

PREDIKSI GAYA BELAJAR MENGGUNAKAN ALGORITMA RANDOM FOREST DENGAN OPTIMASI HYPERPARAMETER TUNING BERBASIS WEB

¹⁾ Risma Bidayatul Hidayah, ²⁾ Khoiriya Latifah, dan ³⁾ Bambang Agus Herlambang

^{1,2,3)} Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas PGRI Semarang

^{1,2,3)}Jl. Dokter Cipto No 24 Semarang – Jawa Tengah - Indonesia

E-mail : rismahdyh162@gmail.com, khoiriyalatifah@upgris.ac.id, bambangherlambang@upgris.ac.id

ABSTRAK

Gaya belajar berperan penting dalam menunjang pencapaian akademik siswa, karena ketidaksesuaian antara metode pembelajaran dan preferensi belajar siswa dapat menurunkan motivasi belajar, menghambat pemahaman materi, serta berdampak negatif terhadap hasil akademik. Menyikapi permasalahan tersebut, sistem prediksi gaya belajar berbasis *web* dikembangkan dalam studi ini dengan menerapkan pendekatan algoritma *Random Forest*. Namun dalam masalah klasifikasi Random Forest sering sekali menghadapi performa prediksi yang kurang optimal. Penelitian ini bertujuan untuk mengoptimalkan algoritma *Random Forest* menggunakan optimasi hyperparameter tuning melalui *grid search* dan *random search* untuk meningkatkan performa klasifikasi prediksi gaya belajar siswa. Penelitian ini menggunakan dataset kuesioner sekunder yang terdiri dari 1.210 sampel data yang telah diklasifikasikan berdasarkan metode VAK (Visual, Auditori, dan Kinestik). Hasil pengujian menunjukkan bahwa model dengan optimasi *hyperparameter* dengan *random search* mampu meningkatkan performa akurasi sebesar 88% dan rata-rata nilai validasi silang sebesar 86%. Dibandingkan dengan model konfigurasi *default* dan *grid search* yang hanya menghasilkan akurasi sebesar 86%. Peningkatan ini juga ditunjukkan dengan pemrosesan penyeimbangan data menggunakan SMOTE dan reduksi data dengan PCA mampu memberikan kontribusi dalam meningkatkan akurasi model. Secara keseluruhan, penelitian ini membuktikan bahwa optimasi hyperparameter *grid search* dan *random search* mampu mengoptimalkan algoritma *Random Forest* dalam klasifikasi gaya belajar siswa secara akurat.

Kata Kunci: Gaya Belajar, Random Forest, Hyperparameter Tuning, VAK, Web.

ABSTRACT

Learning styles play an important role in supporting students' academic achievement, because a mismatch between learning methods and student learning preferences can reduce learning motivation, hinder material comprehension, and have a negative impact on academic results. To address this issue, a web-based learning style prediction system was developed in this study by applying the Random Forest approach. However, in classification problems, Random Forest often faces suboptimal classification results. This study seeks to enhance the effectiveness of the Random Forest algorithm by applying hyperparameter tuning techniques, specifically through grid search and random search methods, in order to improve the accuracy of classifying students' learning styles. This study uses a secondary questionnaire dataset consisting of 1,210 data samples classified based on the VAK model of learning preferences (Visual, Auditory, and Kinesthetic). The test results show that the model with hyperparameter optimization using random search can improve accuracy performance by 88% and the average cross-validation score by 86%. Compared to the default configuration model and grid search, which only achieved an accuracy of 86%, this improvement was also demonstrated by data balancing using SMOTE and data reduction using PCA, which contributed to enhancing model accuracy. Overall, this study demonstrates that grid search and random search hyperparameter optimization can optimize the Random Forest algorithm for accurately classifying students' learning styles.

Keyword: Learning Styles, Random Forest, Hyperparameter Tuning, VAK, Web.

PENDAHULUAN

Gaya belajar merupakan salah satu indikator penting dalam mencapai keberhasilkan prestasi akademik siswa, melalui gaya belajar yang sesuai, siswa dapat lebih mudah menyerap informasi, memahami materi pelajaran, serta menguasai materi dalam proses pembelajaran[1]. Tiap siswa mempunyai

metode yang berbeda dalam memahami, mengingat, serta mengolah informasi yang diterima selama proses pembelajaran. Perbedaan ini dipengaruhi oleh berbagai faktor, seperti pengalaman belajar, lingkungan belajar, serta karakteristik pribadi[2].

Dalam praktik di kelas, kerap ditemukan siswa memerlukan waktu lebih lama untuk

memahami suatu materi, atau bahkan merasa jemu saat mengikuti pembelajaran. Kondisi tersebut biasanya menunjukkan bahwa metode pengajaran yang digunakan kurang sesuai dengan preferensi belajar siswa. Akibatnya, motivasi belajar menurun, pemahaman terhadap materi menjadi terbatas, dan pencapaian hasil akademik tidak optimal[3]. Untuk itu, sangat penting bagi siswa untuk mengenali preferensi gaya belajar mereka secara mandiri. Dengan mengetahui gaya belajar yang paling sesuai, siswa dapat memilih strategi belajar yang lebih efektif.

Metode pembelajaran yang banyak digunakan dalam gaya belajar adalah VAK, yang mengelompokkan individu ke dalam kategori visual, auditori, dan kinestetik. Ketiga kategori tersebut diklasifikasikan menurut indera yang paling berperan atau dominan dalam memproses informasi. Grafik dan diagram dapat membantu siswa dengan gaya belajar visual memahami konsep yang rumit. Sementara pembelajar kinestetik berkembang pesat saat diberi kesempatan untuk melakukan gerakan fisik dan berlatih, pembelajar auditori lebih mudah memahami penjelasan lisan dan diskusi kelompok.

Dengan semakin berkembangnya teknologi, penggunaan machine learning mulai banyak diterapkan di dunia pendidikan, termasuk untuk membantu menganalisis gaya belajar siswa. *Machine learning* memungkinkan komputer belajar dari data yang ada, kemudian menghasilkan prediksi yang dapat digunakan untuk berbagai kebutuhan, termasuk di bidang pendidikan[4]. Di antara berbagai algoritma machine learning, *Random Forest* dikenal luas sebagai metode yang unggul karena performanya yang konsisten dan tingkat keakuratan yang tinggi. Algoritma ini membentuk beberapa pohon keputusan, kemudian mengombinasikan hasilnya untuk memperoleh prediksi akhir yang lebih optimal.

Selain itu, *Random Forest* dapat menangani data rumit dengan baik sekaligus meminimalkan risiko *overfitting*[5].

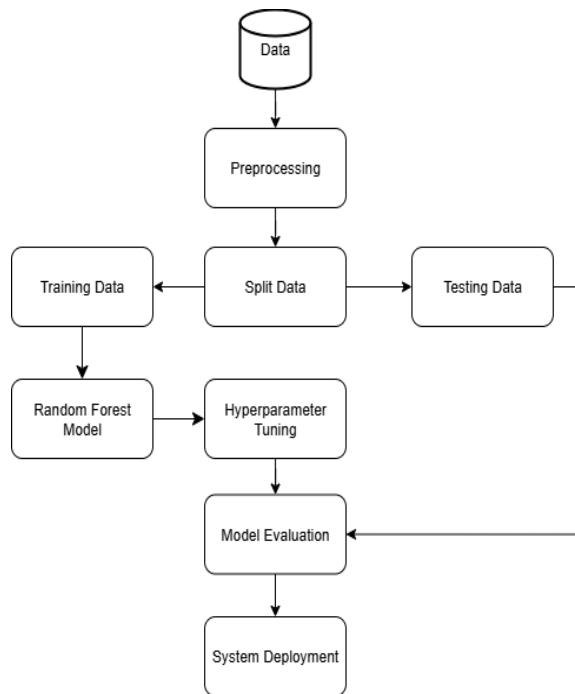
Berdasarkan hasil beberapa penelitian sebelumnya, algoritma *Random Forest* terbukti mampu memberikan performa yang unggul dalam mengidentifikasi preferensi gaya belajar. Misalnya penelitian oleh Sayed et al. berhasil mengembangkan model klasifikasi gaya belajar berdasarkan aktivitas siswa dalam sistem pembelajaran daring. Penelitian tersebut menunjukkan bahwa algoritma *Random Forest* dengan penerapan *hyperparameter tuning* secara manual menghasilkan akurasi tertinggi sebesar 98%, mengungguli algoritma lainnya seperti *Support Vector Machine* (96%), *Naive Bayes* (94%), *Regresi Logistik* (87%), dan *K-Nearest Neighbors* (73%)[6]. Selain itu, penelitian oleh Anistyasari et al. juga menunjukkan bahwa *Random Forest* mampu memprediksi tingkat kemahiran siswa SMK dalam pemrograman komputer dengan akurasi sebesar 88%[7]. Penelitian oleh Kisnu Darmawan dan Makruf memperlihatkan hasil yang baik dalam mendekripsi gaya belajar siswa di lingkungan pembelajaran *virtual*, dengan menggunakan algoritma *Naïve Bayes* menghasilkan akurasi sebesar 98%, mengungguli algoritma *Decision Tree C4.5* yang hanya mencapai 96%[8]. Sementara itu, Penelitian oleh Lokare dan Jadhav membandingkan 3 algoritma klasifikasi seperti *Decision Tree*, *Random Forest* dan *Ensemble learning*, hasil penelitian menunjukkan bahwa algoritma *Random Forest* lebih unggul dengan akurasi sebesar 87,5%[9].

Meskipun algoritma *Random Forest* telah banyak digunakan dan menunjukkan performa yang unggul dalam klasifikasi gaya belajar. Belum ditemukan penelitian yang membandingkan metode tuning seperti *grid search* dan *random search* dalam klasifikasi gaya belajar. Oleh karena itu, penelitian ini

bertujuan untuk mengisi kesenjangan tersebut dengan membandingkan dua metode tuning pada algoritma *Random Forest* serta membangun sistem prediksi gaya belajar berbasis web menggunakan kerangka kerja *flask*.

Melalui sistem ini, siswa diharapkan dapat mengenali preferensi gaya belajarnya secara mandiri. Serta informasi yang dihasilkan dari sistem dapat dimanfaatkan sebagai dasar dalam merancang strategi pembelajaran yang lebih tepat serta sesuai dengan karakteristik individu.

METODE



Gambar 1. Kerangka Kerja Penelitian

Data Collection

Penelitian ini menggunakan data sekunder berupa kuesioner dari Mendeley Data yang berjudul “*Student Learning Preferences*” dan tersedia dengan lisensi Creative Commons Attribution 4.0 International (CC BY 4.0)[10]. Dataset ini terdiri dari 1.210 entri data yang memuat atribut seperti jenis kelamin, usia, serta 14 pertanyaan dengan skor penilaian dalam rentang 1–5 untuk setiap pernyataan yang

dirancang berdasarkan model gaya belajar VAK. Data ini memiliki 17 atribut yang terdiri dari 15 atribut numerik dan 2 atribut kategorikal. Atribut *Learner* dikategorikan berdasarkan model VAK yaitu Visual (V), Auditori (A), Kinestik (K).

Beberapa atribut menggambarkan kecenderungan siswa dalam memahami materi melalui aktivitas visual, seperti membaca tulisan guru di papan tulis atau membaca petunjuk. Atribut lainnya menunjukkan preferensi auditori, seperti memahami pelajaran melalui penjelasan lisan dari guru atau orang lain. Selain itu, terdapat pula fitur yang merepresentasikan kinestetik, misalnya melalui praktik langsung di kelas atau bereksperimen.

Preprocessing

Pemrosesan data ini bertujuan agar data tetap dalam kondisi yang optimal dan tidak bias, sehingga dapat meningkatkan kinerja model yang akan dibangun[11]. Pada penelitian ini memiliki beberapa pemrosesan yang dilakukan antara lain:

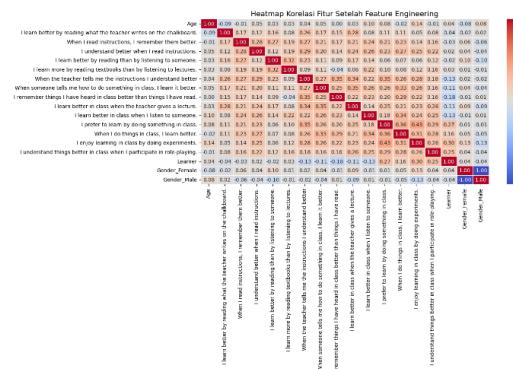
a. Pembersihan Data

Tahap awal dalam pemrosesan data dimulai dengan proses pembersihan terhadap data yang memiliki nilai hilang atau *missing values*. Data yang tidak lengkap dapat memengaruhi akurasi model.

b. Encoding

Dilakukan proses *encoding* untuk mengubah fitur kategorikal menjadi nilai numerik. Pada fitur *Gender* dikodekan menggunakan metode *one-hot encoding*, karena termasuk kategori data nominal yang memiliki nilai setara[12]. Sementara itu, fitur *learner* yang terdiri dari tiga kategori kelas Visual (V), Auditori (A), dan Kinestik (K), dikonversi menggunakan *ordinal encoding*. Metode ini digunakan agar label gaya belajar dapat ditampilkan dalam bentuk numerik berdasarkan urutan kategorinya[13].

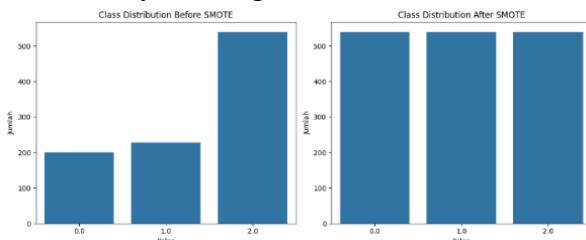
c. Analisis Korelasi Fitur



Gambar 2. Korelasi antar Fitur

Analisis korelasi antar fitur dilakukan untuk mengevaluasi karakteristik keterkaitan antara variabel independen dengan variabel target. Berdasarkan hasil analisis, nilai korelasi tertinggi sebesar 0,30 pada fitur “*I enjoy learning in class by doing experiments*”, sedangkan nilai korelasi terendah sebesar -0,02 terlihat pada fitur “*I learn better by reading than by listening to someone*”. Hal ini menunjukkan bahwa sebagian besar hubungan antar fitur bersifat lemah atau tidak terlalu signifikan secara statistic terhadap fitur target. Meskipun demikian, seluruh fitur tetap dipertahankan dalam proses klasifikasi. Keputusan ini didasarkan pada hasil uji coba sebelumnya yang menunjukkan bahwa penghapusan fitur dengan korelasi rendah justru menyebabkan penurunan akurasi model. Oleh karena itu, seluruh fitur tetap digunakan pada tahap pemrosesan untuk memperoleh hasil kinerja klasifikasi yang lebih optimal.

d. Penyeimbangan Data



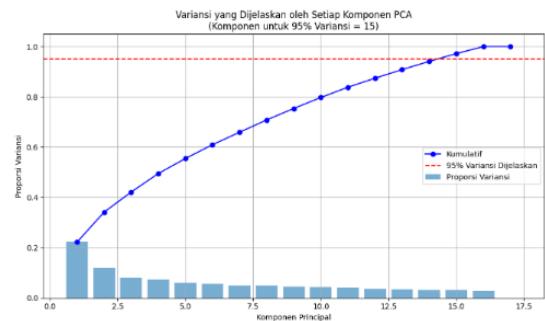
Gambar 3. Distribusi Kategori Kelas

Distribusi kelas pada variable *learner* menunjukkan adanya ketidakseimbangan data

seperti yang ditunjukkan pada Gambar 3. Kelas yang paling mendominasi adalah kelas kinestetik (2.0) dengan 540 data, diikuti oleh kelas auditori (1.0) sebanyak 228 data, dan kelas visual (0.0) sebanyak 200 data. Karena ketidakseimbangan ini, model mungkin lebih mengutamakan kelas dominan dan kurang memprediksi kelas minoritas. Oleh karena itu, kami menerapkan teknik SMOTE (*Synthetic Minority Oversampling Technique*) untuk menangani permasalahan di kelas minoritas dengan menambahkan data sintetis baru. Penggunaan pendekatan ini diterapkan untuk meratakan distribusi kelas, sehingga model mampu mengenali pola dari masing-masing kelas secara lebih efektif dan memberikan hasil klasifikasi yang lebih optimal[14]. Setelah proses SMOTE diterapkan, jumlah entri pada masing-masing kelas berhasil disamakan menjadi 540 data.

e. Reduksi Dimensi Data

Metode *Principal Component Analysis* (PCA) digunakan sebagai teknik reduksi dimensi yang mentransformasikan fitur-fitur ke ruang berdimensi lebih rendah secara linier. Dalam penelitian ini, pengaplikasian PCA menggunakan proporsi 95% ditunjukkan dengan visualisasi pada Gambar 4. Menjelaskan bahwa bahwa sebanyak 15 komponen dari total 17 fitur awal dipertahankan dalam proses ini. Pengurangan fitur pada tahap ini dapat membantu menyederhanakan model dan mengurangi risiko overfitting[15].



Gambar 4. Variansi Komponen PCA

Split Data

Pembagian data adalah proses membagi himpunan data menjadi dua yaitu 80:20, dimana 80% data dimanfaatkan untuk proses pelatihan dan 20% digunakan dalam tahap pengujian. Tahap ini sangat penting untuk menentukan apakah model dapat memprediksi data baru dan dapat mencegah terjadinya *overfitting* dan data *leakage*(kebocoran data) pada model [16].

Tabel 1. Hasil Distribusi Pembagian Data

Distribusi Pembagian Data	Jumlah
Data Latih	968
Data Uji	242

Random Forest Model

Pada penelitian ini, algoritma *Random Forest* digunakan sebagai model utama dalam mengklasifikasikan preferensi gaya belajar siswa. Serta terdapat tiga pendekatan pemodelan yang dibandingkan, yaitu model dengan konfigurasi default, serta dua model yang dioptimasi menggunakan metode *hyperparameter tuning* melalui *grid search* dan *random search*. Pendekatan ini bertujuan untuk mengevaluasi sejauh mana pengaruh optimasi parameter terhadap peningkatan performa model dalam proses klasifikasi gaya belajar.

Hyperparameter Tuning

Optimasi hyperparameter tuning dilakukan oleh peneliti untuk meningkatkan kinerja model. Dalam penelitian ini, pemodelan menggunakan algoritma *Random Forest* yang dioptimasi melalui metode *grid search* dan *random search*. *Grid search* merupakan teknik pencarian parameter optimal yang dilakukan secara menyeluruh dengan menguji semua kombinasi nilai parameter yang telah ditentukan sebelumnya dalam ruang pencarian[17]. Setiap kombinasi nilai parameter yang dievaluasi bertujuan untuk menentukan performa model terbaik.

Meskipun *grid search* dapat menghasilkan parameter yang optimal, metode ini memiliki keterbatasan dari waktu komputasi yang kurang efisiensi. Hal ini disebabkan karena seluruh kombinasi diuji satu per satu, sehingga membutuhkan waktu komputasi yang lama apabila dataset berukuran besar atau jumlah parameter yang diuji cukup banyak. Oleh karena itu, pendekatan selanjutnya menggunakan metode optimasi yang lebih efisien dalam hal waktu komputasi, yaitu *random search*. Berbeda dengan *grid search* yang mengevaluasi seluruh kombinasi parameter, *random search* hanya mengevaluasi sejumlah kombinasi acak.[18].

Penyesuaian parameter melalui pencarian *grid search* dan *random search* dilakukan dengan mengevaluasi berbagai kombinasi nilai *hyperparameter*. Rentang nilai yang digunakan ditentukan berdasarkan hasil dari beberapa kali uji coba pelatihan sebelumnya, dimana kombinasi nilai parameter pada Tabel 2 menunjukkan performa model yang lebih stabil dalam hal akurasi dibandingkan nilai-nilai yang telah diuji sebelumnya.

Tabel 2. Nilai Parameter

Parameter	Nilai yang Diuji
n_estimators	[100, 200, 300]
max_depth	[None, 10, 20, 30]
min_samples_split	[2, 5, 10]
min_samples_leaf	[1, 2, 4]
max_features	[sqrt, log2]

Model Evaluation

Proses evaluasi dalam tahap ini digunakan untuk menilai tingkat kinerja model dalam menghasilkan prediksi saat diberikan data uji yang belum digunakan dalam pelatihan[19]. Penilaian ini mencakup berbagai metrik evaluasi, antara lain tingkat akurasi, nilai presisi, sensitivitas (*recall*), *F1-score*, serta matriks kebingungan yang berperan dalam mengukur kinerja keseluruhan sistem.

Confusion matrix atau matriks kebingungan berfungsi untuk menggambarkan jumlah prediksi yang benar akurat maupun tidak. Matriks ini terdiri dari empat komponen penting, yaitu prediksi benar untuk kelas positif (TP), prediksi benar untuk kelas negatif (TN), kesalahan prediksi positif (FP), dan kesalahan prediksi negatif (FN).

Untuk memastikan evaluasi dilakukan secara adil dan tidak bias, diterapkan validasi silang menggunakan *stratified K-fold* sebanyak 5 lipatan[20]. Evaluasi ini membagi dataset menjadi k subset dengan proporsi kelas yang seimbang di setiap lipatan.

Selain metrik evaluasi, penelitian ini juga menggunakan metode interpretabilitas model menggunakan SHAP (*SHapley Additive exPlanations*). SHAP digunakan untuk menjelaskan kontribusi setiap fitur terhadap prediksi yang dihasilkan oleh model.

System Deployment

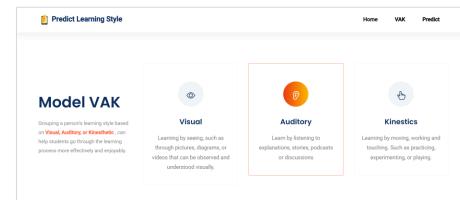
Dalam penerapan sistem prediksi gaya belajar berbasis *web*, penelitian ini menggunakan *framework flask* untuk mengaplikasikan model *Random Forest* dengan pendekatan model terbaik yang telah diperoleh melalui evaluasi model. *Flask* dipilih karena merupakan *framework python* yang populer dalam pengembangan aplikasi *web* yang ringan, efisien, dan mudah diimplementasikan dengan model *machine learning*[21]. Desain antarmuka pengguna pada sistem ini dirancang secara informatif dan mudah diakses oleh pengguna.



Gambar 5. Tampilan Halaman Beranda

Halaman beranda berfungsi sebagai tampilan awal yang memberikan informasi

umum terkait tujuan pengembangan sistem. Pada halaman ini dirancang untuk memberikan pemahaman kepada pengguna tentang pentingnya mengetahui gaya belajar dominan yang dimiliki.



Gambar 6. Tampilan Halaman Model VAK

Halaman model VAK menyajikan informasi mengenai klasifikasi preferensi gaya belajar berdasarkan 3 kategori, yakni visual, auditori, serta kinestik. Setiap kategori memiliki ciri khas dan perbedaan dari masing-masing gaya belajar. Informasi ini sangat penting sebagai dasar pemahaman pengguna sebelum melanjutkan ke tahap pengisian formulir prediksi.

Gambar 7. Tampilan Halaman Prediksi

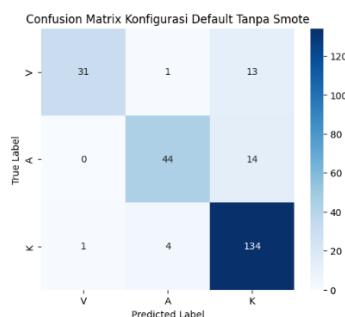
Halaman formulir prediksi merupakan fitur utama dalam sistem ini yang digunakan

untuk memprediksi preferensi gaya belajar siswa. Proses prediksi dilakukan menggunakan pemodelan algoritma *Random Forest*, yang menganalisis data input pengguna untuk memberikan hasil klasifikasi. Pengguna diminta untuk mengisi data pribadi seperti jenis kelamin dan usia, serta menjawab sejumlah pernyataan yang berkaitan dengan kebiasaan dalam aktivitas belajar. Setiap pernyataan disertai dengan skala penilaian dari 1 hingga 5 yang mencerminkan tingkat kesetujuan terhadap aktivitas tersebut. Setelah seluruh data diinputkan, hasil prediksi gaya belajar akan ditampilkan secara otomatis di bagian bawah formulir.

HASIL

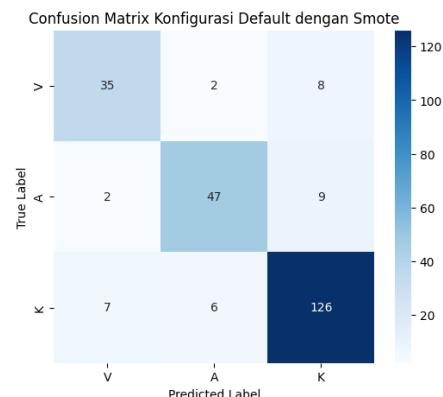
Penyeimbangan Data

Berdasarkan hasil pengujian, diketahui bahwa akurasi model sebelum dan sesudah penerapan SMOTE sama-sama mencapai 86%. Meskipun tidak terdapat peningkatan akurasi secara langsung pada data uji, perbedaan terlihat pada hasil evaluasi menggunakan validasi silang. Rata-rata nilai *cross validation* setelah diterapkan SMOTE meningkat menjadi 85%, sedangkan sebelum SMOTE hanya sebesar 83%. Peningkatan ini menunjukkan bahwa model menjadi lebih stabil dan generalisasi terhadap data baru menjadi lebih baik setelah distribusi data diseimbangkan. Selain itu, evaluasi *confusion matrix* menunjukkan bahwa model sebelum SMOTE menghasilkan prediksi yang kurang optimal pada kelas minoritas. Hal ini dapat dilihat pada visualisasi yang disajikan dalam Gambar 8 dan Gambar 9.



Gambar 8 Confusion Matrix Tanpa Smote

Model cenderung mengalami kesalahan dalam memprediksi kelas Visual sebanyak 13 data salah diklasifikasikan ke kelas Kinestetik. Sementara itu, kelas Auditori juga menunjukkan kesalahan prediksi sebesar 14 data ke kelas Kinestetik. Hal ini menunjukkan bahwa model cenderung bias terhadap kelas dengan jumlah data yang lebih dominan, yaitu kelas Kinestetik.



Gambar 9 Confusion Matrix Dengan Smote

Dari hasil setelah penerapan SMOTE, terlihat bahwa kesalahan klasifikasi antar kelas berkurang secara signifikan. Model menjadi lebih adil dalam mengenali kelas Visual dan Auditori yang sebelumnya tergolong minoritas. Hal ini mengindikasikan bahwa SMOTE mampu meningkatkan kemampuan model dalam mengenali pola dari masing-masing kelas secara seimbang, walaupun akurasi keseluruhan tidak berubah.

Hyperparameter Tuning

Hasil dari pencarian nilai hyperparameter menggunakan grid search dan random search disajikan pada Tabel 2. Kedua pendekatan tersebut menghasilkan kombinasi parameter terbaik berdasarkan ruang pencarian yang telah ditentukan sebelumnya.

Tabel 3. Hasil Tuning Hyperparameter Terbaik

Parameter	Grid Search	Random Search
n_estimators	300	100
max_depth	10	20

min_samples_split	5	5
min_samples_leaf	1	1
max_features	sqrt	log2

Evaluasi Model

Berdasarkan hasil evaluasi menunjukkan model *Random Forest* yang dikonfigurasi dengan optimasi *random search* menunjukkan kemampuan yang lebih konsisten dan akurat dalam mengklasifikasikan preferensi gaya belajar siswa. Proses evaluasi ini menggunakan sejumlah indikator penilaian tingkat akurasi, *precision*, *recall*, *F1-score*, confusion matrix, serta validasi silang *stratified K-fold* 5 lipatan. Hasil evaluasinya dapat dilihat pada Tabel 4.

Tabel 4. Perbandingan Evaluasi Model

Metrik	Default	Grid Search	Random Search
Akurasi	86%	86%	88%
Precision	84%	84%	87%
Recall	83%	83%	84%
F1-Score	84%	84%	86%
Cross Validation	83%	85%	86%

Model *Random Forest* dengan konfigurasi default dan grid search menghasilkan akurasi yang sama sebesar 86%, sementara model random search menunjukkan peningkatan akurasi menjadi 88%. Selain itu, precision, recall dan F1-score pada *random search* juga lebih tinggi, dibandingkan model konfigurasi *default* dan *grid search*.

Validasi silang menggunakan *stratified K-fold* sebanyak 5 lipatan menunjukkan bahwa *random search* memperoleh akurasi rata-rata sebesar 86%, lebih tinggi dibandingkan grid search (85%) dan konfigurasi default (83%). Hasil ini menunjukkan bahwa penerapan optimasi hyperparameter tuning menggunakan *random search* memberikan performa akurasi yang lebih unggul dan stabil dalam klasifikasi.

Hasil evaluasi menggunakan *confusion*

matrix pada kedua model yang telah dioptimasi *grid search* dan *random search* menunjukkan distribusi klasifikasi yang baik, dengan tingkat kesalahan klasifikasi yang relatif kecil. Hal ini menandakan bahwa model mampu mengenali pola dari masing-masing kelas dengan akurat. Kinerja tersebut ditingkatkan menggunakan penerapan penyeimbangan data dengan SMOTE dan proses reduksi dimensi melalui PCA. Kedua pendekatan ini berkontribusi dalam meningkatkan akurasi prediksi serta mengurangi risiko overfitting dan potensi bias yang mungkin terjadi selama proses pelatihan.

SHAP



Gambar 10 Visualisasi SHAP Value

Visualisasi pada Gambar 10 melalui SHAP summary plot untuk model dengan optimasi *random search* menunjukkan kontribusi masing-masing fitur terhadap hasil prediksi model. Fitur yang memiliki nilai SHAP value tertinggi terletak dibagian atas. Warna titik pada plot berwarna merah menandakan nilai fitur yang tinggi, sedangkan titik biru merepresentasikan nilai fitur yang rendah terhadap kelas tertentu. Dari visualisasi tersebut, dapat disimpulkan bahwa prediksi gaya belajar lebih banyak dipengaruhi oleh pernyataan yang berkaitan dengan fitur preferensi aktivitas belajar, dibandingkan

dengan karakteristik pribadi seperti usia dan jenis kelamin.

Analisis Keterbatasan Sistem

Salah satu keterbatasan dalam penelitian ini terletak pada penggunaan data kuesioner sekunder yang terbatas hanya 1.210 sampel. Jumlah data yang digunakan relatif kecil, yang dapat mempengaruhi kemampuan model dalam mengenali pola karakteristik siswa secara lebih luas. Selain itu, penelitian ini juga terbatas pada siswa yang memiliki gaya belajar dominan lebih dari satu. Hal ini menyebabkan model cenderung mengelompokkan siswa ke dalam satu kategori, tanpa mempertimbangkan kecenderungan multigaya belajar yang mungkin dimiliki oleh siswa.

KESIMPULAN

Berdasarkan hasil implementasi optimasi hyperparameter menggunakan *Grid Search* dan *Random Search* untuk meningkatkan performa prediksi, dapat disimpulkan bahwa sebelum dilakukan optimasi hyperparameter dengan penerapan teknik SMOTE dan PCA, model dengan konfigurasi default menghasilkan akurasi sebesar 86% dan rata-rata validasi silang sebesar 83%. Setelah dilakukan optimasi model dengan *grid search* memperoleh akurasi yang sama, namun nilai rata-rata validasi silang meningkat menjadi 85%. Sementara itu, *random search* berhasil meningkatkan akurasi menjadi 88% dan rata-rata validasi silang menjadi 86%. Hasil ini menunjukkan bahwa penerapan *hyperparameter tuning*, khususnya dengan metode *random search*, terbukti lebih efektif dalam meningkatkan performa model dalam memprediksi gaya belajar siswa. Model terbaik kemudian diimplementasikan ke dalam sistem prediksi gaya belajar berbasis web menggunakan *framework flask*. Implementasi ini bertujuan agar sistem dapat diakses oleh pengguna secara langsung untuk memperoleh hasil klasifikasi gaya belajar dengan lebih cepat

dan efisien.

DAFTAR PUSTAKA

- [1] S. Kumar, M. Agarwal, and N. Agarwal, “Defining and measuring academic performance of Hei students-a critical review,” *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 3091–3105, 2021.
- [2] S. Alshmrany, “Adaptive Learning Style Prediction in E-Learning Environment Using Levy Flight Distribution Based CNN Model,” *Cluster Comput.*, vol. 25, pp. 523–536, 2022, doi: 10.1007/s10586-021-03403-3.
- [3] Y. Simanungkalit, N. Natalisa, B. Sembiring, and S. Napitupulu, “Pentingnya Mengetahui Gaya Belajar Peserta Didik Di UPT SD Negeri 064012 Petisah Dalam Kegiatan Pembelajaran di Kelas,” vol. 1, no. 5, pp. 327–331, 2024.
- [4] Y. N. Aryo Sasi Kirono, “Perbandingan Algoritma Machine Learning Dalam Analisis Penyebab Penyakit Gagal Jantung,” *JEPIN*, vol. 10, no. 2, pp. 161–166, 2024, doi: 10.38204/tematik.v9i2.1070.
- [5] H. A. Salman, A. Kalakech, and A. Steiti, “Random Forest Algorithm Overview,” *Babylonian J. Mach. Learn.*, vol. 2024, pp. 69–79, 2024, doi: 10.58496/bjml/2024/007.
- [6] A. R. Sayed, M. H. Khafagy, M. Ali, and M. H. Mohamed, “Exploring the VAK model to predict student learning styles based on learning activity,” *Intell. Syst. with Appl.*, vol. 25, p. 200483, 2025.
- [7] Y. Anisyasari, S. C. Hidayati, R. Harimurti, and Ekoheriadi, “A Random Forest Algorithm for Predicting Computer Programming Skill Associated with Learning Styles,” *Inst. Electr. Electron. Eng. Inc.*, vol. 5, pp. 162–166, 2023, doi: 10.1109/ICVEE59738.2023.10348199.
- [8] A. Kisnu Darmawan and M. Makruf, “Deteksi Gaya Belajar Siswa SMA pada Virtual Based Learning Environment(VBLE) dengan Decision Tree C4.5 dan Naive Bayes,” *Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 5, pp. 532–544, 2023, [Online]. Available: <https://djournals.com/klik>
- [9] V. T. Lokare and P. M. Jadhav, “An AI-based learning style prediction model for personalized and effective learning,” *Think. Ski. Creat.*, vol. 51, p. 101421, 2024, doi: <https://doi.org/10.1016/j.tsc.2023.101421>.
- [10] A. M. Altamimi, M. Azzeh, and M. Albashayreh, “Predicting students’ learning styles using regression techniques,” *arXiv Prepr. arXiv2209.12691*, 2022.
- [11] A. Amato and V. Di Lecce, “Data preprocessing impact on machine learning algorithm performance,” *Open Comput. Sci.*, vol. 13, no. 1, 2023, doi: 10.1515/comp-2022-0278.
- [12] A. Y. Hussein, P. Falcarin, and A. T. Sadiq,

- “Enhancement performance of random forest algorithm via one hot encoding for IoT IDS,” *Period. Eng. Nat. Sci.*, vol. 9, no. 3, pp. 579–591, 2021, doi: 10.21533/pen.v9i3.2204.
- [13] S. Gnat, “Impact of categorical variables encoding on property mass valuation,” *Procedia Comput. Sci.*, vol. 192, pp. 3542–3550, 2021, doi: 10.1016/j.procs.2021.09.127.
- [14] T. Wongvorachan, S. He, and O. Bulut, “A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining,” *Information*, vol. 14, no. 1, 2023, doi: 10.3390/info14010054.
- [15] M. N. Fadhlurrahman and M. Y. Fitrah, “Optimizing Amazon Reviews Using Principal Component Analysis, Feature Selection On Random Forest Classifier,” *Media J. Gen. Comput. Sci.*, vol. 2, no. 1, pp. 12–17, 2025.
- [16] P. A. Prayesy, “STUDI PERBANDINGAN METODE SUPPORT VECTOR MACHINE, RANDOM FOREST, DAN CONVOLUTIONAL NEURAL NETWORKKUNTUK KLASIFIKASI PENYAKIT KULIT,” vol. 4, no. 1, pp. 70–76, 2025.
- [17] M. K. Suryadi, R. Herteno, S. W. Saputro, M. R. Faisal, and R. A. Nugroho, “A Comparative Study of Various Hyperparameter Tuning on Random Forest Classification with SMOTE and Feature Selection Using Genetic Algorithm in Software Defect Prediction,” *J. Electron. Electromed. Eng. Med. Informatics*, vol. 6, no. 2, pp. 137–147, 2024, doi: 10.35882/jeeemi.v6i2.375.
- [18] M. R. Hossain and D. Timmer, “Machine learning model optimization with hyper parameter tuning approach,” *Glob. J. Comput. Sci. Technol. D Neural Artif. Intell.*, vol. 21, no. 2, p. 31, 2021.
- [19] H. Wang, L. Zhang, K. Yin, H. Luo, and J. Li, “Landslide identification using machine learning,” *Geosci. Front.*, vol. 12, no. 1, pp. 351–364, 2021, doi: <https://doi.org/10.1016/j.gsf.2020.02.012>.
- [20] S. Prusty, S. Patnaik, and S. K. Dash, “SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer,” *Front. Nanotechnol.*, vol. 4, p. 972421, 2022.
- [21] M. Lathkar, *Building Web Apps with Python and Flask: Learn to Develop and Deploy Responsive RESTful Web Applications Using Flask Framework (English Edition)*. BPB Publications, 2021.