

Klasifikasi Kondisi Kesehatan Mental Mahasiswa Menggunakan Algoritma Logistic Regression

Dini Pratiwi ¹, Yulia Roslina ²

^{1,2}Teknologi Rekayasa Internet, Politeknik Negeri Lampung

INFORMASI ARTIKEL

Diterima 3 Januari 2026
Direvisi 27 Februari 2026
Diterbitkan 28 Februari 2026

Kata kunci:

kesehatan mental;
mahasiswa;
klasifikasi;
regresi logistik;
machine learning

ABSTRAK

Kesehatan mental mahasiswa merupakan isu penting dalam pendidikan tinggi karena berdampak langsung terhadap kesejahteraan dan performa akademik. Tekanan akademik, sosial, dan psikologis yang dihadapi mahasiswa meningkatkan risiko gangguan kesehatan mental seperti depresi dan kecemasan. Penelitian ini bertujuan untuk mengklasifikasikan kondisi kesehatan mental mahasiswa, khususnya risiko depresi, menggunakan algoritma Regresi Logistik serta membandingkan kinerjanya dengan model baseline dan K-Nearest Neighbors (KNN). Dataset yang digunakan adalah Student Mental Health yang diperoleh dari platform Kaggle, terdiri dari 101 data mahasiswa dengan variabel demografis, akademik, dan psikologis. Tahapan penelitian meliputi pra-pemrosesan data, pembagian data latih dan data uji dengan rasio 80:20, pemodelan klasifikasi, serta evaluasi menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Regresi Logistik memberikan performa terbaik dibandingkan model lainnya dengan akurasi sebesar 0,85, precision 1,00, recall 0,57, dan F1-score 0,73. Model baseline memiliki akurasi 0,65 namun tidak mampu mendeteksi kasus depresi, sedangkan KNN ($k = 5$) menghasilkan akurasi 0,55 dengan performa yang lebih rendah. Analisis lebih lanjut menunjukkan bahwa faktor psikologis seperti Marital, Treatment, dan Anxiety berkontribusi signifikan terhadap prediksi depresi pada mahasiswa. Berdasarkan hasil tersebut, Regresi Logistik dinilai efektif dan relevan untuk klasifikasi risiko depresi pada mahasiswa serta berpotensi digunakan sebagai alat bantu deteksi dini kesehatan mental di lingkungan pendidikan tinggi.

Classification of University Students' Mental Health Conditions Using the Logistic Regression Algorithm

ARTICLE INFO

Received January 3, 2026
Revised February 27, 2026
Published February 28, 2026

Keyword:

Mental health;
students;
classification;
logistic regression;
machine learning

ABSTRACT

Student mental health has become a critical issue in higher education, as it directly affects students' well-being and academic performance. Academic, social, and psychological pressures faced by university students increase the risk of mental health disorders such as depression and anxiety. This study aims to classify students' mental health conditions, particularly the risk of depression, using the Logistic Regression algorithm and to compare its performance with a baseline model and the K-Nearest Neighbors (KNN) algorithm. The dataset used in this study is the Student Mental Health dataset obtained from the Kaggle platform, consisting of 101 student records with demographic, academic, and psychological variables. The research process includes data preprocessing, splitting the dataset into training and testing sets with an 80:20 ratio, classification modeling, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The results show that Logistic Regression achieves the best performance compared to the other models, with an accuracy of 0.85, precision of 1.00, recall of 0.57, and an F1-score of 0.73. The baseline model achieves an accuracy of 0.65 but fails to detect any depression cases, while KNN ($k = 5$) produces a lower accuracy of 0.55. Further analysis indicates that psychological factors such as Marital, Treatment, and Anxiety significantly contribute to the prediction of depression among students. Based on these findings, Logistic Regression is considered an effective and relevant approach for classifying depression risk among university students and has the potential to support early detection of mental health problems in higher education environments.

This work is licensed under a [Creative Commons Attribution 4.0](#)



Corresponding Author:

Corresponding Dini Pratiwi, Politeknik Negeri Lampung
Email: dinipratiwi0183@gmail.com

1. PENDAHULUAN

Dalam beberapa tahun terakhir, isu kesehatan mental menjadi perhatian utama di berbagai sektor, termasuk pendidikan tinggi, seiring meningkatnya prevalensi gangguan psikologis pada kelompok usia produktif. Mahasiswa sebagai bagian dari kelompok tersebut menghadapi berbagai tuntutan akademik dan non-akademik yang berpotensi memengaruhi kondisi psikologis serta kualitas hidup mereka secara keseluruhan.

Kesehatan mental mahasiswa telah menjadi perhatian krusial dalam lingkungan pendidikan tinggi, mengingat dampaknya terhadap performa akademik dan kesejahteraan secara keseluruhan [1]. Tekanan akademis, sosial, dan personal yang terus meningkat berkontribusi pada prevalensi masalah kesehatan mental di kalangan mahasiswa, sehingga menegaskan pentingnya upaya deteksi dan intervensi dini [2]. Kondisi ini diperparah oleh temuan bahwa mahasiswa cenderung memiliki tingkat depresi yang lebih tinggi dibandingkan kelompok usia lain, baik pada fase awal maupun akhir masa perkuliahan, yang masing-masing memiliki tantangan dan stresor tersendiri [3].

Seiring berkembangnya teknologi, pendekatan berbasis machine learning semakin banyak dimanfaatkan dalam penelitian kesehatan mental untuk mendukung proses klasifikasi dan prediksi risiko gangguan psikologis. Pengembangan metode klasifikasi yang akurat menjadi sangat relevan,

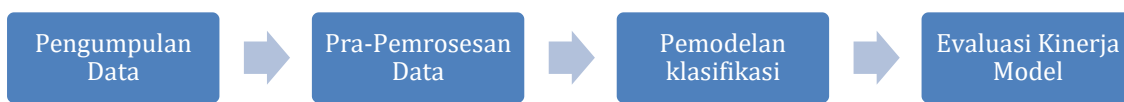
khususnya dengan memanfaatkan algoritma Regresi Logistik, yang dikenal efisien dalam memodelkan hubungan antara variabel dependen biner dan variabel independen serta memiliki tingkat interpretabilitas yang baik [4], [5].

Beberapa penelitian terdahulu telah mengeksplorasi penggunaan berbagai algoritma machine learning dalam klasifikasi kesehatan mental mahasiswa, seperti Random Forest, Support Vector Machine, dan Naïve Bayes Classifier, dengan hasil performa yang bervariasi [6], [7]. Meskipun demikian, masih terdapat keterbatasan dalam kajian yang secara komprehensif menilai performa Regresi Logistik dibandingkan algoritma lain, khususnya dalam konteks keseimbangan antara akurasi prediktif dan kemudahan interpretasi hasil model.

Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi secara komprehensif performa algoritma Regresi Logistik dalam mengklasifikasikan kondisi kesehatan mental mahasiswa, dengan fokus pada identifikasi kelebihan dan kekurangannya dibandingkan dengan algoritma machine learning lain yang relevan. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode klasifikasi yang lebih efektif serta menjadi dasar pertimbangan bagi institusi pendidikan dalam merancang strategi deteksi dini dan intervensi kesehatan mental mahasiswa.

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komparatif. Pendekatan ini bertujuan untuk membangun dan mengevaluasi model klasifikasi dalam mengidentifikasi risiko gangguan kesehatan mental mahasiswa, serta membandingkan performa algoritma Regresi Logistik dengan model pembelajaran mesin pembanding. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, pemodelan klasifikasi, dan evaluasi kinerja model, seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian dalam Pengembangan dan Evaluasi Model Klasifikasi

2.1. Dataset dan Variabel Penelitian

Dataset yang digunakan dalam penelitian ini adalah dataset publik Student Mental Health yang diperoleh dari platform Kaggle. Nama dataset Nama dataset Kaggle: Student Mental Analysis[EDA + ML]Dataset ini berisi data hasil survei kondisi kesehatan mental mahasiswa yang mencakup aspek demografis, akademik, dan psikologis.

Dataset terdiri dari 101 observasi dengan 11 variabel yang ditunjukkan pada Tabel 1. Variabel *Depression* digunakan sebagai variabel target (label) dalam pemodelan klasifikasi, sedangkan variabel lainnya digunakan sebagai variabel prediktor. Variabel *Timestamp* tidak digunakan dalam pemodelan karena tidak memiliki relevansi prediktif terhadap kondisi kesehatan mental mahasiswa.

Tabel 1. Deskripsi Variabel pada Dataset Student Mental Health

Variabel	Deskripsi	Tipe Data
Gender	Jenis kelamin mahasiswa	Kategorik
Age	Usia mahasiswa	Numerik
Course	Program studi mahasiswa	Kategorik
Year	Tahun akademik mahasiswa	Kategorik
CGPA	Indeks Prestasi Kumulatif (rentang nilai)	Kategorik
Marital	Status pernikahan mahasiswa	Kategorik
Depression	Indikator depresi (Yes/No)	Kategorik
Anxiety	Indikator kecemasan (Yes/No)	Kategorik
Panic	Indikator serangan panik (Yes/No)	Kategorik
Treatment	Upaya mencari bantuan profesional (Yes/No)	Kategorik
Timestamp	Waktu pengisian kuesioner	Waktu

2.2. Pra-pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk memastikan kualitas dan kesiapan data sebelum proses pemodelan. Proses pembersihan data dilakukan dengan menghapus observasi yang memiliki nilai usia tidak wajar agar mengurangi potensi *noise*. Dataset tidak mengandung nilai hilang, sehingga proses imputasi tidak diperlukan.

Seluruh variabel kategorik diubah ke dalam bentuk numerik menggunakan teknik label encoding. Variabel *Depression* sebagai target dikodekan menjadi nilai 0 untuk kondisi tidak depresi dan 1 untuk kondisi depresi. Variabel CGPA yang semula berbentuk rentang nilai dikonversi menjadi nilai numerik dengan mengambil nilai rata-rata dari batas bawah dan batas atas setiap rentang.

Selanjutnya, fitur numerik *Age* dan CGPA dinormalisasi menggunakan metode *StandardScaler* sehingga memiliki rata-rata nol dan simpangan baku satu. Proses normalisasi ini dilakukan untuk meningkatkan stabilitas model dan menghindari dominasi fitur tertentu, khususnya pada algoritma yang sensitif terhadap skala data.

2.3. Pembagian Data

Dataset yang telah melalui tahap pra-pemrosesan dibagi menjadi data latih dan data uji dengan perbandingan 80% data latih dan 20% data uji. Pembagian data dilakukan menggunakan teknik *stratified sampling* untuk menjaga proporsi kelas depresi dan tidak depresi tetap seimbang pada kedua subset data.

2.4. Pemodelan Klasifikasi

Penelitian ini menggunakan tiga model klasifikasi, yaitu satu model baseline dan dua model utama. Model baseline dibangun menggunakan *DummyClassifier* dengan strategi *most frequent*, yang berfungsi sebagai tolok ukur performa minimum.

Model utama yang digunakan adalah Regresi Logistik, yang dipilih karena kemampuannya dalam memodelkan probabilitas kejadian biner serta kemudahan interpretasi hasil model. Sebagai pembanding, digunakan algoritma K-Nearest Neighbors (KNN) dengan nilai *k* sebesar 5. Seluruh model dilatih menggunakan data latih yang sama untuk memastikan perbandingan kinerja yang adil [8].

2.5. Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, digunakan *confusion matrix* untuk menganalisis kesalahan klasifikasi secara lebih rinci. Perbandingan kinerja antar model juga disajikan dalam bentuk tabel dan grafik pada bagian hasil dan pembahasan untuk mempermudah interpretasi hasil.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pemodelan Klasifikasi

Dataset hasil preprocessing dibagi menjadi data latih dan data uji dengan perbandingan 80%:20% menggunakan *train-test split* dengan *stratified sampling* sehingga proporsi kelas depresi dan tidak depresi tetap seimbang. Variabel *Depression* digunakan sebagai target, sedangkan fitur terdiri dari *Gender*, *Age*, *Course*, *Year*, *CGPA*, *Marital*, *Anxiety*, *Panic*, dan *Treatment*.

Sebagai tolok ukur awal, dibangun baseline model menggunakan *DummyClassifier* dengan strategi *most frequent*, yaitu model yang selalu memprediksi kelas mayoritas. Hasil evaluasi menunjukkan bahwa baseline memiliki akurasi sebesar 0,65, namun nilai *precision*, *recall*, dan *F1-score* bernilai 0,00. Hal ini menandakan bahwa meskipun akurasinya tampak cukup tinggi, model sama sekali tidak mampu mengenali kasus mahasiswa yang mengalami depresi karena seluruh prediksi diarahkan pada kelas tidak depresi.

Selanjutnya, dibangun dua model klasifikasi, yaitu Logistic Regression sebagai model utama dan K-Nearest Neighbors (KNN) dengan *k* = 5 sebagai model pembanding. Kedua model dilatih

menggunakan data latih yang sama dan dievaluasi pada data uji. Ringkasan hasil evaluasi ditunjukkan pada Tabel 2.

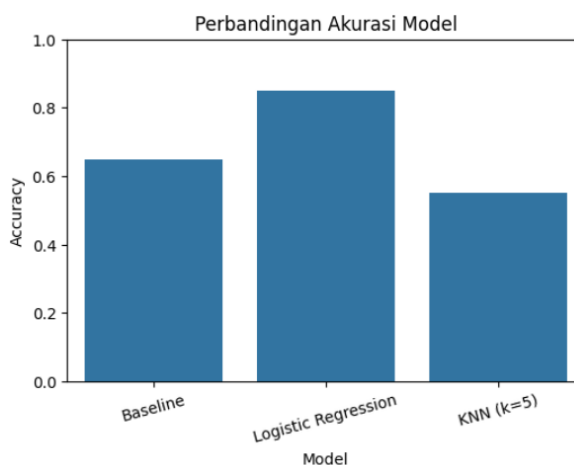
Tabel 2. Hasil evaluasi baseline dan model klasifikasi

Model	Akurasi	Presisi	Recall	F1-score
Baseline	0,65	0,00	0,00	0,00
Logistic Regression	0,85	1,00	0,57	0,73
KNN (k = 5)	0,55	0,33	0,29	0,31

Dari Tabel 2 terlihat bahwa Logistic Regression memberikan performa terbaik dengan akurasi sebesar 0,85 dan F1-score 0,73. Nilai precision yang mencapai 1,00 menunjukkan bahwa seluruh mahasiswa yang diprediksi depresi oleh model memang benar-benar termasuk dalam kelas depresi di data uji (tidak ada *false positive*). Namun, nilai recall 0,57 mengindikasikan bahwa masih terdapat sebagian mahasiswa depresi yang belum terdeteksi oleh model.

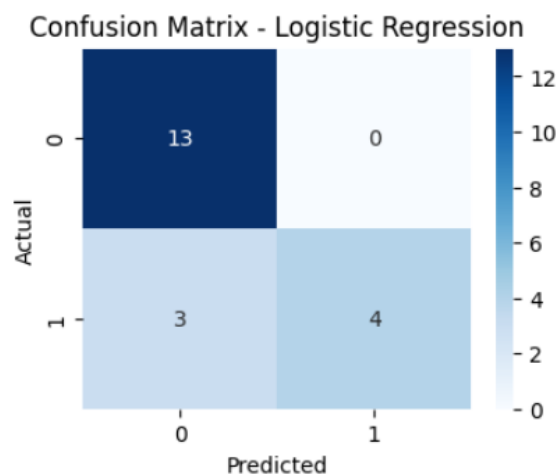
Untuk memvisualisasikan perbedaan kinerja antarmodel, dibuat grafik perbandingan akurasi seperti ditunjukkan pada Gambar 2. Grafik tersebut memperlihatkan bahwa Logistic Regression memiliki akurasi tertinggi (0,85), diikuti baseline (0,65), sedangkan KNN (k = 5) memiliki akurasi paling rendah (0,55). Perbedaan ini menegaskan bahwa model yang benar-benar belajar dari pola data memberikan peningkatan performa yang signifikan dibanding baseline yang hanya mengikuti kelas mayoritas.

Untuk memvisualisasikan perbandingan kinerja antarmodel, grafik akurasi ditampilkan pada Gambar 2.



Gambar 2. Perbandingan akurasi antara baseline, Logistic Regression, dan KNN

Distribusi prediksi model Logistic Regression dianalisis lebih lanjut menggunakan confusion matrix yang ditampilkan pada Gambar 3. Confusion matrix menunjukkan bahwa model berhasil mengklasifikasikan 13 mahasiswa tidak depresi secara benar (*true negative*) dan 4 mahasiswa depresi secara benar (*true positive*). Namun masih terdapat 3 mahasiswa depresi yang diprediksi sebagai tidak depresi (*false negative*), sedangkan tidak ditemukan kasus mahasiswa tidak depresi yang salah diprediksi depresi (*false positive* = 0). Dalam konteks deteksi dini kesehatan mental, keberadaan *false negative* ini perlu mendapat perhatian karena berpotensi menyebabkan mahasiswa yang sebenarnya membutuhkan bantuan tidak teridentifikasi oleh sistem.



Gambar 3. Confusion matrix untuk model Logistic Regression pada dataset Student Mental Health

3.2. Pembahasan

Hasil penelitian menunjukkan bahwa Regresi Logistik memberikan performa terbaik dalam klasifikasi kondisi kesehatan mental mahasiswa dibandingkan dengan model baseline dan K-Nearest Neighbors (KNN). Tingginya nilai akurasi dan F1-score yang diperoleh menunjukkan bahwa model ini mampu menangkap pola hubungan antara variabel prediktor dan status kesehatan mental mahasiswa secara efektif.

Performa Regresi Logistik yang lebih unggul dapat dijelaskan oleh karakteristik model tersebut yang sesuai untuk data dengan variabel biner dan kategorik serta jumlah data yang relatif terbatas. Selain itu, model ini memiliki kemampuan interpretasi yang baik sehingga memudahkan analisis pengaruh masing-masing variabel terhadap probabilitas terjadinya depresi. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa Regresi Logistik masih relevan dan kompetitif dalam klasifikasi kesehatan mental karena menghasilkan prediksi yang stabil dan mudah diinterpretasikan [8].

Jika dibandingkan dengan model baseline, terlihat bahwa baseline hanya “menebak” kelas mayoritas sehingga meskipun menghasilkan akurasi sebesar 0,65, model tersebut tidak mampu mendeteksi satu pun kasus depresi. Hal ini ditunjukkan oleh nilai precision, recall, dan F1-score yang bernilai 0,00. Temuan ini menegaskan bahwa penggunaan baseline hanya berfungsi sebagai acuan performa minimum dan tidak dapat diandalkan untuk tugas klasifikasi kesehatan mental.

Model Regresi Logistik menunjukkan performa yang lebih seimbang dengan akurasi sebesar 0,85. Nilai precision yang mencapai 1,00 mengindikasikan bahwa model sangat berhati-hati dalam memberikan label depresi, sehingga tidak terdapat mahasiswa yang tidak depresi namun salah diklasifikasikan sebagai depresi. Karakteristik ini penting dalam konteks kesehatan mental karena kesalahan positif berpotensi menimbulkan stigma atau intervensi yang tidak diperlukan. Namun demikian, nilai recall yang belum maksimal (0,57) menunjukkan bahwa masih terdapat sebagian mahasiswa depresi yang belum terdeteksi oleh model, sehingga peningkatan sensitivitas model dapat menjadi fokus penelitian selanjutnya.

Sementara itu, model KNN ($k = 5$) menunjukkan performa yang lebih rendah dengan akurasi sebesar 0,55 dan F1-score sebesar 0,31. Kinerja yang kurang optimal ini dapat disebabkan oleh jumlah data yang relatif kecil, distribusi fitur yang tidak terpisah secara jelas antar kelas, serta sensitivitas algoritma KNN terhadap skala data dan kedekatan antar titik data. Hal ini menyebabkan KNN kurang mampu membedakan pola kompleks pada dataset kesehatan mental mahasiswa dibandingkan Regresi Logistik.

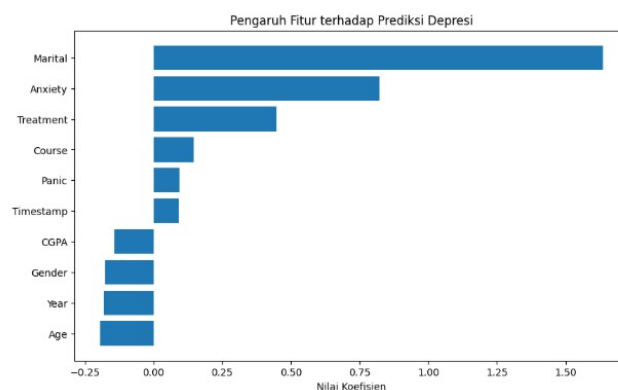
Berdasarkan grafik Koefisien Logistic Regression, terlihat bahwa beberapa variabel memiliki pengaruh yang berbeda terhadap prediksi depresi pada mahasiswa. Variabel *Marital* menunjukkan nilai koefisien positif paling tinggi (sekitar 1,55), yang berarti status pernikahan memiliki pengaruh paling kuat dalam meningkatkan probabilitas mahasiswa terklasifikasi sebagai depresi dalam model.

ini. Semakin tinggi nilai atau kategori yang dikodekan pada variabel tersebut, semakin besar kemungkinan mahasiswa diprediksi mengalami depresi.

Selain itu, variabel *Treatment* dan *Anxiety* juga memiliki koefisien positif yang cukup besar (sekitar 0,77). Hal ini menunjukkan bahwa mahasiswa yang pernah menjalani *treatment* atau memiliki tingkat kecemasan (*anxiety*) lebih tinggi cenderung memiliki peluang lebih besar untuk terklasifikasi sebagai depresi. Variabel *Panic* juga bernilai positif (sekitar 0,28), namun pengaruhnya relatif lebih kecil dibandingkan *Anxiety* dan *Treatment*, sehingga kontribusinya terhadap prediksi depresi tidak sebesar kedua variabel tersebut.

Sebaliknya, beberapa variabel seperti *Gender* (sekitar -0,55) dan *Age* (sekitar -0,48) memiliki koefisien negatif. Ini menunjukkan bahwa peningkatan pada variabel tersebut justru menurunkan probabilitas mahasiswa diprediksi mengalami depresi menurut model. Variabel *CGPA*, *Course*, dan *Year* juga memiliki nilai koefisien yang kecil dan cenderung mendekati nol, sehingga pengaruhnya terhadap prediksi depresi relatif lemah dibandingkan variabel psikologis.

Secara keseluruhan, model menunjukkan bahwa faktor psikologis seperti *Anxiety*, *Panic*, dan riwayat *Treatment* memiliki kontribusi yang lebih dominan dalam proses klasifikasi depresi dibandingkan faktor akademik seperti *CGPA* atau *Year*. Dengan demikian, dapat disimpulkan bahwa variabel psikologis merupakan prediktor yang lebih kuat dalam mendeteksi potensi depresi pada mahasiswa berdasarkan hasil Logistic Regression pada Gambar 4.



Gambar 4. Pengaruh Fitur terhadap Prediksi Depresi

Mahasiswa yang melaporkan adanya kecemasan dan serangan panik memiliki kecenderungan lebih tinggi untuk diklasifikasikan sebagai depresi. Temuan ini konsisten dengan kondisi empiris kesehatan mental mahasiswa di Indonesia, di mana prevalensi gangguan psikologis pada mahasiswa tergolong tinggi. Penelitian sebelumnya melaporkan bahwa 51% mahasiswa di Daerah Istimewa Yogyakarta menunjukkan gejala depresi dan 27,4% mahasiswa di Aceh mengalami gangguan kecemasan [9]. Konsistensi ini memperkuat validitas hasil penelitian serta menegaskan urgensi penerapan model klasifikasi sebagai alat bantu deteksi dini kesehatan mental mahasiswa.

Namun demikian, perlu diperhatikan bahwa ukuran dataset yang relatif kecil (101 data) dapat memengaruhi generalisasi model terhadap populasi mahasiswa yang lebih luas. Dataset dengan jumlah sampel terbatas berpotensi menyebabkan model kurang stabil ketika diterapkan pada data baru. Selain itu, variabel yang digunakan masih terbatas pada faktor demografis dan psikologis dasar, sehingga belum mencakup aspek lingkungan sosial, dukungan keluarga, maupun faktor ekonomi yang dalam beberapa penelitian terbukti turut berkontribusi terhadap kondisi kesehatan mental mahasiswa [10].

Berdasarkan keseluruhan hasil dan analisis tersebut, Regresi Logistik dipilih sebagai model final dalam penelitian ini. Model ini tidak hanya memberikan performa klasifikasi terbaik, tetapi juga menawarkan interpretabilitas yang jelas melalui koefisien regresi, sehingga dapat digunakan untuk memahami pengaruh masing-masing faktor terhadap risiko depresi. Dengan demikian, Regresi Logistik berpotensi menjadi dasar pengembangan sistem skrining awal untuk membantu

institusi pendidikan dalam mengidentifikasi mahasiswa yang berisiko mengalami gangguan kesehatan mental secara lebih efektif dan tepat sasaran.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengklasifikasikan kondisi kesehatan mental mahasiswa, khususnya risiko depresi, dengan menggunakan algoritma Regresi Logistik. Dataset Student Mental Health dari platform Kaggle digunakan sebagai dasar pemodelan, dengan memanfaatkan variabel demografis, akademik, dan psikologis mahasiswa. Proses penelitian meliputi tahap pra-pemrosesan data, pemodelan klasifikasi, serta evaluasi kinerja model menggunakan metrik akurasi, precision, recall, dan F1-score.

Hasil penelitian menunjukkan bahwa algoritma Regresi Logistik memberikan performa terbaik dibandingkan model baseline dan K-Nearest Neighbors (KNN), dengan akurasi sebesar 0,85 dan F1-score sebesar 0,73. Nilai precision yang tinggi menunjukkan bahwa model mampu meminimalkan kesalahan klasifikasi *positif*, sedangkan nilai recall yang belum optimal mengindikasikan adanya beberapa kasus depresi yang belum terdeteksi. Selain itu, analisis fitur menunjukkan bahwa faktor psikologis seperti Marital, Treatment, dan Panic memiliki pengaruh yang signifikan terhadap prediksi depresi pada mahasiswa.

Berdasarkan temuan tersebut, dapat disimpulkan bahwa Regresi Logistik merupakan metode yang efektif dan relevan untuk klasifikasi risiko depresi pada mahasiswa serta berpotensi digunakan sebagai alat bantu deteksi dini kesehatan mental di lingkungan pendidikan tinggi. Meskipun demikian, penelitian ini masih memiliki keterbatasan pada jumlah data yang relatif kecil dan cakupan variabel yang terbatas. Penelitian selanjutnya disarankan untuk menggunakan dataset dengan ukuran yang lebih besar, menambahkan variabel yang lebih beragam, serta mengeksplorasi algoritma pembelajaran mesin lain guna meningkatkan sensitivitas dan akurasi model.

DAFTAR PUSTAKA

- [1] A. N. M. Pudjianto dan E. Y. Hidayat, "Perbandingan Prediksi Depresi Mahasiswa dengan Linear Regression, Random Forest, dan Gradient Boosting," *SINTECH Sci. Inf. Technol. J.*, vol. 7, no. 3, hlm. 180-189, Des 2024, doi: 10.31598/sintechjournal.v7i3.1729.
- [2] Y. Feng, J. Li, T. Liu, Y. Wei, dan N. Li, "Construction of a mental health risk model for college students with long and short-term memory networks and early warning indicators," *J. Intell. Syst.*, vol. 33, no. 1, Jan 2024, doi: 10.1515/jisys-2023-0318.
- [3] M. P. Kayla dan R. A. Saputra, "Klasifikasi Penyakit Gangguan Jiwa menggunakan Metode Logika Fuzzy," *Telematika*, vol. 20, no. 3, hlm. 416, Nov 2023, doi: 10.31315/telematika.v20i3.11789.
- [4] O. S. D. Fadhillah, J. H. Jaman, dan C. Carudin, "Perbandingan Naive Bayes, Support Vector Machine, Logistic Regression Dan Random Forest Dalam Menganalisis Sentimen Mengenai Tiktokshop," *J. Inform. Dan Tek. Elektro Terap.*, vol. 13, no. 1, Jan 2025, doi: 10.23960/jitet.v13i1.5746.
- [5] M. Pulungan, A. Purnomo, dan A. Kurniasih, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 10, hlm. 1493-1502, Des 2023, doi: 10.25126/jtiik.1077989.
- [6] M. Fachriza, "Analisis Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Naive Bayes, Support Vector Machine, Random Forest," *Anal. Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Naive Bayes Support Vector Mach. Random For.*, vol. 0, no. 0, Jul 2024, Diakses: 14 Desember 2025. [Daring]. Tersedia pada: <https://digilib.esaunggul.ac.id/UEU-Undergraduate-20190801251/34308>
- [7] L. Zhang, S. Zhao, Z. Yang, H. Zheng, dan M. Lei, "An artificial intelligence tool to assess the risk of severe mental distress among college students in terms of demographics, eating habits, lifestyles, and sport habits: an externally validated study using machine learning," *BMC Psychiatry*, vol. 24, no. 1, hlm. 581, Agu 2024, doi: 10.1186/s12888-024-06017-2.
- [8] R. Damanhuri dan V. A. Husein, "Analisis Sentimen pada Ulasan Aplikasi Access by KAI Berbahasa Indonesia Menggunakan Word-Embedding dan Classical Machine Learning," *J. Masy. Inform.*, vol. 15, no. 2, hlm. 97-106, Nov 2024, doi: 10.14710/jmasif.15.2.62383.
- [9] A. Shabrina, A. G. Prathama, dan R. H. Ninin, "Persepsi Stigmatisasi Dan Intensi Pencarian Bantuan Kesehatan Mental Pada Mahasiswa S1," *J. Psikol.*, vol. 17, no. 1, hlm. 80, Jun 2021, doi: 10.24014/jp.v17i1.11399.
- [10] "Mental Health and Academic Success in College." Diakses: 20 Februari 2026. [Daring]. Tersedia pada: <https://www.degruyterbrill.com/document/doi/10.2202/1935-1682.2191/html>