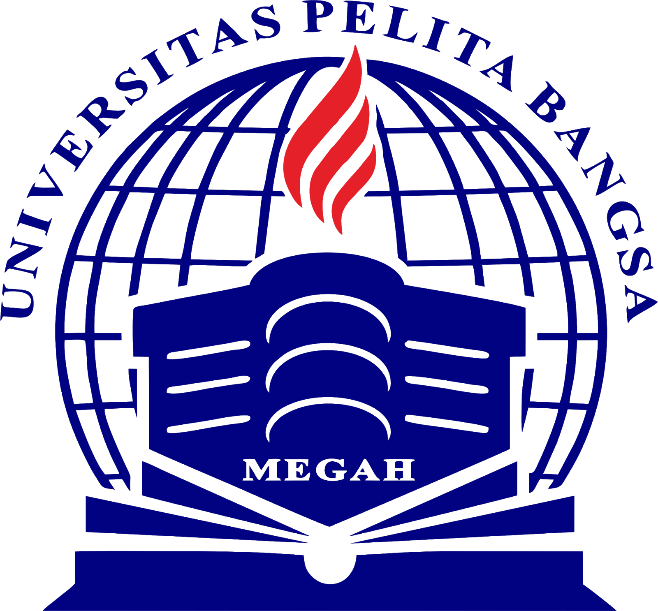
**LAPORAN PROYEK AKHIR PRAKTIKUM DATA MINING 2025**

**Aplikasi Prediksi Stres Mahasiswa Berbasis Streamlit Menggunakan Regresi Linier**

****

Disusun oleh :

1. **Syifa Aurellia Rahma (312210009)**
2. **Ainun Dwi Permana (312310013)**
3. **Arya Wiratama (312310224)**

**PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNIK**

**UNIVERSITAS PELITA BANGSA BEKASI**

**2025**

# LEMBAR PENGESAHAN PROYEK

**LAPORAN PROYEK AKHIR PRAKTIKUM DATA MINING 2025**

**Aplikasi Prediksi Stres Mahasiswa Berbasis Streamlit Menggunakan Regresi Linier**

Laporan proyek akhir ini telah disusun dan diselesaikan oleh:

1. **Syifa Aurellia Rahma (312210009)**
2. **Ainun Dwi Permana (312310013)**
3. **Arya Wiratama (312310224)**

Sebagai salah satu syarat untuk menyelesaikan Praktikum Data Mining pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

|  |  |  |
| --- | --- | --- |
| **Syifa Aurellia Rahma** | **Ainun Dwi Permana** | **Arya Wiratama** |

Laporan ini telah diperiksa dan disetujui untuk disahkan oleh Dosen Pengampu Dosen Pengampu

|  |
| --- |
| **Najamuddin Dwi Miharja, S.Kom., M.Kom.** |
| **NIDN. 0425098802** |

# KATA PENGANTAR

Segala puji dan syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, sehingga kami dapat menyelesaikan laporan proyek akhir praktikum yang berjudul **“Aplikasi Prediksi Stres Mahasiswa Berbasis Streamlit Menggunakan Regresi Linier”** ini dengan baik dan lancar.

Laporan ini disusun sebagai bentuk dokumentasi dan pemahaman kami atas materi yang telah dipelajari dalam **Praktikum Data Mining**, khususnya dalam mengimplementasikan algoritma data mining ke dalam aplikasi berbasis web interaktif menggunakan Streamlit. Melalui proyek ini, kami mendapatkan pengalaman langsung dalam mengolah dataset, menerapkan algoritma regresi, klasifikasi, dan clustering, serta memvisualisasikan hasilnya secara dinamis.

Dalam penyusunan laporan ini, kami menyadari bahwa keberhasilan yang dicapai tidak lepas dari bantuan, dukungan, dan arahan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan terima kasih, kami menyampaikan apresiasi kepada:

1. Bapak Najamudin Dwi Miharja, S.Kom., M.Kom., selaku dosen pengampu praktikum, atas bimbingan, arahan, dan ilmu yang telah diberikan selama proses pembelajaran.
2. Rekan-rekan mahasiswa di kelas Praktikum Data Mining 2025 yang turut berbagi pengetahuan dan pengalaman selama sesi praktikum.
3. Semua pihak yang telah membantu, baik secara langsung maupun tidak langsung, dalam penyusunan laporan dan pengembangan proyek ini.

Kami menyadari bahwa laporan ini masih memiliki keterbatasan. Oleh karena itu, kami sangat terbuka terhadap kritik dan saran yang membangun demi penyempurnaan di masa yang akan datang. Semoga laporan ini dapat memberikan manfaat bagi pembaca serta menjadi referensi tambahan dalam penerapan data mining secara praktis dan aplikatif.

Cikarang, 13 Juni 2025

Kelompok 1

**DAFTAR ISI**

[LEMBAR PENGESAHAN PROYEK i](#_bookmark0)

[KATA PENGANTAR ii](#_bookmark1)

[BAB I 1](#_bookmark2)

* 1. [Latar Belakang 1](#_bookmark3)
  2. [Tujuan 1](#_bookmark4)
  3. [Manfaat 1](#_bookmark5)

[BAB II 3](#_bookmark6)

* 1. [Tinjauan Jurnal Pertama 3](#_bookmark7)
  2. [Tinjauan Jurnal Kedua 3](#_bookmark8)
  3. [Tinjauan Jurnal Ketiga 4](#_bookmark9)

[BAB III 5](#_bookmark10)

* 1. [Arsitektur Aplikasi 5](#_bookmark11)
  2. [Dataset 5](#_bookmark12)
  3. [Alur Sistem 5](#_bookmark13)
  4. [Kode Program 6](#_bookmark14)

[BAB IV 7](#_bookmark15)

* 1. [Tampilan Aplikasi 7](#_bookmark16)
  2. [Evaluasi Model 8](#_bookmark17)
  3. [Visualisasi Prediksi 9](#_bookmark18)
  4. [Koefisien Model 9](#_bookmark19)

[BAB V 11](#_bookmark20)

* 1. [Kesimpulan 11](#_bookmark21)
  2. [Saran 11](#_bookmark22)

[DAFTAR PUSTAKA 12](#_bookmark23)

## Latar Belakang

**BAB I PENDAHULUAN**

Stres merupakan salah satu masalah psikologis yang umum dialami oleh mahasiswa, terutama saat menghadapi tekanan akademik, sosial, maupun tekanan dari kehidupan pribadi. Tingginya tingkat stres yang tidak ditangani dengan baik dapat berdampak negatif terhadap performa akademik dan kesehatan mental mahasiswa. Oleh karena itu, dibutuhkan upaya preventif dan prediktif untuk mengidentifikasi tingkat stres sejak dini.

Kemajuan teknologi informasi dan ilmu data memberikan peluang besar dalam pengembangan aplikasi yang mampu menganalisis data secara cepat dan akurat. Salah satu metode yang dapat digunakan untuk melakukan prediksi adalah algoritma **regresi linier**, yang mampu memodelkan hubungan antara variabel-variabel numerik dan target yang ingin diprediksi, seperti tingkat stres.

Dengan memanfaatkan framework Streamlit, pengembangan aplikasi web berbasis Python dapat dilakukan secara efisien dan interaktif. Aplikasi ini tidak hanya memberikan hasil prediksi secara real-time, tetapi juga memberikan visualisasi data yang memudahkan pengguna dalam memahami hasil analisis. Melalui pengembangan aplikasi prediksi stres berbasis regresi linier ini, diharapkan mahasiswa dan pihak kampus dapat lebih proaktif dalam menangani masalah stres mahasiswa.

## Tujuan

* + 1. Mengembangkan aplikasi berbasis web untuk memprediksi tingkat stres mahasiswa menggunakan algoritma regresi linier.
    2. Menerapkan algoritma regresi linier dalam menganalisis dataset yang berisi faktor- faktor penyebab stres mahasiswa.
    3. Menyediakan visualisasi dan hasil prediksi secara interaktif menggunakan framework Streamlit.

## Manfaat

* + 1. Memberikan sarana prediksi tingkat stres yang mudah diakses dan digunakan oleh mahasiswa maupun pihak kampus.
    2. Membantu meningkatkan kesadaran terhadap kondisi mental mahasiswa melalui pendekatan berbasis data.
    3. Menjadi contoh penerapan praktis algoritma data mining dalam menyelesaikan permasalahan nyata di lingkungan pendidikan.

**BAB II TINJAUAN PUSTAKA**

## Tinjauan Jurnal Pertama

Penelitian yang dilakukan oleh Ainindzi Nur Meiza Pudjianto dan Erwin Yudi Hidayat dalam jurnal berjudul “Perbandingan Prediksi Depresi Mahasiswa dengan Linear Regression, Random Forest, dan Gradient Boosting” bertujuan untuk membandingkan performa tiga model regresi dalam memprediksi tingkat depresi mahasiswa berdasarkan skor kecemasan dan stres dari kuesioner DASS-21. Penelitian ini dilatarbelakangi oleh tingginya tekanan akademik pada mahasiswa, terutama pada mata kuliah menantang seperti Algoritma dan Struktur Data, yang dapat memicu stres dan kecemasan berkelanjutan hingga berkembang menjadi depresi. Dengan melibatkan 206 mahasiswa Teknik Informatika Universitas Dian Nuswantoro sebagai responden, data dikumpulkan melalui kuesioner DASS-21 dan dianalisis menggunakan Multiple Linear Regression, Random Forest Regressor, serta Gradient Boosting Regression.

Hasil evaluasi model menunjukkan bahwa metode Gradient Boosting memiliki performa terbaik dengan nilai MAE sebesar 4.363 dan RMSE sebesar 5.386. Penelitian ini juga mengintegrasikan visualisasi interaktif berbasis Streamlit untuk mendukung eksplorasi data yang lebih mudah. Penelitian ini berkontribusi dalam mengembangkan pendekatan prediktif berbasis machine learning untuk mendeteksi potensi depresi mahasiswa dan dapat dijadikan acuan dalam penyusunan intervensi kesehatan mental di lingkungan perguruan tinggi [1].

## Tinjauan Jurnal Kedua

Penelitian oleh Vionota Oktaviani dkk. dalam jurnal “Perbandingan Kinerja Random Forest dan SMOTE Random Forest dalam Mendeteksi dan Mengukur Tingkat Stres pada Mahasiswa Tingkat Akhir” membahas klasifikasi tingkat stres mahasiswa tingkat akhir dengan membandingkan performa algoritma Random Forest dan SMOTE Random Forest. Data diperoleh dari kuesioner 14 pertanyaan yang ditujukan pada mahasiswa tingkat akhir yang sedang mengerjakan skripsi.

Hasil penelitian menunjukkan bahwa metode Random Forest dengan oversampling SMOTE mampu meningkatkan kinerja klasifikasi, dengan akurasi 71%, precision 72%, dan recall 71%. Penelitian ini menunjukkan bahwa pemilihan algoritma dan teknik

penyeimbangan data sangat berpengaruh terhadap akurasi prediksi tingkat stres mahasiswa, serta dapat menjadi referensi dalam pengembangan aplikasi prediksi stres berbasis machine learning di lingkungan kampus [2].

## Tinjauan Jurnal Ketiga

Penelitian oleh Alfiyan Aldi dkk. dalam jurnal berjudul “Deteksi Dini Tingkat Stres Pada Mahasiswa Menggunakan Metode Iterative Dichotomiser 3 dan K-Nearest Neighbour” mengkaji deteksi dini tingkat stres mahasiswa dengan menggunakan dua metode klasifikasi, yaitu ID3 dan K-Nearest Neighbour (KNN). Penelitian ini menggunakan data dari 300 mahasiswa yang dikumpulkan melalui kuesioner dan wawancara pakar.

Hasil penelitian menunjukkan bahwa kedua metode mampu mengklasifikasikan tingkat stres dengan baik, namun KNN memberikan hasil yang lebih akurat dibandingkan ID3. Penelitian ini menegaskan potensi metode klasifikasi dalam data mining untuk deteksi stres mahasiswa secara dini dan dapat menjadi alternatif algoritma selain regresi linier dalam pengembangan aplikasi prediksi stres berbasis machine learning [3].

# BAB III

**IMPLEMENTASI**

## Arsitektur Aplikasi

Aplikasi ini dibangun menggunakan bahasa pemrograman Python dan framework Streamlit sebagai antarmuka pengguna berbasis web. Untuk kebutuhan prediksi, digunakan algoritma **Regresi Linier** dari pustaka scikit-learn, sedangkan visualisasi dilakukan menggunakan pustaka matplotlib. Aplikasi bersifat interaktif dan dirancang untuk mengolah file dataset .csv yang diunggah pengguna secara langsung.

## Dataset

Dataset yang digunakan diambil dari platform Kaggle dan berisi data mahasiswa dengan berbagai atribut seperti:

* + - **Sleep Duration** (durasi tidur)
    - **Quality of Sleep** (kualitas tidur)
    - **Physical Activity Level** (tingkat aktivitas fisik)
    - **Age** (usia)
    - **Gender**
    - **Stress Level** (target prediksi)

Dari data tersebut, aplikasi akan secara otomatis memilih semua fitur bertipe numerik sebagai variabel input dan menggunakan kolom stress\_level sebagai variabel target (output). Dataset diunggah melalui fitur file\_uploader pada Streamlit.

## Alur Sistem

1. Pengguna mengunggah dataset dalam format .csv
2. Aplikasi menampilkan data awal dalam bentuk tabel
3. Aplikasi melakukan pemisahan antara fitur dan target (fitur: semua kolom numerik; target: stress\_level).
4. Data dibagi menjadi dua: data latih dan data uji dengan rasio 80:20.
5. Model Regresi Linier dilatih menggunakan data latih.
6. Aplikasi menampilkan hasil evaluasi model berupa MSE dan R-squared.
7. Hasil prediksi divisualisasikan dalam bentuk scatter plot.
8. Aplikasi menampilkan nilai koefisien regresi dan intercept.

## Kode Program

Berikut adalah cuplikan kode program utama:

model = LinearRegression() model.fit(X\_train, y\_train)

...

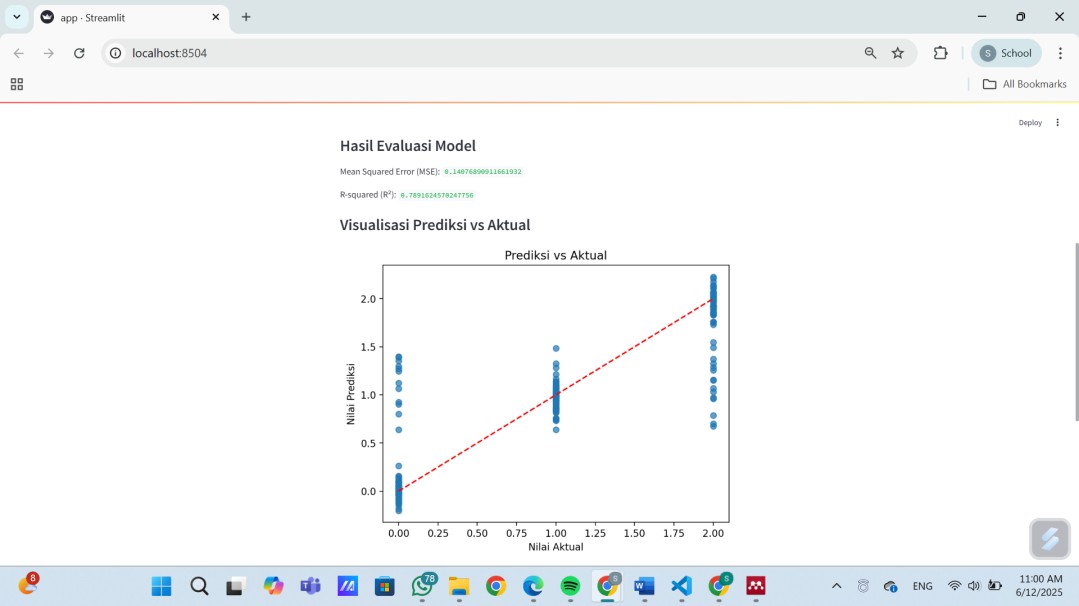
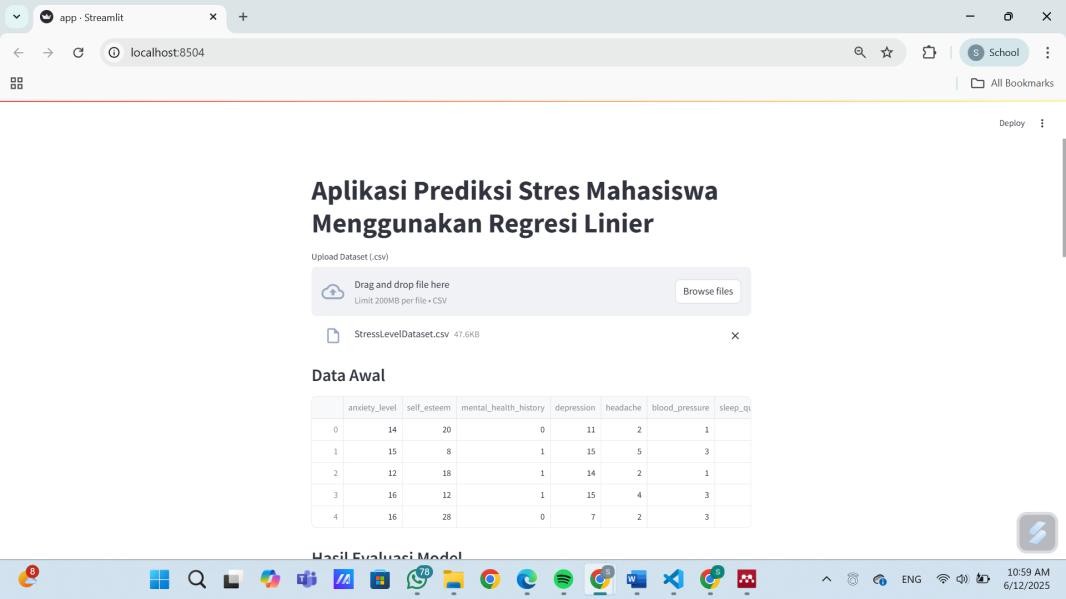
st.write("R-squared (R²):", r2\_score(y\_test, y\_pred))

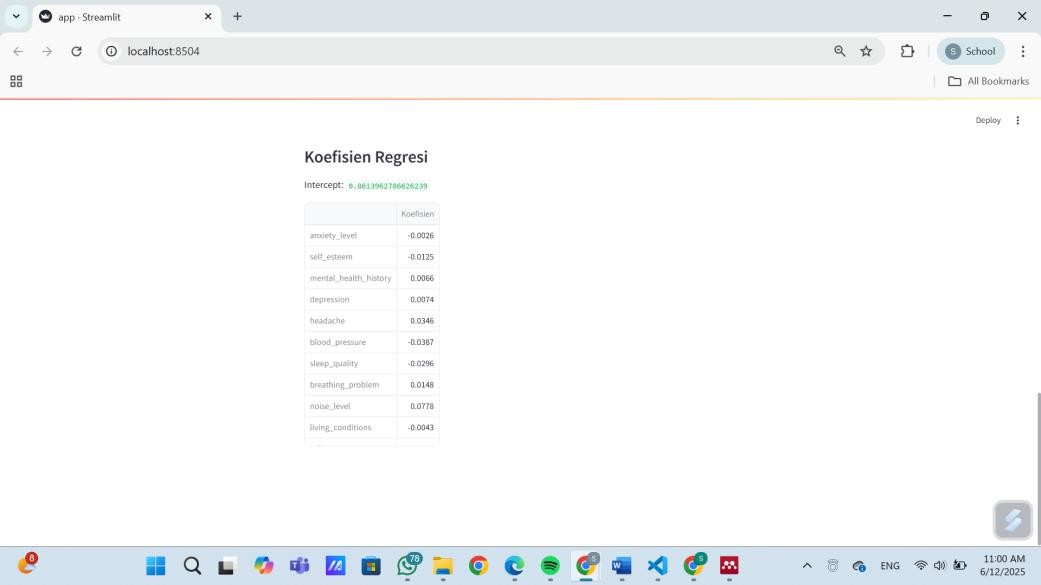
# BAB IV

**HASIL DAN PEMBAHASAN**

## Tampilan Aplikasi

Setelah pengguna mengunggah dataset melalui Streamlit, aplikasi secara otomatis menampilkan preview data awal dan nama-nama kolom yang tersedia. Sistem kemudian memproses seluruh fitur numerik sebagai masukan untuk prediksi tingkat stres mahasiswa (stress\_level). dijalankan dan dataset diunggah, aplikasi akan menampilkan data awal dalam bentuk tabel, serta menampilkan dua fitur utama: jam\_belajar dan jam\_tidur, yang menjadi input prediksi terhadap variabel target, yaitu tingkat\_stres. Berikut adalah output atau tampilan dari Aplikasi Prediksi Stress Mahasiswa Menggunakan Regresi Linier:





## Evaluasi Model

Model Regresi Linier dievaluasi menggunakan dua metrik utama:

* + - **Mean Squared Error (MSE)**: mengukur rata-rata kuadrat selisih antara nilai aktual dan prediksi.
    - **R-squared (R²)**: mengukur seberapa besar variasi target (stress\_level) dapat dijelaskan oleh fitur-fitur yang digunakan.

Contoh hasil evaluasi:

* + - **MSE**: 0.14076890911661932
    - **R²**: 0.7891624570247756

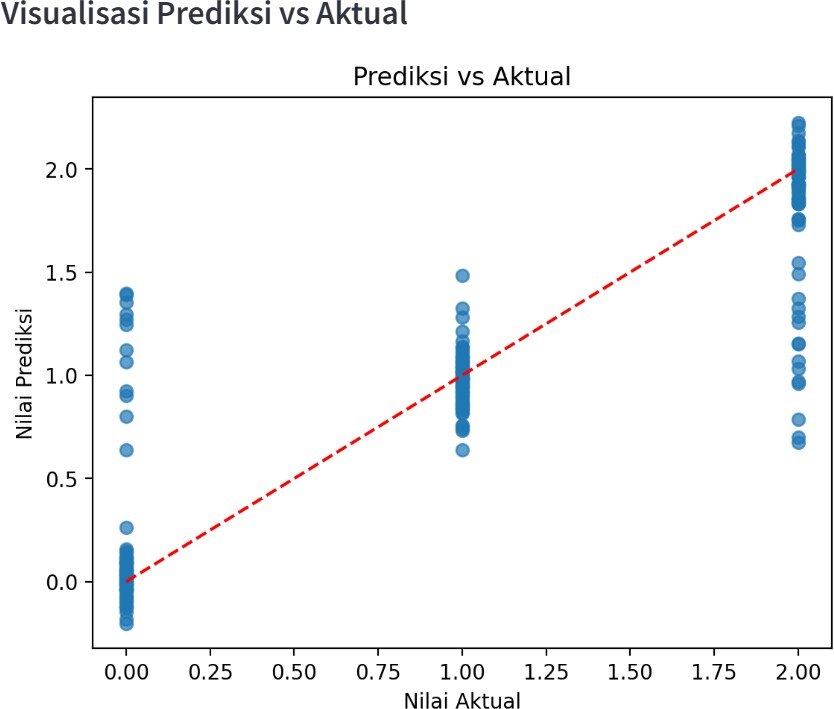
*(Nilai ini tergantung dataset yang digunakan dan dapat bervariasi)*

**

Nilai R² mendekati 1 menunjukkan model memiliki kemampuan yang cukup baik dalam menjelaskan variasi tingkat stres berdasarkan fitur-fitur input.

## Visualisasi Prediksi

Aplikasi menampilkan grafik scatter antara nilai aktual (y\_test) dan prediksi (y\_pred). Garis diagonal merah (y = x) digunakan sebagai acuan prediksi sempurna. Titik-titik yang dekat dengan garis tersebut menunjukkan tingkat akurasi prediksi yang tinggi.



## Koefisien Model

Hasil regresi menampilkan nilai intercept dan koefisien untuk masing-masing fitur. Koefisien menunjukkan pengaruh masing-masing fitur terhadap nilai prediksi stress\_level. Sebagai contoh:

* + - Intercept: 4.72 0.8613962786626239
    - Koefisien:
      * anxiety\_level: -0.0026
      * self\_esteem: -0.0125
      * mental\_health\_history: 0.0066
      * depression: 0.0074
      * headache: 0.0346
      * blood\_pressure: -0.0387
      * sleep\_quality: -0.0296
      * breathing\_problem: 0.0148
      * noise\_level: 0.0778
      * living\_conditions: -0.0043
      * safety: -0.0507
      * basic\_needs: -0.0640
      * academic\_performance: -0.0428
      * study\_load: 0.0542
      * teacher\_student\_relationship: 0.0080
      * future\_career\_concerns: 0.0111
      * social\_support: -0.0289
      * peer\_pressure: 0.0195
      * extracurricular\_activities: 0.0693
      * bullying: 0.0690



Artinya, semakin tinggi durasi tidur, kualitas tidur, dan aktivitas fisik, tingkat stres mahasiswa cenderung menurun (nilai negatif menunjukkan hubungan berbanding terbalik).

**BAB V PENUTUP**

## Kesimpulan

Berdasarkan hasil implementasi aplikasi prediksi tingkat stres mahasiswa, dapat disimpulkan bahwa:

* + 1. Aplikasi berhasil melakukan pemrosesan data, pelatihan model regresi linier, serta menampilkan hasil evaluasi dan visualisasi dengan baik.
    2. Model menunjukkan performa prediksi yang cukup baik berdasarkan metrik MSE dan R².
    3. Antarmuka interaktif berbasis Streamlit sangat memudahkan pengguna dalam menjalankan prediksi tanpa harus memahami pemrograman secara mendalam.

## Saran

* + 1. Untuk meningkatkan akurasi, disarankan menambahkan fitur relevan lain seperti beban tugas akademik, kondisi sosial, atau asupan nutrisi.
    2. Model dapat disempurnakan dengan melakukan hyperparameter tuning, teknik validasi silang, atau eksplorasi model lain seperti Random Forest atau Gradient Boosting.
    3. Aplikasi dapat dikembangkan lebih lanjut dengan menambahkan fitur simpan data prediksi ke database atau pelaporan visual untuk keperluan konseling mahasiswa di kampus.

# DAFTAR PUSTAKA

1. A. Nur, M. Pudjianto, and E. Y. Hidayat, “Perbandingan Prediksi Depresi Mahasiswa dengan Linear Regression , Random Forest , dan Gradient Boosting,” vol. 7, no. 3, pp. 180–189, 2024.
2. V. Oktaviani, N. Rosmawarni, and M. P. Muslim, “Perbandingan Kinerja Random Forest Dan Smote Random Forest Dalam Mendeteksi Dan Mengukur Tingkat Stres Pada Mahasiswa Tingkat Akhir,” *Inform. J. Ilmu Komput.*, vol. 20, no. 1, pp. 43–49, 2024, doi: 10.52958/iftk.v20i1.9158.
3. A. Aldi, S. R. C. Nursari, and F. Maspiyanti, “Deteksi Dini Tingkat Stres Pada Mahasiswa Menggunakan Metode Iterative Dichotomiser 3 dan K-Nearest Neighbour,”

*J. Informatics Adv. Comput.*, vol. 1, no. 1, pp. 1–7, 2020.