

**ANALISIS KOMPARATIF LOGISTIC REGRESSION DAN
RANDOM FOREST DALAM MEMPREDIKSI RISIKO
DEPRESI PADA MAHASISWA**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi SI Informatika



disusun oleh

SEKAR TAJI PALUPI

22.11.4716

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS KOMPARATIF LOGISTIC REGRESSION DAN
RANDOM FOREST DALAM MEMPREDIKSI RISIKO
DEPRESI PADA MAHASISWA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi SI Informatika



disusun oleh

SEKAR TAJI PALUPI

22.11.4716

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS KOMPARATIF LOGISTIC REGRESSION DAN RANDOM
FOREST DALAM MEMPREDIKSI RISIKO DEPRESI PADA
MAHASISWA**

yang disusun dan diajukan oleh

Sekar Taji Palupi

22.11.4716

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Desember 2025

Dosen Pembimbing,



Ferian Fauzi Abdulloh, S.Kom., M.Kom.
NIK. 190302276

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS KOMPARATIF LOGISTIC REGRESSION DAN RANDOM
FOREST DALAM MEMPREDIKSI RISIKO DEPRESI PADA
MAHASISWA

yang disusun dan diajukan oleh

Sekar Taji Palupi

22.11.4716

Telah dipertahankan di depan Dewan Penguji
pada tanggal 12 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Nuri Cahyono, M.Kom.
NIK. 190302278

Ike Verawati, S.Kom., M.Kom.
NIK. 190302237

Ferian Fauzi Abdulloh, S.Kom., M.Kom.
NIK. 190302276



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 12 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Sekar Taji Palupi
NIM : 22.11.4716

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Komparatif Logistic Regression Dan Random Forest Dalam Memprediksi Risiko Depresi Pada Mahasiswa

Dosen Pembimbing : Ferian Fauzi Abdulloh, S.Kom., M.Kom,

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya,
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 12 Desember 2025

Yang Menyatakan,

Sekar Taji Palupi

KATA PENGANTAR

Puji Syukur ke hadirat Tuhan Yang Maha Esa atas segala Rahmat, karunia, dan kemudahan yang diberikan sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Penyusunan skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Penulis menyampaikan terima kasih dan penghargaan yang setulus-tulusnya kepada Bapak Ferian Fauzi Abdulloh, S.Kom., M.Kom., selaku Dosen Pembimbing, yang telah memberikan bimbingan, arahan, dan dukungan selama proses penyusunan skripsi ini. Ucapan terima kasih juga penulis sampaikan kepada Bapak Nuri Cahyono, M.Kom. dan Ibu Ike Verawati, S.Kom., M.Kom., selaku Dosen Penguji, yang telah memberikan masukan, saran, serta kritik yang sangat berharga dalam penyempurnaan skripsi ini.

Rasa terima kasih yang mendalam juga penulis sampaikan kepada kedua orang tua tercinta atas doa, kasih sayang, semangat, dan dukungan moral maupun material yang tiada henti. Penulis juga berterima kasih kepada keluarga, sahabat, dan seluruh pihak yang telah membantu secara langsung maupun tidak langsung dalam proses penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan untuk perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan menjadi kontribusi positif bagi pengembangan ilmu pengetahuan.

Yogyakarta, 11 Desember 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
KATA PENGANTAR	vi
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
INTISARI	xii
<i>ABSTRACT</i>	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	12
2.2.1 Depresi pada Mahasiswa	12
2.3.1 Machine Learning	13
2.2.3 Algoritma	13
2.2.4 Variabel	14
2.2.5 Feature	15

2.2.6	Label (Target)	15
2.2.7	Supervised Learning	16
2.2.8	Klasifikasi	16
2.2.9	Data Training dan data Testing	17
2.2.10	Linear dan Non-linear	18
2.2.11	Logistic Regression	19
2.2.12	Decision Tree	20
2.2.13	Ensemble Learning	21
2.2.14	Random Forest	21
2.2.15	Preprocessing Data	22
2.2.16	Overfitting dan Underfitting	23
2.2.17	Evaluasi Model	24
BAB III METODE PENELITIAN		26
3.1	Objek Penelitian	26
3.2	Alur Penelitian	26
3.3	Alat dan Bahan	29
3.3.1	Data Penelitian	29
3.3.2	Alat dan Instrumen Penelitian	30
BAB IV HASIL DAN PEMBAHASAN		32
4.1.	Exploratory Data Analysis (EDA)	32
4.2	Preprocessing Data	37
4.3	Split Data	38
4.4	Transformasi Variabel Kategorikal (Encoding)	39
4.5	Feature Selection	42
4.6	Model Logistic Regression	45

4.6.1	Baseline Logistic Regression.....	45
4.6.2	Hyperparameter Tuning Logistic Regression.....	47
4.6.3	Treshold Tuning Logistic Regression.....	49
4.6.4	5 Fold Cross Validation Logistic Regression.....	52
4.7	Model Random Forest.....	53
4.7.1	Baseline Random Forest.....	53
4.7.2	Hyperparameter Tuning Random Forest.....	55
4.7.3	Threshold Tuning Random Forest.....	57
4.7.4	Optimalisasi Random Forest.....	59
4.7.5	5 Fold Cross Validation Random Forest.....	61
4.8	Perbandingan Performa Logistic Regression dan Random Forest.....	62
4.9	Hasil Implementasi Model dalam Aplikasi Web.....	64
BAB V PENUTUP.....		66
5.1	Kesimpulan.....	66
5.2	Saran.....	66
REFERENSI.....		68

DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	8
Tabel 4.1. Hasil Cramer'sV	36
Tabel 4.2. Preprocessing Data	37
Tabel 4.3. Pemetaan Encoding Variabel Kategorikal	39
Tabel 4.4. Pengelompokkan Variabel Degree	40
Tabel 4.5. Pemetaan Encoding Variabel Sleep Duration	41
Tabel 4.6. Pemetaan Encoding Variabel Dietary Habits	41
Tabel 4.7. Pemetaan Encoding Variabel Degree	41
Tabel 4.8. Hasil Evaluasi Baseline Logistic Regression	45
Tabel 4.9. Cek Overfitting Baseline Logistic Regression	46
Tabel 4.10. Hyperparameter Grid Logistic Regression	47
Tabel 4.11. hasil Hyperparameter Tuning Logistic Regression	48
Tabel 4.12 Hasil Evaluasi Logistic Regression setelah Hyperparameter Tuning	49
Tabel 4.13 Hasil Evaluasi Logistic Regression setelah Threshold Tuning	50
Tabel 4.14 Hasil Cek Overfitting Logistic Regression (Threshold 0.47)	52
Tabel 4.15 Hasil Evaluasi 5 Fold Cross Validation Logistic Regression	53
Tabel 4.16 Hasil Evaluasi Baseline Random Forest	54
Tabel 4.17 Hasil Cek Overfitting Random Forest	55
Tabel 4.18 Kombinasi Hyperparameter Tuning Random Forest	56
Tabel 4.19 Hasil Evaluasi Hyperparameter Tuning Random Forest	57
Tabel 4.20 Hasil Evaluasi Threshold Tuning Random Forest	58
Tabel 4.21 Hasil Hasil Cek Overfitting Random Forest	59
Tabel 4.22 Hasil Evaluasi Optimalisasi Random Forest	60
Tabel 4.23 Hasil Cek Overfitting Random Forest setelah Optimasi	61
Tabel 4.24 Hasil 5 Fold Cross-Validation Random Forest	61
Tabel 4.25 Perbandingan Hasil Evaluasi Kedua Model	63

DAFTAR GAMBAR

Gambar 3.1. Alur Penelitian	26
Gambar 4.1. Distribusi Depresi	33
Gambar 4.2. Distribusi Gender	33
Gambar 4.3. Distribusi Sleep Duration	34
Gambar 4.4. Distribusi Dietary Habits	34
Gambar 4.5 Distribusi Family History of Mental Illness	35
Gambar 4.6. Distribusi Have you ever had suicidal thoughts?	35
Gambar 4.7. Korelasi Heatmap Antar Variabel	35
Gambar 4.8. Visualisasi Split Data	38
Gambar 4.9. Hasil Feature Importance Random Forest	43
Gambar 4.10. Hasil Feature Importance Logistic Regression	44
Gambar 4.11. Grafik ROC-AUC Logistic Regression	51
Gambar 4.12. Perbandingan Performa LR dan RF	62

INTISARI

Depresi adalah salah satu isu Kesehatan mental yang sering dialami oleh mahasiswa, yang disebabkan oleh tekanan medis, tuntutan prestasi, masalah keuangan, dan gaya hidup yang tidak seimbang. Kondisi ini dapat mengakibatkan penurunan motivasi belajar, gangguan konsentrasi, menurunnya prestasi akademik, serta memengaruhi kesejahteraan psikologis secara keseluruhan. Jika tidak ditangani sejak dini, depresi dapat berkembang menjadi masalah yang lebih serius dan mengganggu kelangsungan studi mahasiswa. Oleh karena itu, perlu ada pendekatan yang dapat membantu mengidentifikasi risiko depresi pada mahasiswa secara lebih awal dan objektif.

Penelitian ini menggunakan pendekatan machine learning untuk memprediksi risiko depresi pada mahasiswa dengan membandingkan dua algoritma klasifikasi, yaitu Logistic Regression dan Random Forest. Tahapan penelitian meliputi pengolahan preprocessing, pembagian dataset menjadi training set dan testing set, pelatihan model, serta evaluasi performa menggunakan metrik *accuracy*, *precision*, *recall*, *f1-score* dan ROC-AUC. Logistic Regression dipilih karena kesederhanaan dan kemudahan interpretasi, sedangkan Random Forest dipilih karena kemampuannya menangani pola data yang lebih kompleks.

Hasil penelitian menunjukkan bahwa Logistic Regression memiliki performa yang lebih unggul dalam mendeteksi depresi mahasiswa, ditunjukkan oleh nilai *recall* yang sangat tinggi yaitu 0.909, serta kemampuan diskriminasi yang lebih baik melalui ROC-AUC 0.922. Sementara itu, Random Forest menunjukkan keunggulan pada *precision* yaitu 0.862 dan menghasilkan *accuracy* yang kompetitif, yaitu 0.834, namun sensitivitasnya lebih rendah dibandingkan Logistic Regression. Pada penelitian ini, Logistic Regression dinilai lebih sesuai untuk dijadikan model utama karena memiliki sensitivitas yang lebih tinggi dan kemampuan diskriminasi yang stabil pada berbagai tahapan pengujian.

Kata kunci: Depresi, Mahasiswa, Machine Learning, Logistic Regression, Random Forest

ABSTRACT

Depression is one of the mental health issues frequently experienced by university students, caused by academic pressure, achievement demands, financial problems, and an unbalanced lifestyle. This condition can lead to decreased learning motivation, impaired concentration, declining academic performance, and overall deterioration of psychological well-being. If not addressed at an early stage, depression may develop into more serious problems and disrupt students' academic continuity. Therefore, an approach is needed to help identify the risk of depression among students in a more objective and early manner.

This study employs a machine learning approach to predict the risk of depression among university students by comparing two classification algorithms, namely Logistic Regression and Random Forest. The research stages include data preprocessing, dataset splitting into training and testing sets, model training, and performance evaluation using accuracy, precision, recall, F1-score, and ROC-AUC metrics. Logistic Regression was selected due to its simplicity and interpretability, while Random Forest was chosen for its ability to handle more complex data patterns.

The results indicate that Logistic Regression demonstrates superior performance in detecting depression risk among students, as reflected by a high recall value of 0.909 and a better discriminative ability with a ROC-AUC of 0.922. Meanwhile, Random Forest shows higher precision, reaching 0.862, and achieves a competitive accuracy of 0.834; however, its sensitivity is lower compared to Logistic Regression. Based on these findings, Logistic Regression is considered more suitable as the primary model due to its higher sensitivity and stable discriminative performance across different evaluation stages.

Keyword: Depression, University Students, Machine Learning, Logistic Regression, Random Forest