

Optimasi Model XGBoost dengan Genetic Algorithm untuk Prediksi Kesehatan Mental Siswa Sekolah Menengah Berbasis Machine Learning

Nor Riduan^{1*}, Alva Hendi Muhammad²

^{1*} Program Studi S2 PJJ Informatika, Universitas Amikom Yogyakarta, Kabupaten Sleman, Daerah Istimewa Yogyakarta, Indonesia.

² Magister Informatika, Universitas Amikom Yogyakarta, Kabupaten Sleman, Daerah Istimewa Yogyakarta, Indonesia.

article info

Article history:

Received 9 April 2025

Received in revised form

10 May 2025

Accepted 1 June 2025

Available online October 2025.

Keywords:

Mental Health Prediction;

Machine Learning; Xgboost

Regressor; Genetic Algorithm;

Hyperparameter Optimization.

Kata Kunci:

Prediksi Kesehatan Mental;

Machine Learning; Xgboost

Regressor; Genetic Algorithm;

Optimasi Hyperparameter.

abstract

Mental health is a vital aspect of human well-being, yet often neglected. Recent studies report a rise in depression, anxiety, and stress among adolescents, especially post-COVID-19. Machine learning has emerged as a powerful tool for predicting mental health conditions. This study employs the XGBoost Regressor using a regression-based ML approach to predict mental health high school students. To enhance accuracy, hyperparameter optimization is conducted using a Genetic Algorithm (GA) to identify the optimal parameter set. The baseline model achieved an MSE of 0.3698, RMSE of 0.6081, and MAPE of 14.09%. After GA optimization, performance improved to an MSE of 0.3092 (16.4% reduction), RMSE of 0.5560 (8.6% reduction), and MAPE of 12.88% (8.6% reduction). These results demonstrate the model's effectiveness for early mental health screening in educational settings, enabling timely interventions by school counselors and healthcare providers.

abstrak

Kesehatan mental merupakan aspek penting dari kesejahteraan manusia yang sering terabaikan. Studi terbaru menunjukkan peningkatan kasus depresi, kecemasan, dan stres pada remaja, terutama setelah pandemi COVID-19. Machine learning telah muncul sebagai alat yang kuat untuk memprediksi kondisi kesehatan mental. Studi ini menggunakan XGBoost Regressor dengan pendekatan machine learning berbasis regresi untuk memprediksi kesehatan mental siswa sekolah menengah. Untuk meningkatkan akurasi, dilakukan optimasi hyperparameter menggunakan Genetic Algorithm (GA) guna menemukan kombinasi parameter terbaik. Model awal menghasilkan MSE sebesar 0,3698, RMSE sebesar 0,6081, dan MAPE sebesar 14,09%. Setelah optimasi dengan GA, kinerja model meningkat dengan MSE sebesar 0,3092 (penurunan 16,4%), RMSE sebesar 0,5560 (penurunan 8,6%), dan MAPE sebesar 12,88% (penurunan 8,6%). Hasil ini membuktikan keandalan model untuk skrining dini di lingkungan sekolah, mendukung intervensi tepat waktu oleh konselor dan tenaga kesehatan.

Corresponding Author. Email: nor.riduan0705@students.amikom.ac.id ^{1}.

Copyright 2025 by the authors of this article. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan Riset). This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. 



ACM Computing Classification System (CCS)

EBSCOhost

Communication and Mass Media Complete (CMMC)

1. Pendahuluan

Kesehatan mental merupakan bagian penting dari kesejahteraan manusia, tetapi kerap terabaikan (Ayuningtyas *et al.*, 2018). Menurut World Health Organization (WHO), lebih dari 264 juta orang di seluruh dunia mengalami depresi, dan sekitar 284 juta orang mengalami kecemasan. Kondisi ini tidak hanya memengaruhi kualitas hidup individu, tetapi juga berdampak pada produktivitas serta beban ekonomi secara global (Śniadach *et al.*, 2021). Kesehatan mental remaja sekolah menengah menjadi semakin krusial, khususnya setelah pandemi COVID-19 (Sifat *et al.*, 2022). Data dari Centers for Disease Control and Prevention (CDC) pada tahun 2021 menunjukkan bahwa lebih dari 37% siswa sekolah menengah di Amerika Serikat melaporkan mengalami gangguan kesehatan mental selama pandemi. Sebanyak 44% menyatakan merasa sedih atau putus asa dalam periode setahun terakhir, dan lebih dari 55% mengalami kekerasan emosional di rumah. Bahkan, 11% di antaranya mengalami kekerasan fisik seperti dipukul atau disakiti (Jones *et al.*, 2022). Sementara itu, studi di Nepal menunjukkan bahwa siswa sekolah menengah berisiko mengalami depresi, kecemasan, dan stres karena berbagai faktor seperti tekanan akademik, latar belakang pendidikan orang tua, cyberbullying, serta tipe institusi pendidikan (Karki *et al.*, 2022).

Akses yang terbatas terhadap layanan psikologis dan adanya stigma sosial terhadap isu kesehatan mental turut menjadi penghalang utama dalam mendapatkan dukungan yang diperlukan. Permasalahan ini tergolong kompleks dan melibatkan banyak faktor (van Venrooij, 2024). Ketika tidak dikenali dan tidak ditangani secara tepat, dampaknya bisa mencakup penurunan produktivitas, meningkatnya biaya layanan kesehatan, bahkan risiko bunuh diri (Stelmach *et al.*, 2022). Oleh sebab itu, pengembangan alat prediksi yang andal dan akurat menjadi kebutuhan yang semakin mendesak (Madububambachu *et al.*, 2024). Perkembangan teknologi machine learning (ML) telah membuka peluang untuk melakukan prediksi kesehatan mental secara lebih presisi (Patt & Meva, 2024). ML merupakan bagian dari kecerdasan buatan yang memungkinkan sistem belajar dari data untuk melakukan prediksi atau pengambilan keputusan

secara otomatis (Sarker, 2021). Metode ini dapat diterapkan pada beragam sumber data seperti hasil survei, catatan medis, hingga media sosial, guna mengenali pola yang mengindikasikan risiko gangguan seperti depresi atau kecemasan (Gil *et al.*, 2022). Sejumlah penelitian terdahulu menunjukkan bahwa XGBoost memiliki performa yang sangat baik dalam tugas prediksi kesehatan mental. Chung dan Teo (2023) menunjukkan bahwa akurasi XGBoost mencapai 87,20% dalam klasifikasi gejala depresi, sedikit lebih tinggi dibandingkan Deep Neural Networks yang berada di angka 86,40%. Temuan lain dari Abdul Rahman *et al.* (2023) mengindikasikan bahwa pendekatan machine learning mampu memprediksi kondisi kesejahteraan mental berdasarkan perilaku kesehatan dengan akurasi yang tinggi. Di sisi lain, penelitian tentang optimasi hyperparameter juga menunjukkan bahwa Genetic Algorithm (GA) dapat meningkatkan performa model secara signifikan. Alibrahim dan Ludwig (2021) melaporkan bahwa GA mampu mencapai akurasi hingga 90,59% pada data yang kompleks dan tidak linier. Pendekatan ini sangat relevan untuk diterapkan pada data siswa sekolah menengah yang cenderung multidimensional dan subjektif, serta tidak selalu mengikuti pola yang linier (Dao, 2025). Dengan mengintegrasikan GA dalam proses penyetelan parameter model prediktif, diharapkan diperoleh konfigurasi yang lebih peka terhadap indikasi awal gangguan psikologis. Penelitian ini menerapkan pendekatan ML berbasis regresi dalam membangun model prediktif yang mampu memperkirakan kondisi kesehatan mental secara akurat. Algoritma utama yang digunakan adalah XGBoost Regressor, yang memiliki kemampuan dalam memodelkan hubungan non-linear antar variabel dan menangkap interaksi fitur secara efisien. Untuk meningkatkan kinerja model, digunakan optimasi berbasis GA dalam proses penyesuaian hyperparameter guna meminimalkan galat prediksi.

2. Metodologi Penelitian

Kesehatan mental merupakan bagian penting dari kesejahteraan manusia, tetapi kerap terabaikan (Ayuningtyas *et al.*, 2018). Menurut World Health Organization (WHO), lebih dari 264 juta orang di seluruh dunia mengalami depresi, dan sekitar 284 juta

orang mengalami kecemasan. Kondisi ini tidak hanya memengaruhi kualitas hidup individu, tetapi juga berdampak pada produktivitas serta beban ekonomi secara global (Śniadach *et al.*, 2021). Kesehatan mental remaja sekolah menengah menjadi semakin krusial, khususnya setelah pandemi COVID-19 (Sifat *et al.*, 2022). Data dari Centers for Disease Control and Prevention (CDC) pada tahun 2021 menunjukkan bahwa lebih dari 37% siswa sekolah menengah di Amerika Serikat melaporkan mengalami gangguan kesehatan mental selama pandemi. Sebanyak 44% menyatakan merasa sedih atau putus asa dalam periode setahun terakhir, dan lebih dari 55% mengalami kekerasan emosional di rumah. Bahkan, 11% di antaranya mengalami kekerasan fisik seperti dipukul atau disakiti (Jones *et al.*, 2022). Sementara itu, studi di Nepal menunjukkan bahwa siswa sekolah menengah berisiko mengalami depresi, kecemasan, dan stres karena berbagai faktor seperti tekanan akademik, latar belakang pendidikan orang tua, cyberbullying, serta tipe institusi pendidikan (Karki *et al.*, 2022). Akses yang terbatas terhadap layanan psikologis dan adanya stigma sosial terhadap isu kesehatan mental turut menjadi penghalang utama dalam mendapatkan dukungan yang diperlukan.

Permasalahan ini tergolong kompleks dan melibatkan banyak faktor (van Venrooij, 2024). Ketika tidak dikenali dan tidak ditangani secara tepat, dampaknya bisa mencakup penurunan produktivitas, meningkatnya biaya layanan kesehatan, bahkan risiko bunuh diri (Stelmach *et al.*, 2022). Oleh sebab itu, pengembangan alat prediksi yang andal dan akurat menjadi kebutuhan yang semakin mendesak (Madububambachu *et al.*, 2024). Perkembangan teknologi machine learning (ML) telah membuka peluang untuk melakukan prediksi kesehatan mental secara lebih presisi (Patt & Meva, 2024). ML merupakan bagian dari kecerdasan buatan yang memungkinkan sistem belajar dari data untuk melakukan prediksi atau pengambilan keputusan secara otomatis (Sarker, 2021). Metode ini dapat diterapkan pada beragam sumber data seperti hasil survei, catatan medis, hingga media sosial, guna mengenali pola yang mengindikasikan risiko gangguan seperti depresi atau kecemasan (Gil *et al.*, 2022).

Sejumlah penelitian terdahulu menunjukkan bahwa XGBoost memiliki performa yang sangat baik dalam tugas prediksi kesehatan mental. Chung dan Teo (2023) menunjukkan bahwa akurasi XGBoost mencapai 87,20% dalam klasifikasi gejala depresi, sedikit lebih tinggi dibandingkan Deep Neural Networks yang berada di angka 86,40%. Temuan lain dari Abdul Rahman *et al.* (2023) mengindikasikan bahwa pendekatan machine learning mampu memprediksi kondisi kesejahteraan mental berdasarkan perilaku kesehatan dengan akurasi yang tinggi. Di sisi lain, penelitian tentang optimasi hyperparameter juga menunjukkan bahwa Genetic Algorithm (GA) dapat meningkatkan performa model secara signifikan. Alibrahim dan Ludwig (2021) melaporkan bahwa GA mampu mencapai akurasi hingga 90,59% pada data yang kompleks dan tidak linier. Pendekatan ini sangat relevan untuk diterapkan pada data siswa sekolah menengah yang cenderung multidimensional dan subjektif, serta tidak selalu mengikuti pola yang linier (Dao, 2025). Dengan mengintegrasikan GA dalam proses penyetelan parameter model prediktif, diharapkan diperoleh konfigurasi yang lebih peka terhadap indikasi awal gangguan psikologis. Penelitian ini menerapkan pendekatan ML berbasis regresi dalam membangun model prediktif yang mampu memperkirakan kondisi kesehatan mental secara akurat. Algoritma utama yang digunakan adalah XGBoost Regressor, yang memiliki kemampuan dalam memodelkan hubungan non-linear antar variabel dan menangkap interaksi fitur secara efisien. Untuk meningkatkan kinerja model, digunakan optimasi berbasis GA dalam proses penyesuaian hyperparameter guna meminimalkan galat prediksi.

$$Obj^{(r)} = \sum_{i=1}^m L(y_i, \widehat{y}_i^{(r)}) + \sum_{k=1}^r \Omega(g_r)$$

Dimana:

i = indeks data ke- i dalam dataset.

m = jumlah total data yang dimasukkan ke dalam pohon.

$l(y_i, \widehat{y}_i^{(t)})$ = fungsi kehilangan yang mengukur selisih antara nilai yang diprediksi ($\widehat{y}_i^{(t)}$) dan nilai sebenarnya (y_i).

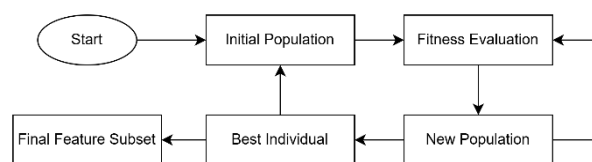
g_r = item struktural dari model pohon keputusan.

$\Omega(f_k)$ = adalah term regularisasi yang membantu mengontrol kompleksitas model pohon keputusan, mencegah overfitting.

Untuk meningkatkan kinerja model XGBoost, digunakan metode Genetic Algorithm (GA). GA pertama kali diperkenalkan oleh John Holland melalui karya ilmiahnya berjudul *Adaptation in Natural and Artificial Systems* pada tahun 1975. Pendekatan ini mengacu pada prinsip evolusi biologis, khususnya seleksi alam sebagaimana dijelaskan oleh Darwin. GA tergolong dalam kategori teknik metaheuristik yang digunakan untuk menemukan solusi optimal dalam berbagai permasalahan optimasi, baik linier maupun non-linier. Dalam pemodelan XGBoost, GA digunakan untuk menyempurnakan konfigurasi hyperparameter. Setiap calon solusi direpresentasikan dalam bentuk struktur simbolik menyerupai kromosom, yang kemudian dievaluasi menggunakan fungsi kinerja (fitness function). Penilaian fitness didasarkan pada nilai Mean Squared Error (MSE), di mana solusi dengan nilai MSE yang lebih rendah dianggap lebih baik. Proses pencarian solusi optimal dilakukan secara iteratif melalui tiga operator utama: seleksi, persilangan (*crossover*), dan mutasi. Operator seleksi berperan dalam memilih individu dengan performa terbaik berdasarkan nilai fitness untuk direproduksi pada generasi berikutnya.

Operator persilangan digunakan untuk menggabungkan dua individu terpilih agar menghasilkan keturunan dengan karakteristik gabungan yang potensial. Sementara itu, mutasi diterapkan untuk menambahkan variasi secara acak guna menghindari stagnasi di solusi suboptimal dan memperluas pencarian dalam ruang solusi. Melalui iterasi berkelanjutan, populasi solusi mengalami perbaikan bertahap. Struktur kromosom diperbarui secara dinamis agar semakin sesuai dengan permasalahan yang sedang dihadapi. Proses ini pada akhirnya menghasilkan konfigurasi yang lebih optimal dibandingkan pendekatan konvensional. GA telah digunakan secara luas dalam berbagai aplikasi, termasuk perencanaan sistem, rekayasa desain, manajemen sumber daya, serta pemrosesan data. Dalam penelitian ini, GA dipilih karena kemampuannya untuk menyesuaikan setiap hyperparameter secara individual dan simultan,

sehingga menghasilkan kombinasi parameter yang lebih efektif. Proses tersebut divisualisasikan pada Gambar 1.



Gambar 1. Langkah Umum Algoritma Genetika

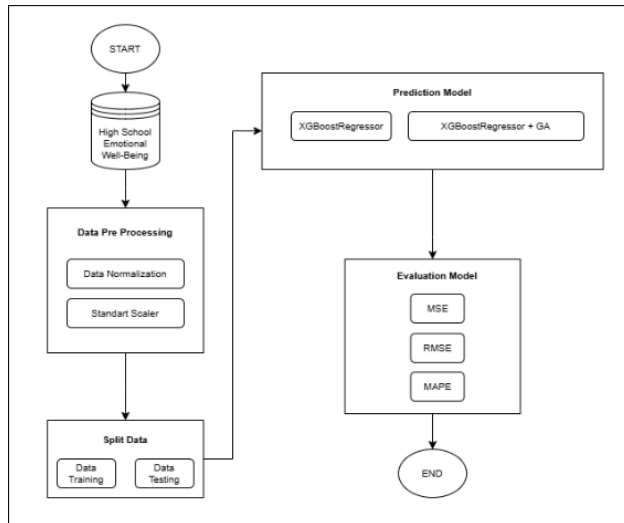
Populasi awal terdiri atas sejumlah solusi kandidat yang dievaluasi menggunakan fungsi fitness untuk menilai kualitas masing-masing individu. Proses seleksi digunakan untuk memilih individu-individu terbaik berdasarkan nilai fitness, sedangkan operator crossover berfungsi menggabungkan gen dari dua individu induk guna membentuk solusi baru yang berpotensi lebih baik. Mutasi kemudian diterapkan untuk memperkenalkan variasi acak, mencegah stagnasi pada solusi lokal. Proses ini terus berlangsung secara berulang hingga kriteria penghentian terpenuhi, misalnya jumlah generasi maksimum atau tercapainya stabilitas nilai fitness antar generasi.

Jenis, Sifat, dan Pendekatan Penelitian

Penelitian ini bersifat kuantitatif dengan tujuan membangun model prediksi kesehatan mental siswa berdasarkan data numerik yang merepresentasikan berbagai variabel terkait. Sifat penelitian ini adalah deskriptif, dengan fokus pada penjabaran langkah-langkah dalam proses pengumpulan, pra-pemrosesan, dan analisis data yang berkaitan dengan kesehatan mental, hingga tahap pengujian model. Pendekatan yang digunakan adalah pendekatan eksperimental berbasis machine learning, di mana proses pemodelan dilakukan menggunakan algoritma regresi XGBoost Regressor untuk memperoleh hasil prediksi dengan tingkat akurasi tertinggi.

Alur Penelitian

Penelitian ini disusun dalam beberapa tahap yang dilakukan secara sistematis guna mengidentifikasi kondisi kesehatan mental siswa. Rangkaian kegiatan tersebut dijelaskan lebih lanjut pada Gambar 2, yang memvisualisasikan struktur tahapan mulai dari studi literatur hingga evaluasi performa model.



Gambar 2. Kerangka Alur Penelitian

Penelitian diawali dengan kajian literatur guna menentukan pendekatan dan algoritma yang sesuai dengan tujuan studi. Setelah itu, dilakukan pengumpulan data. Sumber data yang digunakan dalam penelitian ini merupakan data sekunder, yakni data terbuka (open data) yang dapat diakses secara publik dan tidak diperoleh secara langsung dari responden. Dataset yang digunakan diperoleh dari platform Kaggle melalui tautan: <https://www.kaggle.com/datasets/jeffreyxshen/high-school-emotional-well-being>. Tahapan pra-pemrosesan (preprocessing) mencakup dua langkah utama, yaitu normalisasi dan standarisasi. Normalisasi dilakukan dengan mengubah atribut kategorikal menjadi format numerik agar dapat diproses oleh algoritma machine learning (Kshirsagar *et al.*, 2022). Sementara itu, proses standarisasi diterapkan menggunakan teknik standard scaler, yang berfungsi untuk mengubah variabel kontinu agar memiliki distribusi dengan rata-rata nol dan standar deviasi satu (Sarker, 2021). Penyesuaian skala fitur ini penting karena dapat memengaruhi performa dan kecepatan konvergensi algoritma yang digunakan

(Ahsan *et al.*, 2021). Setelah data dibersihkan dan disiapkan, dataset dibagi menjadi dua bagian: 80% untuk pelatihan (training) dan 20% untuk pengujian (testing). Pembagian ini bertujuan untuk memastikan bahwa validasi model dilakukan secara proporsional dan representatif terhadap keseluruhan data. Tahapan pemodelan terdiri atas dua skenario utama. Skenario pertama menggunakan algoritma XGBoost Regressor sebagai model dasar (baseline), sedangkan skenario kedua menerapkan XGBoost Regressor dengan penyetelan hyperparameter melalui Genetic Algorithm (GA). Parameter yang disesuaikan melalui proses optimasi ini meliputi:

Model yang telah dibangun kemudian dievaluasi dengan tiga metrik pengukuran performa:

- 1) *Mean Squared Error* (MSE)
Mengukur rata-rata kuadrat kesalahan, dengan memberikan bobot lebih besar terhadap kesalahan yang besar.
- 2) *Root Mean Squared Error* (RMSE)
Merupakan akar kuadrat dari MSE dan memberikan nilai kesalahan dalam satuan yang sama dengan target.
- 3) *Mean Absolute Percentage Error* (MAPE)
Mengukur kesalahan prediksi sebagai persentase dari nilai aktual.

3. Hasil dan Pembahasan

Hasil Dataset

Berikut adalah ringkasan dari penjelasan tentang atribut dataset yang menggambarkan faktor-faktor yang memengaruhi kesejahteraan emosional dan kesehatan mental siswa, dengan total 468 entri dan 22 atribut. Atribut dataset dirinci pada Tabel 1.

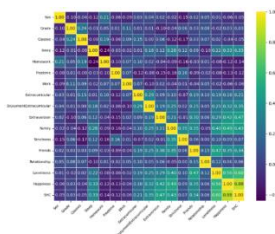
Tabel 1. Atribut Dataset

No	Atribut	Deskripsi
1	Parental Consent	Menunjukkan apakah orang tua siswa memberikan persetujuan untuk partisipasi dalam penelitian.
2	Sex	Jenis kelamin siswa, yang berpengaruh terhadap kesejahteraan emosional.
3	Grade	Kelas atau tingkat pendidikan siswa saat ini (misalnya, kelas 9, 10, atau 11).

4	Classes	Jenis mata pelajaran yang diambil oleh siswa.
5	Sleep	Kualitas atau durasi tidur siswa, berhubungan dengan kesejahteraan emosional dan kesehatan mental.
6	Homework	Waktu yang dihabiskan untuk mengerjakan pekerjaan rumah, berhubungan dengan stres akademik.
7	Freetime	Jumlah waktu yang dimiliki siswa untuk kegiatan di luar sekolah.
8	Work	Jika siswa memiliki pekerjaan paruh waktu atau tidak.
9	Extracurricular	Partisipasi siswa dalam kegiatan ekstrakurikuler seperti olahraga atau seni.
10	Enjoyment Extracurricular	Seberapa menikmati siswa kegiatan ekstrakurikuler mereka.
11	Extraversion	Tingkat ekstrasversi siswa berdasarkan dimensi kepribadian.
12	Family	Struktur dan kualitas hubungan keluarga siswa.
13	Strictness	Tingkat ketegasan yang diterapkan oleh orang tua terhadap siswa.
14	Friends	Jumlah atau kualitas hubungan pertemanan siswa.
15	Relationship	Hubungan siswa dengan teman sebaya atau pasangan.
16	Loneliness	Tingkat perasaan kesepian yang dialami siswa.
17	Happiness	Tingkat kebahagiaan yang dirasakan siswa.
18	Happiness Comparison	Perbandingan tingkat kebahagiaan siswa dengan teman-teman atau rata-rata populasi.
19	Characterization Happiness	Bagaimana siswa menggambarkan kebahagiaan mereka.
20	Characterization Depression	Deskripsi bagaimana siswa menggambarkan perasaan depresi atau kesedihan mereka.
21	Mental Health	Status umum kesehatan mental siswa.
22	SHC (Self-Reported Health Condition)	Kondisi kesehatan yang dilaporkan sendiri oleh siswa, mencakup masalah fisik atau psikologis.

Preprocessing

Tahap *preprocessing* awal dilakukan dengan mengonversi atribut kategorikal “*sex*” ke format numerik. Untuk atribut kategorikal *Parental Consent* (Persetujuan Orang Tua), proses penghapusan dilakukan secara manual dengan menghilangkan poin ini dari dataset karena hasil kategorikalnya selalu “Yes” sehingga tidak memberikan variasi yang relevan untuk analisis. Kolom ‘*Happiness Comparison*, *Characterization Happiness*, *Characterization Depression*, *Mental Health*’ dihapus karena merupakan penilaian yang bersifat pasti dalam perhitungan SHC. Gambar 3 menunjukkan pearson correlation dari dataset, antara berbagai fitur dalam dataset.



Gambar 3. Korelasi Pearson antara Semua Fitur

Warna yang lebih terang (kuning) menunjukkan korelasi positif yang tinggi, sedangkan warna gelap (ungu tua) menunjukkan korelasi negatif atau hubungan yang lemah. Dari analisis *heatmap*, terlihat bahwa SHC memiliki korelasi yang sangat tinggi dengan *Happiness* (0.88), yang berarti bahwa individu dengan tingkat kebahagiaan subjektif yang tinggi cenderung memiliki skor SHC yang tinggi pula. Selain itu, SHC juga memiliki korelasi positif dengan beberapa faktor lain seperti *Enjoyment of Extracurricular*, *Extraversion*, *Family*, dan *Friends*, yang menunjukkan bahwa keterlibatan dalam aktivitas ekstrakurikuler, tingkat ekstroverasi, serta dukungan sosial dari keluarga dan teman berkontribusi terhadap perasaan kebahagiaan subjektif seseorang. Sebaliknya, variabel seperti *Loneliness* menunjukkan korelasi negatif dengan SHC, yang berarti semakin tinggi tingkat kesepian seseorang, semakin rendah kebahagiaan subjektifnya.

Normalisasi dan Standardisasi

Dataset ini kemudian dibagi menjadi data training (80%) dan data testing (20%) untuk memastikan validasi model dilakukan secara akurat. Proses pemodelan melibatkan dua pendekatan utama: *XGBoost Regressor* sebagai model baseline dan *XGBoost Regressor* dengan optimasi *hyperparameter* menggunakan GA. Proses pemodelan *XGBoost Regressor* ditunjukkan pada Kode 1.

Kode 1. Pemodelan *XGBoost Regressor*

```
# Membuat model XGBoost
xgboost_model
=xgb.XGBRegressor(objective='reg:squarederror',
random_state=42)

# Melatih model pada data latih
xgboost_model.fit(x_train, y_train)

# Memprediksi menggunakan data uji
y_pred = xgboost_model.predict(x_test)

# Hitung MSE
mse = mean_squared_error(y_test, y_pred)

# Hitung RMSE
rmse = np.sqrt(mse)

# Hitung MAPE
mape = np.mean(np.abs((y_test - y_pred) /
y_test)) * 100

# Tampilkan hasil
print(f"MSE: {mse}")
print(f"RMSE: {rmse}")
print(f"MAPE: {mape}%")
```

Dari pemrosesan kode di atas diperoleh nilai *MSE*: 0.3698, *RMSE*: 0.6081, *MAPE*: 14.09%. Selanjutnya dilakukan implementasi algoritma genetika untuk mengoptimalkan suatu fungsi atau mencari solusi terbaik dalam suatu ruang pencarian dengan parameter utama seperti pada Kode 2.

Kode 2. Parameter utama GA

```
# Genetic Algorithm Parameters
POP_SIZE = 30 # Ukuran populasi
N_GENERATIONS = 50 # Jumlah generasi
MUTATION_RATE = 0.1 # Tingkat mutasi
```

Ukuran populasi, jumlah generasi, dan tingkat mutasi merupakan tiga parameter utama dalam GA. *POP_SIZE* = 30 menunjukkan bahwa dalam setiap generasi, algoritma akan bekerja dengan 30 individu atau solusi kandidat untuk mencari solusi terbaik. Kemudian, *N_GENERATIONS* = 50 menentukan

bahwa proses evolusi akan berlangsung selama 50 generasi, di mana populasi akan terus diperbarui dan berevolusi untuk menemukan solusi yang optimal. Sedangkan, *MUTATION_RATE* = 0.1 mengindikasikan bahwa terdapat peluang sebesar 10% bagi setiap individu untuk mengalami mutasi pada setiap generasi. Mutasi ini berfungsi untuk menjaga keberagaman dalam populasi dan mencegah algoritma terjebak dalam solusi lokal yang tidak optimal. Parameter-parameter ini memainkan peran penting dalam kinerja algoritma genetika, karena mereka menentukan seberapa cepat dan seberapa baik algoritma dapat menemukan solusi yang optimal dalam suatu ruang pencarian. Setelah implementasi algoritma genetika, diperoleh parameter terbaik yaitu *Best Hyperparameters*: {'n_estimators': 355, 'max_depth': 3, 'learning_rate': 0.020599781401660078, 'subsample': 0.5437685563048382}. Parameter yang didapat diimplementasikan pada pemodelan *XGBoost Regressor* seperti pada Kode 3.

Kode 3. *XGBoost Regressor* dengan best *hyperparameter* yang didapat dari GA

```
# Buat model dengan hyperparameter terbaik
optimized_model = xgb.XGBRegressor(
n_estimators=best_params["n_estimators"],
max_depth=best_params["max_depth"],

learning_rate=best_params["learning_rate"],
subsample=best_params["subsample"],
objective='reg:squarederror',
random_state=42
)

# Latih model pada data latih
optimized_model.fit(x_train, y_train)

# Evaluasi pada data uji
y_pred = optimized_model.predict(x_test)
mse = mean_squared_error(y_test, y_pred)

# Hitung RMSE
rmse = np.sqrt(mse)

# Hitung MAPE
mape = np.mean(np.abs((y_test - y_pred) /
y_test)) * 100

# Tampilkan hasil
print(f"MSE: {mse}")
print(f"RMSE: {rmse}")
print(f"MAPE: {mape}%")
```

Dari pemrosesan kode di atas diperoleh nilai *MSE*: 0.3092, *RMSE*: 0.55602, *MAPE*: 12.88%.

Hasil penelitian ini menunjukkan bahwa penggunaan *XGBoost Regressor* dengan optimasi *hyperparameter* menggunakan *GA* mampu meningkatkan akurasi dalam prediksi kesehatan mental siswa sekolah menengah. Model baseline *XGBoost Regressor* tanpa optimasi *hyperparameter* memiliki nilai *MSE* sebesar 0.3698, *RMSE* sebesar 0.6081, dan *MAPE* sebesar 14.09%. Setelah dilakukan optimasi dengan *GA*, nilai *MSE* menurun menjadi 0.3092 (penurunan 16.4%), *RMSE* menjadi 0.5560 (penurunan 8.6%), dan *MAPE* menjadi 12.88% (penurunan 8.6%).

Pembahasan

Hasil penelitian menunjukkan bahwa penerapan *XGBoost Regressor* yang disertai optimasi *hyperparameter* menggunakan *Genetic Algorithm (GA)* mampu meningkatkan performa model dalam memprediksi kesehatan mental siswa. Penurunan nilai *MSE*, *RMSE*, dan *MAPE* secara berturut-turut mengindikasikan bahwa kesalahan prediksi model berkurang secara signifikan setelah proses optimasi dilakukan. Kombinasi parameter optimal yang diperoleh dari *GA*, seperti jumlah estimasi pohon (*n_estimators*), kedalaman maksimum (*max_depth*), tingkat pembelajaran (*learning_rate*), dan proporsi *subsample*, terbukti memberikan keseimbangan antara kompleksitas model dan kemampuan generalisasi terhadap data baru. Keberhasilan pendekatan ini sejalan dengan penelitian Alibrahim dan Ludwig (2021) yang menyatakan bahwa *GA* unggul dibandingkan metode optimasi lainnya seperti *Grid Search* dan *Bayesian Optimization* dalam menghadapi data non-linier dan multidimensional. Selain itu, studi oleh Ayodele *et al.* (2024) membuktikan bahwa *XGBoost Regressor* yang dioptimalkan dengan *GA* mampu menghasilkan nilai *R-squared* hingga 0,9 dalam prediksi tingkat depresi, jauh melampaui performa *linear regression* dan *Random Forest* tanpa *tuning*. Kinerja tersebut menjadikan *GA* sebagai pendekatan yang layak untuk diterapkan dalam pengembangan model prediksi kesehatan mental berbasis *machine learning*. Penelitian ini juga menguatkan temuan Chung dan Teo (2023), yang menunjukkan bahwa *XGBoost* merupakan salah satu model paling akurat dalam klasifikasi gejala depresi, bersaing ketat dengan *Gradient Boosting* dan *Deep Neural Networks*. Meskipun model yang digunakan dalam penelitian ini berbasis regresi, kemampuan prediksinya tetap dapat dibandingkan dengan

pendekatan klasifikasi pada studi lain, terutama setelah disempurnakan melalui proses *tuning*. Selain itu, Shen (2023), yang menggunakan *dataset* serupa namun menerapkan model *Neural Network*, hanya memperoleh nilai *MSE* sebesar 0,4088 dan *R²* sebesar 0,5176, sedangkan model *linear regression* mencatat performa yang hampir sama. Hasil-hasil tersebut menegaskan bahwa penggunaan *GA* dalam optimasi *hyperparameter XGBoost* dapat memberikan hasil yang lebih unggul. Beberapa keterbatasan perlu diperhatikan, seperti ukuran sampel yang relatif kecil serta potensi bias kultural dan geografis pada atribut tertentu, misalnya variabel *Strictness* yang mungkin ditafsirkan berbeda dalam konteks budaya Asia (van Venrooij, 2024). Selain itu, keterbatasan akses terhadap layanan kesehatan mental, terutama di negara-negara berkembang, juga menjadi hambatan dalam penerapan sistem ini secara luas (Ayuningtyas *et al.*, 2018; Ali *et al.*, 2016).

Untuk memperluas potensi penerapan model, penggabungan data pasif dari perangkat *wearable* seperti *Fitbit* dapat menjadi strategi pelengkap guna mengurangi bias dari pelaporan mandiri (Mullick *et al.*, 2022). Selain itu, integrasi model prediktif ke dalam sistem informasi pendidikan seperti *e-rapor* dapat mendukung upaya deteksi dini secara otomatis berdasarkan indikator perilaku siswa, seperti pola tidur atau performa akademik. Langkah-langkah ini dapat menjadi alternatif preventif yang efisien dan adaptif untuk mengatasi permasalahan kesehatan mental remaja di lingkungan sekolah, sejalan dengan rekomendasi berbagai studi global mengenai peningkatan deteksi dan intervensi dini terhadap gangguan psikologis pada kelompok usia muda (Śniadach *et al.*, 2021; Stelmach *et al.*, 2022).

4. Kesimpulan dan Saran

Penelitian ini menunjukkan bahwa penggunaan *XGBoost Regressor* mampu menghasilkan prediksi kesehatan mental siswa dengan tingkat akurasi yang baik. Setelah dilakukan optimasi *hyperparameter* menggunakan *Genetic Algorithm (GA)*, performa model meningkat secara signifikan, dengan penurunan nilai *MSE* sebesar 16,4%, *RMSE* sebesar 8,6%, dan *MAPE* sebesar 8,6% dibandingkan model dasar. Hasil tersebut mengindikasikan bahwa *GA* efektif dalam

menyusun konfigurasi parameter yang optimal untuk meningkatkan kinerja model. Temuan ini memperkuat potensi penerapan *machine learning* (ML) sebagai alat bantu deteksi awal kesehatan mental di lingkungan pendidikan. Dengan pemanfaatan sistem berbasis data seperti ini, intervensi dapat dilakukan lebih cepat dan tepat sasaran. Penggunaan *dataset* yang lebih besar dan lebih representatif di masa mendatang diharapkan mampu meningkatkan generalisasi model terhadap populasi yang lebih luas. Keterbatasan utama dari studi ini terletak pada penggunaan data sekunder yang kemungkinan belum mencerminkan kondisi nyata secara keseluruhan. Oleh karena itu, penelitian lanjutan dapat mempertimbangkan integrasi data *real-time*, seperti dari media sosial atau perangkat *wearable*, guna meningkatkan akurasi model. Selain itu, penggabungan dengan pendekatan lain seperti *Neural Networks* atau pemrosesan bahasa alami (*Natural Language Processing* / *NLP*) dapat menjadi jalur pengembangan untuk meningkatkan efektivitas prediksi pada kasus-kasus yang lebih kompleks, khususnya dalam memantau dan memahami dinamika psikologis remaja secara lebih mendalam.

5. Daftar Pustaka

- Abdul Rahman, H., et al. (2023). Machine learning-based prediction of mental well-being using health behavior data from university students. *Bioengineering*, 10(5). <https://doi.org/10.3390/bioengineering10050575>.
- Ahsan, M. M., Mahmud, M. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*, 9(3), 52.
- Ali, G. C., Ryan, G., & De Silva, M. J. (2016). Validated screening tools for common mental disorders in low and middle income countries: a systematic review. *PLoS one*, 11(6), e0156939.
- Alibrahim, H., & Ludwig, S. A. (2021). Hyperparameter optimization: Comparing genetic algorithm against grid search and Bayesian optimization. *IEEE*. <https://doi.org/10.1109/CEC45853.2021.9504761>.
- Ambarwari, A., Adrian, Q. J., & Herdiyeni, Y. (2020). Analisis pengaruh data scaling terhadap performa algoritme machine learning untuk identifikasi tanaman. *RESTI*, 4(3), 117–122. <https://doi.org/10.29207/RESTI.V4I1.1517>.
- Arora, A., Bojko, L., Kumar, S., Lillington, J., Panesar, S., & Petrungaro, B. (2023). Assessment of machine learning algorithms in national data to classify the risk of self-harm among young adults in hospital: A retrospective study. *International Journal of Medical Informatics*, 177. <https://doi.org/10.1016/j.ijmedinf.2023.105164>.
- Ayodele, A., Adetunla, A., & Akinlabi, E. (2024). Prediction of depression severity and personalised risk factors using machine learning on multimodal data. *International Journal of Online and Biomedical Engineering*, 20(9), 130–143. <https://doi.org/10.3991/ijoe.v20i09.47581>.
- Bajpai, P., & Kumar, M. (2010). Genetic algorithm—an approach to solve global optimization problems. *Indian Journal of computer science and engineering*, 1(3), 199–206.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>.
- Chung, J., & Teo, J. (2023). Single classifier vs. ensemble machine learning approaches for mental health prediction. *Brain Informatics*, 10(1). <https://doi.org/10.1186/s40708-022-00180-6>.
- Dao, D. M. (2025). Prevalence and patterns of anxiety, depressive, hyperactivity, and inattention symptoms among secondary and high school students: A cross-sectional study. *Multidisciplinary Reviews*, 8(5). <https://doi.org/10.31893/multirev.2025161>.

- Gen, M., & Cheng, R. C. H. (2000). *Genetic algorithms & engineering optimization*. Wiley.
<https://doi.org/10.1002/9780470172261>.
- Gil, M., Kim, S.-S., & Min, E. J. (2022). Machine learning models for predicting risk of depression in Korean college students: Identifying family and individual factors. *Frontiers in Public Health*.
<https://doi.org/10.3389/fpubh.2022.1023010>
- Kshirsagar, P. R., Manoharan, H., Selvarajan, S., Alterazi, H. A., Singh, D., & Lee, H. N. (2022). Perception exploration on robustness syndromes with pre-processing entities using machine learning algorithm. *Frontiers in Public Health*, 10.
<https://doi.org/10.3389/fpubh.2022.893989>.
- Li, C., et al. (2022). Genetic algorithm based hyper-parameters optimization for transfer convolutional neural network. In R. Tiwari (Ed.), *International Conference on Advanced Algorithms and Neural Networks (AANN 2022)* (p. 1228512). SPIE.
<https://doi.org/10.1117/12.2637170>.
- Mullick, T., Radovic, A., Shaaban, S., & Doryab, A. (2022). Predicting depression in adolescents using mobile and wearable sensors: Multimodal machine learning-based exploratory study. *JMIR Formative Research*, 6(6). <https://doi.org/10.2196/35807>.
- Patt, R., & Meva, D. (2024). Advancements in machine learning-based mental health prediction: A comprehensive review. In S. Tanwar, P. K. Singh, M. Ganzha, & G. Epiphaniou (Eds.), *Proceedings of Fifth International Conference on Computing, Communications, and Cyber-Security* (pp. 497–507). Springer Nature Singapore.
https://doi.org/10.1007/978-981-97-2550-2_36.
- Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *Springer*.
<https://doi.org/10.1007/s42979-021-00815-1>
- Shen, Y. (2023). High school emotional well-being: Identifying factors and creating a working model for prediction. *Journal of Student Research*.
<https://doi.org/10.47611/jsrhs.v12i3.4743>.
- Suratno, T., & Rarasati, N. (2019). Optimization of genetic algorithm for implementation designing and modeling in academic scheduling. *EKSAKTA*, 20.
<https://doi.org/10.24036/eksakta/vol20-iss01/166>.
- Verma, S., Sharma, C., Aggarwal, G., & Upadhyay, P. (2024). Artificial intelligence-based approach for classification and prediction of mental health. In 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 708–713).
<https://doi.org/10.1109/Confluence60223.2024.10463203>.