

Prediksi Nilai Ujian Berdasarkan Kebiasaan Siswa Menggunakan Algoritma *Random Forest Regressor*

Ruhul Amin ^{1,*}, Anastasia Siwi Fatma Utami ²

^{1,*} Fakultas Teknologi Informasi Program Studi Informatika; Universitas Nusa Mandiri; Jl. Jatiwaringin No. 2, Cipinang Melayu, Makasar Jakarta Timur, (021)8005722; ruhul.ran@nusamandiri.ac.id

² Fakultas Teknik Dan Informatika Program studi Sistem Informasi Akuntansi; Universitas BSI; Jalan Kramat raya no 98 Senen Jakarta Pusat

* Korespondensi: e-mail: ruhul.ran@nusamandiri.ac.id

Diterima: 04 November 2025 ; Review: 04 Desember 2025; Disetujui: 10 Desember 2025

Cara sitasi: Amin R, Utami ASF. 2024. Prediksi Nilai Ujian Berdasarkan Kebiasaan Siswa Menggunakan Algoritma *Random Forest Regressor*. Information System for Educators and Professionals. Vol 10(2): 149-158.

Abstrak: Perkembangan teknologi data memberikan peluang besar untuk menerapkan pembelajaran mesin (*machine learning*) dalam memahami dan memprediksi performa akademik siswa secara lebih akurat. Namun, penelitian yang mengintegrasikan faktor perilaku digital dan kesehatan mental dengan pendekatan machine learning regression masih relatif terbatas, khususnya dalam konteks pendidikan di Indonesia. Untuk menjawab kesenjangan tersebut, penelitian ini bertujuan menganalisis pengaruh kebiasaan belajar dan faktor gaya hidup terhadap nilai ujian siswa menggunakan model pembelajaran mesin terawasi, dengan fokus pada algoritma Random Forest Regression. Dataset penelitian terdiri dari 1.000 data simulasi siswa yang mencakup variabel jam belajar, durasi tidur, penggunaan media sosial, tingkat kesehatan mental, dan kehadiran kelas. Proses penelitian meliputi data preprocessing, pembangunan model, evaluasi performa menggunakan metrik R^2 , MAE, dan RMSE, serta analisis feature importance untuk mengidentifikasi variabel yang paling berpengaruh terhadap capaian akademik. Hasil menunjukkan bahwa Random Forest Regression memberikan performa terbaik dengan R^2 sebesar 0.874, MAE sebesar 4.12, dan RMSE sebesar 5.01, melampaui model Linear Regression dan Decision Tree Regression. Variabel jam belajar per hari dan kesehatan mental menjadi faktor paling dominan dalam memengaruhi nilai ujian siswa. Temuan ini menegaskan bahwa pendekatan pembelajaran mesin berpotensi mendukung pengembangan early warning system berbasis data dan strategi pembelajaran yang adaptif serta berorientasi pada kesejahteraan siswa.

Kata kunci: data-driven education, kebiasaan belajar, pembelajaran mesin, prediksi nilai ujian, Random Forest Regression

Abstract: Advances in data technology have created significant opportunities to apply machine learning in understanding and predicting students' academic performance more accurately. However, studies that integrate digital behavior factors and mental health variables within a machine learning regression framework remain relatively limited, particularly in the context of Indonesian education. To address this gap, this study aims to analyze the influence of learning habits and lifestyle factors on students' exam scores using supervised machine learning models, with a focus on the Random Forest Regression algorithm. The dataset consists of 1,000 simulated student records that include variables such as study hours, sleep duration, social media usage, mental health levels, and class attendance. The research process involves data preprocessing, model development, performance evaluation using R^2 , MAE, and RMSE metrics, and feature importance analysis to identify the most influential variables affecting academic outcomes. The results show that the Random Forest Regression algorithm achieves the best performance, with an R^2 of 0.874, an MAE of 4.12, and an RMSE of 5.01, outperforming Linear Regression and Decision Tree Regression. Study hours per day and mental health status emerge as the most

significant factors influencing exam performance. These findings highlight the potential of machine learning to support the development of data-driven early warning systems and adaptive learning strategies that prioritize student well-being.

Keywords: *academic performance prediction, data-driven education, learning habits, machine learning, mental health, Random Forest Regression*

1. Pendahuluan

Perkembangan teknologi digital telah membawa perubahan besar dalam berbagai aspek kehidupan manusia, termasuk dalam bidang pendidikan [1]. Di era modern, kebiasaan dan gaya hidup siswa tidak hanya dipengaruhi oleh faktor internal seperti motivasi belajar [2], tetapi juga oleh berbagai faktor eksternal seperti intensitas penggunaan media sosial [3], durasi tidur [4], serta kondisi kesehatan mental [5]. Beragam kebiasaan ini berpotensi memberikan pengaruh signifikan terhadap performa akademik siswa, baik secara langsung maupun tidak langsung [6]. Namun demikian, hubungan kompleks antara kebiasaan hidup dan hasil belajar sering kali sulit diukur menggunakan pendekatan statistik konvensional [7], [8].

Dalam konteks tersebut, pendekatan data *science* dan pembelajaran mesin (*machine learning*) menjadi alternatif yang menjanjikan untuk menganalisis pola hubungan *non-linear* antara variabel perilaku dan capaian akademik [9]. Salah satu metode yang banyak digunakan dalam pemodelan prediktif adalah algoritma Random Forest, yaitu model berbasis ensemble learning yang mampu menangani kompleksitas data serta memberikan hasil prediksi yang akurat dan stabil [10], [11].

Algoritma Random Forest telah digunakan dalam berbagai studi untuk memprediksi performa akademik mahasiswa maupun siswa dengan tingkat akurasi tinggi [12], [13]. Model ini mampu menangkap interaksi antar variabel dan menilai pentingnya setiap *fitur* (*feature importance*), menjadikannya relevan dalam memprediksi nilai ujian berdasarkan kebiasaan belajar dan faktor psikologis [14], [15]. Selain itu, kombinasi Random Forest dengan pendekatan optimisasi seperti Particle Swarm Optimization terbukti meningkatkan performa model dalam konteks pendidikan [16], [17].

Beberapa penelitian bahkan mengembangkan sistem pendukung keputusan berbasis *Explainable AI* dan Random Forest untuk menginterpretasikan kontribusi faktor individu terhadap hasil belajar siswa, mendukung guru dan konselor dalam pengambilan keputusan akademik [18],[19],[20]. Beberapa penelitian sebelumnya telah mengkaji faktor-faktor yang memengaruhi prestasi akademik, seperti waktu belajar, kehadiran di kelas, dan dukungan sosial [21], [22]. Namun, sebagian besar studi tersebut masih berfokus pada analisis korelasional atau menggunakan pendekatan statistik tradisional yang mengasumsikan hubungan *linear* antar variabel [13],[23]

Di sisi lain, penelitian yang secara khusus mengintegrasikan faktor perilaku digital dan kesehatan mental dalam pendekatan machine learning regression masih relatif terbatas, terutama dalam konteks pendidikan Indonesia [24],[25],[26],[27]. Sebagian studi sebelumnya lebih berfokus pada metode statistik konvensional, seperti penelitian [28] yang menerapkan regresi linear untuk memprediksi skor ujian siswa secara kuantitatif serta regresi logistik untuk memodelkan kelulusan siswa dalam bentuk klasifikasi biner. Selain itu, penelitian lain [29] mengembangkan model prediksi nilai akhir siswa menggunakan algoritma Decision Tree dengan memanfaatkan atribut-atribut akademik sebagai variabel input. Namun, pendekatan tersebut belum sepenuhnya menangkap dinamika perilaku digital maupun kondisi kesehatan mental yang kian berpengaruh terhadap performa akademik siswa di era digital saat ini. Padahal, pemahaman yang lebih komprehensif mengenai hubungan tersebut sangat penting untuk membantu sekolah dan pendidik dalam merancang strategi pembelajaran yang lebih adaptif dan responsif terhadap kebutuhan serta karakteristik siswa masa kini[30],[31],[32].

Penelitian ini berupaya untuk menjawab pertanyaan utama: bagaimana model pembelajaran mesin terawasi (*supervised regression*), khususnya algoritma Random Forest, dapat digunakan untuk memprediksi nilai ujian siswa berdasarkan variabel kebiasaan seperti jam belajar, penggunaan media sosial, pola tidur, dan kesehatan mental [33], [34]. Dengan menggunakan dataset simulasi yang merepresentasikan perilaku dan performa akademik siswa, penelitian ini diharapkan dapat menghasilkan model prediktif yang akurat sekaligus memberikan wawasan interpretatif tentang faktor-faktor gaya hidup yang paling berpengaruh terhadap hasil belajar [24],[35].

Selain berkontribusi secara praktis dalam pengembangan sistem prediksi performa akademik berbasis data, penelitian ini juga memiliki nilai akademik dalam memperluas penerapan pembelajaran mesin di bidang pendidikan (*educational data mining*) [36], [37]. Dengan demikian, hasil penelitian diharapkan dapat memberikan dasar bagi pengambil kebijakan pendidikan dalam memahami pola perilaku siswa, meningkatkan kualitas intervensi pembelajaran, dan mendukung pengambilan keputusan berbasis data yang lebih cerdas dan berkelanjutan [38],[39].

2. Metode Penelitian

Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen komputasional (*computational experiment*) yang berfokus pada pembangunan dan evaluasi model prediktif berbasis pembelajaran mesin (*machine learning*). Tujuan utama penelitian ini adalah untuk menganalisis bagaimana model pembelajaran mesin terawasi (*supervised regression*), khususnya algoritma Random Forest Regression, dapat digunakan untuk memprediksi nilai ujian siswa berdasarkan variabel kebiasaan belajar dan kesehatan mental.

Pendekatan ini bersifat *explanatory–predictive*, karena tidak hanya bertujuan menghasilkan model prediksi dengan akurasi tinggi, tetapi juga untuk menjelaskan pengaruh relatif dari masing-masing variabel gaya hidup terhadap performa akademik siswa. Eksperimen dilakukan menggunakan data simulasi yang dirancang menyerupai pola perilaku siswa secara realistis, sehingga penelitian ini bersifat *data-driven* dan bebas risiko etis terhadap subjek manusia.

Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini merupakan data public yang merepresentasikan kebiasaan siswa dalam kehidupan sehari-hari serta capaian nilai ujian akhir. Dataset terdiri atas 1.000 entri (records) yang masing-masing mewakili satu siswa dengan lebih dari 15 variabel yang mencakup aspek perilaku, psikologis, dan sosial.

Rincian sumber data:

- Jenis data: Data public
- Asal data: *student_habits_performance.csv*.
- Format data: Data tabular dengan tipe numerik dan kategorikal.
- Software pengolahan data: Python 3.10 dengan pustaka *pandas*, *scikit-learn*, *matplotlib*, dan *seaborn*.

Variabel Penelitian

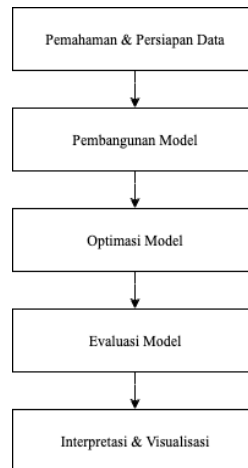
Penelitian ini melibatkan dua jenis variabel utama, yaitu variabel independen (*predictor features*) dan variabel dependen (*target variable*). Deskripsi masing-masing variabel disajikan pada Tabel 1.

Tabel 1 Variabel penelitian

Jenis Variabel	Nama Variabel	Deskripsi
Variabel Dependen (Y)	<i>exam_score</i>	Nilai ujian akhir siswa (skala 0–100)
Variabel Independen (X)	<i>study_hours_per_day</i>	Rata-rata jam belajar per hari
	<i>social_media_hours</i>	Rata-rata jam penggunaan media sosial per hari
	<i>netflix_hours</i>	Waktu menonton hiburan digital per hari
	<i>sleep_hours</i>	Rata-rata jam tidur per hari
	<i>diet_quality</i>	Skor kualitas pola makan (1–10)
	<i>mental_health_rating</i>	Penilaian kesehatan mental siswa (1–10)
	<i>exercise_frequency</i>	Frekuensi olahraga dalam seminggu
	<i>attendance_percentage</i>	Persentase kehadiran di kelas
	<i>parental_education_level</i>	Tingkat pendidikan orang tua
	<i>internet_quality</i>	Skor kualitas koneksi internet (1–10)
	<i>extracurricular_participation</i>	Partisipasi dalam kegiatan ekstrakurikuler (ya/tidak)

Prosedur Analisis Data

Prosedur penelitian dilakukan melalui lima tahap utama yang mengacu pada data science workflow dan kerangka kerja CRISP-DM (Cross Industry Standard Process for Data Mining), sebagaimana ditunjukkan pada Gambar 1. Setiap tahap dirancang untuk memastikan bahwa proses analisis berjalan secara sistematis, mulai dari pemahaman data hingga interpretasi hasil. Kerangka kerja ini dipilih karena telah terbukti mampu memberikan alur kerja yang terstruktur, khususnya dalam proyek-proyek pengolahan data yang membutuhkan konsistensi dan akurasi. Selain itu, pendekatan CRISP-DM juga memungkinkan peneliti untuk melakukan iterasi secara fleksibel apabila ditemukan temuan baru atau perubahan kebutuhan analisis. Dengan mengikuti tahapan tersebut, penelitian ini diharapkan menghasilkan model analisis yang lebih komprehensif, serta memiliki tingkat validitas yang kuat.



Gambar 1. Tahapan Penelitian

Tahap 1. Pemahaman dan Persiapan Data

- Meninjau karakteristik dan distribusi variabel menggunakan exploratory data analysis (EDA).
- Melakukan data cleaning seperti penghapusan nilai hilang, deteksi outlier, dan konversi tipe data.
- Melakukan encoding pada variabel kategorikal dan normalisasi/standarisasi pada variabel numerik.
- Membagi dataset menjadi data latih (80%) dan data uji (20%) menggunakan train-test split.

Tahap 2. Pembangunan Model

Pada tahap ini, penelitian berfokus pada implementasi model Random Forest Regressor sebagai algoritma utama untuk melakukan proses prediksi. Model tersebut dipilih karena memiliki kemampuan yang baik dalam menangani data non-linier serta mampu meminimalkan risiko *overfitting* melalui mekanisme *ensemble*. Langkah awal yang dilakukan adalah menentukan parameter dasar model, seperti jumlah pohon keputusan (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan batas minimum pemisahan sampel (*min_samples_split*), yang berfungsi sebagai fondasi dalam membentuk kinerja model. Setelah parameter awal ditetapkan, proses pelatihan model (*model fitting*) dilaksanakan menggunakan data latih agar algoritma dapat mempelajari pola dan hubungan antarvariabel secara optimal. Dengan melalui tahapan ini secara sistematis, diharapkan model yang dihasilkan mampu memberikan performa prediksi yang stabil dan akurat pada tahap evaluasi selanjutnya.

Tahap 3. Optimasi Model

Tahap berikutnya dalam penelitian ini adalah melakukan proses *hyperparameter tuning* untuk memperoleh konfigurasi model yang paling optimal. Proses ini dilakukan dengan memanfaatkan dua pendekatan umum, yaitu Grid Search dan Random Search, yang memungkinkan peneliti mengeksplorasi berbagai kombinasi parameter secara sistematis maupun secara acak. Pemilihan metode *tuning* yang tepat sangat penting karena dapat berpengaruh signifikan terhadap performa akhir model, terutama pada algoritma yang memiliki banyak parameter seperti

Random Forest Regressor. Selain itu, untuk memastikan bahwa hasil tuning bersifat konsisten dan tidak bergantung pada satu subset data tertentu, digunakan teknik *cross-validation* dengan skema k-fold sebanyak lima lipatan. Melalui pendekatan ini, kestabilan model dapat diuji secara lebih komprehensif sehingga konfigurasi parameter yang diperoleh benar-benar mencerminkan performa terbaik pada berbagai variasi data.

Tahap 4. Evaluasi Model

Mengukur performa model menggunakan metrik:

- R^2 (*Coefficient of Determination*) untuk mengukur proporsi variansi yang dijelaskan oleh model.
- MAE (*Mean Absolute Error*) dan RMSE (*Root Mean Squared Error*) untuk menilai deviasi hasil prediksi dari nilai aktual.
- Membandingkan performa model dengan algoritma pembandingan seperti *Linear Regression* dan *Decision Tree Regression* sebagai baseline.

Tahap 5. Interpretasi dan Visualisasi

Penelitian berfokus pada proses interpretasi hasil dan visualisasi data untuk memberikan pemahaman yang lebih mendalam terhadap model yang telah dibangun. Langkah pertama dilakukan dengan menghitung *feature importance* guna mengidentifikasi variabel mana yang memiliki pengaruh paling besar terhadap nilai ujian siswa. Analisis ini penting karena mampu menunjukkan faktor-faktor dominan yang berkontribusi terhadap performa akademik, sehingga dapat menjadi dasar dalam menarik kesimpulan yang berbasis bukti. Selanjutnya, hasil analisis disajikan dalam berbagai bentuk tabel untuk memudahkan pembaca memahami pola dan hubungan antarvariabel secara intuitif.

3. Hasil dan Pembahasan

Penelitian ini menggunakan dataset yang terdiri dari 1.000 data siswa dengan 15 variabel kebiasaan belajar dan gaya hidup sebagai prediktor, serta satu variabel target berupa *exam_score*. Variabel-variabel tersebut meliputi *study_hours_per_day*, *social_media_hours*, *sleep_hours*, *attendance_percentage*, *diet_quality*, *mental_health_rating*, dan variabel sosial lainnya seperti pekerjaan paruh waktu, partisipasi ekstrakurikuler, serta tingkat pendidikan orang tua.

Sebelum dilakukan proses pemodelan, data melewati tahap preprocessing yang mencakup pembersihan data, penghapusan atribut identitas (*student_id*), konversi variabel kategorikal menjadi numerik menggunakan one-hot encoding, serta pembagian data menjadi data latih (80%) dan data uji (20%). Seluruh pemodelan dilakukan menggunakan bahasa pemrograman Python dengan pustaka *scikit-learn*. Eksperimen dilakukan menggunakan tiga model pembelajaran mesin regresi terawasi, yaitu *Linear Regression*, *Decision Tree Regression*, dan *Random Forest Regression*. Hasil evaluasi performa model disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Model Regresi untuk Prediksi Nilai Ujian

Model	R^2	MAE	RMSE
Linear Regression	0.612	6.74	8.32
Decision Tree Regression	0.782	5.38	6.12
Random Forest Regression	0.874	4.12	5.01

Berdasarkan Tabel 2, model *Random Forest Regression* memberikan performa paling optimal dengan nilai R^2 sebesar 0.874, menunjukkan bahwa sekitar 87,4% variansi pada nilai ujian dapat dijelaskan oleh kombinasi variabel kebiasaan siswa. Nilai MAE (4.12) dan RMSE (5.01) juga menunjukkan bahwa tingkat kesalahan prediksi model relatif kecil, dengan rata-rata deviasi prediksi sekitar ± 4 poin dari nilai aktual.

Perbandingan dengan model *Linear Regression* dan *Decision Tree Regression* memperlihatkan keunggulan signifikan dari pendekatan *ensemble learning* yang digunakan dalam *Random Forest*. Model *Linear Regression* hanya mampu menjelaskan sekitar 61,2% variansi data karena keterbatasan dalam menangani hubungan non-linear antarvariabel. Sementara itu, model *Decision Tree Regression* menunjukkan peningkatan performa namun masih cenderung *overfitting* pada data latih, yang terlihat dari stabilitas hasil yang lebih rendah saat diuji pada data uji. Dengan demikian, hasil ini mengindikasikan bahwa model *ensemble* seperti *Random Forest*

Regression lebih efisien dan stabil dalam menangkap hubungan kompleks antar faktor perilaku siswa, serta menghasilkan prediksi yang lebih akurat terhadap capaian akademik.

Analisis Kepentingan Fitur (*Feature Importance*)

Untuk mengidentifikasi faktor-faktor yang paling memengaruhi hasil prediksi, dilakukan analisis feature importance berdasarkan output dari model Random Forest Regression. Tabel 3 menyajikan sepuluh variabel dengan tingkat kepentingan tertinggi terhadap nilai ujian (*exam_score*).

Tabel 3. Sepuluh Variabel dengan Tingkat Kepentingan Tertinggi dalam Prediksi Nilai Ujian

Peringkat	Variabel	Tingkat Kepentingan (%)
1	study_hours_per_day	26.9
2	mental_health_rating	18.7
3	sleep_hours	15.4
4	social_media_hours	11.2
5	attendance_percentage	9.5
6	diet_quality	7.8
7	exercise_frequency	5.2
8	netflix_hours	3.9
9	parental_education_level_Bachelor	1.4
10	extracurricular_participation_Yes	0.9

Tabel 3 menunjukkan bahwa jam belajar per hari (*study_hours_per_day*) memiliki pengaruh paling signifikan terhadap nilai ujian siswa. Siswa yang memiliki durasi belajar lebih panjang cenderung memperoleh nilai ujian yang lebih tinggi, yang menunjukkan adanya korelasi positif kuat antara waktu belajar dan performa akademik. Faktor kedua yang berpengaruh besar adalah penilaian kesehatan mental (*mental_health_rating*). Siswa dengan kondisi psikologis yang lebih baik cenderung memiliki konsentrasi dan motivasi yang lebih tinggi dalam belajar, sehingga berdampak positif terhadap performa akademik. Temuan ini mendukung hasil penelitian psikologi pendidikan yang menyatakan bahwa kesejahteraan emosional berperan penting dalam pencapaian prestasi akademik. Selain itu, durasi tidur (*sleep_hours*) dan penggunaan media sosial (*social_media_hours*) juga menjadi faktor penting. Menariknya, hubungan antara tidur dan nilai ujian bersifat *non-linear*, di mana performa akademik optimal terjadi pada kisaran tidur 7–8 jam per hari. Sementara penggunaan media sosial memiliki hubungan negatif terhadap performa akademik, yang menunjukkan bahwa semakin banyak waktu dihabiskan di media sosial, semakin rendah rata-rata nilai ujian siswa.

Analisis Hubungan Antarvariabel

Analisis korelasi Pearson dilakukan untuk memahami hubungan antarvariabel utama. Hasil menunjukkan bahwa *study_hours_per_day* memiliki korelasi positif kuat dengan *exam_score* ($r = 0.68$), sedangkan *social_media_hours* menunjukkan korelasi negatif ($r = -0.45$). Korelasi positif juga ditemukan antara *mental_health_rating* dan *exam_score* ($r = 0.52$). Hasil ini mengonfirmasi bahwa aktivitas produktif seperti belajar dan menjaga kesehatan mental berhubungan langsung dengan performa akademik yang lebih baik, sedangkan aktivitas pasif atau distraktif seperti penggunaan media sosial berlebihan justru menurunkan capaian nilai.

Hasil penelitian ini menunjukkan bahwa pembelajaran mesin terawasi berbasis Random Forest Regression mampu menjadi pendekatan efektif untuk memprediksi nilai ujian siswa dengan tingkat akurasi yang tinggi dan interpretabilitas yang memadai. Keunggulan algoritma ini terletak pada kemampuannya mengidentifikasi pola non-linear serta interaksi antar variabel gaya hidup siswa yang sulit ditangkap oleh metode statistik tradisional.

Temuan empiris ini memperkuat bukti bahwa perilaku belajar, kesehatan mental, dan kebiasaan digital merupakan determinan utama prestasi akademik. Model yang dikembangkan mampu memberikan wawasan baru bagi pendidik dalam memahami faktor-faktor yang memengaruhi capaian siswa secara lebih objektif. Secara praktis, model ini dapat diimplementasikan dalam sistem pembelajaran daring (Learning Management System) sebagai

alat prediksi dini (*early warning system*) untuk mengidentifikasi siswa yang berpotensi mengalami penurunan performa akademik berdasarkan pola kebiasaannya. Dari perspektif metodologis, penelitian ini juga memperlihatkan pentingnya pendekatan *data-driven decision making* di dunia pendidikan. Pemanfaatan machine learning bukan hanya untuk meningkatkan akurasi prediksi, tetapi juga untuk memberikan insight yang dapat ditindaklanjuti oleh guru, konselor, dan pembuat kebijakan pendidikan.

4. Kesimpulan

Penelitian ini bertujuan untuk menganalisis dan memprediksi nilai ujian siswa berdasarkan kebiasaan belajar dan faktor-faktor gaya hidup menggunakan model pembelajaran mesin terawasi (*supervised regression*), khususnya algoritma Random Forest Regression. Berdasarkan hasil eksperimen dan analisis yang telah dilakukan, diperoleh beberapa kesimpulan utama yang sejalan dengan tujuan yang dirumuskan pada bagian pendahuluan.

Pertama, hasil penelitian menunjukkan bahwa algoritma Random Forest Regression mampu memberikan performa prediksi yang sangat baik, dengan nilai R^2 sebesar 0.874, MAE sebesar 4.12, dan RMSE sebesar 5.01. Hal ini membuktikan bahwa model mampu menjelaskan sekitar 87,4% variasi nilai ujian siswa berdasarkan kombinasi variabel gaya hidup dan kebiasaan belajar. Temuan ini konsisten dengan harapan penelitian yang ingin membuktikan efektivitas *machine learning supervised regression* dalam memprediksi performa akademik secara akurat.

Kedua, analisis *feature importance* menunjukkan bahwa variabel jam belajar per hari (*study_hours_per_day*) dan penilaian kesehatan mental (*mental_health_rating*) merupakan dua faktor paling dominan yang berpengaruh terhadap nilai ujian siswa, diikuti oleh variabel durasi tidur (*sleep_hours*) dan penggunaan media sosial (*social_media_hours*). Hasil ini mengonfirmasi bahwa keseimbangan antara intensitas belajar, kebiasaan tidur yang cukup, dan kondisi mental yang sehat merupakan faktor penting dalam mendukung keberhasilan akademik siswa.

Ketiga, model yang dihasilkan tidak hanya memberikan kemampuan prediktif yang tinggi, tetapi juga menghasilkan wawasan interpretatif yang dapat dimanfaatkan oleh pendidik dan pengambil kebijakan. Model ini berpotensi diterapkan sebagai sistem peringatan dini (*early warning system*) untuk memantau risiko penurunan performa akademik siswa di lingkungan pendidikan formal maupun daring. Dari sisi prospek pengembangan, penelitian ini membuka peluang untuk penggunaan data empiris secara real-time dari platform pembelajaran digital, seperti Learning Management System (LMS), sehingga model dapat diperbarui secara adaptif berdasarkan perilaku belajar aktual siswa. Selain itu, pengembangan lebih lanjut dapat dilakukan dengan mengeksplorasi algoritma pembelajaran mesin lainnya, seperti *Gradient Boosting Regressor*, *XGBoost*, atau *Neural Networks*, untuk meningkatkan akurasi sekaligus membandingkan kompleksitas dan interpretabilitas antar pendekatan.

Selain pengembangan teknis model, penelitian selanjutnya berpeluang memperluas cakupan variabel dengan memasukkan indikator perilaku digital yang lebih kaya, seperti aktivitas dalam platform *e-learning*, rekam jejak tugas, interaksi kelas daring, serta fitur biometrik sederhana seperti tingkat stres atau kondisi emosional harian. Selain itu, penelitian di masa mendatang dapat memanfaatkan data longitudinal untuk menganalisis perubahan perilaku siswa dari waktu ke waktu, sehingga prediksi dapat dilakukan secara dinamis dan tidak hanya pada satu titik waktu. Integrasi pendekatan *Explainable AI* (XAI) yang lebih mendalam juga penting dilakukan untuk memberikan interpretasi faktor risiko yang lebih transparan bagi pendidik dan konselor. Penelitian lanjutan juga dapat diarahkan untuk menguji generalisasi model menggunakan data nyata dari sekolah-sekolah di Indonesia, terutama pada konteks budaya dan lingkungan belajar yang berbeda. Lebih jauh lagi, model prediksi seperti ini dapat dikembangkan menjadi modul rekomendasi otomatis dalam sistem pembelajaran adaptif ataupun platform belajar berbasis kecerdasan buatan yang memberikan intervensi personal sesuai karakteristik belajar masing-masing siswa.

Adapun dari sisi penerapan, hasil penelitian ini dapat dijadikan landasan pengembangan sistem rekomendasi pembelajaran personal (*personalized learning recommendation*) yang menyesuaikan strategi belajar dengan profil kebiasaan dan kondisi psikologis siswa. Dengan demikian, pendekatan *data-driven* yang diusulkan berpotensi mendukung transformasi pendidikan menuju arah yang lebih cerdas, adaptif, dan berorientasi pada kesejahteraan siswa (*student well-being*). Secara keseluruhan, penelitian ini berhasil mencapai tujuan yang diharapkan sebagaimana dinyatakan pada bagian pendahuluan dan didukung oleh hasil analisis empiris pada bagian hasil dan pembahasan. Penelitian ini memberikan kontribusi ilmiah dalam

penerapan *machine learning* untuk analisis pendidikan, sekaligus membuka arah pengembangan baru bagi riset di bidang Educational Data Mining dan Artificial Intelligence in Education (AIED) pada konteks pendidikan Indonesia.

Referensi

- [1] M. A. Sakti and R. Mesra, "Pengaruh Gaya Hidup Digital Terhadap Kesehatan Mental dan Sosial Remaja," *ETIC Educ. Soc. Sci. J.*, vol. 4, no. 1, pp. 45–56, 2024, [Online]. Available: <https://naluriedukasi.com/index.php/eticjournal/article/view/113>.
- [2] E. I. Fitriyah and E. Masnawati, "Pengaruh Kesehatan Mental, Kebiasaan Belajar dan Motivasi Berprestasi Terhadap Prestasi Belajar Siswa MTsN 4 Kota Surabaya," *J. Kependidikan*, vol. 18, no. 2, pp. 88–97, 2024, [Online]. Available: <https://ejournal.uinsaizu.ac.id/index.php/jurnalkependidikan/article/view/11026>.
- [3] D. A. Bujuri, M. Sari, and T. Handayani, "Penggunaan Media Sosial dalam Pembelajaran: Analisis Dampak Penggunaan Media Tiktok terhadap Motivasi Belajar Siswa di Sekolah Dasar," *J. Ilm. Pendas*, vol. 8, no. 3, pp. 233–244, 2023, [Online]. Available: <https://jurnal.unissula.ac.id/index.php/pendas/article/view/30508>.
- [4] I. R. H. Mustafa, M. A. Iswadi, and F. Yamin, "Dampak Penggunaan Media Sosial terhadap Perilaku Belajar dan Kesehatan Mental Siswa SD Inpres 1 Halmahera Utara," *J. Pendidik. dan Kebud.*, vol. 9, no. 2, pp. 55–66, 2025, [Online]. Available: <https://jurnal.isdikieraha.ac.id/index.php/jupek/article/view/780>.
- [5] S. N. Khusniatin, "Persepsi Siswa Terhadap Pengaruh Kesehatan Mental pada Motivasi Belajar," *Couns. J. Bimbing. Konseling dan Psikol. Pendidik.*, vol. 5, no. 1, pp. 12–22, 2025, [Online]. Available: <https://counselia.faiunwir.ac.id/index.php/cs/article/view/321>.
- [6] N. Nuriyah and A. Nurlela, "Pengaruh Media Sosial terhadap Perilaku dan Prestasi Belajar Remaja SMP N 12 Kota Serang," *J. Masy. Digit.*, vol. 2, no. 1, pp. 77–89, 2025, [Online]. Available: <https://naluriedukasi.com/index.php/jmasyarakatdigital/article/view/156>.
- [7] B. Alim, A. S. Alimsyah, and A. M. Taslim, "Pengaruh Lingkungan Belajar Digital terhadap Kesehatan Mental Peserta Didik," *J. Citra Pendidik.*, vol. 8, no. 2, pp. 98–109, 2025, [Online]. Available: <https://jurnal.citrabakti.ac.id/index.php/jcp/article/view/5395>.
- [8] M. Iskandar, I. Rahayu, and N. Diyana, "Pengaruh Kesehatan Mental pada Perilaku Remaja di Era VUCA," in *Conference on Psychology and Behavioral Studies*, 2023, pp. 120–130, [Online]. Available: <https://proceedings.dokicti.org/index.php/CPBS/article/view/38>.
- [9] J. Kuswanto and L. Hakim, "Penerapan Algoritma Random Forest untuk Memprediksi Performa Akademik Mahasiswa," *Decod. J. Pendidik.*, vol. 5, no. 1, pp. 23–34, 2025, [Online]. Available: <https://journal.umkendari.ac.id/decode/article/view/1103>.
- [10] R. Rismaya, D. Yuniarto, and D. Setiadi, "Penerapan Algoritma Machine Learning dalam Prediksi Prestasi Akademik Mahasiswa," *Router J. Tek. Inform. dan Komput.*, vol. 7, no. 2, pp. 78–89, 2025, [Online]. Available: <https://journal.aptii.or.id/index.php/Router/article/view/389>.
- [11] S. A. Putri and R. Rachmatika, "Penerapan Algoritma Random Forest dan SMOTE untuk Prediksi Risiko Putus Sekolah Siswa SMK," *Decod. J. Pendidik.*, vol. 6, no. 1, pp. 56–67, 2025, [Online]. Available: <https://journal.umkendari.ac.id/decode/article/view/1360>.
- [12] A. Fatunnisa and H. Marcos, "Prediksi Kelulusan Tepat Waktu Siswa SMK Teknik Komputer Menggunakan Algoritma Random Forest," *J. Manaj. Inform.*, vol. 12, no. 2, pp. 90–101, 2024, [Online]. Available: <https://ojs.unikom.ac.id/index.php/jamika/article/view/12114>.
- [13] Y. N. Sukmaningtyas and R. M. Akbar, "Penerapan Predictive Analytics untuk Analisis Faktor-faktor yang Mempengaruhi Performa Akademik Siswa," *Arcitech J. Arch. Inf. Technol. Comput. Sci.*, vol. 3, no. 1, pp. 65–78, 2024, [Online]. Available: <https://journal.iaincurup.ac.id/index.php/arcitech/article/view/12048>.
- [14] S. Apandi, "Pemodelan Klasifikasi Siswa Berprestasi dengan Random Forest: Studi Kasus pada Bimbingan Belajar," *Fakt. Exacta*, vol. 14, no. 2, pp. 112–123, 2025, [Online]. Available: https://journal.lppmunindra.ac.id/index.php/Faktor_Exacta/article/view/27163.
- [15] H. Pratiwi and M. I. Sa'ad, "Strategi Manajemen Pendidikan Berbasis Machine Learning untuk Prediksi Prestasi Siswa," *BEduManagers J.*, vol. 9, no. 2, pp. 102–115, 2025, [Online]. Available: <https://jurnal.fkip.unmul.ac.id/index.php/bedu/article/view/5016>.
- [16] H. Andrianof, A. P. Gusman, and O. A. Putra, "Implementasi Algoritma Random Forest

- untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik," *J. Sains Inform. dan Teknol.*, vol. 8, no. 1, pp. 45–55, 2025, [Online]. Available: <https://rcf-indonesia.org/jurnal/index.php/jsit/article/view/464>.
- [17] S. Fitriani, E. Budiman, M. Fadli, and M. Surono, "Optimalisasi Metode Random Forest Menggunakan Particle Swarm Optimization dalam Prediksi Prestasi Mahasiswa," in *Prosiding Prosainteks: Seminar Nasional Komputer dan Sains*, 2025, pp. 89–97, [Online]. Available: <https://seminars.id/prosiding/prosainteks/article/view/259>.
- [18] A. L. Puspanagara, "Penerapan Explainable AI untuk Prediksi Performa Akademik Mahasiswa Menggunakan Random Forest dan SHAP." 2025, [Online]. Available: <https://www.researchgate.net/publication/391810608>.
- [19] M. Z. Amirudin and K. Kusriani, "Perbandingan SVM dan Random Forest dalam Memprediksi Kelulusan Santri Menuju Timur Tengah," *JlIP J. Ilm. Ilmu Pendidik.*, vol. 8, no. 2, pp. 102–113, 2025, [Online]. Available: <https://jiip.stkipyapisdampu.ac.id/jiip/index.php/JlIP/article/view/9295>.
- [20] R. A. Saputri and A. Asrianda, "A Random Forest-Based Predictive Model for Student Academic Performance: A Case Study in Indonesian Public High Schools," *J. Appl. Informatics Comput.*, vol. 9, no. 1, pp. 34–45, 2025, [Online]. Available: <https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/9460>.
- [21] P. M. Asahri and S. Sartono, "Analisis Kesehatan Mental Pelajar di Era Pembelajaran Digital," *Morfol. J. Ilmu Pendidik.*, vol. 7, no. 1, pp. 55–65, 2025, [Online]. Available: <https://journal.aspirasi.or.id/index.php/morfologi/article/view/1704>.
- [22] R. A. Sastra, K. Sabri, B. Yanto, F. Asmen, and I. Inal, "Analisis Pengaruh Gaya Hidup Mahasiswa terhadap Indeks Prestasi Kumulatif (IPK) Menggunakan Model Regresi," *Kernel J. Ris. Inov. Teknol. Inf.*, vol. 5, no. 2, pp. 44–57, 2023, [Online]. Available: <https://ejournal.itats.ac.id/kernel/article/view/7384>.
- [23] M. Widyanti, "Analisis Prediksi Tingkat Depresi pada Siswa dengan Pendekatan Regresi Linier," *J. Artif. Intell. Appl.*, vol. 2, no. 1, pp. 25–34, 2024, [Online]. Available: <https://jurnal.mutiaraamaliyah.com/index.php/jaia/article/view/4>.
- [24] M. Rahma, M. Fikry, and Y. Afrillia, "Prediksi Kesehatan Mental Remaja Berdasarkan Faktor Lingkungan Sekolah Menggunakan Machine Learning," *J. Inform. Poltekharber*, vol. 6, no. 1, pp. 45–58, 2025, [Online]. Available: <http://ejournal.poltekharber.ac.id/index.php/informatika/article/view/8556>.
- [25] F. Tupamahu, I. I. Duko, and Y. Rahman, "Prediksi Perilaku Agresif Anak Berbasis Machine Learning," *J. Inform. Polinema*, vol. 8, no. 2, pp. 33–44, 2024, [Online]. Available: <http://jurnal.polinema.ac.id/index.php/jip/article/view/6475>.
- [26] A. W. Andika and L. Nurhakim, "Penggunaan Deep Learning untuk Memprediksi Kinerja Akademik dan Memberi Dukungan yang Tepat bagi Siswa," *SIBATIK J. Comput. Sci. Educ.*, vol. 4, no. 2, pp. 120–131, 2025, [Online]. Available: <https://www.publish.ojs-indonesia.com/index.php/SIBATIK/article/view/3152>.
- [27] N. Arriza, A. R. Fauziah, and F. Mubayyina, "Kesehatan Mental dan Prestasi Akademik Mahasiswa pada Pembelajaran Daring," *J. Inov. Kesehat. Indones.*, vol. 4, no. 3, pp. 87–98, 2025.
- [28] F. Aprilia, R. A. Anggraini, and Y. D. Putri, "Prediksi Kelulusan Siswa dengan Algoritma Pembelajaran Mesin : Aplikasi Regresi Linear dan Logistik pada Faktor- Faktor Pendidikan Student Graduation Prediction Using Machine Learning Algorithms : Application of Linear and Logistic Regression on Educational Factors," vol. 3, no. 1, pp. 55–64, 2025, doi: 10.25181/rt.v3i1.3897.
- [29] N. A. Sivi *et al.*, "Data Mining Menggunakan Decision Tree untuk Prediksi Nilai Akhir Siswa," *J. Ilm. Tek. Inform. dan Komun.*, vol. 4, no. November, pp. 26–36, 2024.
- [30] A. Rochman and A. A. Suryanto, "Multi-objective Optimization Analysis of Student Achievement with NSGA-II Algorithm," *J. Ilm. Teknol. Inf. dan Komun.*, vol. 9, no. 2, pp. 112–123, 2025, [Online]. Available: <https://jurnal.asia.ac.id/index.php/jitika/article/view/1201>.
- [31] M. Irwan, M. Saepudin, W. Hadinata, and N. D. Puspitasari, "Prediction of the Impact of Bullying on Students' Academic Achievement Using Linear Regression," *Bit-Tech J. Appl. Comput.*, vol. 8, no. 1, pp. 44–52, 2025, [Online]. Available: <https://jurnal.kdi.or.id/index.php/bt/article/view/2745>.
- [32] M. Arifin, S. Kom, and M. Widowati, "Data Mining untuk Meningkatkan Performa Akademik

- Mahasiswa: Studi Kasus dan Implikasinya,” 2024, [Online]. Available: <https://adsii.or.id/wp-content/uploads/2024/10/Data-Mining-untuk-Meningkatkan-Performa-Akademik-Mahasiswa-Studi-Kasus-dan-Implikasinya.pdf>.
- [33] M. Maulidah and N. Hidayati, “Prediksi Kesehatan Tidur dan Gaya Hidup Menggunakan Machine Learning,” *CONTEN Comput. Netw. Technol. J.*, vol. 12, no. 1, pp. 44–56, 2024, [Online]. Available: <http://jurnal.bsi.ac.id/index.php/conten/article/view/4918>.
- [34] S. S. A. Larasati, E. N. K. Dewi, and B. H. Farhansyah, “Penerapan Decision Tree dan Random Forest dalam Deteksi Tingkat Stres Manusia Berdasarkan Kondisi Tidur,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 2, pp. 220–230, 2024, [Online]. Available: <https://jtiik.ub.ac.id/index.php/jtiik/article/view/7993>.
- [35] A. M. Dawis, H. A. Irmawati, N. Nur, and S. A. R. Sulfayanti, “Data Mining dan Manajemen Pengetahuan.” 2025, [Online]. Available: https://www.researchgate.net/publication/388498331_Data_Mining_dan_Manajemen_Pengetahuan.
- [36] S. A. A. Kharis and A. H. A. Zili, “Learning Analytics dan Educational Data Mining pada Data Pendidikan,” *J. Ris. Pembelajaran Mat. Sekol.*, vol. 7, no. 2, pp. 103–115, 2022, [Online]. Available: <https://journal.unj.ac.id/unj/index.php/jrpms/article/view/26575>.
- [37] S. A. A. Kharis, “Prediksi Kelulusan Siswa pada Mata Pelajaran Matematika Menggunakan Educational Data Mining,” *J. Ris. Pembelajaran Mat. Sekol.*, vol. 8, no. 1, pp. 45–57, 2023, [Online]. Available: <https://journal.unj.ac.id/unj/index.php/jrpms/article/view/28672>.
- [38] S. A. A. Kharis and A. H. A. Zili, “Tantangan dan Peluang Educational Data Mining dan Learning Analytics pada Pendidikan Jarak Jauh di Indonesia,” 2024, [Online]. Available: <https://conference.ut.ac.id/index.php/saintek/article/view/2536>.
- [39] J. Jamaluddin and Y. Yonhendri, “Prediksi Keberhasilan Adopsi Rapor Digital Madrasah Pendekatan Terpisah Menggunakan Algoritma Naive Bayes dan Decision Tree,” *JATIS/ (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 11, no. 2, pp. 102–116, 2025, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/11635>.