

Prediksi Siswa Berprestasi di SMP Negeri 4 Kediri Menggunakan Algoritma *Random Forest*

^{1*} **Apliana Sriyanti Bili**, ² **Intan Nur Farida**,
^{1,2} Teknik Informatika, Universitas Nusantara PGRI Kediri
E-mail: ¹aplbianabili040@gmail.com, ²intanf@unpkediri.ac.id
Penulis Korespondens : Apliana Sriyanti Bili

Abstrak—Penentuan siswa berprestasi merupakan kegiatan penting di sekolah karena dapat menjadi dasar dalam pembinaan dan pemberian penghargaan kepada siswa. Namun, proses penilaian yang masih dilakukan secara manual sering membutuhkan waktu lama dan berpotensi menghasilkan keputusan yang kurang objektif. Penelitian ini menerapkan algoritma *Random Forest* pada sistem prediksi siswa berprestasi di SMP Negeri 4 Kediri. Data yang digunakan meliputi nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan keaktifan ekstrakurikuler. Tahapan penelitian meliputi pengumpulan data, *preprocessing*, pelatihan model, pengujian, evaluasi, dan implementasi sistem berbasis web. Hasil evaluasi menunjukkan bahwa model memperoleh *accuracy* sebesar 87,50%, *precision* sebesar 91,67%, *recall* sebesar 78,57%, dan *F1-score* sebesar 84,62%. Hasil tersebut menunjukkan bahwa sistem dapat membantu sekolah mengidentifikasi siswa berprestasi secara lebih cepat, objektif, dan berbasis data.

Kata Kunci— klasifikasi, *machine learning*, prediksi, *Random Forest*, siswa berprestasi

Abstract— *Determining outstanding students is an important activity in schools because it can serve as a basis for student guidance and award allocation. However, manual assessment often takes more time and may produce less objective decisions. This study applies the Random Forest algorithm to a prediction system for outstanding students at SMP Negeri 4 Kediri. The data used include report card scores, attendance, discipline, competition achievements, and extracurricular activity. The research stages include data collection, preprocessing, model training, testing, evaluation, and web-based system implementation. The evaluation results show that the model achieved accuracy of 87.50%, precision of 91.67%, recall of 78.57%, and F1-score of 84.62%. These results indicate that the system can help schools identify outstanding students more quickly, objectively, and based on data.*

Keywords— *classification, machine learning, prediction, Random Forest, outstanding students*

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Perkembangan teknologi informasi memberikan peluang besar bagi sekolah dalam mengelola data siswa. Data yang dimiliki sekolah tidak hanya dapat disimpan sebagai arsip, tetapi juga dapat diolah menjadi informasi yang membantu proses pengambilan keputusan. Data seperti nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan keaktifan ekstrakurikuler dapat memberikan gambaran yang lebih lengkap mengenai kemampuan siswa [1]

Penentuan siswa berprestasi tidak sebaiknya hanya dilihat dari nilai akademik. Siswa yang memiliki kehadiran baik, disiplin, aktif mengikuti kegiatan sekolah, dan memiliki prestasi di luar kelas juga perlu mendapat perhatian. Beberapa indikator yang dapat digunakan dalam penilaian siswa berprestasi meliputi nilai rata-rata rapor, kehadiran, kedisiplinan, prestasi atau

penghargaan, serta partisipasi dalam kegiatan ekstrakurikuler [2]. Dengan menggunakan beberapa indikator tersebut, proses penilaian dapat menjadi lebih menyeluruh.

Pada praktiknya, penentuan siswa berprestasi masih sering dilakukan secara manual. Cara tersebut memang dapat digunakan, tetapi memiliki beberapa kelemahan. Proses manual membutuhkan waktu lebih lama, terutama jika jumlah data siswa cukup banyak. Selain itu, hasil penilaian juga dapat berbeda karena bergantung pada pertimbangan masing-masing penilai.

Machine learning dapat digunakan sebagai salah satu pendekatan untuk membantu proses tersebut. Dalam *machine learning*, data yang sudah tersedia dapat dipelajari oleh sistem untuk menemukan pola tertentu. Salah satu teknik yang sesuai untuk penelitian ini adalah klasifikasi, yaitu proses mengelompokkan data ke dalam kelas tertentu berdasarkan pola atau karakteristik pada data sebelumnya [3].

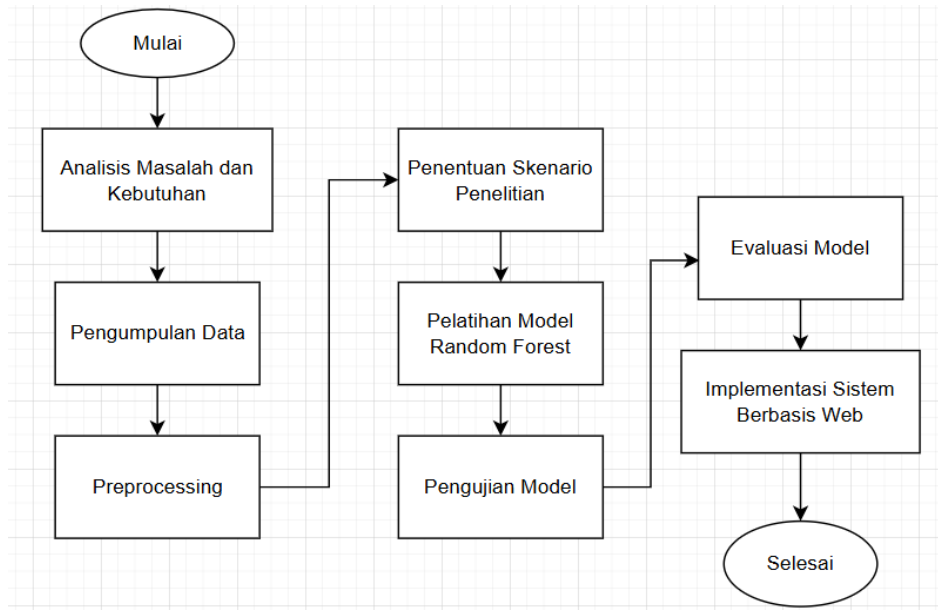
Algoritma yang digunakan dalam penelitian ini adalah *Random Forest*. *Random Forest* merupakan metode klasifikasi yang membangun beberapa pohon keputusan, kemudian menentukan hasil akhir berdasarkan voting mayoritas [4]. Metode ini banyak digunakan karena mampu mengolah beberapa atribut sekaligus dan menghasilkan prediksi yang stabil. Pada bidang pendidikan, *Random Forest* telah digunakan untuk membantu prediksi performa siswa dan mendukung proses bimbingan akademik [5].

Beberapa penelitian sebelumnya juga menunjukkan bahwa *Random Forest* dapat digunakan untuk mengolah data pendidikan. Beberapa penelitian menjelaskan bahwa data akademik dan non-akademik dapat membantu analisis performa siswa [6]. Penelitian ini menunjukkan bahwa *Random Forest* dapat digunakan untuk menangani data pendidikan dengan beberapa atribut berbeda [7]. Selain itu, penelitian lain juga menunjukkan bahwa *Random Forest* mampu digunakan dalam klasifikasi siswa terbaik dengan hasil yang baik [8].

II. METODE

Penelitian ini menggunakan metode pengembangan sistem berbasis web dengan menerapkan algoritma *Random Forest* sebagai metode klasifikasi. Sistem dirancang untuk memprediksi siswa ke dalam dua kategori, yaitu Berprestasi dan Tidak Berprestasi. Tahapan penelitian meliputi pengumpulan data, *preprocessing* data, pelatihan model, pengujian model, evaluasi model, dan implementasi sistem.

Alur penelitian dibuat untuk memberikan gambaran tahapan proses yang dilakukan dalam penelitian sistem prediksi siswa berprestasi menggunakan algoritma *Random Forest*. Proses penelitian dimulai dari analisis masalah dan kebutuhan [9], pengumpulan data siswa, kemudian dilakukan *preprocessing* data agar dapat diproses oleh sistem. Selanjutnya data dibagi menjadi data *training* dan *testing* untuk proses pelatihan dan pengujian model. Setelah itu dilakukan evaluasi performa model sebelum sistem diimplementasikan dalam bentuk aplikasi berbasis *web*. Alur penelitian tersebut ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

Gambar 1. menunjukkan alur penelitian sistem prediksi siswa berprestasi menggunakan algoritma *Random Forest*. Penelitian dimulai dari pengumpulan data siswa, kemudian dilakukan preprocessing data, pembagian data *training* dan *testing*, pelatihan model, pengujian model, evaluasi model, hingga implementasi sistem berbasis *web* untuk membantu proses prediksi siswa berprestasi.

Skenario pengujian pada penelitian ini dilakukan untuk mengetahui kemampuan algoritma *Random Forest* dalam mengklasifikasikan siswa ke dalam kategori Berprestasi dan Tidak Berprestasi. Dataset yang digunakan berjumlah 105 data siswa kelas VIII SMP Negeri 4 Kediri tahun ajaran 2025/2026. Data tersebut dibagi menjadi data *training* 73 siswa (70%) dan data *testing* 32 siswa (30%). Data *training* digunakan untuk melatih model agar dapat mengenali pola dari atribut nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan keaktifan ekstrakurikuler, sedangkan data *testing* digunakan untuk menguji kemampuan model dalam memprediksi siswa baru yang belum pernah digunakan pada proses pelatihan.

Pengujian dilakukan dengan beberapa tahapan. Pertama, data siswa melalui proses *preprocessing*, yaitu mengubah data mentah menjadi bentuk numerik sesuai rubrik penilaian yang telah ditentukan. Setelah data siap, model *Random Forest* dilatih menggunakan data *training* sebanyak 73 siswa dengan parameter $n_estimators = 50$, $max_depth = 4$, dan $random_state = 42$. Selanjutnya, model diuji menggunakan 32 data *testing*. Hasil prediksi dari data *testing* kemudian dibandingkan dengan label sebenarnya untuk memperoleh *confusion matrix*, dari mana dihitung *accuracy*, *precision*, *recall*, dan *F1-score* sebagai ukuran performa model dalam melakukan klasifikasi siswa berprestasi.

Algoritma *Random Forest* bekerja dengan membentuk beberapa pohon keputusan (*decision tree*). Setiap pohon mempelajari pola dari data *training* berdasarkan fitur yang digunakan, yaitu nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan keaktifan ekstrakurikuler. Ketika data

testing dimasukkan ke dalam model, setiap pohon menghasilkan prediksi masing-masing. Hasil akhir klasifikasi ditentukan berdasarkan voting mayoritas dari seluruh pohon keputusan. Jika mayoritas pohon memprediksi siswa sebagai Berprestasi, maka hasil akhir sistem adalah Berprestasi. Sebaliknya, jika mayoritas pohon memprediksi siswa sebagai Tidak Berprestasi, maka hasil akhir sistem adalah Tidak Berprestasi. Dengan mekanisme ini, *Random Forest* mampu menghasilkan prediksi yang lebih stabil dibandingkan menggunakan satu pohon keputusan, sekaligus mencerminkan performa realistik pada data testing terbaru.

Sistem prediksi siswa berprestasi dikembangkan dalam bentuk aplikasi berbasis *web*. Sistem dibuat menggunakan bahasa *pemrograman Python*, *framework Flask*, *database MySQL*, serta beberapa library pendukung seperti *pandas*, *scikit-learn*, *joblib*, dan *mysql.connector*. Fitur utama dalam sistem meliputi pengelolaan data siswa, *input data* siswa, *import CSV*, proses prediksi, riwayat prediksi, dan evaluasi model.

Evaluasi model dilakukan untuk mengetahui kinerja algoritma *Random Forest* dalam melakukan klasifikasi. Metrik evaluasi yang digunakan adalah *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. *Accuracy* digunakan untuk melihat tingkat ketepatan model secara keseluruhan. *Precision* digunakan untuk mengetahui ketepatan model dalam memprediksi siswa berprestasi. *Recall* digunakan untuk melihat kemampuan model dalam menemukan siswa yang benar-benar berprestasi. *F1-score* digunakan untuk melihat keseimbangan antara *precision* dan *recall* [10], [11]

Accuracy, *precision*, *recall*, dan *F1-score* digunakan untuk mengevaluasi kinerja model *Random Forest*. Rumusnya adalah sebagai berikut:

1. Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times \dots\dots\dots (1)$$

Keterangan:

- TP = *True Positive*
- TN = *True Negative*
- FP = *False Positive*
- FN = *False Negative*

2. Precision

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

3. Recall

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (3)$$

4. F1-Score

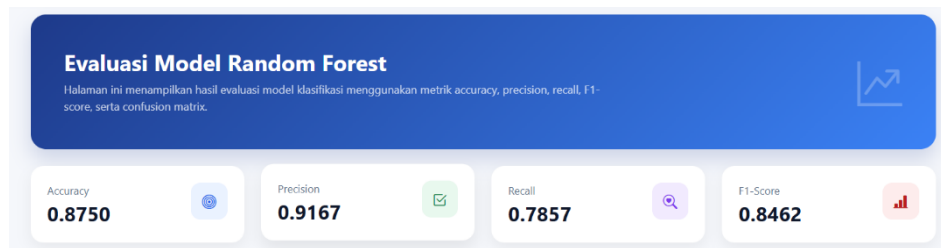
$$F1 \text{ Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \dots\dots\dots (4)$$

III. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan sistem prediksi siswa berprestasi berbasis web dengan menerapkan algoritma *Random Forest Classifier*. Sistem digunakan untuk membantu pihak sekolah dalam mengklasifikasikan siswa ke dalam kategori Berprestasi dan Tidak Berprestasi berdasarkan data akademik dan non-akademik.

Sistem yang dikembangkan memiliki beberapa fitur utama, yaitu pengelolaan data siswa, input data siswa, *import data CSV*, prediksi siswa, riwayat prediksi, dan evaluasi model. Data siswa yang dimasukkan ke dalam sistem akan melalui proses *preprocessing* terlebih dahulu. Pada tahap ini, data seperti nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan keaktifan ekstrakurikuler dikonversi menjadi data numerik agar dapat diproses oleh model *Random Forest*.

Setelah data diproses, sistem melakukan prediksi untuk menentukan apakah siswa termasuk kategori Berprestasi atau Tidak Berprestasi. Hasil prediksi tersebut kemudian dapat disimpan sebagai riwayat prediksi. Selain menghasilkan prediksi, sistem juga menampilkan hasil evaluasi model untuk mengetahui tingkat keberhasilan algoritma dalam melakukan klasifikasi.



Gambar 2. Evaluasi model *Random Forest*

Gambar 2. menunjukkan halaman evaluasi model *Random Forest* yang menampilkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Tampilan ini digunakan untuk melihat performa model setelah proses pelatihan dan pengujian dilakukan.

Berdasarkan evaluasi model *Random Forest* tersebut, diperoleh hasil perhitungan *accuracy*, *precision*, *recall*, dan *F1-score* sebagai berikut.

Accuracy digunakan untuk mengukur tingkat ketepatan model dalam memprediksi seluruh data.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

$$Accuracy = \frac{11 + 17}{11 + 17 + 1 + 3}$$

$$Accuracy = \frac{28}{32}$$

$$Accuracy = 0.875 \quad 87,50\% \\ \text{atau}$$

Precision digunakan untuk mengukur ketepatan model ketika memprediksi siswa sebagai kategori Berprestasi.

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

$$Precision = \frac{11}{11 + 1}$$

$$Precision = \frac{11}{12}$$

$$Precision = 0.9167 \quad 91.67\% \\ \text{atau}$$

Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh siswa yang benar-benar berprestasi.

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{11}{11 + 3}$$

$$Recall = \frac{11}{14}$$

$$Recall = 0,7857 \quad 78,57\% \\ \text{atau}$$

F1-score digunakan untuk melihat keseimbangan antara *precision* dan *recall*.

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

$$F1 \text{ Score} = \frac{2 \times 0.9167 \times 0.7857}{0.9167 + 0.7857}$$

$$F1 \text{ Score} = F1 \text{ Score} = 0.8462 \quad 84.62\% \\ \text{atau}$$

Tabel 1. Evaluasi Model

No	Evaluasi Model	
	Metrik Evaluasi	Nilai
1	Accuracy	87,50%
2	Precision	91,67%
3	Recall	78,57%
4	F1-score	84,62%

Berdasarkan Tabel 1, model *Random Forest* memperoleh *accuracy* sebesar 87,50%, *precision* sebesar 91,67%, *recall* sebesar 78,57%, dan *F1-score* sebesar 84,62%. Nilai *accuracy* sebesar 87,50% menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data testing dengan benar. Nilai *precision* sebesar 91,67% menunjukkan bahwa prediksi siswa berprestasi yang dihasilkan model memiliki tingkat ketepatan yang tinggi. Nilai *recall* sebesar 78,57% menunjukkan bahwa sebagian siswa berprestasi tidak terdeteksi oleh model (*False Negative*), sehingga hasil prediksi lebih realistis. Sementara itu, nilai *F1-score* sebesar 84,62% menunjukkan bahwa model memiliki keseimbangan yang baik antara *precision* dan *recall*.

IV. KESIMPULAN

Penelitian ini menerapkan algoritma *Random Forest* dalam sistem prediksi siswa berprestasi di SMP Negeri 4 Kediri. Data yang digunakan merupakan data siswa kelas VIII tahun ajaran 2025/2026 dengan jumlah keseluruhan 105 data siswa. Atribut yang digunakan meliputi nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan keaktifan ekstrakurikuler. Sistem yang dikembangkan berbasis *web* dan dapat membantu proses klasifikasi siswa ke dalam kategori Berprestasi dan Tidak Berprestasi secara lebih cepat, terstruktur, dan berbasis data.

Hasil evaluasi menunjukkan bahwa model *Random Forest* memperoleh *accuracy* sebesar 87,50%, *precision* sebesar 91,67%, *recall* sebesar 78,57%, dan *F1-score* sebesar 84,62%. Nilai *F1-score* tersebut menunjukkan bahwa model memiliki keseimbangan yang baik antara *precision* dan *recall*. Berdasarkan hasil evaluasi ini, algoritma *Random Forest* dapat digunakan sebagai metode klasifikasi untuk membantu sekolah dalam mengidentifikasi siswa berprestasi secara lebih objektif dan akurat, meskipun beberapa siswa berprestasi masih tidak terdeteksi (*False Negative*).

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing atas bimbingan, arahan, dan masukan yang diberikan sehingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] D. Hartina and A. Siahaan, "Manajemen Kegiatan Ekstrakurikuler dalam Meningkatkan Prestasi Non Akademik Siswa di SMA Negeri 1 Aek Natas," *Journal of Education Research*, vol. 5, no. 2, p p. 2024–2033, Jun. 2024, doi: 10.37985/jer.v5i2.1117.

- [2] A. Muhaimin, M. Amin Hariyadi, and M. I. Imamudin, "Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor dan Kedisiplinan dengan Metode K-Nearest Neighbor," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 1, pp. 193–202, Feb. 2024, doi: 10.55338/jikoms.v7i1.2865.
- [3] C.-G. Wong, G.-K. Tong, and S.-C. Haw, "Exploring Customer Segmentation in E-Commerce using RFM Analysis with Clustering Techniques," *Journal of Telecommunications and the Digital Economy*, vol. 12, no. 3, pp. 97–125, Sep. 2024, doi: 10.18080/jtde.v12n3.978.
- [4] A. Ferdita Nugraha, R. F. A. Aziza, and Y. Pristyanto, "Penerapan metode Stacking dan Random Forest untuk Meningkatkan Kinerja Klasifikasi pada Proses Deteksi Web Phishing," *Jurnal Infomedia*, vol. 7, no. 1, p. 39, Jun. 2022, doi: 10.30811/jim.v7i1.2959.
- [5] R. Amin and A. S. F. Utami, "Prediksi Nilai Ujian Berdasarkan Kebiasaan Siswa Menggunakan Algoritma Random Forest Regressor," *INFORMATION SYSTEM FOR EDUCATORS AND PROFESSIONALS: Journal of Information System*, vol. 10, no. 2, p. 149, Dec. 2025, doi: 10.51211/isbi.v10i2.3722.
- [6] T. Pattiasina and D. Rosiyadi, "COMPARISON OF DATA MINING CLASSIFICATION ALGORITHM FOR PREDICTING THE PERFORMANCE OF HIGH SCHOOL STUDENTS," *Jurnal Techno Nusa Mandiri*, vol. 17, no. 1, pp. 22–30, Mar. 2020, doi: 10.33480/techno.v17i1.1226.
- [7] S. Sobari, A. I. Purnamasari, A. Bahtiar, and K. Kaslani, "MENINGKATKAN MODEL PREDIKSI KELULUSAN SANTRI TAHFIDZ DI PONDOK PESANTREN AL-KAUTSAR MENGGUNAKAN ALGORITMA RANDOM FOREST," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5704.
- [8] A. Darmawan, I. Yudhisari, A. Anwari, and M. Makruf, "Pola Prediksi Kelulusan Siswa Madrasah Aliyah Swasta dengan Support Vector Machine dan Random Forest," *Jurnal Minfo Polgan*, vol. 12, no. 1, pp. 387–400, May 2023, doi: 10.33395/jmp.v12i1.12388.
- [9] Sanjaya. A, *Front-End Development : Membangun Antarmuka Pengguna yang interaktif dan Responsif*. PT. Litnus Malang Indonesia, 2025. ISBN 978-634-234-721-8
- [10] P. Ramadhan, Yuhandri, and J. Veri, "Eksplorasi Algoritma Decision Tree untuk Penentuan Siswa Berprestasi," *bit-Tech*, vol. 7, no. 3, pp. 826–833, Apr. 2025, doi: 10.32877/bt.v7i3.2210.
- [11] H. Barus, I. N. Fajri, and Y. Pristyanto, "Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2552–2561, Oct. 2025, doi: 10.30871/jaic.v9i5.10524.