JURNAL KOMPUTER INDONESIA
e-ISSN 2962-3626

Implementation Of Xgboost For Classifying Student Interest In Informatics At SMP Negeri 30 Bengkulu Selatan

Implementasi Algoritma Xgboost Dalam Klasifikasi Minat Siswa Terhadap Mata Pelajaran Teknik Informatika

Ocy Lara Putri [1)]; Devi Sartika [2)]; Dimas Aulia Trianggana [3)]

[1)] Study Program of Informatics, Faculty of Computer Science, Universitas Dehasen Bengkulu
[2,3)] Department of Informatics, Faculty of Computer Science, Universitas Dehasen Bengkulu
Email: 1) Ocyylaraputri@gmail.com

How to Cite :

Putri , O, L., Sartika , D., Trianggana , D, A. (2026). Implementation of XGBoost for Classifying Student Interest in Informatics at SMP Negeri 30 Bengkulu Selatan. Jurnal Komputer Indonesia, 5(3).

ARTICLE HISTORY

Received [30 Juli 2026]
Revised [09 September 2026]
Accepted [11 September 2026]

KEYWORDS

Explainable Artificial Intelligence; Student Learning Interest; XGBoost.

This is an open access article under the CC–BY-SA license

ABSTRAK

Identifikasi minat belajar siswa pada mata pelajaran Informatika masih cenderung bergantung pada pengamatan guru, sehingga diperlukan pendekatan berbasis data yang lebih objektif. Penelitian ini bertujuan membangun model klasifikasi minat siswa menggunakan XGBoost dan menjelaskan faktor pembentuk keputusan model melalui SHAP. Penelitian kuantitatif berbasis data analytics mengikuti tahapan CRISP-DM dengan 27 data siswa dan tujuh fitur, yaitu nilai tugas, ulangan harian, praktik, UTS, UAS, kehadiran, dan sikap. Data dibagi 80% untuk pelatihan dan 20% untuk pengujian. Pada enam data uji, model mengklasifikasikan dua siswa kategori Minat dan empat siswa kategori Tidak Minat tanpa kesalahan, sehingga Accuracy, Precision, Recall, dan F1-Score masing-masing mencapai 100% dengan AUC 1,00. Analisis SHAP menunjukkan Ulangan Harian sebagai fitur paling dominan dengan nilai kontribusi 1,1179, disusul nilai tugas, UAS, praktik, dan UTS. Hasil ini menunjukkan bahwa XGBoost dan SHAP dapat mendukung identifikasi minat siswa secara terukur dan transparan, meskipun generalisasi hasil perlu mempertimbangkan ukuran dataset yang terbatas.

ABSTRACT

Student interest in Informatics is still commonly identified through teacher observation, creating a need for a more objective data-driven approach. This study develops an XGBoost classification model and explains the factors underlying its decisions using SHAP. A quantitative data-analytics design followed the CRISP-DM stages using 27 student records and seven features: assignment scores, daily tests, practical scores, midterm scores, final scores, attendance, and attitude. The dataset was split into 80% training and 20% testing data. On six held-out cases, the model correctly classified two students as Interested and four as Not Interested, producing 100% Accuracy, Precision, Recall, and F1-Score, with an AUC of 1.00. SHAP analysis identified daily-test scores as the strongest contributor (1.1179), followed by assignment, final-exam, practical, and midterm scores. These findings indicate that XGBoost combined with SHAP can support measurable and transparent identification of student interest, although the limited dataset requires caution when generalizing the performance.

PENDAHULUAN

Pemanfaatan teknologi informasi dalam pendidikan telah berkembang dari sekadar media pembelajaran menjadi instrumen analisis yang mendukung keputusan akademik. Di tingkat sekolah menengah, data nilai, kehadiran, dan perilaku belajar sebenarnya menyimpan pola yang dapat digunakan untuk mengidentifikasi kondisi belajar siswa secara lebih objektif. Data akademik

mencerminkan capaian siswa dari berbagai aktivitas evaluasi dan dapat menjadi dasar untuk memetakan kecenderungan performa maupun kebutuhan pendampingan (Marcydiaz et al., 2024). Pemanfaatan data tersebut menjadi penting ketika guru harus menangani karakteristik siswa yang beragam dalam waktu pembelajaran yang terbatas.

Pada pembelajaran Informatika di SMP Negeri 30 Bengkulu Selatan, pengamatan lapangan menunjukkan adanya perbedaan keterlibatan dan capaian siswa. Sebagian siswa aktif mengikuti diskusi dan praktik, sedangkan sebagian lainnya memperlihatkan hasil belajar yang lebih rendah dan membutuhkan pendampingan lebih intensif. Kondisi tersebut relevan dengan karakter pembelajaran Informatika yang menggabungkan pengetahuan konseptual dengan aktivitas praktik. Pangesthi dan Triyani (2022) menjelaskan bahwa pembelajaran Informatika pada jenjang SMP diarahkan pada pemahaman konsep teknologi informasi, pengolahan data, dan penerapannya melalui kombinasi teori dan praktik. Dengan demikian, variasi capaian pada tugas, ulangan, praktik, UTS, UAS, kehadiran, dan sikap dapat digunakan untuk membaca pola keterlibatan siswa selama pembelajaran.

Minat belajar merupakan salah satu aspek yang berhubungan dengan ketertarikan, perhatian, kesadaran, dan konsentrasi siswa dalam mengikuti proses pembelajaran. Minat yang kuat cenderung ditunjukkan melalui keterlibatan yang lebih konsisten, sedangkan minat rendah dapat berkaitan dengan kebosanan, kurangnya motivasi, atau strategi pembelajaran yang kurang sesuai (Putu et al., 2022). Studi deskriptif pada konteks sekolah menengah juga menunjukkan bahwa metode pembelajaran, kondisi lingkungan belajar, motivasi, serta dukungan sarana dapat berperan terhadap kesulitan atau rendahnya keterlibatan siswa (Putri Eka Adek, 2024; Alfiko, 2025). Namun, penilaian minat yang hanya mengandalkan observasi guru berpotensi menjadi subjektif karena keputusan sangat bergantung pada interpretasi individu dan tidak selalu memanfaatkan seluruh data akademik yang tersedia.

Penelitian prediktif dalam pendidikan menunjukkan bahwa data akademik dapat diolah menggunakan machine learning untuk menghasilkan identifikasi yang lebih sistematis. Ekawati et al. (2025) menunjukkan bahwa integrasi data akademik dengan informasi pendukung dapat digunakan untuk memprediksi ketuntasan belajar, sedangkan Florent et al. (2025) memperlihatkan bahwa data akademik dan perilaku dapat dimodelkan untuk memperkirakan kinerja siswa SMP. Wibowo dan Aryanti (2025) juga menunjukkan bahwa data prestasi siswa dapat dikelompokkan secara komputasional untuk mendukung klasifikasi kondisi akademik. Meskipun demikian, sebagian penelitian pendidikan masih berfokus pada prediksi hasil belajar atau pengelompokan prestasi dan belum secara khusus menggabungkan model klasifikasi berkinerja tinggi dengan mekanisme interpretasi yang menjelaskan alasan di balik keputusan model.

XGBoost dipilih karena merupakan algoritma ensemble berbasis gradient boosting yang membangun pohon keputusan secara bertahap sehingga kesalahan model sebelumnya dapat dikoreksi pada tahap berikutnya. Mekanisme regularisasi membuat algoritma ini efektif untuk menangani pola hubungan fitur yang kompleks (Wardhani & Akbar, 2022). Akan tetapi, akurasi model saja belum cukup untuk konteks pendidikan. Guru membutuhkan penjelasan mengenai faktor apa yang mendorong suatu siswa diklasifikasikan ke dalam kategori tertentu. Oleh karena itu, penelitian ini mengintegrasikan SHAP (SHapley Additive exPlanations) sebagai pendekatan Explainable Artificial Intelligence untuk mengukur kontribusi masing-masing fitur terhadap keluaran model. Penerapan XGBoost dan SHAP pada penelitian lain telah memperlihatkan kombinasi antara performa prediksi dan interpretabilitas, baik pada analisis risiko akademik mahasiswa maupun bidang nonpendidikan (Istiwana Putri Agnes, 2026; Safii et al., 2025).

Kesenjangan yang ditangani penelitian ini terletak pada kebutuhan klasifikasi minat siswa SMP pada mata pelajaran Informatika yang tidak hanya akurat, tetapi juga dapat dijelaskan melalui kontribusi fitur akademik dan nonakademik. Kebaruan penelitian diarahkan pada integrasi tujuh indikator siswa ke dalam model XGBoost dan penggunaan SHAP untuk menampilkan faktor dominan secara transparan melalui sistem berbasis web.

LANDASAN TEORI

Minat Belajar dan Data Akademik

Minat belajar mengacu pada dorongan yang membuat siswa merasa tertarik, memberi perhatian, dan terlibat secara konsisten dalam kegiatan pembelajaran. Dalam praktik pendidikan, minat tidak berdiri sendiri karena dipengaruhi oleh faktor internal dan eksternal. Putu et al. (2022) menempatkan ketertarikan, perhatian, kesadaran, dan konsentrasi sebagai aspek yang dapat digunakan untuk membaca minat belajar. Pada penelitian ini, konsep tersebut tidak diukur melalui kuesioner psikologis tersendiri, tetapi didekati melalui pola data akademik dan nonakademik yang telah tersedia di sekolah, yaitu nilai tugas, ulangan harian, praktik, UTS, UAS, kehadiran, serta sikap.

Pemanfaatan data akademik memungkinkan penilaian dilakukan secara lebih terukur. Marcydiaz et al. (2024) menegaskan bahwa data nilai akademik dapat dirancang menjadi sumber informasi terstruktur untuk mendukung analisis sekolah. Data semacam ini juga dapat digunakan sebagai indikator kondisi belajar ketika terdapat pola nilai yang rendah atau tidak konsisten (Wibowo & Aryanti, 2025). Dalam konteks penelitian ini, ketujuh fitur tidak diperlakukan sebagai ukuran minat secara terpisah, melainkan sebagai sinyal yang secara bersama-sama dipelajari oleh model klasifikasi. Pendekatan ini penting karena satu indikator, misalnya kehadiran, belum tentu cukup untuk menjelaskan minat apabila tidak dibaca bersama capaian tugas, ulangan, dan aktivitas praktik.

Machine Learning, XGBoost, dan SHAP

Machine learning digunakan untuk menemukan pola dari data dan menghasilkan fungsi prediksi tanpa harus menetapkan seluruh aturan keputusan secara manual. Dalam pengembangan analisis berbasis Python, dukungan pustaka pengolahan data dan machine learning memudahkan proses pembersihan data, pemodelan, serta evaluasi secara terintegrasi (Alfarizi et al., 2023). Penggunaan Python juga relevan untuk penelitian berbasis data karena ekosistemnya mendukung pengolahan numerik, manipulasi tabel data, pembangunan model, dan visualisasi hasil (Kasliono et al., 2023).

XGBoost merupakan metode gradient boosting yang menggabungkan sejumlah decision tree secara berurutan. Setiap pohon berupaya memperbaiki kesalahan residual dari struktur sebelumnya, sedangkan regularisasi digunakan untuk mengendalikan kompleksitas model. Karakteristik tersebut membuat XGBoost mampu memodelkan hubungan nonlinier dan interaksi antarfitur pada data tabular (Wardhani & Akbar, 2022). Pada data pendidikan yang melibatkan beberapa komponen nilai, XGBoost relevan karena pola minat tidak harus mengikuti hubungan linear sederhana antara satu nilai dan kelas target.

Keterbatasan utama model ensemble adalah rendahnya keterbacaan mekanisme keputusan apabila hanya dilihat dari output prediksi. SHAP menjawab masalah tersebut dengan menghitung kontribusi setiap fitur berdasarkan konsep nilai Shapley. Nilai positif atau negatif menunjukkan arah kontribusi fitur terhadap prediksi, sedangkan besar absolutnya menunjukkan kekuatan pengaruh.

Dalam konteks pendidikan, interpretasi ini memungkinkan hasil model diterjemahkan menjadi informasi yang dapat dipahami guru. Istiwana Putri Agnes (2026) menggunakan pendekatan explainable machine learning untuk mengidentifikasi faktor dominan risiko akademik, sedangkan Safii et al. (2025) menunjukkan bahwa integrasi XGBoost dan SHAP dapat mempertahankan performa prediksi sekaligus menyediakan interpretasi kontribusi variabel.

CRISP-DM dan Evaluasi Model

Alur penelitian mengikuti CRISP-DM karena kerangka ini menyediakan tahapan sistematis mulai dari pemahaman masalah, pemahaman data, persiapan data, pemodelan, evaluasi, hingga penyajian hasil. Struktur tersebut sesuai dengan penelitian data science yang menuntut keterkaitan antara kebutuhan pengguna dan proses teknis pengolahan data (Massahiro et al., 2024).

Penggunaan CRISP-DM juga mencegah pemodelan dilakukan secara terpisah dari konteks masalah pendidikan karena tujuan analisis ditetapkan sejak tahap awal dan diperiksa kembali saat evaluasi.

Kinerja klasifikasi dianalisis menggunakan confusion matrix, Accuracy, Precision, Recall, F1-Score, serta ROC-AUC. Confusion matrix menunjukkan distribusi prediksi benar dan salah pada masing-masing kelas, sementara metrik turunan memberikan gambaran ketepatan klasifikasi dari sudut yang berbeda. Dalam penelitian dengan dua kelas Minat dan Tidak Minat, ROC-AUC digunakan untuk menilai kemampuan model membedakan kedua kategori pada berbagai threshold. Evaluasi tidak hanya menjadi ukuran keberhasilan teknis, tetapi juga dasar untuk menilai apakah model layak digunakan sebagai alat bantu analisis pada data yang memiliki karakteristik serupa.

Penelitian Terdahulu dan Posisi Penelitian

Penelitian terdahulu memperlihatkan dua kecenderungan. Pertama, kajian minat dan kesulitan belajar pada sekolah menengah banyak menggunakan pendekatan deskriptif untuk mengidentifikasi faktor motivasi, metode pembelajaran, dan lingkungan belajar (Putri Eka Adek, 2024; Alfiko, 2025). Pendekatan tersebut penting untuk memahami konteks, tetapi belum menghasilkan model otomatis yang dapat memproses data akademik secara konsisten. Kedua, penelitian machine learning mulai memanfaatkan data akademik untuk memprediksi ketuntasan atau kinerja siswa. Ekawati et al. (2025) menggunakan Naive Bayes dalam prediksi ketuntasan belajar, sedangkan Florent et al. (2025) menggunakan data akademik dan perilaku untuk memodelkan kinerja siswa SMP.

Posisi penelitian ini berada pada irisan kedua pendekatan tersebut. Model tidak hanya diarahkan untuk menghasilkan klasifikasi, tetapi juga menerjemahkan faktor akademik yang paling dominan melalui SHAP. Dengan demikian, keluaran penelitian mencakup dua lapisan informasi: kelas prediksi siswa dan penjelasan kontribusi fitur. Kombinasi ini membedakan penelitian dari pendekatan yang hanya menghasilkan label tanpa interpretasi, sekaligus mempertahankan konteks pendidikan melalui penggunaan variabel yang berasal dari aktivitas pembelajaran sehari-hari.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis data analytics. Subjek penelitian adalah siswa kelas VII dan VIII yang mengikuti mata pelajaran Informatika di SMP Negeri 30 Bengkulu Selatan. Data penelitian berasal dari dokumen akademik sekolah dan informasi pendukung hasil observasi serta wawancara guru. Dataset pemodelan terdiri atas 27 data siswa dengan tujuh fitur masukan. Variabel target pada implementasi model terdiri atas dua kelas, yaitu Minat dan Tidak Minat. Penggunaan target biner mengikuti struktur dataset dan implementasi pada Bab IV, sehingga evaluasi model dilakukan sebagai klasifikasi dua kelas.

Pengumpulan data dilakukan melalui observasi proses pembelajaran, wawancara terstruktur dengan guru Informatika, dokumentasi data akademik, dan studi literatur. Observasi digunakan untuk memahami aktivitas siswa, interaksi guru-siswa, partisipasi diskusi dan praktik, serta kondisi kelas. Wawancara menggali pemahaman siswa terhadap materi, kendala belajar, kehadiran, kedisiplinan, minat dan motivasi, serta strategi pembelajaran. Dokumentasi menjadi sumber utama fitur numerik karena memuat komponen nilai dan data pendukung yang digunakan model. Studi literatur digunakan untuk membangun landasan konseptual, memilih algoritma, dan menentukan teknik evaluasi.

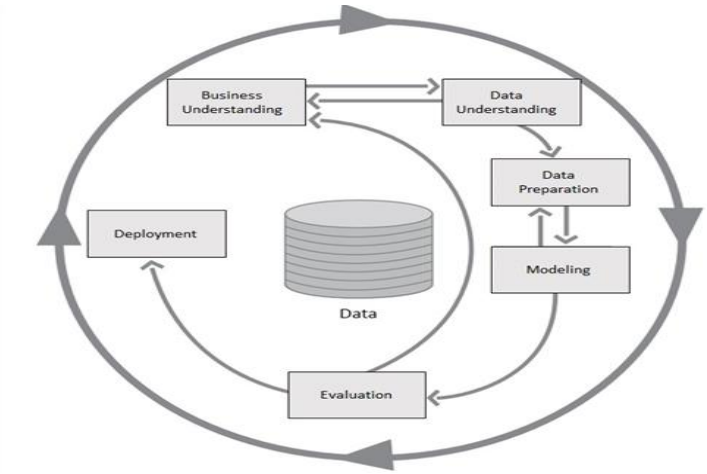
Proses analisis mengikuti tahapan CRISP-DM. Pada business understanding ditetapkan kebutuhan identifikasi minat siswa secara objektif. Data understanding digunakan untuk memeriksa struktur, distribusi, dan kelengkapan data. Pada data preparation dilakukan pembersihan, penanganan data tidak lengkap, transformasi format bila diperlukan, serta penetapan fitur dan label. Data kemudian dibagi menggunakan train-test split sebesar 80% untuk data latih dan 20%

untuk data uji. Dengan 27 data, pembagian ini menghasilkan 21 data untuk proses pelatihan dan 6 data untuk pengujian.

Pemodelan dilakukan menggunakan XGBoost dengan parameter n_estimators=100, max_depth=4, learning_rate=0,1, dan random_state=42. Parameter tersebut digunakan sebagaimana implementasi pada sistem penelitian. Setelah model dilatih, performa dievaluasi menggunakan confusion matrix, Accuracy, Precision, Recall, F1-Score, ROC Curve, dan AUC. Tahap interpretasi menggunakan SHAP untuk menghitung kontribusi fitur terhadap keputusan klasifikasi. Seluruh proses diintegrasikan ke dalam aplikasi berbasis web menggunakan Python sehingga pengguna dapat mengunggah dataset, membangun model, melihat evaluasi, menampilkan hasil SHAP, dan memprediksi data baru secara terstruktur.

Tabel 1. Variabel yang digunakan dalam pemodelan

No.	Variabel	Peran dalam Model	Sumber Data
1	Nilai Tugas	Fitur akademik	Dokumentasi sekolah
2	Ulangan Harian	Fitur akademik	Dokumentasi sekolah
3	Nilai Praktik	Fitur akademik	Dokumentasi sekolah
4	Nilai UTS	Fitur akademik	Dokumentasi sekolah
5	Nilai UAS	Fitur akademik	Dokumentasi sekolah
6	Kehadiran	Fitur nonakademik	Dokumentasi/observasi
7	Sikap	Fitur nonakademik	Dokumentasi/observasi
8	Minat/Tidak Minat	Target klasifikasi	Label dataset



Gambar 1. Tahapan CRISP-DM yang digunakan dalam penelitian (Massahiro et al., 2024)

Tabel 2. Konfigurasi utama model XGBoost

Parameter	Nilai
n_estimators	100
max_depth	4
learning_rate	0,1
random_state	42
Train-test split	80% : 20%

HASIL DAN PEMBAHASAN

Implementasi Sistem dan Dataset

Sistem klasifikasi dibangun sebagai aplikasi berbasis web yang mengintegrasikan proses unggah dataset, preprocessing, pelatihan XGBoost, evaluasi model, interpretasi SHAP, dan prediksi data baru. Dataset training yang digunakan berjumlah 27 siswa dan memuat tujuh atribut masukan. Setelah dataset berhasil dimuat, sistem menampilkan ringkasan jumlah data, fitur, statistik dasar, serta pembagian training dan testing. Implementasi ini mengubah proses identifikasi yang sebelumnya bergantung pada pembacaan nilai secara manual menjadi alur analisis yang terdokumentasi dan dapat dijalankan ulang dengan struktur data yang sama.

Pembagian 80:20 menghasilkan enam data uji yang tidak digunakan ketika model membangun pola dari data training. Pada data uji tersebut terdapat dua siswa pada kelas Minat dan empat siswa pada kelas Tidak Minat. Seluruh enam data berhasil diprediksi sesuai label aktual. Dengan demikian, confusion matrix tidak menghasilkan False Positive maupun False Negative. Hasil ini memperlihatkan separasi kelas yang sangat kuat pada sampel pengujian yang digunakan, tetapi besarnya performa harus dibaca bersama ukuran data uji yang kecil agar interpretasinya tetap proporsional.

Kinerja Klasifikasi XGBoost

Tabel 3. Hasil evaluasi model pada data testing

Metrik	Nilai	Interpretasi
Accuracy	100%	Seluruh data uji diklasifikasikan benar
Precision	100%	Prediksi kelas positif tidak menghasilkan false positive
Recall	100%	Seluruh data positif pada data uji teridentifikasi
F1-Score	100%	Keseimbangan precision dan recall sangat tinggi
AUC	1,00	Pemisahan kedua kelas pada data uji berada pada kategori sangat baik

Tabel 4. Confusion matrix pada enam data testing

Prediksi/Aktual	Minat	Tidak Minat
Minat	2	0
Tidak Minat	0	4

Nilai Accuracy, Precision, Recall, dan F1-Score yang mencapai 100% menunjukkan bahwa pada data testing penelitian, pola yang dipelajari XGBoost mampu memisahkan dua kelas secara tepat. Nilai AUC sebesar 1,00 menguatkan hasil tersebut karena probabilitas prediksi menghasilkan pemisahan kelas yang sangat baik pada threshold yang diuji. Secara algoritmik, performa tersebut dapat dikaitkan dengan kemampuan gradient boosting dalam memperbaiki kesalahan secara bertahap dan menangkap interaksi antarvariabel yang tidak harus linear (Wardhani & Akbar, 2022). Pada dataset ini, komponen nilai akademik menyediakan sinyal yang konsisten terhadap label target sehingga pohon-pohon keputusan dapat membangun batas klasifikasi dengan efektif.

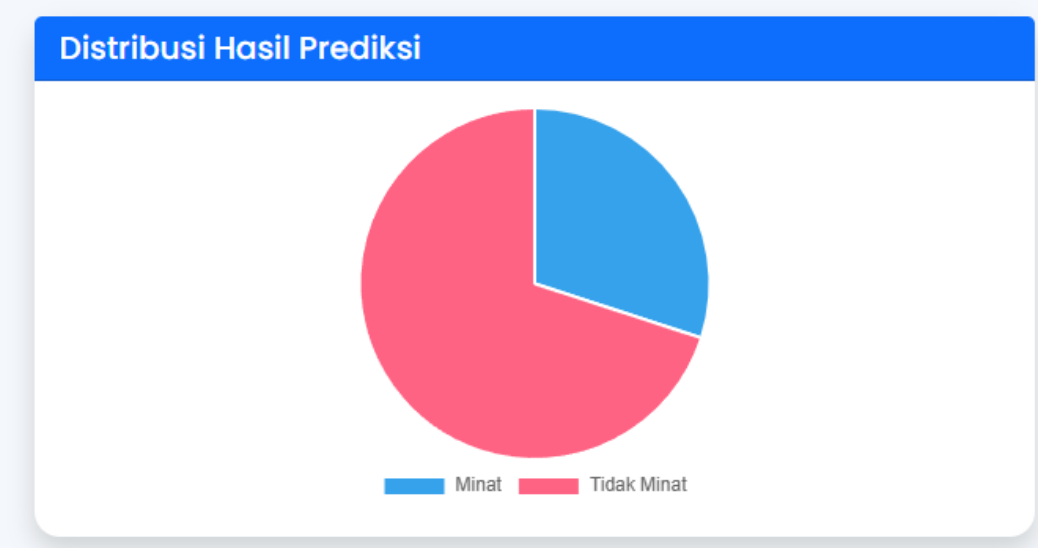
Temuan ini sejalan dengan kecenderungan penelitian machine learning pada pendidikan yang menunjukkan bahwa data akademik dapat digunakan untuk mengidentifikasi pola capaian siswa. Ekawati et al. (2025) menunjukkan kemampuan data akademik dan absensi dalam mendukung prediksi ketuntasan, sedangkan Florent et al. (2025) menemukan bahwa data akademik dan perilaku siswa dapat digunakan untuk memodelkan kinerja siswa SMP. Penelitian ini memperluas arah tersebut dengan memfokuskan target pada minat siswa dan menggunakan XGBoost sebagai model

klasifikasi. Namun, angka 100% pada enam data uji tidak dapat dipandang sebagai jaminan bahwa performa identik akan diperoleh pada populasi yang lebih luas. Ukuran sampel, distribusi kelas, dan karakteristik sekolah tetap menjadi faktor yang dapat memengaruhi generalisasi model.

Pada tahap implementasi prediksi data baru, sistem juga diuji untuk memproses 30 data testing terpisah sebagai demonstrasi penggunaan model. Hasil prediksi mengelompokkan 9 siswa ke kategori Minat dan 21 siswa ke kategori Tidak Minat. Hasil tersebut menunjukkan bahwa modul prediksi dapat bekerja pada dataset baru dan menyajikan distribusi kelas serta probabilitas prediksi. Data 30 siswa ini digunakan sebagai demonstrasi keluaran sistem, bukan sebagai dasar perhitungan metrik validasi utama yang berasal dari pembagian dataset training 27 siswa.

Tabel 5. Distribusi hasil prediksi pada 30 data baru

Keluaran Prediksi Data Baru	Jumlah	Persentase
Minat	9	30%
Tidak Minat	21	70%
Total	30	100%

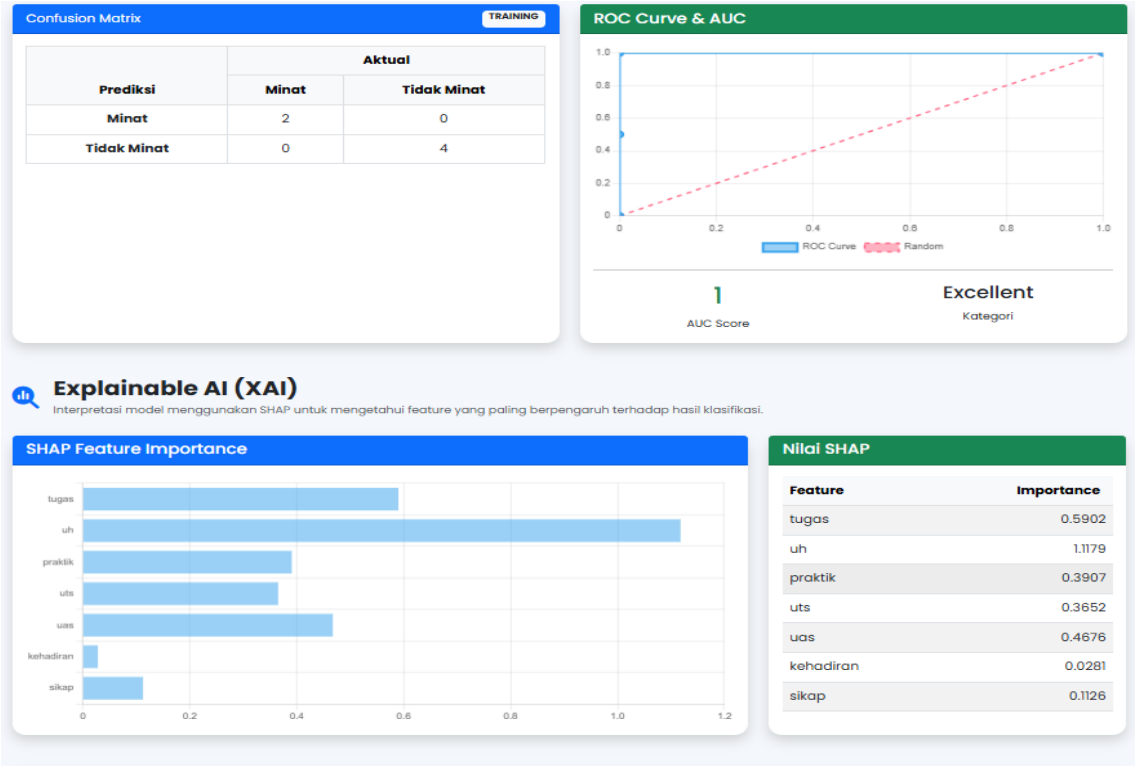


Gambar 2. Distribusi hasil prediksi pada 30 data baru

Interpretasi Faktor Menggunakan SHAP

Tabel 6. Urutan kontribusi fitur berdasarkan SHAP

Peringkat	Fitur	Nilai Kontribusi SHAP
1	Ulangan Harian	1,1179
2	Nilai Tugas	0,5902
3	Nilai UAS	0,4676
4	Nilai Praktik	0,3907
5	Nilai UTS	0,3652
6	Sikap	0,1126
7	Kehadiran	0,0281



Gambar 3. Ringkasan evaluasi XGBoost dan interpretasi SHAP pada sistem penelitian

Analisis SHAP menunjukkan bahwa Ulangan Harian merupakan fitur paling dominan dengan nilai kontribusi 1,1179. Nilai Tugas berada pada posisi kedua sebesar 0,5902, kemudian UAS sebesar 0,4676, Praktik sebesar 0,3907, dan UTS sebesar 0,3652. Dua variabel nonakademik, yaitu Sikap dan Kehadiran, memiliki kontribusi lebih rendah masing-masing 0,1126 dan 0,0281. Pola tersebut menunjukkan bahwa pada dataset penelitian, keputusan model lebih banyak dibentuk oleh performa akademik yang berulang selama proses belajar daripada indikator kehadiran atau penilaian sikap semata.

Dominannya Ulangan Harian dapat dijelaskan karena komponen tersebut merekam respons siswa terhadap materi secara lebih sering dibandingkan ujian semester. Ulangan harian berada dekat dengan ritme pembelajaran dan dapat mencerminkan kontinuitas pemahaman, kesiapan, serta keterlibatan siswa pada materi yang baru dipelajari. Nilai tugas juga memiliki kontribusi besar karena tugas menuntut keterlibatan siswa di luar momen evaluasi formal. Sementara itu, UAS, praktik, dan UTS tetap berperan karena ketiganya merepresentasikan capaian pada bentuk evaluasi yang berbeda. Dengan demikian, model tidak hanya bertumpu pada satu jenis penilaian, tetapi membangun keputusan dari kombinasi beberapa bukti akademik.

Temuan SHAP memberi nilai tambah dibandingkan model klasifikasi yang hanya menyajikan label. Pada konteks pendidikan, guru perlu mengetahui alasan yang dapat ditindaklanjuti. Ketika Ulangan Harian dan Nilai Tugas menjadi faktor utama, intervensi pembelajaran dapat diarahkan pada pemantauan pemahaman berkala, pemberian umpan balik tugas, dan pendampingan lebih cepat sebelum siswa memasuki evaluasi semester. Pendekatan ini konsisten dengan tujuan explainable machine learning, yaitu mengurangi sifat black-box dengan menjelaskan kontribusi variabel terhadap keluaran model (Istiwana Putri Agnes, 2026). Hasil penelitian Safii et al. (2025) juga menunjukkan bahwa SHAP dapat mendampingi XGBoost untuk menjelaskan variabel dominan tanpa menghilangkan fokus pada performa prediksi.

Kontribusi Kehadiran yang relatif kecil tidak berarti kehadiran tidak penting bagi proses pembelajaran. Nilai SHAP hanya menunjukkan besarnya pengaruh fitur terhadap keputusan model pada dataset yang digunakan. Dalam dataset ini, variasi nilai akademik tampak memberikan informasi pemisah yang lebih kuat daripada variasi kehadiran. Demikian pula, nilai Sikap tetap

berkontribusi tetapi tidak sedominan komponen hasil belajar. Hal tersebut memperlihatkan pentingnya membedakan interpretasi model dengan hubungan kausal. SHAP menjelaskan bagaimana model menggunakan fitur untuk menghasilkan prediksi, bukan membuktikan bahwa perubahan satu fitur secara langsung menyebabkan perubahan minat siswa.

Secara logis, kebaruan penelitian terletak pada penyatuan proses klasifikasi dan interpretasi dalam satu sistem. Kajian terdahulu pada sekolah menengah banyak menekankan faktor minat melalui deskripsi kondisi belajar (Putri Eka Adek, 2024; Alfiko, 2025), sedangkan penelitian prediktif lebih banyak berfokus pada ketuntasan atau performa akademik (Ekawati et al., 2025; Florent et al., 2025). Penelitian ini menempatkan data akademik sebagai basis klasifikasi minat dan menambahkan lapisan SHAP agar faktor dominan dapat ditelusuri. Keluaran tersebut lebih sesuai untuk penggunaan sebagai alat bantu guru karena keputusan model dapat disertai alasan yang terlihat pada tingkat fitur.

Implikasi, Validitas, dan Keterbatasan

Dari sisi praktis, sistem dapat digunakan sebagai alat bantu penyaringan awal untuk mengidentifikasi siswa yang memerlukan perhatian lebih pada pembelajaran Informatika. Penggunaan sistem tidak dimaksudkan menggantikan penilaian profesional guru, melainkan melengkapi observasi dengan pola yang ditarik dari data. Informasi SHAP memungkinkan guru memeriksa apakah klasifikasi tertentu terutama dipengaruhi oleh ulangan, tugas, praktik, atau komponen lain, sehingga tindak lanjut dapat disesuaikan dengan kebutuhan siswa.

Keterbatasan utama penelitian adalah ukuran dataset pemodelan yang hanya 27 siswa dari satu sekolah dan satu konteks mata pelajaran. Data testing hasil pembagian 20% hanya berjumlah enam kasus, sehingga performa sempurna pada pengujian perlu ditafsirkan secara hati-hati. Validasi pada data yang lebih besar, beberapa sekolah, dan beberapa tahun ajaran diperlukan sebelum model dianggap memiliki generalisasi luas. Selain itu, fitur yang digunakan terutama berasal dari capaian akademik dan dua indikator nonakademik. Faktor lain seperti motivasi intrinsik, dukungan orang tua, penggunaan teknologi pembelajaran, aktivitas ekstrakurikuler, dan lingkungan belajar belum dimasukkan, padahal dapat memperkaya penjelasan minat siswa.

KESIMPULAN DAN SARAN

Kesimpulan

Penerapan XGBoost pada data siswa SMP Negeri 30 Bengkulu Selatan berhasil menghasilkan model klasifikasi minat pada mata pelajaran Teknik Informatika dengan memanfaatkan tujuh fitur berupa nilai tugas, ulangan harian, praktik, UTS, UAS, kehadiran, dan sikap. Pada enam data testing, model mengklasifikasikan dua siswa kategori Minat dan empat siswa kategori Tidak Minat tanpa kesalahan, sehingga Accuracy, Precision, Recall, dan F1-Score masing-masing mencapai 100% dengan AUC 1,00. Interpretasi SHAP menunjukkan bahwa Ulangan Harian merupakan faktor yang paling dominan dengan nilai kontribusi 1,1179, diikuti Nilai Tugas, UAS, Praktik, dan UTS, sedangkan Sikap dan Kehadiran memiliki kontribusi lebih kecil. Hasil tersebut menjawab tujuan penelitian bahwa kombinasi XGBoost dan SHAP dapat digunakan untuk menghasilkan klasifikasi yang terukur sekaligus penjelasan fitur yang transparan. Meskipun demikian, capaian performa harus dibaca dalam batas dataset yang kecil dan bersumber dari satu sekolah, sehingga temuan belum dapat digeneralisasikan tanpa pengujian tambahan.

Saran

Pengembangan berikutnya disarankan menggunakan jumlah data yang lebih besar, mencakup beberapa tahun ajaran dan sekolah berbeda, serta menerapkan validasi silang atau pengujian eksternal agar kestabilan performa model dapat dinilai lebih kuat. Variabel dapat diperluas dengan memasukkan motivasi belajar, dukungan orang tua, lingkungan belajar, aktivitas ekstrakurikuler, dan penggunaan teknologi pembelajaran. Pada sisi implementasi, hasil klasifikasi sebaiknya tetap

diposisikan sebagai alat bantu keputusan yang dikombinasikan dengan observasi guru dan informasi kontekstual siswa. Pengembangan antarmuka selanjutnya juga dapat menyediakan penjelasan SHAP pada tingkat individu agar guru memperoleh alasan spesifik untuk setiap prediksi dan dapat merancang intervensi pembelajaran yang lebih tepat sasaran.

DAFTAR PUSTAKA

Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python sebagai bahasa pemrograman untuk machine learning dan deep learning. Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID), 2(1), 1–6.

Alfiko, F. (2025). Analisis kesulitan belajar siswa pada mata pelajaran Informatika kelas X Perhotelan 2 SMK N 7 Kota Bengkulu. Computer and Informatics Education Review-CIER, 2025(1), 49–56.

Ekawati, R. R., Atina, V., & Maulidar, J. (2025). Prediksi ketuntasan belajar siswa menggunakan Naive Bayes dengan integrasi data akademik. Jurnal Pendidikan dan Teknologi Indonesia (JPTI), 5(4), 1161–1173.

Florent, A., Rizal, M., & Oka, S. (2025). Prediksi kinerja siswa SMP berdasarkan data akademik dan perilaku menggunakan machine learning. Teknomatika: Jurnal Informatika dan Komputer, 18(2), 67–76.

Istiwana Putri Agnes. (2026). Pendekatan explainable machine learning untuk analisis faktor drop out mahasiswa menggunakan XGBoost. Machine Learning and Knowledge Extraction, 5(1), 169–170. https://doi.org/10.3390/make5010010

Kasliono, K., Suharmono, E., Povi, P., Meriani, R., & Candraningrum, N. (2023). Analisis regresi dan korelasi untuk proyeksi produksi minyak bumi dan gas alam Indonesia menggunakan bahasa pemrograman Python. Jurnal Teknologi Informatika dan Komputer, 9(2), 1297–1313. https://doi.org/10.37012/jtik.v9i2.1756

Marcydiaz, A. H., Fitriya, F. N., Hutagaol, A. S., Trisnawarman, D., & Beng, J. T. (2024). Perancangan datamart nilai akademik siswa pada SMA Z Bekasi. INTECOMS: Journal of Information Technology and Computer Science, 7(6), 2040–2047. https://doi.org/10.31539/intecoms.v7i6.12916

Massahiro, A., Instituto, S., Tecnológicas, D. P., Paulo, D. S., Univer-, F., Paulo, S., Cordeiro, R., & Paulo, S. (2024). The evolution of CRISP-DM for data science: Methods, processes and frameworks. https://doi.org/10.5753/reviews.2024.3757

Pangesthi, G., & Triyani. (2022). Buku interaktif informatika: Untuk SMP/MTs. Intan Pariwara.

Putri Eka Adek. (2024). Analisis faktor-faktor kesulitan belajar siswa pada mata pelajaran geografi materi bumi sebagai ruang kehidupan di kelas X SMA Negeri 4 Pariaman. Jurnal Pendidikan Tambusai, 8(1), 11724–11730. https://doi.org/10.31004/jptam.v8i1.14151

Putu, N., Yuniarti, M., Ayu, G., & Sukma, P. (2022). Efektivitas media pembelajaran terhadap minat belajar siswa Videoscribe Connected. 5(1), 160–171.

Safii, M., Husain, Kirana, I. O., Leana, S. A., & Gultom, Y. I. (2025). Model prediksi penjadwalan produksi energi terbarukan dengan algoritma XGBoost dan analisis interpretatif menggunakan SHAP. Jurnal Sistem Informasi Triguna Dharma (JURSI TGD), 4(4), 794–801. https://doi.org/10.53513/jursi.v4i4.11443

Wardhani, K. D. K., & Akbar, M. (2022). Diabetes risk prediction using Extreme Gradient Boosting (XGBoost). Jurnal Online Informatika, 7(2), 244–250. https://doi.org/10.15575/join.v7i2.970

Wibowo, E. A., & Aryanti, R. (2025). Penerapan metode clustering K-Means menggunakan RapidMiner untuk klasifikasi prestasi siswa di sekolah swasta. Journal of Information Technology and Informatics Engineering, 1(1), 20–24.