

Klasifikasi Gaya Belajar Siswa dengan Random Forest dan XGBoost

Classification of Learning Styles of Junior High School Students Using Random Forest and XGBoost Algorithm

Christine Eirene IS*, Dian Syafitri Chani Saputri, Neny Sulistianingsih, Khasnur Hidjah, Hairani Hairani

Universitas Bumigora, Mataram, Indonesia

Informasi Artikel:

Diterima: 23 Februari 2025, Direvisi: 31 Mei 2025, Disetujui: 13 Juni 2025

Abstrak-

Latar Belakang: Mengetahui secara akurat gaya belajar siswa agar pendidik dapat menyesuaikan metode pembelajaran yang sesuai merupakan Tantangan dalam dunia pendidikan. Namun, belum adanya pemanfaatan *Machine learning* dalam klasifikasi gaya belajar pada sekolah di Kota Mataram merupakan permasalahan yang dapat diteliti.

Tujuan: Penelitian ini bertujuan untuk mengklasifikasikan gaya belajar siswa Sekolah Menengah Pertama (SMP) Negeri 2 Mataram menggunakan algoritma *Random Forest* dan *XGBoost* dalam bidang *Machine learning*.

Metode: Pengumpulan data dilakukan melalui kuesioner yang diisi oleh siswa kelas 7, 8, dan 9. Hasil eksplorasi data (EDA) menunjukkan ketidakseimbangan data pada kelas yang dikumpulkan.

Hasil: Hasil ini menunjukkan bahwa kedua algoritma mampu melakukan klasifikasi dengan baik, dengan *XGBoost* menunjukkan performa yang lebih unggul. Namun, akurasi yang diperoleh belum optimal, kemungkinan disebabkan oleh jumlah *dataset* yang terbatas. Untuk mengatasi ketidakseimbangan data, dilakukan teknik SMOTE. Evaluasi awal menunjukkan bahwa kedua algoritma, *XGBoost* dan *Random Forest*, memiliki akurasi yang sama sebesar 80%. Setelah dilakukan *hyperparameter tuning*, akurasi *XGBoost* meningkat menjadi 84%, sedangkan *Random Forest* mencapai 82%.

Kesimpulan: Penelitian ini memberikan kontribusi dalam pemanfaatan *Machine learning* untuk mengidentifikasi gaya belajar siswa, meskipun diperlukan pengembangan lebih lanjut untuk meningkatkan akurasi model.

Kata Kunci: Gaya Belajar, Klasifikasi, *Random Forest*, *XGBoost*.

Abstract-

Background: Accurately identifying students' learning styles so that educators can adjust their teaching methods accordingly is a challenge in the field of education. However, the application of *Machine learning* for learning style classification has not yet been implemented in schools in Mataram City.

Objective: This study aims to classify the learning styles of students at Junior High School (SMP) Negeri 2 Mataram using *Random Forest* and *XGBoost* algorithms. Accurate identification of learning styles is essential for developing individuals who align with their educational environment.

Methods: Data were collected through questionnaires completed by students in grades 7, 8, and 9. Data exploration (EDA) results show data imbalance in the collected classes.

Result: These results indicate that both algorithms performed well in classifying learning styles, with *XGBoost* showing slightly better performance. However, the accuracy obtained is not yet optimal, likely due to the limited dataset size. The SMOTE technique was applied to address data imbalance. Initial evaluation showed that both *XGBoost* and *Random Forest* achieved an accuracy of 80%. After *hyperparameter tuning*, the accuracy of *XGBoost* increased to 84%, while *Random Forest* reached 82%.

Conclusion: The findings suggest that visual and kinesthetic are the most dominant learning styles among SMP Negeri 2 Mataram students. This study contributes to the application of *Machine learning* in the education sector and highlights the need for further research to enhance model performance.

Keywords: Classification, Learning Styles, *Random Forest*, *XGBoost*.

How to Cite: C. Eirene IS, D. S. C. Saputri, N. Sulistianingsih, K. Hidjah, & H. Hairani, "Klasifikasi Gaya Belajar Siswa dengan Random Forest dan XGBoost," *Jurnal Bumigora Information Technology (BITe)*, vol. 7, no. 1, pp. 27-36, Jun 2025. doi: 10.30812/bite.v7i1.4913.

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

Penulis Korespondensi:

Christine Eirene IS,
Universitas Bumigora, Mataram, Indonesia,
Email: christineeirene@gmail.com

1. PENDAHULUAN

Salah satu faktor yang memengaruhi hasil belajar siswa adalah minat dan motivasi, yang berperan sebagai pendorong utama dalam meningkatkan prestasi akademik [1]. Motivasi yang tinggi dapat menciptakan lingkungan belajar yang kompetitif dan mendorong siswa mencapai potensinya. Dalam upaya meningkatkan kualitas pendidikan, pemanfaatan teknologi menggunakan *machine learning*, menjadi solusi inovatif untuk memahami kebutuhan belajar siswa secara lebih akurat [2]. Gaya belajar, yang mencakup visual, auditori dan kinestetik yang berperan penting dalam proses belajar mengajar, sehingga pemahaman terhadap preferensi ini memungkinkan tenaga pendidik merancang strategi pembelajaran yang lebih efektif [3], [4]. Beberapa penelitian sebelumnya telah menerapkan algoritma *Decision Tree* dan *Naïve Bayes* dalam klasifikasi gaya belajar dengan akurasi tinggi, namun penelitian menggunakan *Random Forest* dan *XGBoost* secara bersamaan dalam klasifikasi gaya belajar masih sangat terbatas [4], [5]. Mengingat sekolah-sekolah negeri di Kota Mataram belum menerapkan pendekatan ini, penelitian ini berfokus pada klasifikasi gaya belajar siswa SMP Negeri 2 Mataram dengan algoritma *Random Forest* dan *XGBoost* untuk memberikan kontribusi terhadap penelitian klasifikasi di bidang pendidikan serta membantu tenaga pendidik dalam menyusun metode pembelajaran yang lebih optimal berdasarkan hasil analisis berbasis *machine learning*.

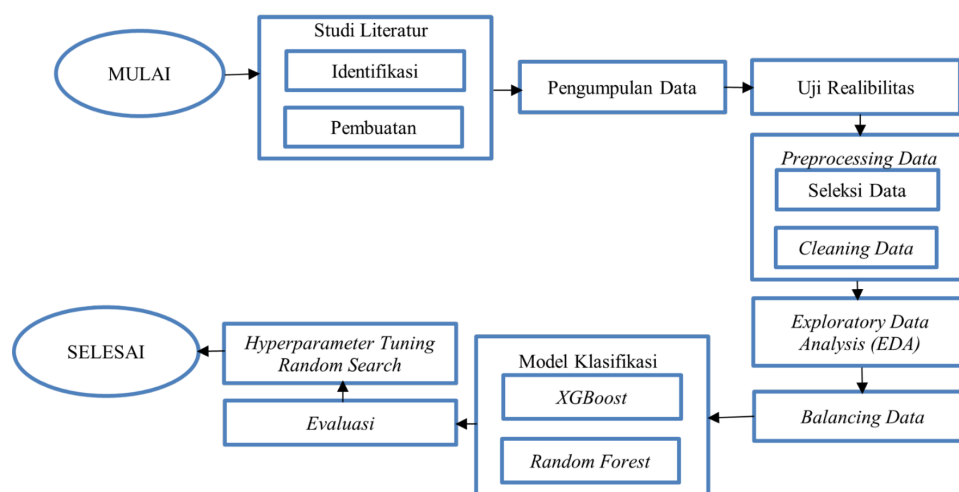
Beberapa penelitian terdahulu telah mengkaji klasifikasi gaya belajar siswa menggunakan berbagai algoritma *machine learning*. Referensi [3] menunjukkan bahwa penggunaan algoritma *Decision Tree* dalam *RapidMiner* memperoleh akurasi sebesar 99,78% dalam mengklasifikasikan gaya belajar siswa. Referensi [6] menunjukkan bahwa penelitian ini membandingkan algoritma *Decision Tree C4.5* dan *Naïve Bayes*, dengan hasil akurasi masing-masing 96% dan 98%, menunjukkan bahwa kedua algoritma ini efektif dalam klasifikasi gaya belajar. Referensi [7] menunjukkan penerapan algoritma *Naïve Bayes* untuk membantu dosen mengidentifikasi gaya belajar mahasiswa berdasarkan data sebelumnya, dengan akurasi mencapai 90%. Referensi [5] menunjukkan penggunaan *dataset* kuesioner dan algoritma *Decision Tree C4.5* menghasilkan akurasi sebesar 80%. Meskipun algoritma *Decision Tree* dan *Naïve Bayes* telah banyak digunakan dalam penelitian sebelumnya, penelitian yang memanfaatkan algoritma *Random Forest* dan *XGBoost* untuk klasifikasi gaya belajar **masih terbatas**, khususnya di lingkungan sekolah negeri. Oleh karena itu, penelitian ini **bertujuan** untuk mengisi *gap research* tersebut dengan menerapkan algoritma *Random Forest* dan *XGBoost* untuk meningkatkan akurasi klasifikasi gaya belajar siswa.

Meskipun berbagai penelitian telah dilakukan dalam klasifikasi gaya belajar siswa menggunakan algoritma *Decision Tree* dan *Naïve Bayes*, penelitian yang secara khusus menggabungkan algoritma *Random Forest* dan *XGBoost* dalam studi kasus klasifikasi gaya belajar belum ditemukan pada lingkungan sekolah negeri seperti di Kota Mataram. **Perbedaan penelitian** ini dengan sebelumnya adalah penggunaan kombinasi algoritma *Random Forest* dan *XGBoost* yang belum banyak di eksplorasi dalam konteks sekolah negeri dengan jumlah siswa besar dan kondisi kelas nyata. Sebagian besar penelitian terdahulu menunjukkan bahwa algoritma berbasis *Decision Tree* memiliki performa yang baik dalam melakukan proses klasifikasi gaya belajar dengan penanganan *dataset* yang tidak seimbang, tetapi belum banyak yang mengeksplorasi keunggulan *Random Forest* dalam mengurangi risiko *overfitting* serta *XGBoost* dalam meningkatkan akurasi melalui proses *boosting* dalam studi kasus gaya belajar siswa. Selain itu, penelitian sebelumnya cenderung menggunakan jumlah responden yang terbatas, sementara penelitian ini berupaya mengatasi keterbatasan tersebut dengan memilih lokasi penelitian di Sekolah Menengah Pertama (SMP) Negeri 2 Mataram yang memiliki jumlah siswa lebih besar dan tingkat

kognitif di atas rata-rata. Oleh karena itu, penelitian ini dapat **berkontribusi** dalam mengembangkan model klasifikasi gaya belajar siswa dengan menerapkan algoritma *Random Forest* dan *XGBoost* guna memperoleh hasil yang lebih akurat dan dapat digunakan sebagai dasar pengembangan strategi pembelajaran yang sesuai dengan preferensi siswa di lingkungan formal.

2. METODE PENELITIAN

Alur penelitian pada Gambar 1 diawali dengan tahap studi literatur untuk mengidentifikasi dan merumuskan permasalahan, kemudian dilanjutkan dengan pengumpulan data. Setelah data terkumpul, dilakukan uji reliabilitas untuk memastikan kualitas data. Tahap selanjutnya adalah praproses data, yang mencakup seleksi dan pembersihan data. Setelah itu, dilakukan analisis data eksploratif (EDA) untuk memahami pola data, diikuti dengan penyeimbangan data (balancing). Data yang telah siap kemudian digunakan dalam pemodelan klasifikasi menggunakan algoritma *XGBoost* dan *Random Forest*. Setelah model terbentuk, dilakukan penyetelan hiperparameter menggunakan metode *Random Search* guna mengoptimalkan kinerja model. Tahap akhir dari penelitian ini adalah evaluasi model sebelum penelitian dinyatakan selesai.



Gambar 1. Tahapan penelitian

2.1. Studi Literatur

Tahap ini memungkinkan peneliti mengidentifikasi kebutuhan, mendefinisikan variabel, dan menyusun kuesioner. Identifikasi masalah dilakukan untuk memahami kebutuhan penelitian, khususnya dalam mengklasifikasikan gaya belajar siswa SMP secara akurat menggunakan metode *machine learning*. Algoritma *XGBoost* dan *Random Forest* dipilih karena menangani berbagai kasus dengan *dataset* beragam. Berdasarkan penelitian sebelumnya, gaya belajar siswa dikategorikan menjadi tiga jenis utama yaitu visual, auditori, dan kinestetik.

Selanjutnya, kuesioner disusun berdasarkan indikator-indikator yang dirujuk dari penelitian [6], yang mencakup tiga dimensi utama gaya belajar yaitu visual (berfokus pada penglihatan seperti grafik dan gambar), auditori (berfokus pada pendengaran seperti penjelasan langsung), dan kinestetik (berfokus pada gerakan fisik atau aktivitas praktis). Kuesioner terdiri dari lima belas indikator, dengan setiap indikator berisi lima hingga tujuh pertanyaan. Penyusunan kuesioner ini bertujuan untuk mengukur preferensi gaya belajar siswa secara komprehensif, sehingga data yang diperoleh dapat digunakan untuk analisis lebih lanjut.

2.2. Pengumpulan Data

Pengumpulan data dilakukan di SMP Negeri 2 Mataram. Data yang diambil adalah data hasil dari kuesioner yang melibatkan sekitar 1.300 siswa yang berjumlah 844 data dan memiliki 95 atribut yang terdiri dari 90 atribut pertanyaan, 3 atribut data diri, 1 atribut *timestamp*, dan 1 atribut LABEL.

2.3. Uji Reliabilitas

Reliabilitas berasal dari istilah “*reability*” yang memiliki arti suatu kestabilan yang dapat dipercaya, sehingga dapat diartikan uji realibilitas merupakan uji suatu alat penelitian yang dapat dibuktikan kekonsistennya/keandalannya dalam melakukan fungsi pengukuran penelitian [8]. Pengujian reliabilitas dalam penelitian ini menggunakan metode *Cornbach's Alpha*. Untuk rumus pada *Cronbach's Alpha* akan dijelaskan pada Persamaan (1).

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum S_i^2}{S_t^2} \right) \quad (1)$$

Dengan, α adalah nilai koefisien reliabilitas (*Cronbach's Alpha*), digunakan untuk mengukur konsistensi internal suatu instrumen (apakah item-item dalam kuesioner mengukur hal yang sama). k adalah jumlah item atau butir pertanyaan dalam instrumen/kuesioner. S_i^2 adalah varians dari skor masing-masing item (i), dihitung satu per satu untuk tiap butir pertanyaan. $\sum S_i^2$ adalah Jumlah total dari varians semua item. S_t^2 adalah varians total dari skor keseluruhan (total skor responden terhadap semua item).

2.4. Preprocessing Data

Pada penelitian ini, *preprocessing* data dilakukan melalui dua tahap utama: seleksi data dan *cleaning* data. Pada tahap seleksi data, *dataset* mentah disaring dengan mengambil atribut-atribut yang relevan untuk penelitian dan menghapus kolom yang tidak diperlukan. *Dataset* awal terdiri dari 95 atribut, termasuk timestamp, data diri (nama lengkap, jenis kelamin, kelas), dan 91 kolom pertanyaan. Kolom *timestamp* dihapus karena tidak diperlukan dalam proses pemodelan. Tahap *cleaning* data dilakukan dengan menghapus data duplikat pada kolom nama lengkap dan menghapus nilai label yang kosong. Proses *cleaning* data dalam penelitian ini relatif sederhana karena *dataset* yang digunakan sudah dalam kondisi yang cukup baik sebelum dilakukan *preprocessing*.

2.5. Exploratory Data Analysis (EDA)

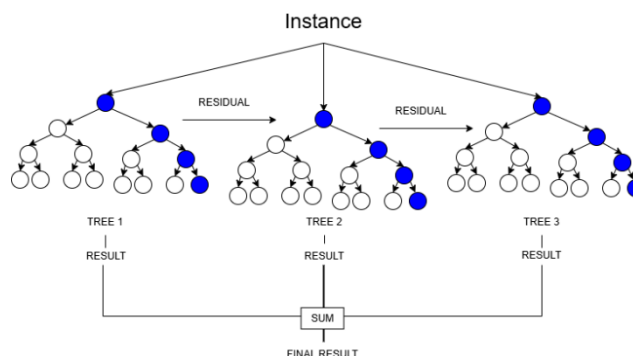
EDA membantu mengidentifikasi pola, hubungan, atau anomali dalam data, sehingga memastikan kualitas dan relevansi data untuk analisis lebih lanjut. Pada penelitian ini, EDA difokuskan pada tiga aspek utama: Pertama, distribusi kelas gaya belajar akan dianalisis untuk melihat sebaran responden berdasarkan tiga label yaitu visual, kinestetik, dan auditori. Analisis ini akan mengidentifikasi gaya belajar dominan di kalangan siswa SMP Negeri 2 Mataram serta mengevaluasi apakah data yang digunakan seimbang atau tidak. Kedua, distribusi responden berdasarkan kelas akan dianalisis untuk memastikan bahwa kuesioner telah diisi secara merata oleh siswa dari berbagai kelas. Hal ini penting untuk memvalidasi representasi data dan keberlanjutan proses sosialisasi kuesioner yang telah dilakukan sebelumnya. Ketiga, distribusi gaya belajar berdasarkan jenis kelamin akan dianalisis untuk melihat apakah terdapat perbedaan signifikan antara gaya belajar siswa laki-laki dan perempuan. Hasil analisis ini dapat menjadi acuan dalam merancang strategi pembelajaran yang lebih inklusif dan efektif.

2.6. Balancing Data

Pada penelitian ini teknik *balancing data* yang digunakan adalah *SMOTE* merupakan teknik *oversampling* yang dapat menghasilkan data tiruan yang baru dalam ruang fitur yang berada di antara data-data yang ada [9]. *SMOTE* digunakan dengan alasan *dataset* yang diperoleh pada proses pengumpulan data relatif kecil hanya sebanyak 844 data sehingga dilakukan *SMOTE* agar tetap bisa menghasilkan sampel sintesis untuk kelas minoritas tanpa mengurangi jumlah *dataset* yang telah ada. Teknik *SMOTE* nantinya akan bekerja dengan menentukan kelas yang minoritas kemudian pada tiap kelas minoritas akan dilakukan pembuatan sampel dengan cara pendekatan antar titik seperti *euclidean distance*, kemudian proses ini akan diulang hingga kelas minoritas menghasilkan sejumlah sampel tambahan yang sama dengan mayoritas.

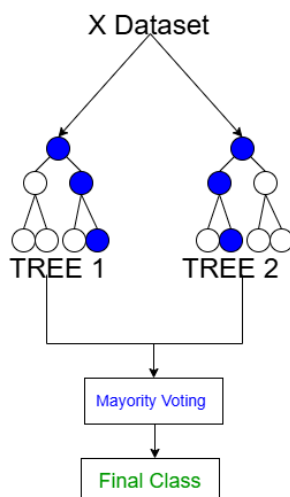
2.7. Model Klasifikasi

XGBoost merupakan salah satu dari algoritma berbasis pohon Keputusan yang berfokus pada peningkatan gradien yang sama dengan menggunakan jumlah peningkatan, kecepatan pembelajaran, rasio sub *sampling* dan kedalaman pohon maksimum sehingga dapat mengontrol *overfitting* dan meningkatkan kinerja yang lebih baik [10]. *XGBoost* menggunakan teknik *ensemble learning* yang menggabungkan hasil prediksi beberapa model pohon keputusan (*Decision Tree*) dan memanfaatkan penurunan gradien untuk mengoptimalkan proses pembelajaran sehingga membuat algoritma ini lebih efisien dibandingkan algoritma *machine learning* lainnya. Konsep kerja model *XGBoost* dalam pembuatan pohon keputusan dapat diilustrasikan seperti Gambar 2 [11].



Gambar 2. Ilustrasi pohon *XGBoost*

Referensi [12] menunjukkan bahwa *Random Forest* adalah salah satu algoritma *Machine learning* yang dipakai dengan memanfaatkan konsep pohon regresi dan klasifikasi sehingga dapat mencapai node akhir dalam struktur pohon melalui metode pemisahan biner rekursif. Referensi [13] menunjukkan pada algoritma *Random Forest* adanya gabungan dari 2 atau lebih pohon keputusan (*Decision Tree*) yang setiap pohon keputusan akan memiliki satu keluaran berupa label terhadap suatu sampel, kemudian untuk seluruh label yang telah dipilih oleh pohon keputusan akan dilihat jumlah terbanyak, label yang mendapat jumlah terbanyak akan menjadi keluaran akhir bagi suatu sample tersebut (lihat Gambar 3).



Gambar 3. Ilustrasi pohon *Random Forest*

2.8. Hyperparameter Tuning

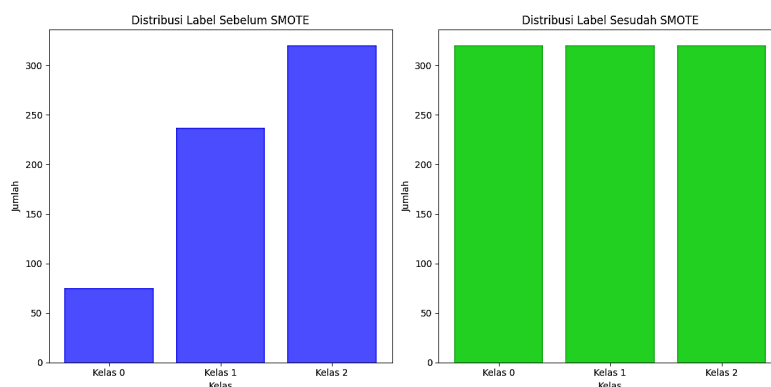
Hyperparameter adalah parameter yang ditetapkan sebelum proses pelatihan model dan berfungsi untuk mengatur struktur model. Namun, untuk beberapa kasus model *machine learning* *hyperparameter* digunakan

untuk meningkatkan sebuah akurasi apabila hasil dari sebuah perhitungan mendapat akurasi rendah yang disebut *hyperparameter tuning* [14].

3. HASIL DAN PEMBAHASAN

Pada bab ini disajikan hasil implementasi penelitian yang diawali dengan proses identifikasi masalah dan penyusunan kuesioner berdasarkan hasil studi literatur. Fokus penelitian telah dijelaskan pada Sub bab sebelumnya mengenai batasan-batasan yang ada salah satunya hanya berfokus pada klasifikasi gaya belajar siswa di SMP Negeri 2 Mataram menggunakan algoritma *XGBoost* dan *Random Forest*. Pengumpulan data dilakukan melalui penyebaran kuesioner via *Google form* yang dilakukan secara online melalui grup *WhatsApp*. Siswa diberikan waktu maksimal satu minggu untuk mengisi kuesioner dengan pemantauan harian melalui absensi di setiap kelas guna memastikan partisipasi. Proses pengambilan data menghasilkan data responden sebanyak 844 data atau sekitar 64% dari populasi di tempat penelitian. Tahapan penelitian yang harus dilakukan setelah berhasil melakukan pengumpulan data adalah melakukan proses uji reliabilitas menggunakan metode *Cronbach's Alpha*. Uji reliabilitas melibatkan data mentah sebanyak 844 data. Penelitian ini karena melibatkan ratusan data maka peneliti menggunakan *google colab* agar dapat efisien. Hasil uji reliabilitas sebesar 0,92 dengan kategori nilai reliabilitas tinggi sehingga telah memenuhi standar dan dapat melanjutkan pada tahapan penelitian selanjutnya.

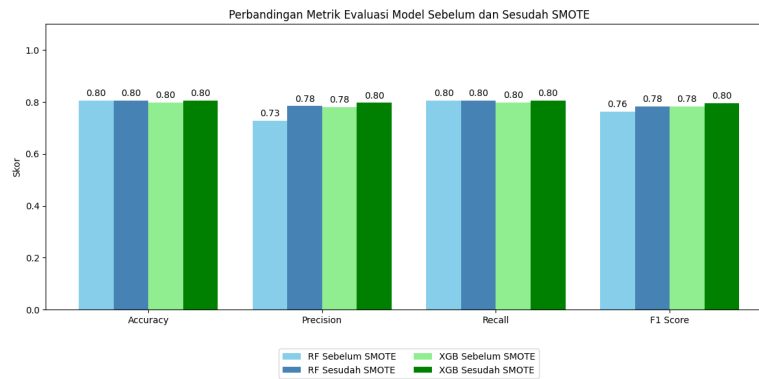
Tahapan penelitian selanjutnya adalah melakukan *preprocessing data* yang meliputi seleksi dan pembersihan data. Dari 844 data awal, 790 data digunakan setelah menghapus kolom yang tidak relevan (seperti *timestamp*) dan data duplikat. Proses EDA menunjukkan ketidakseimbangan distribusi kelas, dengan gaya belajar visual sebagai yang dominan, diikuti kinestetik dan auditori. Mayoritas responden berasal dari kelas 7, dan tidak terdapat perbedaan signifikan dalam gaya belajar antara siswa laki-laki dan perempuan. Untuk mengatasi ketidakseimbangan data, teknik SMOTE digunakan. Proses ini menghasilkan 960 data *training* setelah menyeimbangkan kelas minoritas. Lebih jelasnya dapat dilihat pada Gambar 4.



Gambar 4. Hasil *Balancing Data*

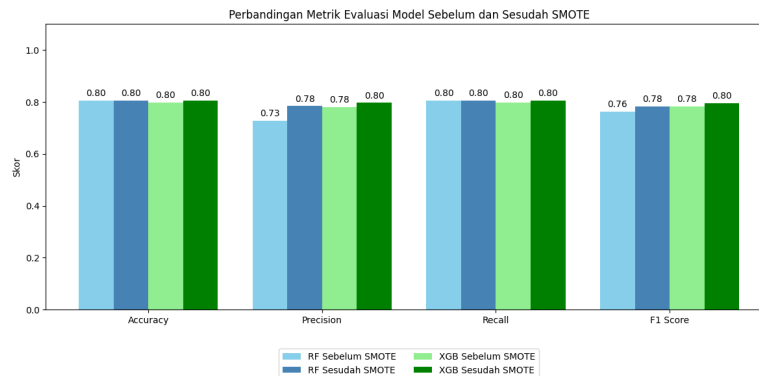
Proses *modelling* menggunakan dua algoritma yakni, *Random Forest* dan *XGBoost* dengan pembagian data *training* dan *testing* masing-masing 80% dan 20% dengan data yang telah di seimbangkan. Proses *SMOTE* hanya diterapkan pada data *training* untuk menghindari data *leakage* pada saat proses *testing* berlangsung. Pada tahapan *modelling* ini melibatkan sebanyak 960 data *training* dan 158 data *testing*. Hasil evaluasi awal menunjukkan bahwa kedua algoritma mendapat akurasi yang sama pada angka 80%, sehingga perlu dilakukan tahapan *hyperparameter tuning* untuk mencari kembali nilai akurasi terbaik kedua model. Grafik perbandingan *modelling* sebelum dan sesudah *SMOTE* dapat dilihat pada Gambar 5.

Langkah selanjutnya adalah melakukan proses *hyperparameter tuning* sebagai tindak lanjut dari hasil evaluasi model konvensional. Dalam penelitian ini, digunakan metode *Random Search* untuk melakukan *hyperparameter tuning*. Sebelumnya, telah dilakukan uji coba dengan metode *Grid Search*, yang memang umum digunakan, namun kedua model yang diuji tidak menunjukkan kecocokan dengan pendekatan tersebut. Pada hasil evaluasi *hyperparameter tuning* diperoleh nilai akurasi untuk algoritma *Random Forest* dan *XGBoost*



Gambar 5. Grafik Sebelum & Sesudah SMOTE

berturut-turut, 82% dan 84% untuk pertimbangan lain diperoleh juga nilai *f1-score* untuk algoritma *Random Forest* dan *XGBoost* berturut-turut yaitu 80% dan 83%. Proses ini memperoleh parameter terbaik untuk algoritma *Random Forest* yaitu: *max_depth* = 10, *min_samples_leaf* = 1, *min_sample_split* = 2, *n_estimators* = 100. Pada algoritma *XGBoost* yaitu *colsample_bytree* = 0.6, *learning_rate* = 0.2, *max_depth* = 5, *n_estimators* = 300, *subsample* = 0.6. Gambar 6 menunjukkan hasil *modelling* model konvensional sebelum dan setelah dilakukan *hyperparameter tuning*.

Gambar 6. Grafik Model Konvensional & Model *Hyperparameter Tuning*

Temuan penelitian ini menunjukkan pada sisi efektivitas penggunaan algoritma *Random Forest* dan *XGBoost* sebagai salah dua dari algoritma *based on Decision Tree* dalam melakukan klasifikasi gaya belajar siswa dengan kategori baik. Berdasarkan hasil penelitian terhadap 790 data yang telah melalui proses *preprocessing* dan *hyperparameter tuning*, algoritma *XGBoost* menunjukkan kinerja lebih baik dibandingkan *Random Forest*, dengan akurasi sebesar 84% dan *f1-score* 83%, sedangkan *Random Forest* mencapai akurasi 82% dan *f1-score* 80%. *XGBoost* lebih unggul dibandingkan *Random Forest* dalam menangani *dataset* tidak seimbang tanpa adanya seleksi *feature* yang dapat memiliki indikasi kemungkinan terjadinya bias, namun *Random Forest* lebih unggul dibandingkan *XGBoost* dengan waktu eksekusi lebih lama 5,20 detik untuk algoritma *XGBoost* dan 1,51 detik untuk algoritma *Random Forest* [15].

Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh referensi [16] yang menunjukkan bahwa algoritma *XGBoost* unggul dalam menangani *dataset* yang tidak seimbang dan menghasilkan akurasi lebih tinggi dalam klasifikasi apabila diperbandingkan dengan algoritma *Random Forest*. Namun, hasil penelitian ini bertentangan dengan penelitian yang dilakukan oleh referensi [4], [6], [7] karena dari penelitian ini dapat dilihat bahwa tidak semua algoritma berbasis *Decision Tree* cocok digunakan dalam studi kasus klasifikasi gaya belajar.

4. KESIMPULAN

Berdasarkan hasil penelitian klasifikasi gaya belajar siswa di SMP Negeri 2 Mataram menggunakan algoritma *XGBoost* dan *Random Forest* dengan 790 data setelah data *preprocessing*, ditemukan bahwa setelah dilakukan *hyperparameter tuning*, *XGBoost* memiliki kinerja lebih baik dibandingkan *Random Forest* dengan akurasi 84% dan *f1-score* 83%, sementara *Random Forest* mencapai akurasi 82% dan *f1-score* 80%. *XGBoost* lebih unggul dalam menangani *dataset* tidak seimbang dibandingkan *Random Forest*, namun memiliki waktu eksekusi lebih lama (5,20 detik vs. 1,51 detik). Teknik *SMOTE* terbukti efektif meningkatkan kinerja model pada *dataset* tidak seimbang, meskipun pada *Random Forest* tidak menunjukkan peningkatan signifikan setelah *SMOTE*. Selain itu, eksplorasi data menunjukkan bahwa mayoritas siswa memiliki gaya belajar visual dan kinestetik. Penelitian selanjutnya disarankan untuk menggunakan *dataset* yang lebih banyak serta melakukan uji signifikansi seperti uji T untuk mengevaluasi kinerja algoritma secara lebih mendalam.

DAFTAR PUSTAKA

- [1] L. Rahmi dan D. Yuswanti, "Meningkatan Motivasi Belajar Menggunakan Model Pembelajaran Picture and Picture pada Mata Pelajaran Ilmu Pengetahuan Alam," *JuDha_PGSD: Jurnal Dharma PGSD*, vol. 1, no. 2, pp. 73–85, 2021.
- [2] Halawa dan Mulyanti, "Faktor-Faktor Yang Mempengaruhi Peningkatan Kualitas Mutu Instansi Pendidikan Dan Pembelajaran," *Inspirasi Dunia: Jurnal Riset Pendidikan dan Bahasa*, vol. 2, no. 2, pp. 57–64, 2023. DOI: [10.58192/insdun.v2i2.757](https://doi.org/10.58192/insdun.v2i2.757).
- [3] I. A. Prasetyo *et al.*, "Metode Decision Tree Dalam Pemilihan Gaya Belajar Pada Siswa Sekolah Dasar," *SAINTEK : Jurnal Ilmiah Sains dan Teknologi Industri*, vol. 5, no. 1, pp. 21–29, 2021. DOI: [10.32524/saintek.v5i1.248](https://doi.org/10.32524/saintek.v5i1.248).
- [4] N. Awaliah, A. Herdiani, dan G. A. A. Wisudiawan, "Deteksi Gaya Belajar berdasarkan Learning Behavior pada Learning Management System menggunakan Naïve Bayes," *e-Proceedings of Engineering*, vol. 7, no. 1, pp. 2322–2342, 2020.
- [5] D. N. Sari, H. Oktavianto, dan I. Saifudin3, "Penerapan Data Mining Untuk Klasifikasi Gaya Belajar Siswa Menggunakan Algoritma C4.5 Application," *Jurnal Smart Teknologi*, vol. 3, no. 2, pp. 184–190, 2022.
- [6] A. K. Darmawan *et al.*, "Deteksi Gaya Belajar Siswa SMA pada Virtual Based Learning Environment(VBLE) dengan Decision Tree C4.5 dan Naive Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 5, pp. 532–544, 2023.
- [7] S. Ramadandi dan J. Jahring, "Student Learning Style Classification Using Naïve Bayes Classifier Method," *Jurnal Teknologi dan Informasi*, vol. 10, no. 2, pp. 170–179, 2020. DOI: [10.34010/jati.v10i2.3096](https://doi.org/10.34010/jati.v10i2.3096).
- [8] V. A. Haq, "Menguji Validitas Dan Reliabilitas Pada Mata Pelajaran Al Qur'an Hadits Menggunakan Korelasi Produk Momenspearman Brown," *An-Nawa : Jurnal Studi Islam*, vol. 4, no. 1, pp. 11–24, 2022. DOI: [10.37758/annawa.v4i1.419](https://doi.org/10.37758/annawa.v4i1.419).
- [9] A. Nugroho dan D. Harini, "Teknik Random Forest untuk Meningkatkan Akurasi Data Tidak Seimbang," *JSITIK: Jurnal Sistem Informasi dan Teknologi Informasi Komputer*, vol. 2, no. 2, pp. 128–140, 2024. DOI: [10.53624/jsitik.v2i2.379](https://doi.org/10.53624/jsitik.v2i2.379).
- [10] K. Davagdorj *et al.*, "Xgboost-based framework for smoking-induced noncommunicable disease prediction," *International Journal of Environmental Research and Public Health*, vol. 17, no. 18, pp. 1–22, 2020. DOI: [10.3390/ijerph17186513](https://doi.org/10.3390/ijerph17186513).

- [11] A. Haris Prayoga dan A. Voutama, “Pengembangan Aplikasi Bank Account Fraud Detection dengan Menggunakan Algoritma XGBoost,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 2916–2922, May 2024. DOI: [10.36040/jati.v8i3.9564](https://doi.org/10.36040/jati.v8i3.9564).
- [12] R. Harahap *et al.*, “Perbandingan Algoritma Random Forest Dan Xgboost Untuk Klasifikasi Penyakit Paru-Paru Berdasarkan Data Demografi Pasien,” *Jurnal Ilmiah Betrik*, vol. 15, no. 02, pp. 130–141, 2024.
- [13] E. Y. Boateng, J. Otoo, dan D. A. Abaye, “Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review,” *Journal of Data Analysis and Information Processing*, vol. 08, no. 04, pp. 341–357, 2020. DOI: [10.4236/jdaip.2020.84020](https://doi.org/10.4236/jdaip.2020.84020).
- [14] M. Haris, “Analysis of The Application Of Hyperparameter Tuning in Machine Learning to Increase The Accuracy of Sales-Level Prediction,” *Jurnal Informatika & Rekayasa Elektronika*, vol. 7, no. 1, pp. 98–106, 2024.
- [15] K. A. A. W. Wardana, A. Mizwar, dan A. Rahim, “Analisis Perbandingan Algoritma XGBoost Dan Algoritma Random Forest Untuk Klasifikasi Data Kesehatan Mental,” *LOGIC : Jurnal Ilmu Komputer dan Pendidikan*, vol. 2, no. 5, pp. 808–818, 2024.

[Halaman ini sengaja dikosongkan.]