

**PREDIKSI KELULUSAN SISWA PADA MAN 3 BANDA ACEH
MENGUNAKAN MODEL RANDOM FOREST DAN
XGBOOST**

TUGAS AKHIR

Diajukan Oleh:

KAUTSAR RIDHA

200705056

Mahasiswa Fakultas Sains dan Teknologi

Program Studi Teknologi Informasi



PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI AR-RANIRY

BANDA ACEH

2026/1447

LEMBAR PERSETUJUAN

PREDIKSI KELULUSAN SISWA PADA MAN 3 BANDA ACEH MENGUNAKAN MODEL RANDOM FOREST DAN XGBOOST

TUGAS AKHIR

Diajukan Kepada Fakultas Sains dan Teknologi
Universitas Islam Negeri (UIN) Ar-Raniry Banda Aceh
Sebagai Salah Satu Beban Studi Memperoleh Gelar Sarjana (S1)
dalam Ilmu Teknologi Informasi

Oleh:

KAUTSAR RIDHA

NIM. 200705056

Mahasiswa Fakultas Sains dan Teknologi
Program Studi Teknologi Informasi

Disetujui untuk Dimunqasyahkan oleh:

Pembimbing I,

Pembimbing II,

جامعة الرانيري

AR - RANIRY

Dr. Hendri Ahmadian M.I.M.

NIP. 198301042014031002

Baihaqi, M.T.

NIP. 198802212022031001



Mengetahui,
Ketua Program Studi Teknologi Informasi

Malahayati, M.T

NIP. 198301272015032003

LEMBAR PENGESAHAN

PREDIKSI KELULUSAN SISWA PADA MAN 3 BANDA ACEH MENGUNAKAN MODEL RANDOM FOREST DAN XGBOOST

TUGAS AKHIR

Telah Diuji Oleh Panitia Penguji Seminar Tugas Akhir
Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh dan Dinyatakan Lulus
Serta Diterima Sebagai Salah Satu Beban Studi Program Sarjana (S-1)
Dalam Ilmu Teknologi Informasi

Pada Hari/Tanggal: Jum'at, 30 Januari 2026 M


30 Rajab 1447 H


Di Darussalam, Banda Aceh

Panitia Ujian Munaqasyah Tugas Akhir:

Ketua,


Sekretaris,



Dr. Hendri Ahmadian M.I.M.
NIP. 198301042014031002


Baihaqi, M.T.
NIP. 198802212022031001

Penguji I,

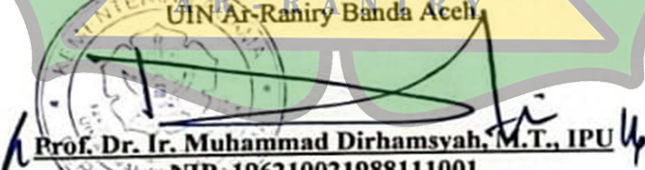
Penguji II,


Malahavati, M.T
NIP. 198301272015032003


Khairan Ar, M.Kom
NIP. 198607042014031001

جامعة الرانيري
Mengetahui:

Dekan Fakultas Sains dan Teknologi
UIN Ar-Raniry Banda Aceh


Prof. Dr. Ir. Muhammad Dirhamsyah, M.T., IPU
NIP. 196210021988111001

LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Kautsar Ridha
Nim : 200705056
Program Studi : Teknologi Informasi
Fakultas : Sains dan Teknologi
Judul : Prediksi Kelulusan Siswa Pada MAN 3 Banda Aceh Menggunakan Model Random Forest dan XGBoost

Dengan ini menyatakan bahwa dalam penulisan tugas akhir ini, saya:

1. Tidak menggunakan ide orang lain tanpa mampu mengembangkan dan mempertanggungjawabkan;
2. Tidak melakukan plagiasi terhadap naskah karya orang lain;
3. Tidak menggunakan karya orang lain tanpa menyebutkan sumber asli atau tanpa izin pemilik karya;
4. Tidak memanipulasi dan memalsukan data;
5. Mengerjakan sendiri karya ini dan mampu bertanggungjawab atas karya ini.

Bila dikemudian hari ada tuntutan dari pihak lain atas karya saya, dan telah melalui pembuktian yang dapat dipertanggungjawabkan dan ternyata memang ditemukan bukti bahwa saya telah melanggar pernyataan ini, maka saya siap dikenai sanksi berdasarkan aturan yang berlaku di Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh.

Demikian pernyataan ini saya buat dengan sesungguhnya dan tanpa paksaan dari pihak manapun.

جامعة الرانيري

A R - R A N I R Y

Banda Aceh,
Yang Menyatakan



Kautsar Ridha

ABSTRAK

Nama : Kautsar Ridha
Nim : 200705056
Program Studi : Teknologi Informasi
Judul : PREDIKSI KELULUSAN SISWA PADA MAN 3 BANDA ACEH MENGGUNAKAN MODEL RANDOM FOREST DAN XGBOOST
Tanggal Sidang : 30 Januari 2026 / 30 Rajab 1447 H
Jumlah Halaman : 78 Halaman
Pembimbing I : Dr. Hendri Ahmadian M.I.M.
Pembimbing II : Baihaqi, M.T.
Kata Kunci : Machine Learning, Prediksi Kelulusan, Random Forest, XGBoost, Streamlit, Sistem Pendukung Keputusan

Penelitian ini bertujuan untuk membangun sistem prediksi kelulusan siswa di MAN 3 Banda Aceh menggunakan algoritma Random Forest dan Extreme Gradient Boosting (XGBoost). Data yang digunakan berupa nilai akademik siswa dari Semester 1 sampai Semester 4 yang kemudian diproses melalui tahapan pra-pemrosesan, pelatihan model, evaluasi, dan implementasi sistem. Hasil pengujian menunjukkan bahwa model XGBoost dan model Hybrid memiliki performa terbaik dengan tingkat akurasi tertinggi sebesar 94,74%, serta nilai precision, recall, dan F1-score yang lebih tinggi dibandingkan Random Forest, sehingga XGBoost dipilih sebagai model utama untuk diimplementasikan ke dalam sistem prediksi. Model tersebut kemudian diintegrasikan ke dalam aplikasi berbasis web menggunakan framework Streamlit dan dideploy pada Streamlit Community Cloud agar dapat diakses secara online oleh pihak sekolah. Sistem yang dibangun mampu menampilkan hasil prediksi kelulusan siswa, ringkasan statistik, serta menyediakan fitur unduh template Excel dan laporan PDF, sehingga dapat digunakan sebagai alat bantu pengambilan keputusan dan sistem peringatan dini bagi sekolah dalam mengidentifikasi siswa yang berpotensi tidak lulus lebih awal.

ABSTRAC

Name : Kautsar Ridha
Nim : 200705056
Department : *Information Technology*
Title : *PREDICTION OF STUDENT GRADUATION AT MAN 3
BANDA ACEH USING THE RANDOM FOREST AND
XGBOOST MODELS*
Date : 30 January 2026 / 30 Rajab 1447 H
Thesis Pages : 78 pages
Supervisor I : Dr. Hendri Ahmadian M.I.M.
Supervisor II : Baihaqi, M.T.
Keywords : Machine Learning, *Graduation Prediction*, Random Forest,
XGBoost, Streamlit, *Decision Support System*

This study aims to build a student graduation prediction system at MAN 3 Banda Aceh using the Random Forest and Extreme Gradient Boosting (XGBoost) algorithms. The data used are student academic grades from Semester 1 to Semester 4 which are then processed through pre-processing, model training, evaluation, and system implementation stages. The test results show that the XGBoost model and the Hybrid model have the best performance with the highest accuracy rate of 94.74%, as well as higher precision, recall, and F1-score values than Random Forest, so that XGBoost is chosen as the main model to be implemented into the prediction system. The model is then integrated into a web-based application using the Streamlit framework and deployed on the Streamlit Community Cloud so that it can be accessed online by the school. The system built is able to display student graduation prediction results, statistical summaries, and provides Excel template download features and PDF reports, so it can be used as a decision-making tool and an early warning system for schools in identifying students who have the potential to fail early.

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Bismillahirrahmanirrahim

Assalamu'alaikum Wr. Wb

Alhamdulillah rabbil'alamin, puja dan puji serta syukur kita ucapkan kehadiran Allah SWT, berkat nikmat dan karunia-Nya yang indah yang masih kita rasakan sampai pada saat ini, nikmat berupa iman, Islam, kesehatan, kesempatan, pengetahuan yang tentunya masih banyak lagi nikmat yang tidak dapat dijabar di atas seluruh kertas ini.

Dalam kesempatan ini penulis bersyukur kepada Allah SWT, berkat Ridho-Nya penulis mampu merampungkan proposal skripsi yang berjudul “Prediksi Kelulusan Siswa Pada MAN 3 Banda Aceh Menggunakan Model Random Forest dan XGBoost” Proposal skripsi ini disusun sebagai kewajiban penulis guna melengkapi tugas dan syarat untuk menyelesaikan pendidikan Strata-I (S1) Program Studi Teknologi Informasi Universitas Islam Negeri Ar-Raniry Banda Aceh, serta memperoleh gelar Sarjana Komputer Universitas Islam Negeri Ar-Raniry Banda Aceh.

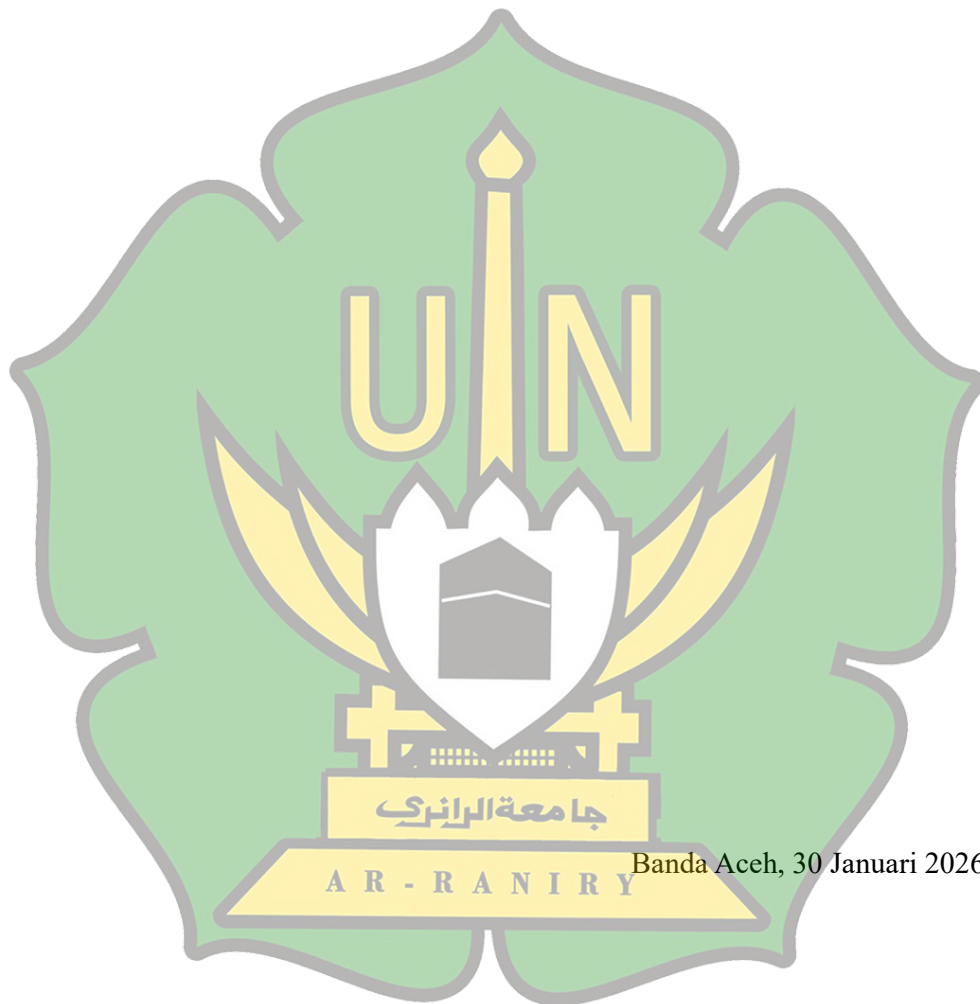
Dalam proses penyusunan proposal ini, penulis telah mendapatkan banyak bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, dengan segala hormat dan rasa Syukur, penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT yang telah memberikan nikmat, rahmat dan kemudahan dalam pengerjaan Tugas Akhir ini.
2. Yang teristimewa, Ayahanda Muhammad Ridha, S.Sos. dan Ibunda Nuraida, S.Pd. serta keluarga besar penulis. Semoga Allah senantiasa melimpahkan rahmat, kesehatan, dan kebahagiaan kepada mereka, serta menjadikan segala amal mereka sebagai jalan menuju ridha dan cinta-Nya.
3. Bapak Dr. Hendri Ahmadian M.I.M. Selaku dosen pembimbing I Tugas Akhir yang telah dengan sabar membimbing, memberikan saran,

kritik, dan masukan berharga selama proses penyusunan tugas akhir ini.

4. Bapak Baihaqi, M.T. selaku pembimbing II Tugas Akhir yang telah banyak meluangkan waktu serta pikiran dalam memberikan bantuan, kritik, dan masukan berharga untuk membimbing penulis demi kesempurnaan Tugas Akhir ini.
5. Ketua dan Sekretaris Program Studi Teknologi Informasi, Ibu Malahayati, M.T dan Bapak Khairan Ar, M.Kom, serta Bapak dan Ibu dosen Program Studi Teknologi Informasi yang telah memberikan ilmu pengetahuan dalam bidang Teknologi Informasi kepada penulis sehingga penulis mampu menyelesaikan skripsi ini.
6. Pembimbing Akademik, bapak Khairan Ar, M.Kom, yang telah membimbing dan memberikan saran selama masa perkuliahan.
7. Staf Prodi Ibu Cut Ida Rahmadiana S,Si. yang telah membantu penulis dalam hal pengurusan administrasi dan surat-surat untuk keperluan penyelesaian skripsi. Dan Seluruh Dosen Teknologi Informasi banyak memberikan ilmu kepada penulis.
8. Kepada para sahabat saya M.Noufal Fazar, Ikhsan Novendra Ramadhan, Johar Nafis, dan teman-teman lainnya yang sering berkumpul di Aromatic Coffe, penulis mengucapkan terima kasih yang sebesar-besarnya atas dukungan dan semangat yang selalu diberikan. Kehadiran kalian tidak hanya menjadi penopang dalam menyelesaikan studi, tetapi juga menjadi bagian berharga dalam perjalanan hidup peneliti. Peneliti sangat bersyukur dan bangga dapat mengenal serta menjalin persahabatan dengan kalian semua.
9. Kepada Hilda Salsabila mahasiswi Pendidikan Matematika, penulis mengucapkan terima kasih yang sebesar-besarnya atas dukungan dan motivasi yang selalu diberikan kepada penulis selama masa perkuliahan.
10. Penulis juga mengucapkan terima kasih yang sebesar-besarnya kepada usaha yang penulis jalankan sejak awal masa perkuliahan hingga sekarang, yang telah berusaha semaksimal mungkin dan pantang menyerah dalam menjalankan perkuliahan.

Peneliti menyadari bahwa Tugas Akhir ini masih jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun selalu peneliti harapkan, demi penyusunan Tugas Akhir yang baik. Peneliti berharap semoga Tugas Akhir ini bermanfaat bagi peneliti, bagi para pembaca maupun peneliti selanjutnya

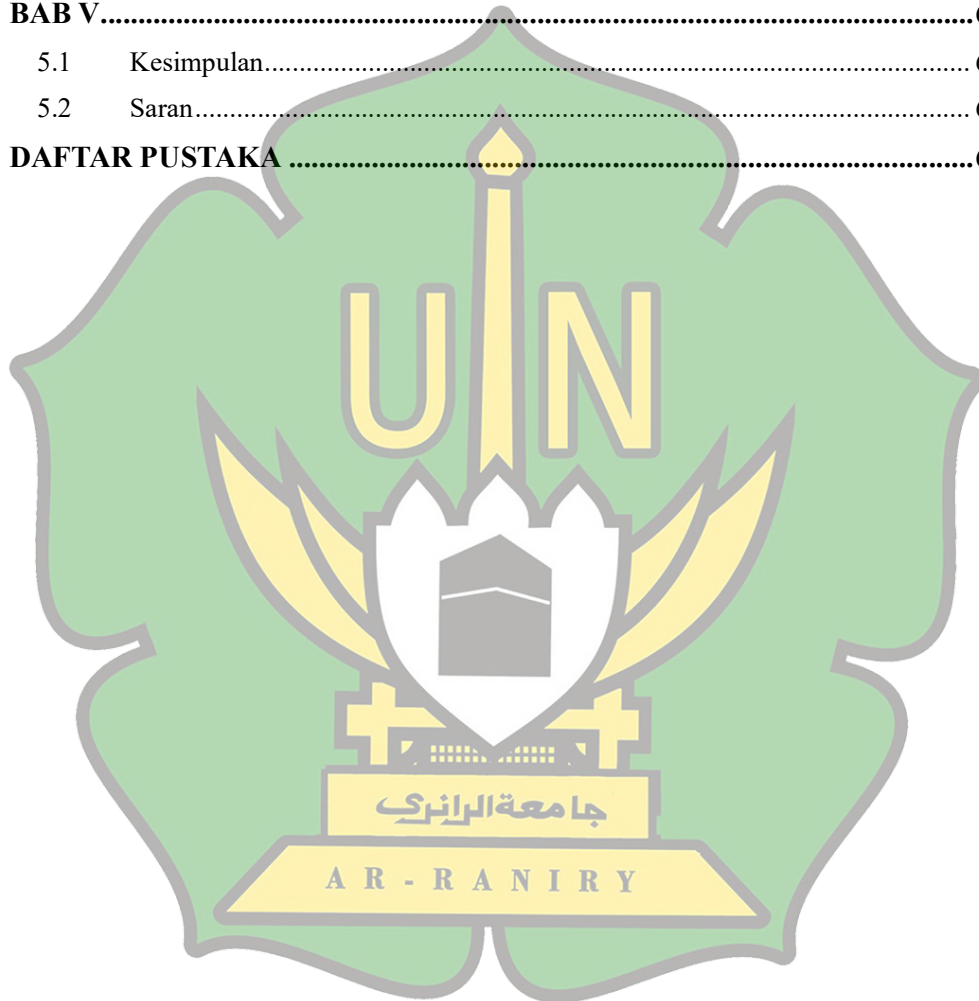


Penulis

DAFTAR ISI

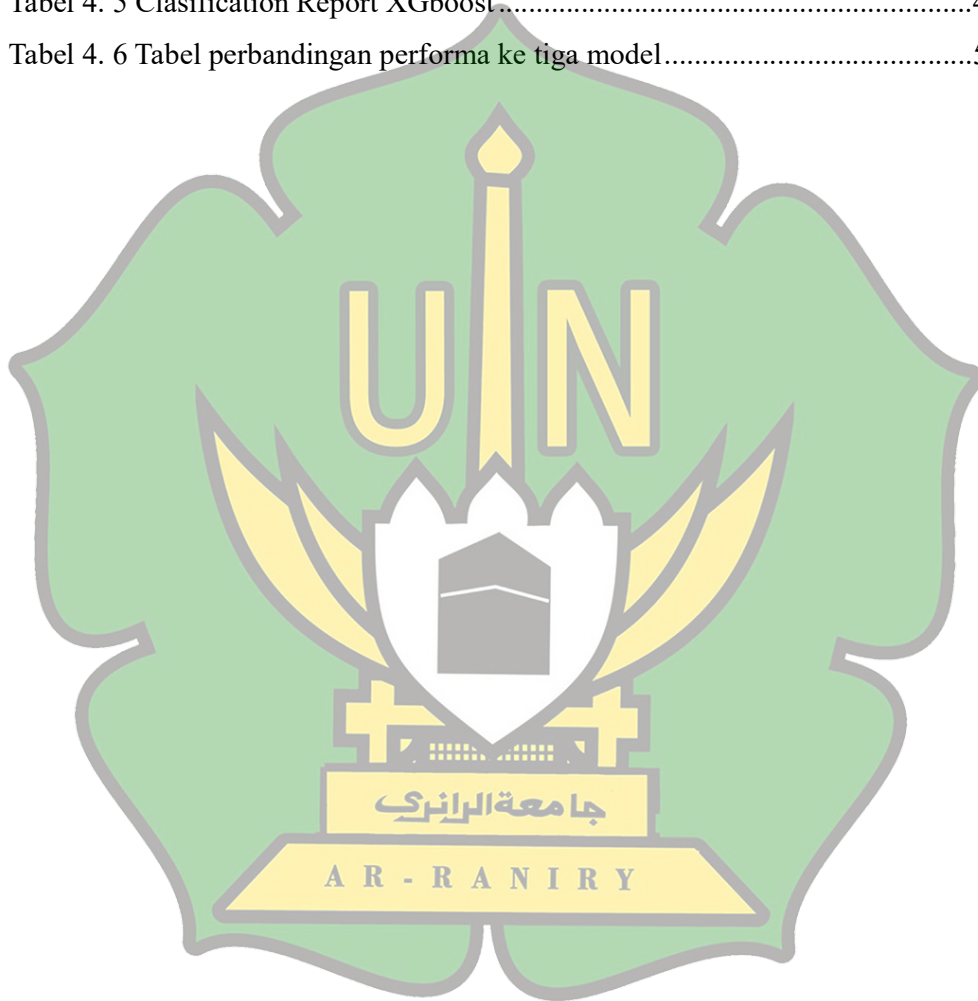
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR	iv
ABSTRAK	v
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiii
BAB I	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Penelitian	5
1.6 Sistematika Penulisan	5
BAB II	7
2.1 Penelitian Terdahulu	7
2.2 Machine Learning dalam Dunia Pendidikan	9
2.3 Random Forest (RF)	10
2.4 Extreme Gradient Boosting (XGBoost)	12
2.5 Model Hybrid dan Ensemble Learning	13
2.6 Evaluasi Model Prediksi	14
2.7 Metode Evaluasi	15
2.8 Streamlit Community Cloud	17
2.9 StreamLit	18
2.10 Python	18
2.11 Google Colaboratory	19
2.12 Kerangka Pemikiran Penelitian / Kerangka teoritis	19
BAB III	21
3.1 Tahapan Penelitian.....	21
3.2 Uji Coba	25
3.3 Alat dan Bahan	25
3.4 Waktu dan Lokasi Penelitian	26
BAB IV	27

4.1	Pengambilan Data	27
4.2	Pra-Processing Data	27
4.3	Implementasi Model Random Forest	31
4.4	Implementasi Model XGBoost.....	36
4.5	Implementasi Model Hybrid (Random Forest + XGBoost).....	42
4.6	Perbandingan Performa Model Random Forest, XGBoost, dan Hybrid (Random Forest + XGBoost	50
4.7	Implementasi Sistem Prediksi Kelulusan Siswa.....	53
4.8	Implementasi Deployment Sistem Prediksi Kelulusan Siswa ke Streamlit Community Cloud.....	57
BAB V	62
5.1	Kesimpulan.....	62
5.2	Saran.....	63
DAFTAR PUSTAKA	65



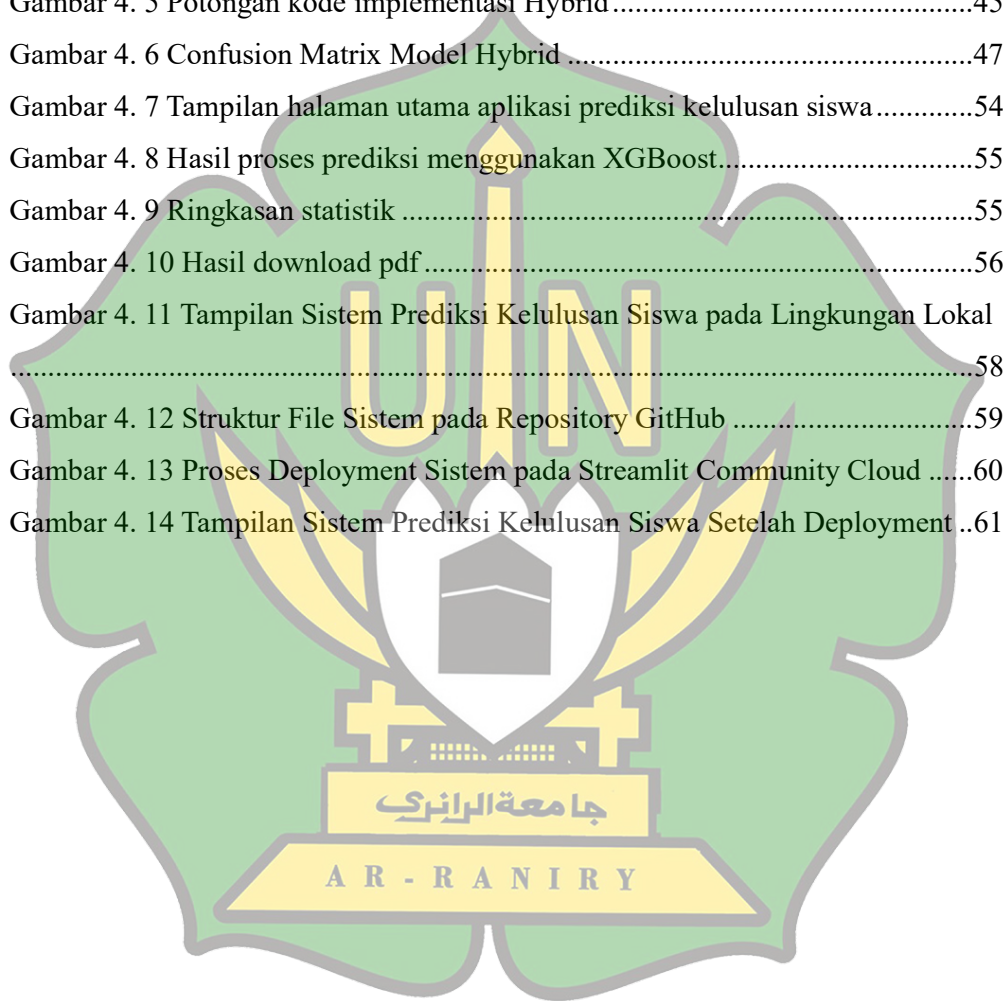
DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu.....	8
Tabel 3. 1 Spesifikasi Perangkat Keras	26
Tabel 3. 2 Spesifikasi Perangkat Lunak	26
Tabel 4. 1 Label Klasifikasi	29
Tabel 4. 2 Contoh data siswa setelah pra-processing.....	30
Tabel 4. 3 Clasification Report Model Random Forest	32
Tabel 4. 4 Clasification Report XGboost	38
Tabel 4. 5 Clasification Report XGboost.....	45
Tabel 4. 6 Tabel perbandingan performa ke tiga model.....	51



DAFTAR GAMBAR

Gambar 1. 1 Wawancara pihak sekolah	3
Gambar 2. 1 Kerangka Teoritis	20
Gambar 3. 1 Tahapan Penelitian	21
Gambar 4. 1 Potongan kode utama proses pelatihan model	32
Gambar 4. 2 Confusion Matrix Random Forest.....	34
Gambar 4. 3 Potongan kode utama proses pelatihan model	38
Gambar 4. 4 Confusion Matrix XGBoost.....	40
Gambar 4. 5 Potongan kode implementasi Hybrid	45
Gambar 4. 6 Confusion Matrix Model Hybrid	47
Gambar 4. 7 Tampilan halaman utama aplikasi prediksi kelulusan siswa.....	54
Gambar 4. 8 Hasil proses prediksi menggunakan XGBoost.....	55
Gambar 4. 9 Ringkasan statistik	55
Gambar 4. 10 Hasil download pdf	56
Gambar 4. 11 Tampilan Sistem Prediksi Kelulusan Siswa pada Lingkungan Lokal	58
Gambar 4. 12 Struktur File Sistem pada Repository GitHub	59
Gambar 4. 13 Proses Deployment Sistem pada Streamlit Community Cloud	60
Gambar 4. 14 Tampilan Sistem Prediksi Kelulusan Siswa Setelah Deployment	61



BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring dengan kemajuan teknologi informasi, analisis data dalam dunia pendidikan telah berkembang dengan pesat. Penerapan data science dan machine learning telah membawa perubahan dalam cara institusi pendidikan mengelola dan menganalisis data akademik. Berbagai algoritma machine learning kini digunakan untuk melakukan prediksi terhadap performa siswa, termasuk dalam hal kelulusan. Di antara algoritma yang populer digunakan adalah Random Forest dan XGBoost. Random Forest dikenal sebagai algoritma berbasis pohon keputusan yang bekerja dengan cara menggabungkan banyak pohon untuk menghasilkan hasil prediksi yang lebih stabil dan akurat (Putra & Erwin Harahap, 2024). Keunggulan utama dari Random Forest adalah kemampuannya dalam menghindari overfitting dan bekerja dengan baik pada dataset yang besar dan kompleks. Sementara itu, XGBoost adalah algoritma yang menggunakan pendekatan boosting untuk meningkatkan akurasi model dengan cara mengoptimalkan kesalahan dari model sebelumnya. Efisiensi waktu pelatihan dan performa yang tinggi membuat XGBoost banyak digunakan dalam berbagai kompetisi data mining dan penerapan industry (Aditya Ananta Wisnu Wardana & Mizwar Rahim).

Namun, meskipun kedua algoritma ini sama-sama unggul dalam aspek tertentu, masing-masing memiliki perbedaan mendasar dalam struktur dan cara kerjanya. Random Forest cenderung lebih sederhana dalam proses tuning parameter dan lebih stabil pada data yang noisy, sedangkan XGBoost menawarkan fleksibilitas lebih dalam hal kontrol parameter dan sering kali menghasilkan akurasi yang lebih tinggi dalam kompetisi nyata. Akan tetapi, performa dari kedua algoritma ini sangat tergantung pada karakteristik data yang digunakan. Beberapa penelitian menunjukkan bahwa pada data pendidikan dengan banyak variabel non-linier dan ketidakseimbangan kelas, XGBoost cenderung unggul. Sebaliknya, dalam konteks lain, Random Forest dapat

memberikan hasil yang lebih baik. Oleh karena itu, perbandingan performa kedua algoritma tersebut pada data kelulusan siswa di lingkungan sekolah tertentu, seperti MAN 3 Banda Aceh, menjadi sangat relevan untuk dilakukan (Naibaho & Zahra, 2023).

Lebih lanjut, dengan kemajuan teknik ensemble learning yang menggabungkan beberapa model dasar (weak learners) untuk menciptakan satu model prediksi yang lebih kuat, akurat, dan stabil, kombinasi dari dua atau lebih model machine learning kini menjadi pendekatan yang semakin populer untuk meningkatkan akurasi prediksi. Teknik seperti stacking, blending, maupun hybrid model memungkinkan hasil prediksi dari beberapa algoritma digabungkan untuk memanfaatkan kekuatan masing-masing model. Dalam konteks prediksi kelulusan siswa, kombinasi antara Random Forest dan XGBoost dapat menjadi strategi yang menjanjikan. Dengan memanfaatkan kekuatan Random Forest dalam stabilitas dan ketangguhannya terhadap noise, serta keunggulan XGBoost dalam akurasi dan efisiensi, diharapkan kombinasi dari keduanya mampu memberikan hasil prediksi yang lebih optimal dibandingkan penggunaan satu model saja. Penelitian ini akan mengeksplorasi potensi kombinasi tersebut sebagai solusi inovatif dalam pengembangan sistem pendukung keputusan di bidang pendidikan. (Agustiningsih dkk.)

Berdasarkan hasil observasi awal di MAN 3 Banda Aceh, jumlah total siswa kelas XII pada tahun ajaran 2024/2025 adalah 199 siswa, yang terdiri dari 104 siswa jurusan MIPA dan 95 siswa jurusan IIS. Penilaian kelulusan siswa masih didasarkan pada nilai ujian yang direkap dalam bentuk legger per kelas, sehingga proses analisis masih dilakukan secara manual oleh guru. Bukti observasi berupa foto wawancara pihak sekolah ditampilkan pada Gambar 1.1.



Gambar 1. 1 Wawancara pihak sekolah

Terakhir, penelitian ini diharapkan tidak hanya memberikan kontribusi akademis dalam bidang data science dan pendidikan, tetapi juga mampu memberikan manfaat langsung bagi sekolah dan peserta didik. Dengan membandingkan performa dua model populer, serta mengeksplorasi kemungkinan kolaboratif di antara keduanya, penelitian ini berupaya menghadirkan pendekatan yang lebih adaptif dan aplikatif dalam mengatasi tantangan pendidikan modern. Selain itu, penelitian ini juga menjadi bagian dari upaya implementasi transformasi digital di lingkungan pendidikan, khususnya di MAN 3 Banda Aceh, sebagai sekolah yang mulai mengadopsi teknologi dalam proses administrasi dan pengambilan keputusan. Diharapkan hasil dari penelitian ini dapat dijadikan referensi untuk pengembangan sistem prediksi kelulusan yang lebih luas di sekolah-sekolah lainnya.

Meskipun pada data aktual seluruh siswa dinyatakan lulus, penelitian ini tetap memfokuskan prediksi pada potensi kelulusan berdasarkan indikator akademik dan kehadiran siswa sebelum kelulusan diumumkan. Dengan demikian, sistem prediksi ini diharapkan dapat menjadi alat bantu peringatan dini bagi sekolah untuk melakukan intervensi akademik, sehingga tingkat kelulusan tetap dapat dipertahankan optimal setiap tahunnya

1.2 Rumusan Masalah

Bedasarkan latar belakang di atas dapat dirumuskan masalah pada penelitian sebagai berikut:

1. Bagaimana perbandingan performa model Random Forest dan XGBoost dalam memprediksi kelulusan siswa di MAN 3 Banda Aceh?
2. apakah Ensemble learning dari kedua model tersebut dapat meningkatkan akurasi prediksi?

1.3 Tujuan Penelitian

Bedasarkan rumusan masalah di atas, maka bisa diambil tujuan dari penelitian ini adalah:

1. Untuk membandingkan performa algoritma Random Forest dan XGBoost dalam memprediksi kelulusan siswa di MAN 3 Banda Aceh, serta mengeksplorasi potensi kombinasi kedua model tersebut guna memperoleh hasil prediksi yang lebih akurat dan optimal.
2. Untuk mengetahui performa dari Ensemble learning kedua model Random Forest dan XGBoost

1.4 Manfaat Penelitian

Dengan adanya penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Penelitian ini memberikan kontribusi bagi perkembangan ilmu pengetahuan, khususnya dalam bidang data science dan pendidikan. Dengan membandingkan dua algoritma populer yaitu Random Forest dan XGBoost, serta mengeksplorasi kemungkinan kombinasi keduanya, penelitian Rini memberikan gambaran yang lebih komprehensif tentang penerapan machine learning dalam konteks prediksi pendidikan. Hasil penelitian ini dapat menjadi referensi bagi mahasiswa, dosen, atau peneliti yang tertarik untuk mengembangkan sistem prediksi berbasis data di dunia pendidikan. Selain itu, penelitian ini juga menambah literatur tentang kelebihan dan kelemahan masing-masing model dalam menangani data klasifikasi, serta bagaimana teknik ensemble dapat diterapkan untuk meningkatkan performa prediktif.

2. Secara praktis, penelitian ini memberikan manfaat langsung bagi MAN 3 Banda Aceh dalam rangka meningkatkan kualitas manajemen pendidikan. Dengan adanya model prediksi kelulusan yang akurat, pihak sekolah dapat lebih mudah mengidentifikasi siswa yang berpotensi tidak lulus sejak dini. Hal ini memungkinkan guru, wali kelas, dan tim bimbingan konseling untuk melakukan tindakan preventif seperti memberikan bimbingan belajar tambahan, pendampingan psikologis, atau melibatkan orang tua dalam proses pembinaan siswa. Dengan strategi yang lebih terarah dan berbasis data, sekolah dapat meningkatkan tingkat kelulusan serta mengurangi jumlah siswa yang tertinggal. Selain itu, hasil penelitian ini juga bisa diadaptasi oleh sekolah lain sebagai bagian dari pengembangan sistem pendukung keputusan (decision support system) yang membantu pengelolaan siswa secara lebih efektif dan efisien.

1.5 Batasan Penelitian

Penelitian ini dibatasi pada penggunaan data siswa MAN 3 Banda Aceh yang mencakup data akademik dan non-akademik yang tersedia, tanpa mempertimbangkan faktor eksternal seperti kondisi psikologis atau lingkungan keluarga. Model yang digunakan hanya terbatas pada algoritma Random Forest, XGBoost, dan kombinasi sederhana dari keduanya.

Selain itu, apabila data aktual menunjukkan seluruh siswa lulus, penelitian ini tetap memprediksi potensi kelulusan siswa berdasarkan indikator akademik dan kehadiran, dengan cara menetapkan label risiko potensi tidak lulus dari nilai ambang batas tertentu, untuk keperluan pelatihan dan pengujian model prediksi.

1.6 Sistematika Penulisan

Untuk dapat mengetahui isi penelitian ini, maka secara singkat akan disusun dalam sistematika penulisan sebagai berikut:

1. BAB I PENDAHULUAN

Bab ini berisi gambaran umum tentang latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat dari sistematika penulisan.

2. BAB II TINJAUAN PUSTAKA

Bab ini memuat materi-materi yang menunjang penelitian ini.

3. BAB III METODE PENELITIAN

Bab ini menjelaskan berbagai metode yang digunakan untuk penelitian ini dapat selesai.

4. BAB IV HASIL DAN PEMBAHASAN

Bab ini akan menjelaskan hasil dari penelitian ini dari awal sampai selesai.

5. BAB V PENUTUP

Bab ini akan menjelaskana Kesimpulan dari penelitian ini dan saran untuk penelitian selanjutnya.



BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian terdahulu yang relevan memiliki peran penting dalam memperluas kerangka penelitian ini. Studi sebelumnya Penelitian oleh Maurino Putra & Erwin Harahap (2024) berjudul *Machine Learning pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Random Forest* menunjukkan bahwa Random Forest memiliki akurasi tinggi dalam memproses data pendidikan dan mampu mengidentifikasi variabel penting yang memengaruhi kelulusan mahasiswa secara efektif.

Sementara itu, Marwa Sulehu, Wisda, First Wanita, dan Markani (2025) dalam penelitiannya yang berjudul "*Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi*" membuktikan bahwa algoritma Random Forest mampu menghasilkan prediksi kelulusan dengan akurasi sebesar 70%, precision hingga 75%, dan recall mencapai 83% pada kelas tertentu. Penelitian ini juga menunjukkan bahwa variabel IPK, motivasi belajar, dan kehadiran merupakan faktor paling berpengaruh dalam menentukan kelulusan mahasiswa, sehingga model ini dianggap cukup efektif sebagai alat bantu untuk mendeteksi mahasiswa yang berisiko tidak lulus lebih awal. Hasil ini memperkuat potensi algoritma Random Forest sebagai pendukung dalam pengambilan keputusan strategis di institusi pendidikan tinggi.

Selanjutnya, Muhammad Labib Mu'tashim, Ati Zaidiah, Bambang Saras Yulistiawan (2023). Yang berjudul *Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta)* Hasil uji coba sampel pada 590 data, algoritma random forest 10 k-fold mendapatkan akurasi 82,64% dan pada gradient boosting 3 k-fold mendapatkan akurasi 79,66%

Selain itu, penelitian oleh Agusti Frananda Alfonsus Naibaho, Amalia Zahra (2023) yang berjudul *Prediksi Kelulusan Siswa Sekolah Menengah Pertama Menggunakan Machine Learning*, hasil dari evaluasi menunjukkan bahwa

algoritma Random Forest memberikan hasil prediksi terbaik dengan akurasi sebesar 99,52% dan nilai ROC-AUC sebesar 99,50%, mengungguli XGBoost yang memperoleh akurasi 99,04% dan Decision Tree dengan akurasi 96,66%.

Abdurrahman Aziz (2024) dalam penelitiannya tentang *Analisis Komparasi Algoritma Machine Learning dalam Prediksi Performa Akademik Mahasiswa: Literature Review*, menyimpulkan bahwa SVM mencatat akurasi tertinggi hingga 94,4% pada dataset dengan margin data yang jelas, sementara Random Forest unggul dalam menangani dataset besar dan kompleks dengan akurasi konsisten sebesar 85%.

Tabel 2. 1 Penelitian Terdahulu

Referensi	Model	Hasil
(Putra dan Erwin Harahap, 2024)	Random Forest	Penelitian ini menggunakan algoritma Random Forest untuk memprediksi kelulusan mahasiswa dengan akurasi mencapai 88%, precision 81%, recall 97%, dan specificity 80%. Data diolah melalui aplikasi web berbasis Streamlit Python.
(Sulehu dkk., 2025)	Random Forest	Random Forest mampu menghasilkan prediksi kelulusan dengan akurasi sebesar 70%, precision hingga 75%, dan recall mencapai 83% pada kelas tertentu.
(Labib Mu'tashim dkk., 2023)	Random Forest	Hasil uji coba sampel pada 590 data, algoritma random forest 10 k-fold mendapatkan akurasi 82,64% dan pada gradient boosting 3 k-fold mendapatkan akurasi 79,66%
(Naibaho dan	Ranndom forest	algoritma Random Forest memberikan

Zahra, 2023)	dan XGBoost	hasil prediksi terbaik dengan akurasi sebesar 99,52% dan nilai ROC-AUC sebesar 99,50%, mengungguli XGBoost yang memperoleh akurasi 99,04% dan Decision Tree dengan akurasi 96,66%.
(Rahman Azis, 2024)	Random Forest	SVM mencatat akurasi tertinggi hingga 94,4% pada dataset dengan margin data yang jelas, sementara Random Forest unggul dalam menangani dataset besar dan kompleks dengan akurasi konsisten sebesar 85%.
(Septiana Rizky et al., 2022)	XGBoost	Rata-rata akurasi XGBoost mencapai 92,57%, lebih tinggi dibandingkan LightGBM yang hanya memperoleh 84,15%. Sementara itu, LightGBM memiliki keunggulan pada aspek spesifisitas dengan rata-rata 80,41%, lebih baik daripada XGBoost yang memperoleh 74,64%.

2.2 Machine Learning dalam Dunia Pendidikan

Machine Learning (ML) merupakan bagian dari kecerdasan buatan *Artificial Intelligence* (AI) yang fokus pada pembuatan algoritma yang memungkinkan komputer belajar dari data dan melakukan prediksi tanpa perlu diprogram secara eksplisit (Sihaloho Dkk.). Dalam konteks pendidikan, ML berperan penting dalam menganalisis data siswa dan memprediksi performa akademik yang kemudian menjadi dasar pengambilan keputusan yang lebih tepat dan terukur (Hery Purnomo dkk.).

Salah satu penerapan paling menonjol dari machine learning di dunia pendidikan adalah dalam memprediksi keberhasilan akademik siswa, termasuk kelulusan. Dengan menggunakan data historis seperti nilai, kehadiran, latar

belakang sosial ekonomi, dan faktor lainnya, machine learning dapat memodelkan kemungkinan siswa lulus tepat waktu (Ababil dkk., 2025).

Seiring meningkatnya jumlah dan kompleksitas data dalam pendidikan, ML menjadi alat bantu penting untuk mengolah informasi tersebut secara efisien. Sistem informasi akademik dan platform e-learning menghasilkan data yang dapat diproses untuk mendeteksi pola dan tren pendidikan siswa secara cepat dan akurat (Muhammad Hermawan dkk.,).

Lebih lanjut, *machine learning* juga mendukung personalisasi pembelajaran. Dengan kemampuan prediksi, sekolah dapat menyediakan intervensi yang tepat sasaran bagi siswa yang mengalami kesulitan belajar, misalnya melalui program bimbingan atau remedial (Asbara dkk., 2024). Ini tidak hanya meningkatkan efektivitas belajar, tetapi juga berpotensi mengurangi angka putus sekolah.

Tak hanya fokus pada performa akademik, machine learning juga dapat digunakan untuk memantau perilaku siswa seperti kecurangan dalam ujian, motivasi belajar, dan partisipasi dalam kegiatan pembelajaran daring. Ini membuktikan bahwa penerapan ML dalam pendidikan bersifat multidimensi menyentuh aspek akademik sekaligus psikososial siswa.

Namun, implementasi ML dalam pendidikan di Indonesia masih menghadapi tantangan, seperti kurangnya infrastruktur teknologi dan rendahnya literasi digital di kalangan pendidik. Oleh karena itu, dibutuhkan pelatihan dan kolaborasi antara pendidik dan ilmuwan data untuk mengembangkan sistem ML yang andal dan layak digunakan di berbagai institusi pendidikan.

Secara keseluruhan, integrasi machine learning dalam pendidikan merupakan langkah strategis menuju sistem pembelajaran yang adaptif, efisien, dan berbasis data. Seiring berkembangnya teknologi, machine learning akan semakin penting dalam mendukung proses belajar-mengajar dan pengambilan kebijakan pendidikan yang lebih cerdas (Naibaho dan Zahra, 2023).

2.3 Random Forest (RF)

Random Forest adalah algoritma machine learning yang termasuk dalam metode *ensemble learning*, yang menggabungkan beberapa pohon keputusan (*decision trees*) untuk meningkatkan akurasi prediksi. Algoritma ini bekerja

dengan membangun sejumlah pohon keputusan pada saat pelatihan dan menghasilkan prediksi berdasarkan agregasi dari hasil masing-masing pohon, seperti suara terbanyak (*majority voting*) untuk klasifikasi atau rata-rata untuk regresi.

Salah satu keunggulan utama Random Forest adalah kemampuannya dalam menangani data dengan jumlah fitur yang besar dan kompleksitas tinggi tanpa mengalami overfitting secara signifikan. Hal ini karena Random Forest menggunakan teknik *bagging* (bootstrap aggregating), yaitu membuat sampel acak dari dataset pelatihan dengan pengembalian dan membangun pohon pada masing-masing sampel tersebut. Selain itu, pemilihan fitur secara acak pada setiap pembentukan node decision tree membantu mengurangi korelasi antar pohon dan meningkatkan keragaman model, yang pada akhirnya meningkatkan akurasi keseluruhan.

Dalam konteks pendidikan, Random Forest telah digunakan untuk memprediksi performa akademik siswa, termasuk dalam memprediksi kelulusan. Misalnya, dalam penelitian yang dilakukan oleh Wildan Abdul Aziz (2022), algoritma Random Forest digunakan untuk klasifikasi sengketa komplain pelanggan perusahaan dan menunjukkan hasil yang baik dengan tingkat akurasi yang tinggi (Wildan Abdul Aziz, 2021).

Selain itu, Random Forest juga digunakan dalam berbagai bidang lain. Azhari Raka Masdian dkk. (2023) menerapkan algoritma ini untuk menganalisis produktivitas padi di Kabupaten Batang, menunjukkan bahwa Random Forest dapat digunakan untuk memodelkan hubungan kompleks antara variabel lingkungan dan hasil pertanian (Raka Masdian dkk., 2023).

Keunggulan lain dari Random Forest adalah kemampuannya dalam mengukur pentingnya setiap fitur dalam proses prediksi. Informasi ini berguna untuk mengidentifikasi variabel-variabel yang paling berpengaruh terhadap hasil prediksi, seperti nilai mata pelajaran, kehadiran, atau latar belakang sosial ekonomi dalam konteks pendidikan. Dengan demikian, institusi pendidikan dapat merancang strategi intervensi yang lebih tepat sasaran bagi siswa yang berisiko tidak lulus tepat waktu.

Namun, Random Forest juga memiliki beberapa keterbatasan. Salah satunya adalah sifatnya yang kurang interpretatif dibandingkan model-model sederhana seperti Decision Tree. Selain itu, model ini memerlukan sumber daya komputasi yang lebih tinggi, terutama ketika digunakan untuk dataset besar dan jumlah pohon yang banyak. Meskipun demikian, dengan dukungan perangkat keras dan perangkat lunak modern, keterbatasan ini dapat diatasi.

Secara keseluruhan, Random Forest merupakan algoritma yang sangat andal untuk prediksi kelulusan siswa karena kestabilannya dalam menangani data kompleks, kemampuannya dalam mengurangi overfitting, serta akurasi yang tinggi. Oleh karena itu, algoritma ini dipilih sebagai salah satu model utama dalam penelitian ini.

2.4 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) adalah algoritma machine learning berbasis *ensemble* yang dikembangkan untuk meningkatkan performa dan efisiensi dari metode *Gradient Boosting*. XGBoost menggabungkan beberapa model pohon keputusan (*decision trees*) secara berurutan, di mana setiap pohon baru dibangun untuk memperbaiki kesalahan dari pohon sebelumnya. Proses ini bertujuan untuk meminimalkan fungsi kerugian (*loss function*) dengan pendekatan gradien, sehingga menghasilkan model yang lebih akurat dan tahan terhadap overfitting.

Salah satu keunggulan utama XGBoost adalah kemampuannya dalam menangani data yang besar dan kompleks dengan efisiensi tinggi. Algoritma ini mendukung paralelisasi dalam proses pelatihan, sehingga mempercepat waktu komputasi. Selain itu, XGBoost memiliki fitur regularisasi yang membantu mengurangi overfitting, serta mampu menangani data yang tidak seimbang dengan baik (Yulianti dkk., 2022).

Dalam konteks pendidikan, XGBoost telah digunakan untuk memprediksi performa akademik siswa. Misalnya, penelitian oleh Arisandi dkk. (2023) menerapkan XGBoost untuk mengklasifikasikan performa pelajar dalam proses pembelajaran. Hasilnya menunjukkan bahwa XGBoost mampu memberikan akurasi tinggi dalam memprediksi performa siswa, yang dapat digunakan sebagai

sistem peringatan dini untuk mengidentifikasi siswa yang membutuhkan intervensi tambahan.

Selain itu, XGBoost juga telah diterapkan dalam berbagai bidang lain di Indonesia. Yulianti dkk. (2022) menggunakan XGBoost untuk klasifikasi nasabah kartu kredit, sementara Kurniawan dkk. (2025) menerapkannya untuk mengklasifikasikan potensi tsunami berdasarkan data gempa bumi. Kedua studi tersebut menunjukkan bahwa XGBoost mampu memberikan hasil prediksi yang akurat dan dapat diandalkan (Andhika Brahmandjati dkk.,2024).

Secara keseluruhan, XGBoost merupakan algoritma yang kuat dan efisien untuk berbagai tugas klasifikasi dan regresi. Keunggulannya dalam menangani data besar, kompleks, dan tidak seimbang menjadikannya pilihan yang tepat untuk diterapkan dalam prediksi kelulusan siswa di MAN 3 Banda Aceh.

2.5 Model Hybrid dan Ensemble Learning

Model Hybrid dan Ensemble Learning adalah pendekatan dalam machine learning yang bertujuan untuk menggabungkan dua atau lebih algoritma pembelajaran guna meningkatkan akurasi, kestabilan, dan generalisasi model prediktif. Pendekatan ini muncul dari pemahaman bahwa satu model tunggal sering kali memiliki kelemahan tertentu, dan dengan menggabungkan model yang berbeda, kelemahan tersebut dapat diminimalkan. Konsep hybrid sangat cocok diterapkan dalam kasus prediksi kelulusan siswa, yang melibatkan berbagai variabel akademik dan non-akademik yang kompleks dan tidak selalu linier sifatnya.

1. Bagging (Bootstrap Aggregating) adalah metode ensemble yang bekerja dengan cara membuat beberapa model dari subset data yang diambil secara acak dan dilatih secara independen. Setiap model memberikan prediksi yang kemudian digabungkan, biasanya melalui proses voting untuk klasifikasi. Algoritma Random Forest merupakan implementasi paling populer dari teknik ini, karena mampu mengurangi variansi model dan tahan terhadap overfitting meskipun data memiliki banyak noise.
2. Boosting adalah teknik yang menyusun model secara berurutan, di mana model baru difokuskan untuk memperbaiki kesalahan dari model sebelumnya. Dalam teknik ini, setiap iterasi akan menambahkan model baru

yang mempelajari kelemahan dari hasil prediksi sebelumnya. XGBoost (Extreme Gradient Boosting) adalah salah satu algoritma boosting paling populer dan efisien karena mampu meminimalkan fungsi kerugian menggunakan teknik gradien dan regularisasi. Algoritma ini sangat efisien dalam menangani data skala besar dan kompleks, serta efektif untuk data yang tidak seimbang, seperti pada kasus klasifikasi kelulusan siswa yang jumlah tidak lulusnya jauh lebih sedikit (Yulianti dkk., 2022).

3. Stacking (Stacked Generalization) adalah teknik ensemble yang menggabungkan hasil dari beberapa model dasar (base learners), kemudian model baru yang disebut meta-learner digunakan untuk mempelajari dan menggabungkan hasil dari model-model sebelumnya. Dalam konteks model hybrid Random Forest dan XGBoost, stacking sangat ideal karena mampu menyatukan kelebihan Random Forest yang kuat dalam kestabilan dan interpretasi dengan keunggulan XGBoost dalam optimasi performa prediktif (Andhika Brahmadjati dkk., 2024).

Beberapa penelitian di Indonesia telah membuktikan bahwa teknik ensemble, khususnya stacking, mampu meningkatkan performa prediksi. Misalnya, studi oleh Arisandi dkk. (2023) menunjukkan bahwa XGBoost secara mandiri sudah sangat akurat dalam memprediksi performa pelajar, sementara penelitian oleh Kurniawan & Wijayanto (2024) menemukan bahwa kombinasi Random Forest dan XGBoost melalui ensemble menghasilkan hasil klasifikasi tenaga kerja yang lebih optimal dibanding penggunaan algoritma tunggal.

Dengan demikian, penggunaan *model hybrid* dalam bentuk stacking antara Random Forest dan XGBoost menjadi sangat relevan untuk digunakan dalam penelitian ini. Kombinasi ini tidak hanya memberikan prediksi yang lebih akurat, tetapi juga dapat meningkatkan pemahaman terhadap faktor-faktor penting yang mempengaruhi kelulusan siswa. Selain itu, pendekatan ini lebih adaptif terhadap berbagai bentuk dan pola data yang muncul dalam konteks pendidikan.

2.6 Evaluasi Model Prediksi

Model hybrid dan *ensemble learning* merupakan pendekatan dalam machine learning yang menggabungkan dua atau lebih algoritma untuk

meningkatkan akurasi dan keandalan prediksi. Tujuannya adalah untuk memanfaatkan kelebihan dari masing-masing algoritma dan mengurangi kelemahannya. Teknik bagging, boosting, dan stacking adalah bentuk umum dari ensemble learning. Bagging bekerja dengan cara melatih beberapa model pada subset data yang berbeda dan menggabungkan hasil prediksinya seperti yang dilakukan algoritma *Random Forest*. Boosting melatih model secara berurutan untuk memperbaiki kesalahan model sebelumnya, contohnya pada XGBoost. Sementara stacking menggabungkan prediksi dari beberapa model awal (base learners), lalu menggunakan model tambahan (*meta-learner*) untuk menghasilkan prediksi akhir yang lebih akurat.

Ketika dua model seperti *Random Forest* dan *XGBoost* digabungkan dalam satu sistem, maka terbentuklah model hybrid. Model ini bertujuan untuk menggabungkan stabilitas dan kecepatan *Random Forest* dengan keakuratan dan efisiensi komputasi dari *XGBoost*. Penelitian oleh Naibaho dan Zahra (2023) menunjukkan bahwa gabungan algoritma machine learning dapat digunakan secara efektif untuk memprediksi kelulusan siswa SMP dengan hasil yang cukup akurat. Selain itu, jurnal lain dari Undiksha (2023) membuktikan bahwa metode ensemble seperti *Random Forest* dan *XGBoost* dapat meningkatkan akurasi prediksi terhadap kinerja akademik siswa berdasarkan data log dari Learning Management System (Ardianti dkk., 2024).

2.7 Metode Evaluasi

Dalam penelitian yang menggunakan algoritma machine learning, khususnya untuk prediksi kelulusan siswa, perlu dilakukan evaluasi terhadap performa model yang digunakan. Beberapa metrik yang umum digunakan dalam evaluasi model klasifikasi adalah Accuracy, Precision, Recall, dan F1-Score. Semua metrik ini dihitung berdasarkan hasil prediksi model yang dikategorikan menjadi True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN).

Definisi TP, FP, TN, FN:

- True Positive (TP): Jumlah data yang sebenarnya positif dan diprediksi positif oleh model.

- False Positive (FP): Jumlah data yang sebenarnya negatif namun diprediksi positif oleh model.
- True Negative (TN): Jumlah data yang sebenarnya negatif dan diprediksi negatif oleh model.
- False Negative (FN): Jumlah data yang sebenarnya positif namun diprediksi negatif oleh model.

Rumus dan Pengertian Metrik Evaluasi

1. Accuracy

Mengukur seberapa banyak prediksi yang benar dibandingkan dengan total prediksi yang dilakukan.

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

2. Precision

Mengukur seberapa tepat model dalam memprediksi kelas positif dari seluruh data yang diprediksi sebagai positif.

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

3. Recall (*Sensitivity*)

Mengukur seberapa baik model dalam menemukan semua data yang benar-benar positif.

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

4. F1-Score

Merupakan rata-rata berbobot dari Precision dan Recall, yang berguna saat data bersifat tidak seimbang.

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

Keterangan:

TP (*True Positive*): Siswa yang diprediksi lulus dan memang benar-benar lulus.

FP (*False Positive*): Siswa yang diprediksi lulus tapi ternyata tidak lulus.

FN (*False Negative*): Siswa yang diprediksi tidak lulus padahal sebenarnya lulus.

TN (*True Negative*): Siswa yang diprediksi tidak lulus dan memang benar tidak lulus.

2.8 Streamlit Community Cloud

Streamlit Community Cloud merupakan layanan hosting berbasis cloud yang disediakan secara resmi oleh Streamlit untuk menjalankan dan mempublikasikan aplikasi web interaktif yang dibangun menggunakan framework Streamlit. Platform ini memungkinkan pengembang untuk melakukan deployment aplikasi secara langsung dari repositori GitHub tanpa memerlukan konfigurasi server yang kompleks, sehingga proses publikasi aplikasi menjadi lebih cepat dan efisien.

Streamlit Community Cloud mendukung berbagai pustaka Python yang umum digunakan dalam pengembangan sistem machine learning dan data science, seperti NumPy, Pandas, Scikit-learn, XGBoost, dan TensorFlow. Dengan dukungan tersebut, platform ini sangat sesuai digunakan untuk implementasi dan demonstrasi sistem prediksi berbasis machine learning dalam bentuk aplikasi web.

Selain itu, Streamlit Community Cloud menyediakan akses publik melalui URL khusus dengan domain `streamlit.app`, sehingga aplikasi dapat digunakan oleh pengguna lain melalui web browser tanpa perlu melakukan instalasi tambahan. Platform ini juga terintegrasi dengan sistem version control GitHub, yang memungkinkan pengelolaan kode secara terstruktur, kolaboratif, dan mudah diperbarui.

Keunggulan utama dari Streamlit Community Cloud terletak pada kemudahan penggunaan, performa yang stabil, serta sifatnya yang gratis untuk proyek bersifat open-source dan akademik. Oleh karena itu, platform ini sangat cocok digunakan sebagai media implementasi sistem prediksi kelulusan siswa berbasis machine learning dalam penelitian ini.

2.9 StreamLit

Streamlit adalah sebuah framework open-source berbasis Python yang dirancang khusus untuk membangun aplikasi web interaktif secara cepat dan sederhana, terutama untuk keperluan analisis data dan pembelajaran mesin (*machine learning*). Streamlit diperkenalkan pertama kali pada tahun 2019 dan dengan cepat mendapatkan popularitas di kalangan ilmuwan data dan pengembang karena pendekatannya yang intuitif dan minim boilerplate code. Dibandingkan framework web tradisional seperti Flask atau Django, Streamlit memungkinkan pengembang membuat aplikasi web hanya dengan menulis skrip Python biasa, tanpa perlu memahami konsep HTML, CSS, atau JavaScript secara mendalam. Dengan hanya beberapa baris kode, pengguna sudah dapat menampilkan grafik, model machine learning, form interaktif, visualisasi data, dan komponen UI lainnya secara real-time.

Keunggulan utama Streamlit terletak pada kemampuannya dalam menyederhanakan proses pembuatan prototipe aplikasi AI atau dashboard data secara cepat. Hal ini sangat berguna dalam konteks penelitian dan eksperimen, di mana peneliti atau pengembang sering kali membutuhkan antarmuka pengguna untuk mendemonstrasikan hasil model atau menganalisis data secara langsung.

2.10 Python

Python adalah bahasa pemrograman yang terus berkembang setiap tahunnya. *Python* diciptakan pada tahun 1991, oleh Guido van Rossum. Bahasa Pemrograman Python adalah pemrograman ABC, karena sifat Bahasa pemrograman Python yang *opensource* yang membuat banyak orang dapat berkontribusi dalam perkembangannya. Python banyak digunakan bukan tanpa alasan seperti mudah dipelajari, serbaguna, sangat populer, banyak tersedia library, banyak digunakan untuk kecerdasan buatan serta banyak keunggulan lainnya (Suryantara, 2024).

Python menyediakan berbagai *library* antara lain:

- *Scikit-Learn*

Scikit-Learn adalah sebuah pustaka (*library*) machine learning yang ditujukan untuk bahasa pemrograman *python*. Pustaka ini dirancang untuk memberikan alat yang sederhana dan efisien dalam analisis data, serta memudahkan pengguna dalam menerapkan berbagai teknik machine

learning. *Scikit-Learn* mencakup beragam algoritma untuk berbagai tugas, termasuk klasifikasi, regresi, dan pengelompokan (*clustering*) (Maesaroh dkk, 2024).

- **XGBoost**

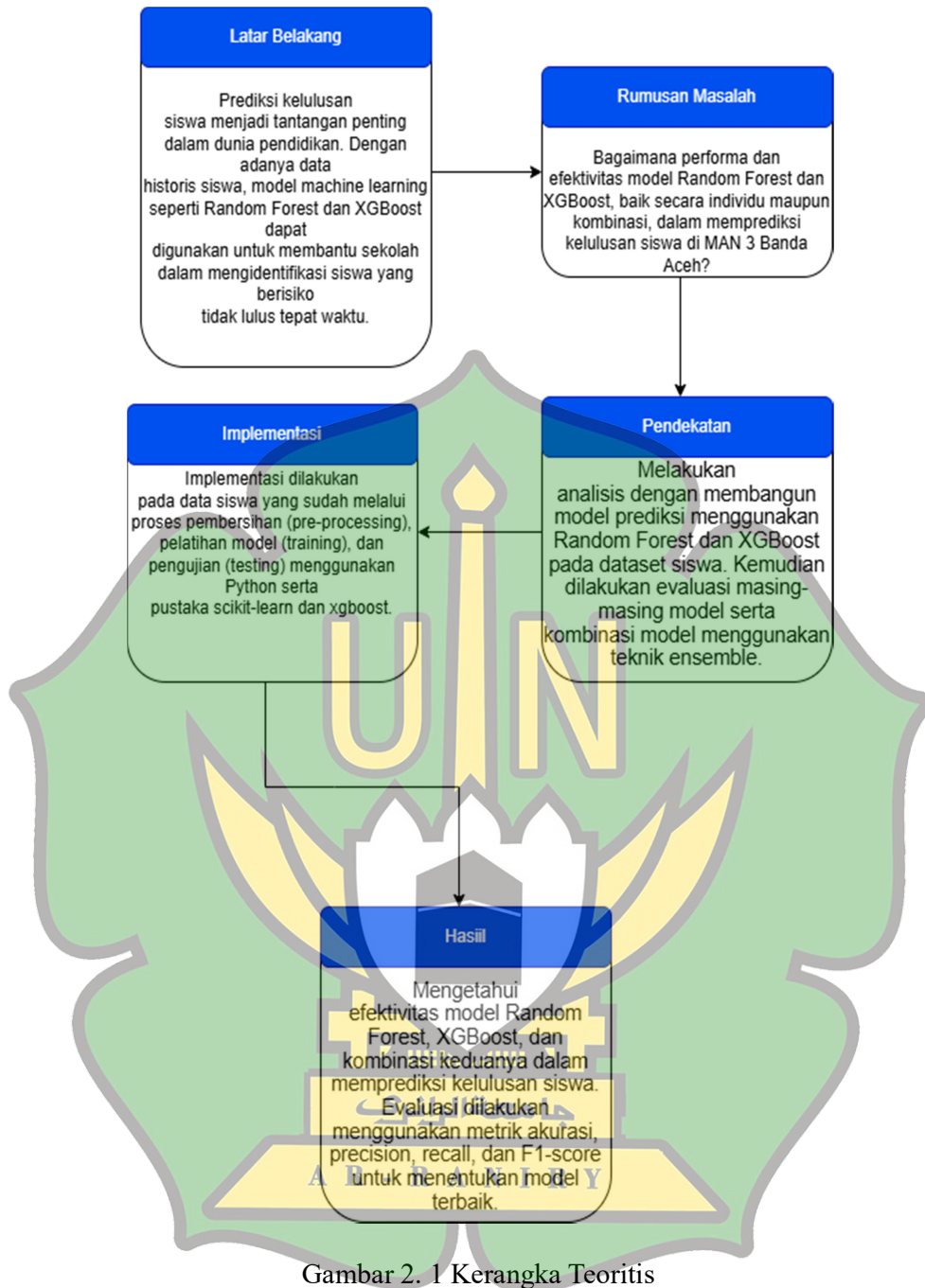
XGBoost (*Extreme Gradient Boosting*) merupakan algoritma machine learning berbasis ensemble yang sangat populer karena kemampuannya dalam meningkatkan akurasi prediksi melalui pendekatan boosting. Algoritma ini bekerja dengan cara membangun model secara bertahap, di mana model berikutnya bertujuan memperbaiki kesalahan dari model sebelumnya. Keunggulan utama XGBoost adalah efisiensinya dalam hal waktu komputasi dan kemampuannya dalam menangani data besar serta kompleks, terutama dalam masalah klasifikasi dan regresi pada data tabular (Muslim.dkk.,2020)

2.11 Google Colaboratory

Google Colab, atau *Google Colaboratory*, adalah platform berbasis cloud yang memungkinkan pengguna untuk menulis, menyimpan, dan berbagi program yang ditulis di *Google Drive*. Mirip dengan *Jupyter Notebook*, *Google Colab* dapat diakses melalui *browser* seperti *Mozilla Firefox* dan *Google Chrome*. Salah satu fitur utamanya adalah kemampuan untuk menjalankan kode *Python* tanpa perlu instalasi atau pengaturan di komputer lokal, karena semua proses dilakukan di cloud (Haana Udtari Anjani dkk., 2024)

2.12 Kerangka Pemikiran Penelitian / Kerangka teoritis

Kerangka Pemikiran menggambarkan struktur paradigma berpikir yang digunakan sebagai landasan oleh penulis untuk memahami, menganalisis dan memecahkan masalah. Kerangka Pemikiran membantu mengorganisir informasi, konsep menjadi satu rangkaian logis.



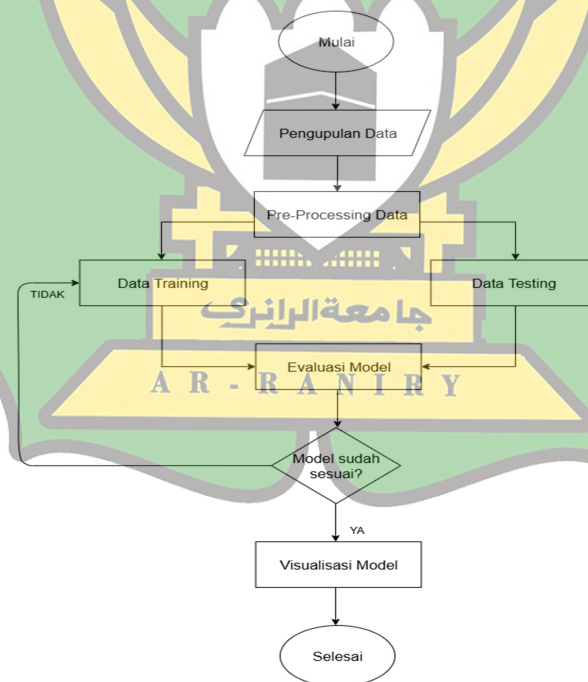
Gambar 2. 1 Kerangka Teoritis

BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Metode penelitian yang digunakan merupakan metode penelitian kuantitatif, yaitu dengan melibatkan pengumpulan data dan analisis data. Penelitian kuantitatif merupakan jenis penelitian yang mengukur suatu fenomena secara objektif dengan menganalisis permasalahan yang terjadi, baik pada peneliti maupun lingkungan sekitarnya (Siyoto dan Ali Sodik, 2015). Data yang dikumpulkan berupa data nilai rapor dari siswa/i MAN 3 Banda Aceh. Penelitian ini menganalisis dan membandingkan performa kedua metode dalam memprediksi kelulusan siswa, menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Adapun tahapan penelitian ini dapat dilihat pada gambar 3.1



Gambar 3. 1 Tahapan Penelitian

3.1.1 Pengumpulan Data

Data dikumpulkan dari MAN 3 Banda Aceh dengan mengacu pada variabel-variabel penting yang berpotensi memengaruhi kelulusan siswa, seperti nilai akademik dan tingkat kehadiran. Data diperoleh melalui sistem informasi sekolah atau hasil rekap manual yang telah tersedia.

a. Sumber Data

Data diperoleh dari MAN 3 Banda Aceh, yang mencakup data akademik siswa, dan kehadiran.

b. Variabel Penelitian

Variabel yang digunakan meliputi variabel input (fitur) seperti nilai rata-rata tiap semester mulai dari semester 1, semester 2, semester 3, dan semester 4, dan variabel output berupa status kelulusan siswa.

c. Teknik Pengumpulan Data

Teknik pengumpulan dilakukan melalui pengunduhan data dari sistem informasi sekolah dan pengumpulan manual dari arsip akademik.

3.1.2 Pra-Processing Data

Tahap pra-processing data dilakukan untuk mempersiapkan data agar dapat digunakan dalam proses pelatihan model *machine learning*. Data yang digunakan dalam penelitian ini terdiri dari dua indikator utama, yaitu nilai rata-rata dari semester 1, semester 2, semester 3, dan semester 4. Beberapa langkah pra-processing yang dilakukan adalah sebagai berikut:

1. Pembersihan Data (*Data Cleaning*)

- a. Memastikan tidak ada data ganda (duplikat).
- b. Menangani data yang kosong (*missing values*), misalnya jika terdapat nilai ujian atau kehadiran yang belum tercatat.

2. Transformasi Data (*Data Transformation*)

- a. Mengubah data ke dalam bentuk numerik agar sesuai dengan kebutuhan algoritma.
- b. Menentukan skala nilai ujian (misalnya 0–100).

3. Seleksi Variabel (*Feature Selection*)

Variabel yang digunakan untuk prediksi, yaitu:

Nilai dari semester 1, semester 2, semester 3, dan semester 4 sebagai indikator kemampuan akademik siswa.

4. Pembagian Data (*Data Splitting*)

- a. Dataset dibagi menjadi data latih (*training set*) dan data uji (*testing set*) dengan perbandingan 80:20.
- b. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengukur performa model.

Dengan pra-processing ini, data menjadi lebih siap untuk digunakan dalam membangun model prediksi menggunakan algoritma *Random Forest*, *XGBoost*, maupun kombinasi keduanya.

3.1.3 Evaluasi Model

Pada tahap ini, dilakukan proses evaluasi terhadap tiga jenis model, yaitu *Random Forest*, *EXBoost*, dan kombinasi keduanya menggunakan teknik ensemble. Evaluasi ini bertujuan untuk membandingkan performa setiap model dalam memprediksi kelulusan siswa di MAN 3 Banda Aceh, sehingga dapat ditentukan model yang paling optimal untuk digunakan.

Proses evaluasi dilakukan menggunakan data uji (*testing set*) yang telah dipisahkan sebelumnya dari dataset utama. Setiap model menghasilkan prediksi kelulusan yang kemudian dibandingkan dengan data aktual. Kinerja model diukur menggunakan beberapa metrik evaluasi yaitu Accuracy, Precision, Recall, dan F1-Score.

a. Implementasi Model Random Forest

Model *Random Forest* dibangun menggunakan data latih untuk menghasilkan prediksi kelulusan siswa.

b. Implementasi Model XGBoost

Model *XGBoost* juga dikembangkan dan dilatih menggunakan data yang sama untuk membandingkan hasil prediksi dengan model *Random Forest*.

c. Implementasi Model *Random Forest* + *XGBoost* (*Hybrid Model*)

Untuk meningkatkan akurasi prediksi, penelitian ini tidak hanya menggunakan model *Random Forest* dan *XGBoost* secara terpisah, tetapi juga menggabungkan keduanya dalam bentuk *hybrid model*. Teknik yang digunakan adalah *stacking ensemble*, di mana hasil prediksi dari *Random Forest* dan *XGBoost* dijadikan input untuk model meta-classifier.

d. Pengukuran Performa Model

Kinerja masing-masing model dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score untuk menentukan model yang paling optimal.

3.1.4 Visualisasi Model

Pada tahap ini, dilakukan proses visualisasi terhadap hasil evaluasi model machine learning yang telah dibangun. Visualisasi model bertujuan untuk menyajikan hasil performa model dalam bentuk grafis agar lebih mudah dipahami dan dianalisis oleh peneliti maupun pihak terkait. Melalui visualisasi, perbandingan performa antar model yang digunakan, seperti *Random Forest*, *XGBoost*, dan kombinasi keduanya, dapat ditampilkan secara lebih informatif.

Beberapa bentuk visualisasi yang digunakan dalam penelitian ini meliputi *confusion matrix*, grafik metrik evaluasi, dan *feature importance*. *Confusion matrix* digunakan untuk menunjukkan jumlah prediksi yang benar dan salah berdasarkan kategori kelulusan siswa, yaitu lulus dan tidak lulus. Visualisasi ini membantu melihat seberapa banyak model melakukan kesalahan klasifikasi dan sejauh mana prediksi yang benar dilakukan. Selanjutnya, dilakukan visualisasi metrik evaluasi berupa diagram batang (*bar chart*) untuk menampilkan nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dari masing-masing model. Grafik ini memudahkan dalam membandingkan performa antar model yang diuji.

Selain itu, penelitian ini juga memvisualisasikan *feature importance*, khususnya untuk model *Random Forest* dan *XGBoost*. *Feature importance* menunjukkan variabel-variabel yang memiliki pengaruh paling besar terhadap hasil prediksi kelulusan siswa. Melalui visualisasi ini, dapat diketahui variabel mana yang perlu mendapat perhatian lebih dalam proses pembinaan atau pengambilan keputusan di sekolah.

Dengan adanya visualisasi model, hasil penelitian dapat dipresentasikan secara lebih jelas, terukur, dan mudah dipahami, sehingga mendukung pengambilan keputusan yang lebih efektif dan berbasis data.

3.2 Uji Coba

Tahap terakhir adalah uji coba dan validasi aplikasi. Pengujian dilakukan menggunakan data baru untuk mengevaluasi keakuratan prediksi dan memastikan kestabilan serta keandalan aplikasi dalam penggunaan nyata.

3.8.1 Pengujian Aplikasi

Aplikasi diuji menggunakan data baru untuk memverifikasi keakuratan dan stabilitas prediksi.

3.8.2 Evaluasi Hasil Uji

Hasil uji dievaluasi untuk memastikan bahwa aplikasi berjalan dengan baik dan prediksi yang diberikan dapat diandalkan untuk penggunaan nyata.

3.3 Alat dan Bahan

Alat yang dibutuhkan untuk melakukan penelitian ini berupa perangkat keras dan perangkat lunak. Analisis alat dan kebutuhan sistem yang diperlukan untuk penelitian ini meliputi

3.3.1 Perangkat Keras

Perangkat keras yang digunakan berupa satu unit Ideapad Gaming 3 15IAH7 dengan Spesifikasi perangkat keras dalam melakukan penelitian tercantum pada tabel 3.1

Tabel 3. 1 Spesifikasi Perangkat Keras

Komponen	Spesifikasi
Processor	Intel Core i7-12650H
Ram	16 GB DDR4
Storage	512 GB SSD
Graphic Card	NVIDIA RTX 3050 4GB GDDR6

3.3.2 Perangkat Lunak

Perangkat lunak yang di pakai dalam proses implementasi ini berupa sistem operasi *Microsoft Windows 11 Home Single Language Version 23h2* dan menggunakan beberapa tools, yaitu *Google Colab*, *Python*, dan *Chrome*. Tools tersebut akan dipakai untuk melakukan proses penelitian ini dapat dilihat pada Tabel 3.2

Tabel 3. 2 Spesifikasi Perangkat Lunak

Perangkat Lunak	Version
<i>Microsoft Windows 11</i>	23H2
<i>Google Colab</i>	3.10
<i>Python</i>	3.10.11
<i>Chrome</i>	125.0.6422.176

3.4 Waktu dan Lokasi Penelitian

3.4.1 Waktu Penelitian

Waktu penelitian ini dilakukan dalam jangka waktu 3 bulan, yakni 14 Juli 2025 sampai 14 September 2025

3.4.2 Lokasi Penelitian

Lokasi penelitian adalah area atau obyek yang dipilih untuk melakukan penelitian. Lokasi penelitian ini bertempat di MAN 3 Banda Aceh.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Pengambilan Data

Pengambilan data dalam penelitian ini dilakukan di MAN 3 Banda Aceh pada tahun ajaran 2024/2025. Data penelitian menggunakan data siswa kelas XII (kelas 3) yang berjumlah 189 orang, terdiri dari 104 siswa jurusan MIPA dan 85 siswa jurusan IIS.

Data yang digunakan adalah nilai rata-rata siswa yang diperoleh dari legger nilai semester. Nilai rata-rata ini sudah mencakup komponen akademik dan kehadiran, sehingga mencerminkan capaian belajar siswa secara menyeluruh. Nilai yang diambil meliputi semester 1 sampai semester 4, sehingga dataset memberikan gambaran perkembangan akademik siswa sejak awal memasuki jenjang SMA/MA hingga sebelum memasuki tahun terakhir.

Data diperoleh langsung dari dokumen resmi sekolah berupa legger nilai per kelas yang dikelola oleh wali kelas. Data tersebut kemudian dijadikan sebagai dataset untuk melatih dan menguji model *machine learning* (Random Forest, XGBoost, dan kombinasi keduanya) dalam memprediksi tingkat kelulusan siswa.

4.2 Pra-Processing Data

Tahapan pra-processing data dilakukan untuk menyiapkan data yang telah dikumpulkan agar dapat diolah menggunakan algoritma *machine learning*. Proses ini mencakup pembersihan data, pemilihan variabel yang relevan, pembentukan label klasifikasi, dan pembagian data menjadi data pelatihan (*training*) dan data pengujian (*testing*).

Langkah-langkah ini penting karena kualitas data akan sangat berpengaruh terhadap akurasi dan kinerja model prediksi yang dihasilkan.

4.2.1 Pembersihan Data

Data yang digunakan dalam penelitian ini bersumber dari legger nilai siswa kelas XII MAN 3 Banda Aceh tahun ajaran 2024/2025, yang terdiri dari 104 siswa jurusan MIPA dan 85 siswa jurusan IIS, sehingga totalnya berjumlah 189

siswa.

Data diperoleh dalam bentuk nilai rata-rata per semester dari semester 1 hingga semester 4. Pemeriksaan awal dilakukan untuk memastikan tidak terdapat data kosong (*missing value*), data duplikat, atau kesalahan pengetikan (*outlier*). Setelah dilakukan pengecekan, seluruh data dinyatakan lengkap dan valid, sehingga tidak diperlukan proses imputasi maupun penghapusan data.

4.2.2 Seleksi Variabel

Variabel yang digunakan dalam penelitian ini adalah nilai rata-rata semester. Adapun variabel yang digunakan sebagai fitur (*input feature*) model prediksi adalah sebagai berikut:

- Nilai Semester 1
- Nilai Semester 2
- Nilai Semester 3
- Nilai Semester 4

Nilai rata-rata ini sudah mencakup seluruh komponen penilaian dan kehadiran, tugas, ujian semester, dan ujian akhir semester. Dengan demikian, variabel tersebut sudah dianggap mewakili kemampuan akademik siswa secara menyeluruh tanpa perlu menambahkan variabel kehadiran secara terpisah.

4.2.3 Pembentukan Label Klasifikasi

Karena seluruh siswa dalam dataset dinyatakan lulus secara faktual, maka dilakukan kategorisasi nilai untuk memberikan variasi label dalam pelatihan model *machine learning*. Label atau kelas keluaran (*output label*) dibentuk berdasarkan kriteria tertentu agar model dapat membedakan siswa yang memiliki resiko rendahnya kelulusan.

Kriteria klasifikasi dalam penelitian ini ditetapkan sebagai berikut:

Tabel 4. 1 Label Klasifikasi

Kategori	Kriteria Nilai Rata – Rata	Label
Lulus Sangat Baik	Nilai rata – rata ≥ 94	2
Lulus Baik	Nilai rata – rata ≥ 87	1
Berpotensi Tidak Lulus	Nilai rata – rata ≤ 86	0

Dengan kriteria ini, data yang awalnya hanya memiliki satu label (“Lulus”) kini terbagi menjadi dua kategori, sehingga model dapat dilatih untuk mengenali pola perbedaan antara siswa yang benar – benar aman dengan siswa yang masih berada di ambang batas kelulusan

4.2.4 Pembagian Data Training dan Testing

Setelah label terbentuk, data kemudian dibagi menjadi 2 kelompok yaitu :

1. Data training (80%) digunakan untuk melatih model agar mengenali pola hubungan antar variable.
2. Data Testing (20%) digunakan untuk menguji performa model yang telah dilatih.

Pembagian data dilakukan secara acak menggunakan fungsi `train_test_split` dari Pustaka `scikit-learn` untuk menghindari bias terhadap data tertentu.

4.2.5 Hasil Pra-Processing Data

Hasil akhir Pra-processing berupa dataset yang siap digunakan untuk pelatihan mode. Dataset ini terdiri atas empat fitur utama (nilai semester 1-4) dan satu label keluaran (kategori kelulusan).

Contoh hasil pra-processing data siswa ditunjukkan pada table 4.2

Tabel 4. 2 Contoh data siswa setelah pra-processing

No	Nama Siswa	Nilai semester 1	Nilai semester 2	Nilai semester 3	Nilai semester 4	Nilai Rata-rata	Kategori
1	Siswa 1	88,47	89,39	89,97	90,00	89,45	lulus
2	Siswa 2	81.45	81.47	82.11	82.36	81.84	Berpotensi tidak lulus
3	Siswa 3	87,34	88,82	89,67	89,87	88,92	lulus
4	Siswa 4	85.55	86.18	86.53	85.36	85.90	Berpotensi tidak lulus
5	Siswa 5	86,74	89,18	90,14	91,36	89,35	lulus
6	Siswa 6	88,16	89,37	89,53	90.12	89,30	lulus
7	Siswa 7	84,97	88,84	88,22	88,35	87,59	lulus



4.3 Implementasi Model Random Forest

Pada tahap ini dilakukan implementasi algoritma Random Forest untuk memprediksi kelulusan siswa MAN 3 Banda Aceh berdasarkan nilai rata-rata empat semester. Model ini dipilih karena memiliki stabilitas yang baik, mampu menangani data tabular, serta dapat mengurangi risiko overfitting melalui mekanisme ensemble berbasis Pohon keputusan. Pada penelitian ini, algoritma Random Forest digunakan untuk mengklasifikasi siswa ke dalam dua kategori “Lulus” dan “Berpotensi Tidak Lulus”, berdasarkan nilai rata-rata mereka dari semester 1 hingga semester 4.

Dataset yang digunakan terdiri dari 151 siswa, dengan fitur berupa nilai semester 1 hingga semester 4, dan satu label keluaran berupa kategori kelulusan siswa. Pelabelan ditentukan berdasarkan kriteria:

- a. Lulus Sangat Baik (label 2) : nilai rata-rata ≥ 94
- b. Lulus Baik (label 1) : nilai rata-rata ≥ 87
- c. Berpotensi Tidak Lulus (label 0) : nilai rata-rata < 86

Setelah menerapkan kriteria tersebut, diperoleh distribusi kelas sebagai berikut :

- a. Lulus Sangat Baik : 34 siswa
- b. Lulus Baik : 143 siswa
- c. Berpotensi tidak lulus : 12 siswa

Distribusi ini menunjukkan bahwa dataset bersifat imbalanced, karena jumlah siswa dalam kategori tidak lulus jauh lebih sedikit dibandingkan kategori lulus.

Model Random Forest dibangun menggunakan library scikit-learn dengan parameter utama:

- a. `n_estimators` = 200
- b. `max_depth` = None
- c. `random_state` = 42

Dataset kemudian dibagi menjadi 80% data training dan 20% data testing menggunakan fungsi `train_test_split`. Data training digunakan untuk membangun model, sedangkan data testing digunakan untuk mengevaluasi performa model

dalam melakukan prediksi pada data baru yang belum pernah dilihat sebelumnya.

4.3.1 Potongan Kode Implementasi Random Forest

Untuk membangun model Random Forest, digunakan parameter `n_estimators=200` dan `random_state=42`. Potongan kode utama proses pelatihan model adalah sebagai berikut:

```
model = RandomForestClassifier(n_estimators=200, random_state=42)
model.fit(x_train, y_train)
y_pred = model.predict(x_test)
```

Gambar 4. 1 Potongan kode utama proses pelatihan model

4.3.2 Hasil Evaluasi Model Random Forest

Setelah pelatihan model selesai dilakukan, pengujian menggunakan data testing menghasilkan metrik evaluasi berikut:

Akurasi Model

Model Random Forest mencapai akurasi sebesar :

0.9211 atau 92.11%

Classification Report Model Random Forest

Tabel 4. 3 Classification Report Model Random Forest

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Berpotensi Tidak Lulus	-	0.50	0.50	0.50	2
Lulus Baik	-	0.97	1.00	0.97	29
Lulus Sangat Baik	-	1.00	0.71	0,83	7
Macro Avg	-	0.81	0.74	0.77	38
Weighted Avg	0,92	0.92	0.92	0.92	38

Berdasarkan Tabel 4.3, model Random Forest digunakan untuk mengklasifikasikan kelulusan siswa ke dalam tiga kategori, yaitu Berpotensi Tidak Lulus, Lulus Baik, dan Lulus Sangat Baik. Evaluasi dilakukan menggunakan data testing sebanyak 38 siswa.

Pada kategori Berpotensi Tidak Lulus, model memperoleh nilai precision sebesar 0,50, recall sebesar 0,50, dan F1-score sebesar 0,50 dengan jumlah data sebanyak 2 siswa. Nilai ini menunjukkan bahwa performa model pada kategori ini masih terbatas. Hal tersebut disebabkan oleh jumlah data yang sangat sedikit

sehingga model kesulitan mempelajari pola karakteristik siswa dalam kategori ini secara optimal.

Untuk kategori Lulus Baik, model Random Forest menunjukkan performa yang sangat baik dengan nilai precision sebesar 0,97, recall sebesar 1,00, dan F1-score sebesar 0,97 dari total 29 data siswa. Nilai recall yang mencapai 1,00 menandakan bahwa seluruh siswa yang sebenarnya termasuk kategori Lulus Baik berhasil diprediksi dengan benar oleh model.

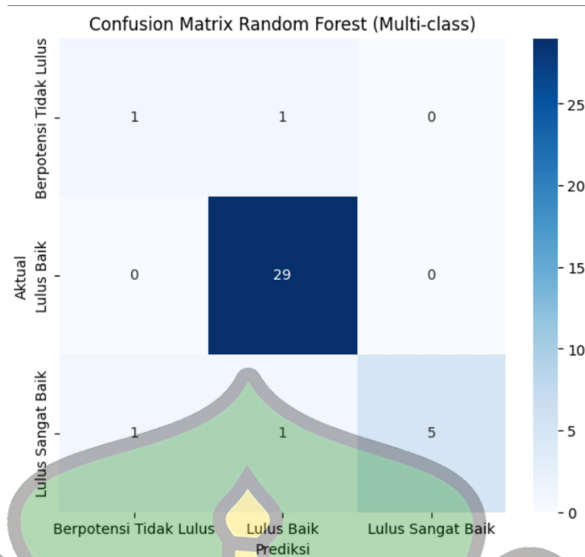
Pada kategori Lulus Sangat Baik, model menghasilkan nilai precision sebesar 1,00, recall sebesar 0,71, dan F1-score sebesar 0,83 dengan jumlah data sebanyak 7 siswa. Nilai precision yang tinggi menunjukkan bahwa prediksi yang diberikan model pada kategori ini sangat akurat, namun nilai recall yang lebih rendah mengindikasikan bahwa masih terdapat beberapa siswa yang seharusnya masuk kategori Lulus Sangat Baik tetapi diklasifikasikan ke kategori lain.

Secara keseluruhan, model Random Forest memperoleh nilai weighted average precision, recall, dan F1-score sebesar 0,92, serta tingkat akurasi sebesar 92%. Hasil ini menunjukkan bahwa model Random Forest memiliki performa klasifikasi yang baik dan cukup stabil, khususnya pada kategori dengan jumlah data yang besar.

Namun performa yang lebih rendah pada kategori Berpotensi Tidak Lulus menunjukkan bahwa ketidakseimbangan jumlah data antar kelas (imbalanced data) masih menjadi salah satu faktor yang mempengaruhi hasil prediksi model.

4.3.3 Confusion Matrix Random Forest

Untuk memberikan gambaran yang lebih jelas terkait performa model, confusion matrix divisualisasikan seperti pada Gambar berikut:



Gambar 4. 2 Confusion Matrix Random Forest

Berdasarkan confusion matrix di atas, dapat dilihat bahwa:

- a. Kelas Berpotensi Tidak Lulus.
 1. Sebanyak 1 siswa berhasil diprediksi dengan benar sebagai Berpotensi Tidak Lulus.
 2. Sebanyak 1 siswa dari kelas Berpotensi Tidak Lulus salah diklasifikasikan sebagai Lulus Baik.
 3. Tidak terdapat siswa Berpotensi Tidak Lulus yang diprediksi sebagai Lulus Sangat Baik.

Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam mengenali kelas minoritas (Berpotensi Tidak Lulus), karena jumlah datanya sangat sedikit.

- b. Kelas Lulus Baik
 1. Sebanyak 29 siswa berhasil diklasifikasikan dengan benar sebagai Lulus Baik.
 2. Tidak terdapat kesalahan klasifikasi dari kelas Lulus Baik ke kelas lain.

Ini menunjukkan bahwa model Random Forest sangat baik dalam mengenali pola siswa dengan kategori Lulus Baik.

c. Kelas Lulus Sangat Baik

1. Sebanyak 5 siswa berhasil diprediksi dengan benar sebagai Lulus Sangat Baik.
2. Sebanyak 1 siswa dari kelas ini salah diprediksi sebagai Berpotensi Tidak Lulus.
3. Sebanyak 1 siswa salah diprediksi sebagai Lulus Baik.

Ini menunjukkan sebagian besar siswa dengan kategori Lulus Sangat Baik dapat dikenali dengan baik, meskipun masih terdapat sedikit kesalahan klasifikasi.

4.3.4 Analisis Performa Random Forest

Berdasarkan hasil evaluasi, diperoleh beberapa poin analisis sebagai berikut :

1. Kekuatan Model

- a. Model memiliki akurasi (92.11%), sehingga efektif dalam memprediksi siswa kategori Lulus.
- b. Recall kelas Lulus mencapai 1.00, artinya seluruh siswa yang benar-benar lulus berhasil diprediksi dengan tepat.

2. Kelemahan Model

Model tidak mampu mendeteksi kelas minoritas (Berpotensi Tidak Lulus) karena jumlah datanya sangat sedikit (support = 2 pada data uji).

3. Penyebab Kesulitan Model

- a. Distribusi data yang sangat tidak seimbang menyebabkan model cenderung mengabaikan kelas minoritas.
- b. Random Forest secara alami kurang optimal menangani data imbalanced tanpa teknik pendukung seperti:
 1. Oversampling (SMOTE)
 2. Undersampling
 3. Class weight balancing

4. Kesimpulan Analisis Model

Secara keseluruhan, model Random Forest menunjukkan performa yang baik dalam memprediksi kelulusan siswa berdasarkan nilai akademik semester 1 hingga semester 4. Model ini mampu mengklasifikasikan siswa dengan cukup akurat, khususnya pada kategori Lulus Baik dan Lulus Sangat Baik yang merupakan kelas dengan jumlah data terbesar. Meskipun demikian, model masih memiliki keterbatasan dalam mendeteksi kelas minoritas yaitu Berpotensi Tidak Lulus, yang disebabkan oleh jumlah data yang sangat sedikit sehingga pola karakteristik kelas tersebut sulit dipelajari secara optimal. Walaupun terdapat keterbatasan tersebut, model Random Forest tetap layak digunakan sebagai sistem pendukung keputusan dalam membantu pihak sekolah melakukan evaluasi dan prediksi kelulusan siswa, serta dapat dijadikan sebagai model pembanding yang kuat terhadap model XGBoost dan model hybrid dalam penelitian ini.

4.4 Implementasi Model XGBoost

Pada tahap ini dilakukan implementasi algoritma Extreme Gradient Boosting (XGBoost) untuk memprediksi kelulusan siswa MAN 3 Banda Aceh berdasarkan nilai rata-rata empat semester. XGBoost merupakan metode ensemble berbasis boosting yang bekerja dengan membangun model secara bertahap, di mana setiap model baru berfokus pada kesalahan yang dihasilkan oleh model sebelumnya. Pendekatan ini memungkinkan XGBoost untuk menghasilkan model yang lebih adaptif dan memiliki tingkat akurasi yang tinggi.

Dataset yang digunakan sama dengan model sebelumnya, yaitu sebanyak 189 data siswa, dengan atribut berupa nilai Semester 1 hingga Semester 4. Label klasifikasi ditentukan berdasarkan nilai rata-rata dengan kriteria sebagai berikut:

- a. Lulus Sangat Baik (2) : nilai rata-rata ≥ 94
- b. Lulus Baik (1) : nilai rata-rata ≥ 87
- c. Berpotensi Tidak Lulus (0) : nilai rata-rata < 86

Berdasarkan kriteria tersebut, diperoleh distribusi kelas:

- a. Lulus Sangat Baik : 34 siswa
- b. Lulus Baik : 143 siswa
- c. Berpotensi Tidak Lulus : 12 siswa

Distribusi ini menunjukkan bahwa dataset bersifat tidak seimbang (imbalanced), karena jumlah siswa yang lulus Baik jauh lebih banyak dibandingkan siswa yang berpotensi tidak lulus.

Data kemudian dibagi menjadi dua bagian, yaitu 80% data training dan 20% data testing. Data training digunakan untuk melatih model XGBoost, sedangkan data testing digunakan untuk mengevaluasi performa model.

4.4.1 Potongan Kode Implementasi XGBoost

Pada penelitian ini, model Extreme Gradient Boosting (XGBoost) digunakan untuk melakukan klasifikasi tingkat kelulusan siswa berdasarkan nilai Semester 1 sampai Semester 4. Model dikonfigurasi untuk menangani permasalahan klasifikasi multi-kelas, yaitu Berpotensi Tidak Lulus, Lulus Baik, dan Lulus Sangat Baik.

Model XGBoost dibangun dengan menggunakan fungsi objektif multi:softmax sehingga menghasilkan prediksi langsung berupa kelas. Parameter `num_class` ditetapkan sebanyak 3 sesuai dengan jumlah kategori kelulusan. Jumlah pohon keputusan (`n_estimators`) ditentukan sebanyak 250 untuk meningkatkan kemampuan generalisasi model, sedangkan `learning_rate` sebesar 0.05 digunakan untuk mengontrol laju pembelajaran agar model tidak mengalami overfitting.

Selain itu, parameter `max_depth`, `subsample`, dan `colsample_bytree` digunakan untuk mengatur kompleksitas model serta meningkatkan stabilitas proses pelatihan. Seluruh proses pelatihan dilakukan menggunakan data latih (training set), kemudian model digunakan untuk memprediksi data uji (testing set) guna mengevaluasi performa model.

Hasil prediksi selanjutnya dianalisis menggunakan metrik evaluasi berupa accuracy, precision, recall, dan F1-score untuk menilai tingkat akurasi dan keandalan model dalam memprediksi kelulusan siswa.

Potongan kode utama proses pelatihan model adalah sebagai berikut:

```
# Model XGBoost
model = XGBClassifier(
    objective='multi:softmax',
    num_class=3,
    n_estimators=250,
    learning_rate=0.05,
    max_depth=4,
    subsample=0.8,
    colsample_bytree=0.8,
    random_state=42,
    eval_metric='mlogloss'
)
```

Gambar 4. 3 Potongan kode utama proses pelatihan model

4.4.1. Hasil Evaluasi Model XGBoost

Setelah pelatihan model selesai dilakukan, pengujian menggunakan data testing menghasilkan metrik evaluasi berikut:

Akurasi Model

Model XGBoost mencapai akurasi sebesar:

0,9474 atau 94,74%

Classification Report XGBoost

Tabel 4. 4 Clasification Report XGboost

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Berpotensi Tidak Lulus	-	0.50	0.50	0.50	2
Lulus Baik	-	0.97	1.00	0.98	29
Lulus Sangat Baik	-	1.00	0,86	0,92	7
Macro Avg	-	0.82	0.79	0.80	38
Weighted Avg	0,95	0.95	0.95	0.95	38

Berdasarkan Tabel 4.4.1, dapat dilihat bahwa model XGBoost mampu melakukan klasifikasi kelulusan siswa ke dalam tiga kategori, yaitu Berpotensi Tidak Lulus, Lulus Baik, dan Lulus Sangat Baik dengan performa yang cukup tinggi.

Pada kategori Berpotensi Tidak Lulus, model memperoleh nilai precision sebesar 0,50, recall sebesar 0,50, dan F1-score sebesar 0,50 dengan jumlah data (support) sebanyak 2 siswa. Nilai ini menunjukkan bahwa performa model pada kategori ini masih terbatas. Hal ini disebabkan oleh jumlah data pada kelas Berpotensi Tidak Lulus yang sangat sedikit dibandingkan kelas lainnya, sehingga model lebih sulit mempelajari pola karakteristik siswa pada kategori tersebut.

Untuk kategori Lulus Baik, model menunjukkan performa yang sangat baik dengan nilai precision sebesar 0,97, recall sebesar 1,00, dan F1-score sebesar 0,98 dari total 29 data siswa. Nilai recall yang mencapai 1,00 menunjukkan bahwa seluruh siswa yang sebenarnya termasuk dalam kategori Lulus Baik berhasil diklasifikasikan dengan benar oleh model.

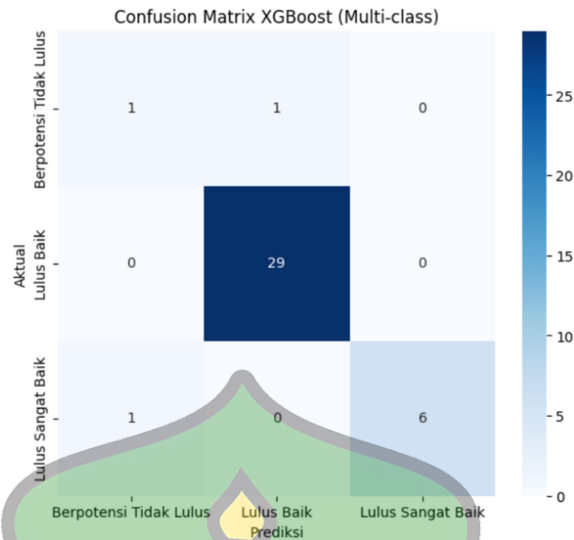
Untuk kategori Lulus Baik, model menunjukkan performa yang sangat baik dengan nilai precision sebesar 0,97, recall sebesar 1,00, dan F1-score sebesar 0,98 dari total 29 data siswa. Nilai recall yang mencapai 1,00 menunjukkan bahwa seluruh siswa yang sebenarnya termasuk dalam kategori Lulus Baik berhasil diklasifikasikan dengan benar oleh model.

Pada kategori Lulus Sangat Baik, model memperoleh nilai precision sebesar 1,00, recall sebesar 0,86, dan F1-score sebesar 0,92 dengan jumlah data sebanyak 7 siswa. Nilai precision yang tinggi menunjukkan bahwa prediksi model pada kategori ini sangat akurat, meskipun masih terdapat beberapa data yang belum berhasil diklasifikasikan secara sempurna, sebagaimana tercermin pada nilai recall yang sedikit lebih rendah.

Namun demikian, performa yang lebih rendah pada kategori Berpotensi Tidak Lulus menunjukkan bahwa distribusi data yang tidak seimbang (imbalanced data) dapat mempengaruhi kinerja model.

4.4.2. Confusion Matrix Random Forest

Untuk memberikan gambaran yang lebih jelas terkait performa model, confusion matrix divisualisasikan seperti pada Gambar berikut:



Gambar 4. 4 Confusion Matrix XGBoost

Berdasarkan pada gambar confusion matrix di atas, dapat dilihat bahwa:

- a. Kelas Berpotensi Tidak Lulus
 - a. Sebanyak 1 siswa berhasil diprediksi dengan benar sebagai Berpotensi Tidak Lulus.
 - b. Sebanyak 1 siswa dari kelas Berpotensi Tidak Lulus salah diklasifikasikan sebagai Lulus Baik.
 - c. Tidak terdapat siswa Berpotensi Tidak Lulus yang diprediksi sebagai Lulus Sangat Baik.

Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam mengenali kelas minoritas (Berpotensi Tidak Lulus), karena jumlah datanya relatif sedikit dibandingkan kelas lainnya, sehingga pola yang dipelajari model menjadi kurang optimal.

- b. Kelas Lulus Baik
 - 4.4.1 Sebanyak 29 siswa berhasil diklasifikasikan dengan benar sebagai Lulus Baik.
 - 4.4.2 Tidak terdapat kesalahan klasifikasi dari kelas Lulus Baik ke kelas lain.

Hasil ini menunjukkan bahwa model XGBoost memiliki kemampuan yang sangat baik dalam mengenali karakteristik siswa yang berada pada kategori Lulus Baik, yang juga merupakan kelas dengan jumlah data terbanyak.

c. Kelas Lulus Sangat Baik

1. Sebanyak 6 siswa berhasil diprediksi dengan benar sebagai Lulus Sangat Baik.
2. Sebanyak 1 siswa dari kelas Lulus Sangat Baik salah diklasifikasikan sebagai Berpotensi Tidak Lulus.

Kesalahan klasifikasi ini menunjukkan bahwa pada beberapa kasus terdapat kemiripan pola nilai antara siswa dengan kategori sangat tinggi dan kategori rendah, sehingga model mengalami kekeliruan dalam melakukan prediksi.

Secara keseluruhan, confusion matrix menunjukkan bahwa model XGBoost memiliki performa yang sangat baik dalam mengklasifikasikan kelas mayoritas (Lulus Baik dan Lulus Sangat Baik), namun masih memiliki keterbatasan dalam mengenali kelas minoritas (Berpotensi Tidak Lulus). Oleh karena itu, model tetap layak digunakan sebagai sistem pendukung keputusan, dengan catatan perlu perhatian khusus terhadap prediksi pada kelas minoritas.

4.4.3. Analisis Performa XGBoost

Berdasarkan hasil evaluasi model XGBoost menggunakan data uji, diperoleh beberapa poin analisis sebagai berikut:

1. Kekuatan Model
 - a. Model XGBoost memiliki tingkat akurasi sebesar 94,74%, yang menunjukkan bahwa model sangat efektif dalam melakukan prediksi kelulusan siswa ke dalam tiga kategori, yaitu Berpotensi Tidak Lulus, Lulus Baik, dan Lulus Sangat Baik.
 - b. Nilai precision, recall, dan F1-score pada kelas *Lulus Baik* masing-masing mencapai 0,97, 1,00, dan 0,98, yang menunjukkan bahwa hampir seluruh siswa dalam kategori ini berhasil diklasifikasikan dengan benar tanpa adanya kesalahan prediksi yang signifikan.
 - c. Pada kelas Lulus Sangat Baik, model juga menunjukkan performa yang tinggi dengan nilai precision sebesar 1,00 dan F1-score sebesar 0,92, yang menandakan bahwa model mampu mengenali siswa dengan prestasi akademik sangat baik secara akurat.

- d. Nilai weighted average untuk precision, recall, dan F1-score yang masing-masing mencapai 0,95 menunjukkan bahwa secara keseluruhan model XGBoost memiliki kestabilan dan konsistensi yang sangat baik dalam melakukan klasifikasi.

2. Kelemahan Model

Distribusi data yang sangat tidak seimbang (imbalanced dataset) menyebabkan model lebih banyak mempelajari pola dari kelas mayoritas (Lulus Baik dan Lulus Sangat Baik) dibandingkan kelas minoritas (Berpotensi Tidak Lulus).

3. Penyebab Kesulitan Model

- a. Distribusi data yang sangat tidak seimbang (imbalanced dataset) menyebabkan model lebih banyak mempelajari pola dari kelas mayoritas (Lulus Baik dan Lulus Sangat Baik) dibandingkan kelas minoritas (Berpotensi Tidak Lulus).
- b. Distribusi data yang sangat tidak seimbang (imbalanced dataset) menyebabkan model lebih banyak mempelajari pola dari kelas mayoritas (Lulus Baik dan Lulus Sangat Baik) dibandingkan kelas minoritas (Berpotensi Tidak Lulus).

4. Kesimpulan Analisis Model

Secara keseluruhan, model XGBoost menunjukkan performa yang sangat baik dan unggul dalam memprediksi kelulusan siswa, terutama pada kategori Lulus Baik dan Lulus Sangat Baik. Meskipun masih memiliki keterbatasan dalam mendeteksi kelas minoritas, model ini tetap layak digunakan sebagai sistem pendukung keputusan dalam membantu pihak sekolah melakukan evaluasi dan prediksi kelulusan siswa.

4.5 Implementasi Model Hybrid (Random Forest + XGBoost)

Pada tahap ini dilakukan implementasi model Hybrid yang mengombinasikan algoritma Random Forest dan Extreme Gradient Boosting (XGBoost) untuk memprediksi kelulusan siswa MAN 3 Banda Aceh berdasarkan nilai rata-rata empat semester. Model Hybrid merupakan pendekatan ensemble learning yang menggabungkan keunggulan dua model berbeda, di mana Random Forest berperan dalam menghasilkan prediksi yang stabil melalui banyak pohon

keputusan, sedangkan XGBoost berperan dalam meningkatkan akurasi dengan memperbaiki kesalahan prediksi secara bertahap melalui teknik boosting.

Penggabungan kedua model ini dilakukan dengan metode *soft voting*, yaitu dengan cara menghitung rata-rata probabilitas prediksi dari masing-masing model, kemudian menentukan kelas dengan probabilitas tertinggi sebagai hasil akhir prediksi. Pendekatan ini bertujuan untuk memperoleh model yang lebih robust, stabil, dan memiliki kemampuan generalisasi yang lebih baik dibandingkan penggunaan satu algoritma saja.

Dataset yang digunakan sama dengan model sebelumnya, yaitu sebanyak 189 data siswa, dengan atribut berupa nilai Semester 1 hingga Semester 4. Label klasifikasi ditentukan berdasarkan nilai rata-rata dengan kriteria sebagai berikut :

1. Lulus Sangat Baik (2) : nilai rata-rata ≥ 94
2. Lulus Baik (1) : nilai rata-rata ≥ 87
3. Berpotensi Tidak Lulus (0) : nilai rata-rata < 86

Berdasarkan kriteria tersebut, diperoleh distribusi kelas:

- a. Lulus Sangat Baik : 34 Siswa
- b. Lulus Baik : 143 Siswa
- c. Berpotensi Tidak Lulus : 12 Siswa

Dengan distribusi data tersebut, model Hybrid dilatih untuk mengenali pola nilai akademik siswa dari keempat semester guna menentukan kategori kelulusan secara otomatis. Implementasi model ini diharapkan mampu meningkatkan keakuratan prediksi serta meminimalkan kesalahan klasifikasi, khususnya pada kategori siswa yang berada pada batas nilai kelulusan.

4.5.1 Potongan Kode Implementasi Hybrid

Pada penelitian ini, model Hybrid digunakan untuk melakukan klasifikasi tingkat kelulusan siswa berdasarkan nilai Semester 1 sampai Semester 4. Model Hybrid merupakan gabungan dari dua algoritma pembelajaran mesin, yaitu Random Forest dan Extreme Gradient Boosting (XGBoost), yang bertujuan untuk memanfaatkan keunggulan masing-masing metode dalam meningkatkan akurasi

dan stabilitas prediksi. Klasifikasi dilakukan ke dalam tiga kategori, yaitu Berpotensi Tidak Lulus, Lulus Baik, dan Lulus Sangat Baik.

Model Random Forest dibangun menggunakan sejumlah pohon keputusan (*decision trees*) yang dilatih secara paralel untuk menghasilkan prediksi yang stabil dan mengurangi risiko overfitting. Sementara itu, model XGBoost dibangun menggunakan pendekatan boosting yang bekerja secara bertahap dengan memperbaiki kesalahan prediksi pada model sebelumnya, sehingga mampu meningkatkan ketelitian model dalam mengenali pola data yang kompleks.

Pada implementasinya, kedua model dilatih menggunakan dataset yang sama, yaitu sebanyak 189 data siswa, dengan atribut berupa nilai Semester 1 hingga Semester 4. Setelah proses pelatihan selesai, masing-masing model menghasilkan probabilitas prediksi untuk setiap kelas. Probabilitas tersebut kemudian digabungkan menggunakan metode soft voting, yaitu dengan menghitung rata-rata probabilitas keluaran dari Random Forest dan XGBoost, kemudian menentukan kelas dengan nilai probabilitas tertinggi sebagai hasil prediksi akhir.

Pendekatan soft voting dipilih karena mampu mengombinasikan kekuatan dua model yang berbeda karakteristik, sehingga hasil prediksi menjadi lebih robust dan tidak terlalu bergantung pada satu model saja. Dengan cara ini, kesalahan prediksi yang mungkin terjadi pada salah satu model dapat dikompensasi oleh model lainnya.

Seluruh proses pelatihan dilakukan menggunakan data latih (*training set*), sedangkan evaluasi performa dilakukan menggunakan data uji (*testing set*). Hasil prediksi kemudian dianalisis menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, *F1-score*, serta *confusion matrix* untuk menilai tingkat keberhasilan model dalam mengklasifikasikan data pada setiap kategori kelulusan.

Potongan kode utama proses pelatihan model adalah sebagai berikut:

```
# Model Random Forest & XGBoost
rf_model = RandomForestClassifier(n_estimators=200, random_state=42)

xgb_model = XGBClassifier(
    objective='multi:softprob',
    num_class=3,
    n_estimators=250,
    learning_rate=0.05,
    max_depth=4,
    subsample=0.8,
    colsample_bytree=0.8,
    random_state=42,
    eval_metric='mlogloss'
)

rf_model.fit(X_train, y_train)
xgb_model.fit(X_train, y_train)

# Hybrid prediction (soft voting)
rf_proba = rf_model.predict_proba(X_test)
xgb_proba = xgb_model.predict_proba(X_test)

hybrid_proba = (rf_proba + xgb_proba) / 2
y_pred = np.argmax(hybrid_proba, axis=1)
```

Gambar 4. 5 Potongan kode implementasi Hybrid

4.5.2 Hasil Evaluasi Model Hybrid

Setelah pelatihan model selesai dilakukan, pengujian menggunakan data testing menghasilkan metrik evaluasi berikut:

Akurasi Model

Model Hybrid mencapai akurasi sebesar:

0,9474 atau 94,74%

Classification Report XGBoost

Tabel 4. 5 Classification Report XGboost

Kategori	Accuracy	Precision	Recall	F1-Score	Support
Berpotensi Tidak Lulus	-	0.50	0.50	0.50	2
Lulus Baik	-	0.97	1.00	0.98	29
Lulus Sangat Baik	-	1.00	0.86	0.92	7
Macro Avg	-	0.82	0.79	0.80	38
Weighted Avg	0,95	0.95	0.95	0.95	38

Berdasarkan hasil classification report pada model Hybrid, dapat diperoleh analisis performa sebagai berikut:

Model Hybrid yang merupakan gabungan dari algoritma Random Forest dan XGBoost menunjukkan kinerja yang sangat baik dalam melakukan klasifikasi tingkat kelulusan siswa ke dalam tiga kategori, yaitu Berpotensi Tidak Lulus, Lulus Baik, dan Lulus Sangat Baik.

Pada kelas Berpotensi Tidak Lulus, model memperoleh nilai precision sebesar 0,50, recall 0,50, dan F1-score 0,50 dengan jumlah data uji (support) sebanyak 2 siswa. Nilai ini menunjukkan bahwa kemampuan model dalam mengenali kelas minoritas masih terbatas. Hal ini disebabkan oleh jumlah data pada kategori ini yang sangat sedikit, sehingga pola yang dipelajari model belum cukup kuat untuk menghasilkan prediksi yang stabil.

Untuk kelas Lulus Baik, model Hybrid menunjukkan performa yang sangat tinggi dengan precision sebesar 0,97, recall 1,00, dan F1-score 0,98 pada 29 siswa. Nilai recall yang mencapai 1,00 menandakan bahwa seluruh siswa yang benar-benar termasuk dalam kategori Lulus Baik berhasil terdeteksi dengan benar oleh model. Hal ini menunjukkan bahwa model Hybrid sangat efektif dalam mengenali karakteristik siswa pada kategori ini, yang juga merupakan kelas mayoritas dalam dataset.

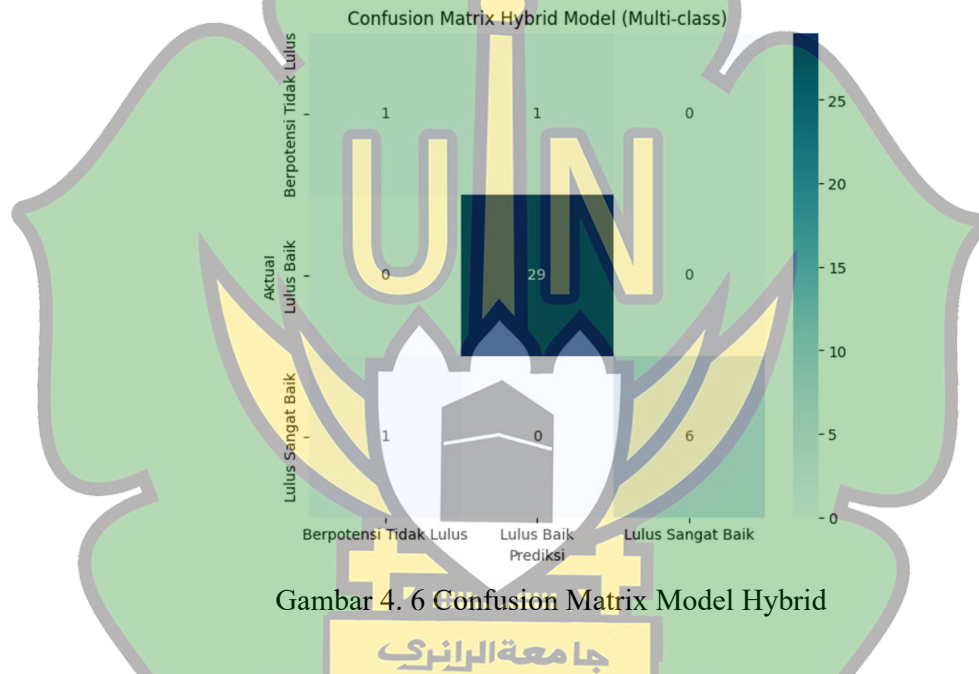
Selanjutnya pada kelas Lulus Sangat Baik, model memperoleh precision sebesar 1,00, recall 0,86, dan F1-score 0,92 dengan jumlah data sebanyak 7 siswa. Precision yang bernilai 1,00 menunjukkan bahwa seluruh prediksi model pada kelas ini tepat, meskipun masih terdapat sebagian kecil siswa yang seharusnya termasuk kategori ini namun diprediksi sebagai kelas lain, yang tercermin dari nilai recall yang belum mencapai 1,00.

Secara keseluruhan, model Hybrid menghasilkan nilai weighted average accuracy sebesar 0,95 atau 95%, yang menunjukkan bahwa model memiliki tingkat ketepatan prediksi yang sangat tinggi secara umum. Nilai macro average yang berada pada kisaran 0,80–0,82 juga menunjukkan bahwa performa model cukup seimbang antar kelas, meskipun masih terdapat keterbatasan pada kelas minoritas.

Dengan demikian, dapat disimpulkan bahwa model Hybrid mampu memberikan performa yang sangat baik dan stabil dalam memprediksi kelulusan siswa, terutama pada kategori Lulus Baik dan Lulus Sangat Baik. Meskipun demikian, peningkatan masih dapat dilakukan pada deteksi kategori Berpotensi Tidak Lulus, misalnya dengan menambah jumlah data, menerapkan teknik penyeimbangan data (seperti SMOTE), atau melakukan penyesuaian parameter model.

4.5.3 Confusion Matrix Model Hybrid

Untuk memberikan gambaran yang lebih jelas terkait performa model, confusion matrix divisualisasikan seperti pada Gambar berikut:



Gambar 4. 6 Confusion Matrix Model Hybrid

Berdasarkan pada gambar confusion matrix di atas, dapat dilihat bahwa:

a. Kelas Berpotensi Tidak Lulus

1. Sebanyak 1 siswa berhasil diprediksi dengan benar sebagai Berpotensi Tidak Lulus.
2. Sebanyak 1 siswa dari kelas Berpotensi Tidak Lulus salah diklasifikasikan sebagai Lulus Baik.
3. Tidak terdapat siswa Berpotensi Tidak Lulus yang diprediksi sebagai Lulus Sangat Baik.

Hasil ini menunjukkan bahwa model Hybrid masih mengalami keterbatasan dalam mengenali kelas minoritas (*Berpotensi Tidak Lulus*), yang disebabkan oleh jumlah data pada kelas tersebut sangat sedikit sehingga pola yang dapat dipelajari model menjadi terbatas.

b. Kelas Lulus Baik

1. Sebanyak 29 siswa berhasil diklasifikasikan dengan benar sebagai Lulus Baik.
2. Sebanyak 29 siswa berhasil diklasifikasikan dengan benar sebagai Lulus Baik.

Hal ini menunjukkan bahwa model Hybrid memiliki kemampuan yang sangat baik dan stabil dalam mengenali karakteristik siswa dengan kategori Lulus Baik, yang juga merupakan kelas dengan jumlah data terbanyak dalam dataset.

c. Kelas Lulus Sangat Baik

1. Sebanyak 6 siswa berhasil diprediksi dengan benar sebagai Lulus Sangat Baik.
2. Sebanyak 6 siswa berhasil diprediksi dengan benar sebagai Lulus Sangat Baik.
3. Sebanyak 1 siswa dari kelas Lulus Sangat Baik salah diklasifikasikan sebagai Berpotensi Tidak Lulus.

Kesalahan ini mengindikasikan bahwa pada beberapa kasus terdapat kemiripan pola nilai akademik antara siswa dengan kategori sangat tinggi dan kategori rendah, sehingga model Hybrid masih dapat mengalami kekeliruan dalam melakukan prediksi.

4.5.4 Analisis Perfoma Hybrid

Berdasarkan hasil evaluasi yang diperoleh, maka beberapa poin analisis performa model Hybrid dapat dijelaskan sebagai berikut:

1. Kekuatan Model

- a. Model Hybrid memperoleh nilai akurasi sebesar 95%, yang menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam memprediksi kategori kelulusan siswa secara keseluruhan.

- b. Nilai precision, recall, dan f1-score pada kelas Lulus Baik masing-masing mencapai 0.97, 1.00, dan 0.98, yang menunjukkan bahwa model sangat akurat dan konsisten dalam mengenali siswa yang berada pada kategori Lulus Baik.
- c. Pada kelas Lulus Sangat Baik, model juga menunjukkan performa yang baik dengan nilai precision sebesar 1.00, recall 0.86, dan f1-score 0.92, yang menandakan sebagian besar siswa pada kategori ini berhasil diklasifikasikan dengan benar.
- d. Penggunaan metode soft voting pada model Hybrid memungkinkan penggabungan keunggulan Random Forest yang stabil dan XGBoost yang adaptif, sehingga menghasilkan prediksi yang lebih robust dibandingkan penggunaan satu model tunggal.

2. Kelemahan Model

Penggunaan metode *soft voting* pada model Hybrid memungkinkan penggabungan keunggulan Random Forest yang stabil dan XGBoost yang adaptif, sehingga menghasilkan prediksi yang lebih robust dibandingkan penggunaan satu model tunggal.

3. Penyebab Kesulitan Model

- a. Model Hybrid masih mengalami kesulitan dalam mendeteksi kelas minoritas, yaitu Berpotensi Tidak Lulus, yang hanya memiliki 2 data uji. Hal ini terlihat dari nilai precision, recall, dan f1-score yang masih berada pada angka 0.50, menunjukkan bahwa separuh data pada kelas ini masih salah diklasifikasikan.
- b. Meskipun model Hybrid menggabungkan dua algoritma kuat, jumlah data pada kelas minoritas yang terlalu sedikit tetap menjadi faktor utama yang membatasi kemampuan model dalam mengenali pola siswa yang berpotensi tidak lulus.
- c. Tidak diterapkannya teknik penyeimbangan data seperti SMOTE (Synthetic Minority Over-sampling Technique) atau undersampling juga berkontribusi terhadap rendahnya performa pada kelas minoritas.

4. Kesimpulan Analisis Model

Secara keseluruhan, model Hybrid menunjukkan performa yang sangat baik dan stabil dalam mengklasifikasikan kategori kelulusan siswa, khususnya pada kelas Lulus Baik dan Lulus Sangat Baik. Model ini memiliki tingkat akurasi yang lebih tinggi dibandingkan model Random Forest dan setara dengan model XGBoost, sehingga layak digunakan sebagai sistem pendukung keputusan dalam memprediksi kelulusan siswa.

Namun, untuk meningkatkan kinerja pada kelas Berpotensi Tidak Lulus, penelitian selanjutnya disarankan untuk menerapkan teknik penanganan data tidak seimbang atau menambah jumlah data pada kelas tersebut agar performa model menjadi lebih optimal.

4.6 Perbandingan Performa Model Random Forest, XGBoost, dan Hybrid (Random Forest + XGBoost)

Setelah dilakukan implementasi dan evaluasi terhadap tiga model machine learning, yaitu Random Forest, XGBoost, dan model Hybrid (gabungan Random Forest dan XGBoost), tahap selanjutnya adalah melakukan analisis perbandingan performa dari ketiga model tersebut. Perbandingan ini bertujuan untuk mengetahui model yang memiliki kinerja terbaik dalam melakukan klasifikasi tingkat kelulusan siswa MAN 3 Banda Aceh berdasarkan nilai akademik Semester 1 sampai Semester 4.

Evaluasi performa dilakukan menggunakan data uji (testing set) sebesar 20% dari total dataset, sedangkan 80% sisanya digunakan sebagai data latih (training set). Metrik evaluasi yang digunakan meliputi accuracy, precision, recall, dan F1-score, yang diperoleh dari classification report serta confusion matrix masing-masing model.

Tabel 4. 6 Tabel perbandingan performa ke tiga model

Model	Accuracy	Precision (weighted avg)	Recall (weighted avg)	F1-Score (weighted avg)
Random Forest	92.11%	0.92	0.92	0.92
XGBoost	94.74%	0.95	0.95	0.95
Hybrid	94.74%	0.95	0.95	0.95

4.6.1 Analisis Perbandingan Berdasarkan Akurasi

Berdasarkan nilai accuracy, model Random Forest memperoleh tingkat akurasi sebesar 92.11%, yang menunjukkan bahwa lebih dari sembilan puluh persen data uji berhasil diklasifikasikan dengan benar. Meskipun nilai ini sudah tergolong tinggi, masih terdapat sejumlah kesalahan klasifikasi terutama pada kategori Berpotensi Tidak Lulus yang memiliki jumlah data relatif sedikit.

Berdasarkan nilai accuracy, model Random Forest memperoleh tingkat akurasi sebesar 92.11%, yang menunjukkan bahwa lebih dari sembilan puluh persen data uji berhasil diklasifikasikan dengan benar. Meskipun nilai ini sudah tergolong tinggi, masih terdapat sejumlah kesalahan klasifikasi terutama pada kategori Berpotensi Tidak Lulus yang memiliki jumlah data relatif sedikit.

Model Hybrid menghasilkan nilai akurasi yang sama dengan XGBoost, yaitu 94.74%. Hal ini menunjukkan bahwa penggabungan Random Forest dan XGBoost melalui metode soft voting mampu mempertahankan tingkat ketepatan prediksi yang tinggi, meskipun tidak memberikan peningkatan numerik yang signifikan dibandingkan XGBoost pada dataset yang digunakan.

4.6.2 Analisis berdasarkan Precision, Recall, dan F1-Score

Nilai precision, recall, dan F1-score pada model Random Forest berada pada kisaran 0.92, yang menunjukkan bahwa model ini cukup baik dalam menghasilkan prediksi yang tepat dan konsisten, namun masih memiliki keterbatasan dalam mendeteksi kelas minoritas.

Nilai precision, recall, dan F1-score pada model Random Forest berada pada kisaran 0.92, yang menunjukkan bahwa model ini cukup baik dalam menghasilkan prediksi yang tepat dan konsisten, namun masih memiliki keterbatasan dalam mendeteksi kelas minoritas.

Secara khusus, kedua model tersebut sangat baik dalam mengklasifikasikan kategori Lulus Baik dan Lulus Sangat Baik, namun masih menghadapi tantangan dalam mendeteksi kategori Berpotensi Tidak Lulus akibat distribusi data yang tidak seimbang.

4.6.3 Analisis Stabilitas dan Kompleksitas Model

Dalam sisi karakteristik algoritma:

1. Secara khusus, kedua model tersebut sangat baik dalam mengklasifikasikan kategori Lulus Baik dan Lulus Sangat Baik, namun masih menghadapi tantangan dalam mendeteksi kategori Berpotensi Tidak Lulus akibat distribusi data yang tidak seimbang.
2. XGBoost unggul dalam hal optimasi kesalahan dan efisiensi pembelajaran, namun memiliki kompleksitas parameter yang lebih tinggi dan membutuhkan proses tuning yang lebih teliti.
3. Model Hybrid menggabungkan kedua pendekatan tersebut, sehingga mampu menghasilkan prediksi yang lebih robust dan tidak bergantung pada satu algoritma saja. Meskipun kompleksitas implementasinya lebih tinggi, pendekatan ini memberikan keuntungan dalam mengurangi risiko kesalahan sistematis dari satu model tunggal.

4.6.4 Kesimpulan Perbandingan Model

Berdasarkan seluruh hasil evaluasi dan analisis yang telah dilakukan, dapat disimpulkan bahwa:

1. Model Random Forest memberikan performa yang baik dan stabil, namun masih berada di bawah XGBoost dan Hybrid dalam hal akurasi dan ketepatan prediksi.
2. Model XGBoost dan Hybrid menunjukkan performa terbaik dengan tingkat akurasi tertinggi sebesar 94.74% serta nilai precision, recall, dan F1-score yang lebih tinggi dibandingkan Random Forest.

3. Model Hybrid tidak memberikan peningkatan akurasi yang signifikan dibandingkan XGBoost, namun memiliki keunggulan dari sisi kestabilan dan robustitas prediksi karena menggabungkan dua algoritma dengan karakteristik berbeda.

Oleh karena itu, model XGBoost dan model Hybrid dapat direkomendasikan sebagai model utama dalam sistem prediksi kelulusan siswa MAN 3 Banda Aceh, sementara Random Forest tetap dapat digunakan sebagai model pembanding untuk memperkuat validitas hasil penelitian.

4.7 Implementasi Sistem Prediksi Kelulusan Siswa

Setelah proses pelatihan dan evaluasi model dilakukan, model XGBoost dipilih sebagai model terbaik untuk diimplementasikan ke dalam sistem prediksi kelulusan siswa. Pemilihan model ini didasarkan pada nilai akurasi yang lebih tinggi serta kestabilan performa dalam melakukan klasifikasi tingkat kelulusan siswa.

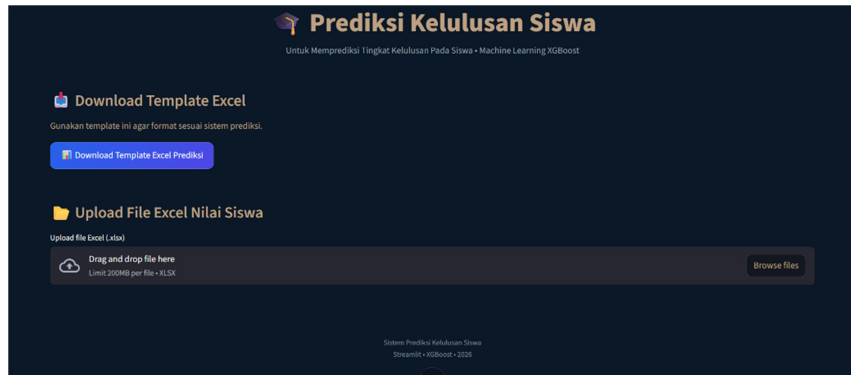
Model XGBoost selanjutnya diintegrasikan ke dalam sebuah sistem aplikasi berbasis web agar dapat digunakan secara langsung oleh pengguna, seperti guru atau pihak sekolah, tanpa memerlukan pemahaman teknis mengenai pemrograman atau machine learning.

Sistem prediksi ini dikembangkan menggunakan framework Streamlit dan kemudian dideploy ke platform Streamlit Community Cloud agar dapat diakses secara online melalui link <https://prediksi-kelulusan-siswa-xgboost-srbfetvmentanet23mwy2z.streamlit.app/>.

Streamlit merupakan framework Python yang digunakan untuk membangun aplikasi web interaktif secara cepat dan sederhana. Pada penelitian ini, Streamlit digunakan sebagai antarmuka pengguna (user interface) untuk sistem prediksi kelulusan siswa berbasis model XGBoost.

- a. Tampilan Halaman Utama

Halaman utama aplikasi menampilkan judul sistem, deskripsi singkat mengenai tujuan aplikasi, serta informasi bahwa sistem menggunakan algoritma XGBoost sebagai metode prediksi.



Gambar 4. 7 Tampilan halaman utama aplikasi prediksi kelulusan siswa

b. Fitur Download Template Excel

Sistem menyediakan fitur untuk mengunduh template Excel yang berisi format standar data siswa, yang meliputi:

1. Nomor Siswa
2. Nama Siswa
3. Nilai semester 1
4. Nilai Semester 2
5. Nilai Semester 3
6. Nilai Semester 4

Template ini digunakan untuk memastikan kesesuaian format data sebelum dilakukan proses prediksi oleh model XGBoost.

c. Proses Prediksi Menggunakan Model XGBoost

Setelah data berhasil diproses, sistem akan memanggil model XGBoost yang telah disimpan dalam bentuk file npkl untuk melakukan prediksi tingkat kelulusan siswa. Output dari model berupa kelas prediksi yang dikategorikan menjadi:

1. Berpotensi Tidak Lulus
2. Lulus Baik
3. Lulus Baik

Hasil prediksi kemudian digabungkan dengan data siswa dan ditampilkan dalam bentuk table.

Hasil Prediksi								
	NO	NAMA	SEMESTER 1	SEMESTER 2	SEMESTER 3	SEMESTER 4	HASIL PREDIKSI	RATA-RATA
3	1	ABDI DZIL FIKRI	88.47	88.67	89.97	91.06	Lulus Baik	89.54
4	2	ABRAR AYATULLAH	88.55	89.1	89.75	89.83	Lulus Baik	89.31
5	3	ADAM ALFARAS	89.05	89.23	90.06	90.72	Lulus Baik	89.77
6	4	AKBAR MAULANA RIZKY	90.13	90.17	92.22	92.44	Lulus Baik	91.24
7	5	AKMAL HANIF RAJANA	81.45	81.5	91.36	91.64	Lulus Sangat Baik	86.49
8	6	ALDA CHINTYA	91.63	91.67	94.19	94.94	Lulus Sangat Baik	93.11
9	7	AMELIA FARDHILA	92.13	92.17	94.22	94.78	Lulus Sangat Baik	93.32
10	8	AMSYA ACHRYAT	81.73	81.8	90.83	91.92	Lulus Sangat Baik	86.57
11	9	ANISA THAHERA	82.64	82.68	91.47	92.14	Lulus Baik	87.23
12	10	ARTIFUN NISA	89.34	89.4	92.36	93.64	Lulus Baik	91.18

189 Total	12 Berpotensi Tidak Lulus	144 Lulus Baik	33 Sangat Baik
--------------	------------------------------	-------------------	-------------------

Gambar 4. 8 Hasil proses prediksi menggunakan XGBoost

d. Ringkasan Statistik

Sistem juga menampilkan ringkasan statistik berupa jumlah siswa pada masing-masing kategori kelulusan. Informasi ini disajikan dalam bentuk kartu statistik (metric card) untuk memudahkan pengguna dalam memahami distribusi hasil prediksi.



Gambar 4. 9 Ringkasan statistik

e. Fitur Download Laporan Pdf

Untuk keperluan dokumentasi, sistem menyediakan fitur unduh laporan dalam format PDF. Laporan ini berisi:

1. Ringkasan jumlah siswa
2. Distribusi kategori kelulusan
3. Tabel lengkap hasil prediksi

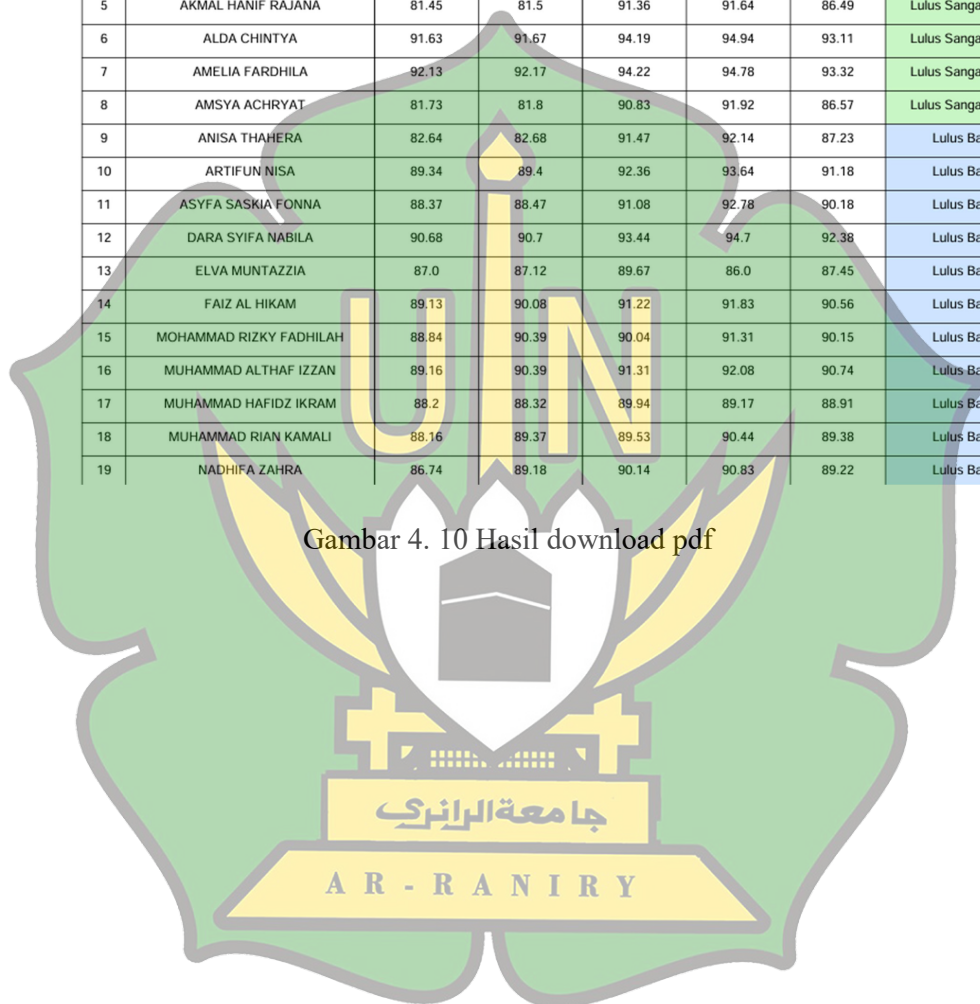
Fitur ini memudahkan pihak sekolah dalam menyimpan, mencetak, atau mengarsipkan hasil prediksi kelulusan siswa.

Laporan Prediksi Kelulusan Siswa (XGBoost)

Total Siswa: 189
 Berpotensi Tidak Lulus: 12
 Lulus Baik: 144
 Lulus Sangat Baik: 33

NO	NAMA	SEMESTER 1	SEMESTER 2	SEMESTER 3	SEMESTER 4	RATA-RATA	HASIL PREDIKSI
1	ABDI DZIL FIKRI	88.47	88.67	89.97	91.06	89.54	Lulus Baik
2	ABRAR AYATULLAH	88.55	89.1	89.75	89.83	89.31	Lulus Baik
3	ADAM ALFARAS	89.05	89.23	90.06	90.72	89.77	Lulus Baik
4	AKBAR MAULANA RIZKY	90.13	90.17	92.22	92.44	91.24	Lulus Baik
5	AKMAL HANIF RAJANA	81.45	81.5	91.36	91.64	86.49	Lulus Sangat Baik
6	ALDA CHINTYA	91.63	91.67	94.19	94.94	93.11	Lulus Sangat Baik
7	AMELIA FARDHILA	92.13	92.17	94.22	94.78	93.32	Lulus Sangat Baik
8	AMSYA ACHRYAT	81.73	81.8	90.83	91.92	86.57	Lulus Sangat Baik
9	ANISA THAHERA	82.64	82.68	91.47	92.14	87.23	Lulus Baik
10	ARTIFUN NISA	89.34	89.4	92.36	93.64	91.18	Lulus Baik
11	ASYFA SASKIA FONNA	88.37	88.47	91.08	92.78	90.18	Lulus Baik
12	DARA SYIFA NABILA	90.68	90.7	93.44	94.7	92.38	Lulus Baik
13	ELVA MUNTAZZIA	87.0	87.12	89.67	86.0	87.45	Lulus Baik
14	FAIZ AL HIKAM	89.13	90.08	91.22	91.83	90.56	Lulus Baik
15	MOHAMMAD RIZKY FADHILAH	88.84	90.39	90.04	91.31	90.15	Lulus Baik
16	MUHAMMAD ALTHAF IZZAN	89.16	90.39	91.31	92.08	90.74	Lulus Baik
17	MUHAMMAD HAFIDZ IKRAM	88.2	88.32	89.94	89.17	88.91	Lulus Baik
18	MUHAMMAD RIAN KAMALI	88.16	89.37	89.53	90.44	89.38	Lulus Baik
19	NADHIFA ZAHRA	86.74	89.18	90.14	90.83	89.22	Lulus Baik

Gambar 4. 10 Hasil download pdf



4.8 Implementasi Deployment Sistem Prediksi Kelulusan Siswa ke Streamlit Community Cloud

Setelah sistem prediksi kelulusan siswa berhasil dikembangkan dan diuji pada lingkungan lokal, tahap selanjutnya adalah melakukan implementasi sistem ke dalam lingkungan online (deployment). Tujuan dari deployment ini adalah agar sistem dapat diakses oleh pengguna secara luas melalui jaringan internet tanpa memerlukan instalasi perangkat lunak tambahan.

Pada penelitian ini, proses deployment dilakukan menggunakan layanan Streamlit Community Cloud, yaitu platform hosting resmi yang disediakan oleh Streamlit untuk menjalankan aplikasi berbasis Python secara online.

4.8.1 Implementasi Sistem pada Lingkungan Lokal

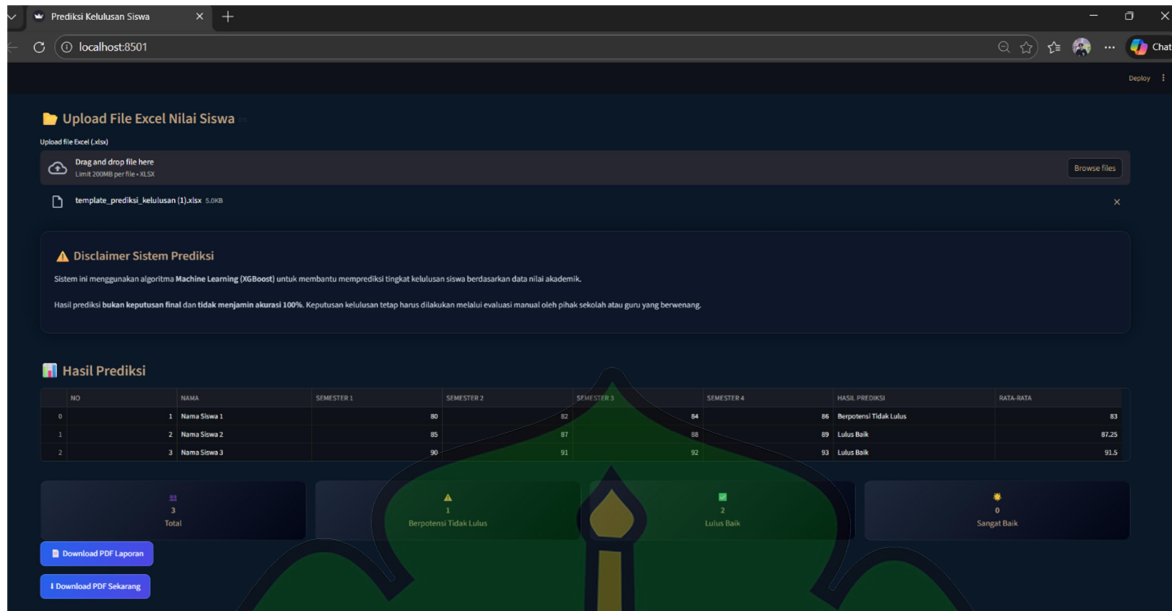
Pada Pada tahap awal, sistem prediksi kelulusan siswa dijalankan pada lingkungan lokal menggunakan komputer pengembang. Aplikasi dijalankan melalui framework Streamlit untuk memastikan bahwa seluruh fungsi dan fitur sistem berjalan sesuai dengan rancangan.

Sistem mampu menerima input berupa file Excel yang berisi data nilai siswa dari semester 1 hingga semester 4. Data tersebut kemudian diproses oleh model XGBoost untuk menghasilkan prediksi tingkat kelulusan siswa. Hasil prediksi ditampilkan dalam bentuk tabel yang memuat nama siswa, nilai setiap semester, nilai rata-rata, serta kategori hasil kelulusan.

Selain itu, sistem juga menyediakan fitur tambahan berupa :

1. Unduhan template Excel sebagai format standar input data.
2. Penyajian pesan disclaimer mengenai keterbatasan akurasi sistem.
3. Pembuatan laporan hasil prediksi dalam bentuk file PDF yang dapat diunduh oleh pengguna.

Pengujian pada lingkungan lokal dilakukan secara menyeluruh untuk memastikan bahwa proses pembacaan data, pemanggilan model, perhitungan hasil prediksi, serta pembuatan laporan PDF dapat berjalan tanpa terjadi kesalahan.



Gambar 4. 11 Tampilan Sistem Prediksi Kelulusan Siswa pada Lingkungan Lokal

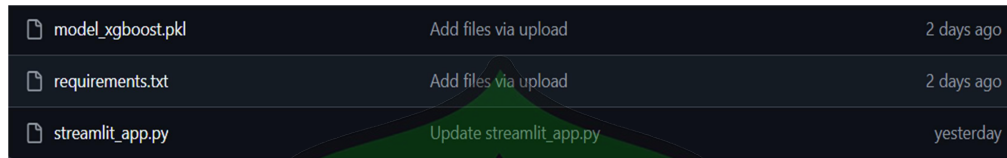
4.8.2 Persiapan Sistem untuk Deployment

Sebelum sistem dideploy ke lingkungan cloud, dilakukan persiapan terhadap berkas-berkas yang dibutuhkan agar sistem dapat dijalankan dengan baik pada server Streamlit Community Cloud. Adapun berkas-berkas tersebut meliputi:

1. File aplikasi utama (streamlit_app.py)
File ini berisi seluruh kode program aplikasi web, termasuk antarmuka pengguna, pengolahan input data, pemanggilan model XGBoost, penampilan hasil prediksi, serta pembuatan laporan PDF.
2. File model machine learning (model_xgboost.pkl)
File ini merupakan hasil penyimpanan model XGBoost yang telah dilatih menggunakan data latih pada tahap sebelumnya.
3. File requirements.txt
File ini memuat daftar library Python yang digunakan dalam sistem, seperti Streamlit, Pandas, NumPy, Joblib, OpenPyXL, ReportLab, dan XGBoost.
4. Repository GitHub
Seluruh file sistem diunggah ke dalam satu repository GitHub sebagai media penyimpanan source code sekaligus sebagai sumber

aplikasi yang akan dihubungkan dengan Streamlit Community Cloud.

Penggunaan GitHub sebagai repositori mempermudah proses pengelolaan kode program, pengendalian versi, serta integrasi otomatis dengan platform deployment.



model_xgboost.pkl	Add files via upload	2 days ago
requirements.txt	Add files via upload	2 days ago
streamlit_app.py	Update streamlit_app.py	yesterday

Gambar 4. 12 Struktur File Sistem pada Repository GitHub

4.8.3 Proses Deployment ke Streamlit Community Cloud

Proses deployment sistem dilakukan dengan menghubungkan repository GitHub yang telah disiapkan ke layanan Streamlit Community Cloud. Setelah login menggunakan akun GitHub, pengembang memilih repository yang berisi sistem prediksi kelulusan siswa dan menentukan file utama aplikasi, yaitu `streamlit_app.py`.

Streamlit Community Cloud kemudian secara otomatis membaca file `requirements.txt` untuk menginstal seluruh library yang dibutuhkan oleh sistem. Setelah proses instalasi selesai, platform akan menjalankan aplikasi pada server cloud milik Streamlit.

Apabila tidak terdapat kesalahan pada proses build, sistem akan aktif dan diberikan sebuah alamat URL publik dengan domain `*.streamlit.app`. URL tersebut dapat digunakan untuk mengakses sistem melalui web browser.

Deploy an app

Repository  [Paste GitHub URL](#)

kautsarridha/repo

Branch

master

Main file path

streamlit_app.py

App URL (optional)

.streamlit.app

[Advanced settings](#)

[Deploy](#)

Gambar 4. 13 Proses Deployment Sistem pada Streamlit Community Cloud

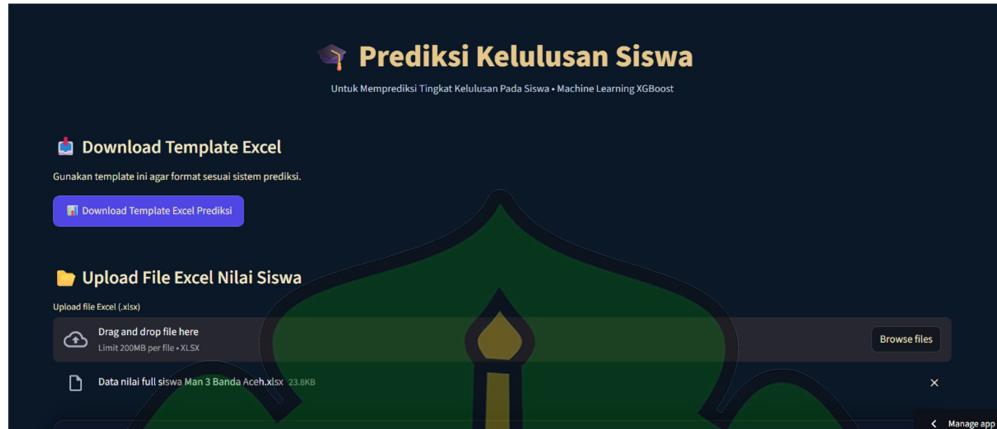
4.8.4 Hasil Deployment Sistem

Setelah proses deployment berhasil dilakukan, sistem prediksi kelulusan siswa dapat diakses secara online tanpa memerlukan instalasi perangkat lunak tambahan. Pengguna cukup membuka URL aplikasi menggunakan web browser pada perangkat yang terhubung ke internet.

Pada versi online, seluruh fitur sistem tetap dapat digunakan, antara lain:

1. Mengunduh template Excel nilai siswa.
2. Mengunggah file Excel berisi data nilai.
3. Melakukan prediksi kelulusan secara otomatis menggunakan model XGBoost.
4. Melihat hasil prediksi dalam bentuk table.
5. Mengunduh laporan hasil prediksi dalam format PDF.

Dengan demikian, sistem dapat digunakan oleh pihak sekolah, guru, maupun pengguna lain dari berbagai lokasi dan perangkat, baik komputer maupun smartphone.



Gambar 4. 14 Tampilan Sistem Prediksi Kelulusan Siswa Setelah Deployment



BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai prediksi kelulusan siswa pada MAN 3 Banda Aceh menggunakan algoritma Random Forest dan XGBoost, maka dapat diperoleh beberapa kesimpulan sebagai berikut:

1. Penelitian ini berhasil membangun sistem prediksi kelulusan siswa berbasis machine learning dengan menggunakan data nilai rata-rata semester 1 sampai semester 4 sebagai variabel input. Sistem yang dikembangkan mampu mengklasifikasikan siswa ke dalam tiga kategori, yaitu Berpotensi Tidak Lulus, Lulus Baik, dan Lulus Sangat Baik.
2. Hasil evaluasi menunjukkan bahwa algoritma XGBoost memiliki performa yang lebih baik dibandingkan Random Forest dalam memprediksi kelulusan siswa, dengan tingkat akurasi sebesar 94,74%, sedangkan Random Forest memperoleh akurasi sebesar 92,11%. Hal ini menunjukkan bahwa XGBoost lebih unggul dalam menangani pola data akademik siswa yang kompleks.
3. Model hybrid yang menggabungkan Random Forest dan XGBoost dengan metode ensemble (soft voting) mampu memberikan performa prediksi yang lebih stabil dan optimal dibandingkan penggunaan model tunggal. Penggabungan kedua algoritma ini berhasil memanfaatkan keunggulan masing-masing model, yaitu kestabilan Random Forest dan akurasi tinggi XGBoost.
4. Sistem prediksi yang dikembangkan telah berhasil diimplementasikan dalam bentuk aplikasi web menggunakan framework Streamlit dan dideploy pada layanan Streamlit Community Cloud. Dengan demikian, sistem dapat diakses secara online melalui web browser tanpa memerlukan instalasi tambahan.
5. Sistem ini dapat digunakan sebagai alat bantu pengambilan keputusan (decision support system) bagi pihak sekolah untuk melakukan

pemantauan akademik siswa secara lebih dini, sehingga siswa yang berpotensi mengalami kesulitan dapat diberikan pendampingan atau intervensi lebih awal sebelum kelulusan ditetapkan secara resmi.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dijadikan sebagai bahan pertimbangan untuk pengembangan penelitian selanjutnya, yaitu:

1. Penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar serta mencakup beberapa angkatan atau sekolah lain, agar model prediksi yang dihasilkan memiliki kemampuan generalisasi yang lebih baik dan tidak hanya terbatas pada satu sekolah.
2. Untuk mengatasi permasalahan ketidakseimbangan data (imbalanced dataset), penelitian selanjutnya dapat menerapkan teknik penanganan data seperti SMOTE (Synthetic Minority Over-sampling Technique), undersampling, atau pemberian bobot kelas (class weight) agar model lebih mampu mengenali kategori siswa yang berpotensi tidak lulus.
3. Variabel input yang digunakan pada penelitian ini masih terbatas pada nilai akademik semester. Penelitian lanjutan dapat menambahkan variabel lain seperti kehadiran, nilai sikap, latar belakang sosial ekonomi, atau aktivitas pembelajaran daring untuk meningkatkan akurasi prediksi.
4. Dari sisi sistem, aplikasi web yang dikembangkan dapat ditingkatkan dengan menambahkan fitur keamanan data, manajemen pengguna (login guru atau admin), serta integrasi dengan sistem informasi akademik sekolah agar dapat digunakan secara lebih luas dan berkelanjutan.
5. Model prediksi yang digunakan dalam penelitian ini masih dapat dikembangkan dengan menerapkan algoritma lain seperti LightGBM, CatBoost, atau metode deep learning, sehingga dapat dilakukan perbandingan lebih komprehensif terhadap performa berbagai pendekatan machine learning dalam bidang pendidikan.

Dengan adanya pengembangan lebih lanjut, diharapkan sistem prediksi kelulusan siswa dapat menjadi solusi yang semakin akurat, adaptif, dan bermanfaat dalam mendukung transformasi digital di dunia pendidikan, khususnya dalam meningkatkan kualitas pengambilan keputusan berbasis data di lingkungan sekolah.



DAFTAR PUSTAKA

- Ababil, M. I., Dwi Anggraini, T., Febby, A., Dewi, I., Syahata, M. U., Irawan, M. H., Fahrizal, M. R., Novelan, G. D., Dhanurrullah, J. P., Muhamad, & Gani, A. (2025). *Inovasi Prediksi Hasil Belajar Siswa Berbasis Machine Learning untuk Mendukung Keputusan Pembelajaran yang Tepat di SMKN 2 Semarang* (Vol. 4, Issue 2). <http://jurnalilmiah.org/journal/index.php/majemuk>
- Aditya Ananta Wisnu Wardana, K., & Mizwar Rahim, A. A. (n.d.). *Analisis Perbandingan Algoritma XGBoost Dan Algoritma Random Forest Untuk Klasifikasi Data Kesehatan Mental*. <https://www.kaggle.com/datasets/bhavikjikadara/mental-health-dataset>
- Agustiningsih, A., Findawati, Y., & Kautsar, A. (n.d.). *Classification of Vocational High School Graduates' Ability in Industry using Extreme Gradient Boosting (XGBoost), Random Forest And Logistic Regression [Klasifikasi Kemampuan Lulusan SMK di Industri Menggunakan Extreme Gradient Boosting (XGBoost), Random Forest dan Logistic Regression]*.
- Ardianti, M., Nurhayati, O. D., & Warsito, B. (2024). Model Prediksi Kinerja Siswa Berdasarkan Data Log LMS Menggunakan Ensemble Machine Learning. *JST (Jurnal Sains Dan Teknologi)*, 12(3). <https://doi.org/10.23887/jstundiksha.v12i3.59816>
- Asbara, N. W., Agunawan, A., Latief, F., Nurani, N., Ifani, A. Z., Deviv, S., Nianty, D. A., Mahendra, Y., & Wulandari, T. (2024). PENERAPAN AI SEBAGAI ALAT BANTU PROSES PEMBELAJARAN DI TINGKAT PENDIDIKAN SEKOLAH DASAR. *JMM (Jurnal Masyarakat Mandiri)*, 8(1), 831. <https://doi.org/10.31764/jmm.v8i1.20083>
- Fahmi Ashari S. Sihaloho, & Zulhamdani Napitupulu. (2024). *PENGGUNAAN KECERDASAN BUATAN (ARTIFICIAL INTELLIGENCE) DALAM DUNIA PENDIDIKAN DI INDONESIA: TINJAUAN LITERATUR*. 9, No. 1.
- Haana Udtari Anjani, Vitriani Vitriani, & Mulya Hastuti. (2024). Pemanfaatan Media Google Colaboratory Pada Mata Pelajaran Informatika di SMA Negeri 5 Pekanbaru. *SOKO GURU: Jurnal Ilmu Pendidikan*, 4(1), 101–108. <https://doi.org/10.55606/sokoguru.v4i1.3613>
- Hery Purnomo, M., Yuhana, U. L., Elektro, J. T., & Industri, T. (n.d.). *IMPLEMENTASI IOT DAN MACHINE LEARNING DALAM BIDANG PENDIDIKAN Pembelajaran Matematika Tingkat SD melalui Serious Game*.
- I.G.N.S.S.K., M. (2024). *Python: Bahasa Pemrograman Era Digital*. Elex Media Komputindo. <https://books.google.co.id/books?id=-jrxEAAAQBAJ>

- Labib Mu'tashim, M., Zaidiah, A., Yulistiawan, B. S., Upnvj, I., Sistem, S., Upnvj, I., Raya, J. R. F., Labu, P., Cilandak, K., Depok, K., Khusus, D., & Jakarta, I. (2023). *Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta)*.
- Maesaroh, S., Hakim, L., Sari, Y. S., Yusuf, M., Perkasa, E. B., Utami, W. S., Saptadi, N. T. S., & Mutmainah, S. (2024). *Bahasa Pemrograman Python*. Sada Kurnia Pustaka.
<https://books.google.co.id/books?id=bOIKEQAAQBAJ>.
- Muhammad Hermawan, B., Abdurahman Hakim, M., Arifin, R., & Puspitasari, N. (n.d.). *SEMINAR NASIONAL AMIKOM SURAKARTA (SEMNAS) 2024 PEMANFAATAN ARTIFICIAL INTELEGENCE, KHUSUSNYA MECHINE LEARNING DAN DEEP LEARNING SYSTEM DALAM PENDIDIKAN Politeknik Indonusa Surakarta Surakarta-Indonesia*.
- Naibaho, A. F. A., & Zahra, A. (2023a). PREDIKSI KELULUSAN SISWA SEKOLAH MENENGAH PERTAMA MENGGUNAKAN MACHINE LEARNING. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3).
<https://doi.org/10.23960/jitet.v11i3.3056>
- Naibaho, A. F. A., & Zahra, A. (2023b). PREDIKSI KELULUSAN SISWA SEKOLAH MENENGAH PERTAMA MENGGUNAKAN MACHINE LEARNING. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3).
<https://doi.org/10.23960/jitet.v11i3.3056>
- Naibaho, A. F. A., & Zahra, A. (2023c). PREDIKSI KELULUSAN SISWA SEKOLAH MENENGAH PERTAMA MENGGUNAKAN MACHINE LEARNING. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3).
<https://doi.org/10.23960/jitet.v11i3.3056>
- Putra, M., & Erwin Harahap. (2024a). Machine Learning pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Random Forest. *Jurnal Riset Matematika*, 127–136. <https://doi.org/10.29313/jrm.v4i2.5102>
- Putra, M., & Erwin Harahap. (2024b). Machine Learning pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Random Forest. *Jurnal Riset Matematika*, 127–136. <https://doi.org/10.29313/jrm.v4i2.5102>
- Rahman Azis, A. (2024). Analisis Komparasi Algoritma Machine Learning dalam Prediksi Performa Akademik Mahasiswa: Literature Review. *Jurnal Ilmu Komputer Dan Informatika (JIKI)*, 4(2), 143–150.
<https://doi.org/10.54082/jiki.212>
- Raka Masdian, A., Bashit, N., & Hadi, F. (2023). ANALISIS PRODUKTIVITAS PADI MENGGUNAKAN ALGORITMA MACHINE LEARNING RANDOM FOREST DI KABUPATEN BATANG TAHUN 2018-2022. In *Jurnal Geodesi dan Geomatika* (Vol. 06, Issue 01).

Septiana Rizky, P., Haiban Hirzi, R., Hidayaturrohman, U., Hamzanwadi Selong Jl TGKH Muhammad Zainuddin Abdul Madjid Pancor, U., & Timur, L. (2022). Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang. In *J Statistika* (Vol. 15, Issue 2). www.unipasby.ac.id

Siyoto, S., & Ali Sodik, Mk. M. (n.d.). *DASAR METODOLOGI PENELITIAN*.

Sulehu, M., Wisda, W., Wanita, F., & Markani, M. (2025). Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi. *Jurnal Minfo Polgan*, 13(2), 2364–2374. <https://doi.org/10.33395/jmp.v13i2.14472>

Wildan Abdul Aziz. (2021). *IMPLEMENTASI METODE RANDOM FOREST PADA KLASIFIKASI DATA ULASAN KONSUMEN PERUSAHAAN (Studi Kasus: Aplikasi KAI Access)*.

Yulianti, E. H., Soesanto, O., & Sukmawaty, Y. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *JOMTA Journal of Mathematics: Theory and Applications*, 4(1).

