *training* menjadi 26.092 data.

Pemodelan Algoritma

Pada tahapan ini dilakukan dengan melatih tiga algoritma *machine learning*, yaitu *Random Forest, XGBoost*, dan *Logistic Regression* dengan menggunakan data *training* yang telah diseimbangkan melalui teknik SMOTE. Ketiga algoritma tersebut diimplementasikan pada *Altair AI Studio* dengan menggunakan parameter terbaik hasil optimasi untuk menghasilkan model klasifikasi depresi mahasiswa. Pada tahap berikutnya, model yang telah dibangun diimplementasikan pada data *testing* untuk memproyeksikan hasil estimasi, yang kemudian diikuti dengan proses penilaian performa model.

1. *Random Forest*

Random forest merupakan sekumpulan pohon untuk klasifikasi atau regresi yang tidak melalui tahap pemangkasan dan dibentuk dengan cara mengambil sampel secara acak dari sekumpulan data. Di samping itu algoritma ini juga mengurangi risiko *overfitting* dengan menggunakan banyak *decision trees*, sehingga menghasilkan model yang lebih konsisten dengan tingkat akurasi prediksi yang lebih tinggi (Maulana et al., 2025).

2. *Extreme Gradient Boosting (XGBoost)*

Extreme Gradient Boosting (XGBoost) merupakan algoritma *machine learning* berbasis ensemble jyang mengintegrasikan metode *gradient boosting decision* (GBDT) demi mengoptimalkan hasil akurasi pada tugas klasifikasi. Algoritma ini memiliki reputasi tinggi terkait

efisiensi komputasi dan ketepatan prediksinya, terutama ketika dihadapkan pada dataset yang berskala besar atau memiliki kompleksitas tinggi (Gunawan et al., 2022).

3. *Logistic Regression*

Logistic Regression merupakan model linear untuk klasifikasi binner yang memprediksi probabilitas kelas positif (1) berdasarkan variabel independennya. Algoritma ini bekerja dengan menghitung kombinasi fitur input secara linear, lalu mengubahnya menggunakan fungsi sigmoid untuk menghasilkan probabilitas antara 0 dan 1 (Afqi et al., 2025).

Evaluasi Model

Tahapan ini dilakukan guna memantau efisiensi prediktif dari setiap algoritma klasifikasi dengan melibatkan total 5.574 data *testing*. Penelitian ini mengimplementasikan instrumen *confusion matrix* yang diwakili oleh metrik uji berupa akurasi, *precision, recall*, serta capaian *F1-Score* (Fadli & Saputra, 2023).

1. Akurasi

Akurasi adalah ukuran yang digunakan untuk menilai seberapa tepat model dalam mengklasifikasikan semua data (Fadli & Saputra, 2023).

$$Accuracy = \frac{TP\ +\ TN}{TP\ +\ TN\ +\ FP\ +\ FN}$$

2. *Precision*

Precision adalah metrik yang berfungsi sebagai tolak ukur untuk menilai tingkat keakuratan model ketika menetapkan label positif (Fadli & Saputra, 2023).

$$Precision = \frac{TP}{TP + FP}$$

3. *Recall*

Recall adalah metrik yang digunakan sebagai tolak ukur untuk mengevaluasi kemampuan suatu model dalam mengidentifikasi seluruh sampel yang termasuk dalam kategori positif (Fadli & Saputra, 2023).

$$Recall = \frac{TP}{TP + FN}$$

4. *F1-Score*

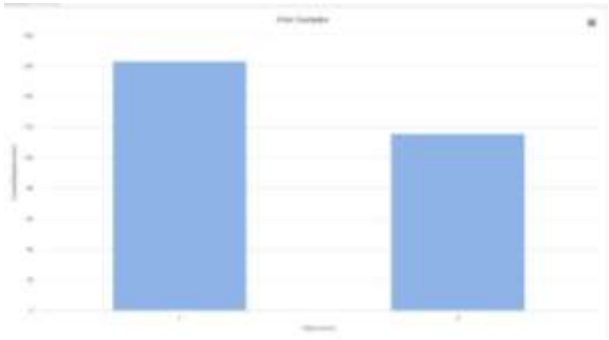
F1-Score adalah rata-rata harmonis antara *precision* dan *recall* yang berfungsi untuk menyeimbangkan kedua ukuran tersebut saat menilai performa model klasifikasi (Fadli & Saputra, 2023).

$$F1\ Score = 2 \times \frac{Recall\ \times Precission}{Recall + Precission}$$

HASIL PENELITIAN DAN PEMBAHASAN

Exploratory Data Analysis (EDA)

Penerapan EDA ini ditunjukkan untuk membedah sekaligus mengenali karakteristik dasar dari dataset sebelum melangkah ke tahapan proses pemodelan. Analisis ini mencakup pemeriksaan statistik deskriptif, distribusi data, dan juga distribusi kelas pada label target depresi. Hasil analisis distribusi kelas label target *Depression* ditunjukkan pada gambar berikut :



Gambar 2. Distribusi Kelas Depression

Berdasarkan gambar 3.1, dataset setelah dilakukan proses filtering terdiri dari 27.870 data mahasiswa dengan distribusi kelas yang tidak seimbang. Distribusi kelas pada label depression = 1 memiliki jumlah data yang lebih banyak dibandingkan dengan kelas Depression = 0. menunjukkan adanya ketidakseimbangan kelas dengan kelas Depression = 1 (mengalami depresi). Kondisi ketidakseimbangan kelas ini menjadi dasar penerapan teknik SMOTE pada tahap selanjutnya untuk menyeimbangkan distribusi kelas sebelum proses pemodelan dilakukan.

Preprocessing Data

Tahapan *preprocessing* data meliputi proses penyaringan data, penanganan *missing values*, tranformasi data kategorikal menjadi numerikal, serta normalisasi menggunakan metode *Z-transformation.* Pada tahap penyaringan data dilakukan berdasarkan atribut *Profession* dengan nilai *Student* menggunakan operator *Filter Examples*, sehingga diperoleh sebanyak 27.870 data mahasiswa. Selanjutnya, proses data *cleaning* dilakukan untuk mengatasi *missing values* pada dataset. Setelah dilakukan analisis data, ditemukan 3 *missing values* pada fitur *Financial Stress. Missing Values* tersebut ditangani dengan menggunakan operator *Replace Missing Values* dengan menggunakan metode *average*, yang berarti mengganti nilai yang kosong dengan nilai rata-rata dari fitur tersebut. Tahap transformasi data dilakukan dengan mengubah fitur-fitur yang bersifat kategorikal menjadi nilai numerik dengan menggunakan operator *Nominal to Numerical* dengan metode *unique integers.* Fitur yang mengalami perubahan meliputi *Gender, City, Degree, Sleep Duration, Dietary Habits, Family History of Mental Ilnes, Have you ever had suicidal thoughts?,* dan *Depression*. Sedangkan proses normalisasi dilakukan menggunakan operator *Normalize* dengan metode *Z-transformation*. Normalisasi bertujuan untuk menyamakan skala nilai di semua fitur numerik, sehingga setiap fitur memiliki rata-rata 0 dan standar deviasi 1.

Splitting Data dan SMOTE

Proses *splitting* data diterapkan guna mendistribusikan dataset ke dalam porsi data *training* dan data *testing* menggunakan rasio perbadingan 80:20. Langkah pemisahan ini mengadopsi

pendekatan *stratified sampling* berparameter *seed* 42 demi memelihara proporsi sebaran kelas agar tetap seimbang. Hasil pembagian data ditunjukkan pada tabel berikut:

Tabel 1. Splitting Data

Data	Jumlah
Training (80%)	22.296
Testing (20%)	5.574
Total	**27.870**

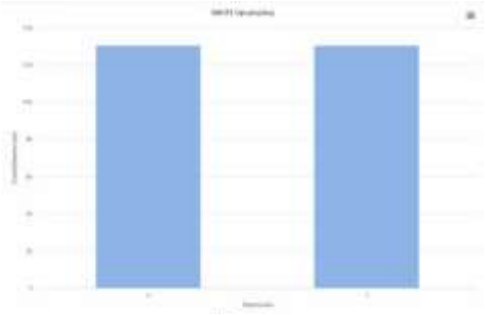
Berdasarkan tabel 1 dataset diberhasil dibagi menjadi 22.296 data *training* dan 5.574 data *testing*. Data *training* digunakan untuk proses pelatihan model, sedangkan data *testing* digunakan untuk mengevaluasi performa model yang dihasilkan.

Perbandingan distribusi kelas sebelum dan setelah SMOTE ditunjukkan pada tabel berikut:

Tabel 2. Hasil SMOTE Upsampling

Kelas	Sebelum SMOTE	Setelah SMOTE
Depression = 1	13.017 (58,45%)	13.046 (50%)
Depression = 0	9.279 (41,55%)	13.046 (50%)
Total	**22.296**	**26.092**

Tabel 2 menunjukkan jumlah kelas sebelum dan setelah diterapkan teknik SMOTE, jumlah data *training* meningkat menjadi 26.092 data. Penambahan jumlah data tersebut terjadi karena SMOTE menghasilkan data sintesis pada kelas minoritas sehingga distribusi kelas menjadi lebih seimbang sebelum proses pemodelan dilakukan.



Gambar 3. Distribusi kelas setelah SMOTE Upsampling

Berdasarkan gambar 3 distribusi kelas pada data training setelah penerapan teknik SMOTE menunjukkan kondisi yang seimbang antara kelas Depression = 1 dan Depression = 0. Masing-masing kelas memiliki jumlah data sebanyak 13.046 data atau sebesar 50%, sehingga ketidakseimbangan kelas (class imbalance) pada data training telah berhasil diatasi sebelum proses pemodelan dilakukan.

Hasil Pemodelan dan Evaluasi

Tabel 3. Hasil Evaluasi

Algoritma	Akurasi	Precision	Recall	F1-Score
Random Forest	83,12%	86,01%	84,98%	85,49%
XGBoost	85,24%	86,70%	88,32%	87,50%
Logistic Regression	85,68%	87,15%	88,60%	87,87%

Berdasarkan hasil evaluasi, algoritma *Logistic Regression* mencapai kinerja optimal dibandingkan kedua algoritma lainnya dengan nilai akurasi tertinggi sebesar 85,68%, *Precision* tertinggi sebesar 87,15%, *Recall* tertinggi sebesar 88,60%, dan *F1-score* tertinggi sebesar 87,87%. Sementara itu, *XGBoost* berada pada posisi kedua dengan akurasi sebesar 85,24%, dengan *F1-score* sebesar 87,50%. Sedangkan *Random Forest* memperoleh akurasi 83,12% dan *F1-score* sebesar 85,49%. Hasil menunjukkan bahwa algoritma *Logictic Regression* memiliki kemampuan yang lebih baik dalam mendeteksi mahasiswa yang mengalami depresi, sehingga mampu meminimalkan kesalahan klasifikasi pada kelas positif (*False Negative*).

Tabel 4. Perbandingan Performa Model

Algoritma	Akurasi	Precision	Recall	F1-score
Tanpa SMOTE				
Random Forest	83,98%	82,35%	78,11%	80,18%
XGBoost	85,38%	83,82%	80,23%	81,99%
Logistic Regression	85,15%	79,44%	86,59%	82,86%
Dengan SMOTE				
Random Forest	83,12%	86,01%	84,98%	85,49%
XGBoost	85,24%	86,70%	88,32%	87,50%
Logistic Regression	85,68%	87,15%	88,60%	87,87%

Berdasarkan tabel 4, penggunaan teknik SMOTE Upsampling telah berhasil mengatasi ketidakseimbangan distribusi kelas pada data *training* dari kondisi tidak seimbang yang mencapai 58,45% dan 41,55% hingga menghasilkan distribusi yang seimbang sempurna, yaitu masing-masing 50%. Meskipun peningkatan akurasi pada ketiga algoritma tidak terlalu signifikan, penerapan SMOTE terbukti meningkatkan nilai *precision, recall,* dan *F1-score* secara konsisten. Secara khusus, pertumbuhan skor *recall* memberikan petunjuk bahwa efektivitas sistem dalam mengidentifikasi mahasiswa dengan gejala depresi mengalami peningkatan yang berarti pasca-penyeimbangan proporsi sampel data. Hal ini sangat krusial dalam konteks klasifikasi depresi, karena kemampuan model unntuk mendeteksi kasus depresi yang sebenarnya lebih diprioritaskan dibandingkan hanya berfokus pada nilai akurasi.

KESIMPULAN

Berdasarkan hasil penelitan mengenai komparasi algoritma *machine learning* untuk prediksi depresi pada mahasiswa menggunakan teknik SMOTE, diperoleh kesimpulan sebagai berikut:

1. Ketiga algoritma yang digunakan, yaitu *Random Forest, XGBoost,* dan *Logistic Regression,* mampu mengklasifikasikan depresi pada mahasiswa dengan baik. Hasil evaluasi menunjukkan bahwa algoritma *Random Forest* meenghasilkan akurasi sebesar 83,12% dengan *F1-score* sebesar 85,49%, algoritma *XGBoost* menghasilkan akurasi sebesar 85,24%, dengan *F1-score* sebesar 87,50%, dan algoritma *Logistic Regression* menghasilkan akurasi sebesar 85,68% dengan *F1-score* sebesar 87,87%.
2. Algoritma *Logistic Regression* menunjukkan hasil terbaik dibandingkan kedua algoritma lainnya dengan tingkat akurasi tertinggi mencapai 85,68%, *precision* sebesar 87,15%, *recall* sebesar 88,60%, dan *F1-score* sebesar 87,87%. Hal ini menunjukkan bahwa hubungan antara

fitur-fitur yang digunakan dengan label depresi bersifat cukup linear sehingga *Logistic Regression* dapat mengidentifikasikannya dengan efektif.

3. Penerapan teknik SMOTE berhasil mengatasi ketidakseimbangan kelas pada data *training*, yaitu dari kondisi awal yang tidak seimbang dengan kelas *Depression* = 1 sebanyak 13.017 data (58,45%) dan kelas *Depression* = 0 sebanyak 9.279 data (41,55%) menjadi seimbang sempurna dengan masing-masing kelas sebanyak 13.046 data (50%). Meskipun penggunaan SMOTE tidak selalu meningkatkan akurasi secara signifikan, SMOTE berhasil meningkatkan nilai *precision*, *recall*, dan *F1-score* pada ketiga algoritma sehingga kemampuan model dalam mendeteksi mahasiswa yang mengalami depresi menjadi lebih baik.

Berdasarkan hasil komparasi, setiap algoritma memiliki karakteristik yang berbeda. Namun, *Logistic Regression* memberikan performa terbaik dan paling sesuai untuk digunakan dalam prediksi depresi pada mahasiswa pada dataset yang digunakan dalam penelitian ini.

DAFTAR PUSTAKA

Afqi, N., Akmal, N., & Budiman, F. (2025). *Analisis Perbandingan Metode Random Forest , XGBoost , dan Logistic Regression Untuk Klasifikasi Deteksi Dini Penyakit Diabetes*. *12*(6), 995–1005. https://doi.org/10.30865/jurikom.v12i6.9392

Alfahrizi, A., Arifiansyah, P., Afandi, M., & Riskianto, D. D. (2026). *Prediksi Depresi Mahasiswa Menggunakan Algoritma Random Forest Berbasis Data Psikososial Depression Prediction Among University Students Using a Random Forest Algorithm Based on Psychosocial Data*. *4*(1), 12–21.

Allorerung, P. P., Erna, A., & Bagussahrir, M. (2024). *Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit*. *9*(3), 178–191.

Chanafy, M., & Sulistiani, H. (2026). *Comparative Analysis Of Machine Learning Algorithm Performance For Student Mental Health Classification*. *16*(79), 19–29.

Fadhilla, M., Wandri, R., Hanafiah, A., Setiawan, P. R., Arta, Y., & Daulay, S. (2025). *Analisis Performa Algoritma Machine Learning Untuk Identifikasi Depresi Pada Mahasiswa*. *5*(1), 40–47. https://doi.org/10.47065/jimat.v5i1.473

Fadli, M., & Saputra, R. A. (2023). *KLASIFIKASI DAN EVALUASI PERFORMA MODEL RANDOM FOREST UNTUK PREDIKSI STROKE Classification And Evaluation Of Performance Models Random Forest For Stroke Prediction*. *12*(02), 72–80.

Gori, T., Sunyoto, A., Fatta, H. Al, Yogyakarta, U. A., Korespondensi, P., Selection, C. F., & Siswa, K. A. (2024). *PREPROCESSING DATA DAN KLASIFIKASI UNTUK PREDIKSI KINERJA DATA PREPROCESSING AND CLASSIFICATION FOR PREDICTING STUDENT*. *11*(1), 215–224. https://doi.org/10.25126/jtiik.20241118074

Gunawan, R., Handika, E. S., & Ismanto, E. (2022). *Pendekatan Machine Learning Dengan Menggunakan Algoritma Xgboost (Extreme Gradient Boosting) Untuk Peningkatan Kinerja Klasifikasi Serangan Syn*. *3*(3), 453–463.

Haryanto, C., Rahaningsih, N., Basysyar, F. M., Cirebon, K., Regression, R. F., & Rumah, H. (2023). *KOMPARASI ALGORITMA MACHINE LEARNING DALAM MEMPREDIKSI HARGA RUMAH*. *1*(1), 533–539.

M, D. P., & Harianto, K. (2025). *Perbandingan Penanganan Missing Value pada Data Numerik Survei Kepuasan Pengguna Lulusan*. *5*(2), 789–798.

Maulana, I. S., Tiara, S. D., Pratama, A. Y., Sari, F. K., Darmawan, I., & Madura, U. (2025).

Prediksi Risiko Depresi pada Mahasiswa Menggunakan Algoritma Random Forest Berdasarkan Data Akademik dan Gaya Hidup. *4*(1), 1–10.

Mualfah, D., Fadila, W., Firdaus, R., Studi, P., Informatika, T., Komputer, F. I., Riau, U. M., & Riau, U. M. (2022). *Teknik SMOTE untuk mengatasi imbalance data pada deteksi penyakit stroke menggunakan algoritma random forest*. *3*(2), 107–113.

Organization, W. H. (n.d.). *Depression*.

RI, K. K. (n.d.). *Laporan Nasional Riskesdas 2018*.