

A Random Forest-Based Predictive Model for Student Academic Performance: A Case Study in Indonesian Public High Schools

Rifa Andriani Saputri ^{1*}, Asrianda ^{2*}, Lidya Rosnita ^{3*}

* Teknik Informatika, Universitas Malikussaleh

srifaandriani@gmail.com ¹, asrianda@unimal.ac.id ², lidyarosnita@unimal.ac.id ³

Article Info

Article history:

Received 2025-04-20

Revised 2025-05-04

Accepted 2025-05-07

Keyword:

*Achievement Accuracy,
Entropy,
Education,
Prediction.*

ABSTRACT

The rapid advancement of information technology has transformed education by providing tools to accurately predict students' academic performance. This study aims to develop a system for predicting academic achievement using the Random Forest algorithm, with a case study at SMAN 1 Aceh Barat Daya and SMAN 3 Aceh Barat Daya. Data from 632 student report cards for grades X and XI in the second semester of the 2023/2024 academic year were used, covering subjects such as Mathematics, Indonesian Language, and others, divided into 80% training data (506 samples) and 20% test data (136 samples). The research methodology involved data preprocessing, training the Random Forest model using entropy and information gain to construct decision trees, and performance evaluation using metrics such as accuracy, precision, and recall. The implementation resulted in a web-based application using Python and Flask, featuring an interactive interface and decision tree visualization. Testing on 136 test samples achieved an accuracy of 87.40%, with 111 correct predictions, 16 false positives, and 0 false negatives, demonstrating the model's reliability in identifying high-achieving students without missing potential. This research is expected to assist schools in identifying outstanding students, making data-driven decisions, and designing more effective educational strategies.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Pendidikan merupakan salah satu faktor kunci dalam pembangunan suatu bangsa karena berperan penting dalam mencetak sumber daya manusia yang berkualitas. Perundang-undangan sistem pendidikan No.20 tahun 2003 mendefinisikan pendidikan sebagai “usaha sadar dan terencana untuk mewujudkan suasana belajar dan pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan sepiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia serta keterampilan yang diperlukan dirinya dan masyarakat” [1].

SMAN 1 Aceh Barat Daya dan SMAN 3 Aceh Barat Daya menghadapi tantangan dalam meningkatkan prestasi akademik siswa akibat keterbatasan dukungan akademik antara guru dan siswa. Oleh karena itu, penerapan algoritma

Random Forest menjadi solusi dalam menganalisis dan memprediksi prestasi akademik secara akurat. Random Forest merupakan teknik pembelajaran mesin yang membangun beberapa pohon keputusan (*decision tree*) untuk memperoleh prediksi yang lebih akurat dengan keunggulan dalam menangani data tidak seimbang, mengurangi risiko *overfitting*, serta memberikan informasi mengenai variabel penting dalam model prediksi. Pada penelitian ini [2]

Sekolah SMAN 1 Aceh Barat Daya dan SMAN 3 Aceh Barat Daya memiliki permasalahan pada akademik dimana kurangnya dukungan akademik antar guru dan siswa pada kedua sekolah tersebut. Sekolah memiliki keterbatasan dalam menentukan prestasi – prestasi akademik pada siswa. Penelitian sebelumnya membuktikan efektivitas algoritma ini, seperti yang dilakukan oleh [3] dalam sistem klasifikasi penjurusan siswa di MAN Sintang dengan akurasi 94,38%,

serta penelitian [4] dalam prediksi rating aplikasi App Store dengan akurasi 86,27%.

Oleh karena itu, penelitian ini bertujuan untuk membangun sistem prediksi prestasi akademik siswa di SMAN 1 Aceh Barat Daya dan SMAN 3 Aceh Barat Daya dengan menerapkan algoritma Random Forest, guna membantu sekolah dalam mengidentifikasi siswa berprestasi, menyusun strategi pendidikan yang lebih baik, serta meningkatkan kualitas pembelajaran berbasis data. Penelitian ini menggunakan data nilai rapor siswa kelas X dan XI dari semester genap tahun ajaran 2023/2024, mencakup berbagai mata pelajaran. Dengan penerapan teknologi machine learning, diharapkan sistem ini mampu menjadi solusi dalam mendukung pengambilan keputusan yang lebih tepat terkait peningkatan prestasi akademik siswa di sekolah.

Data *mining*, menurut para ahli, adalah analisis data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui sebelumnya dan dua teknik baru untuk meringkas data agar mudah dipahami dan bermanfaat untuk pemilih data [5].

Random Forest merupakan Salah satu teknik klasifikasi dan regresi adalah kebun acak. Metode ini menggabungkan berbagai metode pembelajaran yang dibangun dan dikombinasikan dengan pohon keputusan sebagai basis klasifikasi [6].

Pada Random Forest, entropy digunakan dalam proses pembentukan pohon keputusan (*decision tree*) untuk mengukur ketidakpastian atau kemurnian suatu node. *Entropy* mengukur ketidakaturan atau ketidakpastian dalam suatu kumpulan data berdasarkan distribusi kelas di dalam *node* tersebut[7].

Nilai *entropy* dihitung dengan menggunakan rumus yang ditemukan dalam persamaan 1, sedangkan nilai gain informasi dihitung dengan menggunakan persamaan 2.

$$Entropy(Y) = -\sum_i p(c/Y) \log^2 p(c/Y) \dots \dots \dots (2.1)$$

Information Gain mengukur berapa banyak informasi yang diperoleh dengan membagi data berdasarkan suatu atribut a.

$$Information\ gain(Y,a) = Entropy(Y) -$$

$$\sum_{v \in Values(a)} \frac{|Y_v|}{|Y|} Entropy(Y_v) \dots \dots \dots (2.2)$$

Prediksi adalah proses menggunakan data historis untuk membuat estimasi tentang hasil atau keadaan di masa depan. Langkah-langkah yang terlibat dalam prediksi meliputi pengumpulan data, pra-pemrosesan data, pemilihan model, pelatihan model, validasi model, prediksi, dan evaluasi [8].

Dalam penelitian, analisis prediksi (peramalan) sangat penting untuk membuat penelitian lebih tepat dan terarah. Prediksi dilakukan untuk mengetahui apa yang paling mungkin terjadi di masa depan. metode prediksi dapat dilakukan secara kuantitatif dengan perhitungan matematis atau secara kualitatif dengan pendapat para pakar. analisis

deret waktu (time series) adalah salah satu teknik prediksi kuantitatif [9].

Prestasi adalah hasil dari suatu kegiatan yang telah dikerjakan, diciptakan, baik secara individual maupun kelompok. Prestasi tidak akan pernah dihasilkan tanpa suatu usaha baik berupa pengetahuan maupun berupa pengetahuan maupun berupa keterampilan [10].

SMAN 1 Aceh Barat Daya didirikan secara resmi pada tanggal 1 Januari 1900 berdasarkan Surat Keputusan Pendirian dengan nomor 421/SK/B.III/1964. Seiring dengan perkembangan waktu, sekolah ini telah mengalami berbagai peningkatan dalam hal kualitas pendidikan, sarana dan prasarana, serta sumber daya manusia, khususnya tenaga pendidik[11].

Jumlah peserta didik di SMAN 1 Aceh Barat Daya saat ini mencapai 651 siswa, yang terbagi dalam dua jurusan utama, yaitu Ilmu Pengetahuan Alam (IPA) dan Ilmu Pengetahuan Sosial (IPS). Kedua jurusan ini telah dirancang untuk memenuhi kebutuhan akademik siswa sesuai dengan minat dan bakat mereka, serta memberikan bekal yang kuat untuk melanjutkan pendidikan ke jenjang yang lebih tinggi. Sebagai bentuk pengakuan atas mutu dan kualitas pendidikan yang diberikan, SMAN 1 Aceh Barat Daya telah memperoleh akreditasi A berdasarkan Surat Keputusan Akreditasi dengan nomor 842/BAN-SM/SK/2019 yang dikeluarkan pada tanggal 7 Oktober 2019 oleh Badan Akreditasi Nasional Sekolah/Madrasah (BAN-S/M)[12].

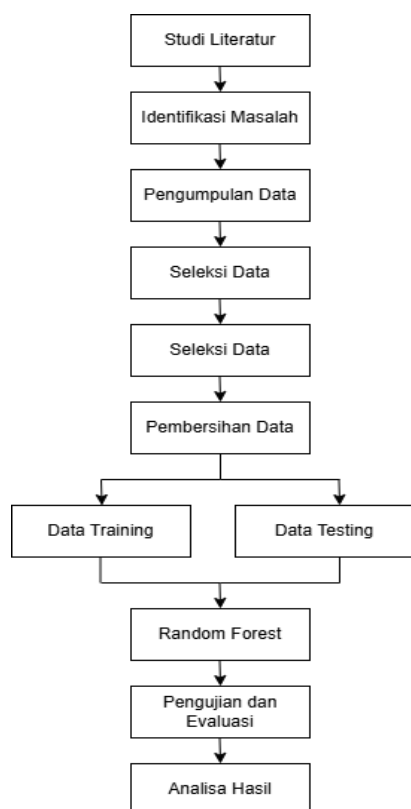
Dalam hal kepemimpinan, saat ini SMAN 1 Aceh Barat Daya dipimpin oleh seorang Kepala Sekolah yang berdedikasi tinggi, yaitu Ibu Rina S., yang secara aktif memimpin dan mengarahkan seluruh kegiatan sekolah agar senantiasa berjalan dengan baik dan sesuai dengan visi dan misi pendidikan nasional. Selain itu, peran penting operator sekolah juga tidak dapat diabaikan, di mana Bapak Hendra Wahyuni selaku operator sekolah bertanggung jawab terhadap pengelolaan data dan administrasi sekolah secara sistematis dan terintegrasi [13].

Sekolah ini secara resmi didirikan pada tanggal 22 Desember 1986 melalui Surat Keputusan Menteri Pendidikan dan Kebudayaan Republik Indonesia dengan Nomor SK 008 F/0/1986. Dalam hal kepemimpinan, SMA Negeri 3 Aceh Barat Daya saat ini dipimpin oleh Ibu Rita Ennjar, oleh Ibu Eralisa, yang bertanggung jawab terhadap pengelolaan sistem data sekolah[14].

Saat ini jumlah murid mencapai 300 siswa. Secara akademik, SMA Negeri 3 Aceh Barat Daya menyelenggarakan program pendidikan dengan dua jurusan utama, yaitu Ilmu Pengetahuan Alam (IPA) dan Ilmu Pengetahuan Sosial (IPS). Selain kegiatan belajar mengajar di dalam kelas, SMA Negeri 3 Aceh Barat Daya juga aktif mengembangkan berbagai program ekstrakurikuler, seperti Pramuka, Paskibra, Palang Merah Remaja (PMR), seni tari, olahraga, dan organisasi siswa seperti OSIS [15].

II. METODE

Penelitian ini mengikuti alur yang sistematis untuk memastikan bahwa setiap langkah dilaksanakan dengan baik dan sesuai dengan tujuan yang telah ditetapkan. Berikut adalah alur penelitian yang akan dilakukan:



Gambar 1 Alur Penelitian

Penelitian ini menggunakan metode *random forest* dengan 10 tahap. Studi Literatur: Setelah merencanakan penelitian, yang dilakukan terlebih dahulu studi Literatur, yaitu mengumpulkan beberapa referensi lalu membaca dan mempelajarinya yang berkaitan dengan *Random Forest*, dan lain yang terkait dengan penelitian yang akan dibuat. Identifikasi Masalah: Identifikasi masalah adalah pengenalan masalah. Masalah dalam penelitian ini adalah ingin melihat apakah algoritma *Random Forest* efektif dalam memprediksi prestasi mahasiswa.

Mengumpulkan data: pengumpulan data dalam penelitian ini dilakukan dengan berbagai teknik untuk memastikan bahwa data yang diperoleh lengkap, akurat, dan relevan dengan tujuan penelitian. Berikut adalah teknik-teknik pengumpulan data yang akan digunakan:

Dokumentasi: mengumpulkan data sekunder yang sudah tersedia di sekolah, seperti nilai rapor siswa, data kehadiran, catatan disiplin, siswa. Data ini akan diperoleh dari arsip sekolah dan sistem informasi akademik yang ada di SMAN 1 Aceh Barat Daya dan SMAN 3 Aceh Barat Daya.

Observasi: melakukan observasi langsung di lingkungan sekolah untuk memahami kondisi belajar-mengajar dan

interaksi siswa dengan guru serta teman sebaya. Observasi ini akan membantu dalam mengidentifikasi faktor-faktor lingkungan yang mungkin berpengaruh terhadap prestasi akademik siswa.

Seleksi Data: seleksi data adalah tahap penting dalam penelitian ini yang bertujuan untuk memastikan bahwa data yang digunakan dalam analisis adalah relevan, akurat, dan berkualitas.

Pembersihan Data: Pembersihan data merupakan langkah penting dan fundamental dalam penelitian ini untuk memastikan bahwa data yang digunakan memiliki kualitas tinggi serta dapat diandalkan untuk analisis lebih lanjut. Proses ini melibatkan berbagai tahapan, seperti mengidentifikasi dan menghapus data yang tidak relevan, menangani data yang hilang atau kosong, memperbaiki inkonsistensi dalam format data, dan mengatasi kesalahan input. Selain itu, pembersihan data juga mencakup penghapusan duplikasi, normalisasi data agar sesuai dengan standar tertentu, dan verifikasi akurasi data melalui pengecekan silang dengan sumber terpercaya. Langkah-langkah ini bertujuan untuk meningkatkan kualitas data sehingga analisis yang dilakukan dapat memberikan hasil yang lebih akurat dan mendukung pengambilan keputusan yang lebih baik.

Data Training: Data training adalah bagian dari dataset yang digunakan untuk melatih model prediksi. Model mempelajari pola dan hubungan antara variabel input dan output dari data ini. Melalui proses ini, model menyesuaikan parameter-parameter internalnya untuk meminimalkan kesalahan prediksi. Data training biasanya mencakup sebagian besar dataset, misalnya 70-80%, untuk memastikan model memiliki cukup informasi untuk belajar dengan baik.

Data Testing: Data testing adalah bagian dari dataset yang digunakan untuk mengevaluasi performa model setelah dilatih dengan data training. Data testing biasanya mencakup sisa dataset, misalnya 20-30%. Model yang telah dilatih diuji dengan data.

Random Forest: setelah dilakukan pengelompokan, data tersebut diolah menggunakan Algoritma Random Forest, yaitu diawali dengan membaca data training. Kemudian hitunglah jumlah dan probabilitasnya. Caranya, hitung jumlah kelas atau label terlebih dahulu.

Pengujian dan Evaluasi: dengan data dari SMAN 1 Aceh Barat Daya dan SMAN 3 Aceh Barat Daya, model prediksi prestasi akademik siswa diuji dan dievaluasi dengan membagi data yang telah dibersihkan menjadi data pelatihan dan data pengujian. Data pelatihan digunakan untuk melatih model dan data pengujian digunakan untuk mengujinya. Untuk mengetahui seberapa baik model dapat memprediksi prestasi akademik siswa berdasarkan data yang diberikan, evaluasi kinerja model dilakukan dengan mengukur *akurasi*, *presisi*, *recall*, *skor F1*, dan *matriks confusion*.

Analisa Hasil: selanjutnya, hasil pengujian dan evaluasi dianalisis untuk mengetahui lebih banyak tentang kinerja

model *Random Forest* dalam memprediksi prestasi akademik siswa di SMAN 1 Aceh Barat Daya dan SMAN 3 Aceh Barat Daya. Ini termasuk menemukan faktor yang paling berpengaruh terhadap prediksi, kecenderungan kesalahan umum, dan kemungkinan perbaikan atau peningkatan model.

Data dikumpulkan melalui dokumentasi nilai rapor, dari SMAN 1 dan SMAN 3 Aceh Barat Daya, serta observasi langsung di lingkungan sekolah. Nilai rapor siswa kemudian diklasifikasikan menjadi tiga kategori prestasi: Berprestasi (nilai ≥ 85), Tidak Berprestasi (nilai < 70), yang menjadi label dalam model klasifikasi. Setelah dilakukan seleksi dan pembersihan data, dataset dibagi menjadi data training sebesar 80% dan data testing sebesar 20%.

Algoritma *Random Forest* diterapkan dengan parameter inti, seperti jumlah pohon keputusan ($n_estimators$) sebanyak 300, kedalaman maksimum pohon (max_depth) sebesar 30, serta kriteria pemisahan menggunakan Gini Index, untuk membangun model yang optimal. Proses pelatihan dilakukan dengan menyesuaikan parameter agar performa model meningkat, diikuti evaluasi menggunakan metrik seperti akurasi, presisi, *recall*, *F1-score*, dan *confusion matrix*.

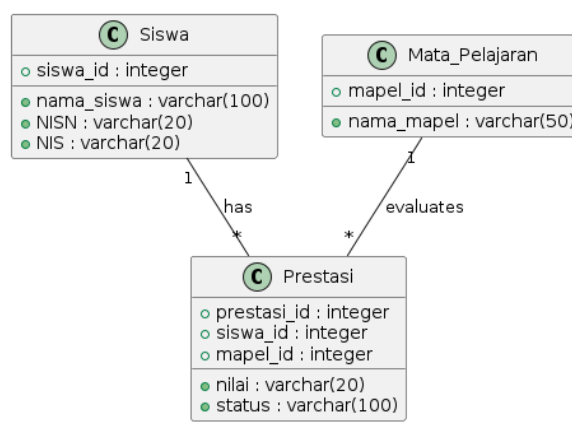
Penelitian ini memiliki beberapa keunikan dan kontribusi baru dibandingkan penelitian-penelitian sebelumnya yang juga menggunakan algoritma *Random Forest*. Pertama, dari segi data, penelitian ini menggunakan dataset nyata dari dua sekolah berbeda, yakni SMAN 1 dan SMAN 3 Aceh Barat Daya, yang terdiri dari 632 data nilai rapor siswa semester genap tahun ajaran 2023/2024. Selain nilai akademik, data juga mencakup variabel seperti kehadiran dan catatan disiplin, yang memungkinkan model memperoleh wawasan yang lebih komprehensif mengenai faktor-faktor yang memengaruhi prestasi siswa.

Dari sisi metodologi, penelitian ini tidak hanya menerapkan algoritma *Random Forest*, tetapi juga melakukan proses analisis mendalam terhadap pembentukan aturan klasifikasi (sebanyak 108 *rule*) serta visualisasi pohon keputusan untuk menginterpretasi hasil model. Pendekatan ini memungkinkan identifikasi faktor-faktor penting dalam menentukan prestasi, seperti kontribusi nilai pada mata pelajaran tertentu terhadap klasifikasi akhir. kontribusi penting lainnya adalah pada aspek pengembangan aplikasi web interaktif berbasis Python dan Flask, yang tidak hanya mampu memproses data dan melakukan prediksi, tetapi juga menyediakan antarmuka visual dan fitur evaluasi model secara real-time. Aplikasi ini dirancang agar dapat digunakan oleh pihak sekolah, terutama guru dan staf akademik, dengan mudah tanpa memerlukan pengetahuan teknis yang mendalam. Fitur seperti unggah data, evaluasi akurasi, visualisasi hasil prediksi, serta ekstraksi aturan keputusan membuat sistem ini lebih aplikatif dan siap digunakan dalam konteks dunia nyata.

III. HASIL DAN PEMBAHASAN

Sistem prediksi prestasi akademik ini dimulai dengan pengguna mengunggah dataset yang berisi variabel akademik siswa, kemudian sistem menampilkan pratinjau untuk memastikan format data yang benar. Data tersebut digunakan untuk melatih model *Random Forest*, yang selanjutnya dievaluasi menggunakan metrik seperti akurasi dan precision. Setelah model terlatih, sistem mengekstraksi aturan untuk memahami pola yang mempengaruhi prestasi siswa serta menyediakan visualisasi data guna mendukung analisis. Akhirnya, model yang telah dikembangkan dapat digunakan untuk memprediksi prestasi akademik siswa berdasarkan data baru, sehingga mendukung pengambilan keputusan berbasis bukti.

Class Diagram



Gambar 2 Class Diagram

Analisis Basis Data

1. Tabel Data Siswa

TABEL I
TABEL DATA SISWA

Nama Kolom	Tipe Data	Length	Keterangan
siswa_id	INTEGER	-	Primary key, ID unik siswa
nama_siswa	VARCHAR	100	Nama lengkap siswa
NISN	VARCHAR	20	Nomor Induk Siswa Nasional
NIS	VARCHAR	20	Nomor Induk Siswa

2. Tabel Mata Pelajaran

TABEL II
TABEL DATA PELAJARAN

Nama Kolom	Tipe Data	Length	Keterangan
mapel_id	INTEGER	-	Primary key, ID unik mata pelajaran
nama_mapel	VARCHAR	50	Nama mata pelajaran

3. Tabel Prestasi

TABEL III
TABEL PRESTASI SISWA

Nama Kolom	Tipe Data	Length	Keterangan
prestasi_id	INTEGER	-	Primary key, ID unik prestasi
siswa_id	INTEGER	-	Foreign key dari

			Tabel Siswa
mapel_id	INTEGER	-	Foreign key dari Tabel Mata Pelajaran
Nilai	VARCHAR	20	Nilai yang dicapai
Status	VARCHAR	100	Status prestasi (e.g., Berprestasi, Tidak Berprestasi)

Berikut adalah tabel hasil dari perhitungan entropy awal :

TABEL IV
HASIL PERHITUNGAN AWAL ENTROPY

Status Prestasi	Frekuensi	p_i	$-p_i \log_2(p_i)$
Berprestasi	11	0.275	0.516
Tidak Berprestasi	13	0.325	0.523
Total	40	1.0	1.568

Proses Pelatihan Model

Dengan memproses pembentukan rule dari data data diatas untuk mendapatkan pohon Keputusan, dimana berikut adalah hasil pembentukan pohon keputusannya.sebagai berikut dimana menghasilkan 108 rule.

Aturan #1:

```
[Root: Status Prestasi]
├── Split on Eko_Baik
│   ├── Eko_Baik = FALSE
│   │   ├── Split on PRA_Cukup
│   │   │   ├── PRA_Cukup = FALSE
│   │   │   │   ├── Split on Eko_Sangat Buruk
│   │   │   │   │   ├── Eko_Sangat Buruk = FALSE
│   │   │   │   │   │   ├── Split on SEI_Cukup
│   │   │   │   │   │   │   ├── SEI_Cukup = FALSE
│   │   │   │   │   │   │   │   ├── Split on PKn_Sangat Buruk
│   │   │   │   │   │   │   │   │   ├── PKn_Sangat Buruk = FALSE
│   │   │   │   │   │   │   │   │   ├── Split on SEB_Buruk
│   │   │   │   │   │   │   │   │   │   ├── SEB_Buruk = FALSE
│   │   │   │   │   │   │   │   │   │   └── [Status Prestasi: Status Prestasi = Berprestasi]
│   │   │   │   │   │   │   └── [Status Prestasi: Status Prestasi = Berprestasi]
│   │   │   │   │   │   └── [Status Prestasi: Status Prestasi = Berprestasi]
│   │   │   └── [Status Prestasi: Status Prestasi = Berprestasi]
│   │   └── [Status Prestasi: Status Prestasi = Berprestasi]
│   └── [Status Prestasi: Status Prestasi = Berprestasi]
```

Aturan #2:

```
[Root: Status Prestasi]
├── Split on Eko_Baik
│   ├── Eko_Baik = FALSE
│   │   ├── Split on PRA_Cukup
│   │   │   ├── PRA_Cukup = FALSE
│   │   │   │   ├── Split on Eko_Sangat Buruk
│   │   │   │   │   ├── Eko_Sangat Buruk = FALSE
│   │   │   │   │   │   ├── Split on SEI_Cukup
│   │   │   │   │   │   │   ├── SEI_Cukup = FALSE
│   │   │   │   │   │   │   │   ├── Split on PKn_Sangat Buruk
│   │   │   │   │   │   │   │   │   ├── PKn_Sangat Buruk = FALSE
│   │   │   │   │   │   │   │   │   ├── Split on SEB_Buruk
│   │   │   │   │   │   │   │   │   │   ├── SEB_Buruk = TRUE
│   │   │   │   │   │   │   │   │   │   └── [Status Prestasi: Status Prestasi]
│   │   │   │   │   │   │   └── [Status Prestasi: Status Prestasi]
│   │   │   │   │   │   └── [Status Prestasi: Status Prestasi]
│   │   │   └── [Status Prestasi: Status Prestasi]
│   │   └── [Status Prestasi: Status Prestasi]
│   └── [Status Prestasi: Status Prestasi]
```

Aturan #3:

```
[Root: Status Prestasi]
├── Split on Eko_Baik
│   ├── Eko_Baik = FALSE
│   │   ├── Split on PRA_Cukup
│   │   │   ├── PRA_Cukup = FALSE
│   │   │   │   ├── Split on Eko_Sangat Buruk
│   │   │   │   │   ├── Eko_Sangat Buruk = FALSE
│   │   │   │   │   │   ├── Split on SEI_Cukup
│   │   │   │   │   │   │   ├── SEI_Cukup = FALSE
│   │   │   │   │   │   │   │   ├── Split on PKn_Sangat Buruk
│   │   │   │   │   │   │   │   │   ├── PKn_Sangat Buruk = TRUE
│   │   │   │   │   │   │   │   │   └── [Status Prestasi: Status Prestasi = Tidak Berprestasi]
│   │   │   │   │   │   │   └── [Status Prestasi: Status Prestasi = Tidak Berprestasi]
│   │   │   │   │   │   └── [Status Prestasi: Status Prestasi = Tidak Berprestasi]
│   │   │   └── [Status Prestasi: Status Prestasi = Tidak Berprestasi]
│   │   └── [Status Prestasi: Status Prestasi = Tidak Berprestasi]
│   └── [Status Prestasi: Status Prestasi = Tidak Berprestasi]
```

TABEL V
HASIL PERCOCOKAN DATA UJI

No	Nama Siswa	Hasil Prediksi
1	Nauwara Annjalyatil Huda	Berprestasi
2	Melda Kusdian Sari	Berprestasi
3	M. Syah Alam	Tidak Berprestasi
4	Fitra Deka Nasywa	Berprestasi
5	T. Sultan Pratama	Berprestasi
6	Ridwan	Berprestasi
7	Ima Safira	Berprestasi
8	Bella Al Taqim	Berprestasi
9	Yusna Dinata Asyura	Berprestasi
10	Ida Sumela	Berprestasi
11	Muhammad Fahri Al Fakran	Berprestasi
12	Alhamsun Asri	Berprestasi
13	Ardian Fahrevi	Berprestasi
14	Davin Nabili	Berprestasi
15	Muhammad Ferdi	Tidak Berprestasi
16	Balinda Wati	Berprestasi
17	Adelia Safira	Berprestasi
18	Mizan Marfino	Berprestasi
19	Mutia Arista	Berprestasi
20	Nazilatul Qamara	Berprestasi
21	Farhan Alfata	Berprestasi
22	Muhammad Razi	Berprestasi
23	Andriyan	Berprestasi
24	T. Abi	Tidak Berprestasi
25	Syakila Amanda	Berprestasi
26	Dinda Yulia Citra	Berprestasi
27	Rafha Alfarazi	Berprestasi
...
...
96	Muhammad Nabawi	Berprestasi
97	Nuri Rahadatul Aisyi	Berprestasi
98	Muhammad Afi Ferdiansyah	Berprestasi
99	Mujibul Rahman	Berprestasi
100	Muhammad Haikal	Berprestasi
101	Nia Yanda Agustia	Berprestasi
102	Mellisa Nasti	Berprestasi
103	Revan Maulana	Berprestasi
104	Rizky Rinaldi	Berprestasi
105	Marwita	Berprestasi
106	Chika Aulia	Berprestasi
107	Suriyadi Saputra	Berprestasi
108	Dinda Malika Prasetya	Berprestasi
109	Zaharatul Munira	Berprestasi
110	M.Miftahul Rianis	Berprestasi
111	Aswatul Pajri	Berprestasi
112	M. Samarkandi	Berprestasi
113	Azizah Khairun Niswah	Berprestasi
114	Hafizah Ulfa	Berprestasi
115	Nelva Patria	Berprestasi
116	Dwi Melfha Azzhara	Berprestasi
117	Taufiq Hidayad	Tidak Berprestasi
118	Novi Muliana	Berprestasi
119	Andrisaputra	Tidak Berprestasi
120	M. Nafis Alfarisi	Berprestasi
121	Muhammad Altat	Berprestasi
122	Putri Fauzi Ani	Berprestasi
123	Mujibur Rahman Al-Bustami	Berprestasi
124	M.Alfandi	Berprestasi
125	Kevin Sujana	Berprestasi
126	Imam Ahmad Dermawan	Tidak Berprestasi
127	Nelda Yuspika	Berprestasi

TABEL VI
TABEL KECOCOKAN

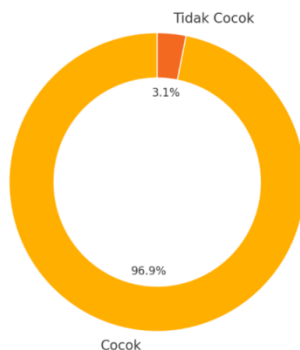
Kategori	Frekuensi	Persentase (%)
Cocok	124	96.88
Tidak Cocok	4	3.13

TABEL VII
TABEL FREKUENSI

Hasil Prediksi	Status Prestasi	Frekuensi
Berprestasi	Berprestasi	108
Berprestasi	Cukup Berprestasi	1
Berprestasi	Tidak Berprestasi	1
Tidak Berprestasi	Cukup Berprestasi	1
Tidak Berprestasi	Tidak Berprestasi	8

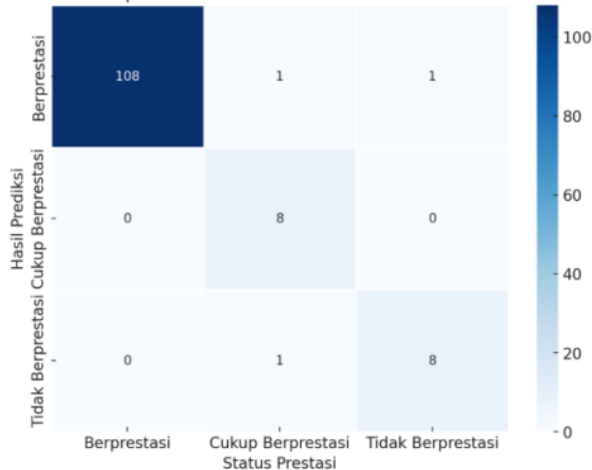
Dimana dapat di visualisasikan sebagai berikut

Grafik Donut Kecocokan Prediksi



Gambar 3 Tabel grafik hasil percocokan

Heatmap Frekuensi Prediksi vs Status Prestasi



Gambar 4 Percocokan berdasarkan status

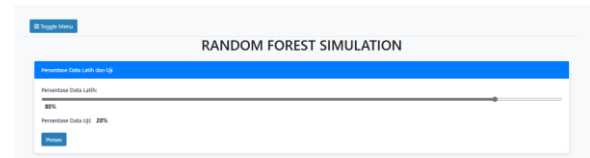
TABEL VIII
IMPLEMENTASI DALAM BENTUK WEB

Fitur	Skenario Pengujian	Hasil	Validitas
Upload Data (/upload)	Mengunggah file Excel dengan format yang benar	File berhasil diunggah	✓ Valid
Ambil Data Latih (/data_latih)	Memeriksa data latih yang telah diunggah	Data berhasil diambil	✓ Valid
Latih Model (/training)	Melatih model dengan data yang diunggah	Model berhasil dilatih dan disimpan	✓ Valid
Evaluasi Model (/evaluating)	Mengevaluasi model dengan data uji	Evaluasi akurasi dan metrik berhasil	✓ Valid
Prediksi Data Uji (/predict)	Melakukan prediksi menggunakan aturan yang diekstrak	Prediksi berhasil dilakukan	✓ Valid
Prediksi Web (/web/predict)	Menginput data secara manual di halaman web	Hasil prediksi ditampilkan di web	✓ Valid
Ekstraksi Aturan (/rules)	Mengambil aturan keputusan dari model	Aturan berhasil diambil dalam JSON	✓ Valid
Visualisasi Pohon (/visualizations)	Menampilkan gambar pohon keputusan	Gambar pohon berhasil dimuat	✓ Valid
Prediksi Data Uji Web (/web/predict_data_uji)	Menampilkan grafik hasil prediksi untuk data uji	Grafik batang berhasil ditampilkan	✓ Valid
Ambil Semua Data (/get_all_data)	Mengambil seluruh data yang telah diunggah	Data berhasil diambil dalam JSON	✓ Valid

B. Hasil Implementasi

Berikut adalah hasil implementasi aplikasi web yang sudah dikembangkan dalam proses yang Sudah dijabarkan sebelumnya dimana nantinya akan menggunakannya untuk memprediksi prestasi.

1. Dashboard Simulasi Random Forest



Gambar 5 Persentase data latih dan uji

[illegible]

Gambar 6 Dataset

Data List														
ID	NO	NAME	STATUS	AGE	SEX	DOB	DOB	DOB	DOB	DOB	DOB	DOB	DOB	DOB
7	509	6547347	Angela R SANTANA	Bank	Bank	Catup	Bank	Catup	Bank	Catup	Bank	Bank	Catup	Bank
108	5233	30759438	COI LYNN	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank
118	1348	9160749	DELLA SANTANA ROSSA	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank
494	13147	6329428	IRNA ALBERTA MURRAY	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank
473	5178	6522228	PURITA GONZALEZ DIAZ	Bank	Bank	Bank	Bank	Catup	Catup	Catup	Bank	Catup	Catup	Cat
388	5321	7987234	YVES DIA PARRERA	Bank	Catup	Bank	Catup	Catup	Catup	Catup	Bank	Bank	Bank	Bank
210	5445	8966328	ISABEL	Catup	Catup	Catup	Bank	Catup	Catup	Bank	Bank	Catup	Bank	Bank
445	13863	6780134	RUAL PINEDAVERDIA	Singtel	Singtel	Bank	Bank	Bank	Bank	Bank	Bank	Singtel	Bank	Bank

Gambar 7 data latih

id	name	status
12794	MONSIEUR HENRI	Expenditure
12823	MR. GUYARD, JEREMY	Expenditure
12857	F. BARRY ALFRED	Expenditure
9122	Berni	Expenditure
12876	BRUNO HENRI MONTELLI	Expenditure
12774	DR. HENRIETTA PIERRE	Expenditure
12739	FR. ADAM JEAN	Expenditure
9108	ADRIANA	Expenditure
12876	BRUNO HENRI	Expenditure
12759	FRANCOIS	Expenditure
12772	HENRIETTE	Expenditure
12876	ADRIANA HENRI	Expenditure

Gambar 8 data uji (hasil prediksi)




Gambar 9 Grafik Distribusi data latih dan data uji

model_performance_metrics		Data UJI	
Model	Value	Model	Value
Accuracy	1.00	Accuracy	0.99
Precision	1.00	Precision	0.99
Recall	1.00	Recall	0.99
F1-Score	1.00	F1-Score	0.99

Gambar 10 Hasil evaluasi model

[illegible]

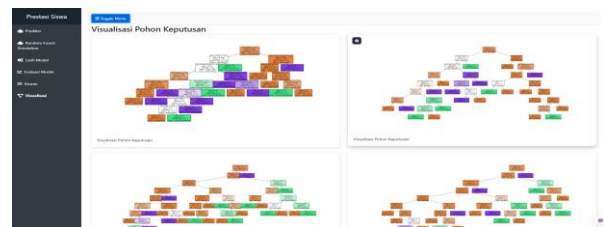
Gambar 11 Aturan yang dihasilkan

 Toggle Menu

Evaluasi Model

Laporan Klasifikasi

Kelas	Precision	Recall	F1-Score	Support
0	0.98	1.00	0.99	551
1	1.00	0.83	0.91	42
2	0.97	0.92	0.95	39
Accuracy	0.98			



Gambar 12 Halaman hasil evaluasi

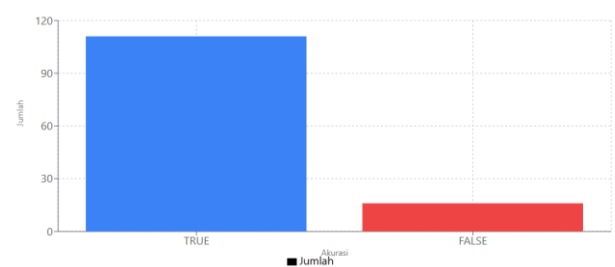
Langkah ini menganalisis akurasi model prediksi performa siswa berdasarkan data 136 siswa. Model memprediksi apakah siswa "Berprestasi" atau "Tidak Berprestasi". Analisis mencakup tabel frekuensi akurasi prediksi, matriks konfusi, dan evaluasi performa model. Hasil menunjukkan tingkat akurasi tinggi, tetapi terdapat beberapa kesalahan prediksi yang dapat dianalisis lebih lanjut.

TABEL IX
FREKUENSI AKURASI

Akurasi	Jumlah
TRUE	111
FALSE	16

TABEL IX
MATRIK KONFUSI

	Prediksi: Berprestasi	Prediksi: Tidak Berprestasi
Asli: Berprestasi	103 (TP)	0 (FN)
Asli: Tidak Berprestasi	16 (FP)	8 (TN)



Gambar 13 Distribusi Akurasi

Berdasarkan matriks konfusi, model memiliki performa sebagai berikut:

1. True Positive (TP): 103 siswa diprediksi "Berprestasi" dan memang berprestasi.
2. True Negative (TN): 8 siswa diprediksi "Tidak Berprestasi" dan memang tidak berprestasi.
3. False Positive (FP): 16 siswa diprediksi "Berprestasi" tetapi sebenarnya tidak berprestasi.

4. False Negative (FN): 0 siswa diprediksi "Tidak Berprestasi" tetapi sebenarnya berprestasi.

Metrik performa model:

1. Akurasi: 87.40% - Menunjukkan persentase prediksi yang benar secara keseluruhan.
2. Presisi: 86.55% - Menunjukkan seberapa banyak prediksi "Berprestasi" yang benar-benar berprestasi.
3. Recall: 100.00% - Menunjukkan seberapa banyak siswa berprestasi yang berhasil diprediksi dengan benar.
4. F1-Score: 92.79% - Menyeimbangkan presisi dan recall untuk evaluasi menyeluruh.

Model sangat akurat dengan akurasi 87.40%, tetapi kesalahan prediksi (FP: 16, FN: 0) menunjukkan bahwa model cenderung lebih sering salah memprediksi siswa "Tidak Berprestasi" sebagai "Berprestasi" (FP) dibandingkan sebaliknya (FN). Ini dapat menjadi fokus perbaikan, misalnya dengan menyesuaikan fitur atau ambang batas klasifikasi model. Analisis menunjukkan bahwa model prediksi memiliki akurasi tinggi (87.40%) dengan 111 prediksi benar dari 136 siswa. Matriks konfusi mengungkapkan bahwa sebagian besar prediksi benar (TP: 103, TN: 8), tetapi terdapat 16 kesalahan (FP: 16, FN: 0). Untuk meningkatkan performa, analisis lebih lanjut terhadap kasus FP dan FN dapat membantu mengidentifikasi pola kesalahan dan memperbaiki model, sehingga lebih akurat dalam konteks pendidikan.

IV. KESIMPULAN

Sistem yang dikembangkan mampu memprediksi prestasi akademik siswa SMAN 1 Aceh Barat Daya dan SMAN 3 Aceh Barat Daya. Model prediksi menunjukkan akurasi tinggi sebesar 87,40%, dengan 111 dari 136 prediksi benar, namun memiliki 16 kesalahan, terutama *false positive* (FP: 16) di mana siswa "Tidak Berprestasi" salah diprediksi sebagai "Berprestasi". Matriks konfusi mengindikasikan performa kuat dengan 103 *true positive* dan 8 *true negative*, tetapi tidak ada *false negative* (FN: 0), menunjukkan kecenderungan model untuk tidak melewatkan siswa berprestasi. Algoritma *Random Forest* terbukti efektif dalam melakukan prediksi prestasi akademik siswa dengan memberikan hasil klasifikasi yang lebih akurat dibandingkan beberapa metode lainnya.

DAFTAR PUSTAKA

- [1] S. Ujud, T. D. Nur, Y. Yusuf, N. Saibi, and M. R. Ramli, "Penerapan Model Pembelajaran Discovery Learning Untuk Meningkatkan Hasil Belajar Siswa Sma Negeri 10 Kota Ternate Kelas X Pada Materi Pencemaran Lingkungan," *J. Bioedukasi*, vol. 6, no. 2, pp. 337–347, 2023, doi: 10.33387/bioedu.v6i2.7305.
- [2] W. Delfiana, "Pemanfaatan Perpustakaan Sekolah Oleh Guru Di SMA Negeri 1 Aceh Barat Daya," 2022.
- [3] F. Mu'Alim and R. Hiday, "Implementasi Metode Random Forest Untuk Penjurusan Siswa Di Madrasah Aliyah Negeri Sintang," *Jupiter*, vol. 14, no. 1, pp. 116–125, 2022. [Online]. Available: <https://www.neliti.com/publications/441871/implementasi-metode-random-forest-untuk-penjurusan-siswa-di-madrasah-aliyah-nege#cite>
- [4] G. A. Sandag, "Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest," *Cogito Smart J.*, vol. 6, no. 2, pp. 167–178, 2020, doi: 10.31154/cogito.v6i2.270.167-178.
- [5] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 690–703, 2021, doi: 10.24176/simet.v11i2.5482.
- [6] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.
- [7] Warjiyono, A. Nur Rais, I. Alfaro, S. Wira Hadi, and W. Kurniawan, "Analisa Prediksi Harga Jual Rumah Menggunakan Algoritma Random Forest Machine Learning," *JURSISTEKNI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 6, no. 2, pp. 416–423, 2024.
- [8] S. Adiguno, Y. Syahra, and M. Yetri, "Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 4, p. 275, 2022, doi: 10.53513/jursi.v1i4.5331.
- [9] W. A. P. Wanto Anjar, "Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan MWanto, A. (2019). Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan Metode Backpropagation. Jurnal & Penelitian Teknik Infor," *J. Penelit. Tek. Inform.*, vol. 2, no. 2, pp. 37–44, 2017.
- [10] A. Matondang, "Pengaruh Antara Minat Dan Motivasi Dengan Prestasi Belajar," *J. Pendidik. Bhs. dan Sastra Indones.*, vol. 2, no. 2, pp. 24–32, 2018, doi: <https://doi.org/10.30743/bahastra.v2i2.1215>.
- [11] A. Maritsa, U. Hanifah Salsabila, M. Wafiq, P. Rahma Anindya, and M. Azhar Ma'shum, "Pengaruh Teknologi Dalam Dunia Pendidikan," *Al-Mutharahah J. Penelit. dan Kaji. Sos. Keagamaan*, vol. 18, no. 2, pp. 91–100, 2021, doi: 10.46781/al-mutharahah.v18i2.303.
- [12] I. Sekolah, "Data Pokok Pendidikan (DAPODIK) SMA NEGERI 1 ACEH BARAT DAYA," pp. 24–26, 2025.
- [13] W. Delfiana, "Pemanfaatan Perpustakaan Sekolah Oleh Guru Di SMA Negeri 1 Aceh Barat Daya," 2022, [Online]. Available: [https://repository.ar-raniry.ac.id/id/eprint/20495/%0Ahttps://repository.ar-raniry.ac.id/id/eprint/20495/1/Winda Delfiana%2C 160503004%2C FAH%2C IP.pdf](https://repository.ar-raniry.ac.id/id/eprint/20495/%0Ahttps://repository.ar-raniry.ac.id/id/eprint/20495/1/Winda%20Delfiana%20160503004%2C%20FAH%2C%20IP.pdf)
- [14] B. Yogyakarta, "ようこそ ! Selamat Datang!," vol. 0022, no. April, p. 2024, 2009.
- [15] Y. Ulfa, "Penerapan model pembelajaran Teams Games Tournament (TGT) terhadap kemampuan pemahaman konsep siswa di SMA Negeri 3 ABDYA," pp. 97–98, 2021.