Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 744−753
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9598

Klasifikasi Siswa Berprestasi Berdasarkan Nilai Akademik dan Non-Akademik dengan Menggunakan Metode Random Forest

Ricky Gunawan*, Yusuf Ramdhan Nasution

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sumatera Utara, Medan
Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara. Indonesia
Email: [1,*]rickygwn674@gmail.com, [2]ramadhannst@uinsu.ad.id
Email Penulis Korespondensi: rickygwn674@gmail.com
Submitted: **06/04/2026**; Accepted: **24/04/2026**; Published: **30/04/2026**

Abstrak−Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi siswa berprestasi dengan mengintegrasikan aspek akademik dan non-akademik menggunakan metode Random Forest. Permasalahan utama yang dihadapi SMA Negeri 1 Natal adalah proses penentuan siswa berprestasi yang masih berfokus pada nilai akademik dan belum mengakomodasi indikator lain seperti kedisiplinan, kehadiran, dan aktivitas ekstrakurikuler secara terpadu. Penelitian ini menggunakan pendekatan kuantitatif dengan teknik pengumpulan data melalui observasi, wawancara, dan studi literatur. Data yang digunakan berasal dari leger rapor semester ganjil tahun ajaran 2024/2025 sebanyak 222 siswa kelas XI. Tahapan penelitian meliputi preprocessing data (data cleaning, transformasi, normalisasi, dan seleksi fitur), pembagian data menggunakan stratified split (70% data latih dan 30% data uji), serta penerapan algoritma Random Forest untuk klasifikasi. Fitur yang digunakan meliputi rata-rata nilai akademik, ketidakhadiran (sakit, izin, alpa), serta aktivitas ekstrakurikuler. Hasil penelitian menunjukkan bahwa model memiliki performa yang sangat baik dengan nilai akurasi sebesar 1,000 pada data uji dan rata-rata akurasi validasi silang sebesar 0,9865. Selain itu, nilai precision, recall, dan F1-score masing-masing mencapai 1,000. Hasil klasifikasi mengidentifikasi 13 siswa sebagai kategori berprestasi, dengan distribusi terbesar berasal dari kelas XI-1. Temuan ini menunjukkan bahwa metode Random Forest mampu menghasilkan klasifikasi yang akurat dan konsisten, serta efektif dalam mengintegrasikan berbagai indikator penilaian. Penelitian ini diharapkan dapat mendukung pengambilan keputusan yang lebih objektif dan komprehensif dalam sistem evaluasi pendidikan, serta memberikan kontribusi terhadap pengembangan model klasifikasi yang lebih holistik untuk menilai keberhasilan siswa di sekolah, tidak hanya berdasarkan prestasi akademik tetapi juga aspek non-akademik yang penting.

Kata Kunci: Klasifikasi; Random Forest; Siswa Berprestasi; Data Mining

Abstract−This study aims to develop a classification system for high-achieving students by integrating academic and non-academic aspects using the Random Forest method. The main problem faced by Natal State High School 1 is that the process of identifying high-achieving students still focuses on academic grades and does not yet comprehensively incorporate other indicators such as discipline, attendance, and extracurricular activities. This study employs a quantitative approach with data collection techniques including observation, interviews, and literature review. The data used were derived from the report cards for the odd-semester of the 2024/2025 academic year, covering 222 eleventh-grade students. The research stages included data preprocessing (data cleaning, transformation, normalization, and feature selection), data splitting using a stratified split (70% training data and 30% test data), and the application of the Random Forest algorithm for classification. The features used include average academic scores, absences (sick, excused, unexcused), and extracurricular activities. The results showed that the model performed very well, with an accuracy of 1.000 on the test data and an average cross-validation accuracy of 0.9865. Additionally, the precision, recall, and F1-score each reached 1.000. The classification results identified 13 students as high achievers, with the largest distribution coming from 11th grade class 1. These findings indicate that the Random Forest method is capable of producing accurate and consistent classifications and is effective in integrating various assessment indicators. This study is expected to support more objective and comprehensive decision-making within educational evaluation systems and to contribute to the development of more holistic classification models for assessing student success in school, based not only on academic achievement but also on important non-academic aspects.

Keywords: Classification; Random Forest; Outstanding Students; Data Mining; Educational Evaluation

1. PENDAHULUAN

Klasifikasi adalah metode yang digunakan untuk mengelompokkan data ke dalam kategori tertentu yang telah ditentukan sebelumnya[1]. Dalam konteks evaluasi pendidikan, klasifikasi memiliki peran yang sangat penting karena dapat membantu dalam penilaian yang lebih objektif, mengidentifikasi siswa berprestasi dengan lebih akurat, serta mengintegrasikan berbagai indikator yang mencakup aspek akademik maupun non-akademik. Dengan demikian, klasifikasi dapat menjadi alat yang efektif dalam mendukung pengambilan keputusan yang lebih komprehensif dalam proses evaluasi siswa[2], [3].

SMA Negeri 1 Natal adalah lembaga pendidikan yang memiliki visi dan misi untuk membentuk generasi yang agamis, terdidik, dan berbudaya (ADIDAYA). Sekolah ini berlokasi di Kecamatan Natal dengan jumlah peserta didik mencapai 712 orang. Pada tahun ajaran sebelumnya, sebanyak 15% siswa di sekolah ini telah berhasil masuk dalam kategori siswa berprestasi berdasarkan pencapaian akademik dan non-akademik. Namun, penilaian berprestasi yang masih berfokus pada aspek akademik menjadi tantangan utama yang ingin diselesaikan dalam penelitian ini.

Di SMA Negeri 1 Natal, proses pemilihan siswa berpresrasi belum maksimalnya sistem evaluasi dalam menentukan siswa berprestasi, karena belum mengintegrasikan berbagai aspek penilaian secara menyeluruh. Selama ini, penilaian lebih berfokus pada dimensi akademik seperti nilai rapor dan hasil ujian, sedangkan aspek

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 744−753
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9598

non-akademik meliputi kedisiplinan, kehadiran, keterlibatan dalam kegiatan ekstrakurikuler 4belum terakomodasi secara sistematis dalam satu sistem penilaian yang terpadu. Kondisi ini berpotensi menimbulkan ketimpangan dalam proses identifikasi siswa berprestasi karena belum menggambarkan potensi siswa secara utuh. Selain itu, keterbatasan dalam analisis data prestasi juga menghambat pihak sekolah dalam melakukan pemetaan kemampuan siswa secara objektif dan berkelanjutan.

Dari masalah tersebut, dibutuhkan sebuah sistem untuk klasifikasi dengan menerapkan algoritma random forest sehingga proses identifikasi siswa berprestasi berdasarkan nilai akademik maupun non-akademik dapat menghaslkan hasil dengan akurasi yang tepat. Dengan demikian, sistem klasifikasi ini diharapkan mampu meningkatkan kualitas sistem evaluasi pendidikan di SMA Negeri 1 Natal yang lebih tepat, berorientasi pada pengembangan potensi peserta didik secara holistik. Penelitian ini fokus pada Klasifikasi siswa berprestasi berdasarkan nilai akademik dan non-akademik di SMA Negeri 1 Natal.

Penerapan metode Random Forest dalam proses klasifikasi siswa berprestasi dipandang sebagai pilihan yang tepat karena algoritma ini memiliki kemampuan unggul dalam mengolah data yang kompleks, berukuran besar, dan bersifat heterogen. Dibandingkan dengan algoritma lain seperti Decision Tree, Naïve Bayes, atau K-Nearest Neighbor (KNN)[4], [5], [6], metode ini lebih efektif dalam mengatasi masalah overfitting, menghasilkan tingkat akurasi yang lebih konsisten, serta mampu melakukan seleksi fitur (feature selection) secara otomatis untuk mengidentifikasi variabel yang paling berpengaruh terhadap hasil klasifikasi. Selain itu, karakteristik ensemble learning pada Random Forest yang mengombinasikan hasil dari sejumlah pohon keputusan (decision tree) menjadikannya lebih andal dan tahan terhadap gangguan atau noise dalam data[7], [8].

Pada Peneletian Sebelumnya oleh Yefta Yosia Asyel pada tahun 2025 dengan judul "Meningkatkan Prestasi Akademik Mahasiswa Teknik Informatika UNSRAT Melalui Optimasi Pembelajaran dengan Random Forest" Penelitian sebelumnya hanya memanfaatkan data akademik tanpa mempertimbangkan faktor non-akademik dengan hasil 70 siswa sebagai siswa terbaik serta 95 siswa sebagai siswa normal. Namun penelitian ini masih memiliki keterbatasan dalam memberikan gambaran komprehensif mengenai siswa terbaik, dan perlu pengembangan lebih lanjut dengan menambahkan variabel non-akademik agar hasil klasifikasi menjadi lebih objektif dan representatif[9].

Pada penelitian ini, penulis memiliki keterbaruan melalui pengembangan variabel yang lebih relavan terhadap kebutuhan yaitu kedisiplinan, kehadiran, keterlibatan dalam kegiatan ekstrakurikuler sehingga mampu memberikan hasil klasifikasi yang lebih objektif.

Berdasarkan permasalahan yang telah diidentifikasi, penelitian ini berfokus pada pengembangan sistem klasifikasi siswa berprestasi yang mengintegrasikan nilai akademik dan non-akademik menggunakan metode Random Forest. Dengan penerapan metode ini, diharapkan dapat memberikan kontribusi signifikan dalam meningkatkan objektivitas dan keakuratan proses evaluasi siswa di SMA Negeri 1 Natal. Penelitian ini juga bertujuan untuk membantu para guru dan staf sekolah dalam mengambil keputusan yang lebih tepat dalam menilai prestasi siswa, sehingga dapat memberikan dukungan yang lebih baik dalam pengembangan potensi siswa secara menyeluruh.

2. METODOLOGI PENELITIAN

2.1 Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan pengumpulan data melalui observasi, wawancara, dan kajian literatur relevan. Pendekatan kuantitatif dipilih karena kemampuannya dalam menguji hubungan antarvariabel serta memverifikasi teori yang telah ada. Data yang diperoleh akan dianalisis untuk mengembangkan model klasifikasi siswa berprestasi menggunakan metode Random Forest. Fokus utama penelitian ini adalah pada analisis data yang mengintegrasikan indikator akademik dan non-akademik untuk menghasilkan sistem klasifikasi yang lebih objektif dan akurat dalam menilai prestasi siswa.

Metode pengembangan yang diterapkan adalah Random Forest, yaitu algoritma berbasis pohon keputusan gabungan yang efektif dalam proses klasifikasi. Melalui metode ini, sistem diharapkan mampu meningkatkan akurasi dan efisiensi dibandingkan metode sebelumnya[10]. Adapun kerangka tahapan penelitian yang ditempuh dalam studi ini disajikan sebagai berikut:

a. Identifikasi Masalah
   Identifikasi masalah dalam penelitian ini berfokus pada proses penilaian siswa berprestasi di SMA Negeri 1 Natal yang masih terbatas pada aspek akademik, tanpa mempertimbangkan faktor non-akademik seperti kedisiplinan, kehadiran, dan aktivitas ekstrakurikuler. Hal ini menyebabkan kurangnya objektivitas dalam penentuan siswa berprestasi, sehingga keputusan yang diambil tidak sepenuhnya mencerminkan potensi siswa secara menyeluruh. Kesenjangan ini menjadi dasar untuk merumuskan penelitian ini yang bertujuan mengembangkan sistem klasifikasi siswa berprestasi dengan mengintegrasikan kedua aspek tersebut melalui metode Random Forest.
b. Pengumpulan Data
   Dalam penelitian ini, terdapat beberapa teknik pengumpulan data yang digunakan, penelitian ini menggunakan pendekatan kuantitatif dengan teknik pengumpulan data melalui observasi, wawancara, dan studi literatur.

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 744−753
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9598

Observasi dilakukan di SMA Negeri 1 Natal untuk memahami proses penentuan siswa berprestasi yang masih dilakukan secara manual. Wawancara dilakukan dengan guru, wali kelas, dan staf tata usaha untuk menggali variabel yang digunakan dalam penilaian siswa, termasuk kriteria nilai akademik, kedisiplinan, kehadiran (sakit, izin, alpa), dan aktivitas ekstrakurikuler. Studi literatur dilakukan untuk memperkuat dasar teori serta mendukung metode yang digunakan dalam penelitian ini. Data yang diperoleh dari ketiga teknik tersebut digunakan untuk mengembangkan model klasifikasi siswa berprestasi dengan mengintegrasikan variabel akademik, kehadiran, dan ekstrakurikuler.

c. Preprocessing Data
Tahap preprocessing data merupakan proses penting untuk memastikan kualitas data yang optimal sebelum dilakukan klasifikasi. Data mentah yang diperoleh seringkali mengandung inkonsistensi, nilai yang hilang, dan perbedaan skala, sehingga perlu dilakukan beberapa langkah pengolahan. Proses dimulai dengan data cleaning untuk menghapus data ganda dan memperbaiki kesalahan input, serta menangani nilai kosong menggunakan teknik rata-rata untuk data numerik dan modus untuk data kategorik. Selanjutnya, data transformation mengubah variabel kategorik, seperti tingkat kedisiplinan dan keaktifan ekstrakurikuler, ke dalam format numerik menggunakan label encoding atau one-hot encoding agar dapat diproses oleh algoritma Random Forest. Data normalization dilakukan dengan teknik Min-Max Scaling untuk menyamakan skala antarvariabel, memastikan bobot atribut yang seimbang dalam klasifikasi. Feature selection kemudian dilakukan untuk memilih atribut yang paling relevan, seperti nilai rapor, kedisiplinan, dan partisipasi ekstrakurikuler, sementara atribut yang tidak signifikan dieliminasi untuk meningkatkan efisiensi model. Terakhir, data splitting dilakukan menggunakan metode Stratified Split, membagi data menjadi 70% untuk training dan 30% untuk testing, guna memastikan proporsi kelas yang seimbang dalam kedua kelompok dan menjaga akurasi model. Dengan melalui tahapan-tahapan ini, data yang digunakan dalam penelitian ini siap untuk diolah oleh algoritma Random Forest, menghasilkan klasifikasi yang lebih akurat dan reliabel.

d. Penerapan
Flowchart algoritma Random Forest dalam penelitian ini menggambarkan tahapan sistematis dalam proses klasifikasi siswa terbaik di SMA Negeri 1 Natal. Proses dimulai dengan input data siswa, yang mencakup variabel akademik dan non-akademik. Data ini kemudian diproses melalui tahap preprocessing yang mencakup pembersihan data (data cleaning), transformasi data (data transformation), normalisasi (data normalization), dan seleksi atribut relevan (feature selection). Setelah data siap, algoritma melaksanakan bootstrap sampling, yaitu teknik pengambilan sampel data secara acak dengan pengembalian untuk membentuk subset data latih bagi setiap pohon keputusan. Pada setiap subset data, algoritma membangun decision tree dengan parameter yang telah ditentukan, seperti jumlah pohon (n_estimators), kriteria pemisahan (seperti Gini Index atau Entropy), dan jumlah fitur acak pada setiap node (max_features). Proses pembangunan pohon ini diulang hingga terbentuk sejumlah pohon keputusan sesuai kebutuhan, dengan parameter n_estimators ditetapkan sebanyak 100 pohon keputusan. Selanjutnya, prediksi dilakukan melalui mekanisme majority voting untuk kasus klasifikasi, di mana hasil prediksi akhir diperoleh dari agregasi hasil seluruh pohon keputusan[11], [12].

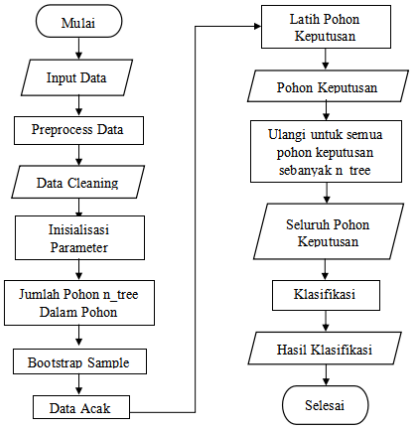
e. Pengujian
Pengujian data dalam penelitian ini dilakukan untuk mengevaluasi kinerja metode dengan memastikan bahwa hasil yang diperoleh konsisten dengan tujuan penelitian serta spesifikasi yang telah ditetapkan. Pengujian model bertujuan untuk mengeksekusi program, mendeteksi kesalahan atau ketidaksesuaian dalam implementasi, serta memastikan bahwa sistem dapat berfungsi dengan baik sesuai kebutuhan pengguna. Kinerja model dievaluasi menggunakan confusion matrix dan metrik evaluasi yang meliputi Akurasi, Precision, Recall, dan F1-Score. Evaluasi ini memberikan gambaran kuantitatif mengenai ketepatan dan keseimbangan performa klasifikasi, serta kemampuan model dalam mengklasifikasikan data dengan benar. Melalui pengujian yang komprehensif ini, efektivitas metode dapat dinilai secara objektif, permasalahan yang muncul dapat segera diperbaiki, dan kualitas serta keandalan sistem dapat ditingkatkan sebelum diimplementasikan secara lebih luas[13], [14].

2.2 Flowchart Algoritma Random Forest

Flowchart Random Forest merupakan representasi visual yang menunjukkan tahapan utama dalam algoritma Random Forest. Diagram ini menggambarkan proses mulai dari pembagian dataset, pembangunan pohon keputusan secara berulang melalui teknik bootstrap sampling, penggabungan hasil dari seluruh pohon, hingga proses pengambilan keputusan akhir melalui mekanisme majority voting atau perhitungan rata-rata[15], [16].

Flowchart algoritma Random Forest pada gambar 1 menjelaskan tahapan sistematis proses klasifikasi siswa terbaik di SMA Negeri 1 Natal. Tahap awal dimulai dengan input data siswa yang mencakup variabel akademik maupun nonakademik, kemudian dilakukan preprocessing data melalui proses pembersihan, transformasi, normalisasi, serta seleksi atribut relevan. Setelah itu, algoritma melaksanakan bootstrap sampling, yaitu pengambilan sampel data secara acak dengan pengembalian untuk membentuk subset data latih. Pada setiap subset tersebut dibangun sebuah decision tree dengan parameter tertentu, seperti jumlah pohon (n_tree) dan kriteria pemisahan, misalnya Gini Index atau Entropy. Proses pembangunan pohon diulang hingga terbentuk sejumlah pohon keputusan sesuai kebutuhan[17], [18]. Berikut ditampilkan algoritma Random Forest pada gambar 1:

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 744−753
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9598



Gambar 1. Flowchart Algoritma Random Forest

Setelah seluruh pohon terbentuk, dilakukan proses majority voting untuk menentukan hasil klasifikasi. Mekanisme ini memastikan bahwa keputusan akhir tidak hanya bergantung pada satu pohon, melainkan merupakan hasil agregasi dari seluruh pohon, sehingga prediksi yang diperoleh lebih stabil dan akurat. Tahap akhir dari flowchart ini adalah menghasilkan output berupa klasifikasi siswa terbaik berdasarkan kombinasi variabel akademik dan nonakademik yang telah dianalisis. Dengan demikian, flowchart ini berfungsi sebagai panduan visual yang memperjelas alur kerja algoritma Random Forest mulai dari input data hingga diperoleh hasil klasifikasi akhir[19], [20].

3. HASIL DAN PEMBAHASAN

Pada bagian ini disajikan hasil penelitian dan implementasi sistem klasifikasi menggunakan metode Random Forest. Model yang dikembangkan mampu mengidentifikasi siswa berprestasi secara otomatis berdasarkan indikator akademik dan non-akademik yang telah diproses melalui tahapan data mining. Proses pengolahan data dimulai dengan pembersihan, transformasi, normalisasi, dan seleksi fitur yang relevan untuk memastikan kualitas data yang optimal. Selanjutnya, algoritma Random Forest diterapkan dengan membagi data menjadi 70% untuk data latih dan 30% untuk data uji, yang menghasilkan hasil klasifikasi yang sangat baik.

Hasil evaluasi model menunjukkan bahwa seluruh data uji berhasil diprediksi sesuai dengan kelas aktualnya, dengan tingkat akurasi, precision, recall, dan F1-score masing-masing sebesar 1.0000 pada data uji, serta rata-rata akurasi sebesar 0.9865 pada validasi silang. Analisis feature importance mengungkapkan bahwa variabel RataAkademik menjadi faktor yang paling dominan dalam klasifikasi, diikuti oleh Alpa, EkskulAktif, dan SkorEkskul.

Untuk menilai efektivitas metode yang diterapkan, hasil ini dibandingkan dengan penelitian sejenis. Sebagai contoh, penelitian oleh Kusworo (2024), yang juga menggunakan algoritma Random Forest untuk memprediksi nilai akhir semester siswa, melaporkan bahwa model tersebut mencapai akurasi prediksi sebesar 93,33%. Perbandingan ini menunjukkan bahwa meskipun Random Forest digunakan dengan dataset yang berbeda, hasilnya menunjukkan konsistensi dalam hal akurasi tinggi, yang memperkuat validitas penerapan metode ini dalam konteks klasifikasi siswa berprestasi. Dalam penelitian ini, dengan mengintegrasikan faktor non-akademik seperti aktivitas ekstrakurikuler dan kedisiplinan, model yang dikembangkan mampu memberikan hasil klasifikasi yang lebih komprehensif dan akurat dibandingkan dengan model yang hanya mengandalkan aspek akademik.
Dari perbandingan tersebut, dapat disimpulkan bahwa penggunaan Random Forest dalam penelitian ini efektif dalam meningkatkan akurasi klasifikasi dengan memanfaatkan berbagai indikator yang relevan, dan dapat menjadi alternatif yang lebih baik dibandingkan dengan model yang hanya bergantung pada aspek akademik.

3.1 Analisis Data

Data yang digunakan dalam penelitian ini berasal dari dokumen resmi leger rapor semester ganjil Tahun Pelajaran 2024/2025 kelas XI SMA Negeri 1 Natal. Dokumen tersebut diperoleh dalam bentuk tujuh file Microsoft Excel yang masing-masing merepresentasikan satu kelas, yaitu XI-1 hingga XI-7. Seluruh file kemudian diintegrasikan menjadi satu dataset terpadu untuk keperluan analisis. Setelah proses integrasi, jumlah keseluruhan data yang diperoleh adalah 222 siswa. Setiap baris data merepresentasikan satu individu siswa, sedangkan setiap kolom menggambarkan atribut atau variabel yang berkaitan dengan performa akademik maupun non-akademik siswa.

3.1.1 Struktur dan Jenis Variabel Awal

a. Preprocessing Data
   Tahap preprocessing data merupakan proses krusial yang harus dilakukan sebelum klasifikasi guna memastikan bahwa data yang digunakan memiliki kualitas optimal. Data mentah biasanya masih mengandung

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 744−753
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9598

inkonsistensi, nilai yang hilang, maupun perbedaan skala, sehingga perlu dilakukan pengolahan awal. Adapun tahapan preprocessing dalam penelitian ini meliputi :

1. Variabel Akademik
   Karakteristik variabel akademik:
   a) Bertipe numerik.
   b) Memiliki rentang nilai yang relatif seragam (umumnya skala 0–100).
   c) Jumlah kolom berbeda antar kelas karena perbedaan komposisi mata pelajaran.
2. Variabel Ketidakhadiran
   Dalam konteks klasifikasi siswa berprestasi, variabel ketidakhadiran berperan sebagai indikator pendukung yang memperkaya analisis, karena prestasi tidak hanya ditentukan oleh nilai akademik, tetapi juga oleh aspek kedisiplinan dan tanggung jawab siswa terhadap kegiatan belajar mengajar.
3. Variabel Ekstrakurikuler
   Data ekstrakurikuler dalam penelitian ini mencerminkan tingkat partisipasi dan kualitas keterlibatan siswa dalam kegiatan sekolah di luar jam pelajaran formal. Informasi yang dicatat meliputi :
   a) Status keaktifan kegiatan.
   b) Kategori penilaian (SB = Sangat Baik, B = Baik).

3.1.2 Distribusi dan Komposisi Data

Distribusi jumlah siswa pada masing-masing kelas relatif seimbang, dengan rentang antara 31 hingga 33 siswa per kelas. Hal ini menunjukkan bahwa tidak terdapat dominasi jumlah siswa yang signifikan dari satu kelas tertentu. Total keseluruhan data adalah 222 siswa, dengan komposisi yang merata antar kelas.

Tabel 1. Jumlah Data Siswa Perkelas

No	Kelas	Jumlah Siswa
1	XI-1	31
2	XI-2	31
3	XI-3	32
4	XI-4	32
5	XI-5	32
6	XI-6	31
7	XI-7	33
Total		222

3.2 Pra-pemrosesan Data

Tahap data cleaning dilakukan untuk memastikan dataset terbebas dari elemen administratif dan siap diproses secara komputasional. Data yang bersumber dari dokumen leger sekolah masih mempertahankan format laporan, sehingga diperlukan penyesuaian agar sesuai dengan kebutuhan analisis. Tahap transformasi dan pembentukan fitur dilakukan untuk mengubah data mentah hasil data cleaning menjadi variabel numerik yang konsisten antar kelas. Enam fitur akhir yang digunakan dalam model adalah RataAkademik, Sakit, Izin, Alpa, EkskulAktif, dan SkorEkskul. Fitur model klasifikasi dapat dijelaskan pada gambar berikut :

Tabel 2. Fitur Dalam Model Klasifikasi

No	Variabel
1	Rata Akademik
2	Sakit
3	Izin
4	Alpa
5	Ekskul Aktif
6	Skor Ekskul

a. Pembentukan Fitur Rata Akademik
   Perhitungan RataAkademik untuk siswa ke-j dirumuskan sebagai berikut:

$$\text{Rata Akademik}_{j} = \frac{\sum_{m=1}^{n_j} x_{jm}}{n_j} \quad (1)$$

   Rumus ini menghitung rata-rata nilai akademik siswa dengan cara menjumlahkan semua nilai mata pelajaran (x_{jm}) yang valid dan membaginya dengan jumlah mata pelajaran (n_j) yang tersedia (tidak kosong). Pendekatan ini memastikan bahwa hanya nilai yang tersedia yang diperhitungkan, sehingga rata-rata yang dihasilkan tetap mencerminkan performa akademik aktual siswa.
b. Transformasi Variabel Ekstrakurikuler

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 744−753
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9598

Variabel ekstrakurikuler pada data awal berbentuk kategorikal dengan dua tingkat kualitas, yaitu SB (Sangat Baik) dan B (Baik). Agar dapat diproses oleh model klasifikasi, dilakukan pemetaan ke dalam bentuk numerik dengan ketentuan sebagai berikut:

SB = 2
B = 1

Berdasarkan pemetaan tersebut, dibentuk dua fitur turunan.

1. Fitur pertama adalah jumlah partisipasi aktif dalam kegiatan ekstrakurikuler:

$$\text{EkskulAktif}_j = \text{SB}_j + \text{B}_j \quad (2)$$

2. Fitur kedua adalah skor rata-rata kualitas partisipasi ekstrakurikuler:

$$\text{Skor Ekskul}_j = \begin{cases} \frac{2\cdot \text{SB}_j + 1\cdot \text{B}_j}{\text{SB}_j + \text{B}_j}, & \text{jika } \text{SB}_j + \text{B}_j > 0 \\ 0, & \text{jika } \text{SB}_j + \text{B}_j = 0 \end{cases} \quad (3)$$

SkorEkskul dihitung sebagai rata-rata tertimbang berdasarkan kualitas partisipasi. Apabila siswa tidak mengikuti kegiatan ekstrakurikuler ($SB_j + B_j = 0$), maka nilai SkorEkskul ditetapkan sebesar 0.

3.2.1 Pembentukan Label Target

Pembentukan label dilakukan setelah seluruh fitur selesai dibentuk. Dengan cara ini, setiap label benar-benar dihasilkan dari kombinasi nilai akademik, absensi, dan aktivitas ekstrakurikuler yang ada pada data siswa.

1. Berprestasi jika RataAkademik >= 88, Alpa = 0, dan SkorEkskul >= 1.
2. Cukup jika tidak termasuk Berprestasi, tetapi memiliki RataAkademik >= 85 dan Alpa <= 1.
3. Kurang jika tidak memenuhi dua kriteria sebelumnya.

Aturan tiga kelas ini dipilih agar hasil klasifikasi lebih informatif dan sesuai dengan kebutuhan penelitian, yakni membedakan siswa dengan capaian tinggi, menengah, dan rendah.

3.2.2 Pembagian Data (Data Splitting)

Pengumpulan data dilakukan dengan membagi sampel sebanyak 222 siswa ke dalam dua kelompok: data latih dan data uji. Pembagian data dilakukan dengan rasio 70:30, di mana 70% dari total data digunakan untuk pelatihan model dan 30% untuk pengujian. Berikut adalah distribusi data yang digunakan dapat dilihat pada tabel 3 berikut.

Tabel 3. Distribusi hasil pengujian model

Kelas	Data Uji (Jumlah)	Prediksi Benar	Prediksi Salah
Berprestasi	4	4	0
Cukup	38	38	0
Kurang	25	25	0

Penggunaan teknik stratified split menjadi penting dalam penelitian ini karena adanya ketidakseimbangan kelas. Dengan metode tersebut, model dilatih dan diuji pada distribusi data yang representatif, sehingga hasil evaluasi kinerja model lebih objektif dan tidak bias terhadap kelas mayoritas. Tahapan pembagian data ini menjadi dasar bagi proses pelatihan model Random Forest yang akan dibahas pada subbab berikutnya.

3.3 Perhitungan Manual Random Forest (Gini Index)

Distribusi Kelas Data Latih untuk Perhitungan Gini hanya menampilkan jumlah data pada tiap kelas, karena rumus Gini membutuhkan distribusi kelas, bukan identitas seluruh siswa satu per satu.

Tabel 4. Distribusi Kelas Data Latih

Kelas	Jumlah Data	Proporsi
Berprestasi	9	0.0581
Cukup	87	0.5613
Kurang	59	0.3806

Berdasarkan data latih terdiri dari 9 siswa Berprestasi, 87 siswa Cukup, dan 59 siswa Kurang.

1. Proporsi kelas pada node akar berturut-turut adalah :

$$\text{p(Berprestasi)} = \frac{9}{155}$$

$$\text{p(Cukup)} = \frac{87}{155}$$

$$\text{p(Kurang)} = \frac{59}{155}$$

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 744−753
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9598

2. Rumus Gini untuk tiga kelas adalah :

$$\text{Gini(D)} = 1 - \sum_{i=1}^{k} p_1^2 \quad (4)$$

Dengan memasukkan ketiga proporsi tersebut ke dalam rumus, diperoleh

$$\text{Gini(D)} = 1 - \left(\frac{9}{155}\right)^2 - \left(\frac{87}{155}\right)^2 - \left(\frac{59}{155}\right)^2$$

$$\text{Gini(D)} = 0{,}536691$$

Nilai ini menunjukkan bahwa node akar masih bercampur sehingga perlu dicari pemisahan terbaik pada langkah berikutnya. Langkah ini penting karena Gini pada node akar menjadi nilai acuan. Setiap kandidat split yang diuji harus menghasilkan Gini split yang lebih kecil daripada Gini awal agar pemisahan dianggap lebih baik.

$$\text{Gini}_{\text{split}}(A) = \frac{|D_1|}{|D|}\text{Gini}(D_1) + \frac{|D_2|}{|D|}\text{Gini}(D_2)$$

$$\text{Gain}(A) = \text{Gini}(D) - \text{Gini}_{\text{split}}(A)$$

Evaluasi Kandidat Split A: RataAkademik <= 84,96

$$p_+ = \frac{13}{222}, \quad p_- = \frac{209}{222}$$

$$\text{Gini(D)} = 1 - \left(\frac{13}{222}\right)^2 - \left(\frac{209}{222}\right)^2$$

$$\text{Gini(D)} = 1 - \frac{169}{49284} - \frac{43681}{49284}$$

$$\text{Gini(D)} = 1 - 0.003428 - 0.886313$$

$$\text{Gini(D)} = 0.110259$$

3.4 Implementasi Python

Implementasi sistem klasifikasi dilakukan menggunakan Python dengan pustaka pandas dan scikit-learn. Program terdiri dari dua file utama, yaitu file preprocessing untuk membentuk dataset dan label, serta file pipeline Random Forest untuk pelatihan dan evaluasi model.

```
[1]: %run bab4_preprocessing_from_excel.py

Preprocessing selesai.
Jumlah data: 222
Distribusi label:
Label
Cukup          125
Kurang          84
Berprestasi     13
Name: count, dtype: int64

[ ]:
```

Gambar 2. Hasil Proses Preprocessing Data

Sistem pada gambar 2. menampilkan informasi hasil preprocessing yang menunjukkan bahwa proses pengolahan data telah berhasil dilakukan. Berdasarkan hasil yang ditampilkan, jumlah data yang digunakan dalam penelitian ini sebanyak 222 data. Selain itu, sistem juga menampilkan distribusi label yang terdiri dari tiga kategori, yaitu Cukup sebanyak 125 data, Kurang sebanyak 84 data, dan Berprestasi sebanyak 13 data. Informasi ini menunjukkan bahwa dataset telah berhasil diproses dan siap digunakan pada tahap pemodelan selanjutnya.

```
[2]: %run bab4_random_forest_pipeline.py

Pipeline selesai.
Train size: 155, Test size: 67
Label source: derived_rule (Label)
Distribusi label train:
Label
Berprestasi     9
Cukup          87
Kurang         59
Name: count, dtype: int64
Distribusi label test:
Label
Berprestasi     4
Cukup          38
Kurang         25
Name: count, dtype: int64
Confusion Matrix:
                    Pred_Berprestasi  Pred_Cukup  Pred_Kurang
Actual_Berprestasi                 4           0            0
Actual_Cukup                       0          38            0
Actual_Kurang                      0           0           25
Accuracy: 1.0000
Precision Macro: 1.0000
Recall Macro: 1.0000
F1 Macro: 1.0000
Mean CV Accuracy (5-fold): 0.9865
Root split tree-1: RataAkademik <= 84.96000
Gini root: 0.536691
Gini split: 0.221767
Gini gain: 0.314924
Prediksi data uji disimpan ke: hasil_prediksi_data_uji.csv
Catatan validitas disimpan ke: catatan_validitas_model.txt
File legacy juga diperbarui: hasil_prediksi_semua_siswa.csv
```

Gambar 3. Hasil Proses Klasifikasi Menggunakan Random Forest

Berdasarkan hasil pengujian yang ditampilkan, model mampu melakukan klasifikasi terhadap tiga kategori label yaitu Berprestasi, Cukup, dan Kurang. Hasil evaluasi model menunjukkan nilai precision, recall, dan F1-

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 744−753
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9598

score sebesar 1.000, serta nilai akurasi rata-rata sebesar 0.9865 berdasarkan proses validasi silang (cross validation). Selain itu, sistem juga menampilkan nilai Gini root sebesar 0.536691, nilai Gini split sebesar 0.221767, serta Gini gain sebesar 0.314924 yang menunjukkan tingkat pemisahan data pada proses pembentukan pohon keputusan.

3.5 Hasil Pengujian

3.5.1 Confussion Matrix

Evaluasi kinerja model pada data uji dilakukan menggunakan confusion matrix multiclass dengan jumlah data uji sebanyak 67 siswa, yang terdiri dari 4 siswa Berprestasi, 38 siswa Cukup, dan 25 siswa Kurang. Berdasarkan tabel confusion matrix, seluruh data uji berhasil diprediksi sesuai dengan kelas aktualnya, di mana pada kelas Berprestasi, 4 data terprediksi benar tanpa adanya kesalahan klasifikasi, pada kelas Cukup, 38 data terprediksi benar tanpa kesalahan, dan pada kelas Kurang, 25 data juga terprediksi dengan benar tanpa adanya kesalahan klasifikasi. Hasil ini menunjukkan bahwa matriks kebingungan hanya berisi nilai pada diagonal utama, sementara seluruh elemen di luar diagonal bernilai nol. Meskipun hasil pengujian pada data uji mencapai 100%, interpretasi ilmiah harus dilakukan dengan hati-hati karena label yang digunakan dibentuk dari aturan berbasis fitur yang sama dengan fitur masukan model, yang berpotensi menyebabkan bias dalam evaluasi. Analisis dominasi siswa berprestasi di kelas XI-1 menunjukkan bahwa kelas ini memiliki proporsi yang lebih tinggi dibandingkan dengan kelas lain. Hal ini mungkin disebabkan oleh faktor-faktor seperti tingkat keterlibatan yang lebih tinggi dalam kegiatan akademik maupun non-akademik, serta kemungkinan adanya dukungan yang lebih kuat dari segi pembelajaran atau lingkungan sosial di kelas tersebut, yang memfasilitasi pencapaian akademik yang lebih baik.

3.5.2 Perhitungan Matrix Evaluasi

Metrik evaluasi dihitung menggunakan pendekatan macro average untuk tiga kelas.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Accuracy} = \frac{4 + 63}{4 + 63 + 0 + 0} = 1.0000$$

Accuracy = 1,0000 menunjukkan seluruh 67 data uji diklasifikasikan dengan benar.
Precision Macro dihitung sebagai rata-rata nilai precision pada kelas Berprestasi, Cukup, dan Kurang.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Precision} = \frac{4}{4 + 0} = 1.0000$$

Precision Macro = 1,0000 menunjukkan ketepatan prediksi rata-rata antar kelas berada pada nilai sempurna.
Recall Macro dihitung sebagai rata-rata nilai recall pada kelas Berprestasi, Cukup, dan Kurang.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{4}{4 + 0} = 1.0000$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{4}{4 + 0} = 1.0000$$

Recall Macro = 1,0000 menunjukkan seluruh anggota tiap kelas berhasil dikenali oleh model.
F1 Macro dihitung sebagai rata-rata harmonik precision dan recall pada seluruh kelas.

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{F1} = 2 \cdot \frac{1.0 \cdot 1.0}{1.0 + 1.0} = 1.0000$$

F1 Macro = 1,0000 dan rata-rata akurasi validasi silang 5-fold sebesar 0,9865 menunjukkan model sangat konsisten pada data penelitian ini.

3.5.3 Sebaran Hasil Prediksi

Berdasarkan hasil klasifikasi terhadap seluruh dataset, model mengidentifikasi 13 siswa sebagai berprestasi. Sebaran siswa tersebut per kelas adalah

1. XI-1: 11 siswa
2. XI-3: 1 siswa
3. XI-5: 1 siswa

Tabel 5. Hasil Prediksi semua Siswa

No	Kelas	Nama	Rata Akademik	S	I	A	Ekskul Aktif	Skor Ekskul	Label	Prediksi
1	XI-2	Aida Nurul Fadillah	86.17	3	4	0	0	0.0	Cukup	Cukup

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 744−753
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9598

No	Kelas	Nama	Rata Akademik	S	I	A	Ekskul Aktif	Skor Ekskul	Label	Prediksi
2	XI-1	Asni Ameilia Nst	87.25	3	0	0	2	2.0	Cukup	Cukup
3	XI-2	Anisa Rahma	85.83	0	1	0	0	0.0	Cukup	Cukup
4	XI-1	Lastiar Pesta Siahaan	91.42	0	0	0	1	2.0	Berprestasi	Berprestasi
5	XI-4	Rian Natasya Simbolon	86.75	1	3	0	0	0.0	Cukup	Cukup
6	XI-3	Juwita Sari	86.42	4	2	0	0	0.0	Cukup	Cukup
7	XI-6	Marsha Deby Amelia Putri	86.75	0	4	0	0	0.0	Cukup	Cukup
8	XI-2	Rika Susanti	85.5	13	6	3	0	0.0	Kurang	Kurang
9	XI-6	Amiruddin	84.5	8	0	2	0	0.0	Kurang	Kurang
10	XI-4	Aura Kasih Lubis	87.83	2	1	0	1	2.0	Cukup	Cukup
…	…	…	…	…	…	…	…	…	…	…
67	XI-5	Putri Melya Sari	87.0	3	0	0	1	1.0	Cukup	Cukup

Distribusi pada tabel 5 menunjukkan bahwa mayoritas siswa berprestasi berasal dari kelas XI-1. Secara deskriptif, hal ini dapat mencerminkan variasi kualitas akademik antar kelas, meskipun interpretasi lebih lanjut memerlukan analisis tambahan terhadap karakteristik masing-masing kelas.

4. KESIMPULAN

Berdasarkan proses penelitian dan pengujian, model klasifikasi siswa berprestasi berhasil dibangun menggunakan algoritma Random Forest berbasis Python melalui Jupyter Notebook, dengan memanfaatkan data sebanyak 222 siswa dari 7 kelas yang telah melalui tahapan preprocessing, meliputi cleaning, transformasi fitur, pembentukan label, serta pembagian data latih dan uji dengan rasio 70:30. Klasifikasi dilakukan menggunakan enam variabel utama, yaitu RataAkademik, Sakit, Izin, Alpa, EkskulAktif, dan SkorEkskul, sehingga penilaian tidak hanya berfokus pada aspek akademik tetapi juga mencakup kedisiplinan dan keterlibatan ekstrakurikuler secara lebih menyeluruh. Hasil evaluasi menunjukkan bahwa model Random Forest memiliki kinerja yang sangat baik dengan tingkat akurasi, precision, recall, dan F1-score sebesar 1.0000 pada data uji, serta rata-rata akurasi sebesar 0.9865 pada validasi silang. Analisis feature importance mengindikasikan bahwa RataAkademik menjadi faktor paling dominan, diikuti oleh Alpa, EkskulAktif, dan SkorEkskul. Secara keseluruhan, metode Random Forest dapat digunakan sebagai pendekatan sistematis untuk membantu pihak sekolah dalam mengklasifikasikan siswa berprestasi secara lebih objektif dan terukur. Sebagai saran pengembangan, penelitian lebih lanjut dapat mempertimbangkan penggunaan data yang lebih bervariasi dan teknik algoritma lainnya untuk meningkatkan akurasi dan relevansi dalam konteks pendidikan yang lebih luas.

REFERENCES

[1] A. Muhaimin, M. A. Hariyadi, and M. Imamudin, “Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor dan Kedisiplinan dengan Metode K-Nearest Neighbor,” J. Ilmu Komput. dan Sist. Inf., vol. 7, no. 1, pp. 193–202, 2024, doi: https://doi.org/10.55338/jikomsi.v7i1.2865.

[2] A. M. Husein and R. E. H. Hutauruk, “Penerapan Algoritma C4.5 Dalam Pemilihan Siswa Berprestasi di SMPN 10 Medan,” Digit. Transform. Technol., vol. 2, no. 1, pp. 8–11, 2022, doi: https://doi.org/10.47709/digitech.v2i1.1768.

[3] S. Apandi, “Pemodelan Klasifikasi Siswa Berprestasi dengan Random Forest : Studi Kasus pada Bimbingan Belajar,” Fakt. Exacta, vol. 18, no. 1, pp. 63–71, 2025, doi: https://doi.org/10.30998/faktorexacta.v18i1.27163 ?

[4] H. Amalia, A. Puspita, A. F. Lestari, and Frieyadie, “APPLICATION OF DECISION TREE AND NAIVE BAYES ON STUDENT PERFORMANCE DATASET,” J. Pilar Nusa Mandiri, vol. 18, no. 1, pp. 53–58, 2022, doi: https://doi.org/10.33480/pilar.v18i1.2714.

[5] R. N. J. S.Intam, Wulandari, A. A. N. Risal, and D. F. Surianto, “Klasifikasi Mahasiswa Berprestasi Menggunakan Fuzzy C-Means Dan Naive Bayes,” J. Ilm. Inform. Glob., vol. 15, no. 1, pp. 9–16, 2024, doi: https://doi.org/10.36982/jiig.v15i1.3666.

[6] P. Ramadhan, Yuhandri, and J. Veri, “Eksplorasi Algoritma Decision Tree untuk Penentuan Siswa Berprestasi,” J. Bit-Tech, vol. 7, no. 3, pp. 826–833, 2025, doi: https://doi.org/10.32877/bt.v7i3.2210.

[7] G. M. Agung, R. A. Zuama, and E. S. Budi, “Analysis of Student Academic Performance Using Random Forest and Support Vector Machines,” Comput. Sci., vol. 6, no. 1, pp. 57–65, 2026, doi: https://doi.org/10.31294/co-science.v6i1.10123.

[8] S. Zawiyah, Lailatul Qodriyah, and M. B. Tamam, “Klasifikasi Prestasi Akademik Mahasiswa Menggunakan Metode Random Forest,” J. Digit. Bus. Inf. Technol. Klasifikasi, vol. 1, no. 2, pp. 61–71, 2024, doi: 10.23971/jobit.v1i2.317.

[9] Y. Y. Asyel, R. S. Mokodompit, J. Ligouw, and A. Yusupa, “Meningkatkan Prestasi Akademik Mahasiswa Teknik

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 744−753
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9598

Informatika UNSRAT Melalui Optimasi Pembelajaran dengan Random Forest," J. Mhs. Tek. Inform., vol. 4, no. 1, pp. 201–209, 2025, doi: https://doi.org/10.35473/jamastika.v4i1.3940.

[10] R. Fitriani, A. P. Windarto, and I. Gunawan, "Penerapan Algoritma Naive Bayes untuk Klasifikasi Tingkat Pemahaman Siswa pada Pembelajaran Daring," J. Media Inform. Budidarma, vol. 6, no. 2, pp. 1040–1047, 2022, doi: https://doi.org/10.30865/mib.v6i2.4001.

[11] E. Amos and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," J. Inf. Syst. Informatics, vol. 3, no. 3, pp. 408–419, 2021, doi: https://doi.org/10.51519/journalisi.v3i3.164.

[12] Selipuri, R. F. Purnomo, and Y. Yuniarthe, "Penerapan Algoritma C4.5 untuk Klasifikasi Tingkat Kedisiplinan Siswa," J. Informatics, Electr. Electron. Eng., vol. 5, no. 1, pp. 33–40, 2024, doi: https://doi.org/10.47065/jieee.v5i1.2630.

[13] B. Q. Husaini and Jemakmun, "Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Penjurusan Siswa," J. Teknol. Inform. dan Komputer2, vol. 9, no. 1, pp. 55–62, 2023, doi: https://doi.org/10.37012/jtik.v9i1.1512.

[14] A. Suciko, Y. Hendriyani, K. Budayawan, and Syafrijon, "Penerapan Data Mining Untuk Klasifikasi Calon Siswa Penerima Program Indonesia Pintar (PIP) Menggunakan Algoritma Naive Bayes," J. Pendidik. Tambusai, vol. 9, no. 2, pp. 15266–15274, 2023, doi: https://doi.org/10.31004/jptam.v9i2.27930.

[15] M. Nachouki and M. A. Naaj, "Predicting Student Performance to Improve Academic Advising Using the Random Forest Algorithm," Int. J. Acad. Res. Bus. Soc. Sci. Educ. Technol., vol. 20, no. 1, pp. 1–17, 2022, doi: 10.4018/IJDET.296702.

[16] Mahyudi, Endaryono, and R. Ristiawan, "Predicting Student Final Grades Using Random Forest Algorithms and Linear Regression," J. Sist. Cerdas, vol. 8, no. 3, pp. 441–448, 2025, doi: https://doi.org/10.37396/jsc.v8i3.618.

[17] D. Kurniasari, R. N. Hidayah, Notiragayu, Warsono, and R. K. Nisa, "CLASSIFICATION MODELS FOR ACADEMIC PERFORMANCE: A COMPARATIVE STUDY OF NAÏVE BAYES AND RANDOM FOREST ALGORITHMS IN ANALYZING UNIVERSITY OF LAMPUNG STUDENT GRADES," J. Tek. Inform., vol. 5, no. 5, pp. 1267–1276, 2024, doi: https://doi.org/10.52436/1.jutif.2024.5.5.2066.

[18] S. Sivakumar and S. Venkataraman, "Evaluating Machine Learning Approaches: A Comparative Study of Random Forest and Neural Networks in Grade Classification," Indones. J. Data Sci., vol. 6, no. 1, pp. 73–80, 2025, doi: https://doi.org/10.56705/ijodas.v6i1.240.

[19] V. N. Wijayaningrum, A. P. Kirana, and I. K. Putri, "STUDENT ACADEMIC PERFORMANCE PREDICTION FRAMEWORK WITH FEATURE SELECTION AND IMBALANCED DATA HANDLING," J. Ilm. Kursor, vol. 12, no. 3, pp. 10–18, 2024, doi: https://doi.org/10.21107/kursor.v12i3.356.

[20] M. Gusnina, Wiharto, and U. Salamah, "Student Performance Prediction in Sebelas Maret University Based on the Random Forest Algorithm," Int. Inf. Eng. Technol. Assoc., vol. 27, no. 3, pp. 495–501, 2022, doi: https://doi.org/10.18280/isi.270317.