Penerapan Algoritma Random Forest Untuk Prediksi Performa Akademik Siswa Sekolah Dasar

Nafi Maula Rif'at[1*], Nibras Faiq Muhammad[2], Wiji Lestari[3]

[1]Fakultas Ilmu Komputer
Universitas Duta Bangsa Surakarta
[1*]220103183@mhs.udb.ac.id

[2]Fakultas Ilmu Komputer
Universitas Duta Bangsa Surakarta
[2]nibras_faiqmuhammad@udb.ac.id

[3]Fakultas Ilmu Komputer
Universitas Duta Bangsa Surakarta
[3]wiji_lestari@udb.ac.id

***Abstrak*— Performa akademik siswa merupakan salah satu indikator penting dalam mengevaluasi keberhasilan proses pembelajaran pada jenjang sekolah dasar. Namun, pemanfaatan data akademik yang tersedia di sekolah sering kali masih terbatas sebagai arsip administratif dan belum dimanfaatkan secara optimal untuk mendukung pengambilan keputusan. Penelitian ini bertujuan menerapkan algoritma Random Forest dalam mengklasifikasikan performa akademik siswa berdasarkan data nilai mata pelajaran dan absensi. Dataset yang digunakan terdiri dari 177 data siswa yang dikelompokkan ke dalam tiga kategori performa akademik, yaitu Tinggi, Sedang, dan Rendah. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, pembangunan model menggunakan Random Forest dengan optimasi hyperparameter menggunakan GridSearchCV berbasis 5-Fold Cross-Validation, serta evaluasi model menggunakan metrik evaluasi dan Confusion Matrix. Hasil validasi Stratified 5-Fold CV menunjukan rata-rata akurasi sebesar 90,79%, dengan standar deviasi 2,82%, mengindikasikan model yang stabil. Optimasi menggunakan GridSearchCV menghasilkan Best CV Score sebesar 92,93%. Evaluasi pada data pengujian menghasilkan akurasi 91,67% dengan nilai precision 92,01%, recall 91,67%, dan F1-score 91,75%. Analisis feature importance menunjukkan bahwa atribut IPAS, Bahasa Indonesia, dan Agama merupakan variabel paling berpengaruh dalam proses klasifikasi. Hasil penelitian menunjukkan bahwa algoritma Random Forest mampu mengklasifikasikan performa akademik siswa dengan baik serta akurat dan dapat dimanfaatkan sebagai alat bantu objektif dalam mendukung pengambilan keputusan pembelajaran berbasis data.**
Kata kunci*— *Random Forest*, Performa Akademik, Prediksi, Klasifiikasi, *Machine Learning, Hyperparameter Tuning.

***Abstract*— Student academic performance is an important indicator in evaluating the success of the learning process at the elementary school level. However, the utilisation of academic data available at schools is often still limited to administrative archives and has not been optimally used to support decision-making. This study aims to apply the Random Forest algorithm to classify students' academic performance based on subject grades and attendance data. The dataset used consists of 177 student data grouped into three categories of academic performance, namely High, Medium, and Low. The research stages include data collection, data preprocessing, model development using Random Forest with hyperparameter optimisation using GridSearchCV based on 5-Fold Cross-Validation, and model evaluation using evaluation metrics and the Confusion Matrix. The results of the Stratified 5-Fold CV validation show an average accuracy of 90.79%, with a standard deviation of 2.82%, indicating a stable model. Optimisation using GridSearchCV resulted in a Best CV Score of 92.93%. Evaluation of the test data yielded an accuracy of 91.67%, with precision 92,01%, recall 91,67%, and F1-score 91,75%. Feature importance analysis indicates that the attributes of Social Science (IPAS), Indonesian Language, and Religion are the most influential variables in the classification process. The results show that the Random Forest algorithm is capable of classifying students' academic performance well and accurately, and can be used as an objective tool to support data-driven learning decision-making.**
Keywords*— *Random Forest, Academic Performance, Predictions,Classification, Machine Learning, Hyperparameter Tuning.

I. PENDAHULUAN

Pendidikan dasar merupakan fondasi krusial dalam pembentukan karakter dan kemampuan akademik siswa. Performa akademik pada siswa sekolah dasar menjadi indikator penting dalam keberhasilan proses pembelajaran serta perkembangan kemampuan siswa yang perlu dipantau secara berkelanjutan [1], [2]. Pada umumnya data nilai dan absensi siswa telah tersedia sebagai bagian dari administrasi akademik sekolah. Namun pemanfaatan data tersebut masih sebatas sebagai arsip administratif dan belum dimanfaatkan secara optimal untuk mendukung pengambilan keputusan dalam pembelajaran. Tanpa adanya sistem yang mampu mengolah dan mengklasifikasikan data tersebut secara sistematis, pihak sekolah belum memiliki gambaran yang objektif mengenai sebaran kategori performa siswa secara keseluruhan [3].

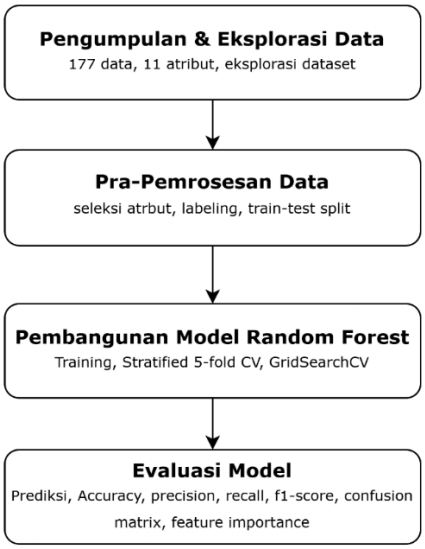
Proses identifikasi performa akademik siswa yang dilakukan oleh guru seringkali bersifat subjektif, umumnya hanya mengandalkan pengamatan langsung oleh guru terhadap nilai dan absensi siswa tanpa dukungan sistem yang mampu mengolah data secara sistematis. Pendekatan ini memiliki keterbatasan seperti rentan subjektivitas dalam penilaian, ketidakmampuan dalam mendeteksi pola hubungan antar data, minimnya visualisasi informasi, serta kurang efisien karena memerlukan waktu yang lebih lama dalam melakukan pengamatan dan penilaian [4]. Kondisi tersebut menunjukkan bahwa tanpa adanya pendekatan berbasis data, proses pemantauan akademik siswa tidak dapat dilakukan secara optimal karena belum mampu memberikan gambaran yang objektif dan menyeluruh. Hal ini menyebabkan pihak sekolah kesulitan dalam memahami pola performa akademik siswa secara sistematis. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengolah data akademik siswa secara sistematis agar pihak sekolah dapat memperoleh gambaran performa siswa secara lebih baik [3], [5].

Algoritma Random Forest dipilih sebagai solusi pada penelitian ini. Random Forest merupakan metode ensemble yang bekerja dengan membangun sejumlah pohon keputusan secara acak kemudian menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan keputusan akhir [6]. Pemilihan algoritma ini didasarkan pada beberapa pertimbangan utama yang relevan dengan karakteristik data dan kondisi dalam penelitian ini. Pertama, Random Forest terbukti mampu menghasilkan prediksi yang stabil meskipun pada dataset yang relatif terbatas, seperti data nilai dan absensi siswa dari satu sekolah. Selain itu, algoritma ini memiliki ketahanan terhadap overfitting yang sering menjadi kelemahan utama decision tree tunggal, sehingga lebih andal untuk digunakan pada data nyata. Lebih lanjut, random forest juga mampu menghasilkan feature importance yang dapat mengidentifikasi variabel data akademik mana yang paling berkontribusi dalam menentukan kategori performa siswa [7], [8].

Berdasarkan uraian tersebut, penelitian ini dilakukan untuk memanfaatkan data nilai dan absensi siswa sebagai dasar untuk penerapan algoritma *Random Forest* dalam melakukan klasifikasi performa akademik siswa sekolah dasar. Hasil penelitian diharapkan dapat digunakan sebagai alat bantu oleh tenaga pendidik atau sekolah dalam memperoleh gambaran performa akademik siswa secara lebih objektif serta mendukung pengambilan keputusan yang berkaitan dengan proses pembelajaran.

II. METODOLOGI PENELITIAN

Dalam penelitian ini terdiri dari beberapa tahapan yang dimulai dari pengumpulan data hingga evaluasi model. Tahapan penelitian digambarkan secara visual melalui diagram alir pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, penelitian diawali dengan pengumpulan dan eksplorasi data untuk memahami karakteristik dataset yang digunakan. Data tersebut kemudian masuk ke tahap pra-pemrosesan yang meliputi seleksi atribut, pembentukan label, pemisahan fitur dan target, serta pembagian dataset. Tahap berikutnya adalah melatih model menggunakan algoritma *Random Forest* yang dioptimasi dengan GridSearchCV berbasis 5-Fold Cross-Validation. Penelitian diakhiri dengan tahap evaluasi menggunakan data testing untuk menguji generalisasi model.

A. Pengumpulan dan Eksplorasi Data

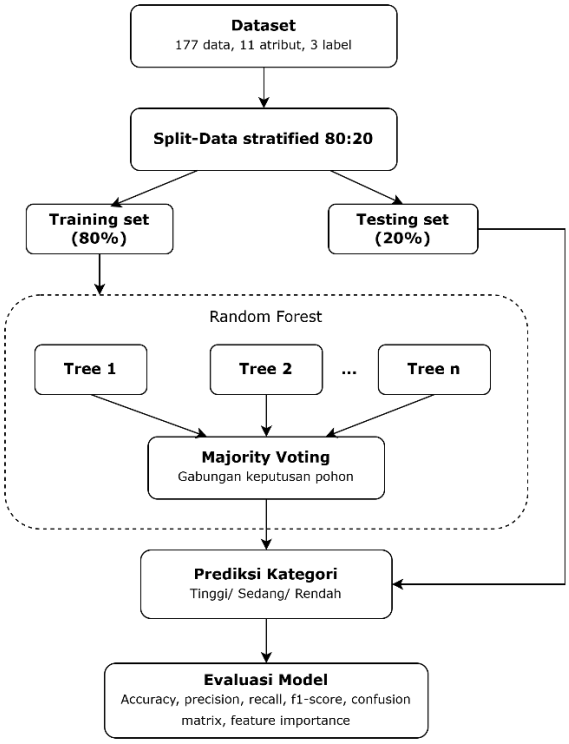
Data penelitian diperoleh dari SD Negeri 03 Jati melalui teknik studi dokumentasi terhadap data akademik siswa berupa data nilai rapor dan rekapitulasi kehadiran siswa pada semester ganjil tahun ajaran 2025/2026. Pada tahap pengumpulan data didapatkan dataset sebanyak 177 data siswa yang berasal dari kelas 3 - 6. Dataset mencakup 10 variabel nilai mata pelajaran yaitu Pendidikan Agama, Pendidikan Pancasila, Bahasa Indonesia, Matematika, IPAS, Seni Rupa, Seni Musik, PJOK, Bahasa Inggris, dan Bahasa Jawa, serta 1 variabel persentase kehadiran (absensi) siswa. Dengan demikian, total fitur yang digunakan dalam penelitian ini adalah 11 fitur. Tahap Eksplorasi data dilakukan untuk memahami karakteristik dan distribusi dataset sebelum masuk ke tahap pemrosesan. Pada tahap ini dilakukan pemeriksaan terhadap struktur data, tipe data setiap variabel, statistik deskriptif, serta keberadaan *missing value.*

B. Pra-Pemrosesan Data

Tahap pra-pemrosesan data bertujuan mempersiapkan data mentah menjadi data yang bersih dan siap digunakan dalam proses pemodelan klasifikasi. Proses yang dilakukan meliputi seleksi atribut dengan menghapus kolom yang tidak digunakan, pembentukan label kategori performa akademik siswa, label mengacu pada standar penilaian sekolah dan dibagi menjadi 3 kategori yaitu, Tinggi untuk siswa dengan rata-rata ≥87, kategori Sedang untuk rata-rata 82–86, dan kategori Rendah untuk rata-rata <82. Label ini merepresentasikan tingkat capaian performa akademik di mana seluruh siswa dalam dataset berada dalam rentang nilai yang memenuhi KKM (≥75). Selanjutnya dilakukan pemisahan fitur (x) dan target (y), fitur terdiri dari 11 atribut, yaitu 10 nilai mata pelajaran dan persentase kehadiran siswa, sedangkan target berupa kategori performa akademik siswa. serta terakhir dilakukan proses *train-test split* atau pembagian dataset dengan metode *stratified random split* dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian. Pendekatan *stratified split* berguna untuk menjaga proporsi tiap kategori kelas tetap seimbang pada kedua subset data.

C. Pembangunan Model Random Forest

Model dibangun menggunakan algoritma Random Forest sebagai metode klasifikasi performa akademik siswa. Untuk memahami bagaimana data diproses di dalam sistem mulai dari input data hingga evaluasi model, ditunjukan arsitektur model Random Forest pada Gambar 2 berikut.



Gambar 2. Arsitektur Model Klasifikasi Performa Akademik Siswa Menggunakan Random Forest

Berdasarkan Gambar 2, dataset akademik siswa dibagi dengan proporsi 80% data pelatihan dan 20% data pengujian. Pada tahap pemodelan, data pelatihan digunakan untuk membangun model *Random Forest*, sedangkan data pengujian digunakan untuk menguji kemampuan klasifikasi model. Hasil prediksi kemudian dievaluasi menggunakan metrik *Accuracy, Precision, Recall, F1-Score, Confusion Matrix,* dan *Feature Importance.*

Model *Random Forest* tersebut dilatih menggunakan data latihan dengan parameter *class_weight* diatur menjadi *balanced* untuk mengatasi ketidakseimbangan distribusi kelas pada dataset. Untuk mengevaluasi stabilitas dan

robustness model secara lebih komprehensif, digunakan teknik Stratified 5-Fold Cross-Validation pada data pelatihan [9]. Dalam proses ini, data pelatihan dibagi menjadi 5 fold yang sama besar, di mana setiap bagian digunakan secara bergantian sebagai data validasi sementara bagian lainnya digunakan sebagai data pelatihan. Hasil akhir CV diperoleh dari nilai rata-rata dan standar deviasi dari 5 iterasi tersebut sebagai indikator kestabilan performa model.

Selanjutnya dilakukan optimasi hyper-parameter menggunakan metode GridSearch Cross-Validation (GridSearchCV) dengan strategi Stratified 5-fold CV. GridSearchCV berfungsi untuk mencari kombinasi hyperparameter terbaik pada model Random Forest. Setiap kombinasi hyperparameter dievaluasi menggunakan teknik stratified 5-Fold CV pada data pelatihan. Kombinasi parameter yang menghasilkan performa validasi terbaik dipilih sebagai model final dan digunakan pada tahap evaluasi menggunakan data pengujian [8].

Stratified dipilih untuk memastikan proporsi setiap kelas tetap terjaga secara seimbang pada setiap fold nya, sehingga evaluasi selama proses tunning lebih representatif terutama pada kondisi distribusi kelas yang tidak seimbang [10]. Parameter yang dioptimasi beserta nilai yang diuji disajikan pada Tabel 1.

Tabel 1. Parameter GridSearch

Parameter	Nilai yang Diuji
n_estimators	50, 100, 150
Max_depth	None, 5, 10, 15
min_samples_split	2, 5, 10
Min_samples_leaf	1, 2, 4

D. Evaluasi Model

Evaluasi model dilakukan untuk mengukur sejauh mana performa dari model Random Forest yang telah dibangun dalam mengklasifikasikan data pengujian secara tepat dan dapat digeneralisasi pada data baru. Pengujian dilakukan secara objektif menggunakan data pengujian/testing. Evaluasi mencakup tiga aspek yaitu metrik evaluasi, confusion matrix, dan analisis feature importance [11].

Performa model diukur menggunakan empat metrik yaitu accuracy, precision, recall, dan F1-score. Seluruh matrik ini dihitung melalui pendekatan *weighted average* untuk mengakomodasi kondisi distribusi kelas yang tidak seimbang pada dataset [12]. Persamaan formal untuk metrik-metrik tersebut adalah sebagai berikut:

Accuracy Mengukur prediksi yang benar terhadap keseluruhan data. Memberikan gambaran umum mengenai sejauh mana model mampu mengklasifikasikan data dengan benar.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision mengukur tingkat keakuratan model dalam memprediksi kelas positif dari seluruh data yang diprediksi sebagai positif.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall mengukur kemampuan model dalam mengidentifikasi seluruh data positif yang sebenarnya. sehingga meminimalkan kesalahan *False Negative*.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

F1-Score merupakan rata-rata harmonik antara precision dan recall yang memberikan gambaran seimbang terutama ketika distribusi kelas tidak seimbang.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \quad (4)$$

Tabel 2

		Keterangan
TP (*True Positive*)	:	Jumlah data yang diprediksi benar untuk kelas positif,
TN (*True Negative*)	:	Jumlah data yang diprediksi benar untuk kelas negatif.
FP (*False Positive*)	:	Jumlah data yang diprediksi salah sebagai kelas positif.
FN (*False Negative*)	:	Jumlah data yang diprediksi salah sebagai kelas negatif.

Confusion Matrix merupakan tabel yang digunakan untuk menganalisis hasil klasifikasi secara lebih rinci dengan membandingkan hasil

prediksi model terhadap data actual secara lebih menyeluruh. Dalam penelitian ini, *confusion matrix* berbentuk matriks 3x3 yang merepresentasikan tiga kategori performa akademik yaitu Tinggi, Sedang, dan Rendah.

Feature Importance dilakukan untuk mengetahui tingkat masing-masing atribut terhadap proses klasifikasi yang dilakukan oleh model Random Forest. Analisis *Feature Importance* digunakan untuk mengidentifikasi variabel akademik mana yang paling berpengaruh dalam proses klasifikasi performa siswa. Nilai *Feature Importance* dihitung berdasarkan seberapa besar kontribusi suatu fitur dalam mengurangi *impurity* pada setiap percabangan pohon keputusan. Fitur yang lebih sering digunakan dalam proses pemisahan data akan memiliki nilai *importance* yang lebih tinggi [13].

III. HASIL DAN PEMBAHASAN

A. *Hasil Pra-Pemrosesan Data*

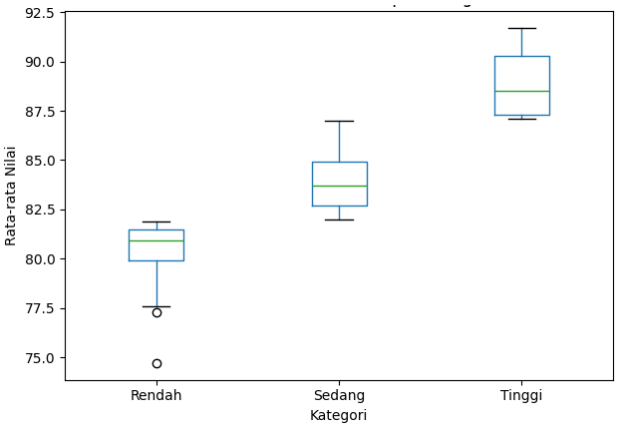
Tahapan awal dilakukan untuk mengetahui karakteristik dari dataset yang digunakan dalam penelitian. Dataset terdiri dari data nilai mata pelajaran dan persentase kehadiran siswa yang telah dikelompokkan ke dalam tiga kategori performa akademik melalui proses pelabelan. Setelah dilakukan proses pelabelan, distribusi data pada setiap kategori ditunjukan pada Tabel 3.

Tabel 3. Distribusi Kategori/Label Dataset

Kategori	Jumlah	Persentase
Tinggi	37	20.9%
Sedang	89	50.3%
Rendah	51	28.8%
Total	**177**	**100%**

Berdasarkan Tabel 2, distribusi dataset menunjukkan bahwa data penelitian masih didominasi oleh kategori Sedang, namun jumlah data pada kategori lain masih cukup representatif untuk digunakan dalam proses klasifikasi.

Untuk memvalidasi hasil pembentukan label, dilakukan visualisasi menggunakan boxplot untuk melihat distribusi rata-rata nilai dari masing-masing kategori performa akademik siswa. Hasil ditunjukan pada gambar dibawah.



Gambar 3. Distribusi Rata-rata Nilai Per Kategori

Berdasarkan Gambar 3, kategori Tinggi memilik median nilai yang lebih besar dibandingkan kategori lain. Kategori Sedang berada pada rentang nilai menengah, sedangkan kategori Rendah pada rentang yang lebih rendah. Perbedaan distribusi nilai yang jelas antar ketiga kategori menunjukkan bahwa atribut yang digunakan memiliki kemampuan untuk membedakan performa akademik siswa.

Selanjutnya dilakukan pemisahan data menjadi fitur (X) dan target (y), dan setelahnya dataset dibagi menjadi data latih dan data uji menggunakan metode *train-test split* dengan rasio 80:20. Distribusi data masing-masing subset ditunjukan pada Tabel 4 dibawah.

Tabel 4. Hasil Pembagian Data Training dan Testing.

Dataset	Jumlah Data	Persentase
Training	141	79.7%
Testing	36	20.3%
Total	**177**	**100%**

Distribusi kategori pada data pelatihan terdiri dari 29 data kategori Tinggi, 71 Sedang, dan 41 Rendah. Sedangkan data pengujian, terdiri dari 8 data kategori Tinggi, 18 Sedang, dan 10 Rendah. Pembagian data dilakukan menggunakan *stratified sampling* sehingga proporsi kategori pada data pelatihan dan pengujian tetap terjaga seimbang.

B. *Hasil Pelatihan dan Optimasi Model*

Dilakukan evaluasi *robustness* model baseline menggunakan metode Stratified 5-Fold Cross-Validation pada data pelatihan untuk mengukur kestabilan dan kemampuan generalisasi model *Random Forest*. Hasil pengujian 5-Fold CV ditunjukkan pada Tabel 5 di bawah.

Tabel 5. Hasil Stratified 5-Fold Cross-Validation

Fold	Akurasi
Fold 1	89.66%
Fold 2	89.29%
Fold 3	96.43%
Fold 4	89.29%
Fold 5	89.29%
Mean	**90.79%**
Std Dev	**2.82%**

Rata-rata akurasi diperoleh 90,79% dengan standar deviasi sebesar 2,82%. Nilai standar deviasi relatif rendah menandakan bahwa performa model cenderung stabil dan konsisten pada setiap fold pengujian, sehingga layak untuk digunakan pada tahap pelatihan dan evaluasi dengan data testing.

Optimasi *hyperparameter* dilakukan menggunakan GridSearchCV dengan *Stratified 5-Fold Cross-Validation* terhadap 108 kombinasi parameter dan menghasilkan 540 proses *fitting.* Proses ini dilakukan untuk mendapatkan kombinasi parameter yang mampu menghasilkan performa terbaik pada model Random Forest. Parameter terbaik yang diperoleh disajikan pada tabel 6.

Tabel 6. Parameter Terbaik Hasil GridSearchCV

Parameter	Nilai Terbaik
n_estimators	100
Max_depth	None
min_samples_split	10
Min_samples_leaf	4
Best CV Score	**92.93%**
Akurasi Training	**97.87%**

Berdasarkan Tabel 6, kombinasi parameter terbaik menghasilkan nilai rata-rata akurasi CV sebesar 92,93%. Setelah model dilatih menggunakan seluruh data pelatihan dengan parameter tersebut, diperoleh akurasi training sebesar 97,87%. Hasil ini menunjukkan bahwa proses optimasi *hyperparameter* mampu menghasilkan model *Random Forest* dengan performa yang lebih baik dan stabil dalam melakukan klasifikasi.

C. Hasil Evaluasi Model

Model *Random Forest* dengan parameter terbaik hasil GridSearchCV dievaluasi menggunakan data pengujian sebanyak 36 siswa. Pengujian dilakukan untuk mengukur kemampuan model dalam melakukan klasifikasi ke dalam kategori yang telah ditentukan. Hasil evaluasi ditunjukan pada Tabel 7.

Tabel 7. Hasil Evaluasi Model Random Forest

Metrik	Nilai
Accuracy	91.67%
Precision (weighted avg)	92.01%
Recall (weighted avg)	91.67%
F1-Score (weighted avg)	91.75%

Model Random Forest menghasilkan *accuracy* sebesar 91,67% pada data pengujian. Nilai masing -masing secara *weighted average* adalah *precision* 92,01%*, recall* 91,67%*,* dan *F1-score* 91,75%, menunjukkan bahwa model mampu mengklasifikasikan ketiga kategori performa akademik secara seimbang dan konsisten.

Hasil evaluasi per kelas disajikan lebih rinci melalui *classification report* pada Tabel 8.

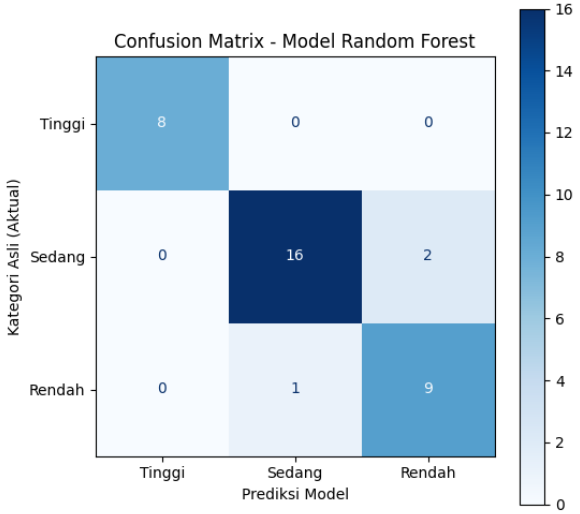
Tabel 8. *Classification Report* per Kelas

Kelas	*Precision*	*Recall*	*F1-Score*	*Support*
Rendah	0.82	0.90	0.86	10
Sedang	0.94	0.89	0.91	18
Tinggi	1.00	1.00	1.00	8
Weighted Avg	**0.92**	**0.92**	**0.92**	**36**

Berdasarkan Tabel 8, kategori Tinggi memperoleh nilai *precision, recall,* dan *F1-score* sempurna sebesar 1,00, seluruh siswa berkategori

Tinggi berhasil diklasifikasikan dengan benar oleh model. Kategori Sedang memperoleh precision 0,94 dan recall 0,89 dengan F1-score 0,91, menunjukkan performa yang sangat baik. Kategori Rendah memperoleh precision terendah sebesar 0,82 dan recall 0,90 dengan F1-score 0,86.

Hasil prediksi model divisualisasikan melalui confusion matrix pada Gambar 4 di bawah.



Gambar 4. *Confusion Matrix* Model *Random Forest*

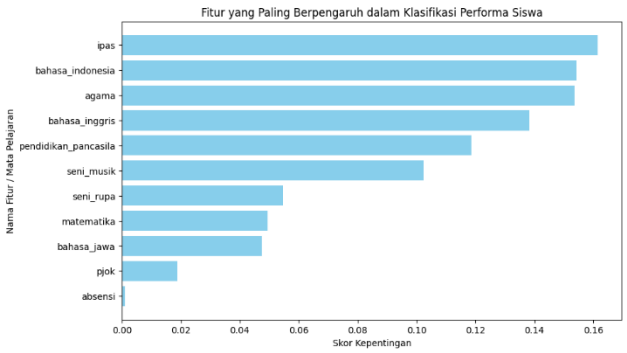
Berdasarkan confusion matrix pada Gambar 4, dari 36 data pengujian terdapat 33 prediksi benar dan 3 prediksi salah. Secara rinci, seluruh 8 siswa berkategori Tinggi diprediksi dengan benar (8/8). Dari 18 siswa berkategori Sedang, 16 diprediksi benar dan 2 diprediksi salah sebagai kategori Rendah. Dari 10 siswa berkategori Rendah, 9 diprediksi benar dan 1 diprediksi salah sebagai kategori Sedang. Kesalahan prediksi yang terjadi pada kategori Sedang dan Rendah dapat disebabkan oleh karakteristik nilai rata-rata kedua kategori tersebut yang berada pada rentang yang berdekatan, sehingga model mengalami kesulitan dalam membedakan keduanya secara tepat.

D. Hasil Analisis Feature Importance

Analisis feature importance dilakukan untuk mengidentifikasi variabel akademik yang paling berpengaruh dalam proses klasifikasi performa siswa. Hasil analisis disajikan pada Tabel 9 dan Gambar 5 berikut.

Tabel 9. Nilai *Feature Importance*

No	Fitur	Skor
1	IPAS	0.161528 (16.15%)
2	Bahasa Indonesia	0.154252 (15.43%)
3	Agama	0.153540 (15.35%)
4	Bahasa Inggris	0.138370 (13.84%)
5	Pendidikan Pancasila	0.118520 (11.85%)
6	Seni Musik	0.102337 (10.23%)
7	Seni Rupa	0.054634 (5.46%)
8	Matematika	0.049411 (4.94%)
9	Bahasa Jawa	0.047559 (4.76%)
10	PJOK	0.018839 (1.88%)
11	Absensi	0.001011 (0.10%)



Gambar 5. Grafik *Feature Importance*

Berdasarkan hasil *feature importance*, mata pelajaran IPAS (Ilmu Pengetahuan Alam dan Sosial) merupakan fitur paling berpengaruh dengan skor kepentingan sebesar 16,15%, diikuti oleh Bahasa Indonesia (15,43%) dan Pendidikan Agama (15,35%). Tiga fitur teratas ini secara bersama-sama menyumbang kontribusi sebesar 47,93% terhadap proses klasifikasi. Hal ini mengindikasikan ketiga fitur tersebut memberikan kontribusi terbesar dalam proses klasifikasi performa akademik.

Sebaliknya, atribut absensi memiliki skor kepentingan paling rendah sebesar 0,10%, yang mengindikasikan bahwa tingkat kehadiran siswa tidak terlalu membedakan kategori performa akademik dalam dataset penelitian ini. Hal ini disebabkan oleh tingkat kehadiran siswa yang relatif homogen sehingga tidak memberikan variasi yang cukup signifikan untuk dijadikan pembeda antar kategori.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Random Forest untuk melakukan klasifikasi performa akademik siswa berdasarkan data nilai 10 mata pelajaran dan persentase kehadiran. Dataset yang digunakan sebanyak 177 data siswa pada semester ganjil tahun ajaran 2025/2026 yang dikelompokkan ke dalam tiga kategori performa akademik, yaitu Tinggi, Sedang, dan Rendah. Hasil validasi menggunakan metode 5-Fold CV menunjukkan rata-rata akurasi sebesar 90,79% dengan standar deviasi sebesar 2,82%. Selanjutnya, dalam proses optimasi *hyperparameter* menggunakan GridSearchCV dihasilkan kombinasi parameter terbaik berupa *n_estimators* = 100*, max_depth = None, min_samples_split* = 10*,* dan *min_samples_leaf* = 4 dengan nilai Best CV Score sebesar 92,93%.

Hasil evaluasi model Random Forest diperoleh *accuracy* sebesar 91,67%, dengan nilai *precision* 92,01%*, recall* 91,67%*,* dan *F1-score* 91,75%. Feature importance menunjukkan bahwa atribut IPAS, Bahasa Indonesia, dan Agama merupakan faktor yang paling berpengaruh dalam proses klasifikasi, sedangkan atribut absensi memiliki pengaruh paling rendah, mengindikasikan bahwa tingkat kehadiran siswa dalam dataset ini relatif homogen sehingga tidak cukup signifikan sebagai pembeda antar kategori performa.

Dengan demikian, model Random Forest yang dibangun terbukti mampu mengklasifikasikan performa akademik siswa sekolah dasar secara objektif dan akurat. Algoritma Random Forest dapat digunakan sebagai salah satu pendekatan untuk membantu mengklasifikasikan performa akademik siswa berdasarkan data nilai dan kehadiran yang tersedia.

REFERENSI

[1] W. Setyowati, Jason Moscato, and Chioke Embre, "Strategi Pendidikan Dasar untuk Menghadapi Tantangan Era Kurikulum Digital dengan Studi Empiris," *Jurnal MENTARI: Manajemen, Pendidikan dan Teknologi Informasi*, vol. 2, no. 1, pp. 43–53, Aug. 2023, doi: 10.33050/mentari.v2i1.379.

[2] A. Rimadhani and M. Abduh, "Upaya Guru dalam Meningkatkan Performa Akademik Siswa Berlatar Belakang Status Sosial Ekonomi Rendah di Sekolah Dasar," *Jurnal Basicedu*, vol. 6, no. 4, pp. 6203–6210, May 2022, doi: 10.31004/basicedu.v6i4.3200.

[3] K. Afandi, M. H. Arief, N. Faizatul Laily, and D. Maulana Nugroho, "Analisis Performa Akademik Mahasiswa Menggunakan Social Network Analysis (Studi Kasus: Prodi Bisnis Digital Universitas dr. Soebandi)," *Journal of Technology and Informatics (JoTI)*, vol. 5, no. 2, pp. 64–69, Apr. 2024, doi: 10.37802/joti.v5i2.514.

[4] Y. Sutikno, "Peran Guru dalam Evaluasi Pembelajaran di Kelas," *Jurnal Maitreyawira*, vol. 4, no. 1, pp. 36–41, Apr. 2023, doi: 10.69607/jm.v4i1.73.

[5] T. Telutci and R. Harman, "PENERAPAN DATA MINING UNTUK MEMPREDIKSI PRESTASI SISWA SEKOLAH DASAR MENGGUNAKAN ALGORITMA C4.5," *Computer Based Information System Journal*, vol. 12, no. 1, pp. 12–23, Mar. 2024, doi: 10.33884/cbis.v12i1.8207.

[6] F. Widjaya, M. Ngete, and D. Tumanggor, "Perbandingan Algoritma Random Forestdengan K-Nearest Neighbor pada Klasifikasi Tingkat Stress Pelajar," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 12, no. 4, pp. 171–187, Dec. 2025.

[7] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Informatik : Jurnal Ilmu Komputer*, vol. 18, no. 3, p. 239, Dec. 2022, doi: 10.52958/iftk.v18i3.4694.

[8] Suci Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng," *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, vol. 4, no. 3, pp. 121–127, Dec. 2022, doi: 10.35580/variansiunm31.

[9] W. Nursahid, B. I. Nugroho, and S. Syefudin, "Optimalisasi Preprocessing Data Menggunakan Pendekatan CRISP-DM untuk Meningkatkan Kualitas Klasifikasi Penyakit Jantung," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 3, pp. 3621–3626, Aug. 2025, doi: 10.31004/riggs.v4i3.2514.

[10] A. Rahman, "Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes," *Jurnal SAINTEKOM*, vol. 13, no. 1, pp. 22–31, Mar. 2023, doi: 10.33020/saintekom.v13i1.349.

[11] F. R. Hidayatullah, A. Alsyar, R. A. Putra, W. A. Ramadhani, and E. Ismanto, "Pemodelan Prediktif Diabetes Menggunakan Pendekatan Multimodel Machine Learning dan Deep Learning," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 6, no. 2, pp. 158–165, Aug. 2025, doi: 10.37859/coscitech.v6i2.9812.

[12] I. D. S. Tarigan, Roni Habibi, and Rd. Nuraini Siti Fatonah, "Evaluasi Algoritma Klasifikasi Machine Learning Kategori Nilai Akhir Tunjangan Kinerja Pegawai," *Jurnal Sistem Cerdas*, vol. 6, no. 3, pp. 251–261, Dec. 2023, doi: 10.37396/jsc.v6i3.246.

[13] F. Ramadhan, D. Herlambang, and A. P. Dipta, "Prediksi Status Kesehatan Berdasarkan Gaya Hidup Menggunakan Metode Decision Tree dan Feature Importance," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 4, pp. 9616–9623, Jan. 2026, doi: 10.31004/riggs.v4i4.5246.