

## Perbandingan Prediksi Depresi Mahasiswa dengan Linear Regression, Random Forest, dan Gradient Boosting

Ainindzi Nur Meiza Pudjianto<sup>1</sup>, Erwin Yudi Hidayat<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro  
Jl. Imam Bonjol No.207, Pendrikan Kidul, Kec. Semarang Tengah, Kota Semarang, Jawa Tengah 50131,  
Indonesia

e-mail: 111202113509@mhs.dinus.ac.id<sup>1</sup>, erwin@dsn.dinus.ac.id<sup>2</sup>

Received : November, 2024

Accepted : December, 2024

Published : December, 2024

### Abstract

*Depression in university students is an important concern in higher education as it can affect academic performance and well-being. Academic pressure in challenging courses, such as Algorithms and Data Structures, often triggers stress and anxiety that can develop into depression. This study aims to analyze the relationship between anxiety and stress scores and depression levels using data from 206 undergraduate Informatics Engineering students at Dian Nuswantoro University collected through the DASS-21 questionnaire. This study aims to apply three prediction models namely Multiple Linear Regression, Random Forest Regressor, and Gradient Boosting Regression to predict the level of depression based on anxiety and stress scores using the DASS-21 questionnaire. . Three prediction models, namely Multiple Linear Regression, Random Forest Regressor, and Gradient Boosting Regression, were used to evaluate the prediction performance of depression level based on anxiety and stress scores. Gradient Boosting Regression showed the best performance with MAE of 4.363 and RMSE of 5.396. This study also conducted additional explorations based on generation, gender, and student living status to provide a more comprehensive understanding. The analysis results were visualized through an interactive application based on Streamlit to facilitate data exploration.*

**Keywords:** prediction, depression, gradient increasing regression

### Abstrak

*Depresi pada mahasiswa menjadi perhatian penting di kalangan pendidikan tinggi karena dapat memengaruhi performa akademik dan kesejahteraan. Tekanan akademis pada mata kuliah yang menantang, seperti Algoritma dan Struktur Data, sering memicu stres dan kecemasan yang dapat berkembang menjadi depresi. Penelitian ini bertujuan untuk menganalisis hubungan antara skor kecemasan dan stres terhadap tingkat depresi menggunakan data 206 mahasiswa Teknik Informatika S1 Universitas Dian Nuswantoro yang dikumpulkan melalui kuesioner DASS-21. Penelitian ini bertujuan untuk mengaplikasikan tiga model prediksi yaitu Multiple Linear Regression, Random Forest Regressor, dan Gradient Boosting Regression untuk memprediksi tingkat depresi berdasarkan skor kecemasan dan stres menggunakan kuesioner DASS-21. . Tiga model prediksi, yaitu Multiple Linear Regression, Random Forest Regressor, dan Gradient Boosting Regression, digunakan untuk mengevaluasi performa prediksi tingkat depresi berdasarkan skor kecemasan dan stres. Gradient Boosting Regression menunjukkan performa terbaik dengan MAE sebesar 4,363 dan RMSE sebesar 5,396. Penelitian ini juga melakukan eksplorasi tambahan berdasarkan angkatan, jenis kelamin, dan status tinggal mahasiswa untuk memberikan pemahaman yang lebih komprehensif. Hasil analisis divisualisasikan melalui aplikasi interaktif berbasis Streamlit untuk mempermudah eksplorasi data.*

**Kata Kunci:** prediksi, depresi, gradient boosting regression

## 1. PENDAHULUAN

Kesehatan mental kini menjadi perhatian utama di Indonesia, terutama terkait gangguan depresi. Riset Kesehatan Dasar (Riskesdas) 2018 menunjukkan bahwa 6,2% remaja pada umur 15-24 tahun dilanda gejala depresi [1]. Fase peralihan remaja dari jenjang SMA ke universitas turut menjadi periode yang penuh tekanan bagi remaja. Pada fase ini, mahasiswa dihadapkan pada berbagai tantangan, di mana mereka harus menyesuaikan diri dengan tuntutan akademis yang tinggi, meningkatkan tanggung jawab pribadi, serta menghadapi tekanan sosial yang semakin kompleks. Berbagai tekanan ini sering kali menjadi pemicu munculnya stres, kecemasan, dan bahkan depresi pada sebagian mahasiswa. Jika kecemasan terus berlanjut dalam waktu berkepanjangan, maka bisa menyebabkan stres kronis, yang pada gilirannya dapat memengaruhi kesehatan mental secara keseluruhan [2]. Ketika stres berlangsung secara berkelanjutan dan tidak tertangani, gangguan stres tersebut dapat berkembang menjadi salah satu bentuk depresi [3]. Bisa disimpulkan bahwa stres dan kecemasan yang berlangsung secara terus-menerus memiliki pengaruh signifikan terhadap munculnya depresi.

Pada Universitas Dian Nuswantoro (UDINUS), program studi Teknik Informatika Strata-1 (S1) adalah program studi dengan jumlah mahasiswa mencapai 608 orang pada tahun 2023 [4]. Dengan populasi yang paling besar, Teknik Informatika-S1 menjadi objek ideal untuk penelitian ini. Mahasiswa Teknik Informatika-S1 wajib mengambil mata kuliah Algoritma dan Struktur Data. Mahasiswa dituntut untuk mempunyai landasan berpikir secara logis dan sistematis, serta mengembangkan kemampuan dalam memecahkan masalah. Pemrograman dianggap sebagai salah satu mata kuliah yang menantang karena mengharuskan mahasiswa menguasai beberapa keterampilan sekaligus, seperti logika, algoritma, dan kemampuan pemecahan masalah menggunakan pendekatan pemrograman [5]. Tantangan akademis ini dapat menjadi salah satu sumber tekanan yang signifikan.

*Depression Anxiety Stress Scales* (DASS) ialah alat ukur yang umum digunakan untuk menilai tingkat depresi, kecemasan, dan stres.

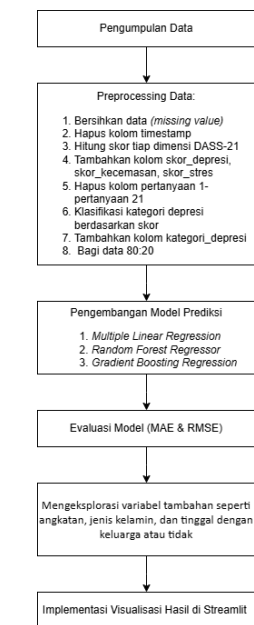
Alat ini dikembangkan oleh Lovibond dan Lovibond pada tahun 1995 dengan tujuan mengukur kondisi emosional negatif secara terpisah [6]. DASS tersedia dalam dua versi, yaitu DASS-21 dan DASS-42. DASS-21 adalah versi singkat yang terdiri dari 21 item, dengan masing-masing 7 item untuk mengukur depresi, kecemasan, dan stres [7]. Alat ini telah terbukti andal di berbagai wilayah, sehingga menjadi standar yang diakui secara internasional, yang mencakup 21 item. DASS-21 dirancang untuk mengatasi keterbatasan alat ukur emosional sebelumnya, dengan psikometrik yang baik dan kemampuan untuk mengidentifikasi masalah kesehatan mental di kalangan mahasiswa [8]. Penelitian [9] menggunakan DASS-42 dan algoritma *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) untuk memprediksi depresi, kecemasan, dan stres berdasarkan variabel psikologis dan demografis. Meskipun pendekatan dan metodenya berbeda, penelitian tersebut menunjukkan relevansi penggunaan DASS sebagai alat ukur kesehatan mental.

Penelitian berikut menunjukkan efektivitas berbagai model prediksi depresi namun dengan fokus dan konteks yang beragam, seperti penelitian [10] menunjukkan bahwa metode *Linear Regression* efektif untuk memprediksi tingkat depresi mahasiswa, dengan mayoritas mahasiswa mengalami depresi ringan. Hasilnya menunjukkan model memiliki akurasi yang cukup baik (MAE 3,84; RMSE 4,7). Penelitian ini menyoroti pentingnya metode prediksi untuk memahami faktor risiko depresi di lingkungan akademik. Sementara itu, metode *Random Forest Regressor* telah berhasil mencapai akurasi tinggi sebesar 80,02%. dalam mendiagnosis gangguan kejiwaan seperti depresi [11]. Selain itu, metode *Gradient Boosting Regression* juga menunjukkan performa terbaik dengan akurasi mencapai 81,17%, dalam prediksi kesehatan mental di lingkungan kerja, menjadikannya metode yang andal dalam aplikasi prediksi kesehatan mental [12]. Dari berbagai penelitian tersebut, terlihat bahwa masing-masing metode memiliki keunggulan dan keterbatasan tergantung pada konteks penerapannya. Oleh karena itu, penelitian ini bertujuan untuk membandingkan ketiga metode prediksi utama, yaitu *Linear Regression*, *Random Forest Regressor*, dan *Gradient Boosting Regression*, dalam satu studi.

Bisa disimpulkan bahwa belum ada kajian yang secara eksplisit membandingkan model regresi seperti *Multiple Linear Regression*, *Random Forest Regressor*, dan *Gradient Boosting Regression* dalam memprediksi tingkat depresi mahasiswa menggunakan DASS-21. Integrasi analisis hasil prediksi dengan aplikasi interaktif berbasis teknologi, seperti Streamlit. Penelitian ini menggunakan pendekatan dengan menerapkan tiga model prediksi, yaitu *Multiple Linear Regression*, *Random Forest Regressor*, dan *Gradient Boosting Regression*, untuk memprediksi tingkat depresi mahasiswa Teknik Informatika S1 UDINUS pada mata kuliah Algoritma dan Pemrograman berdasarkan skor kecemasan dan stres dari DASS-21. Penelitian ini

## 2. METODE PENELITIAN

Adapun tujuan dalam penelitian ini yaitu menganalisis dan memprediksi tingkat depresi. Proses penelitian melibatkan beberapa tahapan utama, yaitu: studi literatur, pengumpulan data, *preprocessing* data, pengembangan model prediksi, evaluasi model, dan implementasi visualisasi hasil di Streamlit. Gambar 1 merupakan uraian dari tahapan dalam metodologi penelitian:



Gambar 1. Tahap Penelitian

### 2.1 Pengumpulan Data

Penulis menyebarkan kuesioner secara *online* kepada mahasiswa Teknik Informatika-S1 UDINUS untuk pengumpulan data. Pengambilan data dilakukan pada 3 hingga 17 Oktober 2024

tidak hanya membandingkan performa model berdasarkan metrik *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE), tetapi juga mengeksplorasi variabel tambahan, seperti jenis kelamin, angkatan, dan status tempat tinggal (tinggal dengan keluarga atau tidak). Hasil penelitian diimplementasikan dalam aplikasi interaktif berbasis Streamlit untuk visualisasi data dan eksplorasi hasil prediksi secara mudah.

Dengan pendekatan ini, penelitian diharapkan dapat memberikan kontribusi signifikan dalam pengembangan model prediksi tingkat depresi mahasiswa, sekaligus menawarkan solusi praktis melalui integrasi teknologi modern dalam pengelolaan kesehatan mental di lingkungan pendidikan tinggi.

yang berjumlah 206 baris. Berikut informasi Dataset ditunjukkan pada Tabel 1.

Tabel 1: Informasi Dataset

No	Kolom	Jumlah Data	Tipe Data
1	timestamp	206	Object
2	angkatan	206	Object
3	jenis_kelamin	202	Object
4	tinggal_dengan_keluarga	206	Object
5	pertanyaan1	206	Int64
6	pertanyaan2	206	Int64
7	pertanyaan3	206	Int64
8	pertanyaan4	206	Int64
9	pertanyaan5	206	Int64
10	pertanyaan6	206	Int64
11	pertanyaan7	206	Int64
12	pertanyaan8	206	Int64
13	pertanyaan9	206	Int64
14	pertanyaan10	206	Int64
15	pertanyaan11	206	Int64
16	pertanyaan12	206	Int64
17	pertanyaan13	206	Int64
18	pertanyaan14	206	Int64
19	pertanyaan15	206	Int64
20	pertanyaan16	206	Int64
21	pertanyaan17	206	Int64
22	pertanyaan18	206	Int64
23	pertanyaan19	206	Int64
24	pertanyaan20	206	Int64
25	pertanyaan11	206	Int64

Kuesioner terdiri dari 21 pertanyaan yang dikelompokkan menjadi 3 dimensi utama, yakni depresi, kecemasan, dan stres yang diambil dari pertanyaan kuesioner DASS-21

yang sudah diterjemahkan ke Bahasa Indonesia dan di sesuaikan yang disesuaikan dengan konteks mata kuliah Algoritma dan Struktur Data. Setiap dimensi memiliki 7 pertanyaan untuk mengukur frekuensi gejala-gejala spesifik yang dialami oleh responden ketika berpartisipasi pada mata kuliah Algoritma dan Struktur Data. Setiap pertanyaan dijawab oleh responden menggunakan skala yang ditunjukkan pada Tabel 2.

Tabel 2: Skala Jawaban

No	Skala	Keterangan
1	0	Tidak Pernah
2	1	Kadang-Kadang
3	2	Cukup Sering
4	3	Sangat Sering

Variabel tambahan seperti angkatan, jenis kelamin, dan status tinggal dengan keluarga juga dikumpulkan untuk memperkaya analisis.

## 2.2 Preprocessing Data

Setelah data sudah terkumpul, maka data dari kuesioner akan diolah lebih lanjut. Langkah-langkah preprocessing meliputi:

1. Pembersihan data yang terdapat *missing value*.
2. Menghapus kolom timestamp.
3. Setiap dimensi dihitung total skornya, yang ditunjukkan pada Tabel 3.

Tabel 3: Dimensi dan Kolom Pertanyaan

No	Dimensi	Kolom Pertanyaan
1	Depresi	pertanyaan3, pertanyaan5, pertanyaan10, pertanyaan13, pertanyaan16, pertanyaan17, dan pertanyaan21
2	Kecemasan	pertanyaan2, pertanyaan4, pertanyaan7, pertanyaan9, pertanyaan15, pertanyaan19, dan pertanyaan20
3	Stres	pertanyaan1, pertanyaan6, pertanyaan8, pertanyaan11, pertanyaan12, pertanyaan14, dan pertanyaan18.

4. Selanjutnya, total skor dari setiap dimensi diperoleh dengan menjumlahkan skor setiap butirnya, kemudian mengalikan hasilnya dengan dua sesuai standar DASS-21. Lalu menambahkan kolom skor\_depresi, skor\_kecemasan, dan

skor\_stres. Berikut ditunjukkan pada Tabel 4

Tabel 4: Aturan Menghitung Total Skor

No	Kolom	Cara Mengitung
1	skor_depresi	(pertanyaan3 + pertanyaan5 + pertanyaan10 + pertanyaan13 + pertanyaan16 + pertanyaan17 + pertanyaan21) × 2
2	skor_kecemasan	(pertanyaan2 + pertanyaan4 + pertanyaan7 + pertanyaan9 + pertanyaan15 + pertanyaan19 + pertanyaan20) × 2
3	skor_stres	(pertanyaan1 + pertanyaan6 + pertanyaan8 + pertanyaan11 + pertanyaan12 + pertanyaan14 + pertanyaan18) × 2

5. Kolom pertanyaan1 sampai pertanyaan21 dihapus.
6. Mengelompokkan kategori depresi sesuai dengan panduan DASS-21, yang mencakup tingkat normal, ringan, sedang, parah, hingga sangat parah.
7. Menambahkan kolom kategori\_depresi.
8. Data dibagi menjadi dua subset, yakni dengan perbandingan 80:20

## 2.3 Pengembangan Model Prediksi

### A. Multiple Linear Regression

*Linear regression* atau regresi linier dibagi menjadi dua jenis. Model regresi linier sederhana digunakan dalam mengidentifikasi hubungan satu variabel independen dengan satu variabel dependen [13]. Berikut persamaan modelnya:

$$Y = a + bX \quad (1)$$

Pada persamaan (1),  $Y$  merepresentasikan variabel hasil, sementara  $a$  merupakan titik potong garis pada sumbu  $y$ . Konstanta pengali dari variabel penentu dilambangkan dengan  $b$  dan  $X$  menunjukkan variabel bebas yang berperan dalam

mempengaruhi nilai  $Y$ . Regresi linier berganda dipilih karena kesederhanaannya dalam menginterpretasikan hubungan linier antara variabel independen dan dependen [14]. Berikut persamaan modelnya:

$$Y = b_0 + b_1 X_1 + \dots + b_n X_n \quad (2)$$

Persamaan (2),  $Y$  mewakili variabel dependen, sedangkan  $X$  merupakan variabel independen. Nilai  $b_0$  adalah intercept atau titik potong pada sumbu  $Y$ , dan  $b$  melambangkan koefisien untuk variabel independen.

### B. Random Forest Regressor

Berdasarkan studi [15], *random forest regressor* adalah algoritma regresi yang memanfaatkan teknik penggabungan prediksi dari beberapa model untuk mencapai hasil prediksi dengan akurasi yang lebih baik dibandingkan model tunggal. Dipilih karena keunggulannya dalam menangani hubungan non-linier serta kemampuannya yang *robust* terhadap *overfitting* [16]. Persamaan yang digunakan dalam *random forest regressor* adalah sebagai berikut:

$$\hat{y}_i = \frac{1}{N_{tree}} \sum_{n=1}^{N_{tree}} \hat{y}_n \quad (3)$$

Dalam persamaan  $\hat{y}_i$  menunjukkan hasil prediksi akhir, yang diperoleh dari rata-rata hasil prediksi semua pohon. Nilai  $N_{tree}$  merepresentasikan jumlah total pohon dalam model, sedangkan  $\hat{y}_n$  adalah hasil prediksi yang dihasilkan oleh pohon ke- $n$ .

### C. Gradient Boosting Regression

*Gradient boosting regression* yaitu salah satu dari algoritma *ensemble* yang bertujuan untuk meningkatkan akurasi prediksi suatu nilai [17]. Model ini memiliki kemampuan untuk menangani pola yang kompleks dalam data serta sering memberikan akurasi tinggi dengan pendekatan *boosting* yang iteratif, sehingga dipilih untuk mengevaluasi performa prediksi pada dataset ini [18][17]. Persamaan yang

digunakan dalam *gradient boosting regression* adalah sebagai berikut:

$$r_{im} = - \left[ \frac{\partial L(y_i, F(x_1))}{\partial F(x_1)} \right]_{F(x)=F_{m-1}(x)} \quad (3)$$

untuk  $i = 1, \dots, n$

## 2.4 Evaluasi Model

Pengujian akurasi model regresi penting dilakukan untuk mengevaluasi sejauh mana model dapat memberikan prediksi yang tepat serta memahami kinerja model dalam konteks data yang ada. Dalam penelitian ini, akurasi model diukur menggunakan dua metrik utama, yaitu MAE dan RMSE. Kedua metrik ini memberikan gambaran tentang tingkat kesalahan prediksi model. MAE menunjukkan rata-rata kesalahan antara *actual value* serta *predicted value*, sering digunakan dalam analisis data deret waktu. Sementara itu, RMSE mengukur tingkat kesalahan secara keseluruhan. Semakin rendah nilai MAE dan RMSE, semakin baik akurasi prediksinya [19].

$$MAE = \sum \frac{|y' - y|}{n} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum (x - y)^2}{n}} \quad (5)$$

Dimana:

$y'$  : nilai prediksi dari model regresi,  
 $y$  : nilai aktual dari variabel dependen,  
 $n$  : jumlah total observasi atau data.

## 2.5 Implementasi Streamlit

Streamlit, sebuah platform berbasis Python yang bersifat *open-source*, yang dapat membuat aplikasi web interaktif yang dirancang untuk mendukung proyek *data science* serta *machine learning* [20]. Pada tahap ini, web dengan *framework* Streamlit digunakan untuk membangun antarmuka pengguna yang interaktif dan menampilkan hasil prediksi dari ketiga model, melihat hasil evaluasi performa model, menampilkan distribusi kategori depresi antara data aktual dan prediksi, serta analisis berdasarkan faktor demografis menggunakan pustaka Plotly untuk memberikan gambaran yang jelas sehingga memudahkan proses interpretasi data.

## 3. HASIL DAN PEMBAHASAN

Penelitian ini memanfaatkan pustaka Python untuk mengimpor dataset,

preprocessing data, melakukan analisis dan visualisasi, serta mempersiapkan dan membangun model prediksi. Setelah dataset diperoleh, langkah pertama adalah preprocessing data, yang mencakup pembersihan *missing values*, menghapus kolom Timestamp, menghitung total skor setiap dimensi (menambah kolom skor\_depresi skor\_kecemasan, dan skor\_stres), mengklasifikasikan kategori depresi (menambah kolom kategori\_depresi) dan menghapus kolom Pertanyaan1 sampai Pertanyaan21. Pada Tabel 5 dibawah menunjukkan data yang telah dilakukan proses *preprocessing*.

Tabel 5: Dataset setelah Preprocessing

No	Column	Jumlah Data	Dtype
1	angkatan	202	Object
2	jenis_kelamin	202	Object
3	tinggal_dengan_keluarga	202	Object
4	skor_depresi	202	Int64
5	skor_kecemasan	202	Int64
6	skor_stres	202	Int64
7	kategori_depresi	202	Int64

Langkah selanjutnya adalah mempersiapkan dataset untuk pelatihan dan pengujian, dengan perbandingan 80:20. Tahap selanjutnya adalah penerapan ketiga metode yang dilatih untuk memprediksi skor\_depresi berdasarkan skor\_stres dan skor\_kecemasan. Kemudian model diuji menggunakan data pengujian untuk menghasilkan prediksi. Evaluasi performa, di mana semakin minim nilai MAE dan RMSE, maka semakin baik akurasi. Tabel 6 merupakan hasil perbandingan ketiga model.

Tabel 6: Evaluasi Model

No	Metode	MAE	RMSE
1	Multiple Linear Regression	5.256242	6.147845
2	Random Forest Regressor	4.781806	6.020678
3	Gradient Boosting Regression	4.363450	5.386264

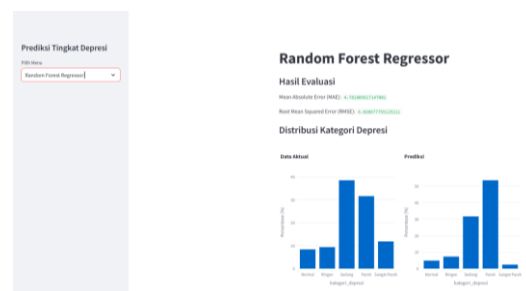
Berdasarkan hasil evaluasi pada Tabel 6, *Gradient Boosting Regression* memiliki performa terbaik dibandingkan model lainnya dengan nilai MAE sebesar 4.363 dan RMSE sebesar 5.386, lebih rendah dibandingkan

Multiple Linear Regression (MAE: 5.256, RMSE: 6.147) dan Random Forest Regressor (MAE: 4.781, RMSE: 5.967). Nilai MAE mencerminkan rata-rata kesalahan prediksi, sedangkan RMSE menunjukkan tingkat penyimpangan prediksi yang lebih besar.

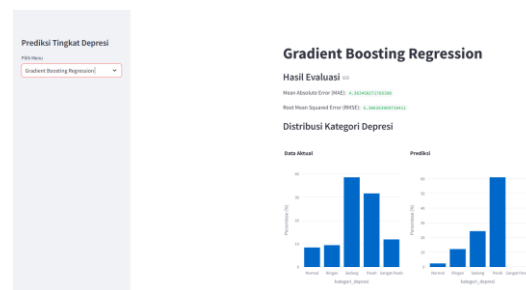
Gambar 3 menunjukkan hasil evaluasi dan visualisasi dari metode *Multiple Linear Regression*. Gambar 4 menunjukkan hasil evaluasi dan visualisasi dari metode *Random Forest Regressor*. Gambar 5 menunjukkan hasil evaluasi dan visualisasi dari metode *Gradient Boosting Regression*.



Gambar 2. Halaman Visualisasi Multiple Linear Regression



Gambar 3. Halaman Visualisasi Random Forest Regressor

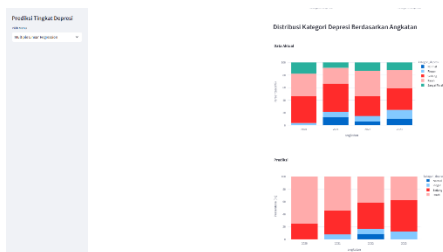


Gambar 4. Halaman Visualisasi Gradient Boosting Regression

Selain memprediksi tingkat depresi menggunakan skor kecemasan dan stres, penelitian ini juga mengeksplorasi distribusi

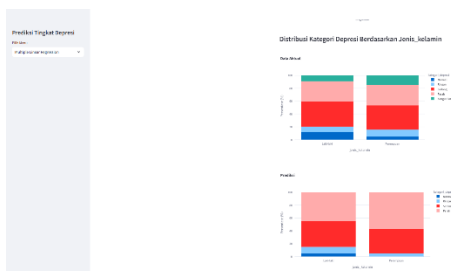
kategori depresi berdasarkan variabel tambahan, seperti jenis kelamin, angkatan, dan status tempat tinggal (tinggal dengan keluarga atau tidak). Hasil analisis menunjukkan pola-pola berikut:

1. Multiple Linear Regression



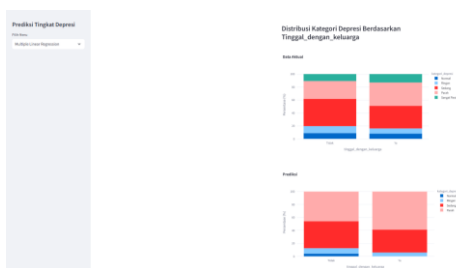
Gambar 5. Distribusi Kategori Depresi Berdasarkan Angkatan Menggunakan Metode Multiple Linear Regression

Berdasarkan Gambar 5, model mampu menangkap dominasi kategori “Sedang” pada semua angkatan, namun kesulitan pada kategori dengan frekuensi kecil seperti “Sangat Parah”.



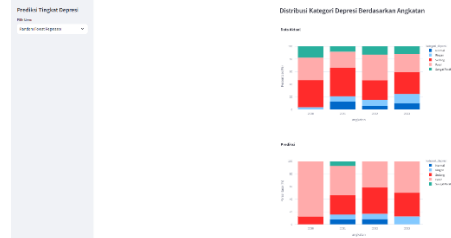
Gambar 6. Distribusi Kategori Depresi Berdasarkan Jenis Kelamin Menggunakan Metode Multiple Linear Regression

Berdasarkan Gambar 6, model memprediksi distribusi yang cukup baik, namun terdapat sedikit *overestimate* pada kategori “Ringan” dan *underestimate* pada “Sangat Parah”.



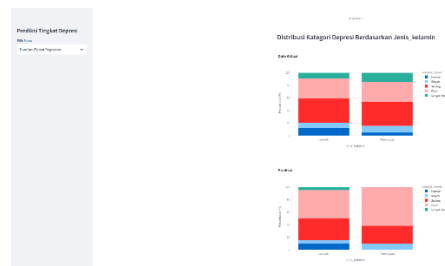
Gambar 7. Distribusi Kategori Depresi Berdasarkan Status Tinggal Menggunakan Metode Multiple Linear Regression

2. Random Forest Regressor



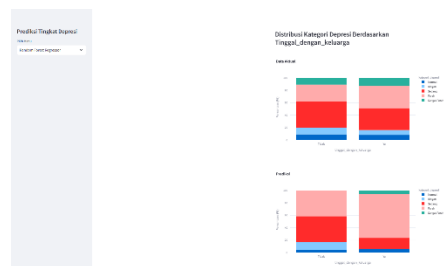
Gambar 8. Distribusi Kategori Depresi Berdasarkan Angkatan Menggunakan Metode Random Forest Regressor

Berdasarkan Gambar 8, distribusi antarangkatan cukup konsisten, namun terdapat *overestimate* pada kategori “Parah” dan *underestimate* pada “Normal”.



Gambar 9. Distribusi Kategori Depresi Berdasarkan Jenis Kelamin Menggunakan Metode Random Forest Regressor

Berdasarkan Gambar 9, model memprediksi pola distribusi dengan baik, meskipun terdapat *overestimate* pada kategori “Parah” untuk perempuan.

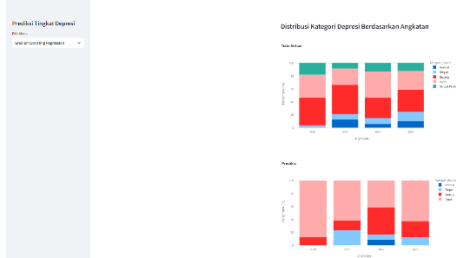


Gambar 10. Distribusi Kategori Depresi Berdasarkan Status Tinggal Menggunakan Metode Random Forest Regressor

Berdasarkan Gambar 10, model berhasil memprediksi kategori Sedang dengan baik, tetapi sedikit *overestimate* pada kategori

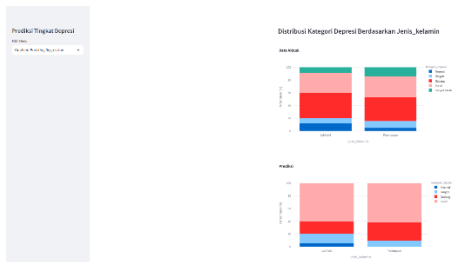
“Parah”, terutama untuk mahasiswa yang tidak tinggal dengan keluarga.

### 3. Gradient Boosting Regression



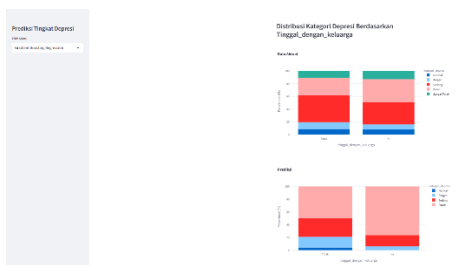
Gambar 11. Distribusi Kategori Depresi Berdasarkan Angkatan Menggunakan Metode Gradient Boosting Regression

Berdasarkan Gambar 11, model menangkap pola distribusi antarangkatan dengan akurasi yang lebih baik dibandingkan dua model lainnya.



Gambar 12. Distribusi Kategori Depresi Berdasarkan Jenis Kelamin Menggunakan Metode Gradient Boosting Regression

Gambar 12 memiliki prediksi yang paling stabil, dengan kesalahan prediksi minimal pada semua kategori.



Gambar 13. Distribusi Kategori Depresi Berdasarkan Status Tinggal Menggunakan Metode Gradient Boosting Regression

Gambar 13 menunjukkan prediksi yang mendekati data aktual, termasuk pada kategori Sangat Parah, meskipun sedikit underestimate.

Dari hasil diatas, hal ini menunjukkan *Gradient Boosting Regression* lebih akurat dan konsisten dalam memprediksi tingkat depresi berdasarkan skor kecemasan dan stres. Model ini berhasil menangkap hubungan kompleks antara variabel independen dan dependen, namun masih memiliki keterbatasan pada kategori depresi yang jarang, seperti “Sangat Parah”. Ketidakeimbangan data dan pola hubungan yang kompleks dapat memengaruhi akurasi prediksi pada kategori tertentu. Bisa disimpulkan bahwa mahasiswa angkatan 2022 dan 2023 menunjukkan kategori depresi sedang yang lebih tinggi. Perempuan cenderung memiliki skor depresi lebih tinggi dibandingkan laki-laki. Mahasiswa yang tinggal jauh dari keluarga juga lebih rentan mengalami depresi dibandingkan yang tinggal bersama keluarga.

Dengan memanfaatkan bahasa pemrograman Python serta alat berbasis web seperti GitHub dan Streamlit, sebuah prototipe aplikasi visualisasi prediksi untuk ketiga model tingkat depresi mahasiswa serta visualisasi hasil eksplorasi distribusi kategori depresi berdasarkan variabel tambahan, seperti jenis kelamin, angkatan, dan status tempat tinggal (tinggal dengan keluarga atau tidak) berhasil dikembangkan dan dapat diakses melalui URL berikut: <https://prediksidepresi.streamlit.app/>.

### 4. KESIMPULAN

Penelitian ini membandingkan performa ketiga model prediksi yaitu *Multiple Linear Regression*, *Random Forest Regressor*, dan *Gradient Boosting Regressor* untuk memprediksi tingkat depresi mahasiswa Teknik Informatika-S1 UDINUS pada mata kuliah Algoritma dan Struktur Data berdasarkan skor kecemasan dan skor stres menggunakan ketiga metode. Dari hasil evaluasi performa, *Gradient Boosting Regression* adalah pilihan terbaik untuk prediksi depresi dalam dataset ini karena performanya yang lebih akurat dalam menangkap pola distribusi utama dengan nilai MAE dan RMSE paling rendah yaitu 4.363 dan 5.386. Model ini mampu menangkap distribusi kategori depresi dengan akurasi yang lebih tinggi dibandingkan kedua model lainnya, terutama pada variabel jenis kelamin, angkatan, dan status tempat tinggal. Namun, model ini masih memiliki keterbatasan, khususnya pada kategori depresi yang jarang terjadi (misalnya “Sangat Parah”), di mana error model masih cenderung meningkat. Penggunaan *framework*



Streamlit memudahkan dalam menampilkan hasil analisis dan prediksi secara interaktif. Hasil analisis ini dapat membantu kampus memberikan perhatian khusus pada mahasiswa berpotensi depresi tinggi, seperti yang tinggal jauh dari keluarga.

Disarankan kepada penelitian selanjutnya agar dapat mengoptimasi model dengan melakukan *hyperparameter tuning* pada *Random Forest* dan *Gradient Boosting Regression* untuk meningkatkan akurasi, menambahkan variabel lain yang relevan untuk menangkap pola lebih kompleks, menciptakan fitur baru dari data yang ada, menggunakan model yang lebih kompleks, menyeimbangkan data, menggunakan *dataset* yang lebih besar dan beragam, serta menyempurnakan visualisasi data agar lebih interaktif pada *framework* Streamlit.

#### PERNYATAAN PENGHARGAAN

Untuk semua pihak yang sudah turut memberikan dukungan, penulis mengucapkan rasa terima kasih sebesar-besarnya sehingga penelitian ini dapat terwujud.

#### DAFTAR PUSTAKA

- [1] F. S. Putri, Z. Nazihah, D. P. Ariningrum, S. Celesta, and C. K. Herbawani, "Depresi Remaja di Indonesia: Penyebab dan Dampaknya," *Jurnal Kesehatan Poltekkes Kemenkes RI PangkalPinang*, vol. 10, no. 2, pp. 99–108, 2022.
- [2] M. Amaludin *et al.*, "Tingkat Depresi, Ansietas Dan Stres Pasien Gagal Ginjal Kronik (Ggk) Dengan Hemodialisa," *Husada Mahakam: Jurnal Kesehatan*, vol. 13, no. 1, pp. 1–7, 2023.
- [3] Y. A. Zakariya, F. Fakhrurrazy, and F. Fujiati, "Literature Review: Korelasi Kadar Malondialdehid Plasma dengan Depresi pada Lanjut Usia," *Homeostasis*, vol. 4, no. 3, pp. 805–814, 2021.
- [4] J. Zeniarja, A. Salam, and F. A. Ma'ruf, "Seleksi Fitur dan Perbandingan Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa," *Jurnal Rekayasa Elektrika*, vol. 18, no. 2, 2022.
- [5] D. Maryono, "Analisis Kesulitan Mahasiswa Prodi PTIK FKIP UNS dalam Penyelesaian Masalah dengan Pemrograman," in *Prosiding Seminar Nasional UNS Vocational Day*, 2016.
- [6] M. Szabo and P. F. Lovibond, "Development and psychometric properties of the DASS-youth (DASS-Y): an extension of the depression anxiety stress scales (DASS) to adolescents and children," *Front Psychol*, vol. 13, p. 766890, 2022.
- [7] M. Makara-Studzińska, E. Tyburski, M. Załuski, K. Adamczyk, J. Mesterhazy, and A. Mesterhazy, "Confirmatory Factor Analysis of Three Versions of The Depression Anxiety Stress Scale (DASS-42, DASS-21, and DASS-12) in Polish Adults," *Front Psychiatry*, vol. 12, p. 770532, 2022.
- [8] I. Marijanović *et al.*, "Use of the Depression, Anxiety and Stress Scale (DASS-21) Questionnaire to Assess Levels of Depression, Anxiety, and Stress in Healthcare and Administrative Staff in 5 Oncology Institutions in Bosnia and Herzegovina During the 2020 COVID-19 Pandemic," *Med Sci Monit*, vol. 27, pp. e930812-1, 2021.
- [9] S. Juwariyah, A. Hulvi, N. Riduan, and K. Kusriani, "Mengukur Faktor Demografi Psikologis: Memprediksi Depresi, Kecemasan, dan Stres dengan menggunakan Machine Learning," *Komputika: Jurnal Sistem Komputer*, vol. 13, no. 2, pp. 149–156, 2024.
- [10] K. F. Khufa and M. Murinto, "Prediksi Kasus Tingkat Depresi Mahasiswa Semester Akhir Menggunakan Regresi Linear Sederhana," *INTEK: Jurnal Informatika dan Teknologi Informasi*, vol. 7, no. 1, pp. 1–6, 2024.
- [11] R. Wajhillah, S. Bahri, and A. Wibowo, "Komparasi Metode Machine Learning pada Diagnosa Gangguan Kejiwaan Depresi," *Syntax*, vol. 9, no. 1, pp. 26–31, 2020.
- [12] F. Firmansyah and A. Yulianto, "Pemodelan Pembelajaran Mesin untuk Prediksi Kesehatan Mental di Tempat Kerja," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 397–407, 2024.
- [13] M. M. SYAIFULLOH, "Prediksi Indeks Standar Pencemaran Udara Di Kota Surabaya Berdasarkan Konsentrasi Gas Karbon Monoksida," *Jambura Journal of Probability and Statistics*, vol. 2, no. 2, pp. 86–95, 2021.
- [14] T. A. Trunfio *et al.*, "Multiple Regression Model to Analyze the Total LOS for

- Patients Undergoing Laparoscopic Appendectomy,” *BMC Med Inform Decis Mak*, vol. 22, no. 1, p. 141, 2022.
- [15] S. Fachid and A. Triayudi, “Perbandingan Algoritma Regresi Linier dan Regresi Random Forest Dalam Memprediksi Kasus Positif Covid-19,” *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, pp. 68–73, 2022.
- [16] K. Kirasich, T. Smith, and B. Sadler, “Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets,” *SMU Data Science Review*, vol. 1, no. 3, p. 9, 2018.
- [17] A. M. Sarah, B. Kurniadi, and E. Warsini, “Implementasi Metode Regresi Linear Dalam Memprediksi Penyakit Anemia Secara Dini,” *Jurnal Teknisi*, vol. 3, no. 1, pp. 14–23, 2023.
- [18] J. Luo, Z. Wei, J. Man, and S. Xu, “TRBoost: a Generic Gradient Boosting Machine Based on Trust-Region Method,” *Applied Intelligence*, vol. 53, no. 22, pp. 27876–27891, 2023.
- [19] B. Pradito and D. S. Purnia, “Komparasi Algoritma Linear Regression dan Neural Network Untuk Memprediksi Nilai Kurs Mata Uang,” *EVOLUSI: Jurnal Sains dan Manajemen*, vol. 10, no. 2, 2022.
- [20] M. Alaeddini, “Emotion Detection in Reddit: Comparative Study of Machine Learning and Deep Learning Techniques,” *arXiv preprint arXiv:2411.10328*, 2024.