



Penerapan Algoritma Random Forest untuk memprediksi Performa Akademik Mahasiswa

Joko Kuswanto^{1*}, Lukmanul Hakim¹

¹Program Studi Informatika, Universitas Baturaja, Indonesia.

Artikel Info

Kata Kunci:

Machine Learning;
Performa Akademik;
Prediksi;
Random Forest.

Keywords:

Machine Learning;
Academic Performance;
Predictions;
Random Forest.

Riwayat Artikel:

Submitted: 12 Maret 2025

Accepted: 30 Maret 2025

Published: 31 Maret 2025

Abstrak: Perkembangan teknologi yang pesat, khususnya di bidang pendidikan, membuka peluang untuk meningkatkan kualitas pembelajaran dan hasil akademik. *Machine Learning*, sebagai salah satu teknologi kecerdasan buatan, memiliki kemampuan untuk membuat prediksi berbasis pola data. Penelitian ini bertujuan untuk menerapkan algoritma *Random Forest* dalam memprediksi performa akademik mahasiswa Program Studi Informatika Universitas Baturaja. Data yang digunakan mencakup variabel akademik (nilai mata kuliah semester 1 dan 2) dan demografis (jenis kelamin, kelas, angkatan), dengan target prediksi berupa Indeks Prestasi (IP) semester 3. Penelitian dilakukan menggunakan metode kuantitatif dengan alat *Google Colab* dan teknik regresi. Evaluasi model dilakukan menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *R-squared* (R^2). Hasil penelitian menunjukkan bahwa model *Random Forest* menghasilkan MAE sebesar 0.27, MSE sebesar 0.2387, dan R^2 sebesar 0.3128, yang menunjukkan akurasi prediksi yang memadai namun dapat ditingkatkan lebih lanjut. Penelitian ini memberikan kontribusi pada pengembangan strategi pembelajaran berbasis data untuk meningkatkan kualitas pendidikan di Program Studi Informatika Universitas Baturaja.

Abstract: Rapid technological developments, especially in the field of education, open up opportunities to improve the quality of learning and academic outcomes. *Machine Learning*, as one of the artificial intelligence technologies, has the ability to make predictions based on data patterns. This study aims to implemenation the *Random Forest* algorithm in predicting the academic performance of students of the Informatics Study Program, University of Baturaja. The data used included academic variables (grades of semester 1 and 2 courses) and demographics (gender, class, class), with a prediction target in the form of an Achievement Index (IP) for semester 3. The research was conducted using quantitative methods with *Google Colab* tools and regression techniques. Model evaluation was carried out using *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), and *R-squared* (R^2). The results showed that the *Random Forest* model produced an MAE of 0.27, an MSE of 0.2387, and an R^2 of 0.3128, which showed adequate prediction accuracy but could be further improved. This research contributes to the development of data-based learning strategies to improve the quality of education in the Informatics Study Program, University of Baturaja

Corresponding Author:

Joko Kuswanto

Email: ko.8515@gmail.com

PENDAHULUAN

Perkembangan teknologi merupakan suatu kondisi yang tidak dapat dihindari, dan justru harus selalu diikuti karena teknologi itu sendiri dapat membantu dalam berbagai aspek dan bidang kehidupan. Begitu juga dengan bidang pendidikan yang tidak dapat lepas dari arus perkembangan teknologi ini. Hal ini dapat disebabkan karena teknologi itu sendiri berperan untuk meningkatkan kualitas pembelajaran dan hasil akademik mahasiswa. Salah satu hasil dari perkembangan teknologi yang biasa diterapkan di bidang pendidikan adalah *machine learning*. *Machine Learning* merupakan cabang dari kecerdasan buatan yang memungkinkan komputer untuk dapat belajar dan berkembang melalui data dan mengidentifikasi pola-pola tersembunyi yang sulit dideteksi oleh manusia (Waruwu et al., 2024) (Rismaya et al., 2025)(Marlina Haiza et al., 2023)(Novrian et al., 2024). Kemampuan inilah yang membuat *machine learning* sering digunakan didalam bidang Pendidikan. Salah satu penerapan *machine learning* adalah untuk memprediksi prestasi akademik mahasiswa, yang bertujuan untuk membantu pendidik memahami pola belajar mahasiswa dan mengambil tindakan proaktif untuk meningkatkan hasil belajar mahasiswa sehingga institusi pendidikan tinggi dapat mengambil keputusan yang lebih tepat terkait intervensi pendidikannya.

Beberapa penelitian terdahulu yang telah menggunakan teknik *machine learning* seperti penelitian yang dilakukan oleh Santoso (Santoso et al., 2020) untuk Prediksi Diagnosis Diabetes dengan menggunakan algoritma Evolving Artificial Neural Network (EANN) dengan hasil akurasi sebesar 83,55%. Penelitian Pitaloka (Pitaloka et al., 2024) untuk memprediksi banjir. Dalam penelitian ini mengkombinasikan algoritma Deep Neural Investigation Network (DNIN), Convolutional Neural Network (CNN) dan Bidirectional Long Short Term Memory (BiLSTM) dengan hasil penelitian diperoleh akurasi 0.9496, precision 0.9346, Recall 0.9833, F1-score 0.9547, AUC 0.9804 dan ROC 0.99. Penelitian Hendri (Hendri et al., 2021) mengategorikan sampah plastik rumah tangga dengan hasil penelitian bahwa sampah plastik dapat dipisahkan menggunakan Deep Learning dengan mengenali citra dan pola dari gambar yang diproses. Penelitian Wardhana (Wardhana et al., 2023) untuk prediksi tingkat kasus penyakit di Indonesia.

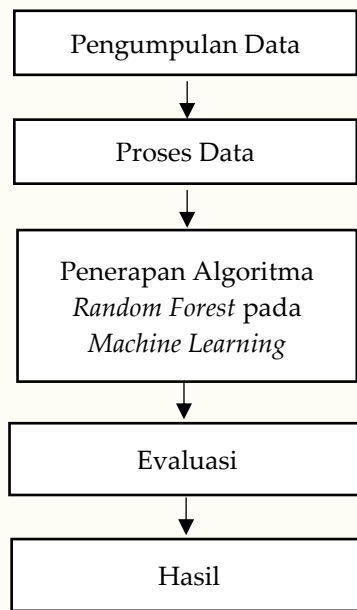
Dalam penelitian ini menerapkan algoritma *Random Forest* untuk memprediksi performa akademik mahasiswa. Algoritma *Random Forest* merupakan salah satu algoritma dalam *Machine Learning*. Algoritma ini memiliki kemampuan untuk menghasilkan suatu prediksi yang akurat melalui proses pengolahan data dengan membentuk beberapa pohon keputusan (Efendi & Zyen, 2024)(Nugroho & Husin, 2022)(Adiyati, 2025). *Random Forest* dapat menggabungkan efek non-linier, dan lebih unggul daripada metode alternatif dalam memodelkan interaksi kompleks ketika interaksi tidak, atau tidak dapat, ditentukan sebelumnya (Dicky Huang & Ito Wasito, 2024). Dalam konteks di dunia pendidikan, *Random Forest* dapat digunakan untuk menganalisis dan memprediksi performa akademik seperti nilai siswa (Kusworo et al., 2024), kehadiran dan aktivitas belajar lainnya.

Penelitian terkait dengan prediksi performa akademik juga telah dilakukan sebelumnya seperti penelitian Budiman (Budiman et al., 2018) menggunakan algoritma C4.5 untuk melakukan klasifikasi terhadap evaluasi aktivitas akademik mahasiswa. Hasilnya adalah algoritma C4.5 dengan rasio data 90% data latih, dan 10% data uji memiliki tingkat akurasi sebesar 78.58%, dan nilai True Positive Rate sebesar 76.72%. Penelitian lain yang dilakukan oleh Purwati (Purwati & Dwi Januanti, 2021), dalam penelitian ini dilakukan prediksi tingkat kelulusan mahasiswa dengan algoritma Naive Bayes. Penelitian Galopo dan Eugene (Perez & Perez, 2021). Dalam penelitian ini dilakukan prediksi mahasiswa yang akan lulus atau gagal pada suatu universitas dengan menggunakan menggunakan algoritma Naïve Bayes. Tingkat akurasi yang dihasilkan adalah sebesar 83.77%, dengan presisi sebesar 80.46%, dan recall sebesar 83.33%. Penelitian yang dilakukan oleh Kusworo (Kusworo et al., 2024), penelitian ini menggunakan algoritma random forest dan menggunakan Dataset nilai siswa. Model Random Forest berhasil diperoleh melalui pelatihan dan evaluasi model menggunakan data uji, menunjukkan akurasi sebesar 93,33%. Penggunaan algoritma random forest untuk memprediksi nilai akhir semester siswa dapat menjadi cara yang tepat untuk menghasilkan akurasi yang tinggi untuk kepentingan pendidik dalam menangani siswanya.

Berdasarkan pemaparan diatas terkait dengan *machine learning* khususnya algoritma *random forest* maka tujuan penelitian ini adalah untuk memprediksi performa akademik mahasiswa dengan menggunakan algoritma *Random Forest* pada program studi Informatika, Universitas Baturaja. Diharapkan melalui penelitian ini dapat ditemukan solusi untuk menyusun strategi pembelajaran yang lebih tepat guna dan efisien.

METODE

Tahapan dalam penelitian ini adalah pengumpulan data, proses data, pembuatan dan pelatihan model prediksi, pengujian model, dan evaluasi. Tahapan penelitian ini bisa dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

1. Pengumpulan Data

Tahap ini melibatkan pengumpulan data yang relevan dari sumber yang sudah ditentukan. Langkah-langkahnya meliputi:

- Akses ke data akademik: Mengakses data dari sistem informasi akademik Program Studi Informatika, Universitas Baturaja dengan persetujuan dari pihak akademik.
- Pengumpulan data mahasiswa: Mengambil data akademik mahasiswa Angkatan 2021, 2022 dan 2023 semester 1 dan 2, yang mencakup nilai akhir disetiap mata kuliah dan data demografis yang berkaitan dengan faktor performa akademik
- Validasi data: Memastikan data yang diambil valid dan sesuai dengan keperluan penelitian, serta memeriksa kelengkapan data.

2. Proses Data

Pada tahap ini, data yang sudah dikumpulkan akan disusun dan diolah menjadi format yang siap digunakan untuk prediksi pada *Machine Learning*. Tahapannya dengan melakukan

- Preprocessing* data: Melakukan pembersihan data untuk mengatasi nilai-nilai yang hilang atau duplikat. Tahapan ini mencakup:
 - Penanganan Data yang Kosong (Missing Values): Menggunakan metode seperti imputasi atau penghapusan data yang tidak relevan.
 - Pengubahan Data Non-Numerik Menjadi Numerik: Misalnya, mengonversi jenis kelamin (laki-laki atau perempuan) menjadi angka (0 dan 1).

- 3) Normalisasi atau Standarisasi Data: Untuk memastikan bahwa semua fitur memiliki skala yang serupa.
- b. Pemilihan fitur (*feature selection*): Menentukan fitur-fitur yang paling relevan dari data yang akan digunakan dalam model prediksi.
3. Penerapan Algoritma Random Forest. Tahap ini melibatkan penerapan algoritma *Random Forest* untuk memprediksi performa akademik mahasiswa berdasarkan data yang telah disusun. Langkah-langkahnya meliputi:
 - a. Pembangunan model: Menggunakan pustaka *Scikit-learn* untuk membangun model prediksi menggunakan algoritma *Random Forest*. Model ini dilatih menggunakan data mahasiswa di semester awal (semester 1 dan 2) sebagai input untuk memprediksi performa akademik di semester berikutnya.
 - b. Pembagian data (*train-test split*): Data dibagi menjadi data pelatihan (*training*) dan data pengujian (*testing*) dengan rasio tertentu (dalam penelitian ini rasion yang digunakan adalah 70% untuk pelatihan dan 30% untuk pengujian).
 - c. Pelatihan model (*training*): Melakukan pelatihan model dengan menggunakan data pelatihan untuk membuat model belajar dari pola-pola yang ada dalam data akademik mahasiswa.
 - d. Pengujian model (*testing*): Menggunakan data pengujian untuk mengevaluasi kinerja model dalam memprediksi performa akademik.
4. Evaluasi: Mengevaluasi hasil prediksi model menggunakan metrik seperti Mean absolute error (MAE), Mean Squared error (MSE) dan R-Squared.

a. *Mean Absolute Error (MAE)*

Digunakan untuk mengukur rata-rata kesalahan absolut antara nilai prediksi (\hat{y}_i) dan nilai aktual (y_i). Kesalahan absolut adalah sebuah jarak antara nilai dari prediksi dan nilai aktual tanpa memperhatikan arah (Positif atau Negatif), teknik ini biasa dirumuskan dengan (Suryanto, 2019) (Febrian et al., 2024):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Di mana:

- y_i adalah nilai aktual pada data ke- i
- \hat{y}_i adalah nilai prediksi pada data ke- i
- n adalah jumlah data.

Contoh kasus: nilai target adalah [10, 15, 20] dan prediksi adalah [12, 14, 18]:

$$MAE = \frac{|10 - 12| + |15 - 14| + |20 - 18|}{3} = \frac{2 + 1 + 2}{3} = 1.67$$

b. *Mean Squared Error (MSE)*

Digunakan untuk mengukur rata-rata dari kuadrat perbedaan antara nilai prediksi (\hat{y}_i) dan nilai aktual (y_i). Dengan mengkuadratkan kesalahan, MSE dapat memberikan penalti lebih besar untuk kesalahan besar, sehingga lebih sensitif terhadap *outlier*. Metode ini dapat dirumuskan dengan (Lestari & Astuti, 2022) (Adiyati, 2025; Nur et al., 2023):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Contoh kasus: Menggunakan nilai [10, 15, 20] dan prediksi [12, 14, 18] maka:

$$MSE = \frac{(10 - 12)^2 + (15 - 14)^2 + (20 - 18)^2}{3} = \frac{4 + 1 + 4}{3} = 3$$

c. *R2 (R-Squared atau Koefisien Determinasi)*

Metode ini digunakan untuk mengukur sejauh mana model regresi dapat menjalankan variasi dalam data target. Nilai R^2 menunjukkan proporsi variasi data yang dapat dijelaskan oleh model dibandingkan dengan variasi total dalam data. Metode ini dapat dirumuskan dengan:

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}}$$

Dimana:

- $SS_{residual} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ (Jumlah kuadrat kesalahan).
 - $SS_{total} = \sum_{i=1}^n (y_i - \bar{y})^2$ (Jumlah kuadrat dari deviasi nilai aktual terhadap rata-rata).
5. Hasil dari model dievaluasi dan diinterpretasikan untuk memahami bagaimana variabel penelitian di awal (semester 1 dan 2) berhubungan dengan prediksi performa akademik di semester berikutnya.

HASIL DAN PEMBAHASAN

1) Pengolahan Data

Pada penelitian ini data yang penulis gunakan adalah data mahasiswa Program Studi Informatika angkatan 2021, 2022 dan 2023 di Universitas Baturaja, dengan total sebanyak 312 data mahasiswa. Setiap individu mahasiswa memiliki 10 fitur penilaian yang mencakup data akademik dan data demografis. Dari jumlah tersebut, 218 data berisi nilai IP semester 2 yang digunakan sebagai variabel target untuk pelatihan model. Sedangkan 94 data lainnya merupakan data mahasiswa yang akan digunakan untuk prediksi atau sebagai data pengujian. Dataset penelitian ini dibuat menggunakan aplikasi Microsoft Excel, dan data tersebut disimpan kedalam format *.xlsx*. Bentuk dataset yang digunakan dalam penelitian ini adalah sebagai berikut:

No.	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	No.	Nama	NPM	Kelas	Angkatan	Nilai I	Nilai II	Nilai III	Nilai IV	Nilai V	Nilai VI	Nilai VII	Jenis Kelamin	IP Semester 2
2	1	EFRIA ANGARA	2133003	A	2021	78.7	78	88.2	86	70.1	80	76.5	Pria	3.74
3	2	KURNIAWAN	2133004	A	2021	71.7	75	86.8	86	87.8	80	81	Pria	4.00
4	3	SITI NURJANAH	2133007	A	2021	71	77	82.1	86	80.6	80	71.3	Wanita	3.74
5	4	IRSYAD RISQULLOH	2133012	A	2021	81.5	75	83.2	86	75.4	80	72.3	Pria	3.58
6	5	ELLA SUTRA	2133013	A	2021	72.8	72	83.7	86	90	80	74.3	Wanita	3.42
7	6	MUHAMMAD AMAR KHULUQUN'ULYA	2133015	A	2021	72.8	57	85.5	86	74.4	78.3	70.8	Pria	3.58
8	7	KENYA SELVIN NATALIA	2133016	A	2021	78	75	82.2	86	90	80	71.5	Wanita	3.68
9	8	FAKHRAH RODHIYAH	2133018	A	2021	76.3	81	82.9	86	89.3	80	75.3	Wanita	3.84
10	9	LUKMANUL HAKIM	2133020	A	2021	74.5	81	85.4	75	91.7	80	81.8	Pria	4.00
11	10	SAPUTRA WIJAYA	2133021	A	2021	74.5	73	87.6	86	84.6	80	80	Pria	3.74
12	11	MARETHA ANGELINA SIANTURI	2133023	A	2021	76.3	75	85.1	86	95	80	82	Wanita	3.89
13	12	ARYA NURHUDA	2133024	A	2021	72.8	70	82	86	81.3	60.3	75.3	Pria	3.42
14	13	NURULIA	2133025	A	2021	73.5	73	83	86	83.3	85	79	Wanita	3.58
15	14	RIA FRANSISKA	2133027	A	2021	72.8	77	85.6	75	85.3	80	70.8	Wanita	3.58
16	15	ELTA SEPTIANA	2133028	A	2021	75.2	74	83.3	86	81.3	80	70	Wanita	3.68
17	16	REXSY OTTA FORNANDO	2133029	A	2021	67.5	73	81.1	75	73.8	60.3	70.5	Pria	3.26
18	17	MIZAKKI ANUGERA HIDAYAT	2133030	A	2021	74.5	67	82.6	86	73.2	80	74.8	Pria	3.16
19	18	RIO AGUSANDA	2133035	A	2021	57	66	80	86	70.6	60	71.5	Pria	2.58
20	19	SDITI VALOVI	2133036	A	2021	78	74	80.8	86	77.9	80	74	Wanita	3.47
21	20	GITA LIA SAPUTRI	2133040	A	2021	78	71	80.2	86	70.7	80	73.5	Wanita	3.42
22	21	DHAVID DWILESON DAYA	2133045	A	2021	67.5	53	72.8	86	70.2	60	62	Pria	3.32

Gambar 1. Dataset Penelitian

a. Fitur-Fitur Dataset

Fitur pada *Dataset* adalah sebuah atribut atau kolom yang digunakan sebagai masukan (*input*) untuk melatih model agar dapat memahami hubungan antara data *input* dengan variabel target. Fitur juga merepresentasikan karakteristik atau informasi penting dari data yang membantu model dalam membuat sebuah prediksi. Pada penelitian ini fitur yang penulis gunakan terdiri dari 10 variabel, yaitu sebagai berikut:

- 1) Nilai per Mata Kuliah: Terdapat 7 mata kuliah yang masing-masing memiliki nilai yang diperoleh mahasiswa selama mengikuti mata kuliah tersebut
- 2) Kelas: Fitur ini mengkategorikan mahasiswa ke dalam dua kelas, yaitu kelas regular A dan Kelas regular B
- 3) Angkatan: Tahun angkatan mahasiswa yang menunjukkan kapan mahasiswa tersebut mulai kuliah.
- 4) Jenis Kelamin: Fitur ini mengidentifikasi jenis kelamin mahasiswa, yaitu Pria atau Wanita.

b. Variabel Target

Variabel target atau yang dikenal juga sebagai label (*output*) adalah sebuah atribut atau kolom dalam *dataset* yang ingin diprediksi oleh model *machine learning*. Didalam penelitian ini, variabel target yang penulis gunakan adalah IP semester 3 mahasiswa. IP semester 3 dipilih sebagai target

prediksi karena mencerminkan performa akademik mahasiswa setelah melewati dua semester. Dari total 196 data, hanya 75 data yang memiliki IP semester 3 yang dapat digunakan sebagai label untuk pelatihan model. Sisa data sebanyak 121 adalah data mahasiswa yang akan diprediksi menggunakan model yang dibangun.

2) Proses Data

Setelah dilakukan pengolahan dataset, langkah selanjutnya adalah tahap preprocessing data yang terdiri dari beberapa tahapan:

a. Pembersihan Data

Tahap awal dalam pra-pemrosesan data adalah pembersihan *dataset* untuk memastikan bahwa data yang digunakan bebas dari masalah seperti nilai kosong atau data redundan.

b. Encoding Data

Beberapa fitur dalam dataset berupa data kategorikal yang perlu diubah menjadi bentuk numerik agar dapat digunakan oleh *algoritma Random Forest*. Karena disini memiliki beberapa fitur yang bersifat kategorikal maka harus menggunakan *Label Encoding* untuk mengubah fitur menjadi nilai numerik. Untuk data Jenis Kelamin, dikodekan menjadi angka, misalnya: Laki-laki = 0, Perempuan = 1. untuk data Kelas: Dikodekan menjadi dua kolom biner, yaitu Kelas A (Regular) dan Kelas B (Weekend). Contoh: Mahasiswa di Kelas A = [1, 0], Mahasiswa di Kelas B = [0, 1]. Untuk Data Angkatan: Dikodekan menjadi kolom biner berdasarkan tahun angkatan mahasiswa.

c. Normalisasi Fitur Numerik

Setelah data kategorikal dienkoding, fitur numerik seperti Nilai Mata Kuliah dan Usia dinormalisasi untuk memastikan setiap fitur memiliki skala yang seimbang. Normalisasi ini dilakukan agar model tidak memberikan bobot berlebih pada fitur dengan skala nilai yang lebih besar.

3) Penerapan Algoritma Random Forest

Setelah selesai melakukan *preprocessing* selanjutnya melakukan pelatihan dan pengujian terhadap dataset, namun sebelumnya perlu membagi dataset menjadi dua kelompok berdasarkan ketersediaan label pada target (IP Semester 3):

- Data Berlabel. Sebanyak 75 data mahasiswa yang memiliki nilai IP semester 3 digunakan untuk melatih model dan evaluasi.
- Data Tanpa Label. Sebanyak 121 data mahasiswa tanpa nilai IP semester 3 digunakan untuk prediksi setelah model selesai dilatih.

Selanjutnya melatih dengan model *Random Forest* berdasarkan data yang telah dipisah. Setelah melatih model, langkah selanjutnya adalah melakukan prediksi terhadap data Uji.

4) Evaluasi

Setelah data didapatkan, kemudian melakukan tahap evaluasi menggunakan 3 matrik yaitu matrik MAE, MSE dan R². Berikut hasil dari evaluasi tersebut:

Tabel 1. Hasil Evaluasi

Metrik	Nilai	Interpresi
<i>Mean Absolute Error (MAE)</i>	0.2704	Rata-rata kesalahan absolut antara prediksi dan nilai aktual.
<i>Mean Squared Error (MSE)</i>	0.2410	Rata-rata kuadrat dari kesalahan, menunjukkan sensitivitas terhadap outlier.
<i>R-squared (R²)</i>	0.3063	Persentase variabilitas data yang dapat dijelaskan oleh model.

a. *Mean Absolute Error (MAE)*

MAE adalah rata-rata dari nilai absolut selisih antara nilai asli dan nilai prediksi. Dalam kasus ini, rata-rata kesalahan prediksi model adalah sekitar 0.27 poin. Nilai MAE yang lebih kecil menunjukkan model yang lebih akurat. Jika IP skala 4, maka MAE 0.27 menunjukkan bahwa prediksi model cukup mendekati nilai asli.

b. *Mean Squared Error (MSE)*

MSE adalah rata-rata dari kuadrat selisih antara nilai asli dan nilai prediksi. Karena menggunakan kuadrat, MSE memberikan bobot lebih besar untuk kesalahan besar. Nilai MSE yang lebih kecil menunjukkan model yang lebih akurat, karena MSE mengkuadratkan error, nilainya sering kali lebih besar daripada MAE. Dalam kasus ini, MSE 0.2387 menunjukkan model memiliki error yang relatif kecil.

c. *R-squared (R^2)*

R^2 adalah metrik yang mengukur seberapa baik model menjelaskan variabilitas dalam data asli. Nilainya berkisar antara 0 hingga 1. Jika semakin mendekati 1, semakin baik model dalam menjelaskan data. Nilai negatif: Jika model lebih buruk dari rata-rata sederhana. Dari tabel di atas dapat ditarik kesimpulan bahwa model *Random Forest* menunjukkan hasil yang cukup baik, tetapi performa model masih dapat ditingkatkan untuk mencapai hasil yang lebih akurat, misalnya dengan mengoptimalkan parameter model atau menambahkan fitur relevan.

Dalam kasus ini, R^2 sebesar 0.3128 menunjukkan bahwa model hanya dapat menjelaskan sekitar 31.28% dari variasi dalam nilai asli IP Semester 2. Sisanya (68.72%) adalah variasi yang tidak dapat dijelaskan oleh model, hal ini dapat terjadi karena dalam kasus di mana variabel yang diprediksi (seperti performa akademik) dipengaruhi oleh banyak faktor yang tidak dapat diukur atau sulit diprediksi (misalnya, faktor sosial, psikologis, atau eksternal lainnya), sehingga nilai R^2 yang lebih rendah bisa dianggap wajar selama tidak mendapat nilai negatif.

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka dapat ditarik sebuah kesimpulan bahwa algoritma *Random Forest* berhasil diterapkan untuk memprediksi Indeks Prestasi (IP) Semester 2 mahasiswa, dengan performa yang diukur menggunakan R^2 , *Mean Absolute Error* (MAE), dan *Mean Squared Error* (MSE), menunjukkan akurasi yang memadai. Model ini mampu mengidentifikasi fitur-fitur penting, seperti nilai mata kuliah dan jenis kelamin, yang memiliki pengaruh signifikan terhadap hasil prediksi. Evaluasi model menunjukkan bahwa MAE sebesar 0.27 mengindikasikan rata-rata kesalahan prediksi yang cukup kecil, sedangkan MSE sebesar 0.2387 menunjukkan *error* model yang relatif rendah. R^2 sebesar 0.3128 mengungkapkan bahwa model dapat menjelaskan sekitar 31.28% variasi dalam data, sementara sisanya masih dipengaruhi oleh faktor lain yang tidak terukur. Hasil ini juga menunjukkan bahwa prediksi model konsisten dengan distribusi data asli, meskipun terdapat potensi ketidakakuratan pada nilai-nilai ekstrem. Dengan demikian, performa model *Random Forest* dinilai cukup baik namun masih dapat ditingkatkan melalui optimasi parameter atau penambahan fitur relevan untuk meningkatkan akurasi prediksi lebih lanjut.

DAFTAR PUSTAKA

- Adiyati, P. A. (2025). Implementasi Algoritma Random Forest. *Jurnal MNEMONIC*, 8(1), 70–73.
- Budiman, E., Haviluddin, Dengan, N., Kridalaksana, A. H., Wati, M., & Purnawansyah. (2018). Performance of Decision Tree C4.5 Algorithm in Student Academic Evaluation. *Lecture Notes in Electrical Engineering*, 488(February), 380–389. https://doi.org/10.1007/978-981-10-8276-4_36
- Efendi, M. S., & Zyen, A. K. (2024). Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk. 5(1), 12–20. <https://doi.org/10.30865/resolusi.v5i1.2149>

- Febrian, S. R., Sunarto, A. A., & Pambudi, A. (2024). Prediksi Penjualan Suku Cadang Motor Dengan Penerapan Random Forest Di PT Terus Jaya Sentosa Motor. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 10507-10513. <https://doi.org/10.36040/jati.v8i5.11100>
- Hendri, H., Hoki, L., Agusman, V., & Aryanto, D. (2021). Penerapan Machine Learning Untuk Mengategorikan Sampah Plastik Rumah Tangga. *Jurnal TIMES*, 10(1), 1–5. <https://doi.org/10.51351/jtm.10.1.2021645>
- Huang, D., & Wasito, I. (2024). Analisis Peran Atlet Dota 2 Dengan Algoritma Random Forest. *Decode: Jurnal Pendidikan Teknologi Informasi*, 4(1), 107-115. <https://doi.org/10.51454/decode.v4i1.216>
- Kusworo, S., Santoso, N. A., & Kurniawan, R. D. (2024). *Prediksi Nilai Akhir Semester Siswa Menggunakan Algoritma Random Forest*. *Jurnal Sains dan Ilmu Terapan*, 7(2), 246-256. <https://doi.org/10.59061/jsit.v7i2.918>
- Lestari, E. S., & Astuti, I. (2022). Penerapan Random Forest Regression Untuk Memprediksi Harga Jual Rumah Dan Cosine Similarity Untuk Rekomendasi Rumah Pada Provinsi Jawa Barat. *Jurnal Ilmiah FIFO*, 14(2), 131. <https://doi.org/10.22441/fifo.2022.v14i2.003>
- Marlina, H., Elmayati, E., Zulus, A., & Wijaya, H. O. L. (2023). Penerapan Algoritma Random Forest Dalam Klasifikasi Penjurusan Di SMA Negeri Tugumulyo. *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, 4(2), 138-143.
- Novrian, R., Agustiani, T., Fikri, M., Hikmatulloh, M. F., Gunawan, M. E., & Firdaus, U. (2024). Penerapan Algoritma Random Forest dalam Prediksi Status Penerima PIP pada Siswa: Studi Kasus pada SMK Amaliah 1. *Karimah Tauhid*, 3(2), 1791–1799. <https://doi.org/10.30997/karimahtauhid.v3i2.11937>
- Nugroho, A., & Husin, A. (2022). Performance Analysis of Random Forest Using Attribute Normalization. *Sistemasi*, 11(1), 186. <https://doi.org/10.32520/stmsi.v11i1.1681>
- Nur, N., Wajidi, F., Sulfayanti, S., & Wildayani, W. (2023). Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga. *Jurnal Komputer Terapan*, 9(1), 58–64. <https://doi.org/10.35143/jkt.v9i1.5917>
- Perez, J. G., & Perez, E. S. (2021). Predicting Student Program Completion Using Naïve Bayes Classification Algorithm. *International Journal of Modern Education and Computer Science*, 13(3), 57–67. <https://doi.org/10.5815/IJMECS.2021.03.05>
- Pitaloka, E., Hartanto, T. B. A., & Sandiwarno, S. (2024). Penerapan Machine Learning Untuk Prediksi Bencana Banjir. *Jurnal Sistem Informasi Bisnis*, 14(1), 62–76. <https://doi.org/10.21456/vol14iss1pp62-76>
- Purwati, N., & Dwi Januanti, A. (2021). Aplikasi Data Mining Dengan Algoritma Naive Bayes Untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal Pepadun*, 2(1), 123–137. <https://doi.org/10.23960/pepadun.v2i1.38>
- Rismaya, R., Yuniarto, D., & Setiadi, D. (2025). Penerapan Algoritma Machine Learning dalam Prediksi Prestasi Akademik Mahasiswa. *Router: Jurnal Teknik Informatika Dan Terapan*, 3(1), 15–23.
- Santoso, R. R., Megasari, R., & Hambali, Y. A. (2020). Implementasi Metode Machinelearning. *Jurnal Aplikasi Dan Teori Ilmu Komputer*, 3(2), 85–97. <https://ejournal.upi.edu/index.php/JATIKOM>
- Suryanto, A. A. (2019). Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi. *Saintekbu*, 11(1), 78–83. <https://doi.org/10.32764/saintekbu.v11i1.298>

- Wardhana, R. G., Wang, G., & Sibuea, F. (2023). Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia. *Journal of Information System Management (JOISM)*, 5(1), 40–45. <https://doi.org/10.24076/joism.2023v5i1.1136>
- Waruwu, M. N., Zega, Y., & Mendrofa, Ratna Natalia Telaumbanua, Y. N. (2024). Implementasi Algoritma Machine Learning untuk Deteksi Performa Akademik Mahasiswa. *Teknimedia*, 5(2), 181–186.